

Updates on large-scale genomic analyses

Daniela Lourenco

M. Bermann, S. Tsuruta, I. Misztal

September 17, 2024



**UNIVERSITY OF
GEORGIA**

**College of Agricultural &
Environmental Sciences**

Dairy data recording

1935



DAIRY HERD IMPROVEMENT RECORD

BUREAU OF DAIRY INDUSTRY, DEPT. OF AGRICULTURE

2023

Trait groups

Trait group	Records	Since
Yield	103 M	1960
Productive life	40 M	1960
Somatic cell score	67 M	1982
Female fertility	94 M	1960
Cow livability	121 M	1960
Heifer livability	6 M	2009
Health	10 M	1985
Gestation length	23 M	1970
Residual feed intake	9,334	2000
Pedigree Animals	~ 95 M	~ 1900

Genotyped animals in 15 years



4 million genotyped animals

4 M



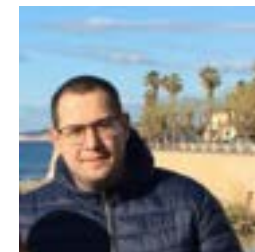
30 M



45 M



<https://www.usda.gov/media/blog/2020/06/18/data-saydairy-has-changed>



J. Dairy Sci. 105:5141–5152
<https://doi.org/10.3168/jds.2021-21805>

© 2022, The Authors. Published by Elsevier Inc. and FASS Inc. on behalf of the American Dairy Science Association.
This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Multibreed genomic evaluation for production traits of dairy cattle in the United States using single-step genomic best linear unbiased predictor

A. Cesarani,^{1*} D. Lourenco,¹ S. Tsuruta,¹ A. Legarra,² E. L. Nicolazal,² P. M. VanRaden,³ and I. Misztal¹

- 5 breeds
- 3 traits
- > 200 M equations

- Did it work?
- 72 hours
- 1 TB RAM
- ssGBLUP

Single-step GBLUP

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} + \mathbf{H}^{-1} \frac{\sigma_e^2}{\sigma_u^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

Single-step GBLUP
(ssGBLUP)

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



Pedigree



Genomic



Pedigree for
genotyped

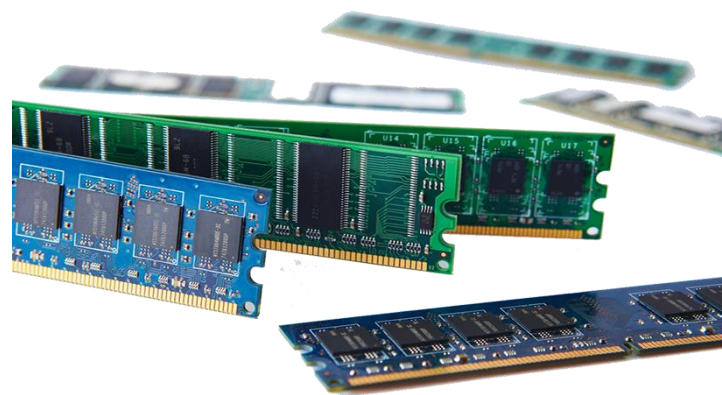
Aguilar et al. (2010)
Christensen and Lund (2010)

“Memory” exercise

- 4 M genotyped individuals

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2 \sum_{i=1}^{SNP} p_i(1 - p_i)}$$

VanRaden (2008)



<https://www.hp.com/us-en/shop/tech-takes/what-are-gigabytes-of-ram>

RAM

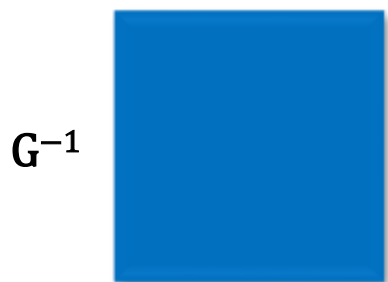
$$\text{RAM}_{\text{TB}} = N * M * 8/1024^4$$

$$\text{RAM}_{\text{TB}} = 4M * 4M * 8/1024^4 = 116 \text{ TB}$$

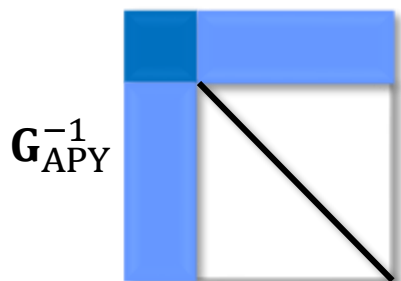
Algorithm for Proven and Young (APY)

- Realized relationship matrix in ssGBLUP

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


 \mathbf{G}^{-1}

Dense $\rightarrow u_i | u_1 + u_2 + u_3, \dots, u_{i-1} = \sum_{j=1}^{n-1} p_{ij} u_j + \varepsilon_i$


 \mathbf{G}_{APY}^{-1}

Sparse $\rightarrow u_i | u_{c1} + u_{c2} + u_{c3}, \dots, u_{ci} = \sum_{j=1}^c p_{ij} u_j + \varepsilon_i$

Condition on a set of features or animals = CORE animals

Misztal et al. (2014)
 Fragomeni et al. (2015)
 Lourenco et al. (2015)


