INTERBULL BULLETIN NO. 56 Leeuwarden, The Netherlands, April 26-30 2021

Interbull genomic evaluation of small Holstein populations: InterGenomics-Holstein (IG-HOL)

***S. Savoia*[1]**

[1] *Interbull Centre, SLU, Box 7023, S-75007 Uppsala, Sweden*

Abstract

In December 2020 Interbull Centre launched a new Service for the international genomic evaluation of small Holstein populations: “InterGenomics-Holstein” (IG-HOL). The IG-HOL Service’s principles are to: a) have two categories of participants: IG-HOL countries (small Holstein populations’ countries), and “Contributors” (larger Holstein populations’ countries); b) pool genotypes provided by both participation categories creating a unique reference population (without sharing any genotypes among participants); c) predict gEBVs on each specific IG-HOL country-trait combination scale (for the traits evaluated in the Interbull MACE Service), by adapting to Holstein specificities the methodology used in the Interbull InterGenomics Brown Swiss Service; d) apply the Interbull method for approximating genomic reliabilities.

In the IG-HOL December 2020 routine run 12 280 male genotypes submitted by 4 IG-HOL countries (IRL, ISR, KOR, SVN) and 4 Contributors (DEU, DFS, FRA, PRT) were used in the genomic evaluation of 30 traits predicting gEBVs for 89 IG-HOL country-trait combinations. The size of the pooled reference population ranged from 4 645 bulls (body condition score - conformation trait) to 7 544 bulls (milk, fat and protein - production traits). As expected, for reference bulls the predicted IG-HOL gEBVs showed high correlations with MACE EBVs, while lower correlations were observed for young bulls comparing their IG-HOL gEBVs to MACE PAs. The reliability gain was satisfactory, with an average increase across all 30 traits of +0.22 for young bulls’ genomic reliabilities vs the MACE PAs’ reliabilities. These results show that IG-HOL is a valuable Service, especially for countries with small Holstein populations which cannot benefit from being part of large international genotype-exchange consortia.

Key words: international evaluation, genomics, reliability gain, InterGenomics, IG-HOL.

Introduction

InterGenomics-Holstein (IG-HOL) is a Service provided by Interbull Centre aiming at providing genomic selection tools to breeding organisations of small Holstein populations which are not part of large international genotype-exchange consortia.

From a technical perspective, the IG-HOL Service has been developed adapting to the Holstein breed the methodology already in place in the “InterGenomics Brown Swiss” Service (IG-BSW), i.e. the International genomic evaluation of Brown Swiss populations, successfully launched by Interbull Centre in 2011 (Jorjani et al., 2011).

However, unlike in the IG-BSW Service where Brown Swiss animals’ genotypes delivered to Interbull Centre are shared among all participating Organisations, in the IG-HOL Service no Holstein genotypes are shared among participants.

The organisations of small Holstein populations participating in the IG-HOL Service, identified as “IG-HOL countries”, submit their genotypes (of A.I. sires, young bulls, cows and heifers) to Interbull Centre which gathers them in the IG-HOL genotypes pool used for the analyses.

During the development of the Service, despite gathering all genotypes provided by the IG-HOL countries, the number of genotyped sires available for the constitution of the reference population turned out not to be optimal.

In order to improve the quality of the evaluation's results, Interbull Centre therefore gave the possibility to larger Holstein populations' organisations to contribute with genotypes to the IG-HOL Service. These organisations, identified as "Contributors", provide to Interbull Centre the genotypes for both nationally proven sires and young bulls in A.I.

As a consequence, IG-HOL countries could benefit from genomic evaluation's results based on a more informative reference population, while Contributors have marketing opportunities for their genotyped young bulls in the IG-HOL countries' national A.I. markets.

Thanks to this collaboration, the first Service test run was performed in September 2020, followed by the first official IG-HOL routine run in December 2020.

The present manuscript describes the methodology applied in the IG-HOL Service as well as the results from the December 2020 routine evaluation.

Materials and Methods

Participants

The IG-HOL countries and Contributors joining the December 2020 routine run are reported in Table 1.

Table 1. – Organisations and respective countries involved in the IG-HOL Service.

