University of Natural Resources
and Life Sciences, Vienna

Department of Sustainable
Agricultural Systems
Division of Livestock Sciences

Genomic Selection Using Small SNP Subsets

Mahlet Teka Anche

Supervisor
Univ. Prof. Dr. Johann Sölkner

Co - Supervisor
Dr. Birgit Gredler
PD. Dr. Birgit Furst-Waltl

Vienna, June 2011

Acknowledgement

First of all I would like to express my deepest gratitude all My supervisors, Prof. Johann, Dr. Birgit Furst Waltl, Dr. Birgit Gredler, I thank you very much for everything. Your support and guidance allow me to face what I thought I can't face and work it through.

My friends thank you for putting a tear of smile on my face, like last night!

My boyfriend, Dessu, thank you very much, you raise me up! You bring out the best of me which I didn't realize I had it before. Amesegenalehuegn!

I'm heartily thankful for the European Union and for Erusmus Mundus for the great opportunity that I offered. If it wasn't for this great chance, I wouldn't be able to meet all the nice person I have meet and all my friends I have make.

Last but not least I would like to thank Zuchtdata for providing me the data for my thesis.

- **Genomic Selection Using Small SNP data**

Anche, M.T[1]., Gredler, B[1,2]., Sölkner, J[1].

[1] University of Natural Resources and Life Sciences, Vienna, Department of Sustainable Agricultural Systems, Division of Livestock Sciences, Vienna

[2] Qualitas AG, Chamerstrasse 56, 6300 Zug, Switzerland

Abstract

Coupling of the available high-density single nucleotide polymorphisms (SNP) chips with efficient genotyping improves the genetic progress that can be achieved by genomic selection compared to the traditional selection method. However, genotyping of all selection candidates with high-density SNP may not be cost effective. Small subset of SNP selected from the high-density SNP chip can be used for prediction of direct genomic breeding value (DGV) for each trait under selection. Here, we attempt to compare the accuracy of DGV using small subset of SNP and the full set of SNP. Subsets comprising different number of SNP were selected from Illumina Bovine 50k Beadchip. SNP were selected based on their absolute effect size and also randomly from the full set of SNP. De-regressed breeding values for fat percentage, protein yield and calving ease for 5556 dual purpose Fleckvieh bulls were used as phenotypic records. We found that BayesB slightly outperformed SNP-BLUP for all traits. SNP selected based on their absolute effect size gave an accuray within the range of 20 % to 64 % and from 20 % to 75 % with SNP-BLUP and BayesB approach, respectively. Whereas an accuracy of 35 % to 71 % (SNP-BLUP) and 35% to 75% (BayesB) were obtained with the randomly selected subsets. GLMSELECT selection procedure resulted in lowest accuray using 100 of the most important SNP regardless of the methods used for prediction of DGV.

Key words: Genomic selection, SNP subsets, Accuracies.

Introduction

The use of molecular markers for the prediction of genetic merit provides faster genetic gain compared to traditional selection schemes only (Meuwissen *et al*. 2001). In the traditional selection strategy, pedigree information and phenotypic records are the basic information sources for the prediction of future genotypic merit of livestock. With the advent of high throughput genotyping technologies to genotype 1000s of SNP, it is now possible to include information about the variation of DNA sequences between individuals in the estimation of genetic merits, which makes the prediction of breeding values more accurate (Vazquez *et al.* 2010). The principle of genomic selection was first proposed by Meuwissen *et al.* (2001), in which markers that are in linkage disequilibrium with the actual gene that affect the trait are used to estimate DGV.

Genomic selection is based on the estimation of DGV. These are estimated as the sum of SNP effects of single markers or haplotypes within a chromosomal segment (Solberg *et al*. 2008)**.** These single genetic markers or haplotypes are assumed to explain most of the genetic variance contributed by the quantitative trait loci (QTL) (Hayes *et al.* 2009). Unavailability of dense marker arrays and the higher cost associated with the genotyping these markers have been the major obstacles for the implementation of genomic selection (Meuwissen *et al*. 2001). Currently, the advancement in the field of molecular genetics' technology enables the wide range application of high density marker chips for selection of livestock species based on their genomic merit.

