Sparse single-step genomic Best Linear Unbiased Prediction in crossbreeding schemes

J. Vandenplas, M.P.L. Calus, J. ten Napel

WAGENINGEN UR
For quality of life

BREED4FOOD

BREED4FOOD

Breed4Food is dedicated to be the leading research consortium in animal breeding, genetics and genomics enabling the Breed4Food partners to breed better products to benefit society's needs.

WAGENINGEN UR
For quality of life

HENDRIX GENETICS

Cobb

TOPIGS

CRV

Single-step GBLUP

Unified BLUP procedure considering simultaneously genotyped and non-genotyped animals

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A_{22}}^{-1} \end{bmatrix}$$

Inverse of the **pedigree-based** relationship matrix

Inverse of the **genomic-based** relationship matrix

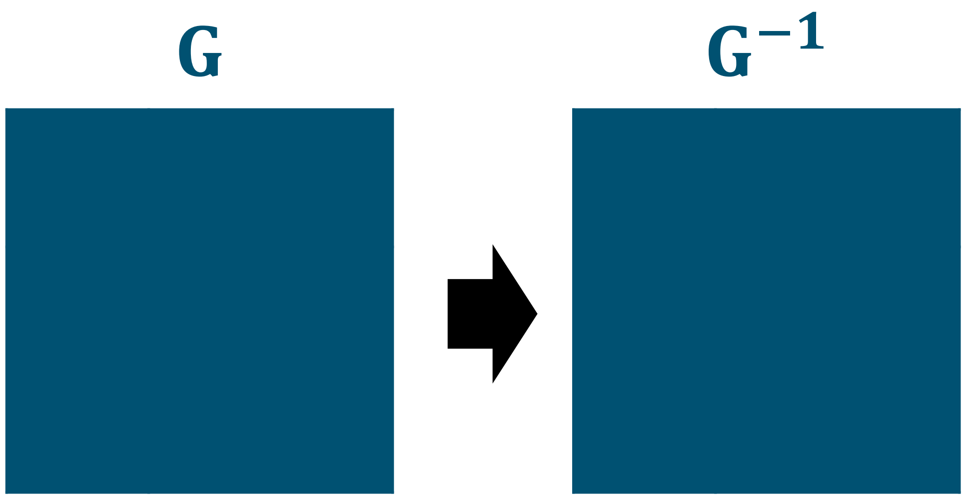
Inverse of the **pedigree-based** relationship matrix associated with **genotyped animals**

Creation and inversion of $\mathbf{G}$ and $\boldsymbol{A}_{22}$

ssGBLUP more and more implemented

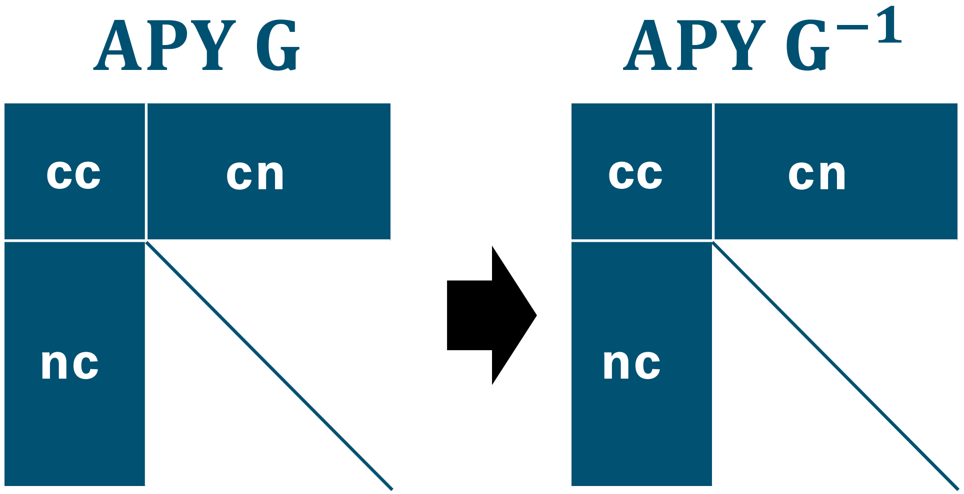
BUT

infeasible for > ~100,000 genotyped animals?



