Accuracy of Genome Wide EBVs;
Using Three Small Breeds as Reference Population

Solomon Antwi Boison

Norwegian University of Life Sciences
Department of Animal and Aquacultural Science
Master Thesis 30 credits 2012

MASTER'S THESIS

In partial fulfilment of the requirements for the Degree of
Master's in Animal Sciences

Accuracy of Genome Wide EBVs; using three small breeds as reference population

Solomon Antwi Boison

Ås, Norway,
May, 2012

Supervisor
Prof. Theodorus H.E. Meuwissen

Department of Animal and Aquacultural Sciences
Norwegian University of Life Sciences
Ås, Norway

EUROPEAN MASTERS IN ANIMAL BREEDING AND GENETICS

Accuracy of Genome Wide EBVs; using three small breeds as reference population

Solomon Antwi Boison

Supervisor
Prof. Theodorus H.E. Meuwissen

Declaration

I hereby declare that this thesis entitled "Accuracy of Genome Wide EBVs; using three small breeds as reference population" is a bona fide record of research work done by me as a part of my Double Degree Program (*European Masters in Animal Breeding and Genetics* -EMABG) from the Norwegian University of Life Sciences (UMB), Ås, Norway and Wageningen University (WUR), The Netherlands.

It has not previously formed the basis for the award to me of any degree, diploma, fellowship or other similar title of any other university or society.

I hereby warrant that the thesis is based on work done by myself jointly with others; I have clearly stated exactly what was done by others and what I have contributed myself.

May 2012 ***Solomon Antwi Boison***

Ås, Norway

……**.....**

Dedicated to my family and friends

Preface

The submission of this master thesis marks the end of my 2 year MSc. program in Animal Breeding and Genetics (*European Masters in Animal Breeding and Genetics* -EMABG). The study was carried out at the Department of Animal and Aquaculture Studies, Norwegian University of Life Sciences.

This thesis was designed to implement genomic selection in small breeding populations using a multibreed reference population. This was because, key findings from genomic selection experiments are that, the reference population used must be very large to subsequently predict accurate genomic estimated breeding values (GEBV); the extent of linkage disequilibrium (LD) between markers and QTL should be high; among others. This meant that, in small populations, to achieve accurate predictions, breeds/populations needs to be combined or a breed with large number of animals could be used as the reference set to predict the breed/populations with the smaller number of animals. But results from predictions derived in one breed do not predict accurate GEBV when applied to other breeds. Thus researchers have suggested that, a multibreed reference population is a potential solution.

We estimated accuracy of GEBV for three Austrian breeds (Braunvieh, Grauvieh and Pinzgauer) with a single and multibreed breed reference population. We used both GBLUP (using genomic relationship matrix and then implementing it in ASReml) and Bayesian methods (Bayes-B and wgt.GBLUP) that increase the weight of certain important SNPs to estimated SNP effect in the prediction equation. Accuracy of GEBV was estimated as the correlation of the estimated GEBV and the EBV provided the Austrian breeding organization. Standard errors of the calculated accuracies were obtained using bootstrapping. Accuracies obtained in the single breed analysis are compared to those obtained from the multibreed analysis. Also the three method used are compared and discussed in the thesis.

Boison, S.A

May, 2012

Ås, Norway

Acknowledgements

To my supervisor Prof. Theodorus Meuwissen, I really appreciate your willingness to accept me as your thesis student, and for the tireless guidance, constructive and invaluable criticisms and comments that lead to the successful completion of this work.

A warm thank you goes to Prof. Johann Sölkner of BOKU (*University of Natural Resources and Applied Life Sciences, Vienna*) for providing me with the data as well as assist in the planning and implementation of this study. I also say thank you to Dr. Gabor Meszarös, a postdoctoral student of Prof. Soelkner for the initial quality control of the data and his keen interest in this thesis.

I am grateful for his time and energy in both travelling to Norway to provide and assist in this thesis. I am very grateful to Xijiang Yu for his assistance and encouragement from the start of this research. I would like also to extend my special thanks to Ruhul Agarwal, Dagnachew Binyam and Kahsay Nirea for your support, guidance and encouragement during this study.

I say thank you to all my friends (Ting Ding, among others) for their encouragement and support throughout this study.

I will also like to thank Keopon Foundation for giving me the funding and the opportunity to study in the Program; European Masters in Animal Breeding and Genetics (EMABG). Without their support none of these would happened.

Finally, to Him who made everything possible, THE ALMIGHTY GOD.

Abstract

Accuracy of genomic breeding values (GEBVs) is largely determined by the number of animals used in training and predicting marker effect. Thus in populations with limited number of animals, there are the need to combine populations or breeds to increase the reference population. The objective of this study was to investigate the accuracy of genomic selection using a single breed and multibreed reference population of the Austrian breeds Braunvieh, Grauvieh and Pinzgauer. Genomic relationship matrix (GBLUP) and Bayesian methods (Bayes-B and wgt.GBLUP) that increase the weight of certain important SNPs were used to predict marker effect. Accuracies were estimated using the 60 youngest bulls and calculated as the correlation between GEBV and published estimated breeding values (EBVs) for single breed and multibreed. Deregressed EBVs were used as phenotypes and a total of 10 traits were analysed. Accuracy of GEBV averaged across the 3 methods and the 10 traits for single breed ranged from 0.46 to 0.52. Two-way combined breed analysis gave an average accuracy of 0.46 and a three-way combined breed analysis was 0.45. Accuracies were not significantly different between methods; GBLUP, Bayes-B and wgt.GBLUP. Multibreed training set yielded maximum gain of about 17% in a both two and three -way analysis. However, on average combining 2 breeds increased accuracy by only 1.9% and a loss of 1.32% for a combination of 3 breeds. Combining breeds to increase the number of animals used in predicting marker effect and estimates GEBV for young bulls increased accuracy but this was not consistent across traits.

Keyword: *GEBV, Genomic selection, Multibreed, Accuracy, GBLUP, Bayes-B*

Norsk sammendrag

Nøyaktighet av genomisk avlsverdier (GEBVs) er i stor grad bestemmes av antall dyr som brukes i opplæring og forutsi markør effekt. Dermed i populasjoner med begrenset antall dyr, er det behovet for å kombinere populasjoner eller raser for å øke referansegruppen. Målet med denne studien var å undersøke nøyaktigheten av genomisk seleksjon ved hjelp av en enkelt rase og multibreed referanse befolkning av den østerrikske raser Braunvieh, Grauvieh og Pinzgauer. Genomisk forhold matrise (GBLUP) og Bayesianske metoder (Bayes-B og wgt.GBLUP) som øker vekten av enkelte viktige SNPs ble brukt til å forutsi markør effekt. Nøyaktigheten ble estimert ved hjelp av de 60 yngste oksene og beregnet som korrelasjonen mellom GEBV og publiserte estimerte avlsverdier (EBVs) for enkelt rase og multibreed. Deregressed EBVs ble brukt som fenotyper og totalt 10 trekk ble analysert. Nøyaktighet av GEBV gjennomsnitt over 3 metoder og de 10 trekkene for enkelt rase varierte 0,46 til 0,52. Toveis kombinert rase analyse ga en gjennomsnittlig nøyaktighet på 0,46 og en tre-veis kombinert rase analyse var 0,45. Nøyaktigheten var ikke signifikant forskjellig mellom metodene, GBLUP, Bayes-B og wgt.GBLUP. Multibreed opplæring sett gitt maksimal gevinst på ca 17% i en både to og tre-veis analyse. Men i gjennomsnitt kombinere 2 raser økt nøyaktighet med bare 1,9% og et tap på 1,32% for en kombinasjon av 3 raser. Kombinere raser for å øke antall dyr brukt i forutsi markør effekt og anslår GEBV for unge okser økt nøyaktighet, men dette var ikke konsekvent på tvers av egenskaper.

Søkeord: *GEBV, genomisk seleksjon, Multibreed, nøyaktighet, GBLUP, Bayes-B*

Table of Content

List of Tables

List of Figures

List of tables in Appendix

1 Introduction

1.1 Background

Animal and plant breeders have long been improving plant and livestock populations by estimating breeding values using phenotypic records and pedigree information. However, the recently developed genomic selection method (Meuwissen *et al*., 2001) have allowed us to use genome wide molecular markers (SNPs, haplotypes, etc) in estimating breeding values for selection candidates.

Genomic selection (**GS**) has been implemented in breeding programs all over the world. In GS, selection of parents for the next generation is based on Genome-wide estimated breeding values (**GEBV**). The implementation of **GS** derives a prediction equation for marker genotypes in a reference population (training dataset) that is genotyped and phenotyped. The estimated marker effect are assumed to be the populations estimates and thus the prediction equation is then used to predict **GEBV's** for selection candidates who have marker genotypes but do not have a trait record (Meuwissen *et al*., 2001; Hayes and Goddard 2010).

As was discussed in 2001, by Meuwissen *et al*., the feasibility of this approach depends on the cost of genotyping plants and animals for a large number of SNPs that are abundant in the genome of most species. The rapid sequencing technology after the year 2000 have discovered many SNPs that span the entire genome at certain marker intervals in human, cattle, pigs, chicken, fish, rice and wheat among others. Species can thus be genotyped with these SNP chips at a fairly low cost.

The fast adoption of **GS** by breeding companies was due to the large reduction in operational cost and the relatively high accuracy of the EBVs predicted for the selection candidates (Schaeffer, 2006). The accuracy of predicting EBV in most species using this method in simulation studies has been high. Meuwissen *et al*., (2001) in a simulation study showed that, accuracy of GEBV's can be 0.73 (BLUP; Best Linear Unbiased Prediction) and as high as 0.85 (Bayes-B). Calus *et al.* (2008) also reported accuracies of 0.83 (traits with h^2=0.5) and 0.66 (traits with h^2=0.10).

However, accuracies reported using real data have slightly been lower than those predicted in simulation studies. De Roos *et al*., (2011) reported high accuracies (average of 0.76) for highly heritable traits like milk yield, fat and protein yield and percentage compared to average accuracy of 0.63 for lowly heritable traits like fertility index, non return rate and longevity in a dairy cattle population from the Netherlands and Flanders. Others have

reported accuracies of 0.52 to 0.64 (Moser *et al*., 2010), 0.71 (Van Raden *et al*., 2009) and 0.83 (Van Raden et al, 2011).

To a large extent, the success of making genome wide predictions in genomic selection depends on the size of the reference population (**RP**), heritability of the traits and the extent of linkage disequilibrium (LD) between markers and QTL (Goddard and Hayes, 2009). Simulation studies have shown that, higher LD's are needed to achieve higher accuracies (Calus *et al*., 2008; Solberg *et al.,* 2008; Meuwissen *et al*., 2001). LD measured as r^2 of not less than 0.20 for adjacent SNP markers has been shown to give accuracies of about 0.8 to 0.9 (Calus *et al*., 2008; Solberg *et al.,* 2008; Meuwissen *et al*., 2001) although these accuracies are slightly lower for traits that are lowly heritable. The idea is that, the lower the extent of LD's in the population, more SNPs are required to make sure that, at least one of them is in complete LD with the QTL (Goddard, 2009). Linkage Disequilibrium is very much dependent on the effective population size (Hayes and Goddard, 2010) of the species under study. Species with small effective population sizes require fewer markers since SNPs will be in greater LD than those with higher effective population sizes (Meuwissen *et al*., 2001; Hayes and Goddard, 2010).

Accuracy of GS is observed to be higher for highly heritable traits than for lowly heritable traits in both simulation studies (Goddard 2008; Daetwyler *et al*., 2008; Calus *et al*., 2008) and studies using real data (De Roos *et al*., 2011; Moser *et al*., 2010; Luan *et al.,* 2009). The accuracy of GEBVs according to the formula of Daetwyler *et al*. (2008) is directly proportional to the heritability or reliability of the traits in the training dataset thus traits with higher heritability gives more accurate estimates of GEBVs than those with lower heritabilities.

The prediction methods used in GS suggest that, a large training dataset is needed to accurately estimate SNP effect and predict GEBVs (Hayes and Goddard, 2010; Meuwissen *et al*., 2001). Accuracy of GEBV increased by 17% through to 21% when the reference population were increased from 500 to 2200 (Meuwissen *et al*., 2001). Van Raden *et al*. (2009) reported 133% increase in accuracy of net merit in North American Holstein bulls when the training population were increased from 1151 to 3576. Luan *et al*., (2009) also reported slightly lower accuracies using 250 daughter yield deviation records in the training dataset than using 400 animals in Norwegian Red bulls.

Due to the large numbers needed in a **RP** for accurate prediction of marker effect in **GS**, implementation in breeds with smaller breeding population will require the aggregation of a

RP across breeds. However, (1) the effect of QTL alleles in one breed may not be the same for the other breed; (2) different QTLs may be segregating across populations; (3) the SNP-QTL LD might not be across breeds (Hayes *et al*., 2009; De Roos *et al*., 2009). Due to the above mentioned reasons, accuracy of GS where only crossbreed predictions (estimating GEBVs of one breed and **RP** from another breeds) were performed have not been very successful. Accuracy of GEBVs in studies of crossbreed prediction have been lower and sometimes negative compared to those of within pure breed prediction (Pryce *et al*., 2011; Hayes *et al*., 2009; Haris *et al*., 2008). Hayes *et al*. (2009) suggests that, we should aggregate breeds into a multibreed **RP** instead of crossbreed prediction, which might reduce some of the above mentioned reason that hamper across breed predictions.

The use of multi-breed **RP** has been studied for highly heritable production traits in dairy cattle production (milk yield, fat and protein yield and percentage) by Pryce *et al*. (2011) and Hayes *et al*. (2009) and in some simulated studies (De Roos *et al*., 2009). They all concluded that, a small accuracy increase for some traits can be achieved when the RP come from a multiple breeds. Accuracies of GEBV were up to 13% higher when the multibreed reference population was used than when a pure breed reference set was used (Hayes *et al*., 2009) although this percentage increase was not consistent across traits. Pryce *et al*. (2011) also reported that, predicting **GEBV's** for a breed that is not in the RP is increased with increasing number of breeds assuming that these breeds are related in the distant past. De Roos *et al*. (2009) in their simulation study noted that, an accurate prediction in this way depends on how divergent or evolutionarily distant the **RP** is from the breed to be predicted. Therefore sufficient marker density and LD between breeds should be high enough to achieve increase in accuracies when different populations or breeds are combined.

1.2 Objectives

The general objective of this paper is to investigate the accuracy of **GS** using a multi-breed **RP** of the Austrian breeds Braunvieh, Grauvieh and Pinzgauer for functional (lowly heritable) and production traits (highly heritable) with both GBLUP (using genomic relationship matrix) and Bayesian methods that increase the weight of certain important SNPs.

The specific aim of this paper is to:

i) Compare the accuracies of GS between using pure breed training dataset and multibreed training dataset.
ii) Compare accuracies using GBLUP and different Bayesian estimates of GEBV
iii) Compare the extent of Linkage Disequilibrium (LD) for marker pairs across breeds

*In this study, accuracy of GS for cross prediction (predicting GEBV from an entirely different population or breed when the **RP** does not contain part or that population or breed) were not investigated basically due to the expected lower and sometimes negative accuracies reported (Pryce et al., 2011 and Hayes et al., 2009)*

2 Material and Methods

2.1 Breed Description

2.1.1 Braunvieh

This breed is popularly known as the "Brown Swiss" in most part of the world. It known to have originated from Switzerland before it spread to other part of Europe (mostly Southern Germany, Italy, France, Slovenia and Austria), the Americas (USA and Canada), Australia, New Zealand and the other part of the world. Currently, the population worldwide counts 7 million head. Braunvieh are milky-type dual purpose cattle. The breed is known to have physical characteristics like unicoloured coat, ranging from brown to grey and beige. Males show darker coats than the female. Other characters included are the dark claws, the black muzzle with a bright edge and bright hairs inside the ears. The horns are bright with dark tips. It has a medium wedge-like body shape and with no emphasis to increase it body size in most breeding goals around the world.

In Austria, it is found within the western and central part with an estimated population size of 162,000 (5,444 herds with over 55,078 cow registered in herd book).

Picture of Braunvieh dam adapted from ZAR (2009) (http://www.zar.at/)

It has been bred for high performance in milk production, functional and reproductive traits. Breeding goal has been to develop total merit index of 48% milk, 5% beef and 47% reproduction and functional traits. Some important performance traits include: Age at first calving (31.4 months), calving interval of approximately 400 days, productive life (3.9 years), and milk production during 305 days (6,856 kg milk yield with 4.11 % fat and 3.41 % protein)(www.rinderzucht-austria.at).

