Integration of foreign estimates of SNP effects into a domestic SNPBLUP

J. Vandenplas, M.P.L. Calus, G. Gorjanc

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

ROSLIN

Introduction

- Genomic evaluation
  - Aim: more accurate genomic EBVs

- SNP-based evaluations under study/testing

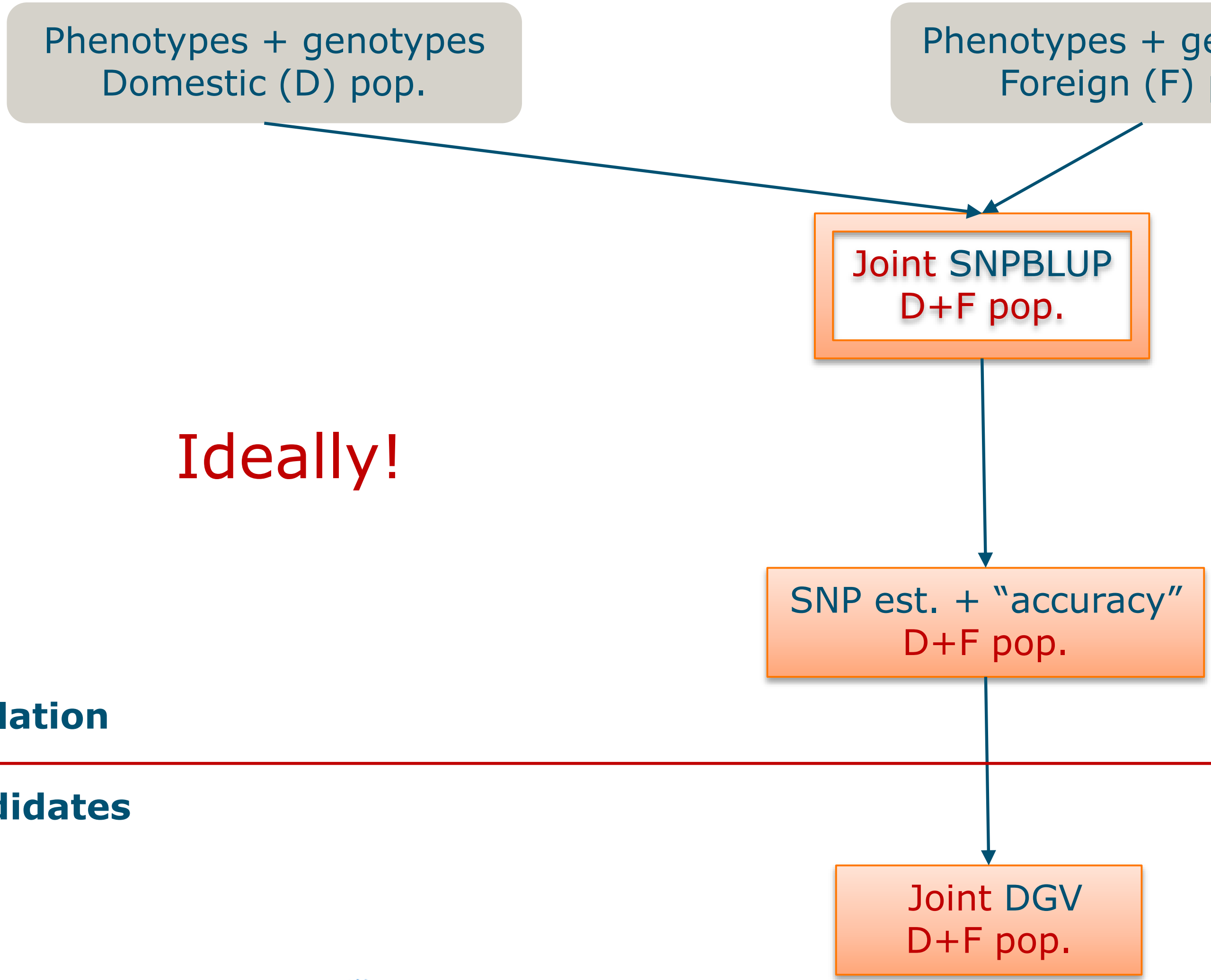
→Future: exchange of estimates of SNP effects?

