Status of Genomic Evaluation in the Brown Swiss Populations

Hossein Jorjani, Jette Jakobsen, Eva Hjerpe, Valentina Palucci and João Dürr
Interbull Centre, Department of Animal Breeding and Genetics,
Swedish University of Agricultural Sciences, Box 7023, S-75007, Uppsala Sweden
e-mail: Hossein.Jorjani@slu.se

Abstract

Genotype data on 7670 bulls from six populations (CHE, DEU-AUT, FRA, ITA, SVN, and the USA) were used to form a pooled reference population for international genomic evaluation of 34 traits (150 population-trait combinations). Genetic correlations between MACE EBV/PA, and the GEBV values were on average around 0.9 and 0.7 for the old and the young bulls, respectively. Validation of the international genomic evaluation model for 16 traits (73 population-trait combinations) was performed. All populations passed the GEBV-test for FAT, MIL, PRO, and FTL. The GEBV-test for ANG, CC2, DLO, FAN, FTP, INT, MSP, RLS, RUH, and STA indicated passing for half of the population-traits, while no population-trait passed the GEBV-test for RAN and SCS. With these results, the first phase of building an efficient international genomic evaluation system is complete, and one can move to extending the system to alternative genomic evaluation models, if one wishes so.

Key words: common genomic reference population, genomic evaluation

Introduction

It has been three years since the Interbull Centre was asked to consider the feasibility of providing an international genomic evaluation service for Brown Swiss populations. From the beginning seven countries (six populations) were part of the project that gradually became known as the InterGenomics project. These populations were Austria and Germany (DEA), Switzerland (CHE), France (FRA), Italy (ITA), Slovenia (SVN), and the United States of America (USA). The genotype data started to be submitted to the Interbull Centre in December 2009, and by April 2010 all populations had submitted the genotypes of the bulls that were considered as the reference population bulls. Preliminary (and mostly descriptive) results were presented in the second Interbull Technical Workshop on Genomics in Paris, France (Jorjani *et al.*, 2010) and in the Interbull Annual Open Meeting in Riga, Latvia (Zumbach *et al.*, 2010). In July 2010, the first genomic evaluation results for protein yield (PRO) were distributed to the participating countries. The Interbull genomic validation test (GEBV-test; Mäntysaari *et al.*, 2010) was implemented in November 2010 for PRO. The number of population-trait combinations included in the Interbull Centre's genomic evaluations increased to 50 by March 2011 (Jorjani, *et al.*, 2011). International genomic evaluation of Brown Swiss populations entered the Interbull Centre's official test evaluation in September 2011 with 140 population-trait combinations. Since September 2011, the Interbull Centre's international genomic evaluation of Brown Swiss populations has gone through the official routine run in November/December 2011, the official test evaluation in January 2012, and the official routine evaluation in April 2012.

One of the guiding principles for the Interbull Centre's plans for the international genomic evaluation was to create a streamlined genomic evaluation system that can, with the

minimum human intervention, perform for as many breeds and traits as possible, and in as a short time as possible. Here, we report the status of Interbull Centre's genomic evaluation system and its application to the international genomic evaluation of the Brown Swiss populations as of April 2012.

Materials and Methods

Data - Genotypes: A total of 8233 genotypes were submitted from AUT, CHE, DEU, FRA, ITA, SVN and the USA (Table 1). All bulls had been genotyped with the different versions of the commercial Illumina chips (Illumina Inc, San Diego, USA). The chips contained different number of SNPs: 54001 (50KV1), 54609 (50KV2), 6909 (LD), and 777962 (HD). Some bulls had been genotyped in more than one country. As the results 7728 unique bull genotypes were available. After quality checks 7670 bull genotypes were used in the analyses (Table 1). Some of the bulls genotyped by the HD chip had been reported as 50KV2, mainly because no routines to submit the whole genotype had yet been developed at the national level.

Table 1. The number bull genotypes submitted from different populations.

