

Approximation of single and multiple-trait ssGBLUP reliabilities with ACCF90GS2

Matias Bermann, Daniela Lourenco, Ignacy Misztal

EAAP - August 2021



UNIVERSITY OF
GEORGIA

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.)

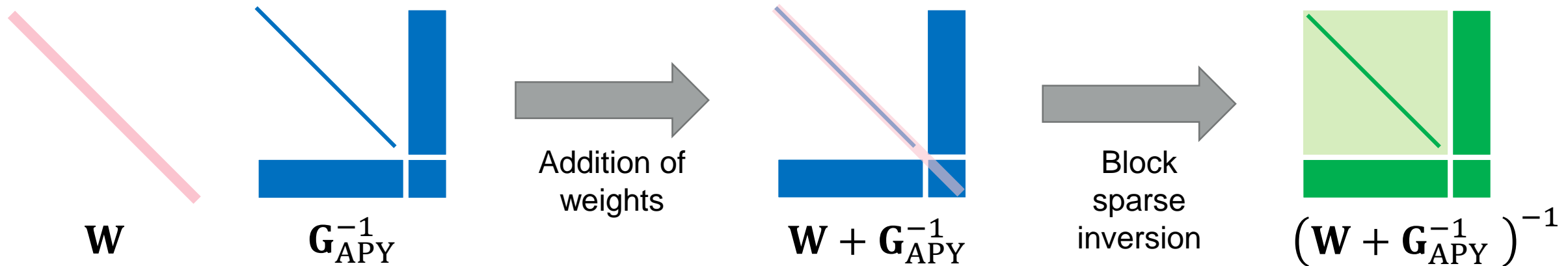
- Core animals:

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

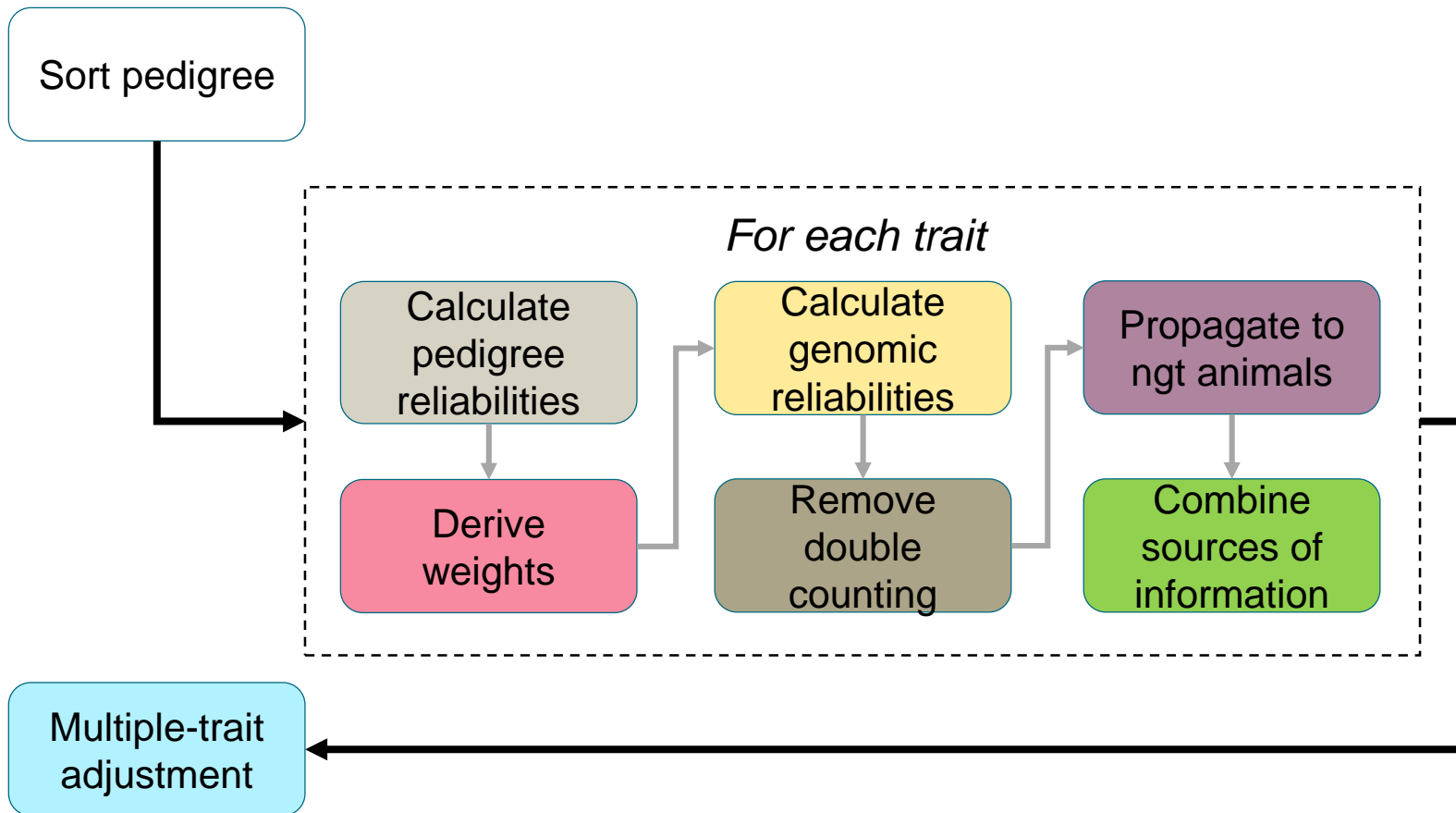
- Non-core animals:

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

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



ACCF90GS2 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)

Testing data sets

Data sets

	Small_st	Small_mt	Large_st	Large_mt
Trait(s)				
Animals in the pedigree				
Animals with records				
Genotyped animals				
Core Non-core				

Data sets

	Small_st	Small_mt	Large_st	Large_mt
Trait(s)	PWG			
Animals in the pedigree	167 k			
Animals with records	76 k			
Genotyped animals	Core Non-core	10 k 40 k		

Data sets

	Small_st	Small_mt	Large_st	Large_mt
Trait(s)	PWG	BW – WW –PWG		
Animals in the pedigree	167 k	172 k		
Animals with records	76 k	78 k		
Genotyped animals	Core Non-core	10 k 40 k	10 k 10 k	

Data sets

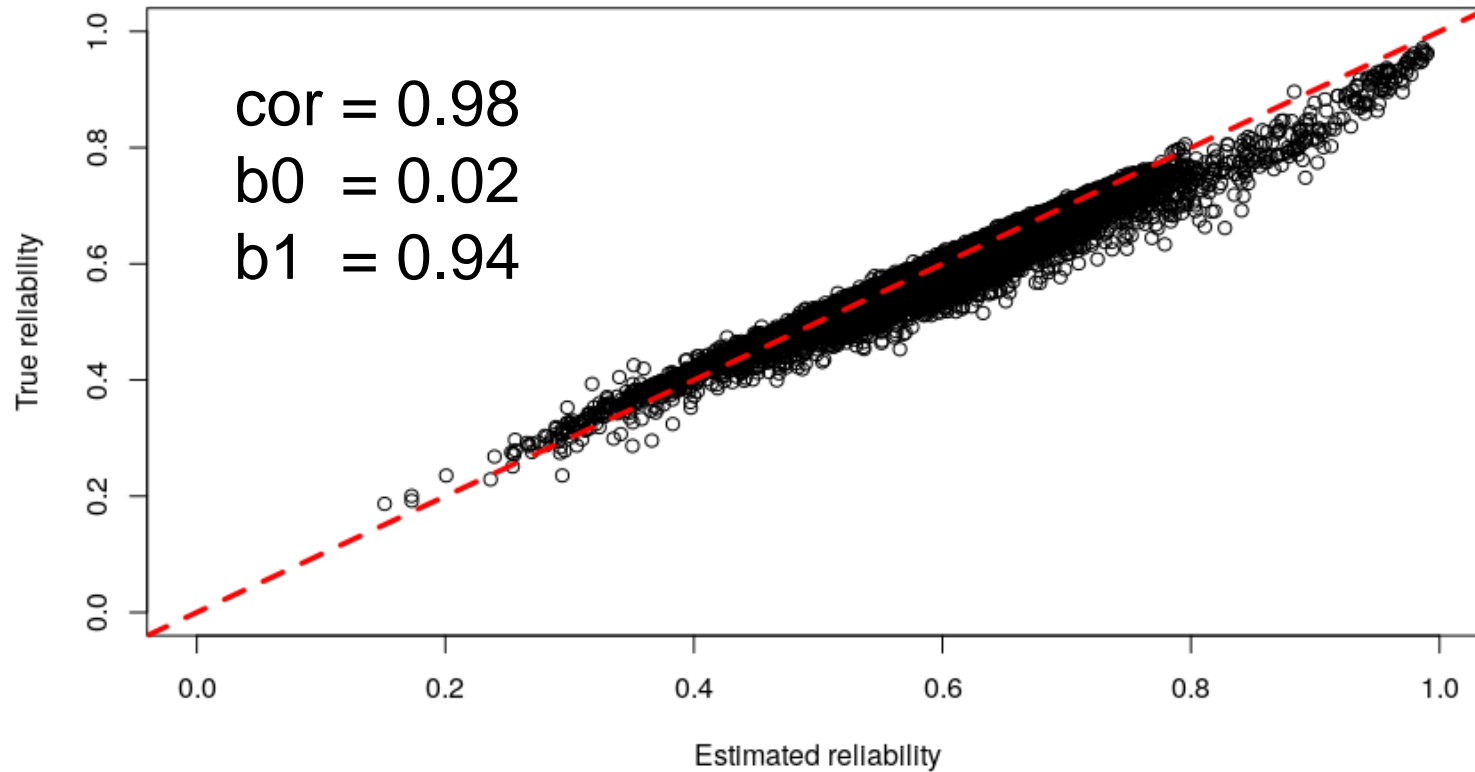
	Small_st	Small_mt	Large_st	Large_mt
Trait(s)	PWG	BW – WW –PWG	PWG	
Animals in the pedigree	167 k	172 k	10.2 m	
Animals with records	76 k	78 k	4.2 m	
Genotyped animals	Core Non-core	10 k 10 k	10 k 324 k	

Data sets

		Small_st	Small_mt	Large_st	Large_mt
Trait(s)		PWG	BW – WW –PWG	PWG	BW – WW – PWG
Animals in the pedigree		167 k	172 k	10.2 m	10.2 m
Animals with records		76 k	78 k	4.2 m	8.6 m
Genotyped animals	Core	10 k	10 k	10 k	10 k
	Non-core	40 k	10 k	324 k	324 k