Organisation[1]	Country	Role
ICBF	IRL	IG-HOL country
ICBA	ISR	IG-HOL country
NIAS	KOR	IG-HOL country
HOL-SLO	SVN	IG-HOL country
VIT	DEU	Contributor
NAV	DFS[2]	Contributor
GENEVAL	FRA	Contributor
ANABLE[3]	PRT	Contributor[3]

[1] ICBF: Irish Cattle Breeding Federation, Ireland; ICBA: Israeli Cattle Breeders Association, Israel; NIAS: National Institute of Animal Science, South Korea; HOL-SLO: Slovenian Holstein Association, Slovenia; VIT: Vereinigte Informationssysteme Tierhaltung, Germany; NAV: Nordic Cattle Genetic Evaluations, Denmark, Finland and Sweden; GENEVAL: GenEval, France; ANABLE: Associação Nacional para o Melhoramento dos Bovinos Leiteiros, Portugal.

[2] Denmark-Finland-Sweden.

[3] joined only during the development of the Service. Its reference bulls' genotypes were used in the evaluations.

Genotypes

A total of 35 431 genotypes were submitted by IG-HOL countries (IRL, ISR, SLO, KOR), Contributors (DEU, DFS, FRA) and PRT (Table 2).

Animals had been genotyped using the following versions of the commercial chip arrays with low (LD), medium (MD) and high (HD) SNP density:

a) LD chips:

- Illumina Bovine3k BeadChip (2 900 SNPs);
- Illumina BovineLD BeadChip (6 900 SNPs);
- ICBF International Dairy and Beef v2 (16 327 SNPs);
- ICBF International Dairy and Beef v1 (16 499 SNPs);

b) MD chips:

- GeneSeek Genomic Profiler v4 (30 105 SNPs);
- EuroGenomics MD v1 (41 913 SNPs);
- GeneSeek Genomic Profiler 50K (47 850 SNPs);
- Weatherbys Scientific Bovine VersaSNP 50K (49 745 SNPs);

- ICBF International Dairy and Beef v3 (52 690 SNPs);
- Illumina BovineSNP50v3 BeadChip (53 218 SNPs);
- Illumina BovineSNP50v1 BeadChip (54 001 SNPs);
- Illumina BovineSNP50v2 BeadChip (54 609 SNPs);

c) HD chips:

- GeneSeek Genomic Profiler HD v1 (76 999 SNPs);
- GeneSeek Genomic Profiler HD v2 (13 480 SNPs);
- Illumina BovineHD BeadChip (777 962 SNPs).

Table 2. - Number of submitted genotypes by chip array SNP density and sex.

Chip array SNP density[1]	Sex		Total
	M	F	
LD	2 463	6 772	9 235
MD	10 055	11 824	21 879
HD	1 172	3 145	4 317
Total	13 690	21 741	35 431

[1] LD: Low SNP density; MD: Medium SNP density; HD: High SNP density.

After checking for consistency between each genotype density and the related chip array, all submitted genotypes were recoded using the IG reference chromosome map (obtained by merging llumina BovineSNP50v1 and Illumina BovineSNP50v2 chips maps, excluding SNPs with uncertain position) to a density of 55 172 SNPs. The markers from the original genotypes not included in the reference map were coded as missing.

All animals' IDs have been validated using the international cross-reference file, used in the Interbull MACE evaluation, in order to detect animals with more than one genotype. In case more than one genotype was available for an animal (either within or across countries), only genotype information with the highest call rate (considering the IG reference chromosome map) was retained. Genotypes belonging to animals without any pedigree information in the Interbull IDEA database (https://idea.interbull.org/) were discarded.

A total of 33 691 validated genotypes (21 411 females' and 12 280 males') were used for the subsequent analyses.

SNP quality control (before imputation)

Before imputation, all available 55 172 SNPs were checked for parentage conflicts. SNPs that showed parentage conflicts (sire/dam and offspring with opposite homozygous) in more than 1% of the IG-HOL genotypes pool were set to missing for all the animals.

Among the 28 643 parent-offspring cases, 1 570 showed more than 2% of SNPs parentage conflicts among homozygous loci on autosomal chromosomes. The pedigree links between these animals were severed.

Imputation

In order to reduce the number of SNPs missing from the IG reference chromosome map, the 33 691 available genotypes were imputed to the 55 172 SNP density following the method proposed by VanRaden et al. (2010) and its associated software (FindHap v. 3, VanRaden 2010).