→How to integrate them into SNPBLUP?

Aim

Developing and testing procedures to integrate estimates of SNP effects and measures of precision from a foreign SNPBLUP into a domestic SNPBLUP

Methods – joint SNPBLUP



Phenotypes + genotypes
Domestic (D) pop.

Phenotypes + genotypes
Foreign (F) pop.

Joint SNPBLUP
D+F pop.

SNP est. + "accuracy"
D+F pop.

Joint DGV
D+F pop.

Ideally!

Training population

Selection candidates

Methods – joint SNPBLUP

$$\begin{bmatrix}\mathbf{y}_d\\ \mathbf{y}_f\end{bmatrix}=\begin{bmatrix}\mathbf{X}_d & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_f\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_d\\ \boldsymbol{\beta}_f\end{bmatrix}+\begin{bmatrix}\mathbf{Z}_d\,\mathbf{W}_d\\ \mathbf{Z}_f\,\mathbf{W}_f\end{bmatrix}\boldsymbol{\alpha}+\begin{bmatrix}\mathbf{e}_d\\ \mathbf{e}_f\end{bmatrix}$$

$$\boldsymbol{\alpha}\sim MVN\left(\mathbf{0},\mathbf{I}\sigma^2_{\alpha_J}\right) \qquad \begin{bmatrix}\mathbf{e}_d\\ \mathbf{e}_f\end{bmatrix}\sim MVN\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix},\begin{bmatrix}\mathbf{R}_d & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_f\end{bmatrix}\sigma^2_e\right)$$

$\mathbf{y}_i$ = *vector of phenotypes*

$\boldsymbol{\beta}_i$ = *vector of fixed effects*

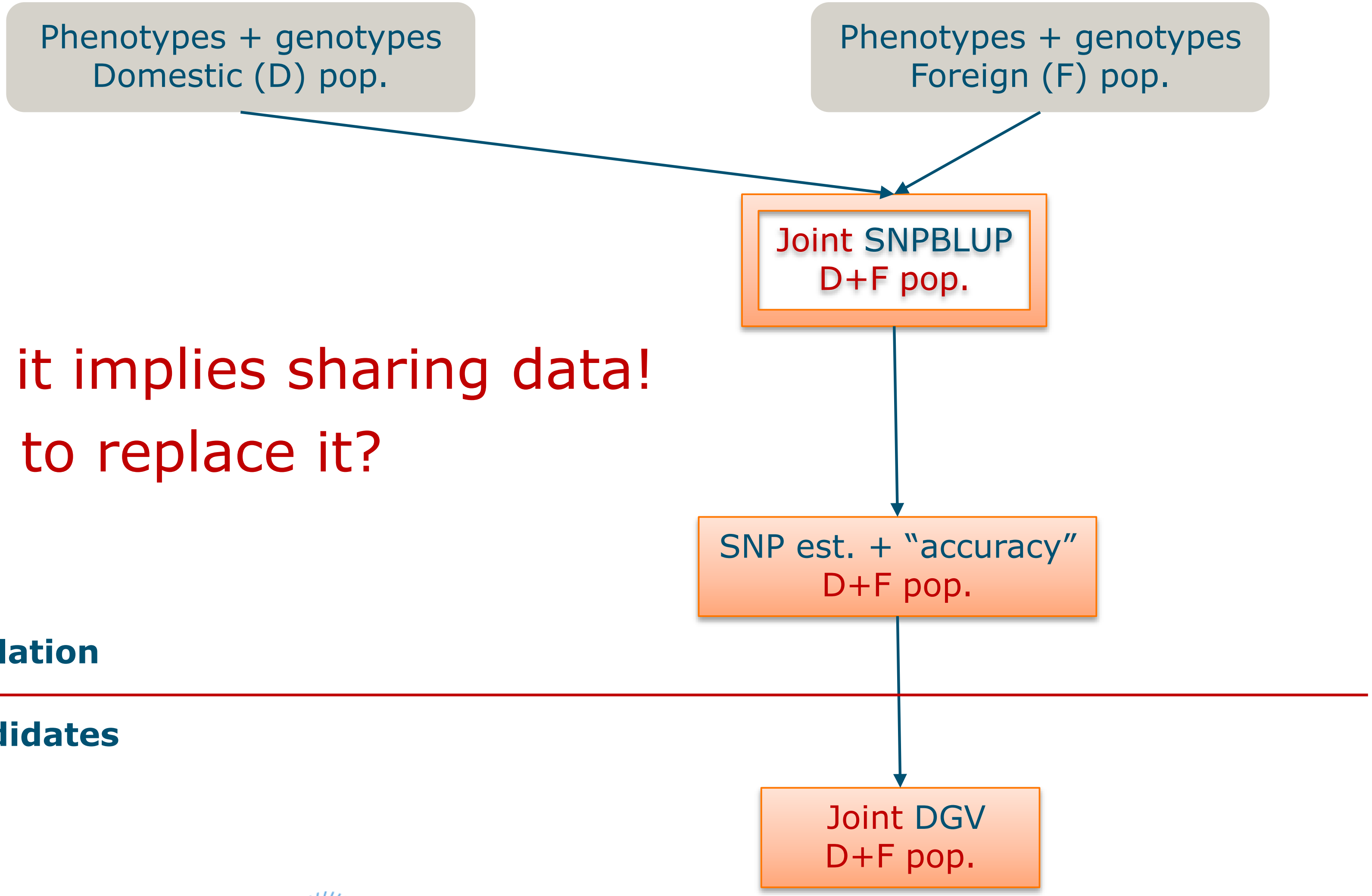
$\boldsymbol{\alpha}_i$ = *vector of SNP effects*