For the implementation of genomic selection reference populations that are phenotyped and genotyped for large sets of markers should be available. The reference populations are used to derive the prediction equation for the estimation of DGV (Moser *et al*. 2010). The prediction

equation is essentially a set of estimated marker-effects that can be used to predict DGV for non-phenotyped individuals based on their genotype. The validation population having both, genotypic and phenotypic information is then used to test the prediction equation for its accuracy. In the following, DGV can be estimated for the selection candidates that have only genotypic information for a given set of markers (Goddard *et al.,* 2007). It was found that accuracy of breeding value that was estimated using only marker data on a simulated data can reach up to 85 % (Meuwissen *et al*. 2001). This will reduce the cost for performance testing or progeny testing of the selection candidates for the traits of interest.

Illumina Inc. currently launched available the BovineHD BeadChip (Illumina. Inc. 2010) with more than 777,000 SNP in addition to the Bovine3K (Illumina. Inc. 2010) and 50K (Illumina. Inc. 2011) BeadChips. Even though the advancement in molecular technology makes it possible to use these high-density SNP markers in genomic selection, there will be a higher cost for genotyping all valuable animals in the population. Therefore, using low density markers for the prediction of DGV becomes an alternative strategy to lower the cost of genotyping a large population.

Different subsets of SNP have been used for estimation of DGV. Depending on the methods used for estimation of DGV, it is possible to achieve a reliable accuracy with small subsets of SNP. Accuracy above 60% was obtained for fat % when using 2000 of SNP with highest effect (Hayes *et al.,* 2010). Moser *et al.,* (2010) used subsets of higher ranking SNP within the range of 100 to 20,000 SNP and lower accuracy was found with small set of SNP. But when increasing the size of SNP from 1000 – 40000, the increase in the accuracy of DGV was insignificant.

This suggests that, selection of the most important markers from the high-density SNP markers is required in order to obtain a reliable accuracy as compared to high-density SNP markers.

The objective of this study was to investigate the accuracy of DGV prediction with subsets of SNP selected based on random selection, their absolute effect size and choice in preliminary runs of LASSO (least absolute shrinkage and selection operator) variable selection for their prediction ability of DGV using BayesB and SNP-BLUP in Fleckvieh (dual purpose Simmental) cattle.

Material and methods

Material

Phenotypic and genotypic data was available for 5556 Fleckvieh (dual purpose Simmental) bulls. Phenotypic and genotypic information were provided by Zuchtdata EDV-Dienstleistungen GmbH, being responsible for the genetic evaluation in Austrian cattle. The bulls were genotyped for the Illumina Bovine 50K Beadchip. After passing quality check procedures such as Minor Allele Frequency (MAF) > 0.5 %, SNP call rate also should be > 75%, GC-score > 0.2, pedigree checking and replacing missing genotypes by average allele frequencies and only 41008 SNP were included for the estimation of direct DGV. De-regressed breeding values for fat percentage, protein yield and male calving ease were used as response variables for estimation of DGV in this study. De-regression of estimated breeding values was carried out according to Garrick et al. (2009) in order to account for the difference in the number of progeny record that exists between bulls. Bulls were split into reference and validation population according to birth year. Bulls born before 2003 were assigned to form the reference population whereas bulls born between 2003 and 2005 were in the validation population. The distribution of bulls across birth year is show in Figure 1. The number of bulls in the reference and validation population, respectively, is given in Table 4.

Method

SNP selection

Subsets of SNP were selected from the full set of SNP applying different SNP selections strategies. SNP were selected randomly and based on the absolute BayesB effect size. In addition 100 SNP were selected using the LASSO variable selection procedure implemented in the SAS 9.2 procedure GLMSELECT (SAS 2009) Institute Inc. (2009) SAS/STAT® User's Guide Version 9.2). The number of SNP selected within each selection strategy is given in Table 2. All subsets selected randomly included the SNP included on the Illumina Bovine 3K Bead Chip. For selecting SNP based on their absolute BayesB effect size, SNP effects of the full set of SNP were first estimated using BayesB. The SNP were ranked based on their effect size for each trait trait. Subsets included 100, 300, 500, 1000, and 3000 SNP with highest BayesB SNP effects (Table 2)

Only 100 of SNPs were selected based on the LASSO variable selection procedure (Tibshirani, 1996). This method of model selection adds and deletes parameters based on a version of ordinary least square where the sum of the absolute regression coefficient is constrained.

