

ACCF90GS2: software for fast approximation of reliabilities of estimated breeding values in single-step GBLUP

Matias Bermann, Daniela Lourenco, Ignacy Misztal

WCGALP - July 2022



Motivation

- Large-scale genetic evaluations can use the Algorithm for Proven and Young (APY)
- Current methods for estimating reliabilities for ssGBLUP are not applicable with APY
- **Objectives:**
 - Develop an efficient algorithm for calculating genomic reliabilities for APY-GBLUP models
 - Use the previous algorithm for estimating reliabilities in single and multiple-traits ssGBLUP models
 - Develop an efficient software for using those methods

Genomic reliabilities

- GBLUP with APY:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \alpha \mathbf{G}_{APY}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

$$\mathbf{G}_{APY}^{-1} = \begin{bmatrix} \mathbf{M}_{nn}^{-1} & \mathbf{G}^{nc} \\ \mathbf{G}^{cn} & \mathbf{G}^{cc} \end{bmatrix} = \begin{array}{c} \diagdown \\ \hline \hline \end{array}$$

$$diag(\mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} + \mathbf{G}_{APY}^{-1})^{-1} \approx diag(\mathbf{W} + \mathbf{G}_{APY}^{-1})^{-1}$$

Genomic reliabilities (cont.)

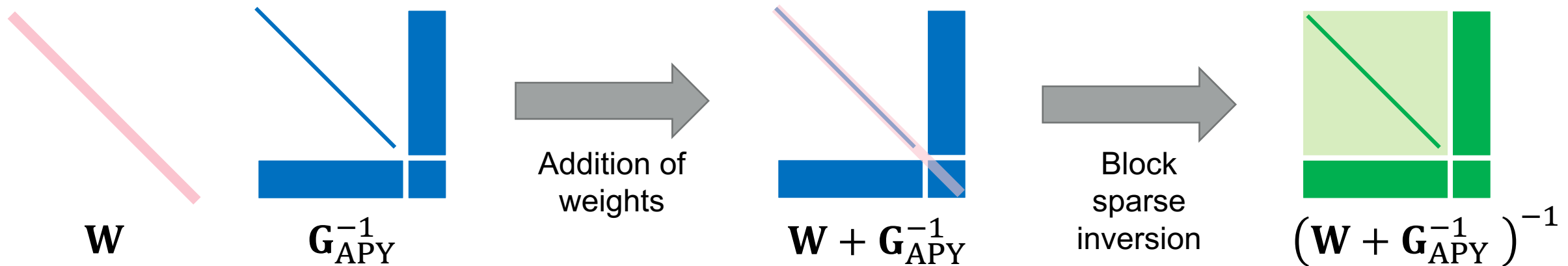
- Core animals:

$$\text{diag}((\mathbf{W}_{cc} + \mathbf{G}^{cc} - \mathbf{G}^{cn}(\mathbf{W}_{nn} + \mathbf{M}_{nn}^{-1})^{-1}\mathbf{G}^{nc})^{-1})$$

- Non-core animals:

$$\text{diag}((\mathbf{W}_{nn} + \mathbf{M}_{nn}^{-1})^{-1} + (\mathbf{W}_{nn} + \mathbf{M}_{nn}^{-1})^{-1}\mathbf{G}^{nc}(\mathbf{W}_{cc} + \mathbf{G}^{cc} - \mathbf{G}^{cn}(\mathbf{W}_{nn} + \mathbf{M}_{nn}^{-1})^{-1}\mathbf{G}^{nc})^{-1}\mathbf{G}^{cn}(\mathbf{W}_{nn} + \mathbf{M}_{nn}^{-1})^{-1})$$

- $\max(O(n_c^3), O(n_n n_c^2))$



ACCF90GS2

Approximation of Prediction Error Variance in Large-Scale Animal Models

I. MISZTAL and G. R. WIGGANS¹
Department of Animal Sciences
University of Illinois
Urbana 61801

Approximating Genomic Reliabilities for National Genomic Evaluation

*Z. Liu¹, P. M. VanRaden², M.H. Lidauer³, M. P. Calus⁴, H. Benhajali⁵, H. Jorjani⁵
and V. Ducrocq⁶*

