Multitrait across country genomic evaluations for Eurogenomics countries

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

Interbull WS, Dubrovnik, August 25, 2018

[1]Natural Resources Institute Finland,
[2]French National Institute for Agricultural Research (INRA), France
[3]Eurogenomics Cooperation, Denmark
[4]Nordic Cattle Genetic Evaluation (NAV), Denmark

Introduction

General idea

Background:

- Interbull has established a SNP MACE project
  - Mike Goddard's group in Melbourne and Interbull Center
  - Countries can share SNP-solutions and LHS matrices, even if they do not share genotypes

  $\Longrightarrow$ MME can be build because pseudo phenotypes are: RHS$=$ inv(LHS) $\times$ SNP solutions
- Eurogenomics countries share also genotypes:

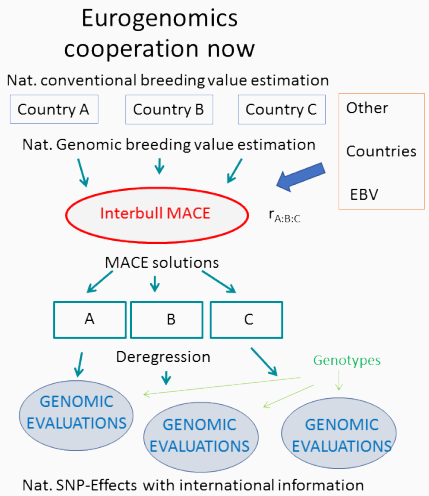
  $\Longrightarrow$ Possible to build the TRUE multi-trait across country SNP BLUP evaluation using pseudo phenotypes from all countries directly

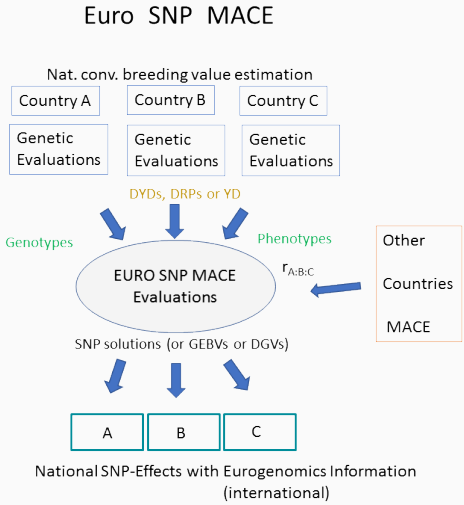
Project – 2 years in 2018-2020

- Financed jointly by Luke, INRA, Eurogenomics COOP and German Livestock Association
  - Research contract signed on April
- Post-doc Hanni Kärkkäinen started 15. May

Our goal is to demonstrate and validate the performance of Eurogenomics SNP MACE

Genomic evaluations with MACE reference *vs.* Eurogenomics SNP MACE





Rough plan

Stage 1 (12 months)

- Run simple MT SNP model across countries
  - Genotypes from 6 countries
  - Phenotypes: deregressed proofs from individual countries
    Protein, somatic cell score, female fertility trait
  - Genetic parameters from Interbull
  - Validate the model results
- Stage 1 extra developments
  - Phenotypes: DYDs are used in place of DRPs
  - Estimation of correlations across countries
  - Considerations of allele frequencies

Rough plan

Stage 2 (12 months)

- Pinpoint the development priorities using the experiences from stage 1
- Individual bull reliabilities from the model
- Handling of external information from third countries (via MACE proofs)
- Different genomic models in different countries

  Residual polygenic effects, different SNPs, haploblock models

Data

Phenotype data

- National genetic evaluation EBVs of AI sampled bulls, that countries[1] send to Interbull
- Kindly provided us by:

  Jutta Jaitner & Zengting Liu, Germany
  Ulrik Sander Nielsen & Gert Pedersen Aamand, Nordic countries
  Julie Promp & Vincent Ducrocq, France
  Pedro Vessies & Gerben de Jong, The Netherlands
  Juan Pena, Spain
  Monika Skarwecka & Andrzej Zarnecki, Poland

[1] Germany (DEU), Nordic countries Denmark, Finland and Sweden (DFS), France (FRA), The Netherlands (NLD), Spain (ESP) and Poland (POL). Order and abbreviations from Interbull practice.