Prediction of DGV

SNP-BLUP (Meuwissen et al. 2001) and BayesB (Meuwissen 2009) were applied for the estimation of SNP effects and the prediction of DGV. In both approaches, the effect of SNP were estimated with the following model:

$$\mathbf{Y} = \boldsymbol{\mu}\mathbf{1}_n + \sum \mathbf{X}_i \hat{\mathbf{g}}_i + \mathbf{e}$$

Where y is the data vector for the traits being analyzed; μ is the overall mean; 1_n is a vector of n ones (n = number of records); g_i is the effect of the i^{th} SNP and X_i is the design matrix that relates the genotype of SNP to all individuals.

BayesB assumes the variance of SNP effect $\delta^2_{gi,}$ to be variable across the genome and comes from a prior distribution (Meuwissen *et al.* 2001). Prior distribution for the variance of SNP effects has an inverted chi-square distribution with a scaled parameter S and number of degrees of freedom v. In accordance to many studies (e.g. Hayes *el al*. 2010), the scaled parameter S = 4.012 was assumed. For the variance associated with the effect of chromosome segment different prior probabilities of 0.5, 0.3, 0.1, 0.01 and 0.001 were considered. A Gibbs chain length of 30,000 cycles was run where the first 10,000 cycles were discarded as burn in. In SNP-BLUP the variance associated with SNP effects is assumed to be the same across the chromosom segement.

Inorder to cross check the prediction ability of SNP-BLUP and BayesB, 100 SNP with highest effect on fat percentage resulted from SNP-BLUP was analyzed with BayesB for estimation of DGV. Vice versa, 100 SNP with highest effect on fat percentage resulted from BayesB approach were analyzed with SNP-BLUP approach for estimation of DGV.

Given the estimates for SNP effect and variance, DGV were calculated as

$$\mathbf{DGV = X \hat{g}}$$

Where **X** is the design matrix that relates the genotype of SNP to all individuals, and $\hat{\mathbf{g}}$ is the estimated SNP effect.

The correlation between the estimated DGV and de-regressed breeding value is calculated as a measure of accuracy of genomic selection with SAS9.2 software package (SAS 2009) Institute Inc. (2009) SAS/STAT® User's Guide Version 9.2).

Results

SNP selection and estimation of SNP effects

Different sizes of SNP subsets were selected based on their absolute effect, importance (LASSO) and randomly from the full set of SNP (Table 2). Within the randomly selected subsets of SNP a large overlap was observed due to the fact that 3K Chip SNP are included in all of the subsets (Table 3). SNP selected based on their absolute effect size were located on different chromosomes. Higher ranking SNP for fat percentage were located on chromosome 14. The top five SNP with highest effect on protein yield and calving ease were located on chromosomes 29, 15, 19, 18, and 16 and on chromosomes 8, 2, 22, 13, and 4, respectively (result not shown).

Accuracy of DGV using randomly selected SNP

The accuracy of genomic selection using different subsets of randomly selected SNP applying BayesB and SNP-BLUP was estimated for fat percentage, protein yield and calving ease for bulls in the validation set. A BayesB result using different subsets of SNP and at prior probabilities of 0.01 and 0.3 for all traits is shown in Table 5.

The accuracy of DGV when using subsets of randomly selected SNP was compared to the accuracy achieved using the full set of SNP markers. Regardless of the methods used for the estimation of SNP effects, the full set of markers including 41008 SNP gave a slightly higher accuracy compared to all other subsets. With the SNP-BLUP approach, using the full set of markers, an accuracy of 0.63 was found for fat percentage, 0.52 protein yield 0.56 and calving ease.

With BayesB approach, an accuracy of 0.70, 0.52, and 0.55 was obtained for fat percentage, protein yield and calving ease respectively at a prior probability of 0.3.

