



Improving computing performance of genomic evaluations by genotype and phenotype truncation

Fernando Bussiman, C. Chen, J. Holl, A. Legarra, I. Misztal, and D. Lourenco



Animal Breeding and Genetics Group

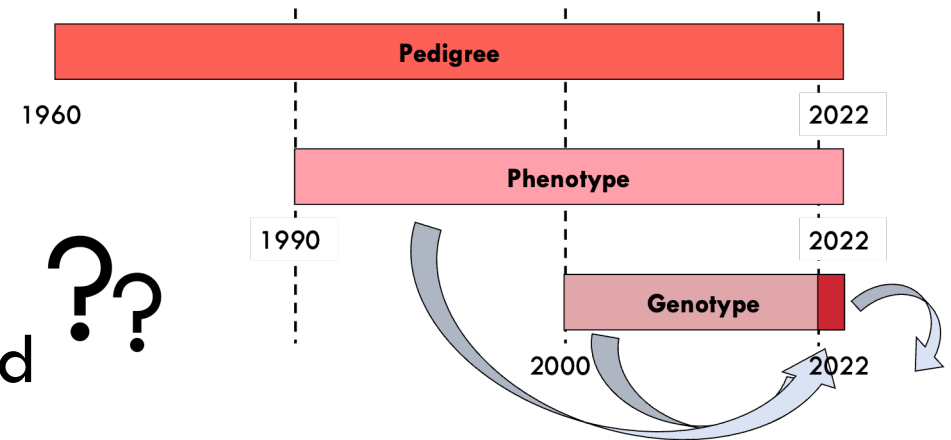
College of Agricultural & Environmental Sciences

UNIVERSITY OF GEORGIA



Introduction

- Large data collection: past x current
- More data, more computing power needed
- Same model to historical and current data may be wrong



- Objective:
 - Investigate data truncation to reduce genomic prediction computing costs
 - young animals
 - as accurate as non-truncated

Truncation Scenarios

- Data truncation is usually applied to phenotypes and pedigree
 - should we use all the genotyped animals?
 - *noInfo*: genotyped animals without own and progeny phenotypes were removed
 - *Age*: old genotyped animals were removed
 - *noInfo* + *Age*: combination of *noInfo* and *Age* scenarios
 - *Allgen*: no genotyped animal was removed
 - Phenotypes were removed every three years
 - Three pedigree depths: 2 and 3 generations back or the entire pedigree



Truncation Scheme



Plus, up to 3 generations of pedigree



Plus, up to 3 generations of pedigree



Plus, up to 3 generations of pedigree



Datasets Provided

- 2 pig lines:

- Terminal Sire – Growth (ADG and BF) and Mortality (WFM)

- > 300K records
 - > 160K genotyped
 - 3.6M pedigree

- Maternal – Reproductive (LS and NS) and Pre-weaning (BW and PWM)

- > 500K records
 - > 170K genotyped
 - 11M pedigree



Validation

- Males and females born in 2019 and subsequent years
 - Phenotyped progeny
- LR
 - Reliability