 \mathbf{G}_{APY}^{-1}


- \mathbf{G}_{APY}^{-1} sparse
- Efficient computations

Masuda et al. (2016)

Efficient Single-step GBLUP

- Realized relationship matrix in ssGBLUP

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

$$\mathbf{G}^{-1} \rightarrow \mathbf{G}_{APY}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{cc}^{-1} \mathbf{G}_{cn} \\ \mathbf{I} \end{bmatrix} \mathbf{M}_{nn}^{-1} \begin{bmatrix} -\mathbf{G}_{nc} \mathbf{G}_{cc}^{-1} & \mathbf{I} \end{bmatrix}$$

Misztal (2016)

$$\mathbf{M}_{nn} = \text{diag}\{g_{ii} - g_{ic} \mathbf{G}_{cc}^{-1} g_{ci}\}$$

$$\mathbf{A}^{-1} \rightarrow \mathbf{A}^{-1} = (\mathbf{I} - \mathbf{P})' \mathbf{M}^{-1} (\mathbf{I} - \mathbf{P})$$

```

Inf=(1., -1., -1.)
Minv=(2., 4/3., 1., 0.)
ainv=0.0

do
  read(1,*,lostat=ic) animal, sire, dam,par_stat
  if (ic /= 0) exit
  p(1)=animal
  p(2)=sire
  p(3)=dam
  do i=0,animal
    do k=1,3
      do l=1,3
        ainv(k,l)=ainv(k,l)+Inf(k)*Inf(l)*Minv(par_stat)
      enddo
    enddo
  enddo
enddo

```

Henderson (1976)

Quaas (1988)

$$\mathbf{A}_{22}^{-1} \rightarrow \mathbf{A}_{22}^{-1} = \mathbf{A}^{22} - \mathbf{A}^{20} (\mathbf{A}^{00})^{-1} \mathbf{A}^{20}$$

Henderson (1976)

Masuda et al. (2017)

ssGBLUP with 4 M genotyped animals

- 4M genotyped – 15k core
- 30M pedigree and 45M records
- 5 breeds

Core animals across breeds

AY = 32 (9.2k)

BS = 182 (47k)

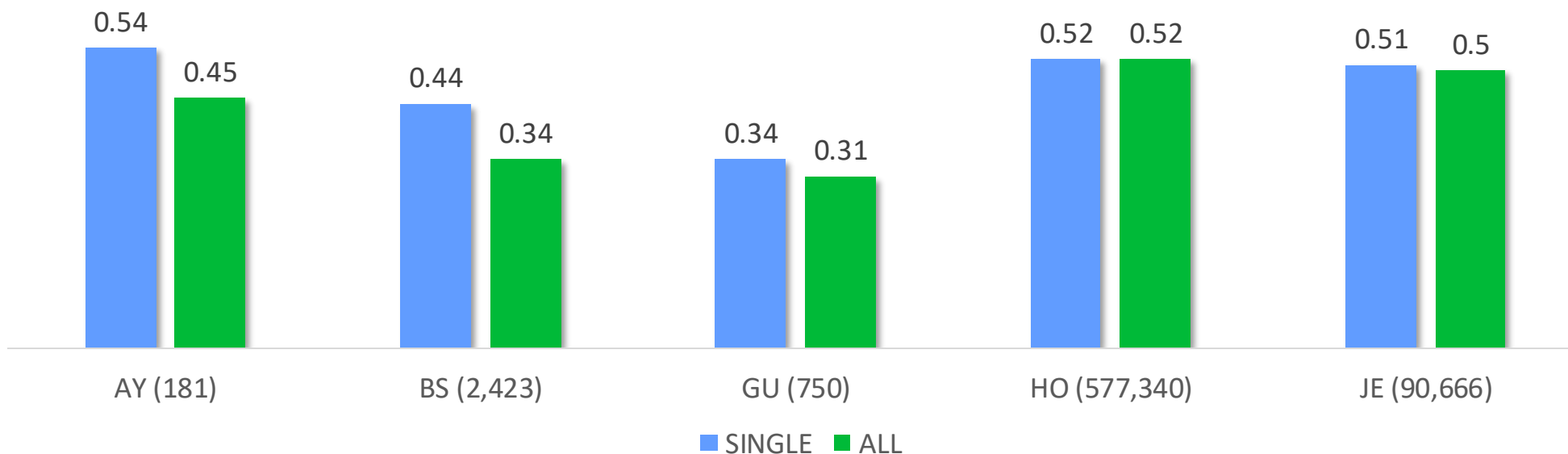
GU = 17 (5k)

HO = 13k (3.4M)

JE = 1.7k (427k)