- **High computational costs**
- **High storage (dense $\mathbf{G}^{-1}$)**

Algorithm for Proven and Young animals



- **Core / non-core animals**
- **Lower computational costs**
- **Lower storage**

→ Limitations removed

Misztal et al., 2016

Aim

APY successfully tested with single-breed data

Applicable for crossbreeding schemes?

e.g., Fragomeni et al., 2015; Masuda et al, 2015; Ostersen et al., 2016

Simulation - pedigree

- 3 breeds (A, B, C)
  - Unrelated / Related
  - 10 discrete generations
    - 30 males – 6,250 females (Ne ~ 120)
- Crossbreds (BC and A(BC))
  - 30 males – 6,250 females from gen. 6 to 9
- Random matings
- Litter size: 2

Simulation - genotypes

- 89,000 genotypes randomly sampled
    - Generations 6 to 10

    - ~21,000 / breed
    - 10,000 BC
    - 16,000 A(BC)

    - 18 chromosomes - ~52,000 SNP

Simulation - phenotypes

- 126,000 phenotypes randomly sampled
    - 1 trait/breed composition
        - Polygenic effect: 5%

    - Generations 5 to 9

    - ~33,333 / breed
    - All genotyped BC and A(BC) = phenotyped

5-trait ssGBLUP

- **G** matrix
  - Allele frequencies across all animals
  - 5% polygenic

- Random core animals
  - A, B, C animals
  - A, B, C, BC, and A(BC) animals
  - Total: 4000 - 13,000 (15,000)