	SNP panel				
	LD	50KV1	50KV2	HD	Sum
CHE	-	1799	948	-	2733
DEA	-	1821	639	-	2445
FRA	-	91	135	-	226
ITA	-	1547	185	-	1644
SVN	-	191	-	-	191
USA	4	938	44	15	994
Used					7670

Genotype imputation – In the April 2012 InterGenomics routine evaluation the bulls genotyped with the LD chip were not used, and the genotypes from the HD chip were reduced to the 50KV2 chip. The chromosome maps from 50KV1 and 50KV2 were merged with each other and a new map with 55172 SNPs, excluding SNPs with uncertain position, was created. The imputation method proposed by VanRaden *et al.* (2010) and its associated program (FindHap; VanRaden, 2010, http://aipl.arsusda.gov/software/findhap/) was used for imputation from 54001 or 54609 SNPs to the common map of 55172 SNPs.

Data - SNPs: The available 55172 SNPs were controlled for data quality, which included calculation of minor allele and genotype frequencies, Chi-square test of departure from expected Hardy-Weinberg equilibrium frequencies, and an extensive across genome control to see if any two loci are in near complete (i.e. 99.5%) phase similarity. After all the edits 45543 SNPs were used in the genomic analysis.

Data - Traits: All 34 traits included in the conventional MACE evaluations were used in the InterGenomics evaluations (Table 2).

Data - Phenotypes: De-regressed international estimated breeding values (MACE EBV) from the April routine MACE evaluations using sire and dam pedigree relationship were used as the phenotype. The international (MACE) EBVs of all animals (even females, ancestors, and genetic groups) were used to calculate Parent averages (PA).

Some very old and very young bulls did not have MACE EBVs. For the very old bulls the year averages for MACE EBV and PA in year 1970 were used. For the very young bulls, MACE EBVs and PAs for the latest 10 years with more than 10 observations were used to calculate a regression line. The regression was then extrapolated to the latest birth years and these values were used as MACE EBV and PA for the very young bulls. Number of bulls with own MACE EBV, constituting the reference population, varied from 2162 to 5320 bulls.

Table 2. The traits included in the genomic evaluations, number of populations per trait, and the size of the reference population.

Trait	Number of Populations	Bulls
ANG	3	2162
BDE	4	4660
CC1	4	3882
CC2	4	3997
CRC	4	4009
CWI	5	4439
DCE	3	4460
DLO	5	4964
FAN	5	4868
FAT	6	5320
FTL	5	4871
FTP	4	3723
FUA	5	3864
HCO	3	2697
HDE	4	4597
INT	2	2782
MAS	5	5067
MCE	3	3554
MIL	6	5320
MSP	3	4653
OCS	5	3584
OFL	4	4517
OUS	4	4579
PRO	6	5320
RAN	5	4872
RLS	5	4872
RTP	4	3560
RUH	5	4868
RUW	5	4868
RWI	4	3959
SCS	5	5067
STA	5	4872
UDE	5	4872
USU	5	4872
Total	150	

Genomic evaluation model: The genomic evaluation model was the GBLUP method used in the USDA and proposed by VanRaden (2008). The GBLUP software was also provided by Paul VanRaden (personal communication, 2009).

An iterative, nonlinear model with heavy-tailed prior for marker effects analogous to Bayes A is used (a curve parameter of 1.05 was used). Base population allele frequencies are subtracted from genotypes, and a polygenic effect (poly) with 10% of additive variance is fit in the model:

DPGM = mean + ∑genotypes*effects + poly + error.

where DPGM stands for de-regressed predicted genetic merit (Because MACE results can be expressed both as EBV and PTA, we use Predicted Genetic Merit (PGM) to include both of these).

The model is basically a single trait model (e.g. for protein yield). However, it is repeated for each population that has MACE evaluation for that trait. In this way all genotyped bulls, irrespective of their country of origin/genotyping, are used for all traits and populations (either included in the reference population or predictee population).

The evaluation model includes international (MACE) evaluation of ancestors (e.g. for calculation of parent averages for bulls born before 1981), and consequently the foreign information for non-genotyped ancestors is included. Genomic evaluation can be propagated to the non-genotyped animals. However, we at the Interbull Centre have not done this propagation, because it is InterGenomics countries' responsibility to do it.