When applying the BayesB approach, we separately observed the change in the accuracy of DGV at a prior probability of 0.01 for all of the traits analyzed. A negligible increase in the accuracy of DGV was obtained with increasing the SNP size from 3000 to 41008. For fat percentage, an accuracy of 0.58 was found when using 3000 SNP and a value of 0.75 with the full set of SNP at the prior probability of 0.01. For protein yield and calving ease, the accuracy of DGV ranged from 0.35 to 0.52 and 0.42 to 0.60, respectively, at the above mentioned prior probability.

Within the different prior probabilities, the accuracy of genomic selection tends to show a slight decrease when increasing the probabilites from 0.001 to 0.5 the accuracy was decreased from 0.58 to 0.55 for fat percentage with 3K Chip. With the full set of SNP, the accuracy of genomic selection decreased from 0.75 to 0.71 for the same triat. Whereas for protein yield and calving ease, no siginificant decrease was observed across the different prior probabilities. And accuracy of DGV at a prior probability of 0.01 is given in Figure 2 separately for all the traits under study.

In gengeral, Protein yield was found to have the lowest accuracy across the different selected subsets selected randomly (Figure 2) and based on their absolute effect size (Table 5) as compared to fat percentage and calving ease .

Accuracy of DGV using SNP selected based on absolute effect size

The number of SNP selected according to absolute effect size within each subset is given in Table 2. The accuracy of DGV showed slightly wider range for subsets of SNP selected based on their absolute values compared to those randomly selected. Using SNP-BLUP for fat percentage the accuracy was 0.57 when using 100 SNP with the highest absolute effect. Accuracies of 0.20 and 0.41 were found for calving ease and protein yield using SNP-BLUP with 100 of most effective SNP (Figure 4).

Using BayesB, the accuracy of DGV showed no difference for different prior probabilities cosidered at lower densities of SNP. For the full set of SNP for fat percentage, the accuracy of DGV decreases from 0.75 to 0.71 with the increase in the prior probability from 0.001 to 0.5. Fat percentage is found to have the highest accuracy of DGV compared to protein yield and calving ease for all subsets. Generally, the change in the accuracy of DGV with the increase in SNP size ranges from 0.61 to 0.75 for fat percentage , from 0.20 to 0.53 for protein yield and from 0.37 to 0.60 for calving ease, respectively, across the different prior probabilities considered. Since no significant difference was found in the accuracy of DGV across the different prior probabilities considered, accuracy only at a prior probability of 0.001 is given in Table 5.

When selecting 100 SNP based on highest effect for fat percentage from the SNP-BLUP, gave accuracy of 0.61 with BayesB approach and 0.55 with SNP-BLUP approach. Whereas with 100 SNP that have highest effect on fat percentage resulting from BayesB approach, the same accuracy of genomic selection was obtained irrespective of the methods used for prediction of DGV (result not shown).

Finally, with LASSO procedure was applied to select 100 of the most important SNP for the trait fat percentage only. A maximum value of 0.28 was obtained for the accuracy of genomic selection across the different prior probabilities considered with the BayesB approach. With the SNP-BLUP approach, an accuracy of 0.29 was found with 100 of most important SNP (result not shown). It was found that only 16 SNP were found in common between those selected based on BayesB approach and LASSO. Whereas between SNP-BLUP and LASSO, there were only 21 SNP overlap. Between BayesB and SNP-BLUP selected SNP subsets, 30 SNP were found in common.

Discussion

According to Hayes *et al.* (2010), the density of markers, the number of loci affecting the trait and the distribution of their effect are some of the factors that affect the accuracy of DGV. Though the main objective of our study was to investigate the prediction ability of different subsets of SNP, the effect of all the above mentioned factors was also reflected.

Even though the traits analyzed and the methods used differed between studies, the results obtained in the current study are in agreement with previous reports regarding SNP size reduction and its consequence on the accuracy of genomic selection (Weigel *et al*. 2009, Moser *et al*.2010, Vazqez *et al*.2010).