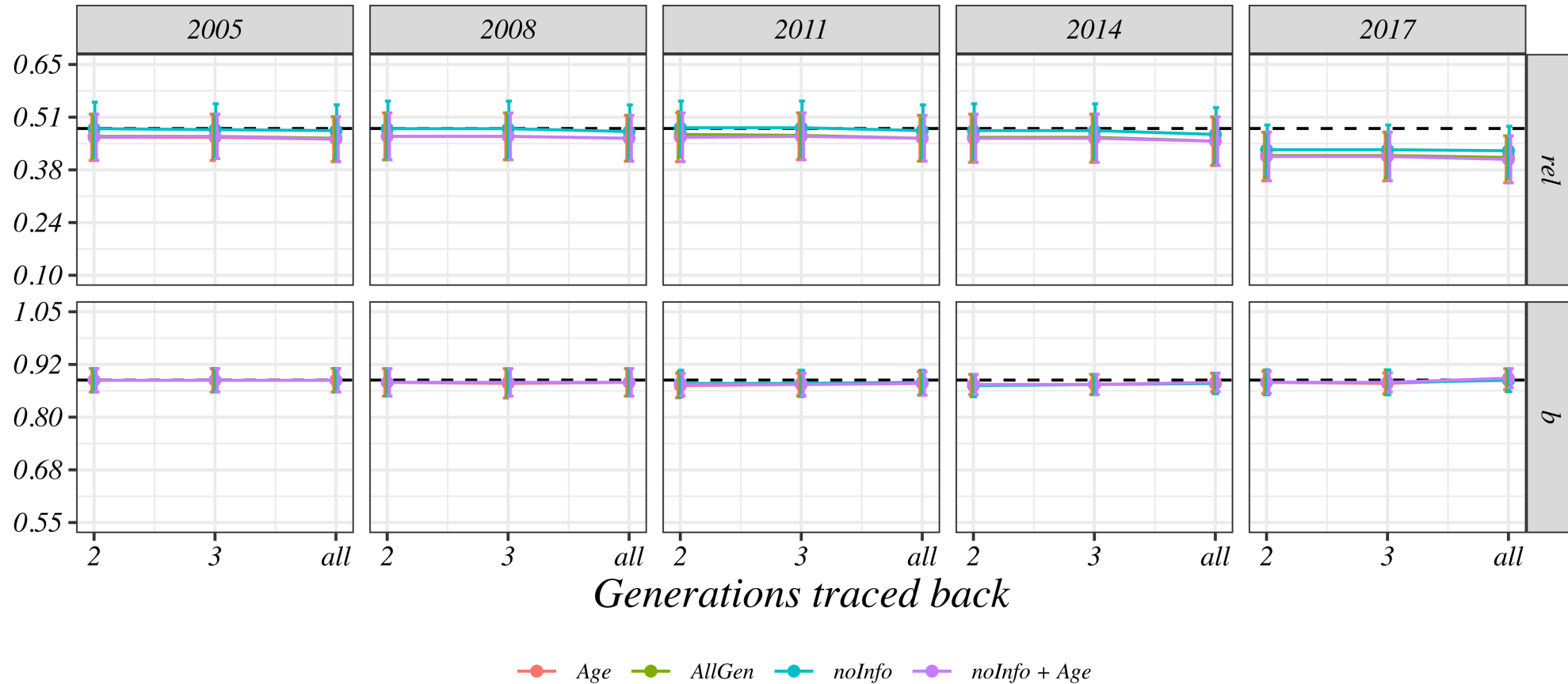
 - Dispersion bias

$$\text{rel} = \frac{\text{cov}(\hat{\mathbf{u}}_w, \hat{\mathbf{u}}_p)}{\sigma_u^{2*}}$$

$$b = \frac{\text{cov}(\hat{\mathbf{u}}_w, \hat{\mathbf{u}}_p)}{\text{var}(\hat{\mathbf{u}}_p)}$$