$\mathbf{e}_i$ = *vector of residuals*

$\mathbf{W}_i$ = *matrix of SNP genotypes*

$\mathbf{X}_i$, $\mathbf{Z}_i$ = *incidence matrices*

Methods – joint SNPBLUP



Phenotypes + genotypes
Domestic (D) pop.

Phenotypes + genotypes
Foreign (F) pop.

Joint SNPBLUP
D+F pop.

Issue: it implies sharing data!

➔How to replace it?

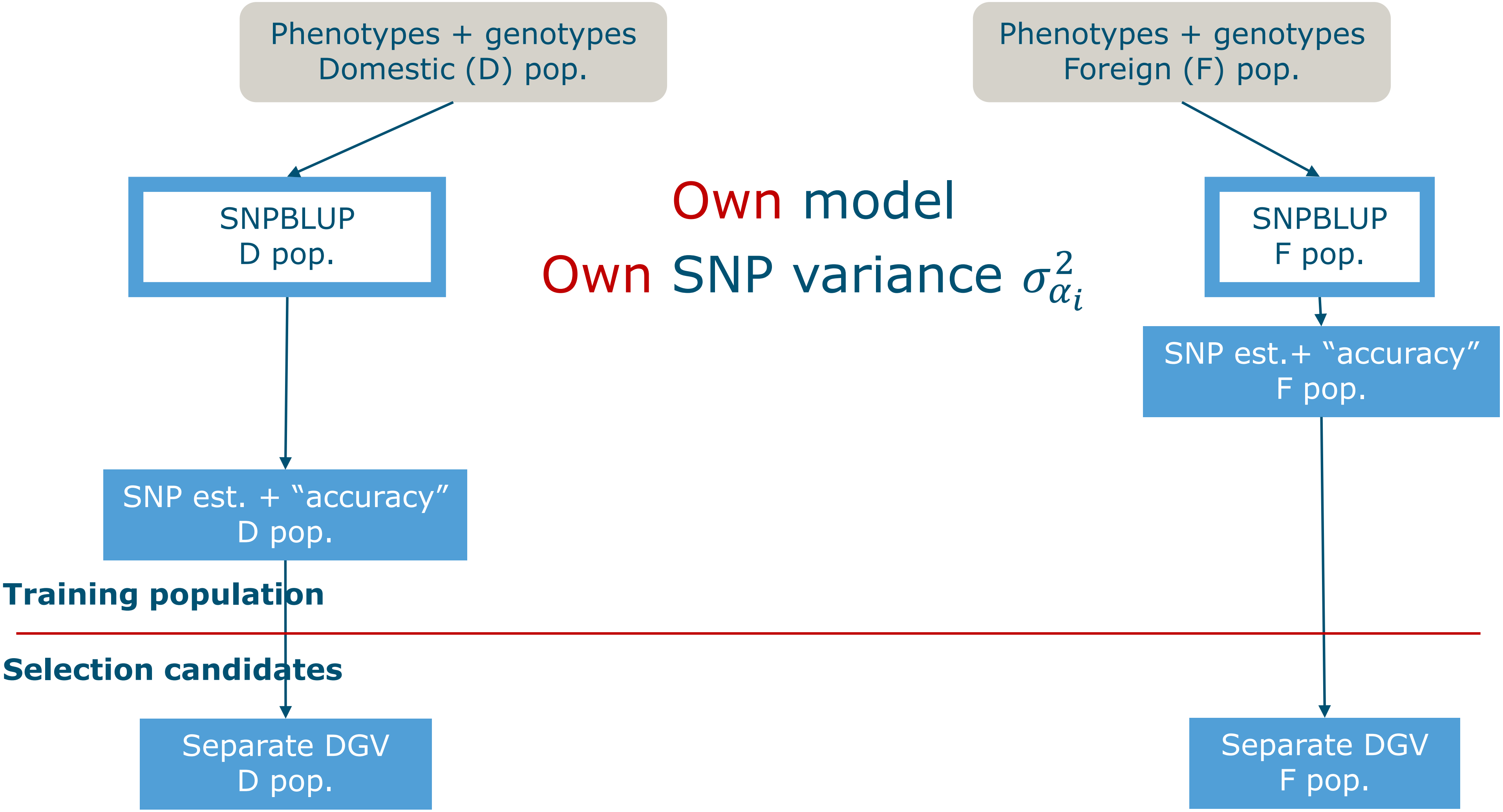
SNP est. + “accuracy”
D+F pop.

Training population

Selection candidates

Joint DGV
D+F pop.

Methods – separate SNPBLUP



Phenotypes + genotypes
Domestic (D) pop.

Phenotypes + genotypes
Foreign (F) pop.

SNPBLUP
D pop.

Own model

Own SNP variance $\sigma^2_{\alpha_i}$

SNPBLUP
F pop.

SNP est.+ "accuracy"
F pop.

SNP est. + "accuracy"
D pop.