Approximate Reliability of Genetic Evaluations Under an Animal Model

BEVIN HARRIS¹ and DAVID JOHNSON
Livestock Improvement, Private Bag 3016, Hamilton, New Zealand

Approximation of reliabilities for multiple-trait model with maternal effects

T. Strabel¹, I. Misztal², and J. K. Bertrand

Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models

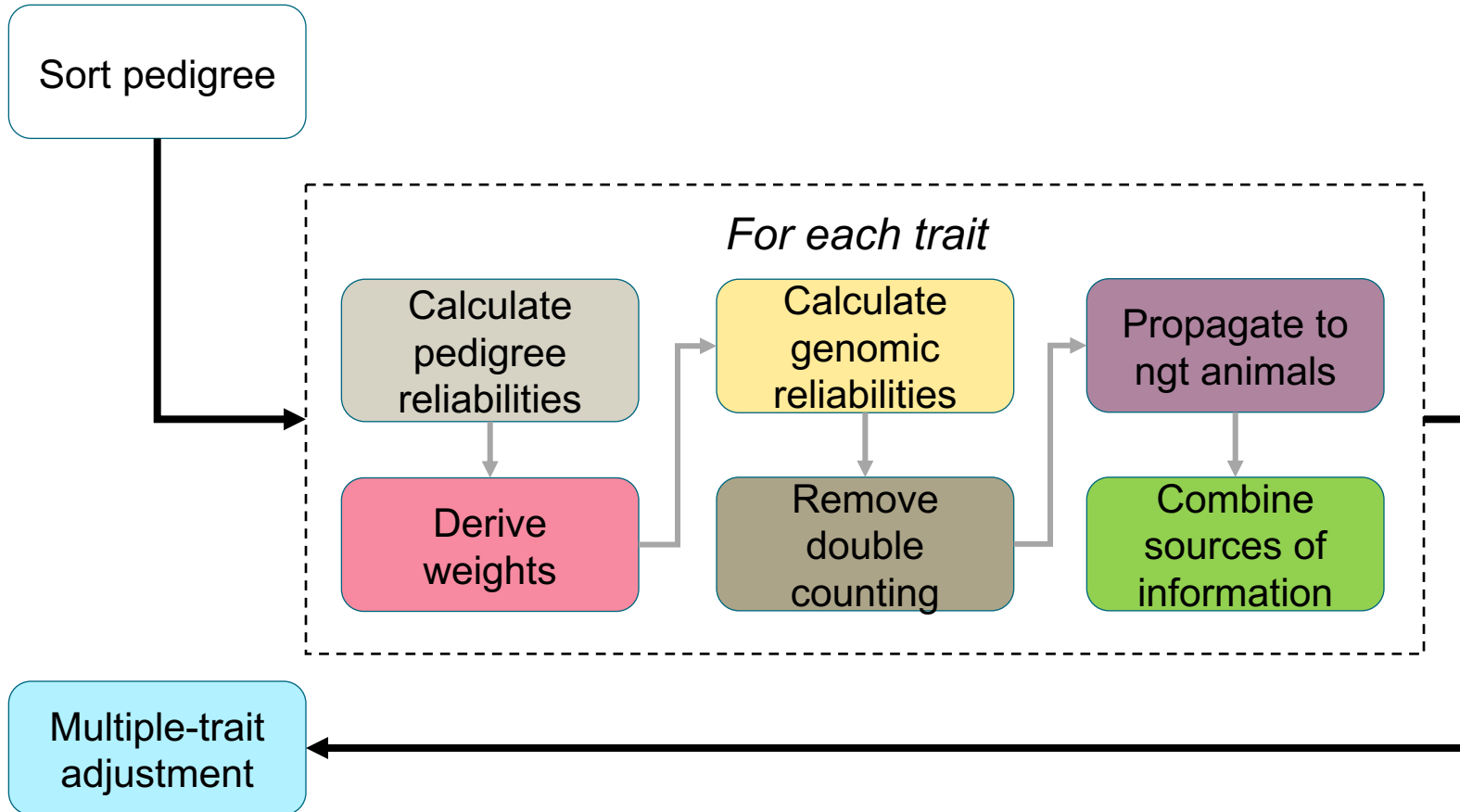
BY B. TIER and K. MEYER

Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the Algorithm for Proven and Young

