Multitrait across country genomic evaluations for EuroGenomics countries

Hanni Kärkkäinen, Napoleon Vargas,
Martin Lidauer, Esa Mäntysaari and
EG SNP MACE working group

Starting point

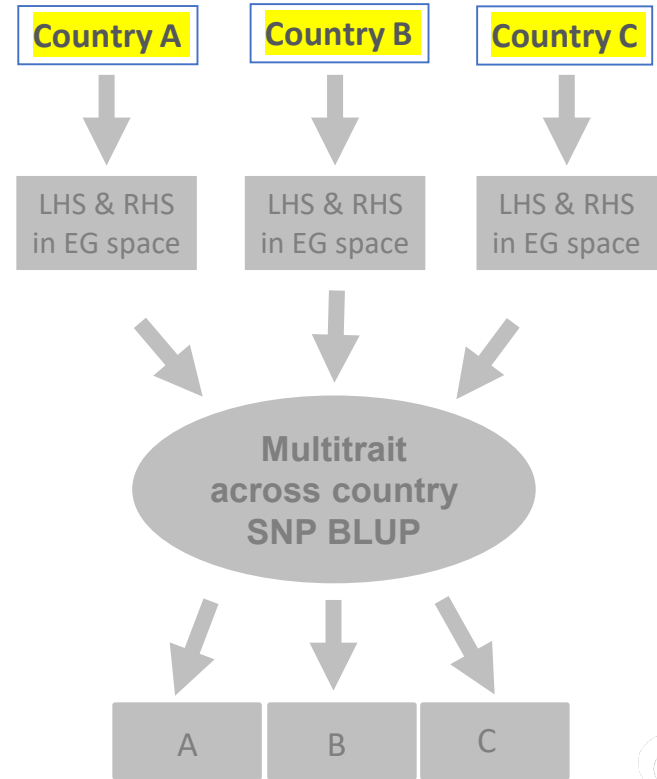The first EuroGenomics SNP MACE –project 5/2018-5/2020

- We developed a multitrait across country SNP BLUP using **EuroGenomics bull genotypes and records directly**

- Model produced on average (only) slightly higher validation reliability and was slightly less biased (higher b1) than the current EG practice

⇨ Countries wanted to **include the full national reference population** (i.e. genotyped cows) into the SNP MACE model

⇨ **Meta analysis, i.e. something similar to the Interbull SNP MACE project**

EG SNP MACE concept

1. Countries perform genomic evaluations with their own data and method $\implies$ National SNP estimates
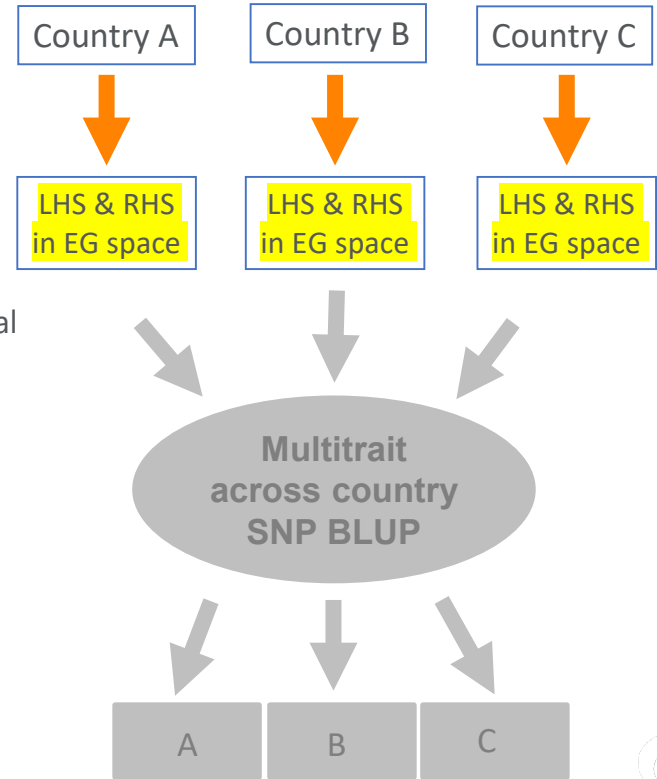
National full reference *genomic* evaluation

