Use of causative variants and SNP weighting in a single-step GBLUP context

Breno Fragomeni[1], Lourenco DAL[1], Legarra A[2], Tooker ME [3], VanRaden PM[3],Lawlor, T.[4], Misztal I[1]

[1]University of Georgia, Athens, USA

[2]INRA, Castanet-Tolosan, France

[3]AGIL ARS-USDA, Beltsville, USA

[4]American Holstein Association, Brattleboro, USA

7 – 11 February 2018
THE GLOBAL STANDARD FOR LIVESTOCK DATA
ICAR
INTERBULL
Annual Conference
icar2018.nz

11-16 February 2018
11th WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION
wcgalp.com

A O T E A   C E N T R E ,   A U C K L A N D ,   N E W   Z E A L A N D

Motivation

- Decreasing costs of whole genome sequence
- Revived interest in causative variants for prediction
- Several authors are finding and using causative variants
  - No improvement :
    - Binsbergen et al., 2015 and Erbe et al., 2016
  - Up to 5% improvement:
    - Brondum et al. 2015 and Vanraden et al., 2017

Motivation

GSE Genetics Selection Evolution

RESEARCH ARTICLE   Open Access

CrossMark

Incorporation of causative quantitative trait nucleotides in single-step GBLUP

Breno O. Fragomeni[1*], Daniela A. L. Lourenco[1], Yukata Masuda[1], Andres Legarra[2] and Ignacy Misztal[1]

- ssGBLUP was able to reach accuracies close to 1 in simulation
  - Simulated QTN position and effects known
  - GWA estimated weights had limited impact
- GWA Methodology – no limitation in minimum and maximum weights (Zhang et al., 2016)

Objective

- Test different SNP weighting methods in GBLUP and ssGBLUP in field data with the inclusion of causative variants

Field Data

- 4M Records for Stature
- 3M Cows
- 4.6M Animals in pedigree
- h^2=0.44
- 27k Genotyped Sires
  - 54k SNP
  - 54k SNP + 17k Causative Variants (VanRaden et al., 2017)

Analysis

- GBLUP
  - Multi-step approach
  - Daughter deviation as phenotypes
  - Genomic Relationship Matrix
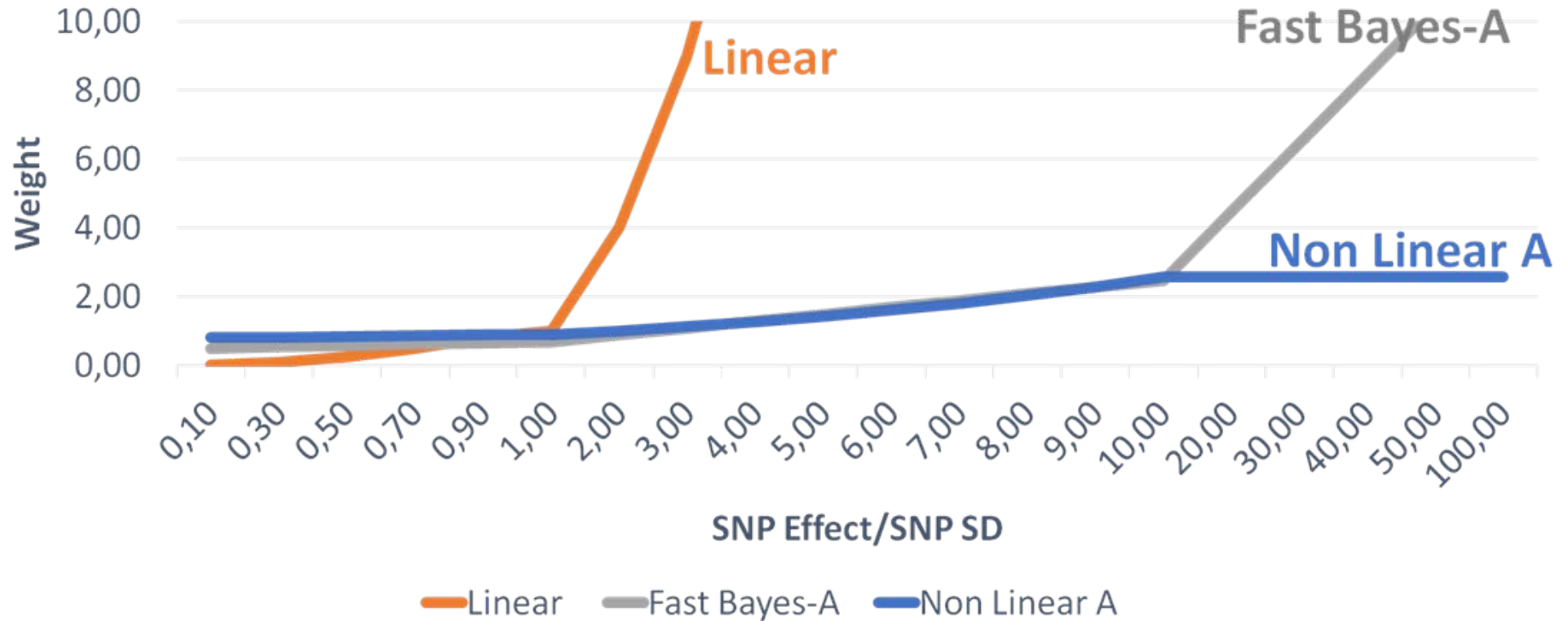 - Homogeneous or heterogeneous residual variance – different reliabilities