Results for the small datasets

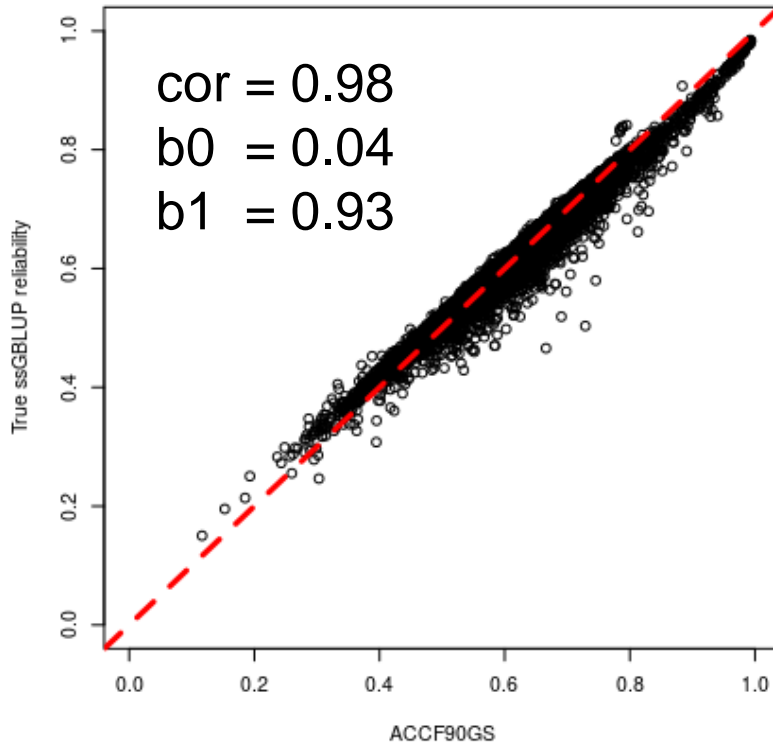
Small st - Results



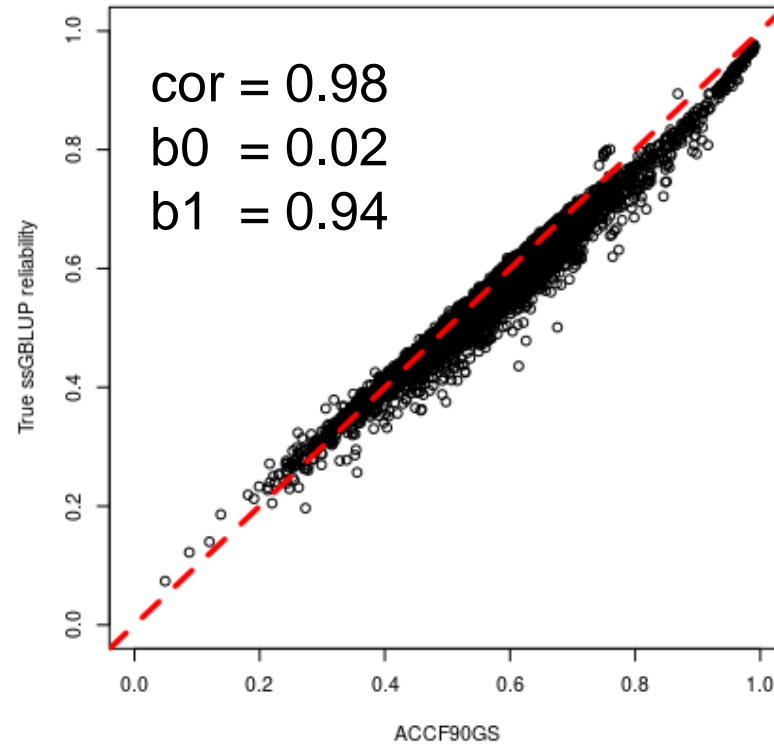
Small mt - Results

Multiple trait model

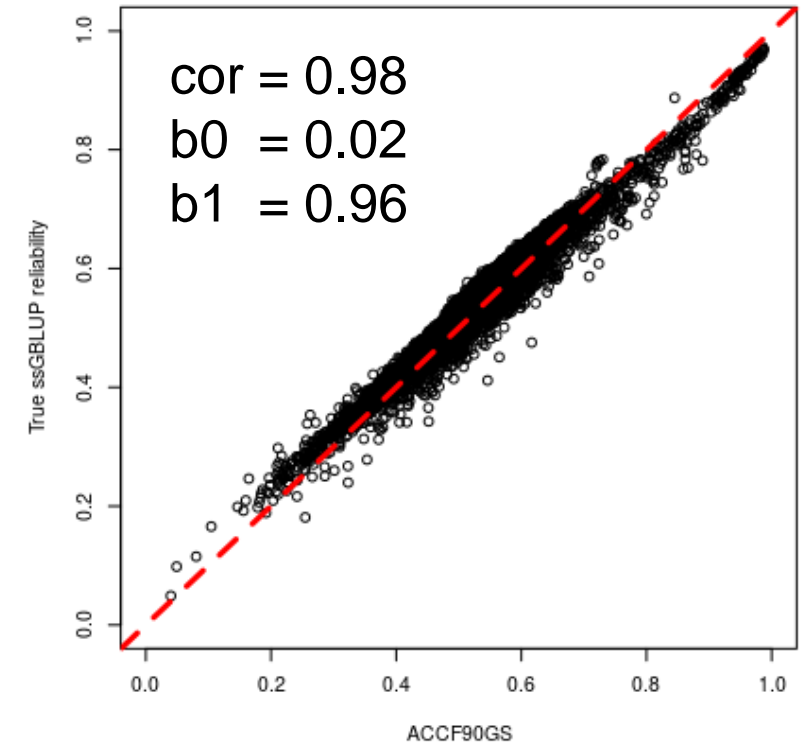
BW



WW



PWG



Computational performance

Computing times (min)

	Small_st	Small_mt	Large_st	Large_mt
Genomic reliabilities				
Other steps				
Total time				

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

Computing times (min)

	Small_st	Small_mt	Large_st	Large_mt
Genomic reliabilities	0.28			
Other steps	0.03			
Total time	0.31			

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

Computing times (min)