Matias Bermann,¹  Daniela Lourenco, and Ignacy Misztal 



ACCF90GS2 flowchart



- Maximum number of generation
- Recursive quicksort
- Harris & Johnson (1998)
- Supplied by user
- Back-solve pedigree reliabilities with a root-finding technique
- Block sparse inversion
- Back-solve pedigree reliabilities and set to zero contributions from ngt animals
- Back-solve genomic reliabilities and use them as weights
- Effective daughter/records contribution
- Strabel & Misztal (2001)

Data sets

	Data1	Data2
Trait(s)		
Animals in the pedigree		
Animals with records		
Genotyped animals	Core	
	Non-core	

Data sets

	Data1	Data2
Trait(s)	BW – WW –PWG	
Animals in the pedigree	172 k	
Animals with records	78 k	
Genotyped animals	Core Non-core	10 k 10 k

Data sets


Trait(s)	Data1	Data2
	BW – WW –PWG	BW – WW – PWG
Animals in the pedigree	172 k	10.2 m
Animals with records	78 k	8.6 m
Genotyped animals	Core 10 k	10 k
	Non-core 10 k	324 k

Data1 - results

Trait	Effect	Correlation	Intercept	Slope	MAC
BW	Direct	0.98	0.04	0.93	0.01
	Maternal	0.94	0.03	0.92	0.01
WW	Direct	0.98	0.02	0.94	0.01
	Maternal	0.95	0.03	0.93	0.01
PWG	Direct	0.98	0.02	0.96	0.01

Computing times (min)

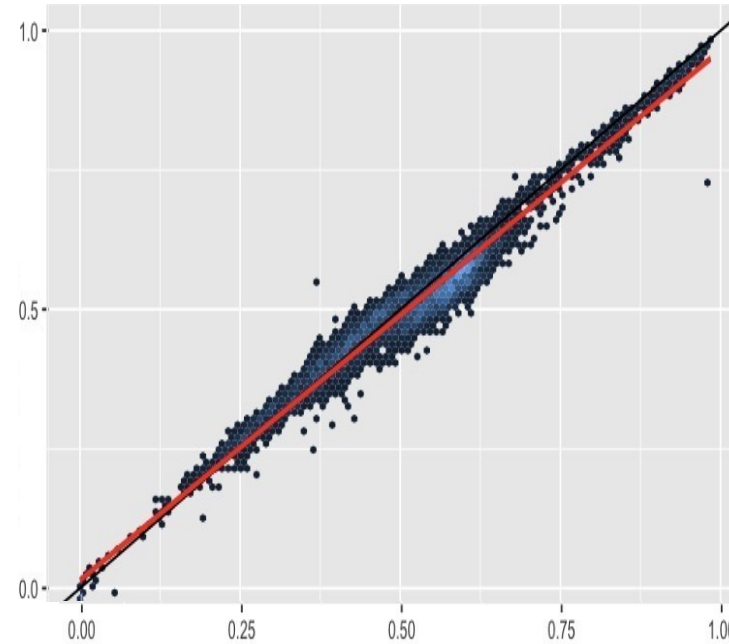
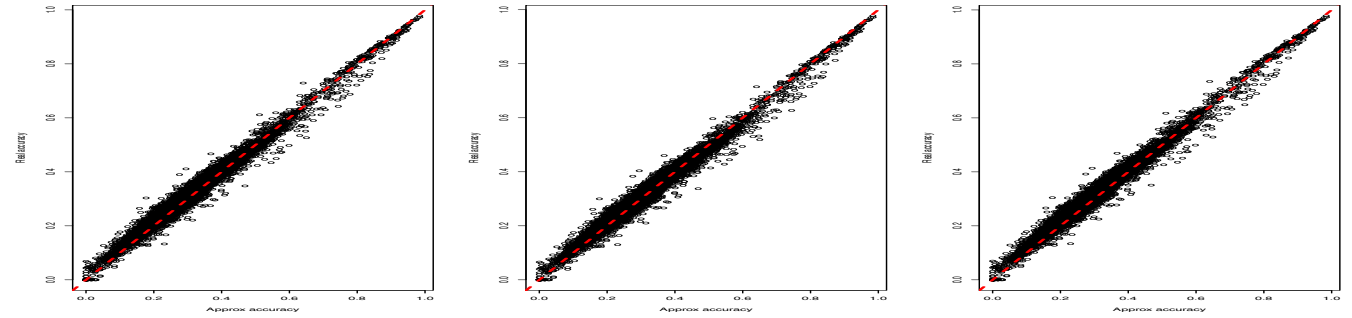
	Data1	Data2
Genomic reliabilities	0.78	12.56
Other steps	0.12	8.24
Total time	0.9	20.8

