Multitrait Across Country Genomic Evaluations for Eurogenomics Countries

Research Plan and First Results

Hanni Kärkkäinen[1], Vincent Ducrocq[2], Sören Borchersen[3], Gert P. Aamand[4], Esa A. Mäntysaari[1]

[1] Natural Resources Institute Finland

[2] French National Institute for Agricultural Research (INRA)

[3] 3Eurogenomics Cooperation, [4]Nordic Cattle Genetic Evaluation (NAV)

Abstract

Six Eurogenomics countries have jointly established a project to demonstrate and validate the performance of a genomic multitrait across country evaluation based on shared genotypes and phenotypes. The project spans two years and was started at May 2018. The first preliminary results indicate that the model is feasible and benefits the participating countries.

Key words: genomic selection, MACE, GBLUP

Introduction

Interbull has established a SNP MACE project with Mike Goddard's group in Melbourne and Interbull Centre (Liu & Goddard, 2018). Their approach is based on the idea that if the countries could share solutions of SNP-effects and the mixed model left hand side (LHS) matrices of a SNP-BLUP model, then the mixed model equations (MME) can be built even if the genotypes are not shared. This is because the LHS and SNP-solutions can be used to resolve pseudo phenotypes: RHS= (LHS) x SNP solutions (Liu & Goddard, 2018).

However, the countries participating in Eurogenomics genotype exchange (Germany (DEU), Nordic countries Denmark, Finland and Sweden (DFS), France (FRA), The Netherlands (NLD), Spain (ESP) and Poland (POL)) share also genotypes, enabling building of a genuine multi-trait across country SNP BLUP evaluation using pseudo phenotypes from all countries directly. The Natural Resources Institute Finland (Luke), French National Institute for Agricultural Research (INRA), Vereinigte Informationssysteme Tierhaltung (Vit), German Livestock Association and the Eurogenomics cooperation, have jointly financed a two-year project to demonstrate and validate the performance of Eurogenomics SNP MACE evaluation.

A rough research plan for the project comprises at the first stage (first 12 months) to run simple multitrait SNP model across countries and validate the results. Extra developments for the first stage include using daughter yield deviations (DYD) instead of deregressed genetic predictions (DRP), estimation of correlations across countries and considerations of optimal choice for allele frequencies. In the second stage (subsequent 12 months) the project will first focus on the key issues revealed during the stage 1, and then proceed on computing individual bull reliabilities from the model, handling of external information from the third countries (via MACE proofs) and handling the different genomic models used in the countries. The project research contract was signed on April, the work started on May 2018.

Materials and Methods

Eurogenomics Data

The phenotypic data consists of national genetic evaluations (EBVs) of AI-bulls, which countries send to Interbull. Also reliabilities of the EBV and effective daughter contributions (EDC) and the trait heritabilities are provided by the countries.

The EBVs, EDCs, pedigrees and country wise heritabilities were used to compute the bull DRPs, which were then used as pseudo phenotypes in the model. The genetic correlations between countries are computed by and acquired from Interbull (http://www.interbull.org/ib/maceev_archive).

The shared Eurogenomics SNP genotypes were received from Nordic Cattle Genetic Evaluations (NAV) and were used as such. The data consists of 46,342 segregating biallelic marker genotypes, with no missing data. The set of markers is same for all of the genotyped animals, with the same 0,1,2 coding.

Traits considered

Here we have concentrated on three traits: protein yield, somatic cell score and female fertility. These traits are not only important selection criteria, but also represent different trait types. Protein yield (pro) is a simple, well behaving high heritability trait (h^2 between 0.28 and 0.48), measured fairly similarly in all of the countries. Female fertility, on the other hand, is quite a challenging trait to analyze. We selected to our analyses the "Interbull fertility trait 4", lactating cow's ability to conceive expressed as an interval trait (cc2). It has very low heritability (0.01-0.08), and it is measured in different manner in different countries: DEU, DFS, FRA and NLD send "interval from first to last insemination", while ESP and POL send "days open". The somatic cell score is between these with h^2 from 0.15 to 0.37. The heritability estimates come from the countries. Total number of animals with observations in the analyses was 35,188, 35,178 and 34,231 for pro, scs and cc2, respectively. Majority of the bulls (90% of those with protein record) have daughters only in their country of origin, but there still are more than 1000 bulls with daughters (with protein yield record) in at least 4 countries.