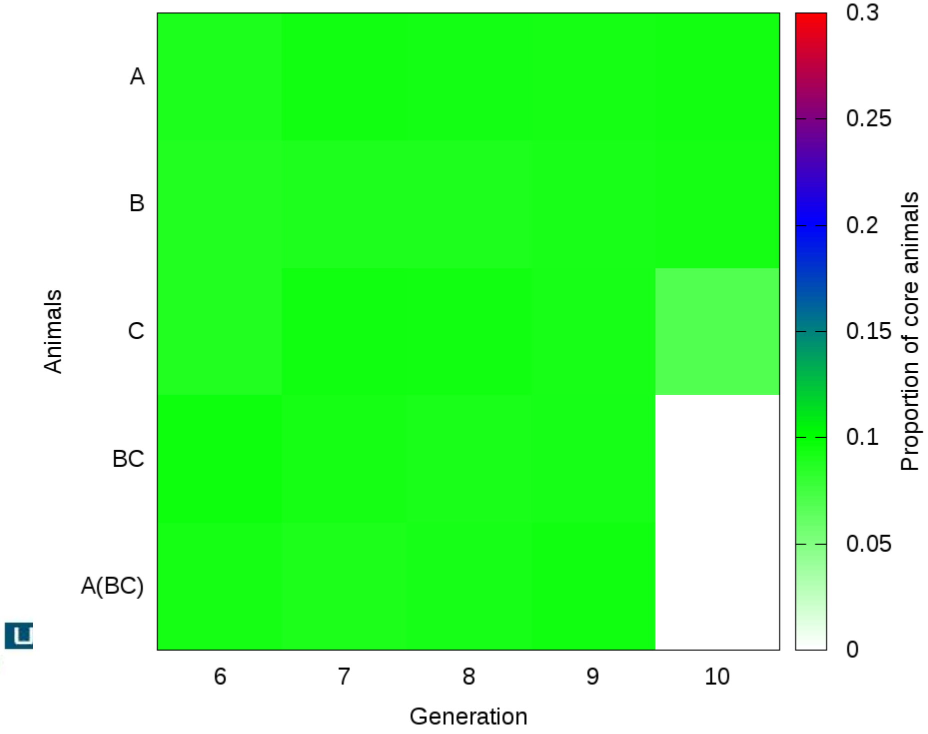
- 5 replicates

Core animals - representation

Due to properties of the genotype dataset

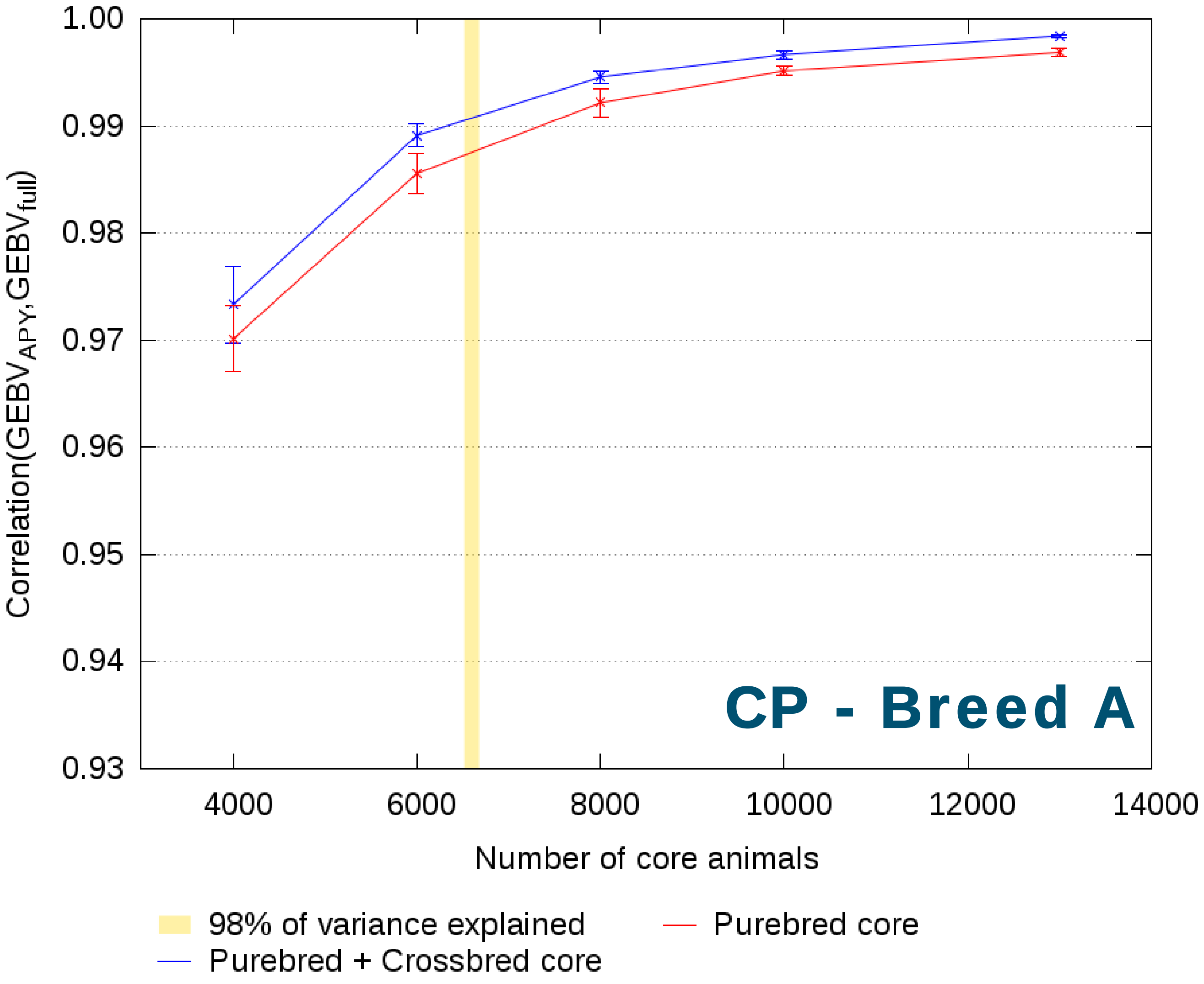
Similar representation of genotyped animals in

all generations (and all breed compositions)