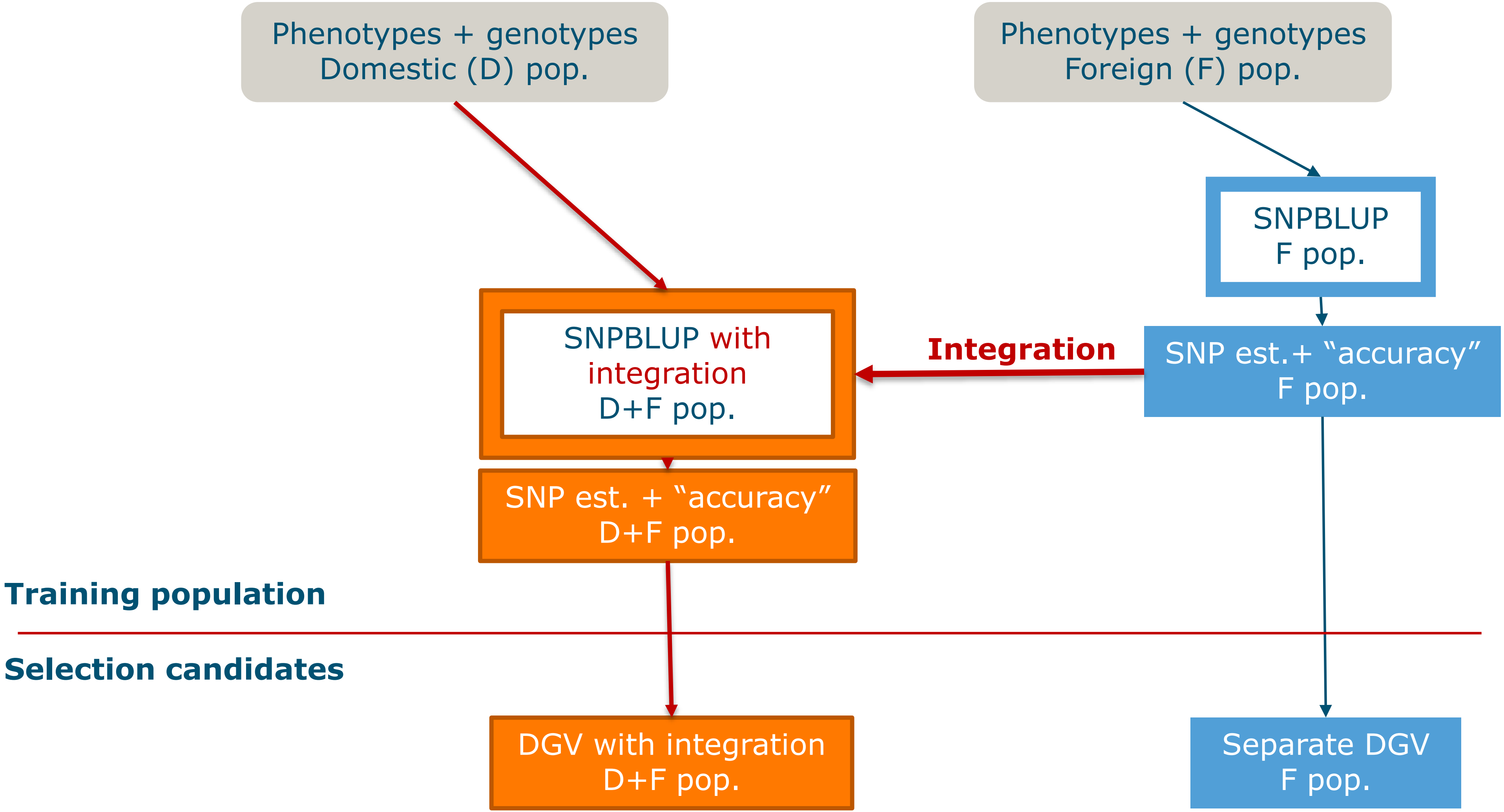
Training population

Selection candidates

Separate DGV
D pop.

Separate DGV
F pop.

Methods – SNPBLUP with integration



Phenotypes + genotypes
Domestic (D) pop.

Phenotypes + genotypes
Foreign (F) pop.

SNPBLUP
F pop.

SNPBLUP with integration
D+F pop.

Integration

SNP est.+ "accuracy"
F pop.

SNP est. + "accuracy"
D+F pop.

Training population

Selection candidates

DGV with integration
D+F pop.

Separate DGV
F pop.

Methods –SNPBLUP with integration

- Assumptions
  - Same model/variances $(\sigma_e^2$ & $\sigma_{\alpha_J}^2)$ as joint SNPBLUP
  - Same genotype (scaling) across all SNPBLUP