SNP MACE Model

The SNP MACE model equation takes form

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Zg} + \mathbf{e},$$

with $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_c)$, where $\mathbf{y}_i \in \mathbb{R}^{n_i}$ is vector of pseudo phenotypes of country *i* with n_i records; $\boldsymbol{\mu} = (\mathbf{1}^{n_1}\mu_1, \dots, \mathbf{1}^{n_c}\mu_c)$, where μ_i is the general mean of country *i* and $\mathbf{1}^{n_i}$ is a vector of n_i ones; $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_c)$, where $\mathbf{Z}_i$ is the genotype matrix of country *i*, (all countries have the same set of *m* markers with same 0,1,2 coding); $\mathbf{g} = (\mathbf{g}_1, \dots, \mathbf{g}_c)$, where $\mathbf{g}_i$ is the vector of marker effects in country *i*; and $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_c)$, where $\mathbf{e}_i$ is the vector of random residual effects of country *i* individuals.

The (co)variance structure of the model consists of $\mathrm{Var}(\mathbf{g}_i) = \sigma^2_{s_i}\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma} = \mathbf{I}^m \times 1/\sum_{j=1}^{m} 2p_j(1-p_j)$, with p_j denoting the allele frequency of locus *j*, $\sigma^2_{s_i}$ is the sire variance of country *i* and $\mathbf{I}^m$ is an identity matrix; $\mathrm{Cov}(\mathbf{g}_i, \mathbf{g}_{i^+}) = \boldsymbol{\Gamma}\,\sigma_{ii^+}$, with σ_{ii^+} denoting the genetic covariance between countries *i* and i^+; $\mathrm{Var}(\mathbf{e}_i) = \sigma^2_{e_i}\mathrm{diag}(1/\mathrm{EDC}_{ik})$, where $\sigma^2_{e_i} = \sigma^2_{s_i}(4-h^2_i)/h^2_i$, with h^2_i denoting the heritability at country *i*, and EDC_{ik} the effective daughter contribution of animal *k* at country *i*; and finally the residual covariance between countries, $\mathrm{Cov}(\mathbf{e}_i, \mathbf{e}_{i^+}) = \mathbf{0}\ \forall i \neq i^+$, as the estimated breeding values of the countries are based on the national records only.

Validation Method

In order to validate the SNP MACE method, the data was split into learning (reference population) and validation sets by bulls' birth date, so that the youngest 10% of the records at each country were assigned to the validation set. Only records with $R^2_{DRP_v} \geq 0.5$ were used in the validation (except for Poland cc2 $R^2_{DRP_v} \geq 0.3$, due to limited number of records), and animals with EDC ≥ 10 in at least 10 herds in learning. Validation animals having a record in other country's learning set were removed from the learning set.

Animal direct genomic values (DGV) were computed from the SNP solutions as $\hat{a}_{ik} = \mathbf{z}_{ik}\hat{\mathbf{g}}_i$ for animal *k* in country *i*. Validation reliability was defined as $R^2_v = \left(\mathrm{Cor}(\mathrm{DRP}_v, \mathrm{DGV}_v)\right)^2 / R^2_{\mathrm{DRP}_v}$, and bias b_1 vas tested by a weighted linear regression of DRP_v on predicted DGV_v, using EDC_v as weights (Mäntysaari *et al*. 2010). The SNP MACE prediction validations were compared to country-wise single trait SNP-BLUP validations.

Results & Discussion

The computations were performed with MiX99 release XI/2017 version 17.1107 (MiX99 Development Team, 2017). We were not perfectly happy with the convergence properties of the model, especially with low heritability trait cc2. The poor convergence resulted in long computation time (around 12h) and possibly in less reliable estimation. We computed also equivalent MACE GBLUP, which showed no problems with convergence and was much faster (around 3h). The GEBVs were practically equal to SNP MACE ones for all the traits and countries, also SNP solutions solved from GBLUP $\hat{a}$:s were consistent with SNP MACE. The results shown and discussed below are based on the equivalent GBLUP models.