2.1.2 Grauvieh (Tiroler Grauvieh)

The breed is also known as the "Tyrol Grey" is certain part or the world. It is believed to have originated from Austria and lived mostly in the Alpine regions. Today the breed is also found in Canada, Italy, Bavaria, and Switzerland among others. The breed is regarded as a rare and endangered species of livestock today, and is therefore part of the Austrian Government gene protection program. Its physical characteristics include: a uni-coloured coat of silver to iron-grey, sometimes brownish-grey, with certain lighter and darker spots. The skin is black. Special characteristics are a red shock of hair, black horn tips and black, hard hooves.

The multi-purpose breed has been mainly breed for milk and beef.

The breed is still found largely in the Alpine regions of Austria with a population of 18,000 (3.809 registered cows). The Breeding goal has been to develop total merit index of 30% milk, 20% beef and 50% fitness traits.

Production performance for milk yield and it component has been 4,837 kg with 3.93% fat and 3.25% protein. It is also highly breed for meat and thus have high quality beef. It calving interval is about 33.8 month and has a productive lifespan of 4.7 years (www.rinderzucht-austria.at).

Picture of Grauvieh dam adapted from ZAR (2009) (http://www.zar.at/)

2.1.3 Pinzgauer

The breed takes its name from the *Pinzgau* district of Salzburg, Austria. It was first developed in the sixth century from Bavarian cattle. It was exported to other part of Europe especially including Romania, Czech Republic, Austria and Yogoslavia. It has then spread to USA, Canada, South Africa and other countries.

In Austria, the breed is also found in the mountainous (alpine) areas and has a population of 47,000 with 7,680 registered cows. Pinzgauer are easily recognisable by their deep chestnut colour with white markings on the back, underside, udder and tail.

The breeding goal is a total merit index of 36% milk, 14% beef and 50% fitness traits.

Picture of Pinzgauer dam adapted from ZAR (2009) (http://www.zar.at/)

Some important performance traits include: Age at first calving (34.1 months), productive life (3.7 years), and milk production during 305 days (5,398 kg milk yields with 3.86 % fat and 3.24% protein) (www.rinderzucht-austria.at).

2.2 Phenotypic data

The phenotypic data (*provided by Zuchtdata EDV- Dienstleistungen GmbH) http://www.zar.at/*) used in estimating SNP effects and predicting GEBV for the ten (10) traits in this study were de-regressed estimated breeding values (DrEBV) of bulls. The method of Garrick *et al*. (2009) was used for the de-regressing the original estimated breeding values (EBV) that were based on routine genetic evaluation of on average 8-10 year old bulls. Parent average effects and the differences in progeny records are removed thereby accounting for the heterogeneous variances or different reliabilities of the EBV. The following ten (10) traits were analysed: milking speed, protein and fat yield and percentage, milk yield, fertility, longevity, persistency and somatic cell count (*see table 1*). The traits are briefly described, however details are found at the *Zuchtdata EDV- Dienstleistungen GmbH* website (http://www.zar.at/).

2.2.1 Production Traits

***Milk yield*:** The EBV for milk yield was estimated with a test day animal BLUP model. Milk yield from a maximum of four lactations were used. The total amount of milk produced per day as a sum of morning and evening lactation and accumulated for the entire lactation.

***Milk composition (Protein and Fat yield and percentage)*:** Daily milk records are analysed for these protein and fat percentage and yield. EBVs are estimated from the records of the 3 lactations using again an animal test day BLUP model.

2.2.2 Reproduction and Functional Traits

***Milking speed*:** The average milking speeds per cow of only the first lactation and milking ability as visual scores from the famer are combined to as phenotype for estimating EBVs. EBVs are calculated together (multivariate) with the somatic cell count taking the genetic correlations into account by an animal BLUP model.

***Fertility*:** EBVs are calculated as the non-return-rate 56 days of heifers and cows, time to first insemination and time from first to last insemination (heifers and cows) using a BLUP animal model. A female fertility index is calculated from non-return rate and time from first to last insemination, which refers to the fertility of the daughters of a bull.

***Longevity*:** An individual animal's productive life EBV is predicted using survival analysis which also accounts for censored animals. It is based on a yield-independent productive (milk yield) life as it serves as a yardstick for evaluating vitality and fertility.

***Persistency*:** Persistency is defined as the decrease or increase from lactation day 60 to day 300. EBVs are then estimated using a test day animal BLUP model.

***Somatic cell count (SCC)*:** The concentration of somatic cells per millilitre for the first three lactations collected during milk recording is used for EBV estimation using a test day BLUP animal model. SCC is considered an auxiliary characteristic for predisposition and resistance to mastitis.

The number of genotyped bulls with DrEBV (discussed under result; *table 3a, 3b and 3c*) varied because bulls with reliabilities of EBVs < 0.30 (r = 0.55) were excluded. This ensures that accurate phenotypes are used to estimate GEBVs accurately especially when the number of genotyped bulls was small.

2.3 Pedigree structure

A total pedigree database of 6057 animals from Braunvieh, 1691 from Grauvieh and 3107 animal from Pinzgauer all in about 8 generations including the genotyped bulls were used in this study. There were approximately 1740, 524, 1136 sires and 3862, 998 and 1851 dams for Braunvieh, Grauvieh and Pinzgauer respectively.

2.4 Genotypic data

There were 202 Braunvieh, 100 Grauvieh and 101 pinzgauer bulls genotyped for 54,001 SNPs markers using the Illumina bovine SNP50 beadchip. In addition, 322 Braunvieh, 120 Grauvieh and 121 pinzgauer bulls were genotype for 777,000 SNPs using the Illumina BovineHD beadchip. The same sets of SNPs of the 54001 markers were extracted from the 777K SNP chips to make a total of 524 Braunvieh, 221 Grauvieh and 221 Pinzgauer bulls.

Initial pedigree checks using the SNP information were done to remove sons of sires with incorrect pedigree (sons with different homozygous alleles than what the sire is carrying;

sons are removed when 1000 alleles are discordant). Genotype quality checking was performed within breed using PLINK (Purcell *et al.*, 2007). Maximal identical – by – state between bulls: 0.999 (in order to get rid of monozygotic twins or double genotyped sires with false ID). SNPs were selected on; minor allele frequency (MAF) > 2%, call rate > 95%, missing genotypes < 1%, Hardy Weinberg Equilibrium (HWE) p-value > 10^{-3}, SNPs mapped to the X chromosome were removed (Hayes *et al*., 2009, De Roos *et al*., 2009). Animals with GenCall score (Illumina Inc., 2008) of less than 0.60 were discarded. The final extracted SNP's segregating across all the three (3) breeds were 35,319.

Table 1: Total number of bulls after both phenotypic (DrEBV) and genotypic SNP editing for the three (3) breeds; Braunvieh, Grauvieh and Pinzgauer

Traits	**Breed**		
	Braunvieh	**Grauvieh**	**Pinzgauer**
Milking Speed	453	181	170
Protein *Kg*	450	190	159
Protein Percent	450	190	159
Fat *Kg*	450	190	159
Fat Percent	450	190	159
Fertility Maternal	387	93	151
Milk Kg	450	190	159
Longevity	423	121	155
Persistency	455	213	196
Somatic Cell Count	455	196	189

2.5 Reference and cross validation dataset

Marker effects were estimated from a reference dataset of bulls depending on their birth years and the traits. The validation dataset consisted of the 60 youngest bulls (forward prediction) with phenotypes for that particular traits evaluated except for Grauvieh where the 30 youngest bulls were used for the traits fertility Maternal and Longevity. The distribution of bulls across birth years for Braunvieh, Grauvieh and Pinzgauer bulls is shown in Figure 1. Two and three way combinations of breeds were used for the multibreed **GS** (see Table 2). The cross validation dataset for the multibreed analysis were the same 60 young bulls used for the pure breed analysis.

Table 2: Overview of the pure bred, two and three way multibreed analysis for both training and cross validation dataset

Analysis	Reference	Validation
Pure breed	REF-BV	VAL-BV
	REF-GV	VAL-GV
	REF-PI	VAL-PI
Two way Multibreed	All GV + REF-BV	VAL-BV
	All BV + REF-GV	VAL-GV
	All PI + REF-GV	VAL-GV
	All GV + REF-PI	VAL-PI
	All PI + REF-BV	VAL-BV
	All BV + REF-PI	VAL-PI
Three way Multibreed	All GV + All PI + REF-BV	VAL-BV
	All BV + All PI + REF-GV	VAL-GV
	All BV + All GV + REF-PI	VAL-PI

REF – Reference dataset *VAL – Cross Validation dataset*

BV – Braunvieh; GV – Grauvieh and PI - Pinzgauer

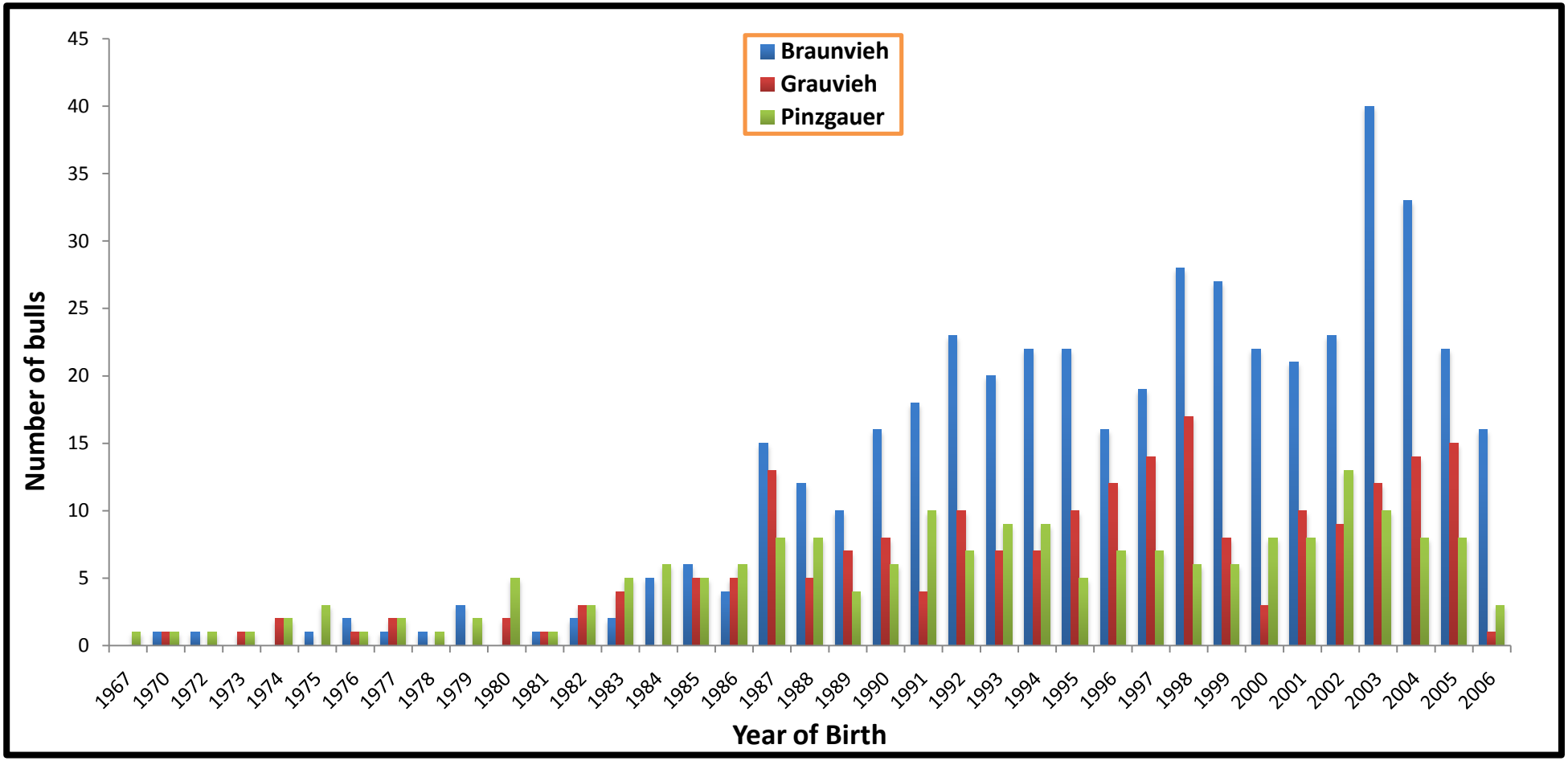

Figure 1: Number of bulls across birth years for the forward prediction in Braunvieh, Grauvieh and Pinzgauer breeds. Validation dataset are the youngest 60 bulls depending on the traits.

2.6 Methods

2.6.1 Statistical analysis

DrEBV for the multibreed analysis were adjusted for fixed effect (breed) before been used as a response variable in the subsequent GBLUP and Bayesian models. We assume

$$y = \mu + Xb + e$$

equation 1.0

y = vector of DrEBV for the traits
μ = is the overall mean
X = is a design matrix relating records to breed
b = is a vector of breed effect
e = vector of random residual errors *N (0, 1)*

In the pure breed analysis, the uncorrected DrEBV phenotypes were used as the response variables. This is because, the algorithm for the analysis of the Bayesian methodologies were developed only to include the mean of phenotypes pre-corrected for their fixed effect.

2.6.1.1 GBLUP

GEBV will be estimated by fitting a polygenic effect assuming that every marker has a constant variance (GBLUP) (Meuwissen *et al*., 2001) *i.e.* assuming that each marker explains an equal proportion of the total genetic variance (σ_g^2). Genomic relationship matrix (G) based on SNP marker genotypes instead of the conventional additive genetic relationship matrix (PBLUP) from pedigree information were used in estimating the GEBV. The GBLUP model assumed was:

$$y = 1_n\mu + Zg + e$$

equation 2.0

y = corrected DrEBV
1_n = vector of 1s
μ = overall mean
Z = design matrix allocating records to breeding values
g = vector of random additive genetic effect using the genomic relationship matrix (G)

coming from $N(0, G\sigma_g^2)$

e = vector of random residual errors $N(0, I\sigma_e^2)$

The genomic relationship matrix (G) is calculated by using SNP marker genotype as described by Yang et al. (2010).

$$G = {XX}/{m}$$

X= matrix of standardised SNP genotypes Xij

m=number of SNPs

X_{ij} denotes the standardised SNP genotypes of animal *i* for SNP *j*

For genotypes 0, 1 and 2

$$X_{ij} : {(0 - 2p_j)}/{\sqrt{H}}; {(1-2p_j)}/{\sqrt{H}}; {(2 - 2p_j)}/{\sqrt{H}}$$

The values of the three SNP genotypes are originally 0, 1 and 2 respectively, but are standardised to a mean of zero and a standard deviation of 1 (by subtracting the mean (2pj) and dividing by the standard deviation √H. $Heterozygosity\ (H) = 2p_j(1 - p_j)$.

Thus the G_{ik} between two animals *i* and *k* were calculated

$$G_{ij} = corr(X_{ij} : X_{kj}) = cov(X_{ij} : X_{kj})$$

The calculated genomic relationship matrix is implemented in the equations to calculate GBLUP breeding values using ASReml v3 software package (Gilmour *et al.*, 2009).

2.6.1.2 Bayesian Methodologies (Bayes-B and Weighted Mixture model)

Bayesian methodology will be used to vary the variance assumption employed across loci instead of a constant variance assumption in GBLUP (Meuwissen *et al*., 2001).

2.6.1.1.1 Bayes-B

This model assumed that some markers had a big effect of variance σ^2 with probability of π, whilst the remaining markers have a small effect with small variance with a probability of (1-π), the variance of which will be assumed to be equal and will be estimated in the model from the data (Luan *et al*., 2009), instead of assuming that these markers and with a variance of 0, had virtually no effect at all (Meuwissen *et al*., 2001). Assuming that, the variance of those

SNPs with no or little effect was not equal to zero but small allows the Gibbs sampler to work slightly faster than usual and allows for many small genes spread across the genome. The prior probability π is unknown and therefore different values are tested till we arrived at the one that gives the largest accuracies of GEBV. Interestingly, varying these prior distributions of the marker effect showed little or no increase in accuracy for most of the traits. The model used was:

$$y = \mu + \sum_{j=1}^{N_m} X_j a_j + e$$

equation 3.0

y = vector of phenotypes

N_m = number of markers fitted

Xj = vector denoting the genotype of the individuals for marker j

a_j = effect of the marker

e = vector of random residual errors $N(0, I\sigma_e^2)$

In detail X_j is calculated from individuals with genotypes $X_{ij} = 0$ if individual i is homozygous for the first allele at locus j. $X_{ij} = 1/\sqrt{H_j}$ if heterozygous. $X_{ij} = 2/\sqrt{H_j}$ if individual i is homozygous for the second allele at locus j, and $X_{ij} = 2q_i/\sqrt{H_j}$ if the marker genotype is missing, where q_j is the frequency of the second marker allele and H_j is the marker heterozygosity. The division by $\sqrt{H_j}$ standardizes the variance of the marker genotype data to 1 (Luan *et al*., 2009).