Accuracy of DGV between subsets of SNP

The difference in the accuracy of DGV between subsets of SNP varies between randomly selected subsets and SNP that were selected based on their absolute effect size. This might be due to the fact that different sizes of SNP were selected based on different methods. Higher accuracy of genomic prediction was obtained with larger sets of SNP. For fat percentage, 100 of higher ranking SNP were sufficient enough to achieve an accuracy of 0.61. Fewer loci affect the trait fat percentage and thus fewer markers are required to explain much of the total

genetic variance that is required to achieve the maximum accuracy. With protein yield and calving ease, slightly lower values were obtained at lower densities (Table 5). The accuracy obtained with 1000 informative SNP for fat percentage was in correspondence with 20,000 SNP for fat percentage. This indicates that few informative SNP that are closer to the gene DGAT1 gene are responsible for such a variation in the accuracy of DGV. 10000-20000 SNP for protein yield and 5000-10000 SNP for calving ease that are selected randomly.

As presented in Table 3, an overlap of SNP between randomly selected SNP subsets increases with the increasing of SNP size. This might be one of the reasons for the accuracy of genomic selection to be much more consistent with the increasing size of randomly selected SNP subsets. Overlapping of SNP causes most of SNP to be commonly shared between different subsets that are used for prediction of DGV across the different subsets. But it can also be due to the increase in information as a result of increased size of SNP set.

Difference between SNP-BLUP and BayesB approach in their prediction of DGV

Between the two approaches used for prediction of DGV, the accuracy only slightly differed. In general, the BayesB approach slightly outperformed the SNP-BLUP approach. This result is in agreement to what is found by many studies (Meuwissen *et al*. 2001, Verbyla *et al*. 2010, Hayes *et al.* 2010). With different subsets of SNP and SNP selection procedures, different results were obtained using the two approaches. Regarding SNP-BLUP, lower values for the accuracy of DGV might have resulted from the assumption of normal distribution for the variance associated with the SNP effect and not including prior knowledge about the effect of markers.

Using BayesB, higher accuracies were obtained for all different subsets compared to the SNP-BLUP approach. For fat percentage, the difference between the two approaches across the different subsets selected randomly and based on their absolute effect size is very negligible.

But this difference between the two approaches becomes very insignificant and both methods achieve similar accuracies at higher densities for all traits. Daetwyler *et al.* (2010) also showed that at higher density of SNP markers, the difference between the two methods become smaller and eventually the same accuray will be obtained.

With regard to SNP selection method, instead of selecting SNP based on their absolute effect size, selecting based on the genetic variance the markers explain would have increased the accuracy of DGV. Increasing the number of genotyped individuals would also be another option to increase the accuracy of DGV (Moser *et al.* 2010).

Conclusion

In genomic selection the use of small density SNP panels becomes an alternative strategy to reduce the cost of genotyping. I this study, smaller sets of SNP that are selected either randomly or based on their absolute effect size give a lower accuracy compared to the full SNP panel of the 50K Bovine BeadChip but differences are becoming marginal for relatively large subsets. The commercially available chip with ~3000 markers is not a viable option per se but imputation of 50K chip from genotypes on the 3K chip could be an alternative strategy for wider application of genomic selection in all commercial farms (Weigel et al.2010). SNP subsets between 20,000 and 30,000 provide an accuracy with is very close to that achieved with full set of 41,008 SNP. This indicates that increasing the SNP subsets from 50K to 777K might not substantially increase the accuracy of genomic selection if not coupled with advanced statistical methods that are not available yet.

Table 1. Number of bulls (N), means and standard deviation (std.dev) for fat %, protein yield and calving ease.

	Validation data set			Reference data set		
Traits	**N**	**mean**	**Std.dev**	**N**	**mean**	**Std.dev**
Fat %	1549	-0.03	0.22	3731	0.20	0.24
Protein yield	1549	9.73	13.87	3731	-5.05	16.47
Calving ease	1800	101.17	12.40	3756	99.56	11.58

Table 2. Number of SNPs selected randomly, based on their absolute effect and importance (GLMSELECT)

Method of selection	**No. of SNPs**
Random selection	3000, 5000, 10,000, 20,000 and 30,000
Based on their absolute effect size	100, 300, 500, 1000, and 3000
GLMSELECT	100 (for fat % only)

Table 3. Number of SNPs that overlap between different subsets of randomly selected SNPs

SNP subsets		size of overlapping SNPs
5000	10,000	3305
	20,000	3816
	30,000	4404
10,000	20,000	6115
	30,000	7931
20,000	30,000	15,106

