Joint Genomic Evaluation of BSW Populations

Hossein Jorjani[1], Birgit Zumbach[1], João Dürr[1] and Enrico Santus[2]
[1]Interbull Centre, Department of Animal Breeding and Genetics,
Swedish University of Agricultural Sciences, Box 7023, S-75007, Uppsala Sweden
[2]Associazone Nazionale Allevatori Bovini Razza Bruna
Loc. Ferlina, 204 , 37012 Bussolengo (Verona) , Italy
e-mail: Hossein.Jorjani@hgen.slu.se

Abstract

Brown Swiss dairy cattle populations from seven countries have joined forces to start a pilot project (Intergenomics) for joint genomic evaluation of these populations. The entire process of genomic evaluation, from data acquisition to the distribution of results, has been catalogued and Intergenomics' Technical Committee has set priorities for the use of sound scientific methods for each step of the data analysis. Descriptive statistics of the bull genotype data are also presented here. Full results are expected to be ready by the end of May 2010.

Key words: genomic evaluation, SNP selection, joint evaluation

Introduction

Interbull Genomic Task force in its report to the Interbull International Workshop on "Genomic Information in Genetic Evaluation" held in Uppsala, Sweden, January 26-29 2009, envisaged five different scenarios for the use of genomic information in international genomic evaluation (Banos *et al.*, 2009). The first of these scenarios (Scenario I) included Interbull Centre's full access to bull genotypes (as well as national EBV's for bulls). Scenario I was judged to be particularly attractive for *small populations*.

Interbull Genomic Task Force did not specify any definition for small population, however it was generally understood that the reference population size of 2500 bulls mentioned by Schaeffer (2006) was probably too small. Brown Swiss cattle populations from nine evaluation centers (including AUT-DEU joint evaluation, summing up to 10 countries) participate in the international genetic evaluations conducted at the Interbull Centre. However, probably none of these populations is large enough to form a competitive reference population.

The European Brown Swiss Cattle Breeders Federation evaluated different options and judged that Scenario I outlined by Interbull Genomic Task Force might be the best option for the Brown Swiss breed. Consequently, they asked Interbull Centre to prepare a proposal for a limited pilot project to study feasibility of implementing scenario I. Later, the European Brown Swiss Cattle Breeders Federation discussed their intentions with the World Federation of Brown Swiss Cattle breeders and invited all Brown Swiss Cattle populations to join the pilot project.

The interested parties, from 7 different countries (AUT, CHE, DEU, FRA, ITA, SVN and the USA) and the Interbull Centre, formed a Management Committee with the participation of one representative from each party. They gradually adopted the name Intergenomics for the project and in their first meeting in fall 2009 appointed a technical committee to make decisions on technical matters. At least one scientist from each of the interested parties was nominated for the Technical Committee.

The aim of this paper is to give a short summary of decisions by the Intergenomics Technical Committee and some preliminary statistics on the shared data.

Material & Method

Intergenomics Technical Committee discussed the proposal from the Interbull Centre project team, in which the workflow was divided into three main areas, as follows:

Genotype data acquisition

- Analysis of BSW pedigree;
- Nomination process;
- Exchange mechanisms;
- Sample and genotype availability;
- SNP-chip usage guidelines;
- SNP-Database;
- Data request;
- Data reception routines;
- File formats.

Data editing

- Devising quality control measures;
- Standard and breed specific criteria;
- SNP-data verification;
- Parentage checks;
- Selection of informative SNPs;
- SNP imputation;
- Search for causative mutations.

Data analysis

- Preparing dependent variables;
- Estimating SNP effects;
- Estimating direct genomic value (DGV);
- SNP and DGV reliability;
- Combining DGV and EBV;
- Validation of national and international genetic evaluation systems;
- Sensitivity analysis.

For each of the steps mentioned above members of the Intergenomics Technical committee a) presented a country report; b) discussed the list of activities; priority and impact; available methods; preferred method; c) appointed a working group and d) decided about the time table.

Data

It was decided that the national EBV reported to the Interbull Centre should be used to derive the dependent variable, which would be the international breeding value of bulls.

For genotype exchange two simple file formats 704 and 705 were used (Appendix I and II, respectively). SNP genotype data were reported from CHE, FRA, ITA, SVN and the USA.

Results and Discussion

SNP usage guidelines and file formats

The working group for preparation of SNP usage data guidelines worked on a general guidelines and the possibility of creating 5 different file formats was discussed.