For protein yield, the MACE GBLUP validation reliabilities varied between 0.53 and 0.59 (Table 1), and the single trait GBLUP reliabilities between 0.36 and 0.51, so the gain from using shared genotyped data and multitrait model is considerable (gain percentage 12-48%).
The somatic cell score has lower heritability than protein yield and, as expected, results on average lower R_v^2 under both genomic MACE and single trait GBLUP (Table 1). However, the values are more dispersed between countries, and *e.g.* Poland has higher validation reliability for scs than for protein.
The model seemed to work well even for the very low heritability female fertility trait (Table 1). The differences in R_v^2 between countries are even larger than with scs: especially Germany and Spain have very high reliabilities under both MACE and single trait model. However, interpretation of country differences needs caution because R_v^2 depends highly on the $R^2_{\mathrm{DRP}_v}$, which were based on h^2 provided by countries.
The bias estimate b_1 indicated slight to moderate inflation of the variances under both models, especially for the low heritability trait.

The single trait validation reference method is an oversimplification, as it refers to a situation where country uses only their own genotypes and phenotypes. Thus, the reference method is expected to result low reliabilities and hence make the genomic MACE model look too good in comparison. This comparison is alike in Lund *et al*. 2011, where the purpose was to quantify the benefits from country cooperation. More illustrative comparison between multitrait versus single trait use of MACE records is underway.

Conclusions

Fitting SNP MACE with individual animal genotypes is feasible, and countries gain from cooperation. Next steps:
1. More realistic validation reference method: we will quantify whether the genomic MACE is better than using (the current practice) single trait GBLUPs on MACE DRPs.
2. Consider models that account better the country wise definitions of genomic evaluations.

Acknowledgements

The project is financed by Natural Resources Institute Finland (Luke), French National Institute for Agricultural Research (INRA), Eurogenomics Cooperation and German Livestock Association.

References

Interbull MACE evaluations archive, April 2018. http://www.interbull.org/ib/maceev_archive

Liu, Z. and Goddard, M.E. 2018. A SNP MACE model for international genomic evaluation: technical challenges and possible solutions. *Proc. WCGALP* 11.393.

Lund, M.S., De Roos, A.P., De Vries, A.G., Druet, T., Ducrocq, V., Fritz, S., Guillaume, F., Guldbrandtsen, B., Liu, Z., Reents, R. and Schrooten, C., 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genetics Selection Evolution, 43(1), p.43.

MiX99 Development Team 2017. MiX99: A software package for solving large mixed model equations. Release XI/2017. Natural Resources Institute Finland (Luke). Jokioinen, Finland. URL:http://www.luke.fi/mix99

Mäntysaari, E.A., Liu, Z. and VanRaden, P., 2010. Interbull validation test for genomic evaluations. Interbull bulletin, (41), p.17.

Table 1. Validation reliability R_v^2 of DGV predicted either by single trait GBLUP or MACE GBLUP, and the gain acquired by using MACE GBLUP.

Trait	Country	Single trait	MACE GBLUP	Gain
pro	DEU	0.489	0.548	0.058
	DFS	0.361	0.534	0.173
	FRA	0.510	0.576	0.066
	NLD	0.492	0.585	0.093
	ESP	0.450	0.549	0.099
	POL	0.400	0.543	0.143
scs	DEU	0.445	0.528	0.083
	DFS	0.435	0.514	0.079
	FRA	0.419	0.500	0.081
	NLD	0.301	0.476	0.175
	ESP	0.299	0.406	0.107
	POL	0.432	0.614	0.182
cc2	DEU	0.628	0.687	0.059
	DFS	0.327	0.460	0.133
	FRA	0.393	0.525	0.132
	NLD	0.412	0.489	0.077
	ESP	0.554	0.619	0.065
	POL	0.254	0.298	0.044