After obtaining the marker effect, Genome wide estimated breeding Values (GEBV) will be predicted as

$$GEBV = \mu + \sum_{j=1}^{N_m} X_{ij} \hat{a}_j$$

e*quation 3.1*

Where μ is the overall mean; X_{ij} is the marker genotype of individual i for marker j; $\hat{a}_i$ is the estimated effect of marker j.

For each trait, the Gibbs sampler of the Markov Chain Monte Carlo (MCMC) method was run on a single chain of 40,000 iterations and 10,000 burn-ins based on a convergence test with the traits milk yield and SCC for using different chain length and burn-ins to estimate parameters.

2.6.1.1.2 Weighted GBLUP (wgt.GBLUP)

The model was the same as in equation 3.0 but the SNP variance assumption changed to $V(a_j) = b_j^2$, where b_j is the solution of the *jth* SNP in the GBLUP model. Thus b_j^2 is seen here as an estimate of the variance due to the *jth* SNP, except that the prediction error variance of b_j is ignored when estimating the variance of the SNP, which implies that the variance is underestimated. This underestimation corrected by scaling up the $V(a_j)$ such that the sum of the overall SNPs equals to the total genetic variance. In a sense this model is Bayes-A model where the variance due to each SNP is estimated. wgt.GBLUP. The model implies that, SNPs with higher GBLUP-SNP effect, b_j are regressed back less than those with lower SNP effect.

Another variant of this wgt.GBLUP which used the marker effect estimated with a multibreed (all 3 breeds) training set as weight for the maker effect estimated in the purebred analysis did not improve accuracy above 1% (these are averaged across breeds and traits; results for both single breed and multibreed analysis are presented in table 16 of Appendix 1). Therefore the earlier mentioned wgt.GBLUP method was used instead.

All the Bayesian methods were programmed in Fortran90 and compiled for Linux and were developed by Theo H.E. Meuwissen (Norwegian University of Life Science, Aas, Norway). These programs (***BAYESGG*** ~ Bayes-B and ***BAYESP*** ~ wgt.GBLUP) were then run on an Intel Core ^{TM} Duo CPU E8500.

2.7 Evaluation of Accuracy of GEBV

Accuracy of GEBV were estimated as the correlation between GEBV and EBV; $r = cor(GEBV, EBV)$. EBVs were obtained from the *Zuchtdata EDV- Dienstleistungen GmbH,* Austria. Therefore, this meant that the theoretical maximum for these accuracies will be the average accuracy of the EBVs obtained from *Zuchtdata EDV- Dienstleistungen*

GmbH, Austria. Note that, unless otherwise stated accuracies are calculated as the correlation between the estimated GEBVs and the EBV obtained from *Zuchtdata EDV- Dienstleistungen GmbH,* Austria without dividing this estimate by the theoretical maximum. As stated earlier, forward prediction (the youngest bulls are used in validation dataset) procedure was used in evaluating the accuracy of GEBV. Since the numbers of bulls in this study were small, the bootstrapping procedure (sampling with replacement) was used to calculate the standard error of the correlation between the GEBV and the EBV.

The estimated GEBV were bootstrapped 10,000 times (this value appeared to give stable results) and the bootstrap GEBVs are correlated to the EBVs. The standard error is calculated from the 10,000 estimated accuracies. This procedure gives us a fair estimate of the degree of dispersion of the estimated correlation. Although other cross validation procedure like random splitting procedures could have been employed; this study chose to use forward prediction which is more relevant to breeding companies. This is because; marker effects will be estimated from older animals and the target selection candidate for the implementation of **GS** might include younger animals or their offspring. Bootstrapping was done by the ***R statistical software package*** (R, Development Core Team, 2011).

2.8 Regression of EBV on GEBV

The regression coefficient was used to measure the predicted bias by regressing the estimated breeding values obtained from *Zuchtdata EDV- Dienstleistungen GmbH* in Austria on the GEBV. An estimated regression coefficient of 1 indicates an unbiased estimator of the true breeding value *i.e.* 1 unit higher predicted GEBV corresponds to 1 unit EBV (De Roos *et al.*, 2011).

2.9 Correlation between GS methods

Pearson correlation coefficient of GEBV estimated with each method for a particular trait was used as a measure of the relationship between prediction methods.

2.10 Extent of Linkage Disequilibrium (LD)

Extent of LD was calculated to help explain the prediction pattern among and between the breeds. As reported by De Roos *et al*. (2008) and Calus *et al*. (2008), accuracy of GS is affected by the LD in that population, since GS relies on markers that are in high LD to the QTL. The persistence of LD was calculated for syntenic (adjacent) marker pairs using genome-wide SNPs (De Roos *et al*., 2008; Hill and Robertson 1968). The r and R^2 representing the measures of LD for 2 syntenic markers will be calculated using PLINK (Purcell *et al*., 2007) as:

$$r = \frac{P_{A1B1}P_{A2B2} - P_{A1B2}P_{A2B1}}{\sqrt{P_{A1}P_{A2}P_{B1}P_{B2}}}$$

To evaluate further the persistence of LD phase across breeds, the correlation of r between breeds were calculated for the mean genomic distance (67 kb) reported for the Illumina BovineHD beadchip.

3 RESULTS

3.1 Summary of phenotypic data

The mean value of DrEBV, reliabilities and number of records for each trait for the bulls in the training dataset of Braunvieh, Grauvieh and Pinzgauer breeds are presented in Table 3a, 3b & 3c respectively. Reliabilities for all DrEBV were higher for Braunvieh breed followed by Pinzgauer and then the Grauvieh breed. Also on average, the number of bulls in the training dataset was highest for Braunvieh, and lowest for Pinzgauer.

Braunvieh breed: The number of bulls in the training dataset ranged from 327 to 395 (Table 3a). The average reliability for all traits was 0.86 and a standard deviation of 0.09. The DrEBV were slightly more reliable (r^2 > 0.92) and of less variation (SD < 0.06) for production traits (milk yield, Protein and Fat Percentage and yield) than for functional and reproductive traits (milking speed, fertility, and Somatic cell count; except persistency).

Table 2a: Braunvieh breed: Traits for the study: number of bulls, mean and standard deviation (SD) of DrEBV and mean reliability (r^2) of DrEBV of bulls in training dataset

Traits	**Number of bulls**	**Mean DrEBV**	**SD of DrEBV**	**Mean r^2**	**SD of r^2**
Milking Speed	393	98.33	10.45	0.83	0.14
Protein *Kg*	390	-13.93	17.30	0.92	0.06
Protein Percent	390	-0.01	0.12	0.92	0.06
Fat *Kg*	390	-13.30	19.71	0.92	0.06
Fat Percent	390	0.03	0.17	0.92	0.06
Fertility Maternal	327	102.9	11.78	0.67	0.19
Milk Kg	390	-356.4	539.7	0.92	0.06
Longevity	363	102.6	15.58	0.75	0.15
Persistency	395	98.09	11.90	0.92	0.06
Somatic Cell Count	395	96.94	13.09	0.87	0.09

Grauvieh breed: the reliabilities of DrEBV for all traits range from 0.57 through to 0.81 (see table 3b) with higher reliabilities associated with production traits and lower reliabilities with functional and reproductive traits. Most of the bulls did not have records on fertility and longevity or reliability of EBV these two traits were < 0.30. This left us with only 93 and 121 bulls to be used for GS, thus the 30 young bulls were used as validation bulls leaving 63 and 91 as training bulls for fertility and longevity.

Table 2b: Grauvieh breed: Traits for the study: number of bulls, mean and standard deviation (SD) of DrEBV and mean reliability (r^2) of DrEBV of bulls in training dataset

Traits	Number of bulls	Mean DrEBV	SD of DrEBV	Mean r^2	SD of r^2
Milking Speed	121	99.01	10.12	0.61	0.17
Protein *Kg*	130	-9.62	13.80	0.81	0.11
Protein Percent	130	-0.03	0.17	0.81	0.11
Fat *Kg*	130	-16.33	28.80	0.81	0.11
Fat Percent	130	-0.14	0.23	0.81	0.11
Fertility Maternal	63	95.10	11.34	0.57	0.18
Milk Kg	130	-241.0	444.7	0.81	0.11
Longevity	91	91.92	121.6	0.59	0.17
Persistency	153	104.7	14.65	0.80	0.12
Somatic Cell Count	136	94.49	12.16	0.68	0.17

Pinzgauer breed: The number of bulls with phenotype varied for different traits. Reliabilities were above 0.65 and ranged from 0.66 to 0.90. The average reliability for all traits was 0.84 ± 0.12. As was the case for Braunvieh and Grauvieh, reliabilities were higher on average for production traits than for functional and reproductive traits.

Table 2c: Pinzgauer breed: Traits for the study: number of bulls, mean and standard deviation (SD) of DrEBV and mean reliability (r^2) of DrEBV of bulls in training dataset

Traits	Number of bulls	Mean DrEBV	SD of DrEBV	Mean r^2	SD of r^2
Milking Speed	110	69.64	10.51	0.77	0.18
Protein *Kg*	99	-10.68	18.74	0.90	0.09
Protein Percent	99	0.04	0.17	0.90	0.09
Fat *Kg*	99	-11.21	25.44	0.90	0.09
Fat Percent	99	0.07	0.24	0.90	0.09
Fertility Maternal	91	101.1	9.85	0.66	0.18
Milk Kg	99	-372.7	631.8	0.90	0.09
Longevity	95	93.31	13.67	0.75	0.17
Persistency	136	106.4	12.96	0.89	0.11
Somatic Cell Count	129	98.46	13.51	0.80	0.17

3.2 Accuracy[1] of GEBV prediction

3.2.1 Accuracy of GEBV in purebred-GS

Table 3, 4 and 5 represent the accuracy of GEBV in the validation dataset for the 10 traits studied in Braunvieh, Grauvieh and Pinzgauer breed respectively. In all, accuracy of GEBVs

[1] *Note that, although we mention differences in accuracy between methods as well as differences in accuracy*

among methods did not differ significantly[2] for all the three breeds. In Braunvieh breed, the correlations between GEBV and EBV ranged from 0.26 to 0.63 for GBLUP, from 0.25 to 0.63 for Bayes-B and from 0.27 to 0.63 for wgt.GBLUP (Table 3). Accuracies were clearly higher for fat and protein percentage when using Bayesian methodology (Bayes-B and wgt.GBLUP) then GBLUP (Table 3). This result agreed with studies by Hayes *et al*. (2009). It is well known that, milk components like fat percentage are influenced a by few QTL with large effects (Grisart *et al*., 2004). However in Grauvieh and Pinzgauer, there were no clear advantages of using any of the Bayesian assumption in predicting these traits known to have some QTLs with large effect.

Table 3: Braunvieh (BV) breed: Accuracy of GEBV using GBLUP and SNP effect form Bayes-B and wgt.GBLUP

Traits	**No of bulls in ref. dataset**	**[2]Accuracy EBV_{valid}**	**r[(*GEBV*,*EBV*)][1]**		
			GBLUP	Bayes-B	wgt.GBLUP
Milking Speed	393	0.89	$0.63_{(0.06)}$	$0.63_{(0.07)}$	$0.63_{(0.06)}$
Protein *Kg*	390	0.91	$0.35_{(0.13)}$	$0.35_{(0.13)}$	$0.35_{(0.12)}$
Protein Percent	390	0.91	$0.38_{(0.10)}$	$0.44_{(0.10)}$	$0.42_{(0.11)}$
Fat *Kg*	390	0.91	$0.48_{(0.09)}$	$0.46_{(0.09)}$	$0.48_{(0.09)}$
Fat Percent	390	0.91	$0.40_{(0.10)}$	$0.42_{(0.09)}$	$0.41_{(0.09)}$
Fertility Maternal	327	0.74	$0.47_{(0.13)}$	$0.48_{(0.12)}$	$0.48_{(0.12)}$
Milk Kg	390	0.91	$0.26_{(0.13)}$	$0.25_{(0.14)}$	$0.27_{(0.13)}$
Longevity	363	0.74	$0.41_{(0.09)}$	$0.48_{(0.10)}$	$0.44_{(0.10)}$
Persistency	395	0.91	$0.57_{(0.09)}$	$0.57_{(0.09)}$	$0.58_{(0.09)}$
Somatic Cell Count	395	0.86	$0.55_{(0.08)}$	$0.54_{(0.08)}$	$0.55_{(0.08)}$
Mean	**-**	***0.87***	***0.45***	***0.46***	***0.46***

[1] *The youngest 60 bulls are used in calculating the accuracies*
[2] *Average accuracies for the EBV's of the validation dataset*

On average, across all 10 traits, accuracies were highest for wgt.GBLUP (0.47) followed by Bayes-B (0.46) and GBLUP (0.45) in Braunvieh. But this was not the case for both Grauvieh and Pinzgauer (Table 4 & 5) where the 2 methods (GBLUP and Bayes-B) did equally well and outperformed wgt.GBLUP.

Table 4: Grauvieh (GV) breed: Accuracy of GEBV using GBLUP and SNP effect form Bayes-B and wgt.GBLUP

Traits	**No of bulls in ref. dataset**	**[2]Accuracy EBV_{valid}**	**r[(*GEBV*,*EBV*)][1]**		
			GBLUP	Bayes-B	wgt.GBLUP

[2] *Standard errors were estimated with 10,000 bootstrap samples of the validation GEBV. Details are stated in a previous section of this paper.*

Milking Speed	121	0.82	$0.59_{(0.09)}$	$0.64_{(0.07)}$	$0.62_{(0.08)}$
Protein *Kg*	130	0.87	$0.31_{(0.13)}$	$0.30_{(0.13)}$	$0.31_{(0.13)}$
Protein Percent	130	0.87	$0.72_{(0.07)}$	$0.67_{(0.07)}$	$0.70_{(0.07)}$
Fat *Kg*	130	0.87	$0.34_{(0.11)}$	$0.31_{(0.12)}$	$0.34_{(0.11)}$
Fat Percent	130	0.87	$0.70_{(0.07)}$	$0.69_{(0.07)}$	$0.70_{(0.07)}$
Fertility Maternal	63	0.76	$0.47_{(0.09)}$	$0.45_{(0.09)}$	$0.31_{(0.13)}$
Milk Kg	130	0.87	$0.44_{(0.11)}$	$0.45_{(0.11)}$	$0.44_{(0.11)}$
Longevity	91	0.73	$0.12_{(0.11)}$	$0.14_{(0.11)}$	$0.13_{(0.11)}$
Persistency	153	0.86	$0.43_{(0.12)}$	$0.40_{(0.12)}$	$0.42_{(0.12)}$
Somatic Cell Count	136	0.82	$0.70_{(0.06)}$	$0.70_{(0.06)}$	$0.71_{(0.06)}$
Mean	**-**	***0.84***	***0.48***	***0.48***	***0.47***

[1] *The youngest 60 bulls are used in calculating the accuracies except for the traits Fertility Maternal and Longevity where the youngest 30bulls are used*
[2] *Average accuracies for the EBV's of the validation dataset*

In Grauvieh, accuracies of GEBV were above 0.66 for fat and protein percentage and somatic cell count across all methods. However, accuracies for the other traits were lower than 0.65 with accuracy of GEBV for longevity as low as 0.12 (Table 4). Values ranged from 0.30 to 0.72 across methods and were higher for those reported for Braunvieh but lower for the highest accuracy of GEBV for Pinzgauer (0.80; Table 5). The three method predicted GEBVs almost equally for most traits except fertility, where accuracies were 34% (r = 0.47) and 31% (r = 0.45) higher for GBLUP and Bayes-B respectively than for wgt.GBLUP (r = 0.31) (Table 4).

Although accuracies were slightly low to moderately high in Pinzgauer for most traits (values ranged for 0.20 to 0.80; Table 5), except for longevity where accuracies were negative, even across methods. The average accuracy of the validation bulls was 0.79 coupled with small number of bulls in training dataset (95) might have resulted in these negative estimate.

Generally, across all breeds and methods, longevity had very low accuracies (Table 3, 4 & 5). Accuracies of GEBVs for fitness and reproduction traits (especially somatic cell count, milking speed and fertility) have been higher in these three populations. On average, accuracy of theoretical EBVs were 47.5% (values ranged from 29% to 71%), 43.3% (values ranged from 12.4% to 85.4%) and 42.2% (values ranged from 12.1% to 76.2%) higher than those estimated in Braunvieh, Grauvieh and Pinzgauer. This implies that, accuracies of GEBVs can improve by the stated percentages.