- ssGBLUP
  - Same model as national evaluation for type traits
  - No deregressions
  - Matrix combining pedigree and genomic information (**H**)

Weighted genomic relationship matrix

$$\mathbf{G} = \mathbf{ZDZ'}\frac{\sigma_s^2}{\sigma_a^2} = \frac{\mathbf{ZDZ'}}{\sum_i 2p_i q_i}$$
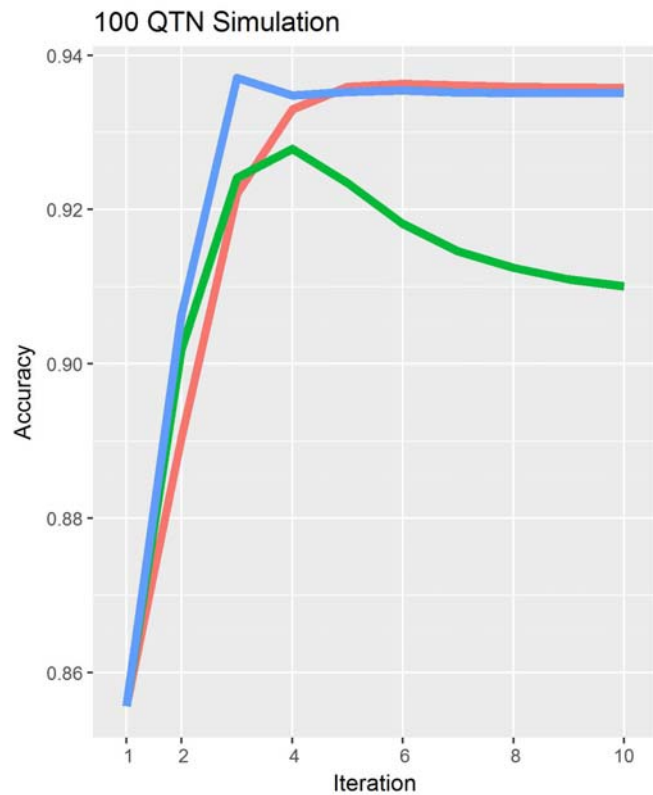
- Default
  - $\hat{d}_i = 1$
    - (VanRaden et al., 2008)

- Linear weights
  - $\hat{d}_i \sim \sigma_{SNP_i}^2$
    - (Zhang et al., 2010)

- Non-linear A weights
  - $\hat{d}_i = 1.125^{\frac{|\widehat{u_i}|}{sd(u)}-2}$
    - (VanRaden et al., 2008)
    - Value capped at 10

- Fast-Bayes A
  - $\hat{d}_i = \frac{\widehat{u}_i^2 + df*S^2}{df+1}$
    - (Sun et al., 2012)