- 
- X60 data size
 - X30 computing time
 - X16 genotypes
 - X22 computing time

1.5 TB of memory - Processor Intel Xeon Gold 6258R - 32 threads

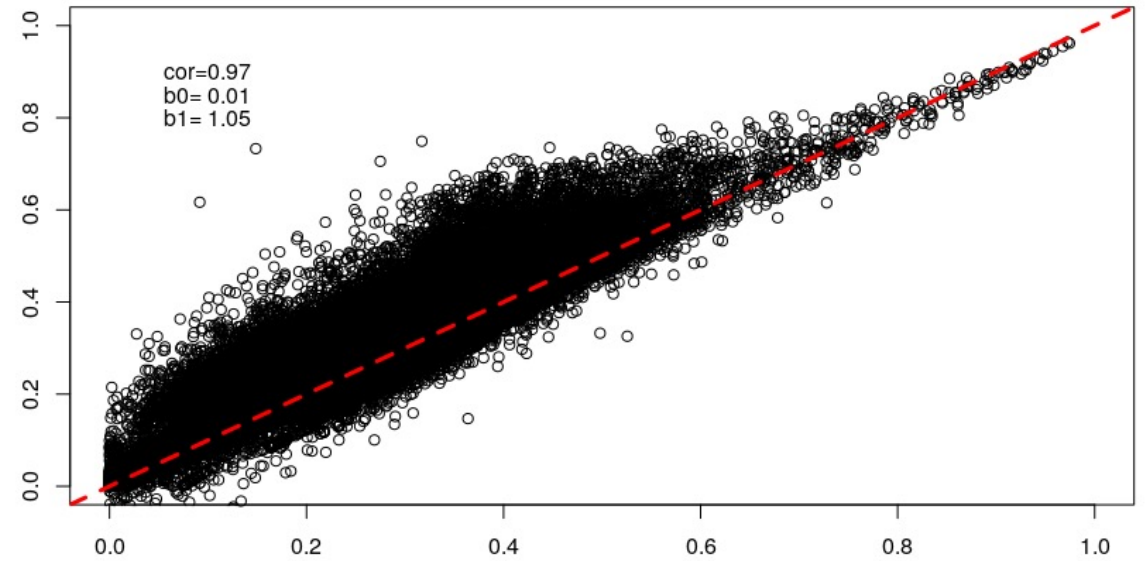
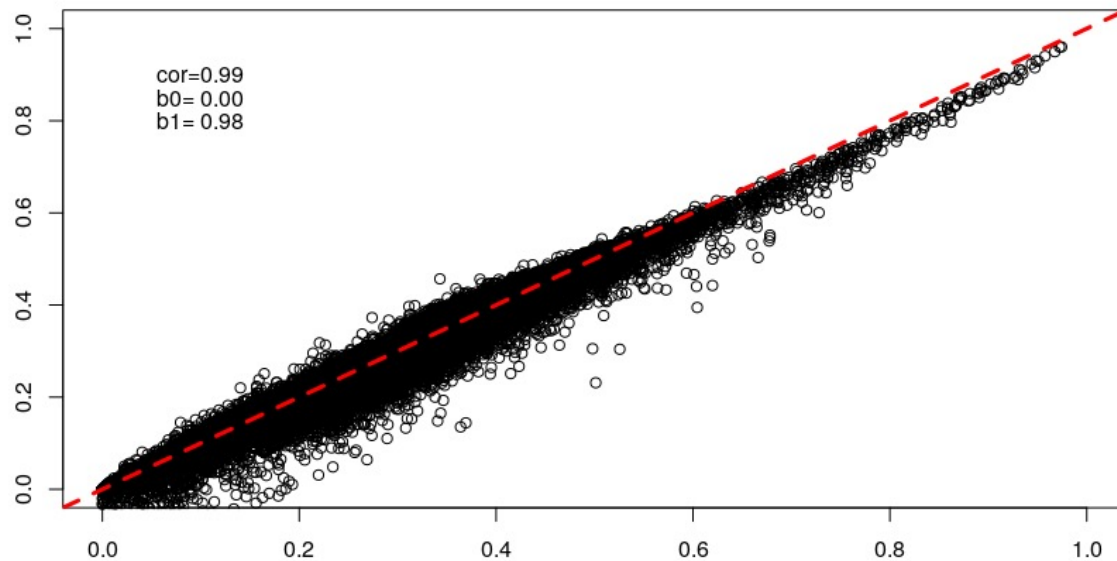
Other implementations