Table 5: Pinzgauer (PI) breed: Accuracy of GEBV using GBLUP and SNP effect form Bayes-B and wgt.GBLUP

Traits	**No of bulls in ref. dataset**	[2]**Accuracy** EBV_{valid}	**r[(*GEBV,EBV*)]**[1]		
			GBLUP	Bayes-B	wgt.GBLUP
Milking Speed	110	0.91	$0.51_{(0.11)}$	$0.51_{(0.11)}$	$0.52_{(0.11)}$
Protein *Kg*	99	0.93	$0.49_{(0.11)}$	$0.48_{(0.11)}$	$0.47_{(0.11)}$
Protein Percent	99	0.93	$0.22_{(0.12)}$	$0.21_{(0.12)}$	$0.22_{(0.12)}$
Fat *Kg*	99	0.93	$0.62_{(0.08)}$	$0.62_{(0.08)}$	$0.61_{(0.08)}$
Fat Percent	99	0.93	$0.43_{(0.12)}$	$0.42_{(0.11)}$	$0.41_{(0.12)}$
Fertility Maternal	91	0.85	$0.79_{(0.05)}$	$0.80_{(0.04)}$	$0.78_{(0.05)}$
Milk Kg	99	0.92	$0.46_{(0.10)}$	$0.47_{(0.10)}$	$0.43_{(0.10)}$
Longevity	95	0.79	$-0.04_{(0.12)}$	$-0.05_{(0.12)}$	$-0.06_{(0.12)}$
Persistency	136	0.92	$0.52_{(0.12)}$	$0.53_{(0.10)}$	$0.53_{(0.10)}$
Somatic Cell Count	129	0.90	$0.66_{(0.08)}$	$0.67_{(0.07)}$	$0.67_{(0.07)}$
Mean	**-**	***0.90***	***0.47***	***0.47***	***0.46***
*Mean**		*0.90*	*0.52*	*0.52*	*0.52*

[1] *The youngest 60 bulls are used in calculating the accuracies*
[2] *Average accuracies for the EBV's of the validation dataset*
**Mean values excluding values of longevity*

3.3 Comparison of accuracy from purebred and multibreed

The estimated accuracies of GEBV for selected traits were plotted for single breed analysis and both 2 & 3 way multibreed analyses. Figures 2, 3 & 4 shows the comparison of accuracies estimated with single and multibreed training dataset for Braunvieh, Grauvieh and Pinzgauer respectively. The values for all traits are represented in Table 6 and 7.

In Braunvieh (Figure 2), accuracies were consistently higher for fertility and somatic cell count than for the other selected traits. In general, using a 2-way multibreed training dataset resulted in slightly greater accuracies than training on purebred, although these differences were not significant. Combining Braunvieh and Pinzgauer gave higher accuracies for fertility and somatic cell count. This type of trend was also observed when Braunvieh was combined with Grauvieh as accuracies of protein percentage and milk yield tend to increase slightly (Figure 2). The above mentioned trend that, increase in accuracies for multibreed training dataset than purebred training dataset were consistent across methods; GBLUP, Bayes-B and wgt.GBLUP (Figure 2a, 2b and 2c).

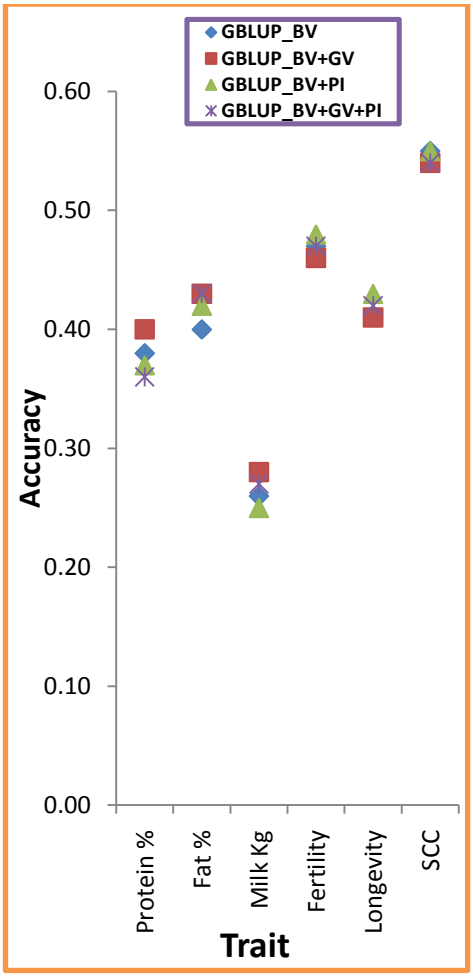

Figure: 2a

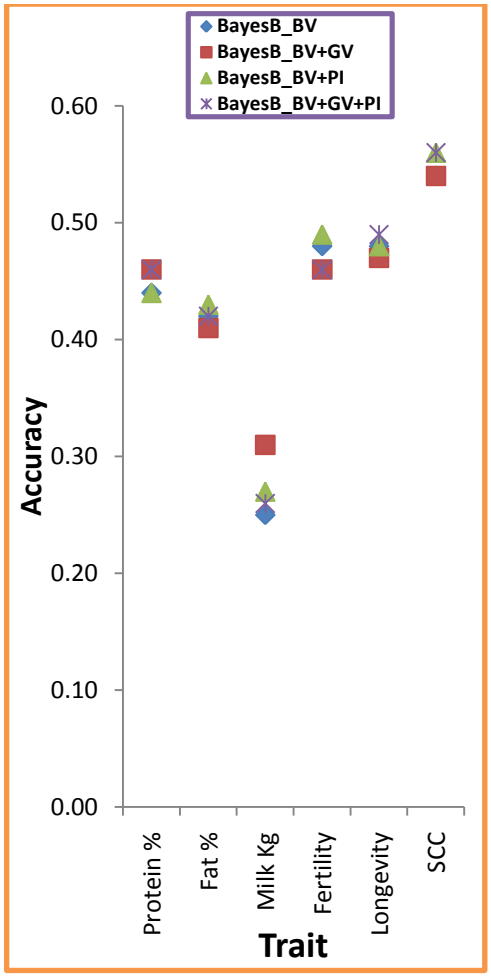

Figure: 2b

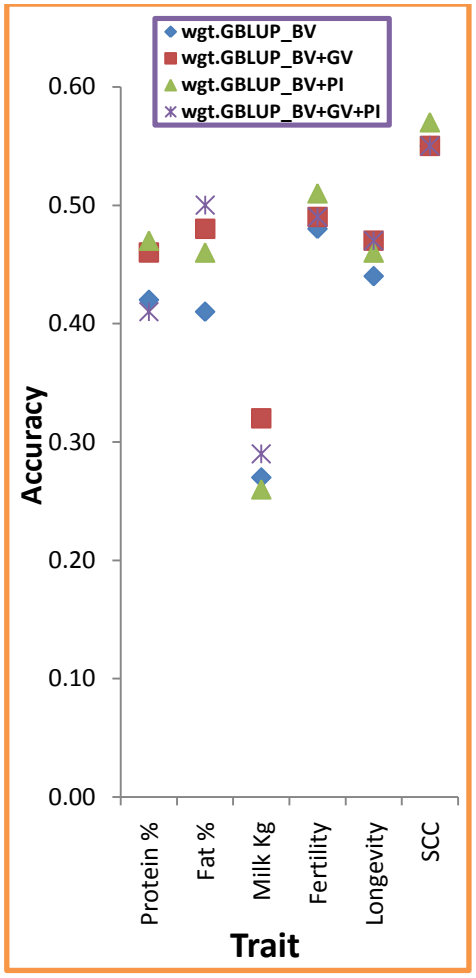

Figure: 2c

Figure 2: Accuracy of GEBV estimated with GBLUP (on the left; 2a), Bayes-B (in the middle; 2b) and wgt.GBLUP (on the right; 2c) when using pure bred and multibreed training dataset in estimating marker effect in Braunvieh

Generally, correlations of predicted GEBVs and EBVs for GBLUP were less variable. Accuracies were almost the same for both purebred and multibreed training dataset compared to wgt.GBLUP where values were highly variable among training dataset (*In Appendix*; Table 6 & 7) and Figure 2.

However, there was a clear trait – by method – by training set interaction. Meaning that, depending on the traits or the method or the training dataset, methods were superior to one another. For example, the clear advantage of Bayes-B to predict GEBV of fat and protein percentage with higher accuracies with the purebred training dataset (Table 1, Figure 2) diminishes when multibreed training dataset is used. This was also observed by Pryce *et al*.

(2011). They reported that although accuracies were higher for Bayes-A method, this wasn't consistent as there were clear instances where GBLUP did better than Bayes-A.

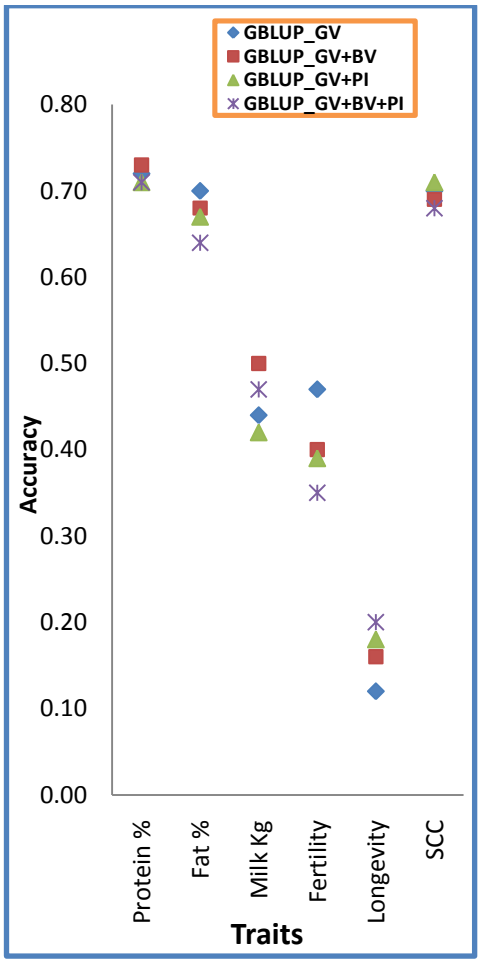

Figure: 3a

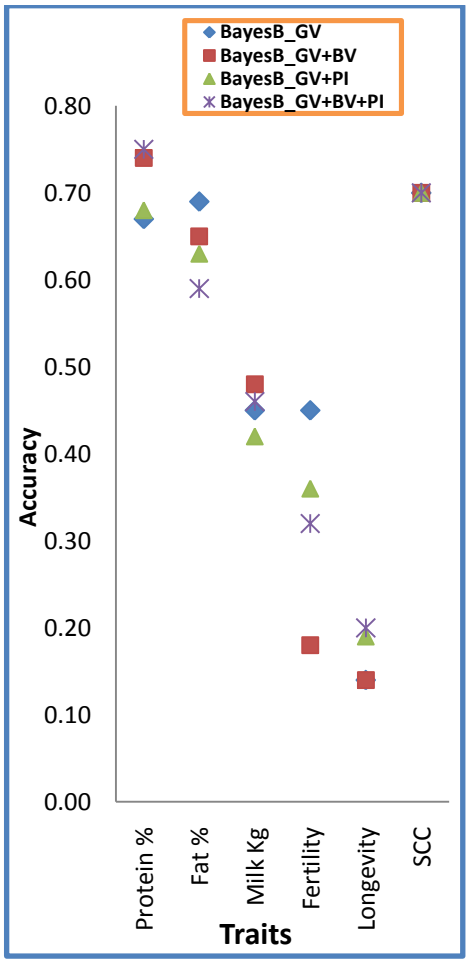

Figure: 3b

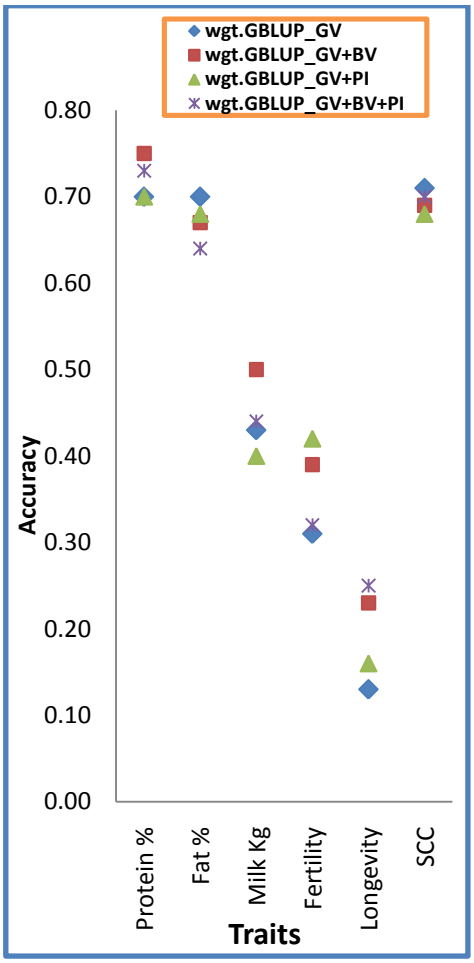

Figure: 3c

Figure 3: Accuracy of GEBV estimated with GBLUP (on the left; 3a), Bayes-B (in the middle; 3b) and wgt.GBLUP (on the right; 3c) when using pure bred and multibreed training dataset f in estimating marker effect in Grauvieh

Figure 3 depicts the comparison of accuracies when using purebred and multibreed training dataset to predict marker effect and estimate GEBV for Grauvieh. As was observed in Braunvieh, combining these two breeds (Grauvieh and Braunvieh) increased accuracy of protein percentage. This trend was seen across methods (Figure 3a, 3b & 3c). Using purebred training datasets in estimating marker effect resulted in higher accuracies than using multibreed training dataset for fat percentage and fertility. Increase in accuracy was observed for longevity, when Grauvieh was combined with Pinzgauer and when the three breeds were combined in the training dataset. This increase was consistent across methods of prediction for this study.

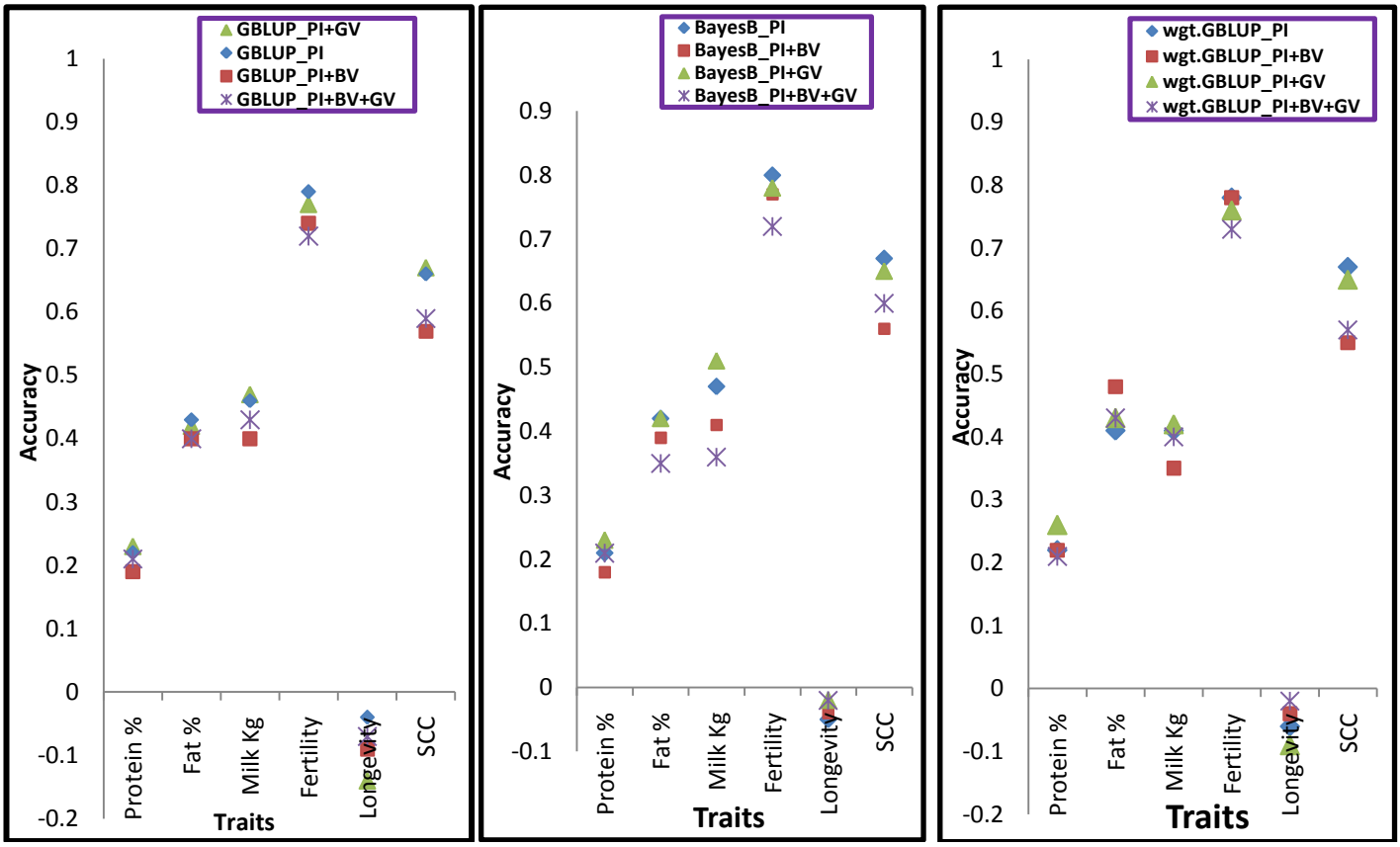

Figure: 4a *Figure: 4b* *Figure: 4c*

Figure 4: Accuracy of GEBV estimated with GBLUP (on the left), Bayes-B (in the middle) and wgt.GBLUP (on the right) when using pure bred and multibreed training dataset in estimating marker effect in Pinzgauer