Methods –SNPBLUP with integration

$$\begin{bmatrix} \mathbf{X}'_d\mathbf{X}_d\,\sigma_e^{-2} & \mathbf{X}'_d\mathbf{Z}_d\,\mathbf{W}_d\,\sigma_e^{-2} \\ \mathbf{W}'_d\mathbf{Z}'_d\mathbf{X}_d\,\sigma_e^{-2} & \mathbf{W}'_d\mathbf{Z}'_d\mathbf{Z}_d\,\mathbf{W}_d\,\sigma_e^{-2} + \underbrace{\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1} - \mathbf{I}\sigma_{\alpha_f}^{-2}}_{\mathbf{W}'_f\mathbf{Z}'_f\mathbf{M}_f\,\mathbf{Z}_f\,\mathbf{W}_f\;\sigma_e^{-2}} + \mathbf{I}\sigma_{\alpha_J}^{-2} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}_d} \\ \widehat{\boldsymbol{\alpha}} \end{bmatrix} =$$

Least-squares part of the foreign SNPBLUP

$$\begin{bmatrix} \mathbf{X}'_d\mathbf{y}_d\,\sigma_e^{-2} \\ \mathbf{W}'_d\mathbf{Z}'_d\mathbf{y}_d\,\sigma_e^{-2} + \underbrace{\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1}\widehat{\boldsymbol{\alpha}_f}}_{\mathbf{W}'_f\mathbf{Z}'_f\mathbf{M}_f\,\mathbf{y}_f\;\sigma_e^{-2}} \end{bmatrix}$$

RHS of the foreign SNPBLUP

➔Several ways to approximate $\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1}$

Methods – approximations of $\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1}$

1) No approximation *(reference)*: $\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1}$

Methods – approximations of $\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1}$

1) No approximation *(reference)*: $\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1}$

2) $\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1} \approx \left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)_{within\_chromosome}\right)^{-1}$

Methods – approximations of $\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1}$

1) No approximation *(reference)*: $\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1}$

2) $\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1} \approx \left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)_{within\_chromosome}\right)^{-1}$

3) $\left(PEC\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1} \approx \left(PEV\left(\widehat{\boldsymbol{\alpha}_f}\right)\right)^{-1}$

Methods – approximations of $\left(PEC(\widehat{\boldsymbol{\alpha}_f})\right)^{-1}$

1) No approximation *(reference)*: $\left(PEC(\widehat{\boldsymbol{\alpha}_f})\right)^{-1}$

2) $\left(PEC(\widehat{\boldsymbol{\alpha}_f})\right)^{-1} \approx \left(PEC(\widehat{\boldsymbol{\alpha}_f})_{within\_chromosome}\right)^{-1}$

3) $\left(PEC(\widehat{\boldsymbol{\alpha}_f})\right)^{-1} \approx \left(PEV(\widehat{\boldsymbol{\alpha}_f})\right)^{-1}$

4) $\left(PEC(\widehat{\boldsymbol{\alpha}_f})\right)^{-1} \approx \left(\boldsymbol{\Lambda}_f \left(f(\mathbf{LD}_f, \mathbf{p})\right) \boldsymbol{\Lambda}_f \sigma_e^{-2} + \mathbf{I}\sigma_{\alpha_f}^{-2}\right)$

$\mathbf{p}$: allele frequencies in the training set

$\mathbf{LD}_f$ computed from foreign selection candidates

$\boldsymbol{\Lambda}_f$: effective number of records per SNP

- Estimated from $PEV(\widehat{\boldsymbol{\alpha}_f^*})$, $\mathbf{LD}_f$, and $\mathbf{p}$