- US Ayrshire
 - 94,000 in the pedigree
 - 9,000 genotyped animals
- US Holstein
 - 60 millions in the pedigree
 - 3.1 genotyped animals
 - 4 hours
- Uruguayan Holstein
 - Test-day model



Challenges

- Complex models
 - Categorical data?
 - Random regression
- Propagation to non-genotyped animals



Conclusion

- ACCF90GS2 gives a good approximation of reliabilities for ssGBLUP
- Great flexibility for different models
- Computing costs are less than or equal to estimation of breeding values

Questions?

Matias Bermann – mbermann@uga.edu



Genomic reliabilities (cont.)

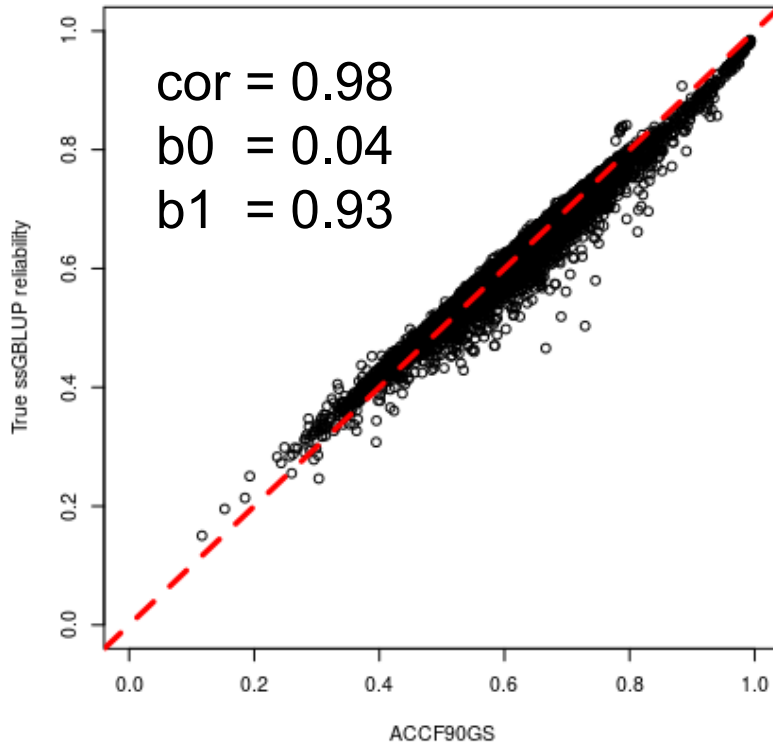
$$\text{diag}(\mathbf{W} + \mathbf{G}_{\text{APY}}^{-1})^{-1} = \frac{\text{diag}((\mathbf{W}_{\text{nn}} + \mathbf{M}_{\text{nn}}^{-1})^{-1} + (\mathbf{W}_{\text{nn}} + \mathbf{M}_{\text{nn}}^{-1})^{-1} \mathbf{G}^{\text{nc}} (\mathbf{W}_{\text{cc}} + \mathbf{G}^{\text{cc}} - \mathbf{G}^{\text{cn}} (\mathbf{W}_{\text{nn}} + \mathbf{M}_{\text{nn}}^{-1})^{-1} \mathbf{G}^{\text{nc}})^{-1} \mathbf{G}^{\text{cn}} (\mathbf{W}_{\text{nn}} + \mathbf{M}_{\text{nn}}^{-1})^{-1})}{\text{diag}((\mathbf{W}_{\text{cc}} + \mathbf{G}^{\text{cc}} - \mathbf{G}^{\text{cn}} (\mathbf{W}_{\text{nn}} + \mathbf{M}_{\text{nn}}^{-1})^{-1} \mathbf{G}^{\text{nc}})^{-1})}$$