Accuracy of GEBVs when using purebred Pinzgauer training dataset and combined breeds in estimating marker effect varied for traits and methods used in this study (*In Appendix*; Table 6 & 7; Figure 4). It was evident (Figure 4a, 4b & 4c) that wgt.GBLUP was superior in estimating marker effect and predicting GEBV with smaller prediction error for protein and fat percentage than GBLUP and Bayes-B especially using multibreed training dataset. As was mentioned before for the Braunvieh and Grauvieh, there was trait – by method – by training set interactions for Pinzgauer. Combining Pinzgauer and Braunvieh in a 2 way multibreed training dataset reduces accuracy, this meant that, accuracies from purebred training set were higher than for combined PI and BV breeds for almost all traits (*In Appendix*; Table 6 &7; Figure 4a, 4b & 4c). This is seen clearly in both GBLUP and Bayes-B but moderately in wgt.GBLUP. There was however an increase in accuracy of GEBV when Pinzgauer was combined with Grauvieh mostly for fitness traits and slightly for milk yield

and milk component. And again this is most clear for the methods GBLUP and Bayes-P than for wgt.GBLUP (Figure 4a, 4b & 4c).

Generally, there were small increases and decreases in accuracies across traits when multibreed training set was used (Table 8a, 8b & 8c). Combining training set of Pinzgauer dataset or Grauvieh or the combination of the 2 to Braunvieh training lead to no or increases in accuracy of production traits (Table 8a). On average accuracy increase about 2.88% - GBLUP, 4.17% – Bayes-B and 11.20% wgt.GBLUP when using a multibreed training set for the selected production traits and about 0.23% - GBLUP, 0.23% – Bayes-B and 3.50% wgt.GBLUP for fitness traits. Increases were higher for wgt.GBLUP followed by Bayes-B.

Table 8a: Percentage increases or decreases in accuracy using multibreed training dataset for Braunvieh breed. Values[3] are average accuracies across three production and fitness traits

	Method	**Braunvieh**				
		BV+GV	**BV+PI**	**BV+PI+GV**	**Mean**	**Overall Mean**[c]
Production traits[a]	GBLUP	6.73	0.00	1.92	***2.88***	***1.56***
Fitness traits[b]		-1.40	2.10	0.00	***0.23***	
Production traits	Bayes-B	6.73	2.88	2.88	***4.17***	***2.2***
Fitness traits		-2.10	2.10	0.70	***0.23***	
Production traits	wgt.GBLUP	15.38	8.65	9.62	***11.2***	***7.36***
Fitness traits		2.80	4.90	2.80	***3.50***	

[a] *Three traits were included: Protein and Fat percentage and Milk yield*
[b] *Three traits were included: Fertility, Longevity and somatic cell count*
[c] *The mean increase in accuracy for both production and fitness traits*
BV – Braunvieh GV - Grauvieh PI – Pinzgauer

Although individual fitness traits were improved (e.g. longevity; Figure 3 & 4) when Grauvieh was combined to Pinzgauer, average accuracies for 3 selected fitness traits, fertility, longevity and somatic cell count, shows an overall decrease in estimated accuracy for the 2 breeds (Table 8b & 8c). Note that the decrease in accuracies was smaller than a combination of Braunvieh and Pinzgauer multibreed training set. In Grauvieh, increase in accuracy of production traits were recorded when Braunvieh training bulls were added. But accuracy decreased for fitness traits. Also any form of combination with Pinzgauer resulted in reduction in accuracies estimated with purebred prediction. Exactly opposite trend was

[3] Values were calculated as :

$$\sum_{j=3} \frac{\left(Method_{multibreed\ ,j} - Method_{purebre\ d,j}\right)}{Method_{purebred\ ,j}} \times 100$$

Where j is the traits; Method is GBLUP or Bayes-B or wgt.MixP

observed for Pinzgauer. An increase and decrease in accuracy for 2 way combination with Grauvieh for production traits and fitness traits respectively. However regardless of the traits under study accuracy for purebred analysis decreased when Braunvieh training bulls were added to pinzgauer training dataset. Average decrease in accuracy was 1.87% - GBLUP and 5.18% – Bayes-B when using a multibreed training set for Grauvieh as well as 8.26% - GBLUP and 6.68% – Bayes-B for Pinzgauer.

Table 8b: Percentage increases or decreases in accuracy using multibreed training dataset for Grauvieh breed. Values are average accuracies across three production and fitness traits

		Grauvieh				
		GV+BV	**GV+PI**	**GV+BV+PI**	**Mean**	**Overall Mean**[c]
Production traits[a]	GBLUP	2.69	-3.23	-2.15	***-0.90***	***-1.87***
Fitness traits[b]		-3.10	-0.78	-4.65	***-2.84***	
Production traits	Bayes-B	3.23	-4.30	-0.54	***-0.54***	***-5.18***
Fitness traits		-20.93	-3.10	-5.43	***-9.82***	
Production traits	wgt.GBLUP	4.48	-2.69	-1.08	***0.36***	***5.22***
Fitness traits		12.40	8.53	9.30	***10.1***	

[a] *Three traits were included: protein and fat percentage and milk yield*
[b] *Three traits were included: fertility, longevity and somatic cell count*
[c] *The mean increase in accuracy for both production and fitness traits*
BV – Braunvieh GV - Grauvieh PI – Pinzgauer

Table 8c: Percentage increases or decreases in accuracy using multibreed training dataset for Pinzgauer breed. Values are average accuracies across three production and fitness traits

		Pinzgauer				
		PI+BV	**PI+GV**	**PI +BV+GV**	**Mean**	**Overall Mean**[c]
Production traits[a]	GBLUP	-10.8	0.90	-6.30	***-5.40***	***-8.26***
Fitness traits[b]		-13.5	-7.80	-12.06	***-11.1***	
Production traits	Bayes-B	-10.8	5.41	-16.21	***-7.20***	***-6.68***
Fitness traits		-9.22	-0.71	-8.51	***-6.15***	
Production traits	wgt.GBLUP	0.90	6.31	0	***2.40***	***-2.11***
Fitness traits		-7.09	-4.96	-7.80	***-6.61***	

[a] *Three traits were included: protein and fat percentage and milk yield*
[b] *Three traits were included: fertility, longevity and somatic cell count*
[c] *The mean increase in accuracy for both production and fitness traits*
BV – Braunvieh GV - Grauvieh PI – Pinzgauer

3.4 Regression of EBVs on predicted GEBVs

Table 9 a,b & c, 10 & 11 shows the regression coefficients of EBV and predicted GEBVs for various methods and traits. Regression coefficient deviated highly from 1 for some traits. On for purebreds there were less bias for Pinzgauer followed by Grauvieh and Braunvieh, this is

averaged across traits. Generally for the purebreds, wgt.GBLUP showed less bias in predicting EBVs compared to Bayes-B and GBLUP. Regression of EBV on GEBV relies on the independence of the prediction errors for EBVs and GEBVs. A prediction error of EBV also depends on their accuracy (Brøndum *et. al*., 2011). Traits with lower accuracies (Table 3, 4 & 5) for the correlation between GEBV and EBV showed higher and lower bias (deviation from 1; Table 9a, b & c).

Table 9a: Regression Coefficient of EBV on GEBV with the GS methods GBLUP, Bayes-B and wgt.GBLUP for purebred analysis in Braunvieh for the 10 traits

Traits	**Braunvieh**		
	GBLUP	***Bayes-B***	***wgt. GBLUP***
Milking Speed	$1.48_{(0.24)}$	$0.95_{(0.15)}$	$1.48_{(0.24)}$
Protein *Kg*	$0.52_{(0.18)}$	$0.47_{(0.17)}$	$0.48_{(0.17)}$
Protein %	$0.71_{(0.23)}$	$0.73_{(0.19)}$	$0.71_{(0.20)}$
Fat *Kg*	$0.82_{(0.20)}$	$0.70_{(0.18)}$	$0.78_{(0.19)}$
Fat %	$0.93_{(0.28)}$	$0.83_{(0.24)}$	$0.90_{(0.26)}$
Fertility	$0.91_{(0.22)}$	$0.56_{(0.13)}$	$0.78_{(0.19)}$
Milk *Kg*	$0.45_{(0.22)}$	$0.38_{(0.19)}$	$0.44_{(0.21)}$
Longevity	$0.62_{(0.18)}$	$0.38_{(0.13)}$	$0.49_{(0.13)}$
Persistency	$1.06_{(0.20)}$	$0.79_{(0.15)}$	$0.86_{(0.16)}$
Somatic cell count	$0.85_{(0.17)}$	$0.66_{(0.14)}$	$0.71_{(0.14)}$
Mean	**0.84**	**0.65**	**0.76**

Standard error for these estimate ranged from 0.13-0.41, estimated from the 10,000 bootstrapping
(*) – Poorly estimated, values were too large or too small.

Multibreed training set predicted EBV worse than purebred training set (Table 10). Regression coefficient showed more deviation from 1. Longevity had the greatest bias of 0.52 averaged across multibreed training set (Table 10). Regression coefficients for milking speed, protein percentage, fertility and somatic cell count seems to be predicting EBVs with less bias in both the 2 way and 3 way multibreed training set. Brøndum *et. al*. (2011), also reported larger deviation from 1 for longevity in Finnish Red diary bulls.

Table 9b: Regression Coefficient of EBV on GEBV with the GS methods GBLUP, Bayes-B and wgt.GBLUP for purebred analysis in Grauvieh for the 10 traits

Traits	**Grauvieh**		
	GBLUP	***Bayes-B***	***Wgt.GBLUP***
Milking Speed	$1.91_{(0.35)}$	$0.87_{(0.14)}$	$1.31_{(0.22)}$
Protein *Kg*	$0.50_{(0.20)}$	$0.48_{(0.20)}$	$0.51_{(0.21)}$
Protein %	$1.62_{(0.21)}$	$1.17_{(0.17)}$	$1.22_{(0.17)}$
Fat *Kg*	$0.54_{(0.19)}$	$0.45_{(0.18)}$	$0.52_{(0.19)}$
Fat %	$1.66_{(0.22)}$	$1.55_{(0.21)}$	$1.58_{(0.21)}$
Fertility	$1.96_{(0.41)}$	$1.12_{(0.29)}$	$0.49_{(0.20)}$

Milk *Kg*	$0.80_{(0.22)}$	$0.63_{(0.16)}$	$0.62_{(0.17)}$
Longevity	*	$0.62_{(0.41)}$	*
Persistency	$0.88_{(0.24)}$	$0.63_{(0.19)}$	$0.76_{(0.21)}$
Somatic cell count	$1.40_{(0.19)}$	$1.05_{(0.14)}$	$1.08_{(0.14)}$
Mean	**1.25**	**0.86**	**0.90**

Table 9c: Regression Coefficient of EBV on GEBV with the GS methods GBLUP, Bayes-B and wgt.GBLUP for purebred analysis in Pinzgauer for the 10 traits

Traits	**Pinzgauer**		
	GBLUP	***Bayes-B***	***wgt.GBLUP***
Milking Speed	$1.17_{(0.26)}$	$1.01_{(0.22)}$	$1.04_{(0.22)}$
Protein *Kg*	$1.15_{(0.27)}$	$1.11_{(0.26)}$	$1.10_{(0.27)}$
Protein %	$0.39_{(0.23)}$	$0.37_{(0.23)}$	$0.38_{(0.23)}$
Fat *Kg*	$1.29_{(0.22)}$	$1.29_{(0.21)}$	$1.27_{(0.21)}$
Fat %	$1.25_{(0.35)}$	$1.20_{(0.37)}$	$1.20_{(0.36)}$
Fertility	$1.38_{(0.14)}$	$1.41_{(0.14)}$	$1.38_{(0.14)}$
Milk *Kg*	$1.24_{(0.31)}$	$1.22_{(0.30)}$	$1.16_{(0.31)}$
Longevity	*	*	*
Persistency	$0.94_{(0.20)}$	$0.94_{(0.20)}$	$0.96_{(0.20)}$
Somatic cell count	$1.24_{(0.18)}$	$1.20_{(0.18)}$	$1.22_{(0.18)}$
Mean	**1.12**	**1.08**	**1.08**

3.5 Comparison of GS prediction Methods

Correlations between methods for the validation set were high (Table 13, 14 & 15, *see appendix 1*) for Braunvieh and Pinzgauer compared to Grauvieh for some selected traits. Correlations were lower when breeds are combined both in the 2-way and 3-way multibreed analysis. Correlations ranged from 0.70 to 1.0. Also correlations were higher for milk yield and somatic cell count. This might be due to the higher reliabilities of DrEBV as well as the number of (Table 3a, 3b & 3c) bulls in the training set. When number of bulls for fertility in Grauvieh was increased from 63 to 450 with the addition of Braunvieh bulls, correlations with other methods increased. Nirea (2009) also reported similar correlation between Bayesian methods and GBLUP for fat and protein percentage and somatic cell count.