The average imputation accuracy across all genotypes was 0.99, ranging from 0.96 for the Illumina Bovine3k BeadChip genotypes (2 900 SNPs; 42 028 SNPs imputed) to 0.99 for the Illumina BovineSNP50v2 BeadChip genotypes (54 609 SNPs, 2 117 SNPs imputed). 48 female ancestors' genotypes were imputed from progeny, with an average accuracy of 0.76.

Females' imputed genotypes were distributed to each submitting IG-HOL Organisation and only males' imputed genotypes were used in the IG-HOL genomic prediction.

SNP quality control (after imputation)

A further quality check was performed on the 55 172 imputed SNPs, which included calculation of minor allele and genotype frequencies. After editing, 45 895 SNPs were used in the genomic analysis.

Traits

30 traits and 89 IG-HOL country-trait combinations (among the ones evaluated in the Interbull MACE December 2020 routine run) were assessed in the IG-HOL December 2020 routine run (Table 3).

Table 3. – Number of IG-HOL countries participating in the MACE traits groups evaluations.

MACE Traits Group[1]	IG-HOL countries
Calving	2
Conformation	3
Fertility	2
Longevity	3
Production	4
Udder Health	4

[1] The list and the description of traits within each MACE traits group is available at https://interbull.org/ib/cop_chap6

Phenotypes

International MACE breeding values (MACE EBVs) from the Interbull MACE December 2020 routine run, using sire and dam pedigree relationship, were used as input data.

MACE EBVs from each specific IG-HOL country scale were standardised and then deregressed following the approach by VanRaden et al. (2009).

Only genotyped bulls receiving a MACE EBV directly from their own daughters' contribution were considered as reference bulls. The number of bulls with own MACE EBV, constituting the reference population, varied from 4 645 bulls (for body condition score - conformation trait) to 7 544 bulls (for milk, fat and protein - production traits).

The MACE EBVs of all available animals (females, ancestors, and genetic groups) were used to calculate MACE parent averages (MACE PAs). The genotyped bulls without own MACE EBV and younger than 7 years of age were identified as "young bulls".

Prediction of genomic breeding values (gEBVs)

The genomic evaluation model and the software used for the Interbull IG-BSW Service described in detail by Jorjani et al. (2012) were used in the IG-HOL Service.

Briefly, an iterative, nonlinear GBLUP model with heavy tailed prior for marker effects analogous to "Bayes A" was used setting a curve parameter of 1.05 (VanRaden, 2008). Base population allele frequencies were subtracted from genotypes, and a polygenic effect accounting for 20% of additive variance was fitted in the following single trait model:

$$\boldsymbol{y} = \boldsymbol{\mu} + \mathbf{Z}\boldsymbol{g} + \mathbf{W}\boldsymbol{u} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is $\mathbf{n} \times 1$ vector of de-regressed MACE EBVs, $\boldsymbol{\mu}$ is the general mean, $\mathbf{Z}$ is $\mathbf{m} \times 1$ design matrix containing regression coefficients on m SNPs, $\boldsymbol{g}$ is $\mathbf{m} \times 1$ vector of SNPs effects, $\mathbf{W}$ is $\mathbf{p} \times 1$ design matrix that relates observations to $\mathbf{p}$ number of animals in the pedigree, $\boldsymbol{u}$ is $\mathbf{p} \times 1$ vector of polygenic effect and $\boldsymbol{e}$ is $\mathbf{n} \times 1$ vector containing the adjusted reliability of the de-regressed MACE EBVs.

The model was repeated for each country-trait combination having a MACE evaluation for the evaluated trait, using all genotyped bulls, irrespective of their country of origin/genotyping (either included in the reference population or as young bulls).

Direct genomic values (DGV) were obtained as the sum of SNPs effects for both references and young bulls. Final gEBVs were obtained combining DGV, PA computed from the subset of genotyped ancestors using traditional relationships and de-regressed MACE EBV (or PA for young bulls) applying the selection index theory (VanRaden et al., 2009).

Approximation of genomic reliabilities applying the Interbull method

Genomic Reliabilities were obtained applying the Interbull method following the guidelines from Liu et al. (2017) and from Liu et al. (2019). The *snp_blup_rel* software (Ben Zaabza et al., 2020) was used in the process. Reference bulls' EDCs were computed in each IG-HOL country scale using bulls' daughter information from all countries participating in MACE evaluation, following the method of Liu (2011) and a dedicated software developed by VIT.