1. Read $\mathbf{G}_{\text{APY}}^{-1}$ and store in memory
2. Approximate \mathbf{W} and overwrite $\mathbf{G}_{\text{APY}}^{-1}$ as $\mathbf{G}_{\text{APY}}^{-1} = \mathbf{W} + \mathbf{G}_{\text{APY}}^{-1} \alpha$
3. Compute $\mathbf{G}^{\text{cn}*} = \mathbf{G}^{\text{cn}} \mathbf{M}_{\text{nn}}$
4. Overwrite \mathbf{G}^{cc} as $\mathbf{G}^{\text{cc}} = \mathbf{G}^{\text{cc}} - \mathbf{G}^{\text{cn}} \mathbf{G}^{\text{nc}*}$
5. Invert \mathbf{G}^{cc}
6. Overwrite \mathbf{G}^{cn} as $\mathbf{G}^{\text{cn}} = -\mathbf{G}^{\text{cc}} \mathbf{G}^{\text{cn}*}$
7. Overwrite $\mathbf{M}_{\text{nn}}^{-1}$ as $\mathbf{M}_{\text{nn}_i}^{-1} = \mathbf{M}_{\text{nn}_i} - (\mathbf{G}_{.i}^{\text{cn}*})' \mathbf{G}_{.i}^{\text{cn}}$
8. Obtain PEV from $\text{diag}(\mathbf{G}^{\text{cc}})$ and $\mathbf{M}_{\text{nn}}^{-1}$ for core and non-core animals, respectively