ALL = 15k core

Accuracy for cows - Protein



ssGBLUP with 4 M genotyped animals

Dimensionality within each breed

AY = 5k

BS = 5k

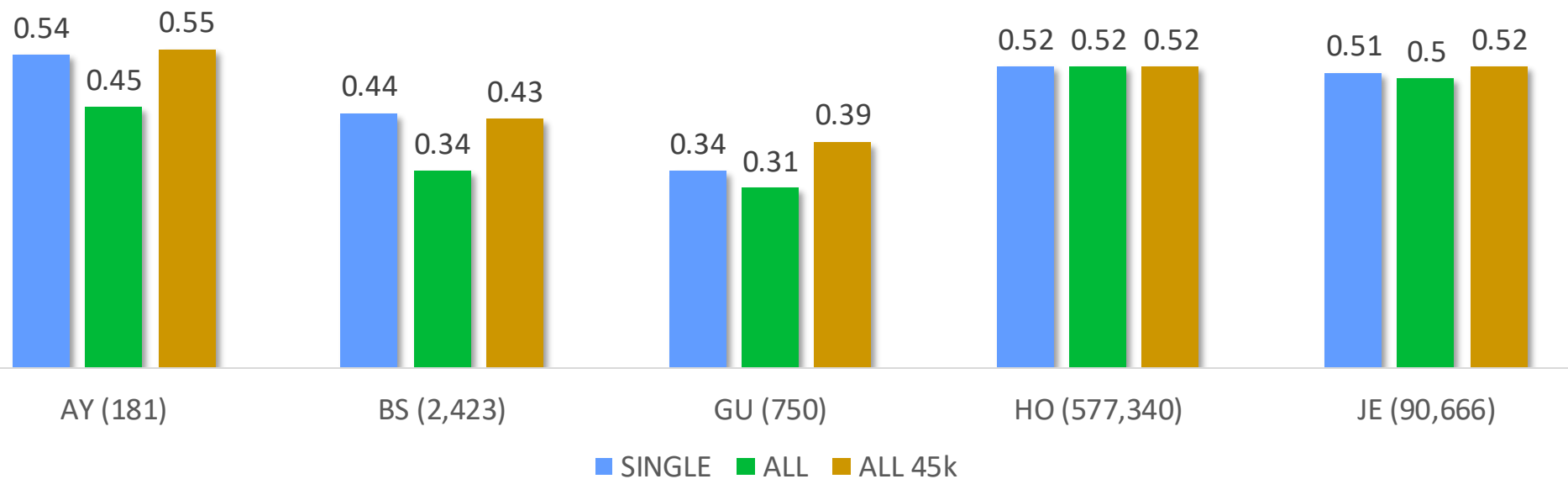
GU = 5k

HO = 15k

JE = 15k

Accuracy for cows - Protein

ALL 45k = 45k core



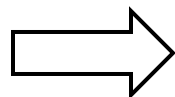
Model

Milk (MY), fat (FY), and protein (PY) yields recorded from January 2000 to June 2020

- Model:
$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_h\mathbf{h} + \mathbf{Z}_a\mathbf{Q}_a\mathbf{g}_a + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_p\mathbf{p} + \mathbf{e}$$



herd × management
 age × parity
 inbreeding
 heterosis



Breed-specific

Breed effect

- Altered QP-transformation for \mathbf{H}^{-1} (Tsuruta et al., 2019)

- UPG: breed, sex, and year

$$\mathbf{H}^* = \mathbf{A}^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -(-\mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ 0 & -\mathbf{Q}'_2(-\mathbf{A}_{22}^{-1}) & \mathbf{Q}'_2(-\mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix}$$

Computing time

	Rounds		Sec / round		Time	
	BLUP	ssGBLUP	BLUP	ssGBLUP	BLUP	ssGBLUP
AY	504	863	0.08	0.08	< 1 min	~ 1 min
BS	364	867	0.18	0.45	1 min	~ 6 min
GU	345	757	0.07	0.07	< 1 min	< 1 min
HO	457	473	21.25	56.31	2.7 h	7.4 h
JE	586	432	2.00	5.58	~ 20 min	~ 40 min
ALL		1,142		64.84		~ 20 h
ALL_45k	643	1,763	27.01	130.68	4.8 h	~ 64 h

- 2.7 days for solutions
- 5 days for computing \mathbf{G}_{APY}^{-1} and \mathbf{A}_{22}^{-1} in ALL_45k