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

ROSLIN

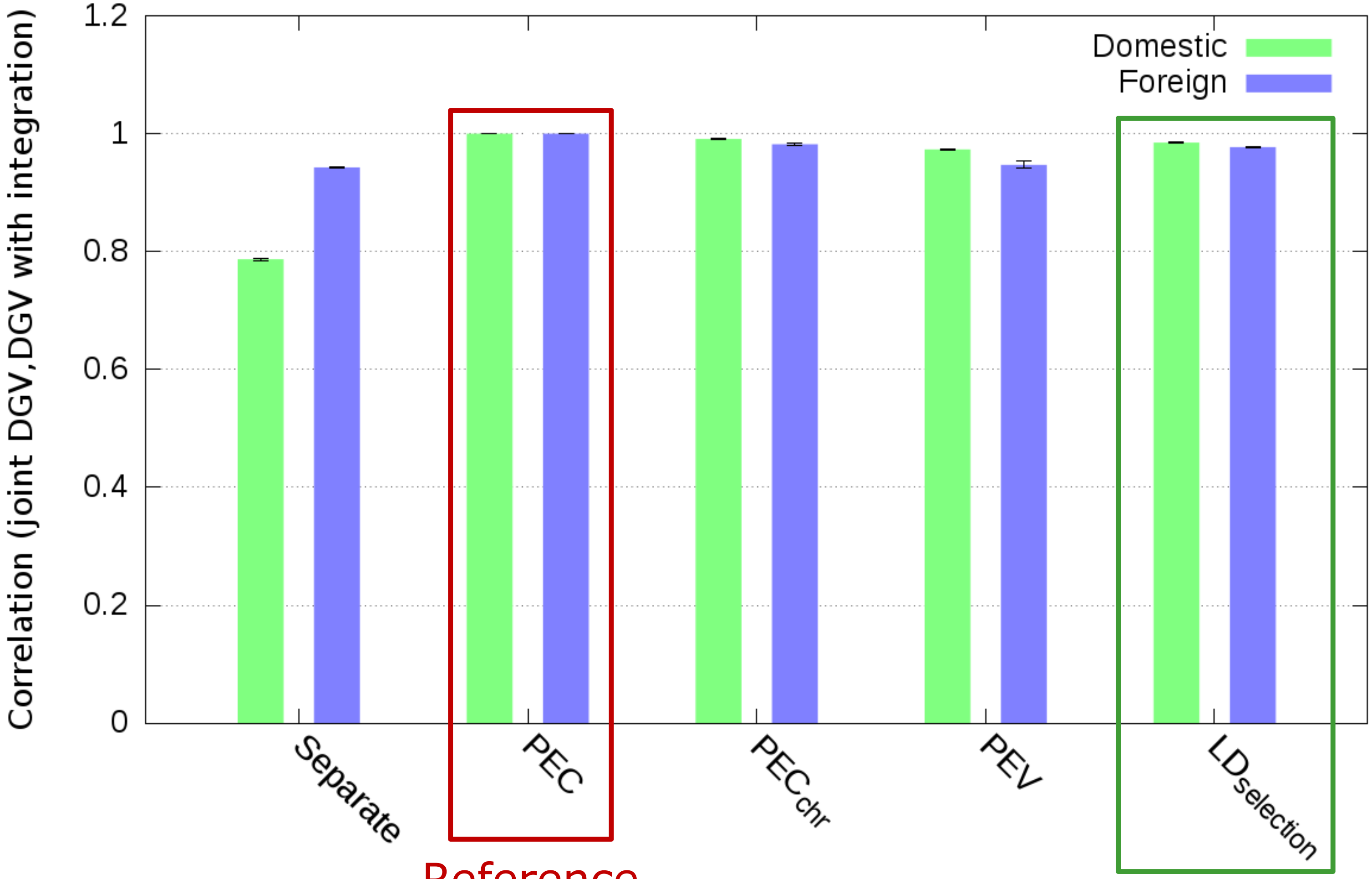
Simulation

- 2 Holstein-like populations
  - 1 trait (h^2 = 0.30 - 60K SNPs)

- Training populations
  - 5,000 animals / population
  - Randomly sampled from gen. 1 to 6
  - Domestic: own performance records
  - Foreign: pseudo-records (~DYD, DRP) + weights