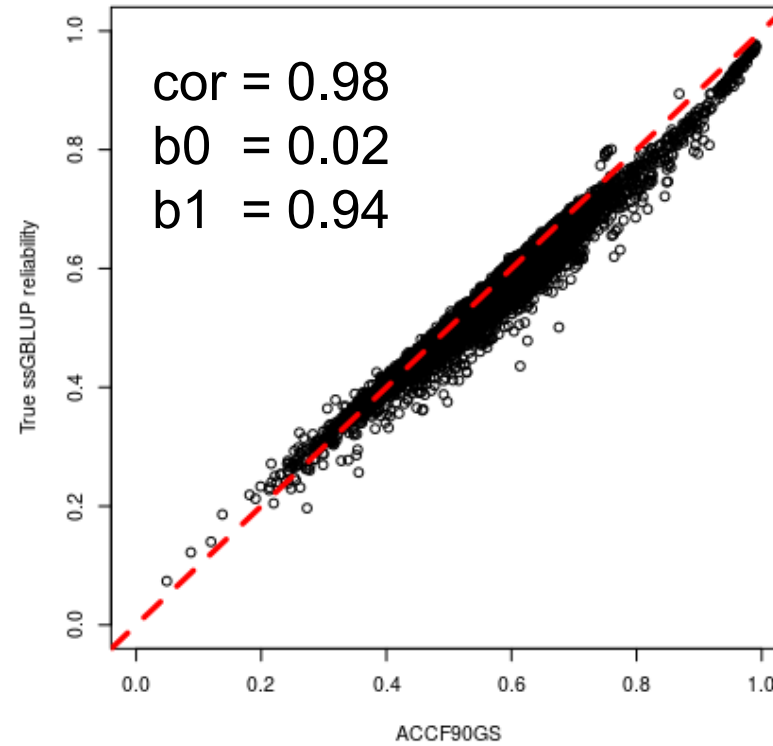
Small mt - Results

Multiple trait model

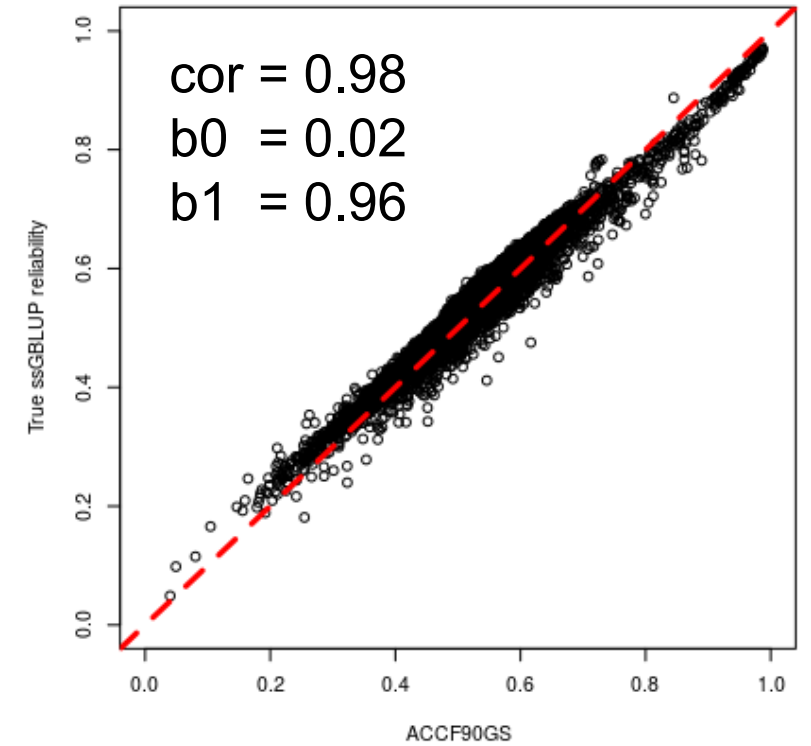
BW



WW

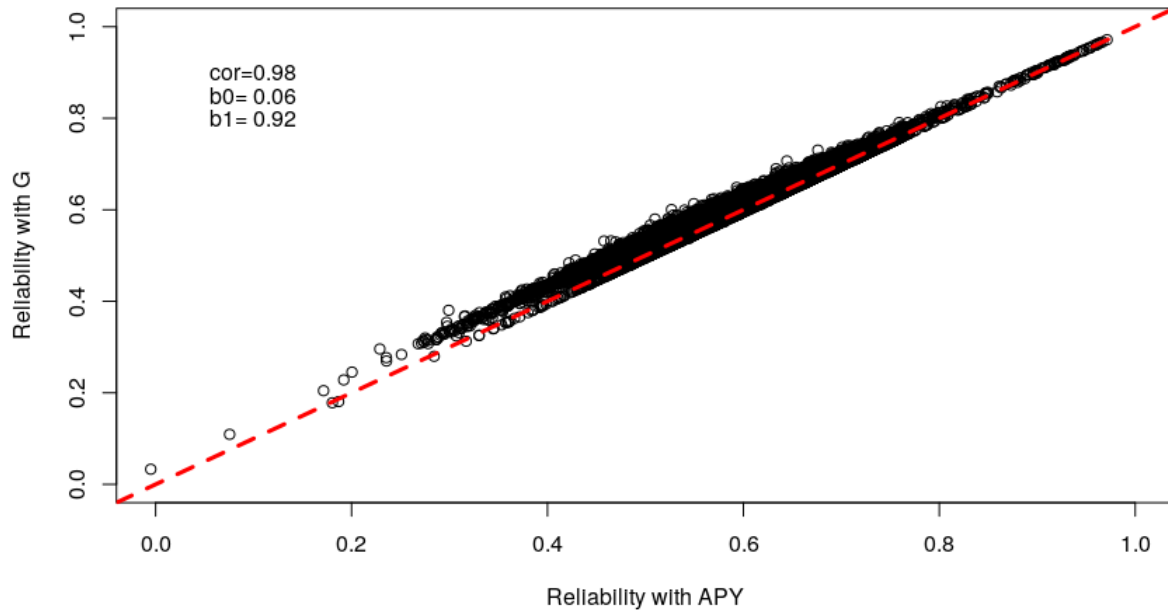


PWG

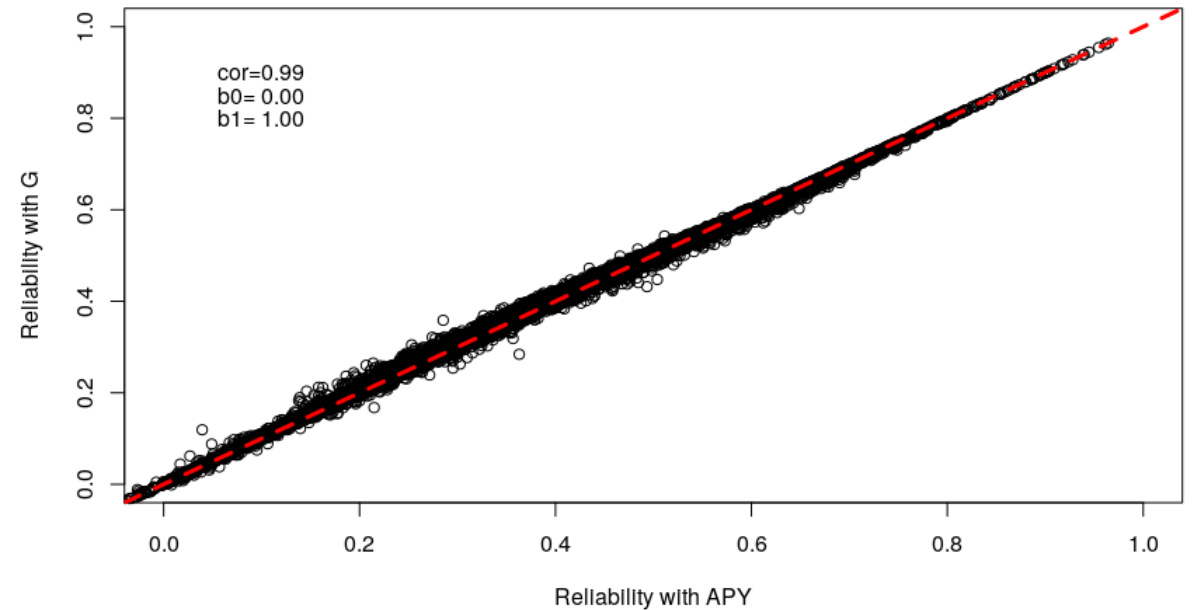


Small st – Exact full G vs. Exact APY

Genotyped



Non-Genotyped



Small st - Results

	Correlation	Intercept	Slope	MSE
Genotyped	0.98	0.02	0.94	0.0002
Non-genotyped	0.97	0.01	1.05	0.002

Small mt - Results

		Correlation	Intercept	Slope	MSE
BW	Genotyped	0.98	0.04	0.93	2.1×10^{-4}
	Non-genotyped	0.97	0.01	1.05	1.1×10^{-3}
WW	Genotyped	0.98	0.02	0.94	2.3×10^{-4}
	Non-genotyped	0.99	0.00	0.97	8.3×10^{-4}
PWG	Genotyped	0.98	0.02	0.96	2.9×10^{-4}
	Non-genotyped	0.99	0.00	1.01	6.6×10^{-4}