Country A	Country B	Country C
LHS & RHS in EG space	LHS & RHS in EG space	LHS & RHS in EG space

Multitrait across country SNP BLUP

| A | B | C |

National SNP-effects with EuroGenomics information

NATURAL RESOURCES INSTITUTE FINLAND

EG SNP MACE concept

1. Countries perform genomic evaluations with their own data and method $\Longrightarrow$ National SNP estimates

2. Countries impute the genotypes they use to the common EG SNP density

3. EG SNP MACE **preprosessor** generates blocks of MME using national full reference and national SNP estimates compliant with common EG SNP set
$\Longrightarrow$ **Meta data: LHS_{EG} and RHS_{EG}**

National full reference *genomic* evaluation

Country A	Country B	Country C

LHS & RHS in EG space | LHS & RHS in EG space | LHS & RHS in EG space

Multitrait across country SNP BLUP

A	B	C

National SNP-effects with EuroGenomics information

EG SNP MACE concept

National full reference *genomic* evaluation

1. Countries perform genomic evaluations with their own data and method $\Rightarrow$ National SNP estimates

2. Countries impute the genotypes they use to the common EG SNP density

3. EG SNP MACE preprosessor generates blocks of MME using national *full reference* and *national SNP estimates* compliant with common EG SNP set
$\Rightarrow$ **Meta data: LHS_{EG} and RHS_{EG}**

4. Countries share the meta data

5. **Meta data is plugged into across country MT SNP BLUP model**
   - Model solved to get SNP-solutions utilizing the full EuroGenomics reference population

```
Country A        Country B        Country C
   ↓                ↓                ↓
LHS & RHS        LHS & RHS        LHS & RHS
in EG space      in EG space      in EG space
   ↘                ↓                ↙
        Multitrait
        across country
        SNP BLUP
   ↙            ↓            ↘
  A             B             C
```
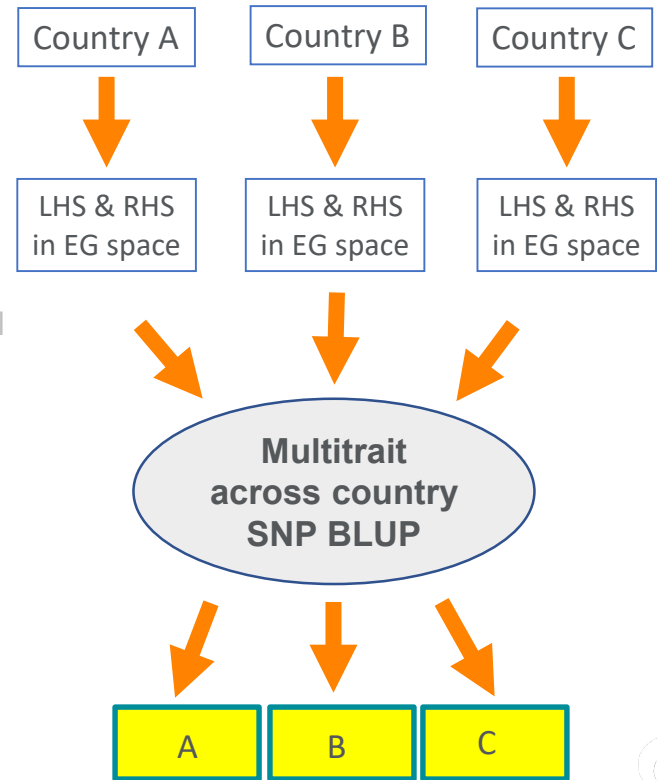
National SNP-effects with EuroGenomics information

EG SNP MACE concept

1. Countries perform genomic evaluations with their own data and method $\Longrightarrow$ National SNP estimates

2. Countries impute the genotypes they use to the common EG SNP density

3. EG SNP MACE preprosessor generates blocks of MME using national *full reference* and *national SNP estimates* compliant with common EG SNP set
$\Longrightarrow$ Meta data: LHS_{EG} and RHS_{EG}