Reliability of direct genomic value is computed from the traditional daughter equivalents plus the genomic daughter equivalents, which differ for each animal depending on its average genomic relationship to the reference population. Final reliability is

computed by selection index using reliabilities of DGV, traditional PGM, and subset PGM. These reliability values differ from "system reliability" that can be calculated from validation R^2. No correction for potential inflation of the reliabilities has been applied to the estimates, but members of the InterGenomics technical committee are aware of the issue and will apply discounts when estimating the GEBVs at the national level.

Calculation of DGV includes foreign information from a subset of the current Interbull evaluation for the genotyped animals.

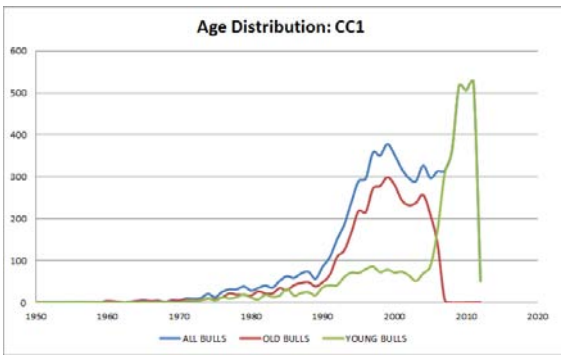
Validation: The last time the Interbull genomic validation test (GEBV-test; Mäntysaari *et al.*, 2010) was implemented was in connection with the January 2012 test run in which 16 traits were included in the validation run. Considering the presence of multiple population scales, the total number of population-trait combinations was 73 (Table 3).

Table 3. The traits included in the genomic validation, number of populations per trait, and the number of bulls used as the reference population.

Trait	Number of Populations	Bulls
ANG	3	1394
CC2	3	2383
DLO	5	3756
FAN	5	3125
FAT	6	3505
FTL	5	3128
FTP	4	2288
INT	2	1637
MIL	6	3505
MSP	3	2980
PRO	6	3505
RAN	5	3129
RLS	5	3129
RUH	5	3125
SCS	5	3319
STA	5	3129

Results & Discussion

Age distribution: The genotyped bulls were born from 1950 to 2012, with few bulls born before 1975 (see the example figures for CC1 and PRO in Figure 1). An increasing number of bulls born after 1981 are being genotyped. However, even for a centrally important trait such as PRO, there are some bulls that have not been genotyped. Compared to our previous reports (e.g. Jorjani *et al.*, 2011) more young bull genotypes have been submitted to the Interbull Centre, which we take as a sign of more open attitude towards genotype sharing, and more importantly as a sign of more trust in the Interbull Centre's genomic evaluations.



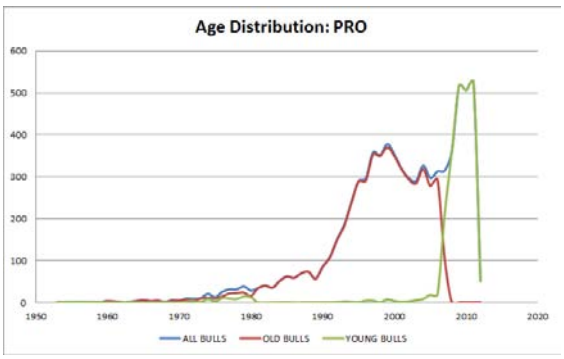


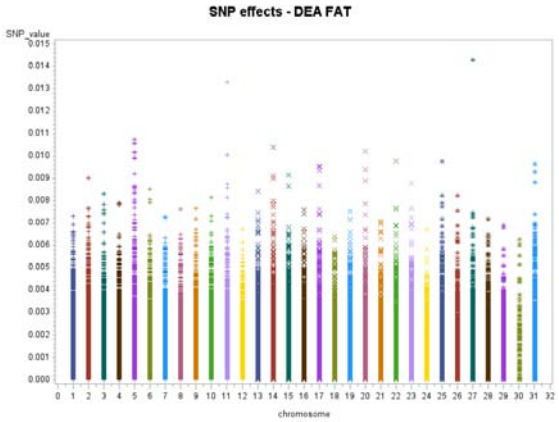
Figure 1. Age distribution and availability of data for the female fertility trait CC1 and PRO.