File formats 701 to 703 were reserved for file formats providing detailed information on assessment of the “call quality from each sample”, “clustering quality” and “sample * SNP calls”, respectively. These file formats are meant to be used in the long term and can also be used for setting standards for genotype reporting, even outside of the Intergenomics project. Without these file formats there is little opportunity to have a quality control over reported genotypes.

File formats 704 and 705 provide summary information on AB-Genotype coding system devised by the chip manufacturing company Illumina. The only quality control measure that can be extracted from this file format is the percentage of non-called SNPs per bull.

Bull genotypes

Data in file formats 704 were received from CHE, FRA, ITA and SVN and in file format 705 from ITA, SVN and USA. Table 1 shows the number of bull genotypes submitted from each country. Bull genotypes from AUT and DEU are due later this year.

Table 1. Number of submitted bull genotypes.

	Number of bulls	
Country	Total	Foreign
CHE	1202	274
FRA	93	
ITA	762	179
SVN	191	14
USA	816	31

Some bulls had been genotyped more than once (Table 2). There were also some bulls genotyped more than once within a country (results not shown).

Table 2. Number of multiple genotyped bulls.

	FRA	ITA	SVN	USA
CHE	4	93	-	54
FRA		4	-	-
ITA			1	70
SVN				-

There were some differences between the multiple genotypes reported for these bulls (Table 3). A variety of reasons can cause the discrepancies in reported genotypes. However, the information content of file format 704 or 705 are not enough for identifying the reasons. For these bulls the genotype with least number of non-called SNPs were used, which is a sub-optimal solution. The reason is that strong quality control in one country may assign many SNPs to non-called, while in another country low quality SNPs may be included in the files. Average numbers of non-called SNPs per bull were between 114 and 3460 SNPs for different countries. An alternative would be to check for sire-son genotype consistency.

Table 3. Percentage of genotype differences for called SNPs among multiple genotyped bulls.

	FRA	ITA	SVN	USA
CHE	1%	1%		6%
FRA		0%		
ITA			3%	7%
SVN				

Intergenomics Technical Committee found it more appropriate to perform genotype imputation for the low quality SNP at the international level and file formats 704 and 705 provided a coding system that could be used for this purpose. However, it was also decided that genotype imputation has low priority at this stage.

There was a peculiar deviance of allele frequencies from 0.5, i.e. the overall average frequency of allele A for all of 54001 SNPs was about 0.46 in all countries (Table 4). Whether the SNPs on the commercial chips are neutral is not clear. However, there is no reason for the A allele be less frequent that the B allele. Because this observation has been valid for HOL and JER, VanRaden (personal communication) speculated that this might have to do with how the AB-genotype coding system is constructed.

Table 4. Frequency of the A allele within the AB-genotype coding system.

	CHE	FRA	ITA	SVN	USA	ALL
MIN	0.00	0.00	0.00	0.00	0.00	0.00
AVE	0.45	0.45	0.46	0.45	0.46	0.46
MAX	1.00	1.00	1.00	1.00	1.00	1.00
SD	0.34	0.35	0.34	0.33	0.35	0.33

Editing the SNP list

To ascertain the total number of usable SNP several checks were performed. First, if the percentage of non-called SNPs were larger than 25% in all bulls, the SNP was discarded. Second, SNPs with minor genotype frequency < 0.0004 (a criterion more restrict than MAF < 0.02) were discarded. Third, the genotypes at each SNP were compared to all other SNP genotypes in the population and if the genotypes were equal (or opposite) in more than 0.995 of the bulls, the SNP was discarded. After the editing, 42437 usable SNPs were remained for analysis.

Even though the overall average allele frequencies were the same across countries, there were many SNPs which showed clear allele frequency differences in different countries. Again, it is not clear if the SNP panel used contains only neutral SNPs, however, differences in allele frequencies for different SNPs, if not caused by pure random drift, may indicate selection for country specific breeding goals.

Hardy-Weinberg test

Departures from Hardy-Weinberg ratios were also checked. There were between 1909 and 3946 SNPs per country that showed highly significant departure from the expected ratios. It is interesting to note that different SNPs were in violation of equilibrium in different countries.

