



**General discussion:**  
**Technical challenges and possible solutions**

Zengting Liu



## Discussion topics

- Statistical methods
- Input data from countries
- Data needed at international level
- International evaluation and validation
- Results provided to countries
- Integration to national publication
- General aspects



## Statistical methods

- Short-term and medium-term goals
  - More accurate genomic prediction and identify useful sequence variants
- Short-term for using current chips (mostly 50K): a SNP MACE model
- Medium-term for using sequence data
- Medium-term project: estimate country correlations more accurately using genomic data than using pedigree



## Medium-term research projects

- Optimal use of 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



## SNP MACE: Input data from countries

- SNP effect estimates from national evaluations
  - SNP BLUP models (single- or multiple steps)
  - GBLUP models (single- or multiple steps)
    - Back solve GEBV to SNP effects
  - Bayesian variable selection SNP models
  - French genomic model



## SNP MACE: Input data from countries

- LHS and RHS of equations for SNP effects:  $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$  and  $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}$ 
  - After correcting for or absorbing all other effects, including the residual polygenic effect
  - Countries may not have the data in national evaluation
  - Standardized method / software
  - Data transfer issue for 50k x 50k matrices



## Input data from countries

- A common set of SNP markers is desirable
- Converting the country data to the common SNP set
  - Conversion of the SNP effects based on equal DGV of reference animals
  - $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$  and  $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}$  of the converted SNP effects



## Data needed at international level

- Generate between-country data: residual covariances
  - Use of MACE EBV of foreign bulls in training set
  - More common bulls lead to higher residual covariances
  - Similar to GMACE (Sullivan, 2016)
- If SNP effects are based on national phenotypes, residual covariances are zero
- Use current country correlations for SNP effects
- No de-regression step of national SNP estimates





# International SNP effect estimation

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



## A special validation

- Cross validation by random selection of data
- Forward prediction by data truncation
- A special validation for sharing reference bulls
  - Scenario 1: countries sharing RP and using MACE data (status quo)
  - Scenario 2: the SNP MACE model applies to the countries using national phenotypes only
  - Scenario 3: the SNP MACE model applies to the countries using MACE EBV of shared reference bulls



## Results 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



## Results 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
- Countries not included in MACE but with own genomics
  - Assuming 'desired' / equal correlations to all the genomic countries



## Integration to national publication

- Considering the residual polygenic effect
  - SNP markers do not explain all genetic variation
- Combining with parental information
  - Not all animals with phenotypes are genotyped
- Countries using single-step models
- Reliabilities for combined GEBV



## General aspects

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



- Thank you for your attention!



## Why the international SNP model?

- Optimal use of cow reference populations
  - Less affected by pre-selection than bulls
  - Many cows give more information on LD than their sires
  - Exchanging genotypes of >100,000 cows infeasible
  - By-passing international cow MACE evaluation
  - Using foreign reference cow information
- For new traits even more advantages
- More accurate evaluation using sequence data and across breeds