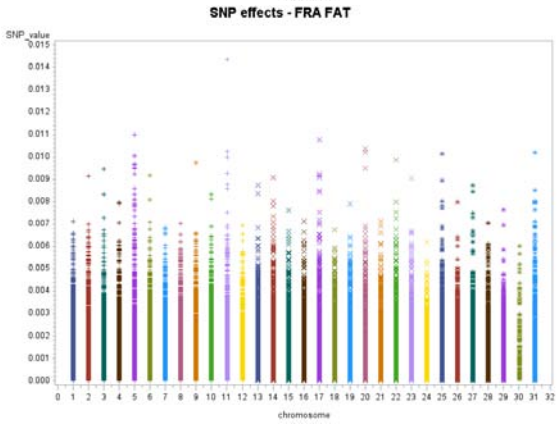
Phenotype availability: National genetic evaluations for all the traits were not available for all the countries. Consequently, the availability of MACE EBV for the pool of the

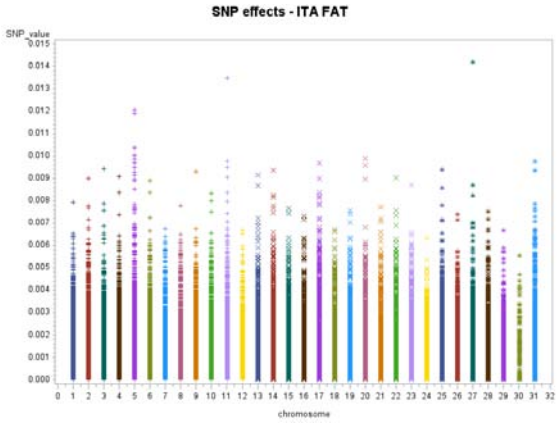
genotyped bulls showed a large variation across trait. For example, while 69% of all genotyped bulls had MACE EBV for production traits (FAT, MIL, and PRO), for ANG the availability of MACE EBV was as low as 28%. Percentage of genotyped bulls without own MACE EBV or MACE PA was only 1% for production traits (FAT, MIL, and PRO), and about 18% for the fertility trait INT.

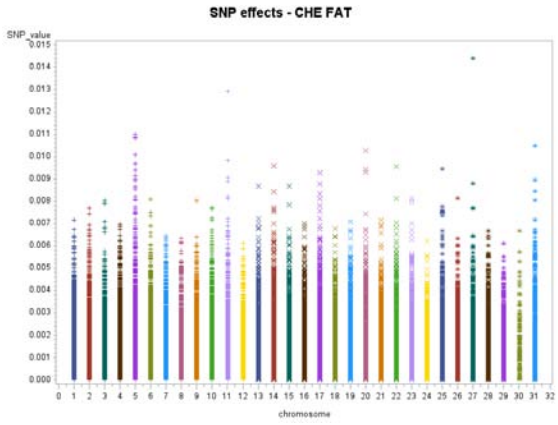
Depending on the number of countries participating in the routine MACE evaluations, there were between 3-6 country-scales available for genomic evaluations. The total number of genomic evaluations amounted to 150 population-trait (scale) combinations (Table 2).

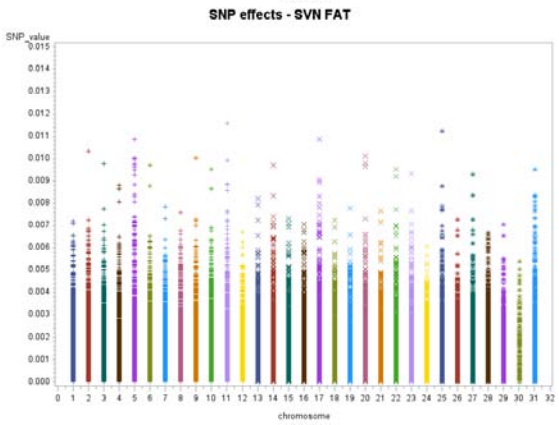
SNP effects: According to the nature of genomic evaluation model used in this study (VanRaden, 2008) SNP effects were estimated for all 45543 SNPs. Manhattan plots of SNP effects for the trait fat yield (FAT) are shown in Figure 2 for the 6 populations included in the analysis of FAT.