Updates in \mathbf{A}_{22} for blending

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



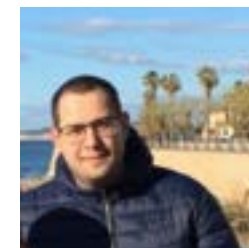
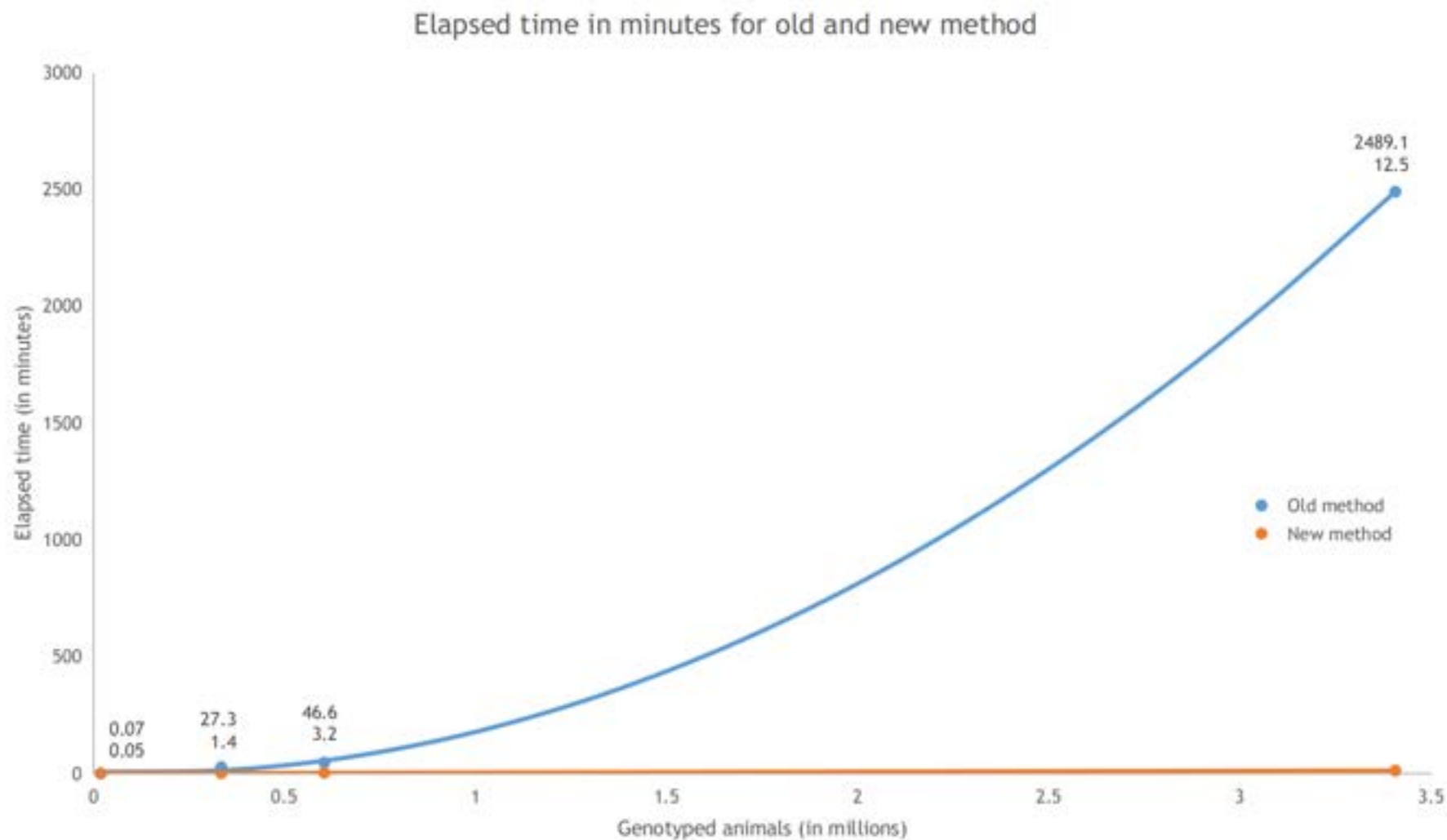
Matias Bermann

\mathbf{A}_{22}^{-1} components: $\mathbf{A}_{22}^{-1} = \mathbf{A}^{22} - \mathbf{A}^{21}(\mathbf{A}^{11})^{-1}\mathbf{A}^{12}$

APY \mathbf{G}^{-1} : $\mathbf{G}_{\text{APY}}^{-1} = \begin{bmatrix} \mathbf{G}_{\text{CC}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{\text{CC}}^{-1}\mathbf{G}_{\text{cn}} \\ \mathbf{I} \end{bmatrix} \mathbf{M}_{\text{nn}}^{-1} \begin{bmatrix} -\mathbf{G}_{\text{nc}}\mathbf{G}_{\text{CC}}^{-1} & \mathbf{I} \end{bmatrix}$

- Blending: $\mathbf{G} = 0.95 \mathbf{G}^* + 0.05 \mathbf{A}_{22}$
 - Colleau (2002)
 - Rearranging Colleau for core and noncore: 12.5 minutes