Table 10: Regression Coefficient of EBV on GEBV with the GS methods GBLUP, Bayes-B and wgt.GBLUP for 2 way multibreed analysis for the following traits: Milking Speed (DMG), Protein Kg (EKG), Protein Percent (EP), Fat Kg (FKG), Fat Percent (FP), Fertility Maternal (FRM), Milk Kg (MKG), Longevity (ND), Persistency (PER) and Somatic Cell Count (ZZ)

REF	VAL	Method	DMG	EKG	EP	FKG	FP	FRM	MKG	ND	PER	ZZ
	BV	*GBLUP*	1.46$_{(0.24)}$	0.51$_{(0.18)}$	0.70$_{(0.21)}$	0.77$_{(0.19)}$	0.93$_{(0.26)}$	0.82$_{(0.21)}$	0.50$_{(0.22)}$	0.65$_{(0.19)}$	1.03$_{(0.20)}$	0.81$_{(0.17)}$
		Bayes-B	0.90$_{(0.15)}$	0.47$_{(0.17)}$	0.76$_{(0.19)}$	0.81$_{(0.22)}$	0.76$_{(0.22)}$	0.51$_{(0.13)}$	0.49$_{(0.20)}$	0.39$_{(0.11)}$	0.79$_{(0.15)}$	0.64$_{(0.13)}$
		wgt.GBLUP	1.11$_{(0.20)}$	0.47$_{(0.17)}$	0.78$_{(0.20)}$	0.76$_{(0.19)}$	0.94$_{(0.23)}$	0.59$_{(0.14)}$	0.52$_{(0.20)}$	0.53$_{(0.13)}$	0.83$_{(0.16)}$	0.65 $_{(0.13)}$
BV+GV												
	GV	*GBLUP*	1.16$_{(0.19)}$	0.51$_{(0.23)}$	1.39$_{(0.17)}$	0.59$_{(0.20)}$	1.71$_{(0.24)}$	1.43$_{(0.36)}$	0.76$_{(0.17)}$	0.67$_{(0.39)}$	0.79$_{(0.23)}$	1.12$_{(0.15)}$
		Bayes-B	0.82$_{(0.13)}$	0.51$_{(0.20)}$	1.23$_{(0.15)}$	0.64$_{(0.22)}$	1.44$_{(0.22)}$	0.26$_{(0.19)}$	0.65$_{(0.15)}$	0.23$_{(0.22)}$	0.56$_{(0.19)}$	0.97$_{(0.13)}$
		wgt.GBLUP	0.88$_{(0.88)}$	0.47$_{(0.20)}$	1.31$_{(0.15)}$	0.55$_{(0.19)}$	1.54$_{(0.22)}$	0.69$_{(0.21)}$	0.66$_{(0.15)}$	0.55$_{(0.30)}$	0.60$_{(0.19)}$	0.93$_{(0.13)}$
	BV	*GBLUP*	1.39$_{(0.23)}$	0.44$_{(0.17)}$	0.65$_{(0.21)}$	0.69$_{(0.18)}$	0.86$_{(0.25)}$	0.85$_{(0.20)}$	0.39$_{(0.20)}$	0.61$_{(0.17)}$	0.99$_{(0.18)}$	0..81$_{(0.16)}$
		Bayes-B	0.93$_{(0.15)}$	0.40$_{(0.16)}$	0.73$_{(0.19)}$	0.60$_{(0.16)}$	0.81$_{(0.22)}$	0.57$_{(0.13)}$	0.39$_{(0.18)}$	0.37$_{(0.11)}$	0.74$_{(0.14)}$	0.68$_{(0.14)}$
		wgt.GBLUP	1.06$_{(0.17)}$	0.38$_{(0.16)}$	0.79$_{(0.20)}$	0.64$_{(0.15)}$	0.87$_{(0.22)}$	0.56$_{(0.12)}$	0.37$_{(0.18)}$	0.40$_{(0.11)}$	0.71$_{(0.14)}$	0.67$_{(0.13)}$
BV+PI												
	PI	*GBLUP*	1.15$_{(0.28)}$	1.01$_{(0.29)}$	0.36$_{(0.24)}$	1.21$_{(0.23)}$	1.13$_{(0.34)}$	2.01$_{(0.29)}$	1.13$_{(0.33)}$	*	1.12$_{(0.25)}$	1.19$_{(0.23)}$
		Bayes-B	0.83$_{(0.20)}$	0.90$_{(0.27)}$	0.32$_{(0.22)}$	1.05$_{(0.23)}$	0.94$_{(0.30)}$	1.33$_{(0.14)}$	1.05$_{(0.31)}$	*	0.90$_{(0.22)}$	1.00$_{(0.20)}$
		wgt.GBLUP	0.96$_{(0.22)}$	0.80$_{(0.28)}$	0.37$_{(0.20)}$	1.02$_{(0.21)}$	1.23$_{(0.30)}$	1.31$_{(0.14)}$	0.91$_{(0.32)}$	*	0.91$_{(0.20)}$	1.00$_{(0.20)}$
	GV	*GBLUP*	1.17$_{(0.20)}$	0.45$_{(0.19)}$	1.49$_{(0.20)}$	0.47$_{(0.18)}$	1.59$_{(0.23)}$	0.82$_{(0.25)}$	0.56$_{(0.16)}$	0.82$_{(0.39)}$	0.72$_{(0.21)}$	1.07$_{(0.13)}$
		Bayes-B	0.83$_{(0.19)}$	0.45$_{(0.16)}$	1.16$_{(0.16)}$	0.50$_{(0.17)}$	1.31$_{(0.21)}$	0.80$_{(0.27)}$	0.53$_{(0.15)}$	0.35$_{(0.24)}$	0.65$_{(0.18)}$	1.03$_{(0.14)}$
		wgt.GBLUP	0.90$_{(0.15)}$	0.41$_{(0.19)}$	1.21$_{(0.16)}$	0.51$_{(0.18)}$	1.47$_{(0.21)}$	0.90$_{(0.25)}$	0.54$_{(0.16)}$	0.69$_{(0.39)}$	0.60$_{(0.18)}$	1.06$_{(0.13)}$
GV+PI												
	PI	*GBLUP*	1.47$_{(0.32)}$	1.13$_{(0.25)}$	0.48$_{(0.26)}$	1.25$_{(0.21)}$	1.38$_{(0.38)}$	1.34$_{(0.15)}$	1.24$_{(0.30)}$	*	1.34$_{(0.30)}$	1.32$_{(0.19)}$
		Bayes-B	0.91$_{(0.22)}$	1.09$_{(0.25)}$	0.40$_{(0.23)}$	1.23$_{(0.20)}$	1.21$_{(0.34)}$	1.32$_{(0.14)}$	1.20$_{(0.26)}$	*	0.88$_{(0.20)}$	1.14$_{(0.17)}$
		wgt.GBLUP	1.07$_{(0.22)}$	1.04$_{(0.25)}$	0.46$_{(0.22)}$	1.20$_{(0.20)}$	1.29$_{(0.35)}$	1.26$_{(0.14)}$	1.01$_{(0.29)}$	*	0.90$_{(0.20)}$	1.15$_{(0.18)}$
Mean			**1.06**	**0.64**	**0.81**	**0.81**	**1.19**	**0.97**	**0.72**	**0.52**	**0.83**	**0.98**

Standard error for these estimate ranged from 0.11 - 0.39, estimated from the 10,000 bootstrapping
(*) – Poorly estimated, values were over (too large) or under estimated (too small).
REF – Reference dataset VAL – Validation dataset
BV – Braunvieh GV - Grauvieh PI – Pinzgauer

Table 11: Regression Coefficient of EBV on GEBV with the GS methods GBLUP, Bayes-B and wgt.GBLUP for 3 way multibreed analysis for the following traits: Milking Speed (DMG), Protein Kg (EKG), Protein Percent (EP), Fat Kg (FKG), Fat Percent (FP), Fertility Maternal (FRM), Milk Kg (MKG), Longevity (ND), Persistency (PER) and Somatic Cell Count (ZZ)

REF	VAL	Method	DMG	EKG	EP	FKG	FP	FRM	MKG	ND	PER	ZZ
	BV	*GBLUP*	$1.40_{(0.24)}$	$0.44_{(0.17)}$	$0.62_{(0.21)}$	$0.67_{(0.18)}$	$0.89_{(0.25)}$	$0.77_{(0.19)}$	$0.42_{(0.20)}$	$0.66_{(0.19)}$	$1.01_{(0.19)}$	$0.78_{(0.16)}$
		Bayes-B	$0.87_{(0.15)}$	$0.40_{(0.16)}$	$0.71_{(0.18)}$	$0.60_{(0.16)}$	$0.73_{(0.21)}$	$0.51_{(0.13)}$	$0.38_{(0.19)}$	$0.39_{(0.10)}$	$0.73_{(0.14)}$	$0.67_{(0.13)}$
		wgt.GBLUP	$1.04_{(0.18)}$	$0.40_{(0.17)}$	$0.68_{(0.20)}$	$0.61_{(0.16)}$	$0.94_{(0.21)}$	$0.52_{(0.12)}$	$0.42_{(0.18)}$	$0.48_{(0.12)}$	$0.76_{(0.14)}$	$0.62_{(0.12)}$
BV+GV+PI	**GV**	*GBLUP*	$1.10_{(0.19)}$	$0.49_{(0.22)}$	$1.35_{(0.18)}$	$0.54_{(0.19)}$	$1.48_{(0.23)}$	$1.08_{(0.39)}$	$0.79_{(0.17)}$	$0.75_{(0.39)}$	$1.01_{(0.22)}$	$1.15_{(0.14)}$
		Bayes-B	$0.80_{(0.13)}$	$0.51_{(0.20)}$	$1.27_{(0.15)}$	$0.47_{(0.13)}$	$1.26_{(0.23)}$	$0.60_{(0.23)}$	$0.63_{(0.16)}$	$0.33_{(0.20)}$	$0.60_{(0.17)}$	$0.87_{(0.12)}$
		wgt.GBLUP	$0.85_{(0.14)}$	$0.49_{(0.21)}$	$1.28_{(0.16)}$	$0.54_{(0.17)}$	$1.44_{(0.23)}$	$0.56_{(0.22)}$	$0.61_{(0.16)}$	$0.53_{(0.27)}$	$0.56_{(0.17)}$	$0.86_{(0.12)}$
	PI	*GBLUP*	$1.15_{(0.28)}$	$1.05_{(0.27)}$	$0.40_{(0.25)}$	$1.22_{(0.22)}$	$1.14_{(0.35)}$	$2.01_{(0.26)}$	$1.15_{(0.32)}$	*	$1.11_{(0.26)}$	$1.27_{(0.23)}$
		Bayes-B	$0.75_{(0.21)}$	$0.88_{(0.24)}$	$0.35_{(0.21)}$	$1.06_{(0.20)}$	$0.81_{(0.28)}$	$1.15_{(0.15)}$	$0.87_{(0.30)}$	*	$0.77_{(0.19)}$	$1.08_{(0.19)}$
		wgt.GBLUP	$0.88_{(0.22)}$	$0.88_{(0.25)}$	$0.36_{(0.22)}$	$1.07_{(0.20)}$	$1.14_{(0.31)}$	$1.12_{(0.14)}$	$0.98_{(0.30)}$	*	$0.85_{(0.19)}$	$0.99_{(0.19)}$
Mean			**0.98**	**0.62**	**0.78**	**0.75**	**1.09**	**0.92**	**0.69**	**0.52**	**0.82**	**0.92**

Standard error for these estimate ranged from 0.10 - 0.39, estimated from the 10,000 bootstrapping
(*) – Poorly estimated, values were over (too large) or under estimated (too small).
REF – Reference dataset VAL – Validation dataset
BV – Braunvieh GV - Grauvieh PI – Pinzgauer

3.6 LD between syntenic markers and Persistence of LD between breeds

We assessed LD in Braunvieh, Grauvieh and Pinzgauer measured as R^2 (Figure 5) for syntenic markers and also the extent of LD between breeds using the correlation of the *r* (Table 12). Marker interval on the Illumina BovineSNP50 BeadChip has a mean of 67 kbp (median of about 50 kkp with most markers in a range of 45-100 kbp). LD (R^2) for these markers with 67 kbp was 0.153, 0.142, 0.133 and 0.126 for Braunvieh, Grauvieh, and Pinzgauer and for combined breeds respectively.

These values were slightly lower than those reported by De Roos *et al*. (2008) studying Holstein and Angus breeds. Recent studies by Lamer *et al.* (2012) unpublished; in Canadian Brown Swiss for Illumina Bovine SNP50 BeadChip on LD, reported average R^2 value of 0.20 for marker distance between 50 – 100 kbp.

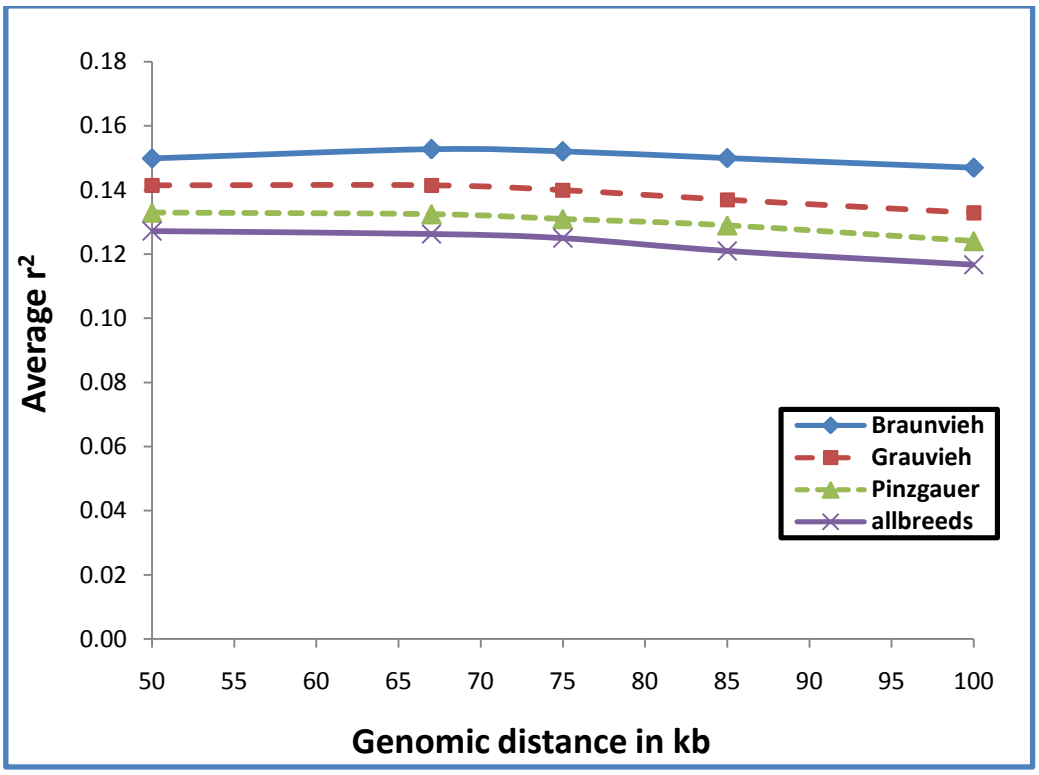

Figure 5: Average LD (r2) for syntenic markers of genomic distances between 50 kb and 85 kb for Braunvieh, Grauvieh, Pinzgauer and combined breeds (multibreed)

The extend of LD was measured as the correlation of LD (measured as *r*) between breeds (Table 12). Correlation of *r* for syntenic markers at 67 kpb's apart were strongest for

Grauvieh – Pinzgauer (0.45) combination followed by Braunvieh – Grauvieh (0.44) and Braunvieh – Pinzgauer (0.41) (Table 12).

Table 12: Average LD (r) for genomic distances of 67 kb and 1000 kb. Correlation of r of marker pairs between breeds: r values above the diagonal are correlation with 67 kb marker distance whiles below the diagonal is the correlation with 1000 kb marker distance.

Breed	**Marker distance (kb)**		**Correlation of r**		
	67	***1000***	***Bruanvieh***	***Grauvieh***	***Pinzgauer***
Bruanvieh	0.102	0.052		0.442	0.408
Grauvieh	0.081	0.029	0.200		0.446
Pinzgauer	0.076	0.030	0.174	0.219	

4 Discussion

The study investigated the accuracy of GEBV for both purebred and multibreed GS approaches for small populations (Braunvieh, Grauvieh and Pinzgauer breeds form Austria). The result shows that accuracy of GEBVs were low to moderately high in these three breeds for both purebred and multibreed analysis. The benefit of combining breeds into a 2 way multibreed training set was only 1.9% and a loss of 1.32% in prediction accuracy for a 3 way multibreed training set, these values are averaged over methods (GBLUP, Bayes-B and wgt.GBLUP) and the 10 traits used in this study.

For the single breed analysis, there were no clear advantages of using any of the Bayesian assumption in predicting traits known to have some QTLs with large effect especially in Grauvieh and Pinzgauer. It is possible that, the number of training animals (190 for Grauvieh and 159 for Pinzgauer) were too small for Bayes-B to identify and locate the QTL. Moreover, the two breeds have been selected for more beef and fitness traits than milk production. Thus the known DGAT allele for milk yield and it component might not be segregating in these two breeds. Generally, across all breeds and methods in the single breed analysis, longevity had very low accuracies (Table 3, 4 & 5), which reflects very low heritabilities associated with this traits and the small number of bulls both use in the training dataset and validation dataset. Accuracies of GEBVs for fitness and reproduction traits (especially somatic cell count, milking speed and fertility) have been higher in these three populations, probably due to the fact that, QTLs for these traits might be segregating in moderate frequency thus explaining larger proportions of the genetic variance (Falconer and Mackay, 1996).

When breeds were combined, there were combinatorial differences between the breeds. Braunvieh and Grauvieh predicted each other with increased in accuracy of 2.1% and 3.6% respectively. Also, addition of Grauvieh to Pinzgauer and vice versa increased accuracy by 8.7% for pinzgauer and 2.0% for Grauvieh compared to using purebred training set but there were loss (5.7%) in accuracy aggregating Braunvieh and Pinzgauer. This implies small to no benefit in combining these three breeds together. An increase in accuracy combining Grauvieh and Pinzgauer (mostly for fitness traits and slightly for milk yield and milk component which was clearly shown using the methods GBLUP and Bayes-B than for wgt.GBLUP; Figure 4a, 4b & 4c) as well as the combination of the three breeds. Hayes *et al*. (2009), Pryce *et al*. (2011), and De Roos *et al*. (2009) all reported a slight to moderate

increase in accuracies when multibreed training set is adopted to estimate marker effect and predict GEBVs for purebreds. They suggested that, this trend is only possible if populations under study are closely related and SNP marker densities were high enough to maintain LD of markers with QTLs. The specific breed – by breed increases and losses in accuracies observed in this study, might be due to the aim of the breeding goals leading to SNP markers not been in LD with QTL in both breeds and or QTL not been in the same phase across breeds for those specific traits. If QTLs are segregating in these breeds and the above mentioned reason is true, increasing the number of animals will expectedly increase accuracy as was observed. As mentioned earlier, Grauvieh and Pinzgauer have been selected for high beef and fitness traits and therefore marker-QTL association might be in the same phase which leads to complementary increase in accuracies due to the increase in the number of training animals compared to Braunvieh which is been highly selected for milk and virtually no selection for beef. Different selection schemes might have lead to different QTL alleles segregating for different population. This view of QTLs been in LD and not in the same phase across population has been shared by Hayes et al. (2009) and Pryce et al. (2011). Both research teams demonstrated that, SNPs in low LD with QTL do not receive an effect in the prediction equation but only those in high LD.