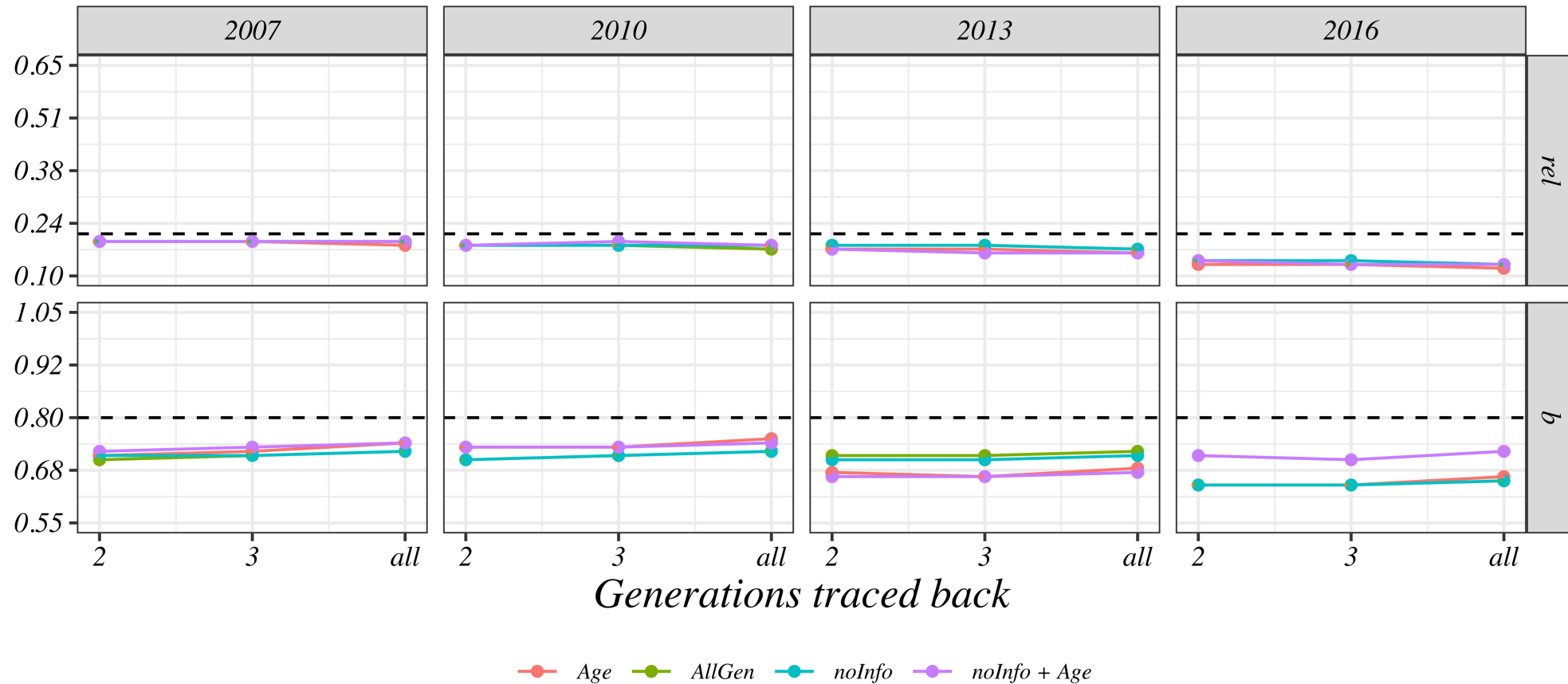
Reliability and Dispersion

Growth traits - Terminal Sire line



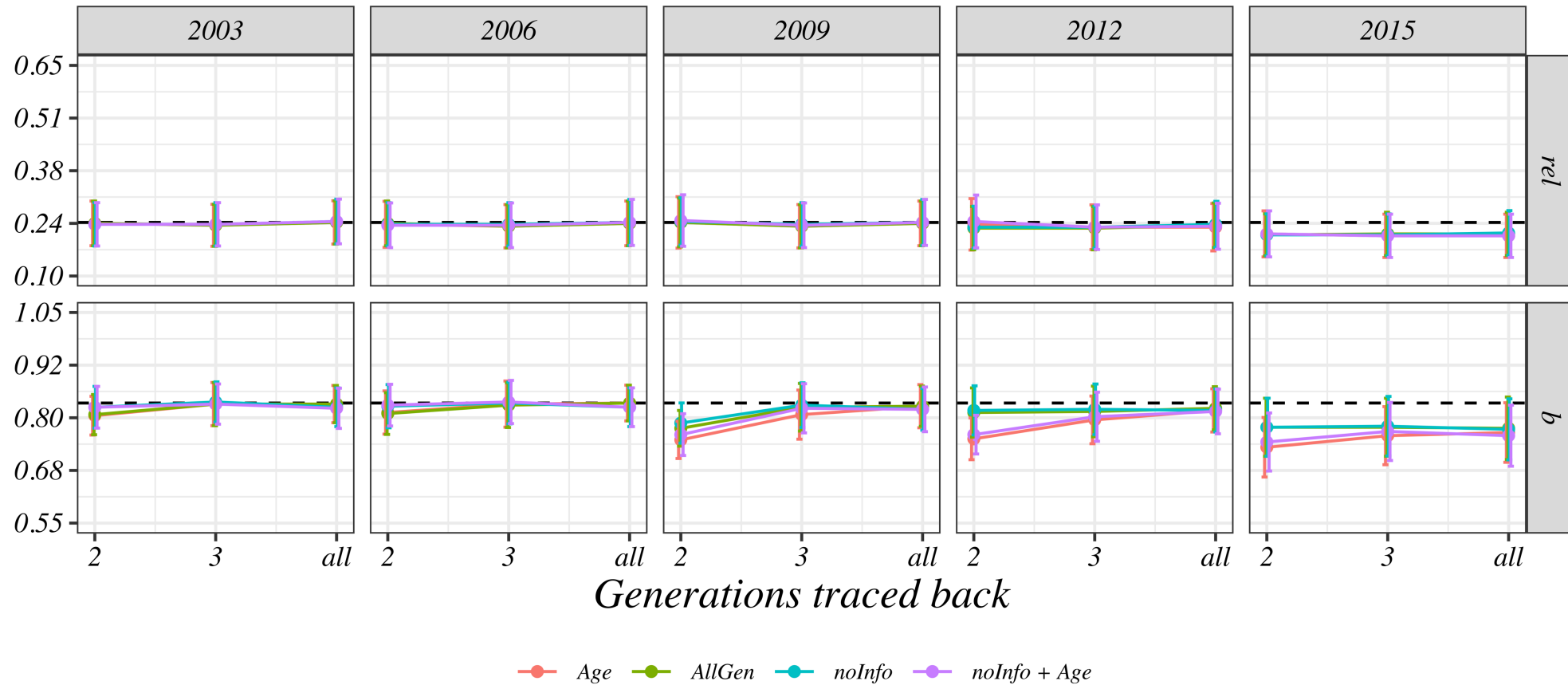
Reliability and Dispersion

Mortality - Terminal Sire line



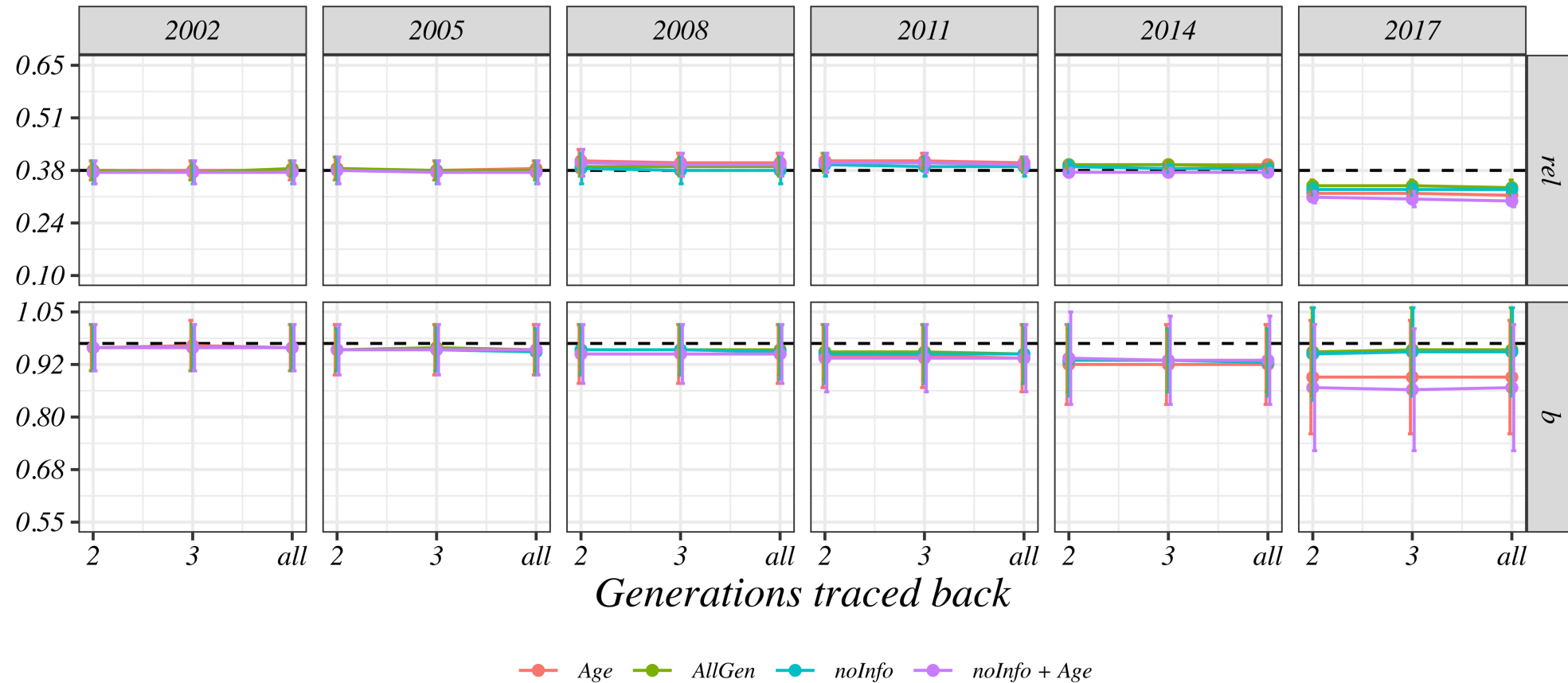
Reliability and Dispersion

Preweaning traits - Maternal line



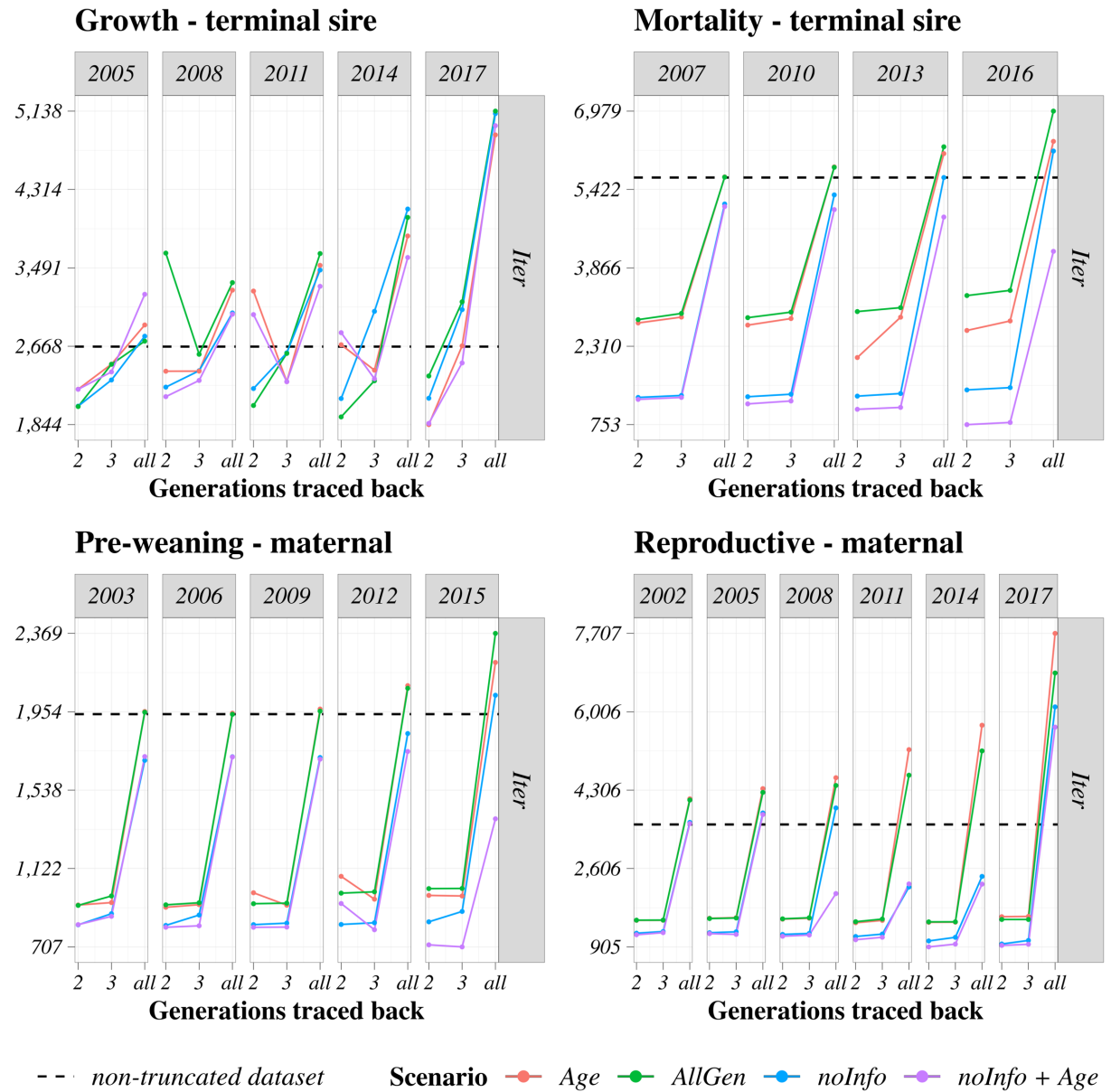
Reliability and Dispersion

Reproductive traits - Maternal line



Number of Iterations

- A positive “side effect”
- 90% gain in efficiency!
- Not compromising accuracy



Conclusions

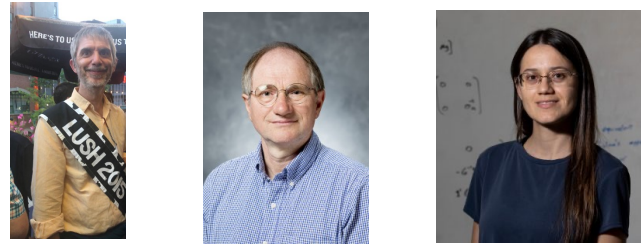
- Tracing up to three generations in the pedigree is enough
- Data truncation can cause a slight drop in accuracy
 - If genotyped animals have no phenotypes
- Keeping all genotyped animals for truncated datasets is not ideal
- Removing unneeded information increases computing efficiency
 - without compromising genomic predictions



Bussiman et al., 2023



Acknowledgements



Animal Breeding and Genetics Group
College of Agricultural & Environmental Sciences
UNIVERSITY OF GEORGIA