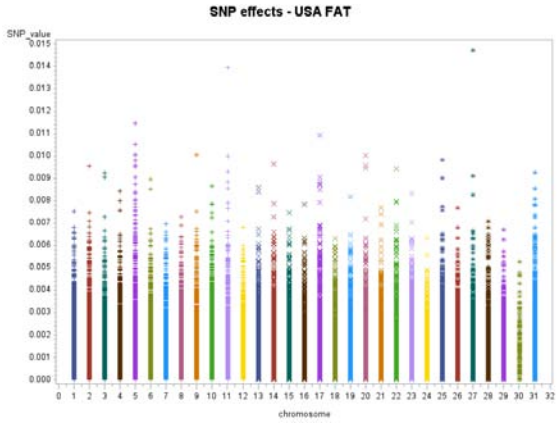


Figure 2. Plot of SNP effects for FAT on the 6 population scales.

Almost all countries with genomic evaluation use the MACE EBVs to convert the EBV of foreign bulls to the local scale. Therefore, the use of MACE EBV in the InterGenomics is by no means unique. However, InterGenomics might be considered as an extreme case of using MACE EBVs. One concern about MACE EBV is that it may smear out the national evaluations and lead to estimation of very similar across country SNP effects. As can be seen in Figure 2 different patterns of SNP effects can be observed in the 6 populations. Admittedly, the patterns of SNP effects for FAT are very similar to each other and less distinct than in many other traits. In fact, the largest SNP effects, say the largest 10 SNPs, have very little overlap with each other (results not shown).

Correlation between EBV/PA and GEBV: Correlations between the EBV and GEBV for old bulls (with own MACE EBV) were, obviously, very high. These bulls were, after all, the reference population and the SNP effects, DGVs, polygenic effects, and GEBVs were calculated from them. For the young bulls (with only a MACE PA) the correlations were reasonably high.

Table 4. Average correlations (R) and their standard deviations (SD) between MACE EBV and GEBV for all, old, and young bulls, for the six populations and the 34 traits of evaluation.

	All bulls		Old bulls		Young bulls	
	R	SD	R	SD	R	SD
CHE	0.88	0.03	0.93	0.03	0.73	0.08
DEA	0.88	0.04	0.93	0.04	0.73	0.09
FRA	0.87	0.04	0.92	0.04	0.73	0.08
ITA	0.89	0.03	0.93	0.03	0.76	0.09
SVN	0.93	0.01	0.95	0.00	0.80	0.02
USA	0.86	0.06	0.91	0.06	0.73	0.07
ANG	0.84	0.05	0.90	0.06	0.78	0.05
BDE	0.88	0.00	0.94	0.01	0.74	0.01
CC1	0.85	0.02	0.90	0.03	0.75	0.01
CC2	0.82	0.03	0.86	0.05	0.71	0.01
CRC	0.85	0.01	0.93	0.01	0.67	0.06
CWI	0.86	0.01	0.92	0.01	0.68	0.02
DCE	0.83	0.01	0.90	0.02	0.62	0.05
DLO	0.86	0.01	0.90	0.01	0.67	0.02
FAN	0.87	0.01	0.91	0.02	0.69	0.02
FAT	0.94	0.01	0.96	0.01	0.82	0.01
FTL	0.90	0.00	0.97	0.00	0.66	0.01
FTP	0.87	0.00	0.95	0.01	0.75	0.02
FUA	0.82	0.08	0.83	0.10	0.79	0.06
HCO	0.79	0.02	0.88	0.02	0.69	0.02
HDE	0.86	0.00	0.91	0.01	0.71	0.03
INT	0.81	0.00	0.91	0.01	0.70	0.04
MAS	0.90	0.01	0.95	0.01	0.64	0.02
MCE	0.80	0.01	0.88	0.03	0.67	0.05
MIL	0.94	0.01	0.96	0.01	0.79	0.01
MSP	0.88	0.00	0.95	0.01	0.64	0.02
OCS	0.88	0.01	0.88	0.01	0.88	0.01
OFL	0.84	0.03	0.85	0.03	0.83	0.03
OUS	0.90	0.01	0.92	0.00	0.84	0.01
PRO	0.94	0.01	0.96	0.00	0.84	0.01
RAN	0.89	0.00	0.96	0.00	0.64	0.01
RLS	0.88	0.00	0.94	0.00	0.67	0.03
RTP	0.86	0.00	0.94	0.01	0.76	0.01
RUH	0.90	0.01	0.92	0.01	0.84	0.02
RUW	0.90	0.01	0.92	0.01	0.84	0.02
RWI	0.86	0.02	0.94	0.02	0.66	0.02
SCS	0.90	0.01	0.95	0.01	0.64	0.02
STA	0.92	0.00	0.96	0.00	0.83	0.00
UDE	0.89	0.00	0.96	0.01	0.66	0.02
USU	0.89	0.01	0.93	0.01	0.76	0.02