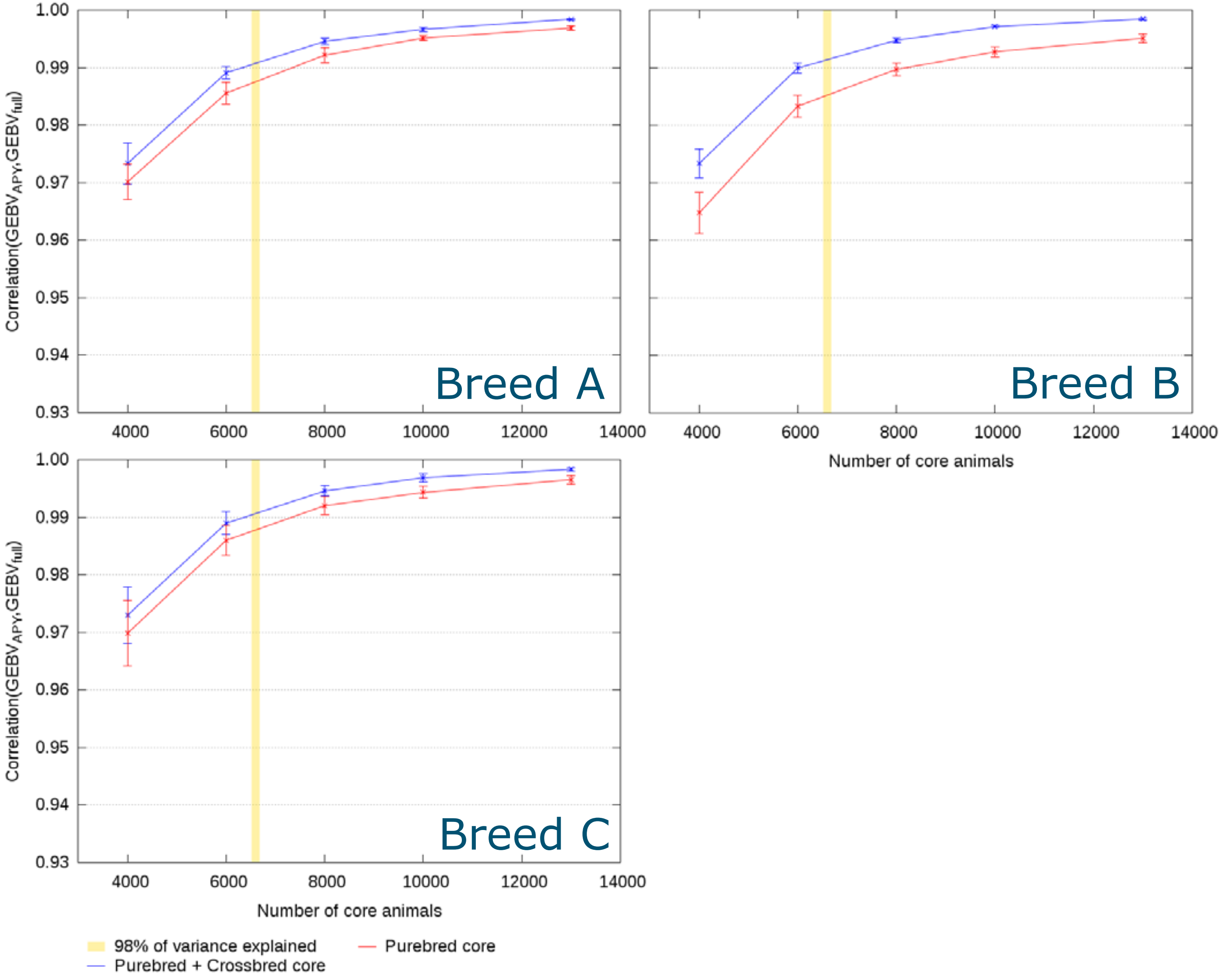
Validation – unrelated

Purebred animals (10th generation) without phenotype



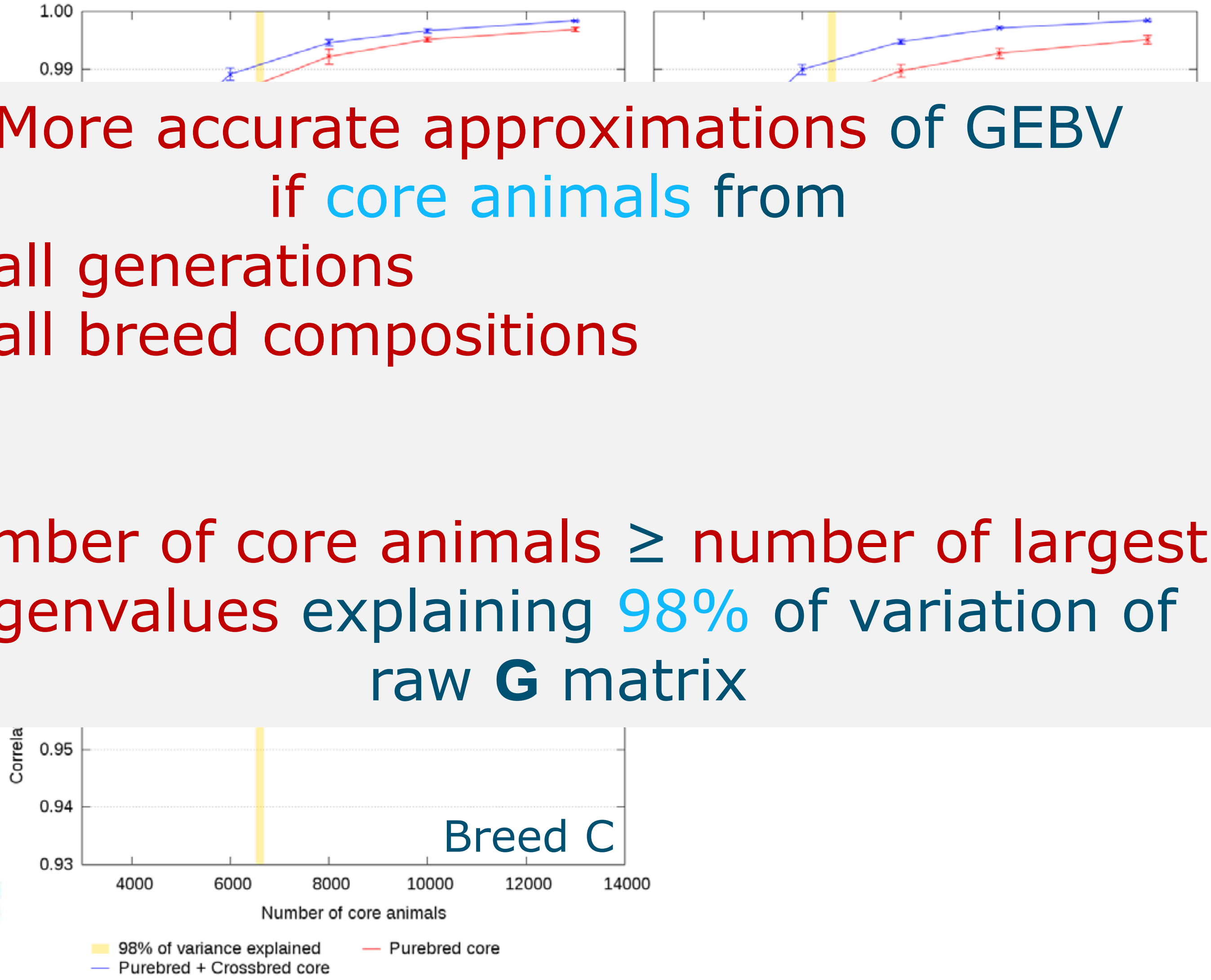
Validation – unrelated