- Selection candidates
  - 10,000 animals from gen. 7 / population

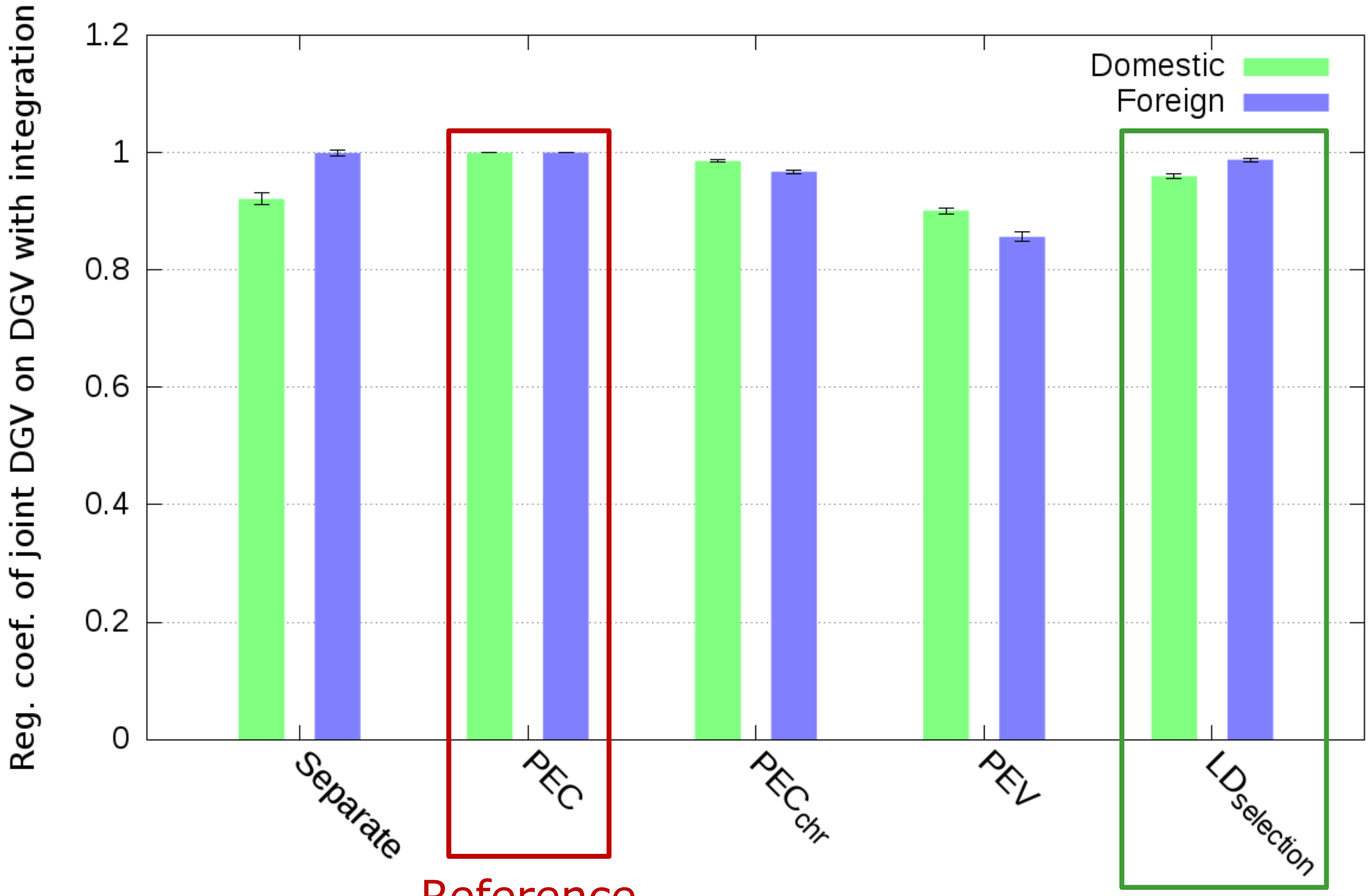
Results – correlations



Reference

- Accurate integration
  - Even with only PEV and LD information

Results – bias



Reference

Almost no bias, except for PEV

Conclusions

- Accurate integration of estimates of SNP effects
  - Without exchanging genotypes/phenotypes

- Procedure similar to integration of foreign EBVs
  - ➔Similar assumptions/issues/solutions

- Easy extensions
  - Multiple populations, multiple traits, ...
  - Special case: SNP-MACE

Thank you!

BREED4FOOD

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

ROSLIN