“SNPMace”
(International SNP Evaluations)
Next steps towards the first pilot run

Simone Savoia
simone.savoia@slu.se
Outline

Background

Interbull SNPACE project

• Aims
• Achievements

Towards the first SNPACE pilot run

• Data call
• Next steps
Selection decisions are taken on the basis of gEBVs calculated from SNP genotypes and their estimated effects.

To improve SNP effects and gEBVs accuracy, increasing the amount of information used in the estimation process is needed.

Ideal solution: increase reference populations’ size combining across countries raw data (genotypes and phenotypes).

Restrictions and privacy regulations may limit countries in sharing such information.

International information is used in national reference populations by using MACE EBVs (genotyped and phenotype females are not included).
Combine across countries SNP effects has been considered a valuable alternative to make **genomic prediction more accurate**.

- No exchange of raw data
- Use of females’ information from national reference populations
- “difficult to measure” traits with ad-hoc genotypes animals may benefit
SNPMace project – aims

“International SNP Evaluations” as a new Interbull Service

“SNPMace” principles (Meta-analysis)

• exploiting summary statistics from countries’ national genomic evaluation:
  - LD matrices
  - SNPs’ frequencies
  - SNPs’ effects
  - error variance in the population

• Considers correlations among different populations

More accurate population-specific SNP effects
The ‘SNPMace’ methodology has been developed and the underlying model presented to the community:

\[
\begin{bmatrix}
\vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots \\
Z_i R^{-1} Z_i + G_{ii} & \cdots & G_{ii}^T \\
G_{ii}^T & \cdots & \cdots \\
\cdots & \cdots & \cdots \\
Z_i^T R^{-1} Z_i + G_i^{i+} & \cdots & G_i^{i+} \\
\end{bmatrix} \times \begin{bmatrix}
\vdots \\
\vdots \\
\vdots \\
\hat{g}_i \\
\vdots \\
\hat{g}_i^+ \\
\end{bmatrix} = \begin{bmatrix}
\vdots \\
\vdots \\
\vdots \\
Z_i^T R^{-1} y_i \\
\vdots \\
Z_i^T R^{-1} y_i^+ \\
\end{bmatrix}
\]

Goddard et al. (2018)*

A. Jighly and M. Goddard (Agriculture Victoria, Melbourne) developed an efficient software (MetaGS) to run the SNPMace model.

The meta-analysis generated gEBVs having correlation in the range of 0.99 to 1 with those obtained by using raw information.

* Additional references at the end of the presentation.
A **pilot run** is considered a crucial step prior to commencing any Interbull service:

- Using data from countries interested in joining SNPMAce
- Test the feasibility of the service and tune up the system, based on countries’ peculiarities
SNPMace data call

- **Feb 2021 – Data call**: focused on **production** and **cma** traits

- **Mar 2021 – Stage 1**: Collecting information on genomic models and set of SNPs used at national level by countries interested in joining the pilot run

- **May 2021 – Stage 2**: Review the received information and provide recommendations regarding the data preparation - SNPMace webinar

- **Jul 2021 – Stage 3**: Participating countries providing input data to Interbull Centre
Data call – Stage 1 (completed)

8 HOL Organizations

• National genomic evaluation model
• Set of SNPs used at national level
• animals’ categories in reference population and criteria for inclusion

<table>
<thead>
<tr>
<th>Country</th>
<th>N SNPs</th>
<th>Genotyped and phenotyped Females*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PROD</td>
<td>CMA</td>
</tr>
<tr>
<td>A</td>
<td>45,613</td>
<td>250,000</td>
</tr>
<tr>
<td>B</td>
<td>46,342</td>
<td>100,000</td>
</tr>
<tr>
<td>C</td>
<td>53,469</td>
<td>270,000</td>
</tr>
<tr>
<td>D</td>
<td>50,855</td>
<td>55,000</td>
</tr>
<tr>
<td>E</td>
<td>42,275</td>
<td>63,000 n.a.</td>
</tr>
<tr>
<td>F</td>
<td>46,216</td>
<td>n.a. n.a.</td>
</tr>
<tr>
<td>G</td>
<td>37,996</td>
<td>270,000 n.a.</td>
</tr>
<tr>
<td>H</td>
<td>46,161</td>
<td>36,000 15,000</td>
</tr>
<tr>
<td>Total</td>
<td>1,044,000</td>
<td>503,500</td>
</tr>
</tbody>
</table>

* In bold females included in reference population
Interbull Centre is currently working on the creation of a reference sets of SNPs (avoiding duplicates), comparing the national sets ones:

✓ “Intersection” set: including only SNPs that are present in all the countries’ sets

✓ “Union” set: including SNPs that are present in at least one country set

Several different possible SNPMace scenarios
Stage 3 provides countries to submit the following data:

I. **National SNP** effect estimates $g_i$ (for each of the investigated trait);

II. **$Z_i \ 'R_i−1Z_i$ matrix** for a measure of prediction error (co)variances of the SNP effect estimates (using an ad-hoc version of MetaGS);

III. **Marker allele frequencies** of the reference SNP allele;

IV. **Variance** of the direct genomic values.

The data received from countries participating in the pilot run will not be shared by Interbull Centre among the participants.
Possible SNPMace scenarios

**Basic** – use of “intersection” set of SNPs

- Countries’ Z’Z and Z’y reduced to the list of “intersection” set of SNPs
- Predictions provided by Interbull Centre rescaled to national sets

**Improved** – use of “union” set of SNPs and “T” matrix

- The genotypes of “n” bulls provided by each country used to design the SNPMace reference T matrix and for imputation of SNPs missing from the “union” set.
- These genotypes will not be shared with other Service Users.

**Optimal** - use of “union” set of SNPs only (investigation ongoing)

- Approximation of the Z’Z full matrix merging the Z’Z provided by each country, without using any additional genotypes.
Expected benefits from SNPMace:

- Improves the accuracy of SNP effect estimates compared to within-country SNP solutions;
- No need of sharing/exchanging/pooling genotypes on individual bulls or cow;
- Use at International level of females’ information from national reference populations.
Please, feel free to contact Interbull@slu.se if you are interested in the SNPMace project.
References