Updates in A_{22} for blending



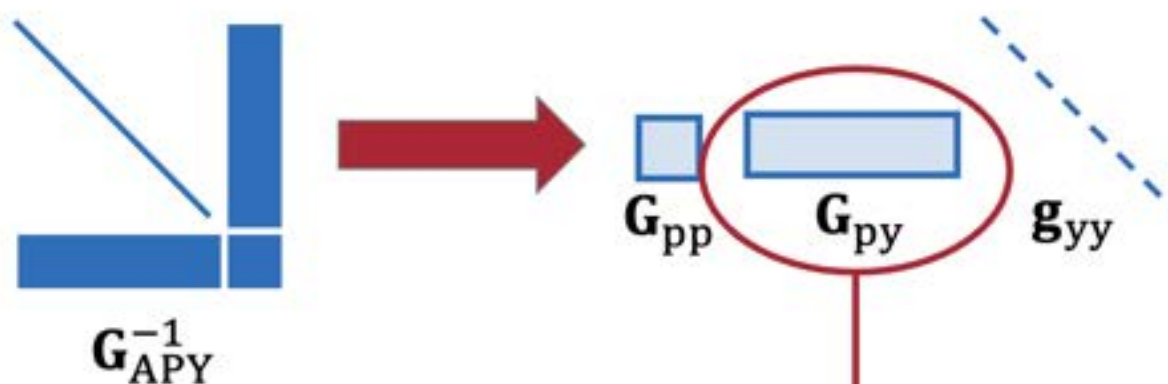
Alberto
Cesarani

1TB of RAM

- From 5 days to 8 hours to compute G_{APY}^{-1} and A_{22}^{-1} in ALL_45k

Better memory management

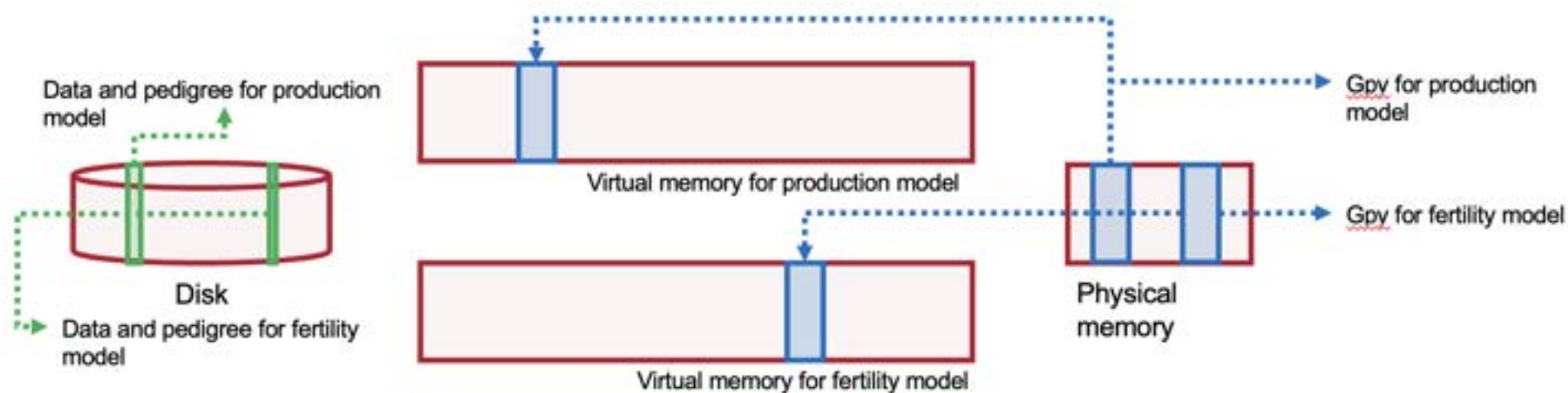
- Iteration on data
 - Data in disk and parallelization by splitting data and pedigree files
 - Genomic matrices in memory
- Large genomic datasets -> APY



- Largest memory usage
- $n_p \times n_y \times 8 \text{ byte} \times 9.31 \times 10^{-10} \text{ GB/byte}$
- For $n_p = 30,000$ and $n_y = 3,500,000$ -> 782.04 GB

Better memory management

OLD



Bermann et al.
(Unpublished)

ssGBLUP with Memory Mapping

- 4 fertility traits: CCR, HCR, DPR, and EFC
- 50M records, 60M in pedigree, ~2M animals genotyped, ~500M equations
- Why only 2M genotyped animals?
 - Many have no records and no progeny records
- ssGBLUP with APY: 45k core
 - Multi-breed
- How to compute GEBV for the remaining 7M genotyped animals?
 - Indirect predictions



Tabet et al.
(Under review)

ssGBLUP with Memory Mapping

Memory mapping

- use “memory mapping” `mmap()` to handle G_{APY}^{-1}
- A **memory-mapped file** is a segment of virtual memory^[1] that has been assigned a direct byte-for-byte correlation with some portion of a file [...] this correlation between the file and the memory space permits applications to treat the mapped portion as if it were primary memory.
- 720 Gb RAM become 720 Gb disk
- modern alternative to “read from file and compute” iteration-on-data

ssGBLUP with Memory Mapping

Running of APY

- PreGSf90: Set up G_{APY}^{-1} (with blending of [5% or 10%] $A_{\Gamma22}$).
 - RAM \approx 720 Gigabytes [not using ~~mmap~~] **It also has memory mapping now!**
- Blup90iod3 (PCG iteration on data)
 - uses “memory mapping” mmap() to handle G_{APY}^{-1}
 - As a result, only 120 Gb (non-genomic parts, including the 4 x 60M animals GEBVs...) are needed for the iteration
- accf90GS2 for reliabilities (Bermann et al 2022a) also uses mmap()





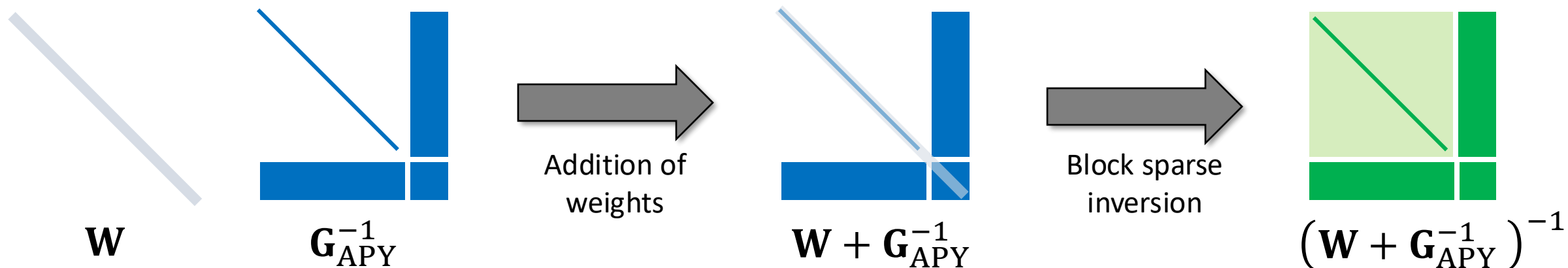
- memory mapping assigns memory to disk space
- 720 GB of RAM -> 720 GB of disk
- 120 GB of RAM