Weight matrix elements

Simulation results
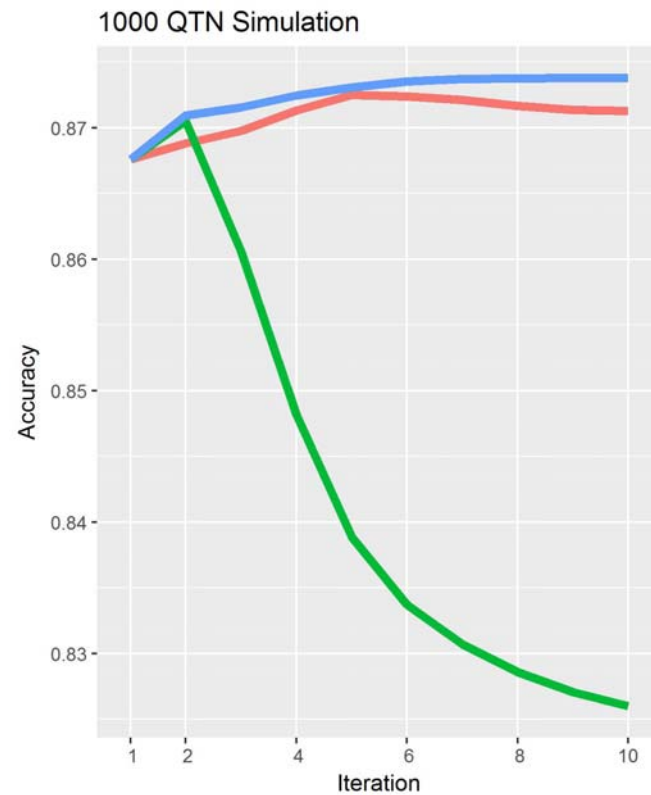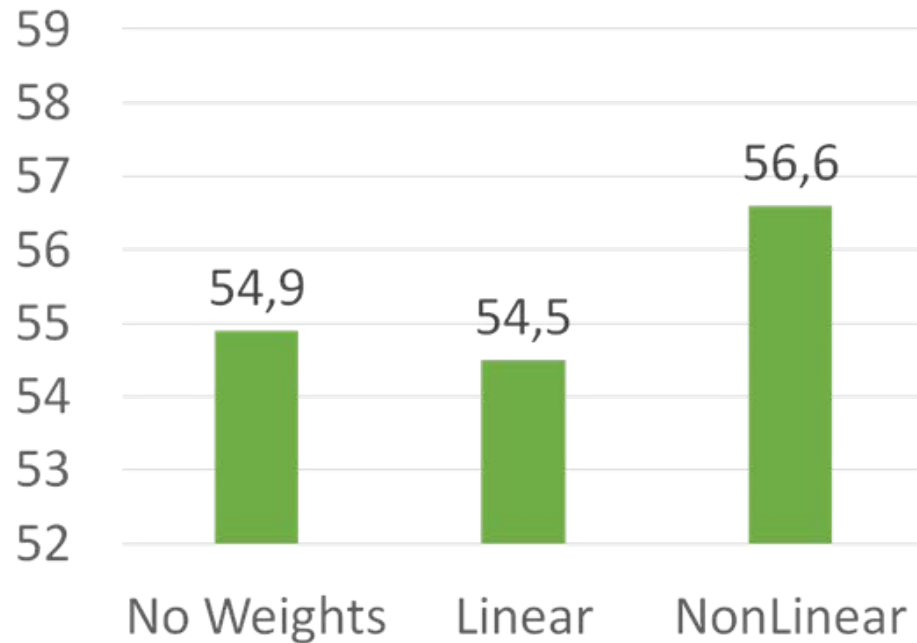
GBLUP – 54K SNP - Reliabilities