Compared to some national genomic evaluations the correlations between MACE EBV and GEBV may seem low. However, this is a consequence of the extensive use of MACE EBV, in which the majority of bulls for any population scale not having any daughters in that population. Therefore, MACE EBVs on average have lower reliability than the domestic bulls of any population.

Gain in reliability: Addition of genotype information to the conventional sources of information (pedigree/parent average, and daughter performance) led to the increase of reliability for both groups of bulls, reference and candidate bulls.

Table 5. Average gain in reliability (G) and its standard deviation (S) for three groups of bulls for the 6 populations and the 34 traits of evaluation.

	All bulls		Old bulls		Young bulls	
	G	S	G	S	G	S
CHE	22	2.9	12	4.7	36	2.9
DEA	22	2.8	12	4.2	36	3.1
FRA	23	3.3	15	5.8	36	4.4
ITA	21	2.8	12	4.0	38	1.9
SVN	21	0.1	12	0.2	41	0.0
USA	23	3.1	16	5.6	35	4.5
ANG	21	0.5	13	5.1	23	2.6
BDE	21	0.6	11	0.8	37	0.4
CC1	26	1.3	20	4.1	32	1.6
CC2	27	0.6	21	1.7	33	0.6
CRC	25	1.5	16	3.4	34	0.5
CWI	24	1.0	15	1.7	36	0.4
DCE	25	2.8	18	5.5	36	1.2
DLO	27	1.5	21	2.5	38	0.9
FAN	25	2.5	18	4.4	37	0.7
FAT	19	1.4	10	1.9	40	0.6
FTL	18	0.2	6	0.2	38	0.1
FTP	22	0.9	7	1.8	35	0.1
FUA	25	1.1	19	4.0	32	1.8
HCO	23	0.4	19	2.8	26	1.6
HDE	26	0.9	19	1.5	36	0.3
INT	23	0.4	14	1.7	29	0.4
MAS	20	1.5	10	2.2	39	0.3
MCE	26	0.6	21	3.5	30	1.9
MIL	18	1.5	9	1.9	40	0.6
MSP	20	0.5	10	0.7	37	0.2
OCS	24	0.6	13	1.6	33	0.3
OFL	26	1.9	19	4.0	36	1.3
OUS	21	0.6	10	1.0	38	0.3
PRO	19	1.4	10	1.8	40	0.5
RAN	19	0.5	8	0.7	38	0.2
RLS	22	0.8	13	1.2	38	0.3
RTP	22	0.5	10	0.8	33	0.1
RUH	23	1.3	14	2.0	38	0.4
RUW	22	2.0	14	3.3	38	0.5
RWI	24	1.8	13	4.1	35	0.6
SCS	19	1.5	10	2.2	39	0.3
STA	18	0.3	6	0.4	38	0.1
UDE	19	1.1	8	1.6	38	0.1
USU	23	1.7	14	2.8	38	0.4

Gain in reliability ranged from 5% to 20% for the old bulls, and from 20% to 40% for the young bulls (depending on the heritability and the population structure).The increase in reliability depended, among other things, on the number of bulls in the reference population and the heritability of trait, and showed large variation.

Validation result: The number of bulls qualifying as test bulls in the different country-scales depended on the calculated EDC values for each country. Because the majority of the bulls with MACE EBV in any country scale have daughters in only one country, the bulls' EDC values were back-calculated from the reliability of the MACE EBV, which depended on the specific country-trait heritability values. The average number of qualifying bulls was 846$\pm$255.