Approximating reliabilities

- Accuracy based on PEV
 - Approximated for large populations
 - Weights based on approximations
 - Block sparse inversion with APY



JOURNAL ARTICLE
 Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the Algorithm for Proven and Young 
 Matias Bermann , Daniela Lourenco, Ignacy Misztal
 Journal of Animal Science, Volume 100, Issue 1, January 2022, skab353,
<https://doi.org/10.1093/jas/skab353>



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

Equivalence ssGBLUP – ssSNPBLUP

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} + \mathbf{H}^{-1} \frac{\sigma_e^2}{\sigma_u^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'_g \mathbf{Z}_g \mathbf{M}_g & \mathbf{X}'_n \mathbf{Z}_n \\ \mathbf{M}'_g \mathbf{Z}'_g \mathbf{X}_g & \mathbf{Q} & \mathbf{M}'_g \mathbf{A}^{gn} \frac{\sigma_e^2}{\sigma_g^2} \\ \mathbf{Z}'_n \mathbf{X}_n & \mathbf{A}^{ng} \mathbf{M}_g \frac{\sigma_e^2}{\sigma_g^2} & \mathbf{Z}'_n \mathbf{Z}_n + \mathbf{A}^{nn} \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \\ \hat{\mathbf{u}}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{M}'_g \mathbf{Z}'_g \mathbf{y}_g \\ \mathbf{Z}'_n \mathbf{y}_n \end{bmatrix}$$

- Equivalent models under same assumptions and data

- Equal estimable functions

- $\hat{\mathbf{u}} = \mathbf{Z}\hat{\mathbf{a}}$

- $\hat{\mathbf{a}}|\hat{\mathbf{u}} = k\mathbf{Z}'\mathbf{G}^{-1}\hat{\mathbf{u}}$

- $Var(\hat{\mathbf{a}}|\hat{\mathbf{u}}) = k\mathbf{Z}'\mathbf{G}^{-1}(\mathbf{G} - \mathbf{C}^{\mathbf{u}_2\mathbf{u}_2})\mathbf{G}^{-1}\mathbf{Z}k$

Stranden and Garrick (2009)

Guladron-Duarte et al. (2014)

- $p\text{-value}_i = 2 \left(1 - \Phi \left(\left| \frac{\hat{a}_i}{sd(\hat{a}_i)} \right| \right) \right) \rightarrow \text{ssGWAS}$

- If using APY in ssGBLUP
 - Equivalent APY ssSNPBLUP model
 - $\hat{\mathbf{u}} = \mathbf{Z}^{\dagger} \hat{\mathbf{a}}$
 - $\hat{\mathbf{a}} | \hat{\mathbf{u}} = k \mathbf{Z}^{\dagger'} \mathbf{G}_{\text{APY}}^{-1} \hat{\mathbf{u}} = k \mathbf{Z}'_c \mathbf{G}_{\text{CC}}^{-1} \hat{\mathbf{u}}_c$
 - $\text{Var}(\hat{\mathbf{a}} | \hat{\mathbf{u}}) = k \mathbf{Z}'_c \mathbf{G}_{\text{CC}}^{-1} (\mathbf{G}_{\text{CC}} - \mathbf{C} \mathbf{u}_{2c} \mathbf{u}_{2c}') \mathbf{G}_{\text{CC}}^{-1} \mathbf{Z}_c k$


 Function of CORE animals



Bernmann et al. *Genetics Selection Evolution* (2022) 54:52
<https://doi.org/10.1186/s12711-022-00744-7>



RESEARCH ARTICLE

Open Access

On the equivalence between marker effect models and breeding value models and direct genomic values with the Algorithm for Proven and Young

Matias Bernmann^{1*}, Daniela Lourenco¹, Natalia S. Fornari^{1,2}, Andres Legarra³ and Ignacio Misztal⁴

Single-step GWAS – now unlimited

- Genomic evaluation process
 - GEBV using APY ssGBLUP + accuracy using block sparse inversion



Leite et al. *Genetics Selection Evolution* (2024) 56:19
<https://doi.org/10.1186/s12711-024-00925-3>

Genetics Selection Evolution

- $\mathbf{C}^{\mathbf{u}_{2c}\mathbf{u}_{2c}} = (\mathbf{W} + \mathbf{G}_{\text{APY}}^{-1})^{-1}$
- $\text{Var}(\hat{\mathbf{a}}|\hat{\mathbf{u}}) = k\mathbf{Z}'_c\mathbf{G}_{\text{CC}}^{-1}(\mathbf{G}_{\text{CC}} - \mathbf{C}^{\mathbf{u}_{2c}\mathbf{u}_{2c}})\mathbf{G}_{\text{CC}}^{-1}\mathbf{Z}_c k$