Validation of IG-HOL gEBVs

IG-HOL gEBVs were validated during the development of the IG-HOL Service, applying the Interbull genomic validation test (GEBV-test; Mäntysaari et al., 2010).

An *ad-hoc* IG-HOL evaluation truncated on the birth year of the reference bulls was performed to get the reduced data sets.

In order to increase the size of the validation population, foreign bulls were also considered among the test ones.

The IG-HOL gEBVs successfully passed the GEBV-test in 92% of the investigated country-trait combinations.

Validation of genomic reliabilities

Formulas based on GMACE input data were developed following the method described by Sullivan and Jakobsen (2014) to get predicted genomic reliabilities in each IG-HOL country-trait combination. The average predicted genomic reliability for each country-trait combination was subtracted from the average genomic reliability of candidate bulls approximated with the Interbull method.

The obtained differences were comparable with the ones observed in countries involved in the Interbull GMACE Service.

Results & Discussion

Age distribution

The distribution of reference and young bulls by year of birth and origin (IG-HOL Organisation or Contributors) for production traits is shown in Figure 1.

The 7 544 genotyped bulls with MACE EBV (for production traits) in MACE December 2020 routine run were born between 1986 and 2016. The bulls with MACE EBV submitted by Contributors represented 42% of the total reference population, most of them being born between 2009 and 2015.

The 2 315 genotyped young bulls without MACE proof (for production traits) were born between 2013 and 2020, and 37% of them was submitted by Contributors.

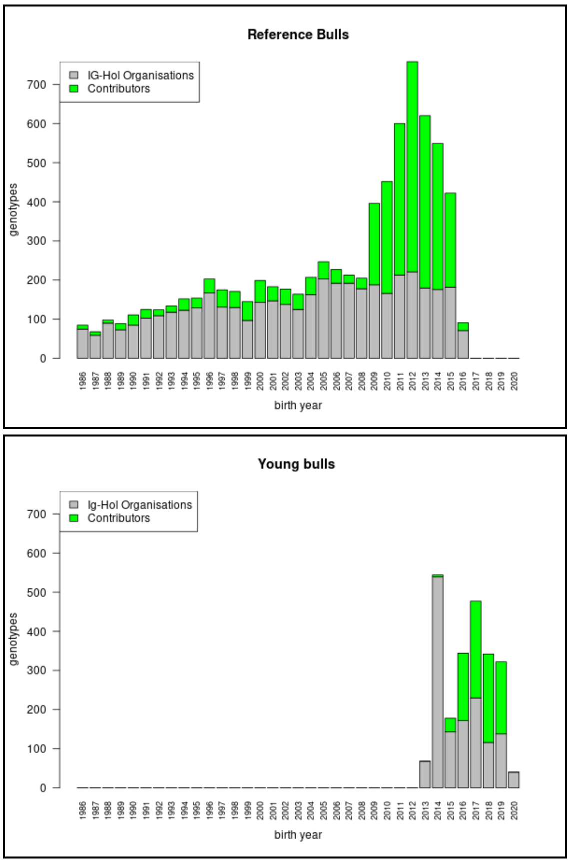

Figure 1. Distribution of reference and young bulls by year of birth and origin (IG-HOL Organisation or Contributors) for production traits.

Phenotypes availability

The IG-HOL countries did not provide national proofs in the December 2020 MACE run for all the evaluated traits. Consequently, the availability of MACE EBVs (phenotypes) on the IG-HOL country scales showed variation across traits, ranging from 1 to 4 country scales per trait.

MACE EBVs on a total of 89 IG-HOL country-trait combinations were used for single trait genomic evaluations.

SNP effects

According to the parameters used in the genomic model, SNP effects were estimated for all 45 895 SNPs retained after quality checks, and for each of the 89 country-trait combinations.

As previously reported by Jorjani et al. (2012), one of the main concerns about using MACE EBVs as input variable for genomic predictions is that it could lead to estimation of very similar SNP effects across countries.

To investigate this potential issue, descriptive statistics and correlations between SNPs effects estimated for the different countries within the same trait were investigated and the descriptive statistics summarised (by trait group) in Table 4.

Table 4. - Correlations between SNP effects estimated in different population scales (within trait), and differences with the across country genetic correlations used in the Interbull MACE Service.