HOMOGENEOUS RESIDUAL VARIANCE

No Weights	54,9
Linear	54,5
NonLinear	56,6

HETEROGENEOUS RESIDUAL VARIANCE

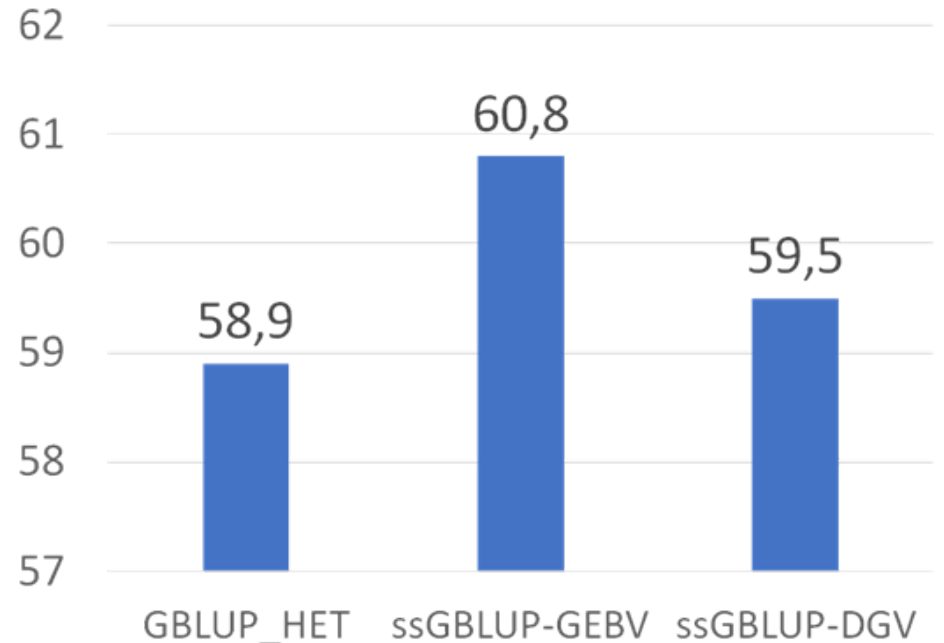|	
No Weights	58,7
Linear	56,3
NonLinear	58,6
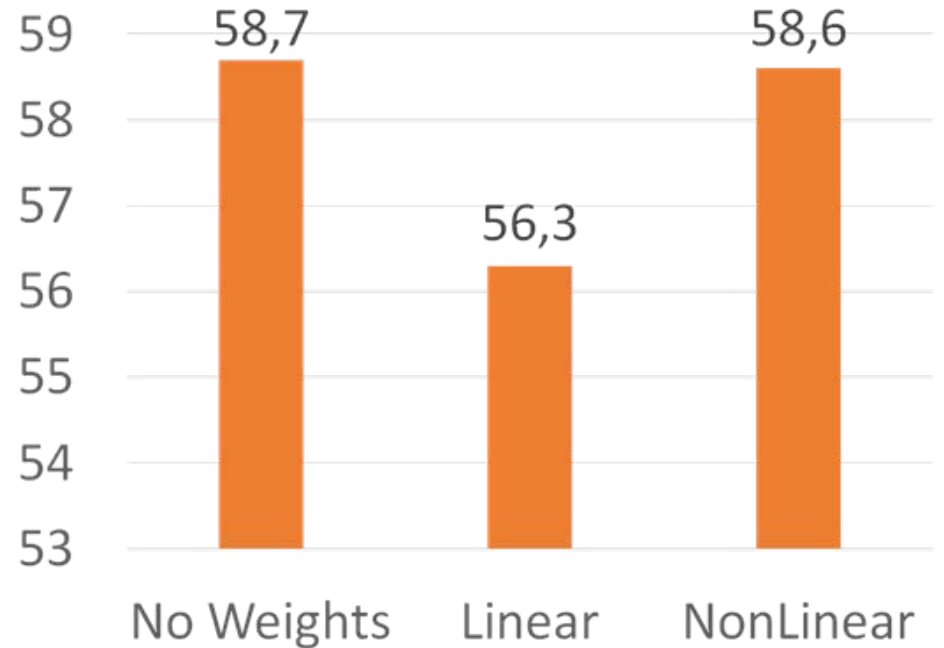
GBLUP and ssGBLUP – 54K SNP - Reliabilities