4. Countries share the meta data

5. Meta data is plugged into across country MT SNP BLUP model
   - Model solved to get SNP-solutions utilizing the full EuroGenomics reference population

6. Results of country specific SNP-solutions are converted back to national SNP set space and blended into national evaluation

National full reference *genomic* evaluation

National SNP-effects with EuroGenomics information

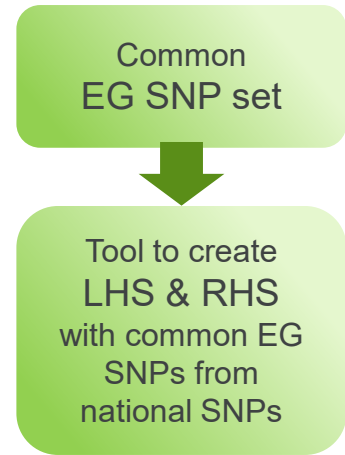# Building blocks for EG SNP MACE

> Common
> EG SNP set

Common EG SNP set

- Two sets were established: **Intersection** of national loci and **Union** of national loci
  - Selected loci must also belong to either Illumina50k v3 or EG MD v2 chip

- **Union** set requires that countries impute the EG SNPs not in their national set

- During the project we used only the **Intersection** set, i.e. loci used by all countries

Building blocks for EG SNP MACE

Tool for transforming national equations into common EG space

- Included in the Preprocessor program
  that builds the meta data equations

- Convert national LHS and RHS into common ones
  that can be used in SNP MACE – and back to national ones

- Program distributed to countries, that compute and send the LHS and RHS

Common
EG SNP set

Tool to create
LHS & RHS
with common EG
SNPs from
national SNPs

Building blocks for EG SNP MACE

SNP MACE solver program

- Uses meta data
- Produces SNP solutions with EuroGenomics information

- Computes also absorption matrices needed for integrating the SNP MACE solutions into national evaluations
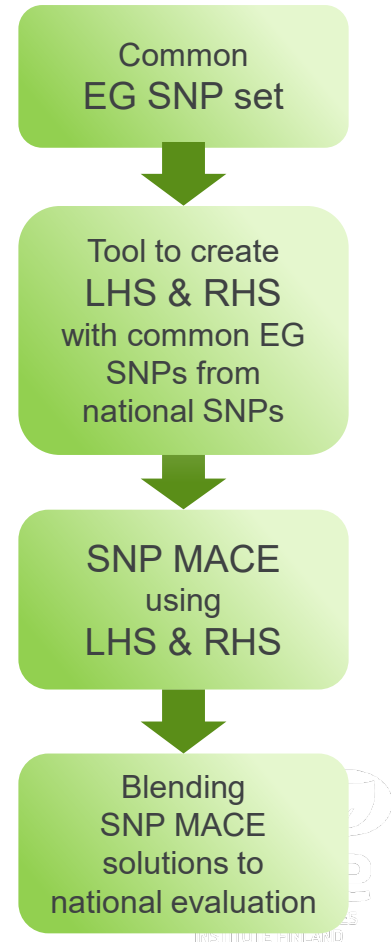
Common
EG SNP set

Tool to create
LHS & RHS
with common EG
SNPs from
national SNPs

SNP MACE
using
LHS & RHS
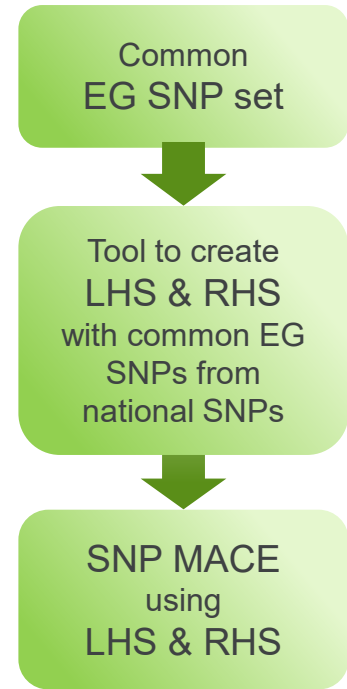
Building blocks for EG SNP MACE

