
# SNP MACE

H. Jorjani, Z. Liu, M. Goddard

# Where are we now (1)?

National genomic evaluations

- Genotype of the reference population animals (bulls) shared;
- Phenotype of the reference population animals (bulls) shared;
  - **→ Residual covariance / double counting**
- Different sets of SNPs are used in different countries;
- Different models are used in different countries;
  - **→ SNP effects are not directly usable in other countries;**
- Size of the bull reference population increasing exponentially;
  - **→ Computational demand of using foreign genotypes ↑**

# Where are we now (2)?

National genomic evaluations (cont'd)

- Cow genotypes are not included in the reference population in some models, and also not exchanged widely (not yet);
- There is no MACE for cows;
  - (Too many problems to solve)
- Losing cow genotypes is drawback (less genomic pre-selection, better estimate of LD, etc.);
- Introduction of new traits is MACE dependent;
- New traits may have a small reference population;
- Multi-trait and multi-breed evaluations not so feasible, especially at the international level;

# Where are we now (3)?

And the most important of all:

The pre-requisites and underlying assumptions of MACE are in danger of being violated.

# What can be done?

- **USE SNP-MACE**!

$$g_{ij}^{NAT} = \mu_i + g_{ij}^{INT} + \varepsilon_{ij}$$

where

$g_{ij}^{NAT}$ represents SNP effects of SNP marker $j$ from country $i$,

$\mu_i$ is general mean of the $i$-th country

$g_{ij}^{INT}$ represents SNP effects from the international evaluation

$\varepsilon_{ij}$ is the residual effect

- SNP-MACE is considered to be a complement to other types of international genetic and genomic evaluations

# Mixed model equations of the SNP MACE model

$$\begin{bmatrix} \ddots & & \cdots & \cdots \\ & \left[\Psi_i + \begin{bmatrix} 0 & 0 \\ 0 & G^{ii} \end{bmatrix}\right] & \left[\Psi_{ii^+} + \begin{bmatrix} 0 & 0 \\ 0 & G^{ii^+} \end{bmatrix}\right] & \\ & \ddots & \left[\Psi_{i^+} + \begin{bmatrix} 0 & 0 \\ 0 & G^{i^+i^+} \end{bmatrix}\right] & \vdots \\ \textit{symm.} & & & \ddots \end{bmatrix} \begin{bmatrix} \cdots \\ \hat{\mu}_i \\ \hat{g}_i \\ \vdots \\ \hat{\mu}_{i^+} \\ \hat{g}_{i^+} \\ \cdots \end{bmatrix} = \begin{bmatrix} \cdots \\ \Delta_i \\ \vdots \\ \Delta_{i^+} \\ \cdots \end{bmatrix}$$

where $\Psi_i = \begin{bmatrix} 1'R_i^{-1}1 & 1'R_i^{-1}W_iZ_i \\ Z_i'W_i'R_i^{-1}1 & Z_i'W_i'R_i^{-1}W_iZ_i \end{bmatrix}$ $\Delta_i = \begin{bmatrix} 1'R_i^{-1}y_i \\ Z_i'W_i'R_i^{-1}y_i \end{bmatrix}$

Residual covariance between countries: $\Psi_{ii^+} = Z_i'W_i'R_i^{-½}R_{i^+}^{-½}W_{i^+}Z_{i^+}$

Depends on national SNP effects including foreign phenotypes (or not), and on common reference animals in case of using MACE data

In case of SNP effects estimated without using MACE data: $\Psi_{ii^+} = 0$

# General aspects and issues

- Impute genotypes of reference animals to a common set of SNP markers
  - National level
- Large, dense matrices from countries $\mathbf{Z'R^{-1}Z}$
  - National level
- Estimating country correlations of SNP effects
  - International level
- Genetic trend validation for country SNP effects
  - Converted to animal DGV for the validation?
- Mendelian sampling variance test
  - National level

# SNP-MACE Goals

- **Medium-term goals:**
  - Develop the method for medium density chips (mainly 50K): a SNP MACE model
  - Complement MACE / GAMCE / InterGEnomics