Phenotype data II

- Reliabilities of the EBV, and EDC also provided
- Deregressed proofs: computed using EBV, EDC and pedigree with MiX99
- Animals with at least 10 EDC in at least 10 herds used in analyses
- Later we will use DYD's (not yet asked from countries)

Genotype data

- EG genotypes received from NAV, used "as is"
- 46342 SNP genotypes for 62628 bulls
- coded as 0,1,2

Big thanks to

Bernt Guldbrandtsen, Aarhus University, Denmark

Traits considered

1. Protein yield
   - High heritability trait (0.28 – 0.48)
2. Somatic cell score
   - Medium high heritability trait (0.15 – 0.37)
3. Female fertility
   - Lactating cow's ability to conceive
     - expressed as an interval trait
     - Interbull fertility trait 4, cc2
   - Low heritability trait (0.01 – 0.08)
   - Countries differ on submitted fertility traits:
     - DEU, DFS, FRA and NLD send "interval from first to last insemination cows (days)"
     - ESP and POL send "days open" as trait 4

Some statistics of the trait records

Trait	Country	NofAnim	heritability	meanEDC	medianEDC
pro	DEU	11322	0.48	783	124
pro	DFS	7079	0.42	496	131
pro	FRA	7969	0.30	1229	126
pro	NLD	6853	0.44	925	158
pro	ESP	4399	0.28	553	172
pro	POL	4953	0.29	237	98
scs	DEU	11308	0.23	760	120
scs	DFS	7096	0.23	445	119
scs	FRA	7949	0.15	1363	139
scs	NLD	6914	0.37	789	139
scs	ESP	4370	0.175	581	179
scs	POL	4913	0.32	212	88
cc2	DEU	11100	0.010	1272	283
cc2	DFS	7019	0.064	432	111
cc2	FRA	7658	0.041	1484	155
cc2	NLD	6857	0.073	901	198
cc2	ESP	3692	0.043	364	147
cc2	POL	3963	0.080	113	51

Country of origin

Table 2: Country of origin of the animals with protein yield record

Country of origin	DEU	DFS	FRA	NLD	ESP	POL	sum
DEU	7157	193	160	299	431	436	8676
DFS	244	5803	19	24	44	129	6263
FRA	330	52	6286	144	509	595	7916
NLD	1842	367	403	5044	906	333	8895
ESP	7	0	25	8	1325	5	1370
POL	0	0	0	0	0	2900	2900
USA	935	457	605	765	630	399	3791
CAN	344	134	340	142	377	85	1422
ITA	209	43	83	62	109	41	547
other	254	30	48	365	68	30	795

Records common to countries

Table 3: Number of protein records common to countries

	DEU	DFS	FRA	NLD	ESP	POL
DEU	11322					
DFS	960	7079				
FRA	870	712	7969			
NLD	1557	925	901	6853		
ESP	1345	837	1326	1211	4399	
POL	1002	718	690	753	836	4953

Table 4: Number and percentage of protein records common to # countries

	1	2	3	4	5	6
count	31643	1784	687	393	355	326
%	89.93	5.07	1.95	1.12	1.01	0.93

- Most of the animals with 5 or 6 records imported from USA
- Other traits show similar patterns

Methods

SNP MACE Model

- Basic SNP MACE model $\mathbf{y} = \boldsymbol{\mu} + \mathbf{Zg} + \mathbf{e}$

$$\iff \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_6 \end{bmatrix} = \begin{bmatrix} \mu_1 \mathbf{1}^{n_1} \\ \vdots \\ \mu_6 \mathbf{1}^{n_6} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1\mathbf{g}_1 \\ \vdots \\ \mathbf{Z}_6\mathbf{g}_6 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_6 \end{bmatrix}$$