MACE Traits Group	avg	min	Max	diff vs MACE
Calving	0.94	0.94	0.94	+0.05
Conformation	0.91	0.75	0.99	+0.16
Fertility	0.83	0.83	0.83	+0.22
Longevity	0.77	0.70	0.84	+0.19
Production	0.88	0.75	0.97	+0.11
Udder Health	0.92	0.85	0.98	+0.11

The average across country correlation within trait for SNP effects was 0.90, ranging from 0.70 (for direct longevity – longevity trait) to 0.99 (for teat length – conformation trait), showing that differences still exist in SNPs solutions for a trait in different populations scales.

However, comparing the SNP effects correlations to the across countries correlations used in the Interbull MACE Service evaluation for the same traits, an overall increase in the correlation intensity was observed in favour of the former. The relationship between across countries SNP effects correlations and correlations used in MACE evaluation is shown in Figure 2.

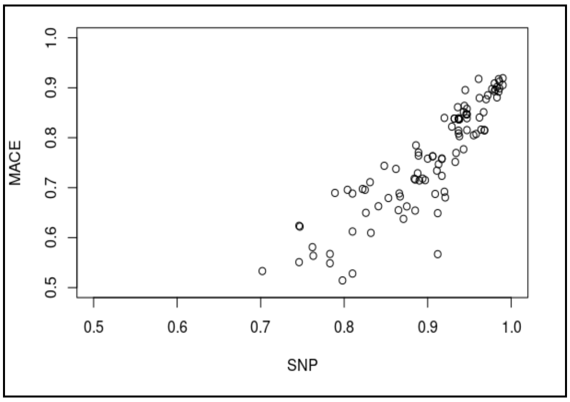

Figure 2. Relationship between SNP effects correlations and across country correlations used in the Interbull MACE evaluation.

Correlation between IG-HOL gEBVs and MACE EBVs/PAs

As defined in the IG-HOL agreement between IG-HOL Organisations and Contributors all animals with a genotype receive a gEBV on all IG-HOL country scales for all traits having a MACE evaluation for the specific IG-HOL country-trait combination. The animals do not receive gEBVs on the Contributors' country scales.

Considering both reference sires and young bulls in all IG-HOL country-trait combinations, a total of 776 346 gEBVs were predicted in the IG-HOL December 2020 routine run.

Correlations between IG-HOL gEBVs and MACE EBVs/PAs are shown in Table 5 and 6.

Table 5. - Average correlations between IG-HOL gEBVs and MACE EVBs for reference bulls by IG-HOL country and traits group.

MACE Traits Group	IG-HOL country A	B	C	D
Calving	0.95	0.96		
Conformation	0.96		0.94	0.94
Fertility	0.96	0.92		
Longevity	0.93	0.93		0.93
Production	0.93	0.95	0.95	0.95
Udder Health	0.94	0.95	0.95	0.94

Table 6. - Average correlations between IG-HOL gEBVs and MACE PAs for young bulls by IG-HOL country and traits group.

MACE Traits	IG-HOL country			
Group	A	B	C	D
Calving	0.62	0.68		
Conformation	0.67		0.68	0.68
Fertility	0.65	0.60		
Longevity	0.71	0.84		0.81
Production	0.81	0.80	0.80	0.74
Udder Health	0.70	0.69	0.62	0.66

As expected, for reference bulls the correlations between IG-HOL gEBVs and MACE EBVs were high in all country-trait combinations, ranging from 0.86 (for locomotion – conformation trait) to 0.98 (for stature – conformation trait).

For the young bulls, correlations between IG-HOL gEBVs and MACE PAs were lower compared to the reference bulls' ones, showing differences when moving from a pedigree based to a genomic based breeding value's prediction for animals without proven proofs. Correlations for young bulls were, on average, close to 0.70, ranging from 0.54 (for chest width – conformation trait) to 0.84 (for direct longevity – longevity trait). Young bulls' correlations between IG-HOL gEBVs and MACE PAs showed across countries consistency within traits groups.

Genomic reliabilities

The use of genomic information in addition to the conventional sources (pedigree and daughter performances) led to an increase of reliability for both reference sires and young bulls, in all IG-HOL country-trait combinations.