	Small_st	Small_mt	Large_st	Large_mt
Genomic reliabilities	0.28	0.36		
Other steps	0.03	0.06		
Total time	0.31	0.42		

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

Computing times (min)

	Small_st	Small_mt	Large_st	Large_mt
Genomic reliabilities	0.28	0.36	1.85	
Other steps	0.03	0.06	1.47	
Total time	0.31	0.42	3.32	

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

Computing times (min)

	Small_st	Small_mt	Large_st	Large_mt
Genomic reliabilities	0.28	0.36	1.85	4.74
Other steps	0.03	0.06	1.47	6.73
Total time	0.31	0.42	3.32	11.11

- X16 genotyped animals
- X13 computing time
- X3 traits
- X3.3 computing time
- X7 genotyped animals
- X7 computing time

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

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 Berman – mbermann@uga.edu



Genomic reliabilities (cont.)

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

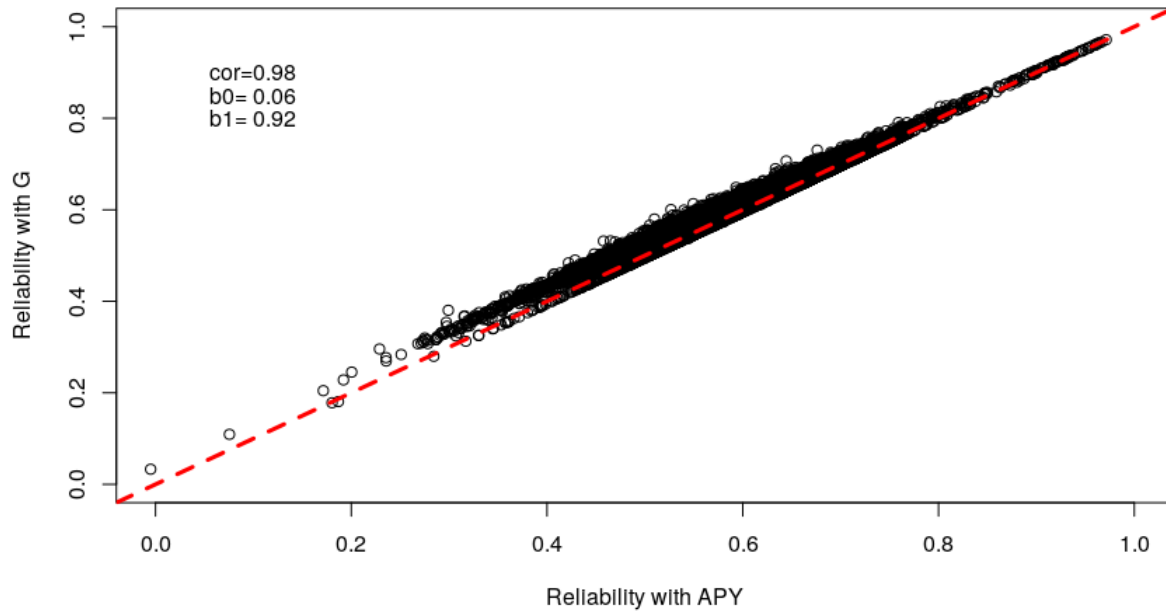
1. Read $\mathbf{G}_{\text{APY}}^{-1}$ and store in memory
2. Approximate \mathbf{W} and overwrite $\mathbf{G}_{\text{APY}}^{-1}$ as $\mathbf{G}_{\text{APY}}^{-1} = \mathbf{W} + \mathbf{G}_{\text{APY}}^{-1} \alpha$
3. Compute $\mathbf{G}^{\text{cn}*} = \mathbf{G}^{\text{cn}} \mathbf{M}_{\text{nn}}$
4. 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 st - Results