- **Long-term goals** for using sequence data
  - Bayesian methods and use of biological information
  - Identify causal sequence variants
  - Reduce the decay of linkage disequilibrium
  - Increase accuracy of SNP effects
  - Increase the variance explained by SNPs
  - Estimate country correlations more accurately (using genomic data instead of using pedigree)

# SNP MACE: Input data from countries

- SNP effect estimates from national evaluations
  - Certain groups of models can be accommodated easily
  - Other groups of models are difficult (or almost impossible) to accommodate
- LHS and RHS of equations for SNP effects: $\mathbf{Z'R^{-1}Z}$ and $\mathbf{Z'R^{-1}y}$
  - After correcting for or absorbing all other effects, including the residual polygenic effect
  - Standardized method / software
  - Data transfer issue for 50k x 50k matrices

# SNP MACE: Data at the international level

- A common set of SNP markers is desirable
- Converting the country data to the common SNP set
  - Conversion of the SNP effects assuming equal DGV of reference animals
  - $\mathbf{Z'R^{-1}Z}$ and $\mathbf{Z'R^{-1}y}$ of the converted SNP effects
- Generate between-country data: residual covariances
  - Use of MACE EBV of foreign bulls in training set
  - More common bulls lead to higher residual covariances
- If SNP effects are estimated using only national phenotypes, residual covariances are zero
- Use current $r_G$ for SNP effects (possible change of A to G)
- No de-regression step of national SNP estimates

# SNP MACE: Results back to countries

- To participating countries
  - MACE SNP effect estimates converted to country own SNP set
    - DGV calculation by countries
  - PEV of the converted MACE SNP effect estimates
    - Full PEV matrix or reliabilities of SNP effect estimates
    - Reliabilities of DGV calculated by countries
  - Complement to GMACE

- To non-participating countries
  - MACE participating countries without genomics
    - SNP effects combined with selection index and country correlations
    - One SNP at a time
    - PEV of the SNP effect estimates
    - Complement to GMACE
  - Countries not included in MACE but with own genomics (e.g. China)
    - Assuming ‘desired’ / equal correlations to all the genomic countries

# Integration to national publication

- Considering the residual polygenic effect
  - Residual polygenic variance > 0
- Combining with parental information
  - Not all animals with phenotypes are genotyped
- Single-step genomic models in national evaluation
  - Updating SNP effects and re-iterate
- Calculation of reliabilities for combined GEBV

# A validation for the SNP MACE model

- Scenario 1: Countries with a common bull reference pop.
  - Phenotypes: MACE de-regressed EBV on each scale
  - SNP effect estimates on own country scale (status quo)
  - DGV of a group common candidates for comparisons

- Scenario 2: Countries with an across-country reference bull population
  - Phenotypes: MACE de-regressed EBV on own country scale
  - SNP effect estimates on own country scale
  - The SNP MACE model vs the status quo

- Scenario 3: Countries with own national reference bull populations
  - Phenotypes: national de-regressed EBV on own country scale
  - SNP effect estimates on own country scale
  - The SNP MACE model vs the Scenario 1 applied to national DRP

# IB SNP MACE Project

What is required of the ITBC for this project?

**(Pending on the approval of IG-BSW (and possibly other sources of data)**

- Estimate rG, but A → G
- Perform a ”national” genomic evaluation for the BSW countries without using foreign genotypes;
- LHS and RHS of equations for SNP effects: $\mathbf{Z'R^{-1}Z}$ and $\mathbf{Z'R^{-1}y}$
- Estimate LD

# SNP MACE: Some challenges

- New solving algorithms for dense matrices
  - No longer sparse matrix as Animal Model
- Parallel computing across SNP markers
- Software development

# Summary

- The SNP MACE model can become a part of ITBC's service portfolio
- The SNP MACE model is a good way to integrate foreigh information into the national genomic models

- Keep the current infra-structure of NGEC
  - Optimal conventional and genomic evaluations for own national data
  - No need for direct access to national genotypes and phenotypes data
- Impact of the bias by genomic pre-selection on bulls may be avoided
- New efficient computing algorithms needed for the dense equations
  - Parallel computing
- Particularly useful for new traits with large-scale genotyped cows


# THANK YOU