Purebred animals (10th generation) without phenotype



Validation – unrelated

Purebred animals (10th generation) without phenotype



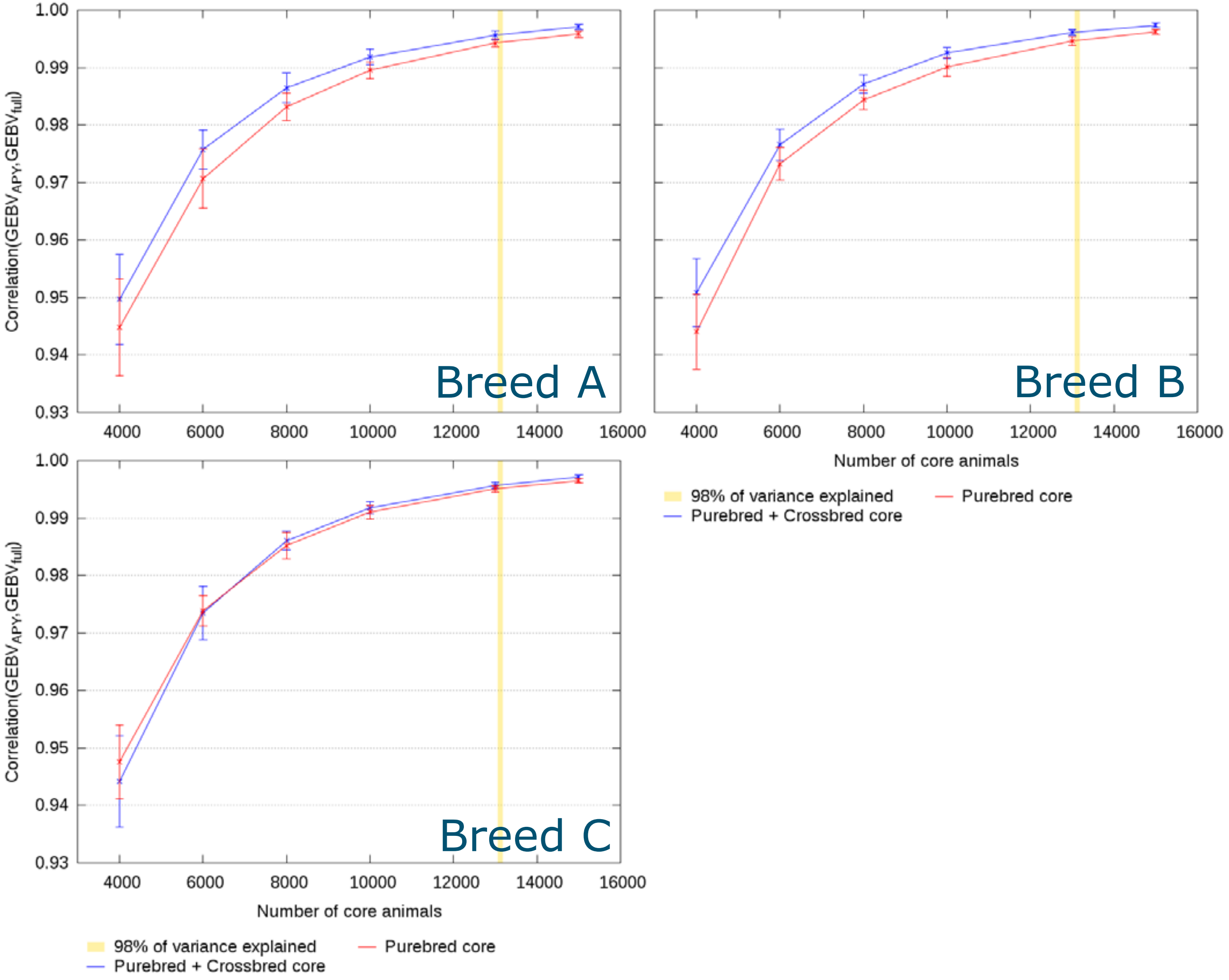
More accurate approximations of GEBV
if core animals from

- all generations
- all breed compositions

Number of core animals ≥ number of largest eigenvalues explaining 98% of variation of raw **G** matrix