Unweighted GRM
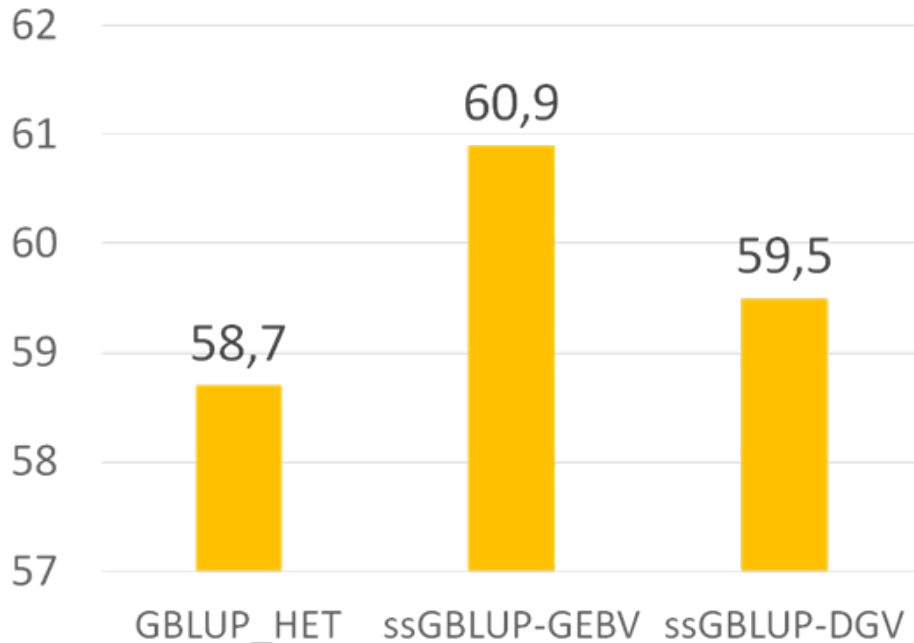
Weighted GRM

$$GEBV = w_1 PA + w_2 DGV$$

Including causative variants

HETEROGENEOUS RESIDUALS

	GBLUP - Homogeneous	GBLUP - No Weights	GBLUP - Linear	GBLUP - NonLinear	ssGBLUP - GEBV
54k	54,9	58,2	56,4	58,4	60,9
70k	55,5	58,7	56,3	58,9	60,8