Moreover, according to the Grauvieh and Pinzgauer breeding companies, breeding goal for fitness traits is about 50% of the total merit index. This suggest that, there might be close genetic ties (*i.e* in terms of alleles that are present) among the two breeds for fitness traits, probably because both have been breed with larger emphasis on those traits than Braunvieh that has been breed for milk production and milk composition. Lund *et al*. (2011) reported an overall increase of 10% in accuracy for multibreed predictions over purebred predictions studying Holstein bulls from the *EuroGenomics* project (animals from France, Denmark, Sweden, Finland, Germany, Netherlands and Flanders). Contrarily, as mentioned earlier in this paper, the combination of Pinzgauer and Braunvieh in a 2 way multibreed training dataset reduced accuracy, this meant that, accuracies from purebred training set were higher than for combined PI and BV breeds for almost all traits (Table 6 &7; Figure 4a, 4b & 4c) using both GBLUP and Bayes-B but moderately in wgt.GBLUP.

LD decays slightly as genomic distance increase for all breeds (Figure 5). LD was substantially higher for Braunvieh followed by Grauvieh. Lamer *et al.* (2012) unpublished; in Canadian Brown Swiss for Illumina BovineSNP50 BeadChip on LD, reported average R^2 of

0.20 for marker distance between 50 – 100 kbp. De Roos *et al*. (2008) also reported average R^2 of 0.167 for markers within 50 – 100 kbp for six different populations of Holsteins, Jersey and Angus breeds. The lower values reported in the Braunvieh breed ("Brown Swiss") compared to the Canadian Brown Swiss suggest that, effective population size might be higher and that Canadian Brown Swiss might have been selected for higher genetic merit (De Roos *et al*., 2008). The implication of these slightly low R^2 for these breed is that, markers might be too distant to predict QTL effect. The combination of all three breeds reduces LD between markers to an average of 0.123, suggesting that, on average combining breeds might not increase accuracy of GEBV. This is seen with the 3 way multibreed training set giving decrease in accuracies average across traits and methods of 1.32%. Correlation of *r* for syntenic markers at 67 kpb's apart were strongest for Grauvieh – Pinzgauer (0.45) combination followed by Braunvieh – Grauvieh (0.44) and Braunvieh – Pinzgauer (0.41) (Table 12). In general these estimates are smaller those reported by De Roos et al. (2008) considering the marker distance used. This means, marker pairs were in LD or in a certain LD phase might not persist. These results depict clearly what was observed when breeds were combined in the 2 way analysis. Generally, accuracies were slightly increased for Braunvieh and Grauvieh, Grauvieh and Pinzgauer multibreed training set and reduced for Braunvieh and Pinzgauer combination.

In addition to the studying LD between markers pairs and extent of LD between breeds, a plot (Figure 6) of the principal component (*PC1 and PC2 using a scale singular value decomposition method*; both explaining about 70% of genotype diversity- *Genomic relationship matrix)* of the SNP genotype for all the 35,319 markers also affirms the result seen with the LD studies. The first principle component on the vertical axis separate Braunvieh from the 2 breeds. This again shows some genetic ties between Grauvieh and Pinzgauer. Some SNPs are only largely present in one breed and might cause a reduction in accuracy if breeds are combined together especially if those SNPs explain part of the genetic variances we see in that breed.

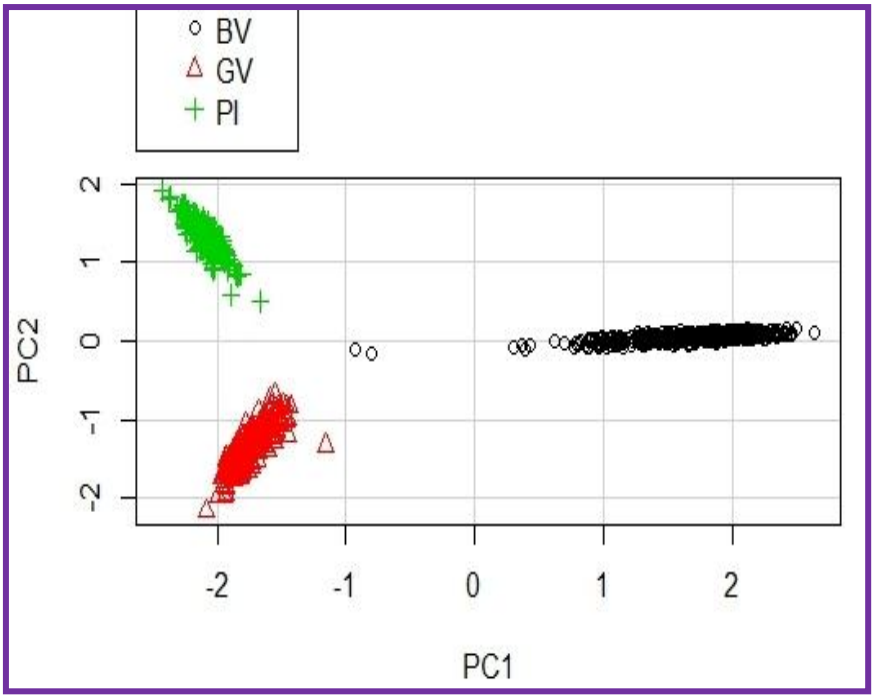

Figure 6: The first 2 principal components (PC1 and PC2) of Braunvieh, Grauvieh and Pinzgauer breeds using the GRM matrix

We assert that for Braunvieh "Brown Swiss" combining breeds from different countries might help improve accuracies of this Austrian breed. There is the evidence that, the Austrian Braunvieh breed are closely related to the once in USA and Canada (they were imported from Switzerland) (http://www.brownswissusa.com/Breed/History), thus combining the animals from these countries will be expected to be more beneficial than an entirely new breed. The reason is that, animals from these breeds have a recent common ancestor making them genetically less distant than with other breeds. As was reported by Lund *et al*. (2011), a multibreed reference set of Holstein bulls from the *EuroGenomics* project increased accuracy of about 10%.

In this study, accuracy did not improve much when the number of animals in the training set of purebreds was increased substantially by combining breeds. This suggests that, breed relationships are much more important than the number of animals in the training set. Kizilkaya *et al*. (2010), Ibánẽz-Escriche *et al*. (2009), Toosi *et al*. (2010) and Harris *et al*. (2008) all stated that, for a highly diverged population (divergence may be due to different selection schemes for different genetic merit), marker density showed be sufficiently high to achieve accurate predictions of marker effect in increases accuracy of GEBV and is much more important than the number of animals in the training set. This implies that in our current

study marker density should be increased from the current 35,319 SNPs to about 500,000 considering the large effective population size of these breeds.

Moreover, combining distant breeds might increase the number of independent chromosome segment (M_e) (Daetwyler *et al*., 2010). Daetwyler *et al*. (2010) defined accuracy (r) for a GBLUP model as; $r = \sqrt{N_T h^2 / N_T h^2 + M_e}$ and Where M_e (Effective number of chromosome segment) was defined as $M_e = 2N_e L / \log(4N_e L)$; N_T is the number of training animals, h^2 is the heritability of the trait, N_e is the effective population size and L is the genome length in Morgan's. Implicitly, an increase in M_e without increasing the number of training animals in the population will lead to a decrease in accuracy. This might the main reason for the loss in accuracy observed for combining breeds especially in the case of the 3 way multibreed analysis.

5 CONCLUSION

The benefit of combining these breeds into a 2 way multibreed training set was overall 1.9% and a loss of 1.32% in prediction accuracy for a 3 way multibreed training set. There were no significant difference in the methods (GBLUP, Bayes-B and wgt.GBLUP) used, both for the within/single breed and the multibreed analysis. Increasing the number of animals in the reference set did not necessary increase accuracies but breed relatedness or diversity among breeds were much more important in increasing accuracy. The result for this study should be interpreted with caution as the number of animals used in the analysis was limited.

Further studies will be undertaken with 777k imputed SNPs using both family information and 50k SNPs.

6 References

1. Brøndum, R. F., Rius-Vilarrasa, E., Strandén, I., Su, G., Guldbrandtsen, B., Fikse, W. F. and Lund M. S. (2011). ***Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations***. *Journal of Dairy Science*, 94:4700–4707

2. Calus, M.P.L., Meuwissen, T.H.E., De Roos, A.P. and Veerkamp R.F. (2008). ***Accuracy of Genomic Selection Using Different Methods to Define Haplotypes***. *Genetics,* 178: 553-561

3. Daetwyler, H.D., Villanueva, B. and Woolliams, J.A. (2008). ***Accuracy of predicting the genetic risk of disease using a genome-wide approach***. *PLoS One*, 3: 3395

4. Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams, 2010. ***The impact of genetic architecture on genome-wide evaluation methods***. *Genetics,* 185: 1021–1031

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

6. De Roos, A.P.W., Schrooten C. and Druet T. (2011). ***Genomic breeding value estimation using genetic markers, inferred ancestral haplotypes, and the genomic relationship matrix***. *Journal of Dairy Science*, 94:4708–4714

7. De Roos, A.P.W., Hayes, B.J. and Goddard, M.E. (2009). ***Reliability of genomic breeding values across multiple populations***. *Genetics,* 183(4): 1545-53

8. Falconer, D.S. and Mackay, T.F.C. (1996). ***Introduction to quantitative genetics***. Fourth (4th) Edition Longman, UK.

9. Garrick, D.J., Taylor, J.F. and Fernando, R.L. (2009). ***De-regressing estimated breeding values and weighting information for genomic regression analyses***. *Genetics Selection Evolution,* 41: 51

10. Gilmour, A.R., Gogel, B.J., Cullis, B.R. and Thompson, R. (2009). ***ASReml User Guide Release*** 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK

11. Goddard, M.E. and Hayes, B.J. (2009). ***Mapping genes for complex traits in domestic animals and their use in breeding programmes***. *Nature Review Genetics*, 10: 381–391

12. Goddard, M.E. (2009). ***Genomic selection: prediction of accuracy and maximisation of long term response***. *Genetica*, 136(2): 245–257

13. Grisart, B., Farnir F., Karim L., Cambisano N., Kim J.J., Kvasz, A., Mni, M., Simon, P., Fre`re, J-M., Coppieters, W. and Georges, M. (2004). ***Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition***. *Proceeding of the National Academy of Sciences, USA*, 101: 2398– 2403

14. Harris, B.L., Johnson, D.L. and Spelman, R.L. (2008). ***Genomic selection in New Zealand and the implications for national genetic evaluation***. Proc. Interbull Meeting, Niagara Falls, Canada.

15. Hayes, B. and Goddard, M. (2010). ***Genome-wide association and genomic selection in animal breeding***. *Genome,* 53: 876–883

16. Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Verbyla, K. and Goddard, M.E., (2009). ***Accuracy of genomic breeding values in multi-breed dairy cattle populations***. *Genetics Selection Evolution,* 41: 51

17. Hill, W.G. and Robertson, A. (1968). ***Linkage disequilibrium in finite populations***. *Theoretical and Applied Genetics,* 38:226–231

18. Ibánẽz-Escriche, N., Fernando, F.L., Toosi, A. and Dekkers J.C. (2009). ***Genomic selection of purebreds for crossbred performance***. *Genetics Selection Evolution,* 41: 12

19. Kizilkaya, K., Fernando, R.L. and Garrick, D.J. (2010). ***Genomic prediction of simulated multi-breed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes***. *Journal of Animal Science,* 88: 544-551

20. Larmer, S., Schenkel, F. and Sargolzaei, M. (2012) unpublished. ***Assessing the extent of Linkage Disequilibrium and Consistency of Linkage Phase in 5 dairy breeds.*** Update for Dairy Gen Funded Project Sep 13/2011

21. Luan, T., Woolliams, J.A., Lien, S., Kent, M. and Meuwissen, T. H. E. (2009). ***The Accuracy of Genomic Selection in Norwegian Red Cattle Assessed by Cross-Validation***. *Genetics,* 183: 1119–1126

22. Lund, M. S., De Roos, S.P.W, de Vries, A.G., Druet, T., Ducrocq, V., Fritz, S., Guillaume, F., Guldbrandtsen, B., Liu, Z., Reents, R., Schrooten, C., Seefried, F. and Su, G. (2011). ***A common reference population from four European Holstein populations increases reliability of genomic predictions***. *Genetics Selection Evolution,* 43:43

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

24. Moser, G., Tier, B., Crump, R.E., Mehar, S., Khatkar, M.S. and Raadsma H.W. (2009). ***A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers***. *Genetics Selection Evolution,* 41:56

25. Nirea, K.G. (2009). ***Comparison of genomic selection methods using genomic data on Austrian Fleckvieh Cattle***. MSc Thesis, *Department of Animal and Aquacultural Sciences, Norwegian university of Life Sciences, Post Office Box 5003, 1432, Aas Norway*

26. Pryce, J.E., Gredler, B., Bolormaa, S., Bowman, P.J., Egger-Danner, C., Fuerst, C., Emmerling, R., Sölkner, J., Goddard, M. E. and Hayes, B.J. (2011). ***Short communication: Genomic selection using a multi-breed, across-country reference population***. *Journal of Dairy Science,* 94: 2625–2630

27. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. and Sham, P.C. (2007). ***PLINK: a toolset for whole-genome association and population-based linkage analysis***. *American Journal of Human Genetics,* 81

28. R Development Core Team (2011). ***R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria***. URL http://www.r-project.org.

29. Schaeffer, L.R., (2006). ***Strategy for applying genome-wide selection in dairy cattle***. *Journal of Animal Breeding and Genetics,* 123: 218–223

30. Solberg, T.R., Sonesson, A.K., Wolliams, J.A. and Meuwissen, T.H.E., (2008). ***Genomic selection using different marker types and densities***. *Journal of Animal Sciences,* 86: 2447-2454

31. Toosi, A., Fernando, R. L. and Dekkers, J.C.M. (2010). ***Genomic selection in admixed and crossbred populations***. *Journal of Animal Science,* 88:32-46

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

33. VanRaden, P.M., O'Connell, J.R., Wiggans, G.R. and Weigel, K.A. (2011). ***Genomic evaluations with many more genotypes***. *Genetics Selection Evolution,* 43:10

34. VanRaden, P.M. (2008). ***Efficient methods to compute genomic prediction***s. *Journal of Dairy Science*, 91:4414-4423

35. www.rinderzucht-austria.at

36. *www.zar.at/* Zentrale Arbeitsgemeinschaft öesterreichischer Rinderzucht (ZAR), (2009)

37. Yang J., Beben B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.F., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E., Visscher P.M. (2010) Missing heritability of human height explained by genomic relationships. Nat. Genet., 42, 565–569.