Results of the GEBV-test showed much variation depending on the population and trait. All populations passed the GEBV-test for FAT, MIL, PRO, and FTL. The GEBV-test for ANG, CC2, DLO, FAN, FTP, INT, MSP, RLS, RUH, and STA indicated passing for half

of the population-traits, while no population-trait passed the GEBV-test for RAN and SCS.

Compared to the previous GEBV-tests conducted the estimated regression coefficient (b_1) were higher. We attribute the increase of b_1 to the larger reference population. There were distinct differences between the regression parameters between the population-traits that passed the test and those that failed it (Table 6).

Table 6. Average regression coefficient (b_1), average standard error of b_1 (SE b_1), and gain in the R-Square (R^2) for the population-trait combinations that passed and failed the GEBV-test.

	Passed		Failed	
	Mean	SD	Mean	SD
b_1	0.94	0.07	0.75	0.10
SE b_1	0.10	0.11	0.04	0.01
R^2 Gain	17.21	7.93	12.98	5.87

The gain in the degree of determination (R^2 Gain) was, understandably smaller than the gain in reliability shown in Table 5.

Conclusions

It is true that MACE EBVs are commonly used in many national genomic evaluation to calculate a converted EBV for foreign bulls. However, there are a lot of differences between the use of MACE EBV and the implementation of the genomic evaluation at the national and international level. Most of the differences arise from the fact that national genetic evaluations contain and produce more results in form of more complete pedigree on the dam side and national EBV for all animals. In using only MACE EBV, we are confronted with a situation in which the majority of bulls have daughters in only one country, and there are a number of international editing rules (such as consideration of birth year cut-off limits), that excludes a large number of bulls from having official MACE EBV. Another obstacle is the calculation of PA from MACE EBVs while the national EBVs from the dams are not included in the MACE analysis. Therefore, there were a number of "feasibility" issues that needed examination. Current results show that:

- Pooling genotypes from countries / populations to form a common reference population is possible and helpful;

- MACE EBVs , despite their limitations, can be successfully used for international genomic evaluations;

- The machinery for the operation of a large scale genomic evaluation operation, with data from several sources, can be created at a central location;

- Results of such genomic evaluations, including the validation results, are comparable with national genomic evaluations of similar size.

Acknowledgements

We would to acknowledge the InterGenomics partners (Arbeitsgemeinschaft der Österreichischen BraunviehZüchter, Austria; Arbeitsgemeinschaft Deutsches Braunvieh, Germany; Associazione nazionale allevatori bovini della razza Bruna, Italy; Brown Swiss Association of the US, USA; Brune Genetique Services, France; Braunvieh Schweiz, Switzerland; and Zveza rejcev govedi rjave pasme Slovenije, Slovenia) to give us permission to use their data in this study. We are also grateful for all the discussions and advices from members of the InterGenomics Technical Committee, especially Beat Bapst, Attilio Rossoni, Paul VanRaden and George Wiggans.

References

Jorjani, H., Zumbach, B., Dürr, J. & Santus, E. 2010. Joint genomic evaluation of BSW populations. *Interbull Bulletin 41*, 8-13. http://www-interbull.slu.se/ojs/index.php/ib/article/view/1133/1124

Jorjani, H., Jakobsen, J., Nilforooshan, M.A., Hjerpe, E., Zumbach, B., Palucci, V. & Dürr, J. 2012. InterGenomics project: Genomic evaluation of BSW populations. Interbull Bulletin 43: 5-8. http://www-interbull.slu.se/ojs/index.php/ib/article/view/1250/1242

Mäntysaari, E., Liu, Z. & VanRaden, P. 2010. Interbull validation test for genomic evaluations. *Interbull Bulletin 41*, 17-21. http://www-interbull.slu.se/ojs/index.php/ib/article/view/1134/1125

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

VanRaden, P.M., O'Connell, J.R., Wiggans, G.R. & Weigel, K.A. 2010. Combining different marker densities in genomic evaluation. *Interbull Bulletin 42*, 113-117. http://www-interbull.slu.se/ojs/index.php/ib/article/view/1168/1159

Zumbach, B., Jorjani,H. & Dürr, J. 2010. Brown Swiss genomic evaluation. *Interbull Bulletin 42*, 44-51. http://www-interbull.slu.se/ojs/index.php/ib/article/view/1156/1147