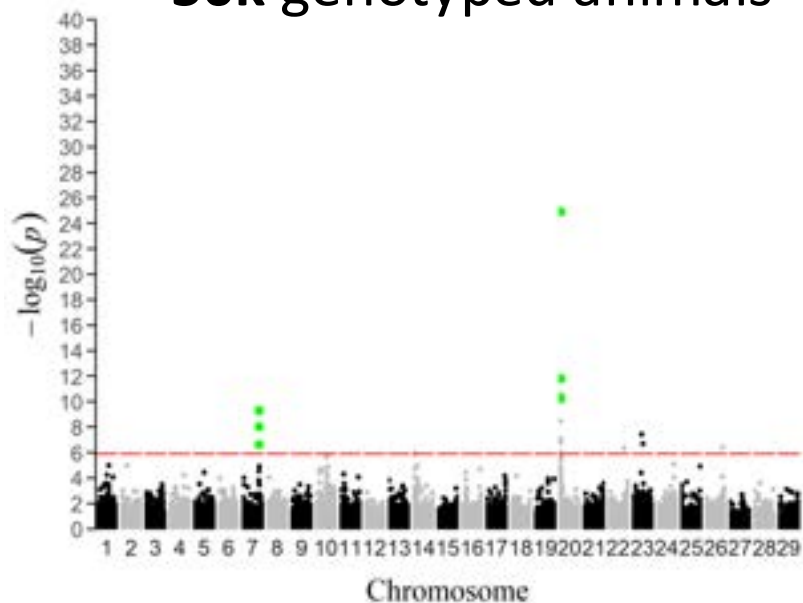
RESEARCH ARTICLE

Open Access

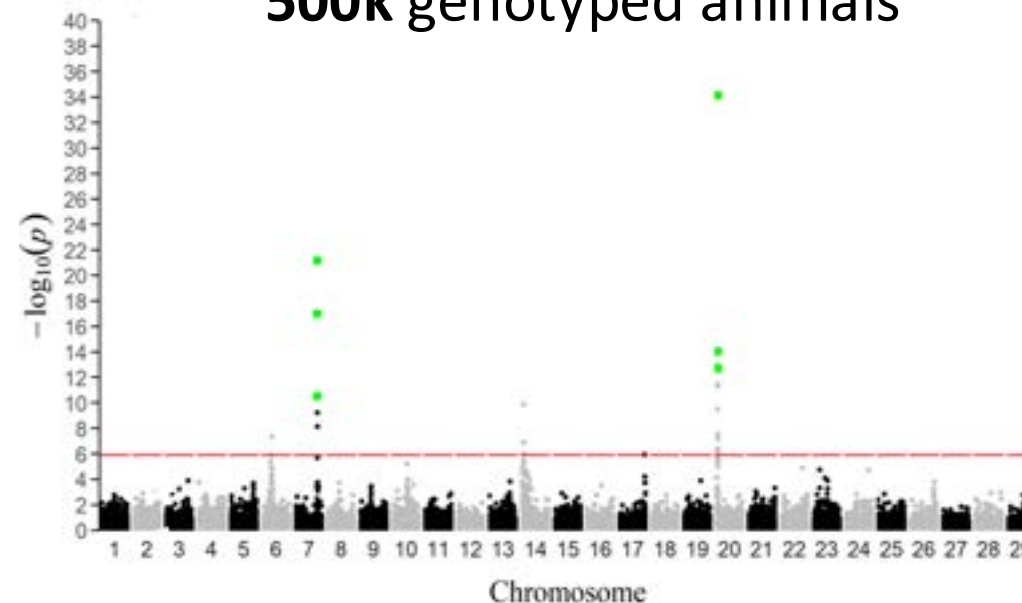
Marker effect p-values for single-step GWAS with the algorithm for proven and young in large genotyped populations

Natália Galoro Leite^{1*}, Matias Bermann¹, Shogo Tsuruta¹, Ignacy Misztal¹ and Daniela Lourenco¹