- $\mathbf{y}_i \in \mathbb{R}^{n_i}$ is the pseudo phenotype (deregressed national breeding value, later DYD) for country $i \in [1, \ldots, 6]$ with n_i observations
- μ_i the general mean for country i
- $\mathbf{Z}_i \in \mathbb{R}^{n_i \times m}$ design matrix for genotypes (m is the number of markers, all countries have the same set of markers with same 0,1,2 coding)
- $\mathbf{g}_i \in \mathbb{R}^m$ estimated SNP effects for country i
- $\mathbf{e}_i \in \mathbb{R}^{n_i}$ residual effects for country i individuals
- $\text{Var}(\mathbf{g}_i) = \sigma^2_{s_i}\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma} = \mathbf{I}^m \times 1/\sum_{j=1}^{m} 2p_j(1-p_j)$ with p_j = allele frequency of locus j , $\sigma^2_{s_i}$ = sire variance of country i and $\mathbf{I}^m \in \mathbb{R}^{m \times m}$ identity matrix
- $\text{Cov}(\mathbf{g}_i, \mathbf{g}_{i^+}) = \sigma_{ii^+}\boldsymbol{\Gamma}$, where $\sigma_{ii^+} = \rho_{ii^+} \times \sigma_{s_i}\sigma_{s_{i^+}}$, with ρ_{ii^+} = genetic correlation between countries i and i^+
- $\text{Var}(\mathbf{e}_i) = \sigma^2_{e_i}\text{diag}(1/\text{EDC}_{ik}) = \mathbf{R}_i$, where $\sigma^2_{e_i} = \sigma^2_{s_i}(4 - h_i^2)/h_i^2 \quad \forall i$, for animals $k \in [1, \ldots, n_i]$
- $\text{Cov}(\mathbf{e}_i, \mathbf{e}_{i^+}) = 0 \quad \forall i \neq i^+$
  - $\star$ $\sigma^2_{s_i}$ and ρ_{ii+} from Interbull
  - $\star$ EDC_{ik} and h_i^2 from countries

SNP MACE Model – Mixed Model Equations

Let the genetic (co)variance var($\mathbf{g}$) = $\mathbf{G}$, and it's inverse $\mathbf{G}^{-1}$ be

$$\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{G}^{11} & \cdots & \mathbf{G}^{16} \\ & \ddots & \vdots \\ \textit{symm.} & & \mathbf{G}^{66} \end{bmatrix}$$

then

$$\begin{bmatrix} \ddots & & \vdots & & \vdots & & \iddots \\ & \begin{bmatrix} \mathbf{1}'\mathbf{R}_i^{-1}\mathbf{1}' & \mathbf{1}'\mathbf{R}_i^{-1}\mathbf{Z}_i \\ \mathbf{Z}_i'\mathbf{R}_i^{-1}\mathbf{1}' & \mathbf{Z}_i'\mathbf{R}_i^{-1}\mathbf{Z}_i + \mathbf{G}^{ii} \end{bmatrix} & \cdots & & \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{ii^+} \end{bmatrix} & \cdots & \\ & & \ddots & & \vdots & & \\ & & & \begin{bmatrix} \mathbf{1}'\mathbf{R}_{i^+}^{-1}\mathbf{1}' & \mathbf{1}'\mathbf{R}_{i^+}^{-1}\mathbf{Z}_{i^+} \\ \mathbf{Z}_{i^+}'\mathbf{R}_{i^+}^{-1}\mathbf{1}' & \mathbf{Z}_{i^+}'\mathbf{R}_{i^+}^{-1}\mathbf{Z}_{i^+} + \mathbf{G}^{i^+i^+} \end{bmatrix} & \cdots & \\ & & & & & & \ddots \end{bmatrix} \times \begin{bmatrix} \vdots \\ \begin{bmatrix} \hat{\mu}_i \\ \hat{\mathbf{g}}_i \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \hat{\mu}_{i^+} \\ \hat{\mathbf{g}}_{i^+} \end{bmatrix} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \begin{bmatrix} \mathbf{1}'\mathbf{R}_i^{-1}\mathbf{y}_i \\ \mathbf{Z}_i'\mathbf{R}_i^{-1}\mathbf{y}_i \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \mathbf{1}'\mathbf{R}_i^{-1}\mathbf{y}_i \\ \mathbf{Z}_{i^+}'\mathbf{R}_i^{-1}\mathbf{y}_{i^+} \end{bmatrix} \\ \vdots \end{bmatrix}$$