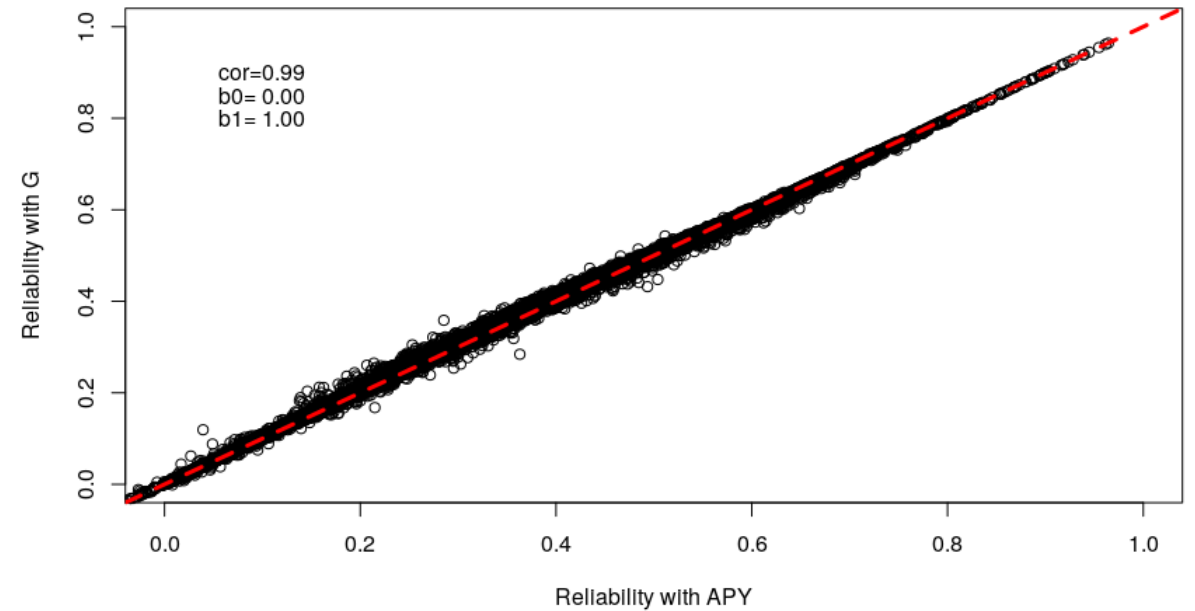
	Correlation	Intercept	Slope	MSE
Genotyped	0.98	0.02	0.94	0.0002
Non-genotyped	0.97	0.01	1.05	0.002

Small st – Exact full G vs. Exact APY

Genotyped



Non-Genotyped



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}