Table 4. Mean, standard deviation(std. dev), minimum(min) and maximum(max) of reliabilities of Estimated breeding values (EBVs)

	Reference data set					Validation data set				
Trait	N	mean	Std.dev	Min	max	N	mean	Std.dev	Min	max
Fat percent	3731	93.69	0.52	74.00	99.00	1549	87.97	5.12	70.00	96.00
Protein yield	3731	93.69	3.52	74.00	99.00	1549	87.97	5.12	70.00	96.00
Calving ease	3756	92.02	4.69	34.00	99.00	1800	89.21	3.17	70.00	99.00

Table 4. Correlation between de-regressed EBV and DGVs for the validation populations based on Bayesian approaches derived from different SNP subsets selected randomly from the full set of SNPs for fat percentage at a prior probability of 0.001. 0.01, 0.1, 0.3, and 0.5

	Fat percentage		Protein yield		Calving ease	
SNP size	0.01	0.3	0.01	0.3	0.01	0.3
3000	0.58	0.56	0.35	0.35	0.42	0.41
5000	0.60	0.58	0.40	0.40	0.50	0.47
10000	0.65	0.64	0.45	0.44	0.54	0.51
20000	0.72	0.70	0.50	0.50	0.57	0.54
30000	0.74	0.70	0.51	0.51	0.60	0.60
41008	0.75	0.70	0.52	0.52	0.59	0.56

Table 5. Correlation between de-regressed EBV and DGVs for the validation populations according to BayesB approaches derived from different SNP subsets selected based on their absolute effect size for fat %, protein yield and calving ease at a prior probability of 0.001.

SNP size	Fat %	Protien yield	Calving ease
100	0.61	0.32	0.40
300	0.68	0.37	0.43
500	0.68	0.43	0.45
1000	0.71	0.48	0.50
3000	0.72	0.53	0.60
41008	0.75	0.20	0.37

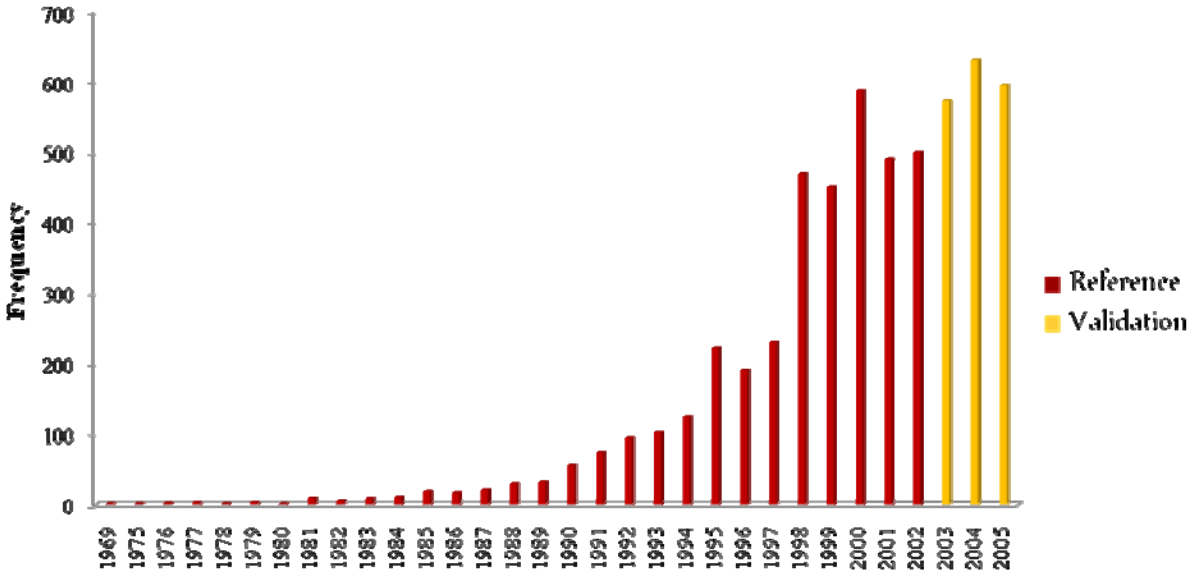

Figure 1. Distribution of all 5556 bulls across the birth year in the reference and validation data set.