Moving from MACE EBVs to IG-HOL gEBVs, the gain in reliability for reference sires ranged from +0.031 (for stature – conformation trait) to +0.097 (rear-leg rear view – conformation trait), with an average increase across all country-trait combinations of +0.065. The trait group which benefited the most was direct longevity (conformation trait), with an average reliability gain of +0.090. Large variation was observed in the genomic reliabilities' values comparing the different traits, which was mainly related to the different sizes of the trait specific reference population and to the traits' heritability.

Regarding young bulls, reliability gain resulting from using IG-HOL gEBVs instead of MACE PAs was, on average across traits, +0.220 and ranged from +0.171 (for overall feet and legs score – conformation trait) to +0.253 (for fat – production trait). As expected, the trait group for which young bulls showed the highest reliability gain was production, with an average increase across countries of +0.250, thanks to the largest reference population and the high heritability of the production traits.

Conclusions

The results presented in the current manuscript show that the machinery used for the Interbull IG-BSW Service, delivered by the Interbull Centre since 2011 for genomic evaluation of the Brown Swiss breed, is adaptable to new breeds and scenarios, providing genomic selection tools to countries with small Holstein populations.

As assessed during the development of the IG-HOL Service, genotypes provided by Contributors nearly doubled the size of the reference population, therewith significantly improving the quality of the Service.

The estimated SNP effects showed to be consistent with the parameters used in the analyses and in agreement with the across countries correlations used in the Interbull MACE Service.

The IG-HOL gEBVs resulted in a more accurate prediction of animals' true breeding values compared to the MACE EBVs/PAs, as confirmed by the higher genomic reliability obtained by applying the Interbull method developed by the appointed Interbull working group.

The presented results show the IG-HOL Service to be a valuable Service, especially for countries with small Holstein populations which cannot benefit from being part of the large international genotype-exchange consortia.

Acknowledgments

Special thanks to the Organisations joining the InterGenomics-Holstein Service, both the IG-HOL countries and the Contributors.

Thanks to the support and the advice from the Interbull IG-HOL WG, appointed by the Interbull Steering Committee, which have been crucial in achieving the release of the IG-HOL Service.

Interbull Technical Committee members are kindly acknowledged for their recommendations and guidance during the development of the IG-HOL Service and the validation of the Service's results.

In particular, the assistance provided by Dr. Zengting Liu during the implementation of the Interbull method for approximating genomic reliabilities was greatly appreciated.

Thanks also to current and former colleagues at Interbull Centre, for their valuable contribution to the development and implementation of this new Interbull Service.

References

Ben Zaabza, H., Mäntysaari, E. A., & Strandén, I. 2020. Snp_blup_rel: software for calculating individual animal SNP-BLUP model reliabilities. Agricultural and Food Science. 29(4), 297–306. https://doi.org/10.23986/afsci.95617

Liu, Z. 2011. Use of MACE Results as Input for Genomic Models. Interbull Bulletin. 43, 1-4.

Liu, Z., VanRaden, P. M., Lidauer, M. H., Calus, M. P., Benhajali, H., Jorjani, H., & Ducrocq, V. 2017. Approximating genomic reliabilities for national genomic evaluation. Interbull Bulletin. 51, 75-85.

Liu, Z., VanRaden, P. M., Lidauer, M. H., Calus, M. P., Benhajali, H., Jorjani, H., & Ducrocq, V. 2019. A technical document on derivation and application of adjustment factor for genomic reliability values.

Jorjani, H., Jakobsen, J., Nilforooshan, M., Hjerpe, E., Zumbach, B., Palucci, V., Dürr, J., 2011. Genomic evaluation of BSW populations, InterGenomics: Results and Deliverables. Interbull Bulletin. 43, 5-8.

Jorjani, H. Jakobsen, J., Hjerpe, E., Palucci, V., Dürr, J., 2012. Status of Genomic Evaluation in the Brown Swiss Populations. Interbull Bulletin. 46, 46-54.

Mäntysaari, E., Liu, Z. & VanRaden, P. 2010. Interbull validation test for genomic evaluations. Interbull Bulletin. 41, 17-21.

Sullivan, P.G. & Jakobsen, J.H. 2014. GMACE Pilot #4: Adjusting the National Reliability Input Data. Interbull Bulletin. 48, 40-45.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science. 91, 4414-4423. https://doi.org/10.3168/jds.2007-0980

VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F., Schenkel, F.S., 2009. Invited review: reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science. 92, 16-24. https://doi.org/10.3168/jds.2008-1514

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.