ssGBLUP - DGV
59,5

■ 54k   ■ 70k   ──
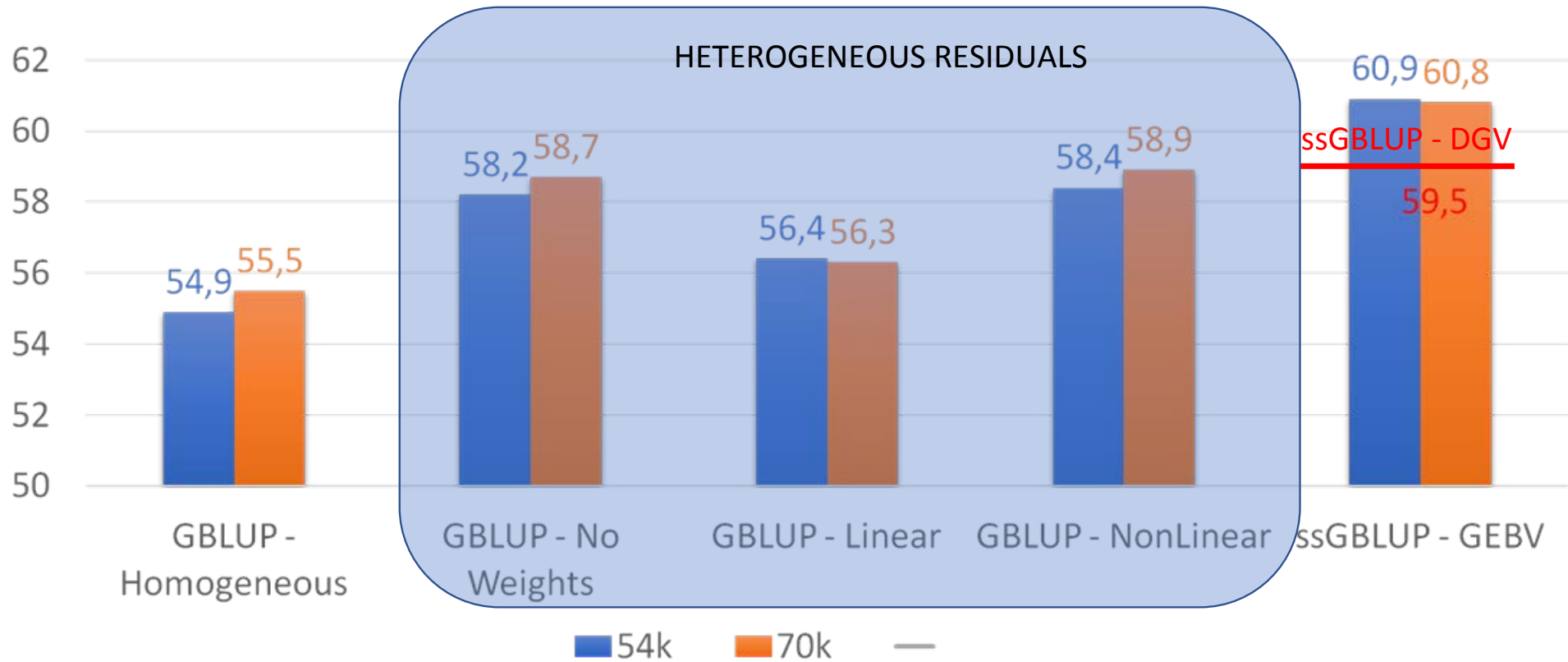
Inflation coefficient: b_1
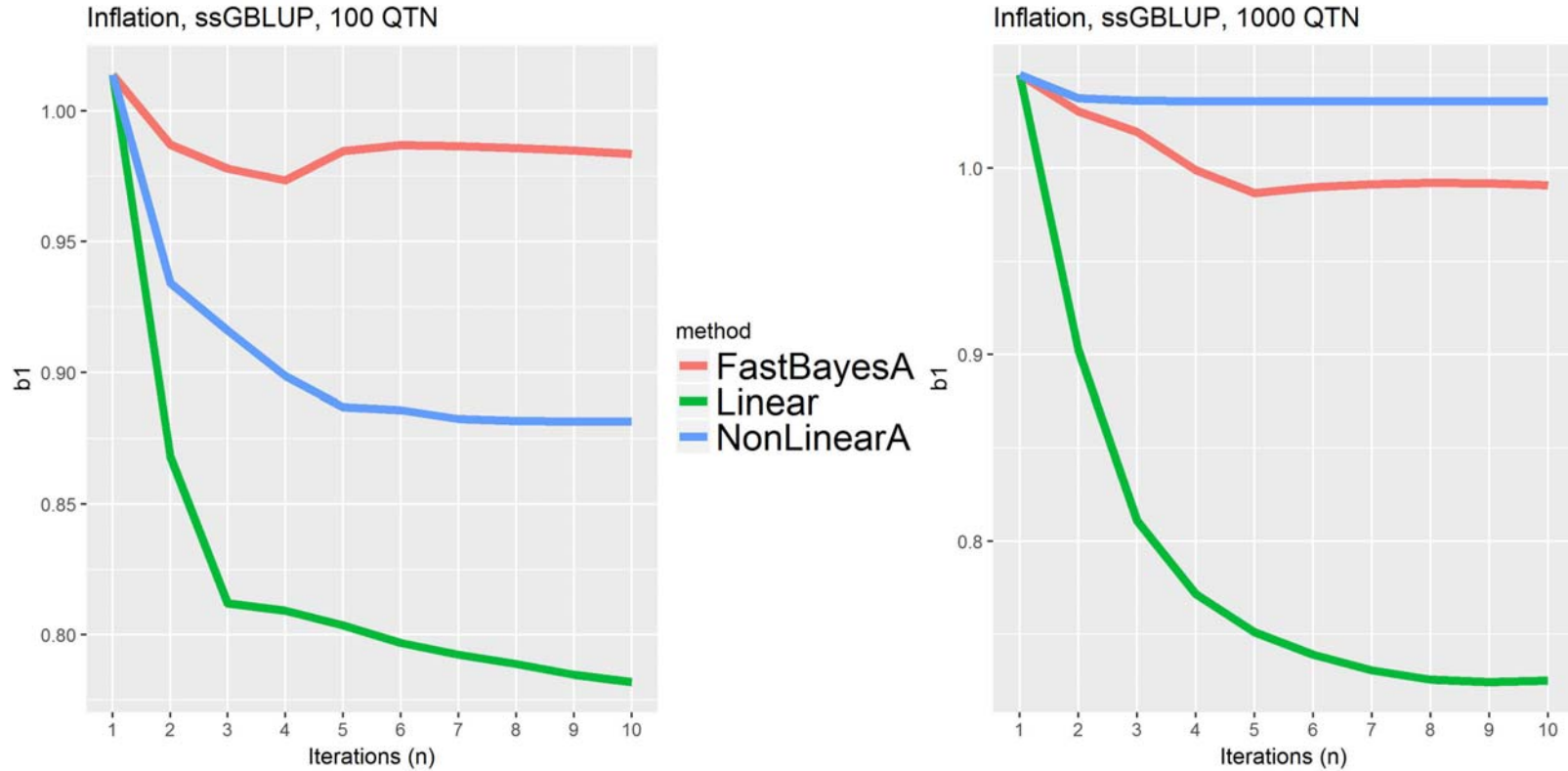
Conclusions

- Gains with causative variants have more impact in GBLUP than in ssGBLUP
  - More data is used in single-step methodology, therefore impact of prior is less important
  - Sequence data might mask or fix methodology problems
- Non-linear methodology is better for weighting marker effects than linear weights
- Estimating weights in single-step GBLUP is still a research topic