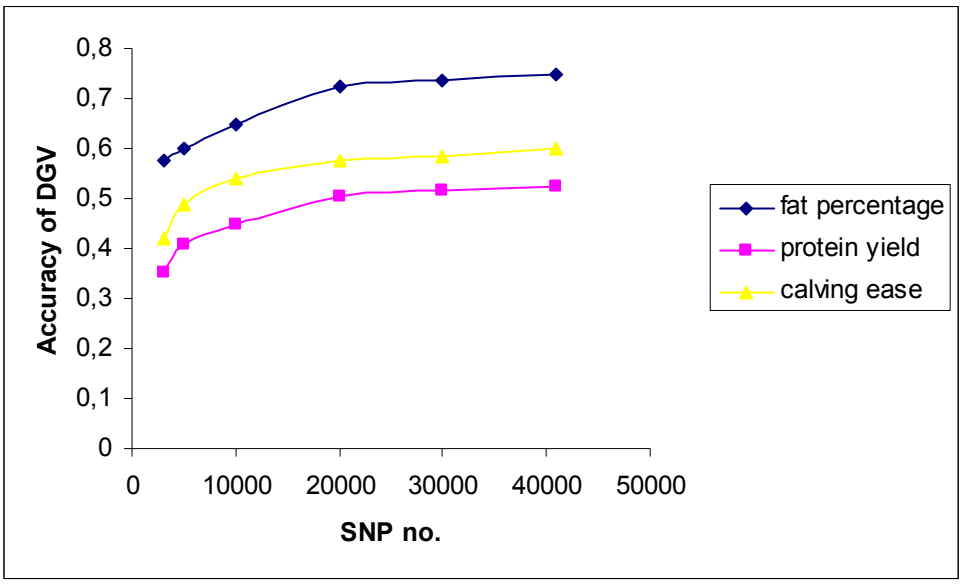

Figure 2. Accuracy of genomic breeding value for fat percentage, protein yield and calving ease derived from BayesB approach at a prior probability of 0.01 for fat percentage, protein yield and calving ease.

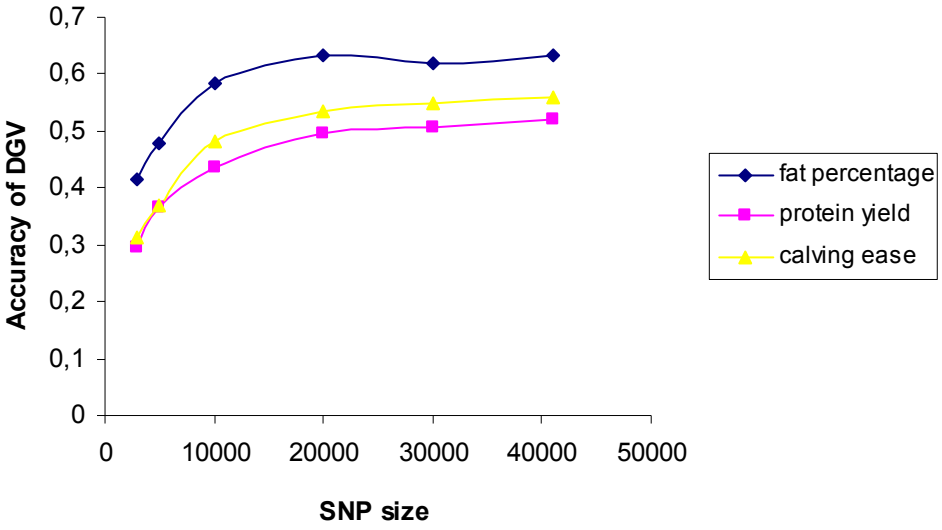

Figure 3. Accuracy of genomic breeding value for fat percentage, protein yield and calving ease derived from the SNP-BLUP approach when SNPs are selected randomly from the full set.