Appendix

Table 6: Accuracies of GEBV with their standard errors (subscript) using GBLUP or SNP effect from Bayes-B, and wgt.GBLUP for the traits: Milking Speed (DMG), Protein Kg (EKG), Protein Percent (EP), Fat Kg (FKG), Fat Percent (FP), Fertility Maternal (FRM), Milk Kg (MKG), Longevity (ND), Persistency (PER) and Somatic Cell Count (ZZ) for 2 way cross predictions of Braunvieh, Grauvieh and Pinzgauer breeds

Reference dataset[1]	**Validation**[2] **dataset**	**Method**	**Traits**									
			DMG	**EKG**	**EP**	**FKG**	**FP**	**FRM**	**MKG**	**ND**	**PER**	**ZZ**
Braunvieh and Grauvieh	**Braunvieh**	*GBLUP*	$0.62_{(0.07)}$	$0.34_{(0.13)}$	$0.40_{(0.10)}$	$0.47_{(0.09)}$	$0.43_{(0.09)}$	$0.46_{(0.13)}$	$0.28_{(0.13)}$	$0.41_{(0.10)}$	$0.55_{(0.09)}$	$0.54_{(0.08)}$
		Bayes-B	$0.61_{(0.07)}$	$0.35_{(0.12)}$	$0.46_{(0.11)}$	$0.43_{(0.10)}$	$0.41_{(0.09)}$	$0.46_{(0.12)}$	$0.31_{(0.13)}$	$0.47_{(0.10)}$	$0.58_{(0.09)}$	$0.54_{(0.08)}$
		wgt.GBLUP	$0.61_{(0.07)}$	$0.34_{(0.12)}$	$0.46_{(0.10)}$	$0.47_{(0.09)}$	$0.48_{(0.08)}$	$0.49_{(0.12)}$	$0.32_{(0.13)}$	$0.47_{(0.10)}$	$0.57_{(0.09)}$	$0.55_{(0.08)}$
	Grauvieh	*GBLUP*	$0.62_{(0.08)}$	$0.29_{(0.13)}$	$0.73_{(0.06)}$	$0.36_{(0.11)}$	$0.68_{(0.08)}$	$0.40_{(0.09)}$	$0.50_{(0.10)}$	$0.16_{(0.10)}$	$0.40_{(0.12)}$	$0.69_{(0.07)}$
		Bayes-B	$0.65_{(0.08)}$	$0.32_{(0.12)}$	$0.74_{(0.06)}$	$0.37_{(0.10)}$	$0.65_{(0.08)}$	$0.18_{(0.12)}$	$0.48_{(0.10)}$	$0.14_{(0.12)}$	$0.37_{(0.13)}$	$0.70_{(0.07)}$
		wgt.GBLUP	$0.61_{(0.07)}$	$0.30_{(0.13)}$	$0.75_{(0.06)}$	$0.37_{(0.11)}$	$0.67_{(0.08)}$	$0.39_{(0.09)}$	$0.50_{(0.10)}$	$0.23_{(0.11)}$	$0.38_{(0.12)}$	$0.69_{(0.07)}$
Braunvieh and Pinzgauer	**Braunvieh**	*GBLUP*	$0.63_{(0.07)}$	$0.32_{(0.13)}$	$0.37_{(0.11)}$	$0.45_{(0.09)}$	$0.42_{(0.09)}$	$0.48_{(0.12)}$	$0.25_{(0.13)}$	$0.43_{(0.09)}$	$0.59_{(0.09)}$	$0.55_{(0.08)}$
		Bayes-B	$0.64_{(0.07)}$	$0.32_{(0.13)}$	$0.44_{(0.11)}$	$0.44_{(0.09)}$	$0.43_{(0.09)}$	$0.49_{(0.12)}$	$0.27_{(0.14)}$	$0.48_{(0.11)}$	$0.57_{(0.10)}$	$0.56_{(0.08)}$
		wgt.GBLUP	$0.63_{(0.07)}$	$0.30_{(0.13)}$	$0.47_{(0.10)}$	$0.48_{(0.08)}$	$0.46_{(0.09)}$	$0.51_{(0.11)}$	$0.26_{(0.14)}$	$0.46_{(0.11)}$	$0.56_{(0.10)}$	$0.57_{(0.08)}$
	Pinzgauer	*GBLUP*	$0.48_{(0.11)}$	$0.42_{(0.10)}$	$0.19_{(0.12)}$	$0.57_{(0.09)}$	$0.40_{(0.11)}$	$0.74_{(0.05)}$	$0.40_{(0.10)}$	$-0.09_{(0.12)}$	$0.50_{(0.10)}$	$0.57_{(0.10)}$
		Bayes-B	$0.47_{(0.11)}$	$0.41_{(0.10)}$	$0.18_{(0.12)}$	$0.52_{(0.10)}$	$0.39_{(0.11)}$	$0.77_{(0.05)}$	$0.41_{(0.09)}$	$-0.04_{(0.13)}$	$0.52_{(0.09)}$	$0.56_{(0.10)}$
		wgt.GBLUP	$0.50_{(0.10)}$	$0.36_{(0.11)}$	$0.22_{(0.12)}$	$0.54_{(0.10)}$	$0.48_{(0.10)}$	$0.78_{(0.05)}$	$0.35_{(0.12)}$	$-0.04_{(0.12)}$	$0.51_{(0.09)}$	$0.55_{(0.10)}$
Grauvieh and Pinzgauer	**Grauvieh**	*GBLUP*	$0.61_{(0.08)}$	$0.29_{(0.14)}$	$0.71_{(0.07)}$	$0.33_{(0.12)}$	$0.67_{(0.08)}$	$0.39_{(0.10)}$	$0.42_{(0.11)}$	$0.18_{(0.11)}$	$0.42_{(0.12)}$	$0.71_{(0.06)}$
		Bayes-B	$0.62_{(0.08)}$	$0.30_{(0.14)}$	$0.68_{(0.07)}$	$0.35_{(0.11)}$	$0.63_{(0.09)}$	$0.36_{(0.10)}$	$0.42_{(0.11)}$	$0.19_{(0.12)}$	$0.42_{(0.12)}$	$0.70_{(0.06)}$
		wgt.GBLUP	$0.62_{(0.08)}$	$0.27_{(0.14)}$	$0.70_{(0.07)}$	$0.35_{(0.11)}$	$0.68_{(0.08)}$	$0.42_{(0.10)}$	$0.40_{(0.12)}$	$0.16_{(0.11)}$	$0.39_{(0.13)}$	$0.68_{(0.07)}$
	Pinzgauer	*GBLUP*	$0.51_{(0.11)}$	$0.51_{(0.11)}$	$0.23_{(0.11)}$	$0.62_{(0.08)}$	$0.42_{(0.11)}$	$0.77_{(0.05)}$	$0.47_{(0.10)}$	$-0.14_{(0.13)}$	$0.51_{(0.10)}$	$0.67_{(0.07)}$
		Bayes-B	$0.48_{(0.11)}$	$0.50_{(0.12)}$	$0.23_{(0.12)}$	$0.62_{(0.08)}$	$0.42_{(0.11)}$	$0.78_{(0.05)}$	$0.51_{(0.10)}$	$-0.02_{(0.13)}$	$0.50_{(0.10)}$	$0.65_{(0.07)}$
		wgt.GBLUP	$0.52_{(0.11)}$	$0.48_{(0.11)}$	$0.26_{(0.11)}$	$0.61_{(0.08)}$	$0.43_{(0.11)}$	$0.76_{(0.06)}$	$0.42_{(0.11)}$	$-0.09_{(0.13)}$	$0.50_{(0.11)}$	$0.65_{(0.08)}$
Mean			***0.58***	***0.36***	***0.46***	***0.46***	***0.51***	***0.54***	***0.39***	***0.19***	***0.49***	***0.62***

[1] *The training dataset contains the full number of bulls in each breed if that breed is not used as the validation dataset*

[2] *The validation dataset contains the youngest 60 bulls in each breed except in Grauvieh for the traits **FRM** and **ND** where the youngest 30 bulls are used*

Table 7: Accuracies of GEBV in a 3 way cross predictions using GBLUP or SNP effect from Bayes-B and wgt.GBLUP for the traits: Milking Speed (DMG), Protein Kg (EKG), Protein Percent (EP), Fat Kg (FKG), Fat Percent (FP), Fertility Maternal (FRM), Milk Kg (MKG) , Longevity (ND), Persistency (PER) and Somatic Cell Count (ZZ) of Braunvieh, Grauvieh and Pinzgauer breeds

Reference dataset[1]	**Validation**[2] **dataset**	**Method**	**Traits**									
			DMG	**EKG**	**EP**	**FKG**	**FP**	**FRM**	**MKG**	**ND**	**PER**	**ZZ**
Braunvieh + Grauvieh + Pinzgauer	**Braunvieh**	*GBLUP*	$0.61_{(0.07)}$	$0.32_{(0.13)}$	$0.36_{(0.10)}$	$0.45_{(0.09)}$	$0.43_{(0.08)}$	$0.47_{(0.12)}$	$0.27_{(0.13)}$	$0.42_{(0.10)}$	$0.57_{(0.09)}$	$0.54_{(0.08)}$
		Bayes-B	$0.60_{(0.07)}$	$0.32_{(0.13)}$	$0.46_{(0.11)}$	$0.43_{(0.09)}$	$0.42_{(0.09)}$	$0.46_{(0.12)}$	$0.26_{(0.13)}$	$0.49_{(0.10)}$	$0.56_{(0.09)}$	$0.56_{(0.08)}$
		wgt.GBLUP	$0.61_{(0.07)}$	$0.30_{(0.13)}$	$0.41_{(0.10)}$	$0.45_{(0.09)}$	$0.50_{(0.07)}$	$0.49_{(0.12)}$	$0.29_{(0.13)}$	$0.47_{(0.10)}$	$0.57_{(0.09)}$	$0.55_{(0.08)}$
	Grauvieh	*GBLUP*	$0.61_{(0.08)}$	$0.28_{(0.13)}$	$0.71_{(0.07)}$	$0.35_{(0.10)}$	$0.64_{(0.08)}$	$0.35_{(0.11)}$	$0.47_{(0.11)}$	$0.20_{(0.10)}$	$0.41_{(0.12)}$	$0.68_{(0.07)}$
		Bayes-B	$0.64_{(0.07)}$	$0.31_{(0.12)}$	$0.75_{(0.06)}$	$0.43_{(0.10)}$	$0.59_{(0.10)}$	$0.32_{(0.11)}$	$0.46_{(0.11)}$	$0.20_{(0.11)}$	$0.42_{(0.12)}$	$0.70_{(0.07)}$
		wgt.GBLUP	$0.62_{(0.08)}$	$0.29_{(0.12)}$	$0.73_{(0.07)}$	$0.38_{(0.10)}$	$0.64_{(0.08)}$	$0.32_{(0.11)}$	$0.44_{(0.11)}$	$0.25_{(0.10)}$	$0.39_{(0.13)}$	$0.70_{(0.07)}$
	Pinzgauer	*GBLUP*	$0.48_{(0.11)}$	$0.45_{(0.10)}$	$0.21_{(0.12)}$	$0.59_{(0.09)}$	$0.40_{(0.11)}$	$0.72_{(0.06)}$	$0.43_{(0.10)}$	$-0.07_{(0.13)}$	$0.49_{(0.10)}$	$0.59_{(0.09)}$
		Bayes-B	$0.43_{(0.11)}$	$0.43_{(0.10)}$	$0.21_{(0.12)}$	$0.57_{(0.09)}$	$0.35_{(0.12)}$	$0.72_{(0.06)}$	$0.36_{(0.10)}$	$-0.02_{(0.13)}$	$0.48_{(0.10)}$	$0.60_{(0.09)}$
		wgt.GBLUP	$0.47_{(0.11)}$	$0.41_{(0.11)}$	$0.21_{(0.11)}$	$0.56_{(0.10)}$	$0.43_{(0.10)}$	$0.73_{(0.06)}$	$0.40_{(0.11)}$	$-0.02_{(0.13)}$	$0.50_{(0.09)}$	$0.57_{(0.10)}$
Mean			***0.56***	***0.35***	***0.45***	***0.47***	***0.49***	***0.51***	***0.38***	***0.21***	***0.49***	***0.61***

[1] *The training dataset contains the full number of bulls in each breed if that breed is not used as the validation dataset*

[2] *The validation dataset contains the youngest 60 bulls in each breed except in Grauvieh for the traits **FRM** and **ND** where the youngest 30 bulls are used*

Table 13: Correlation of 3 prediction methods (GBLUP, Bayes-B and wgt.GBLUP) for GEBV of the selected traits Fertility, Milk yield and Somatic cell Count (SCC) for the purebred

Trait	*Method*	*REF- Braunvieh*		*REF- Grauvieh*		*REF- Pinzgauer*	
		VAL (Braunvieh)		*VAL (Grauvieh)*		*VAL (Pinzgauer)*	
		Bayes-B	*wgt.GBLUP*	*Bayes-B*	*wgt.GBLUP*	*Bayes-B*	*wgt.GBLUP*
	GBLUP	0.93	0.99	0.98	0.70	1.00	0.99
Fertility	*Bayes-B*		0.96		0.73		0.99
	GBLUP	0.98	1.00	0.99	0.99	0.99	0.99
Milk *Kg*	*Bayes-B*		0.99		1.00		0.98
	GBLUP	0.99	0.99	0.97	0.98	0.99	0.99
SCC	*Bayes-B*		1.00		1.00		0.99

BV – Braunvieh GV - Grauvieh PI – Pinzgauer
REF – Training dataset VAL – Validation dataset

Table 14: Correlation of 3 prediction methods (GBLUP, Bayes-B and wgt.GBLUP) for GEBV of the selected traits Fertility, Milk yield and Somatic cell Count for selected two way multibreed GS

Trait	*Method*	*REF (BV+GV)*		*REF (GV+BV)*		*REF (PI+GV)*	
		VAL (Braunvieh)		*VAL (Grauvieh)*		*VAL (Pinzgauer)*	
		Bayes-B	*wgt.GBLUP*	*Bayes-B*	*wgt.GBLUP*	*Bayes-B*	*wgt.GBLUP*
	GBLUP	0.93	0.93	0.74	0.87	0.96	0.95
Fertility	*Bayes-B*		0.98		0.78		0.98
	GBLUP	0.96	0.97	0.98	0.98	0.98	0.96
Milk *Kg*	*Bayes-B*		0.98		0.97		0.95
	GBLUP	0.99	0.98	0.98	0.96	0.99	0.98
SCC	*Bayes-B*		0.99		0.98		0.99

BV – Braunvieh GV - Grauvieh PI – Pinzgauer
REF – Training dataset VAL – Validation dataset

Table 15: Correlation of 3 prediction methods (GBLUP, Bayes-B and wgt.GBLUP) for GEBV of the selected traits Fertility, Milk yield and Somatic cell Count for three way multibreed GS

Trait	*Method*	*REF (BV+PI+GV)*		*REF (GV+BV+PI)*		*REF (PI+BV+GV)*	
		VAL (Braunvieh)		*VAL (Grauvieh)*		*VAL (Pinzguaer)*	
		Bayes-B	*wgt.GBLUP*	*Bayes-B*	*wgt.GBLUP*	*Bayes-B*	*wgt.GBLUP*
	GBLUP	0.94	0.92	0.95	0.88	0.96	0.95
Fertility	*Bayes-B*		0.99		0.95		0.98
	GBLUP	0.98	0.97	0.95	0.97	0.97	0.96
Milk Kg	*Bayes-B*		0.98		0.96		0.97
	GBLUP	0.99	0.98	0.98	0.97	0.98	0.97
SCC	*Bayes-B*		0.99		0.99		0.98

BV – Braunvieh GV - Grauvieh PI – Pinzgauer
REF – Training dataset VAL – Validation dataset

Table 16: Accuracies of GEBV with their standard errors (subscript) using GBLUP or SNP effect from Bayes-B, and wgt.GBLUP for the traits: Milking Speed (*DMG*), Protein Kg (*EKG*), Protein Percent (*EP*), Fat Kg (*FKG*), Fat Percent *(FP*), Fertility Maternal (*FRM*), Milk Kg (*MKG*), Longevity (*ND*), Persistency (*PER*) and Somatic Cell Count (*ZZ*) for single and Multibreed predictions

REF[1]	VAL[2]	DMG	EKG	EP	FKG	FP	FRM	MKG	ND	PER	ZZ
Bruanvieh	BV	0.63	0.35	0.42	0.48	0.41	0.48	0.27	0.44	0.58	0.55
Grauvieh	GV	0.62	0.31	0.70	0.34	0.70	0.31	0.44	0.13	0.42	0.71
Pinzgauer	PI	0.52	0.44	0.22	0.57	0.42	0.76	0.40	0.02	0.50	0.64
Bruanvieh	BV	0.61	0.34	0.39	0.46	0.33	0.44	0.30	0.40	0.57	0.56
Grauvieh	GV	0.63	0.29	0.73	0.36	0.65	0.30	0.50	0.15	0.39	0.69
Bruanvieh	BV	0.62	0.33	0.37	0.43	0.37	0.43	0.29	0.44	0.58	0.54
Pinzgauer	PI	0.47	0.42	0.19	0.53	0.39	0.66	0.37	-0.10	0.53	0.52
Grauvieh	GV	0.65	0.29	0.70	0.33	0.69	0.41	0.41	0.17	0.37	0.69
Pinzgauer	PI	0.50	0.45	0.18	0.54	0.39	0.71	0.38	-0.06	0.45	0.62
Bruanvieh	BV	0.60	0.32	0.37	0.40	0.37	0.39	0.29	0.43	0.59	0.55
Grauvieh	GV	0.63	0.25	0.70	0.32	0.65	0.22	0.44	0.16	0.39	0.67
Pinzgauer	PI	0.46	0.43	0.14	0.53	0.35	0.64	0.37	-0.05	0.50	0.52

BV – Braunvieh GV - Grauvieh PI – Pinzgauer
REF – Training dataset VAL – Validation dataset