SNP MACE solver program

- Uses meta data
- Produces SNP solutions with EuroGenomics information

- Computes also absorption matrices needed for integrating the SNP MACE solutions into national evaluations

Blending the SNP MACE solutions into national evaluations

- Mathematical theory developed, not tested in practice

```
Common
EG SNP set
   ↓
Tool to create
LHS & RHS
with common EG
SNPs from
national SNPs
   ↓
SNP MACE
using
LHS & RHS
   ↓
Blending
SNP MACE
solutions to
national evaluation
```

Theory is easy – practical implementation not so

Everything works great in controlled situation

- We pilot tested and validated the method using the shared bull data from the 1st project

With real full reference meta data

- The national evaluation systems differ in many ways, that must be understood and considered
- Otherwise, the meta data equations are not compatible with each other

- The method was validated using three countries' data,
  that seemed to produce reasonably logical results together

- But we weren't able to harmonize all meta data sets within the project time

Validation

Limitation to validation set up:
Full reference data equations consists of all animals,
and can't be divided into reference/validation sets*

Purpose of the validation is to find out:
Does it pay to include genotyped cows
(=full national reference population) into the SNP MACE

$\longrightarrow$ **Compare to shared bull data –based SNP MACE**
(*i.e.*, to the method developed in the 1st EG SNP MACE project)

*more realistic validation would be possible but would require extra work from the countries: They should compute national SNP solutions and compile equations also for truncated data set, leaving the validation set out.
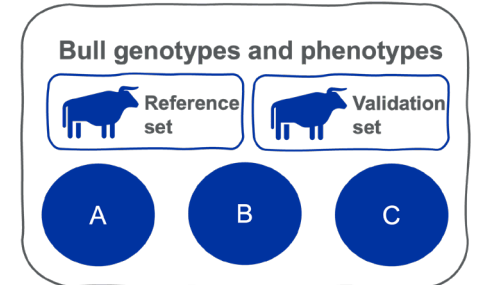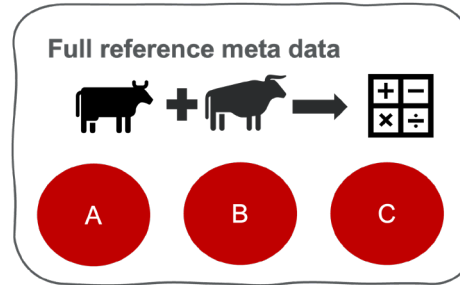
Validation scheme
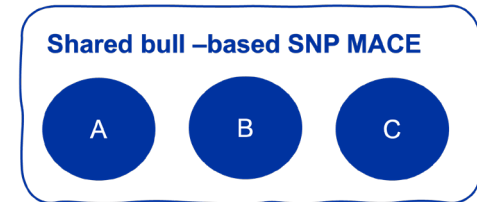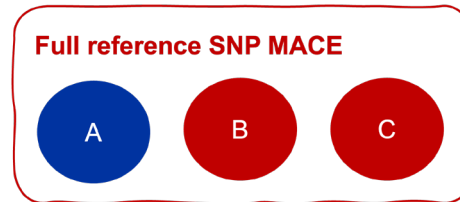## for the value of full reference compared to shared bulls

For each *Country* **at the time**, we

a. Compute SNP solutions using
    1) **Full reference** pseudo **data from** <u>other countries</u> and
    2) 90% (**reference group**) of the **shared bull data of *The Country***
       (because the full reference equations can't be divided into validation and reference groups)

b. **As competing method,** we compute SNP solutions with a **shared bull data –based SNP MACE**
    • *i.e.,* the method developed in the 1st EG SNP MACE project

c. **Predict youngest of the shared bulls of *The Country*** (10% of the data)
   • Using SNP solutions from the methods **a.** or **b.**