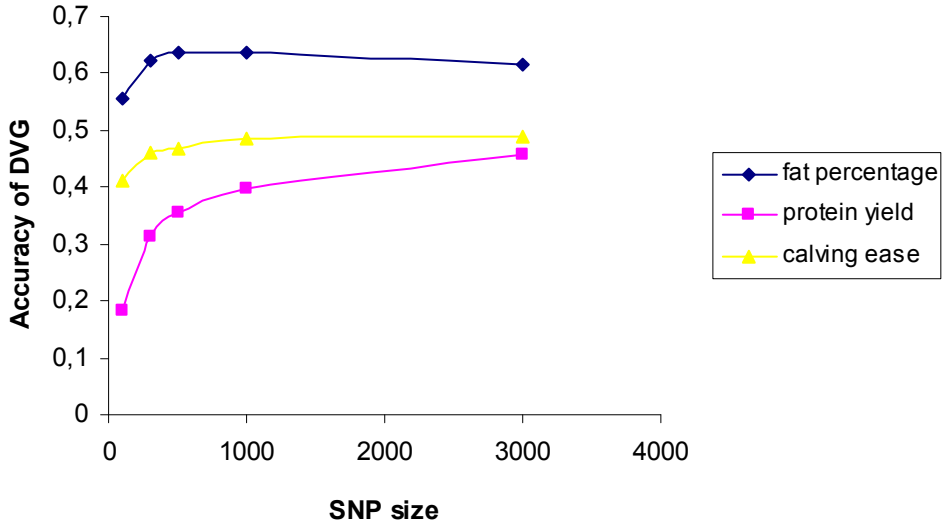

Figure 4. Accuracy of genomic breeding value for fat percentage, protein yield and calving ease derived from the SNP-BLUP approach when SNPs selected based on their absolute effect size.

References

BovineHD Genotyping BeadChip. Illumina 2010.

BovineSNP50 Genotyping BeadChip. Illumina. Inc. 2011.

Daetwyler, H.D., Pong-Wong, R., Villanueva, B., Woolliams, J.A., 2010. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. Genetics 185, 1021-1031.

Garrick, D., Taylor, J., Fernando, R., 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution 41, 55.

Goddard, M.E., Hayes, B.J., 2007. Genomic selection. Journal of Animal Breeding and Genetics 124, 323-330.

Golden Gate Bovine3K Genotyping BeadChip. Illumina Inc. 2010.

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

Hayes, B.J., Pryce, J., Chamberlain, A.J., Bowman, P.J., Goddard, M.E., 2010. Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-Fat Percentage, and Type in Holstein Cattle as Contrasting Model Traits. PLoS Genet 6, e1001139.

Institute Inc. (2009) SAS/STAT® User's Guide Version 9.2).

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157, 1819-1829.

Moser, G., Khatkar, M., Hayes, B., Raadsma, H., 2010. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. Genetics Selection Evolution 42, 37.

Steinbock, L.,Nasholm, B.,Berglund, B.,Johansson, K.,Philipsson, J.,2003 Genetic effect on stillbirth and calving difficulty in Swedish Holsteins at first and second calving. Journal of Dairy Science 86, 2228-2235.

Solberg, T.R., Sonesson, A.K., Woolliams, J.A., Meuwissen, T.H.E., 2008. Genomic selection using different marker types and densities. J. Anim Sci. 86, 2447-2454.

Vazquez, A.I., Rosa, G.J.M., Weigel, K.A., de los Campos, G., Gianola, D., Allison, D.B., 2010. Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins. Journal of Dairy Science 93, 5942-5949.

Verbyla, K., Bowman, P., Hayes, B., Goddard, M., 2010. Sensitivity of genomic selection to using different prior distributions. BMC Proceedings 4, S5.

Weigel, K.A., de los Campos, G., González-Recio, O., Naya, H., Wu, X.L., Long, N., Rosa, G.J.M., Gianola, D., 2009. Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. Journal of Dairy Science 92, 5248-5257.

Weigel. K. A., de los Compos. G., Vazquez. A. I., Van Tassell. C. P., Rosa. G. J. M, O'Connell. J. R., VanRaden. P. M., Wiggans. G. R., 2010. Genomic Selection and its effect on Dairy Cattle Breeding Programs. 10th World Congress of Genetics Applied in Livestock Production, Leipzig, Germany.