Validation Method

- Data was split into learning and validation sets by bulls' birth date
  - The youngest 10% from each country $\rightarrow$ validation set
- Animal solutions (DGV) were computed as $\hat{a}_{ik} = \mathbf{z}_{ik}\hat{\mathbf{g}}_i$ for animal k in country i
- Validation reliability was defined as $R_v^2 = (\text{cor}(\text{DRP}_v, \text{DGV}_v))^2 / R^2_{\text{DRP}_v}$,
  (where subscript v refers to validation set records)
- The bias b_1 was tested with a weighted linear regression of DRP_v on predicted DGV_v, using EDC_v as weights
- The SNP MACE prediction set solutions were compared to country-wise single trait SNP-BLUP solutions

Results

SNP MACE with MiX99

- Computations were performed with MiX99 release XI/2017 version 17.1107
- Not happy for the convergence properties of the model

  $\Longrightarrow$ Long computation time (around 12h for SNP MACE)
  Possibly result suffers slightly, esp. with low heritability trait cc2
- We computed also equivalent G-BLUP MACE
  - No problems with convergence
  - *Much* faster, around 3h
  - GEBVs practically equal to SNP MACE ones
    - correlation ≥ 0.98 for all traits & countries
  - SNP solutions were solved from G-BLUP $\hat{a}$:s,
    - solutions were consistent with SNP MACE

1. Protein yield

Table 5: Validation reliability R_v^2 of DGV predicted either by single trait SNP-BLUP or SNP MACE, and the gain acquired by using SNP MACE

Country	Single trait	SNP MACE	Gain
DEU	0.514	0.570	0.056
DFS	0.457	0.563	0.106
FRA	0.505	0.579	0.075
NLD	0.491	0.607	0.116
ESP	0.448	0.549	0.101
POL	0.389	0.541	0.152

- Generally reliabilities from 0.54 to 0.61 from MT
- The gain is considerable, gain percentage 11–39%
- Variances may be slightly inflated (b_1 in range $0.80 - 0.90$)

2. Somatic cell score

Table 6: Validation reliability R_v^2 of DGV predicted either by single trait SNP-BLUP or SNP MACE, and the gain acquired by using SNP MACE

Country	Single trait	SNP MACE	Gain
DEU	0.457	0.529	0.072
DFS	0.407	0.510	0.103
FRA	0.386	0.480	0.094
NLD	0.445	0.560	0.115
ESP	0.311	0.448	0.138
POL	0.478	0.611	0.133

- Lower $h^2 \Rightarrow$ slightly lower values than with protein
- Generally reliabilities from 0.45 to 0.61
- Variances may be slightly inflated (b_1 in range $0.77 - 0.89$)

3. Female fertility – cc2

Table 7: Validation reliability R_v^2 of DGV predicted either by single trait SNP-BLUP or SNP MACE, and the gain acquired by using SNP MACE

Country	Single trait	SNP MACE	Gain
DEU	0.582	0.687	0.105
DFS	0.289	0.392	0.103
FRA	0.255	0.338	0.083
NLD	0.382	0.440	0.058
ESP	0.505	0.590	0.085
POL	0.130	0.212	0.083

- Very low h^2, still method seemed to work
- Relative gain from MT approach was bigger when the single trait R_v^2 was low
- Some variances inflated (b_1 in range $0.62-0.92$)

Conclusion

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


EURO
Genomics
for reliable cattle breeding


Luke
NATURAL RESOURCES
INSTITUTE FINLAND


BUNDESVERBAND
RIND UND SCHWEIN E.V.

Thank You !


INRA
SCIENCE & IMPACT


vit
Service & Daten aus einer Quelle