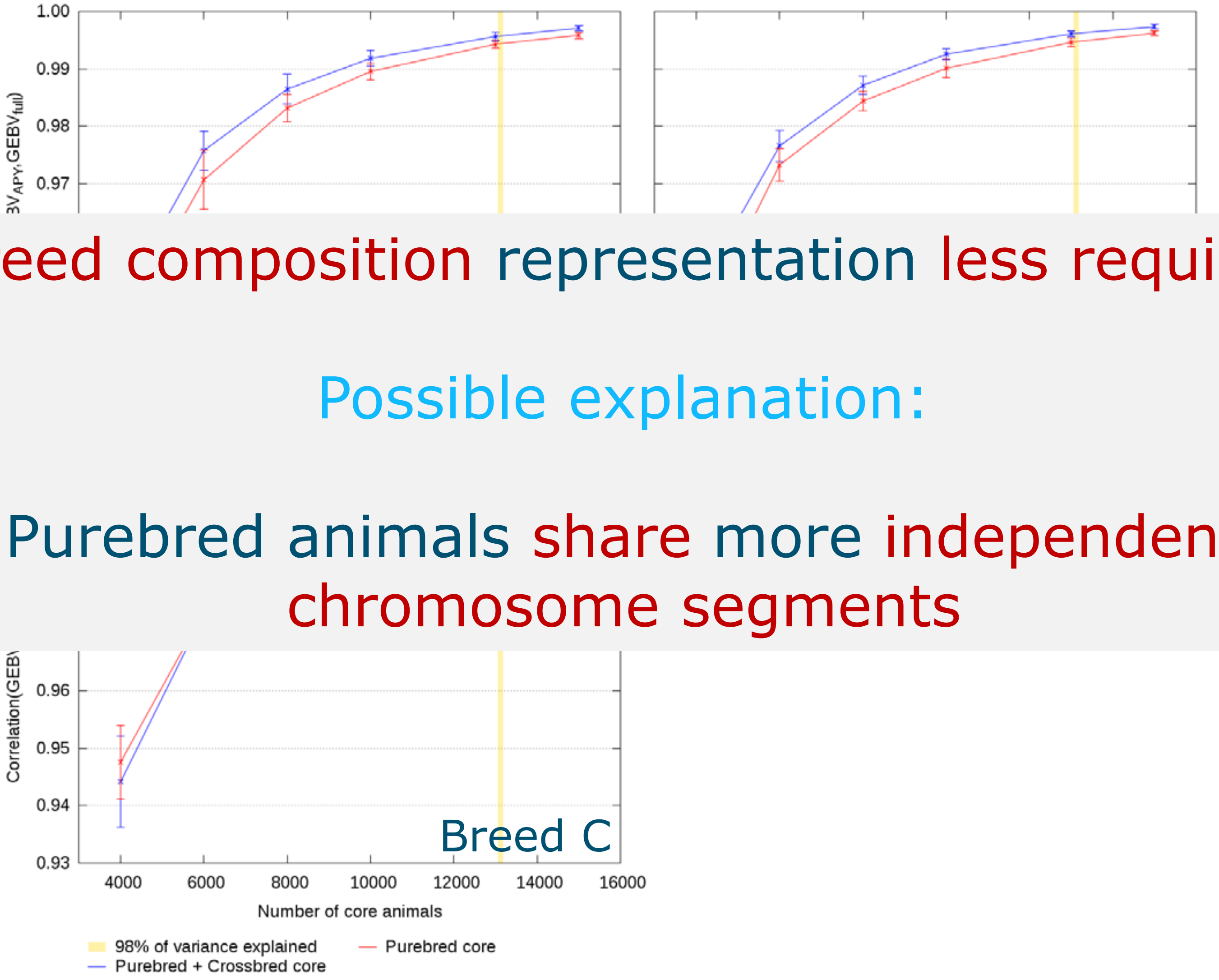
Validation – related

Purebred animals (10th generation) without phenotype



Validation – related

Purebred animals (10th generation) without phenotype



Breed composition representation less required

Possible explanation:

Purebred animals share more independent chromosome segments

Conclusions

- APY is applicable in (simulated) crossbreeding schemes
- Number of core animals
  - ≥ Number of largest eigenvalues explaining 98% of variation of raw **G** matrix
- Core animals must represent
  - all generations (Ostersen et al., 2016)
  - all breed compositions

Thank you!

BREED4FOOD