Based on the decision made by the Intergenomics Technical Committee the default method of analysis of data, and the base for comparison with other methods is the methods proposed by VanRaden (2008). The analysis of genomic data has started and fine-tuning of different computer programs is underway. Full results of analyzing the data by the default method are expected to be ready by the end of May 2010.

Conclusions

Genomic evaluation of dairy cattle population is still in its infancy and it is Intergenomics' hope to contribute significantly to its development. The results will be foremost useful Brown Swiss farmers around the world. However, building up a new evaluation system from scratch at the Interbull Centre will add a very important part to the Interbull Centre's of repertoire of analysis tools that can be helpful for farmers of all breeds.

Acknowledgments

Authors wish to express their appreciation for the initiative taken by the European Federation of Brown Swiss Cattle Breeders, Intergenomics Management Committee and Intergenomics Technical Committee.

References

Banos, G., Calus, M., Ducrocq, V., Dürr, J., Jorjani, H., Liu, Z., Mäntisaari, E., Sullivan, P. & VanRaden, P. 2009. Preliminary report from Interbull Task Force on the role of genoimc information in genetic evaluations. Proceedings of the Interbull international workshop on "Genomic information in genetic evaluations" *Interbull Bulletin 39,* 66-71.

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

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

Appendix I

File Format for SNP Data Delivery to the Interbull Centre for the International Genomic Evaluation

National SNP Data (Short version) **(record length=42)**

Starting Position	Field Description	Format	Example
1	Record type [1]	integer 3	704
4	Country sending this information [2]	character 3	AUS
7	Breed of evaluation [3]	character 3	BSW
	Animal international identification [4]		
11	Breed of animal [5]	character 3	BSW
14	Country of first registration of animal [2]	character 3	AUS
17	Sex	character 1	M
18	ID number of animal [6]	character 12	000000A12345
	National SNP data [7]		
31	SNP Index [8]	integer 8	54001
40	Allele 1 – AB [9]	Character 1	A
42	Allele 2 – AB [9]	Character 1	B

[1] Record type is always 704 for this File Format
[2] ISO 3166-1 alpha-3 codes (3 characters, capital letters)
[3] Breed of evaluation (3 characters, capital letters, BSW, GUE, HOL, JER, RDC, SIM)
[4] Please see Interbull Bulletin 28.
[5] Breed of animal (3 characters, capital letters)
[6] Alpha-numerical, always 12 characters long
[7] File Format 704 is the short version of File Format 703 (to be distributed in 2010)
[8] SNP Index from 1 to 54001 based on the standard order of Illumina Bovine SNP50 BeadChip
[9] Acceptable values are A/B (called genotypes), M/N (imputed genotypes), and Z (not called genotypes)

Appendix II

File Format for SNP Data Delivery to the Interbull Centre for the International Genomic Evaluation

National SNP Data (Short version) (record length=54031)

Starting Position	Field Description	Format	Example
1	Record type [1]	integer 3	705
4	Country sending this information [2]	character 3	AUS
7	Breed of evaluation [3]	character 3	BSW
	Animal international identification [4]		
11	Breed of animal [5]	character 3	BSW
14	Country of first registration of animal [2]	character 3	AUS
17	Sex	character 1	M
18	ID number of animal [6]	character 12	000000A12345
	National SNP data [7]		
31	AB – Genotype for SNP Index 1 [8,9]	integer 1	0
32	AB – Genotype for SNP Index 2 [8,9]	Integer 1	1
…	AB – Genotype for SNP Index … [8,9]	Integer 1	2
54031	AB – Genotype for SNP Index 54001 [8,9]	Integer 1	5

[1] Record type is always 704 for this File Format
[2] ISO 3166-1 alpha-3 codes (3 characters, capital letters)
[3] Breed of evaluation (3 characters, capital letters, BSW, GUE, HOL, JER, RDC, SIM)
[4] Please see Interbull Bulletin 28.
[5] Breed of animal (3 characters, capital letters)
[6] Alpha-numerical, always 12 characters long
[7] File Format 705 is an alternative for File Format 704
[8] SNP Index from 1 to 54001 based on the standard order of Illumina Bovine SNP50 BeadChip
[9] Acceptable values are according to the following table:

AB – Genotype	Allele 1 – AB	Allele 1 – AB	Genotype call
0	B	B	Assigned
1	A	B	Assigned
2	A	A	Assigned
5	Not assigned	Not assigned	
7	B	B	Imputed
8	A	B	Imputed
9	A	A	Imputed