Caveat: Solutions of *The Country* do not use that Country's own cow data

Data sets:

Full reference meta data

A   B   C

Bull genotypes and phenotypes

Reference set   Validation set

A   B   C

Example: validation for country A:

Full reference SNP MACE

A   B   C

Shared bull –based SNP MACE

A   B   C

Validation reliability:
Impact of Full reference population

Protein yield

Stature

Fertility*
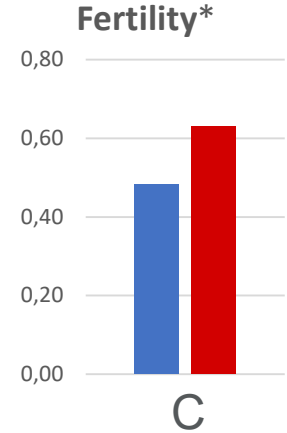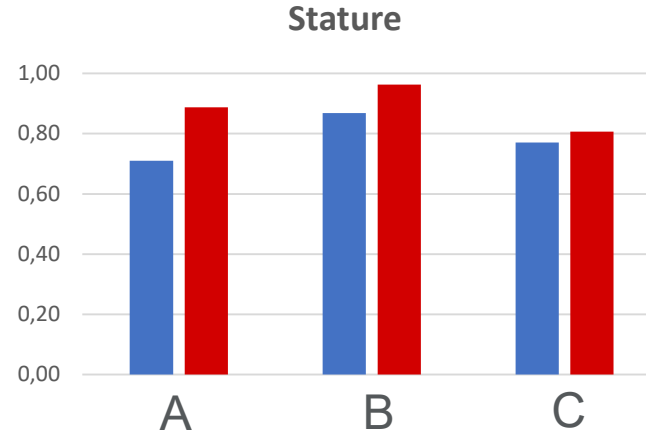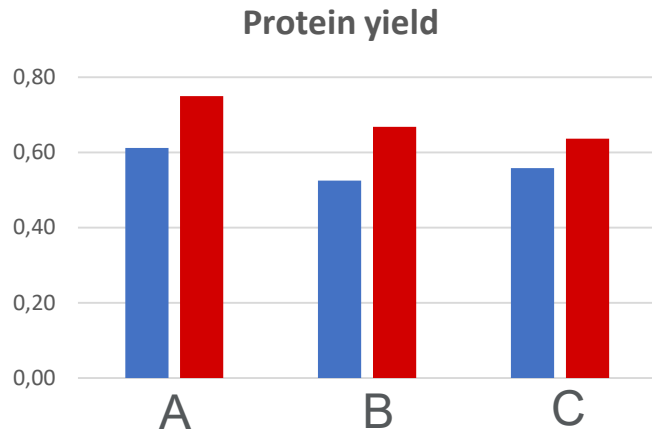
ITB validation reliability $\frac{R^2}{weight}$

- Where R^2 is the coefficient of determination from a weighted linear regression of DGV on DRP

- Only validation set animals with weight > 0.5 used

 Shared bull –based SNP MACE

 Full reference SNP MACE

*there was an issue with some of the fertility meta data, hence validation only for one country

Conclusions

- **Full reference SNP MACE can be done using meta data**

- The validation showed that adding full reference information into SNP MACE **improves accuracy greatly**

- We have accomplished

  1. **SNP MACE solver program**
  2. **Method to include the SNP MACE solutions into national evaluations**
  3. **Program that converts the national equations into the common EG SNP set space and the SNP MACE solution back to national space**
  4. **Common EG SNP set**
  5. **Lots of understanding of the national evaluation systems, and their impact to the combined evaluation**

- Main thing still needed is **more understanding** of the national systems and a **method to account for** that information

Luke
NATURAL RESOURCES
INSTITUTE FINLAND

Take home:

1. **Full reference SNP MACE can be done using meta data and it seems to improve accuracy greatly**

2. **Requires lots of understanding of the national systems and methodology to account for the differences and changes**

Thank you!