50k genotyped animals



500k genotyped animals



Single-step GWAS – now unlimited

- Genotypes

JE: 528,638

HO: 1,794,100

GU: 3,774

BS: 10,417

AY: 1,940

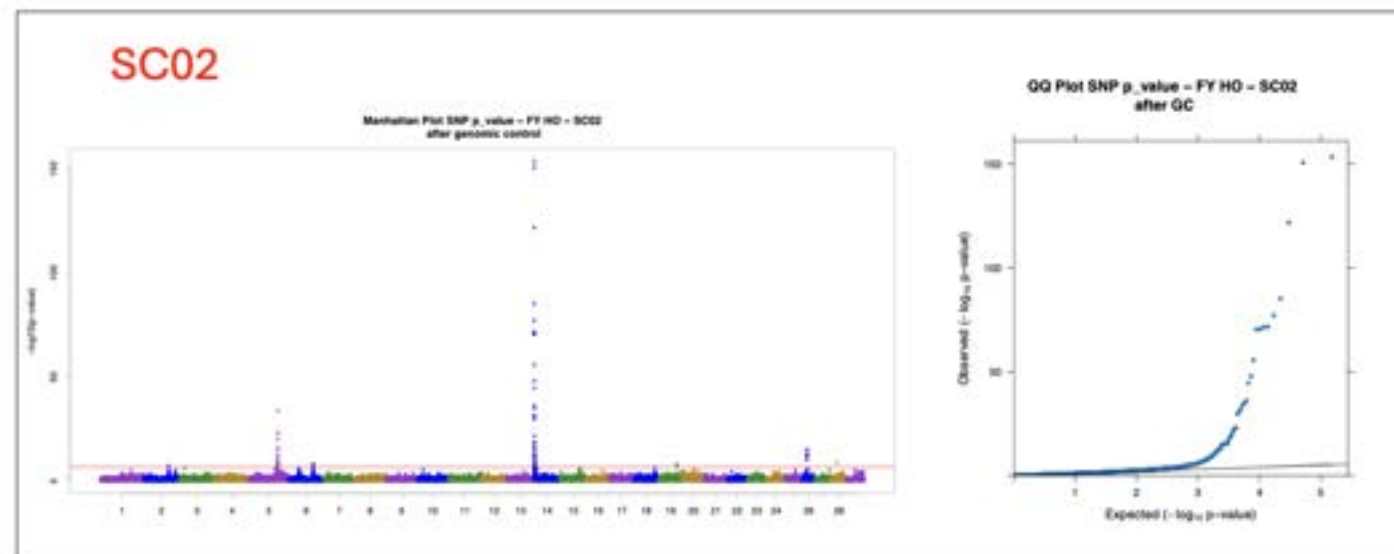
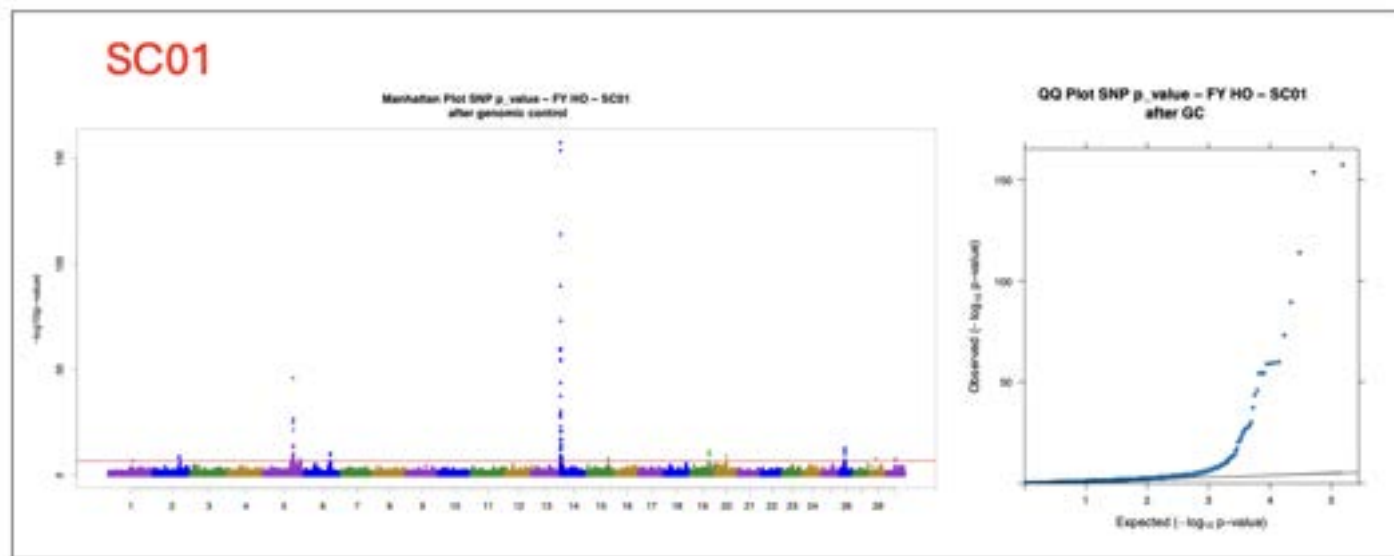
MULTI: 2,338,869

- Pedigree (all breed)

52,667,746

- Phenotypes (all breed)

106,982,064



Carrara et al.
(In progress)

Reliability of IP

- Reliability of Indirect Predictions

$$\text{REL}_{\text{IP}_j} = \frac{\mathbf{z}_j \text{var}(\hat{\mathbf{a}}) \mathbf{z}_j'}{\sigma_{\mathbf{u}}^2}$$

Liu et al. (2017)

$$\text{Var}(\hat{\mathbf{a}}) = k \mathbf{Z}'_c \mathbf{G}_{cc}^{-1} (\mathbf{G}_{cc} - \mathbf{C}^{\mathbf{u}_{2c} \mathbf{u}_{2c}}) \mathbf{G}_{cc}^{-1} \mathbf{Z}_c k$$

?

$$\mathbf{C}^{\mathbf{u}_{2c} \mathbf{u}_{2c}} = (\mathbf{W} + \mathbf{G}_{\text{APY}}^{-1})^{-1}$$

Reliability of GEBV using block sparse inversion

Holstein Dataset

- **Pedigree:** 2,240,568 animals
- **Milk Yield:** 1,422,330 Records
- **Genotypes:** Total: 33,338

Training: 32,570 bulls

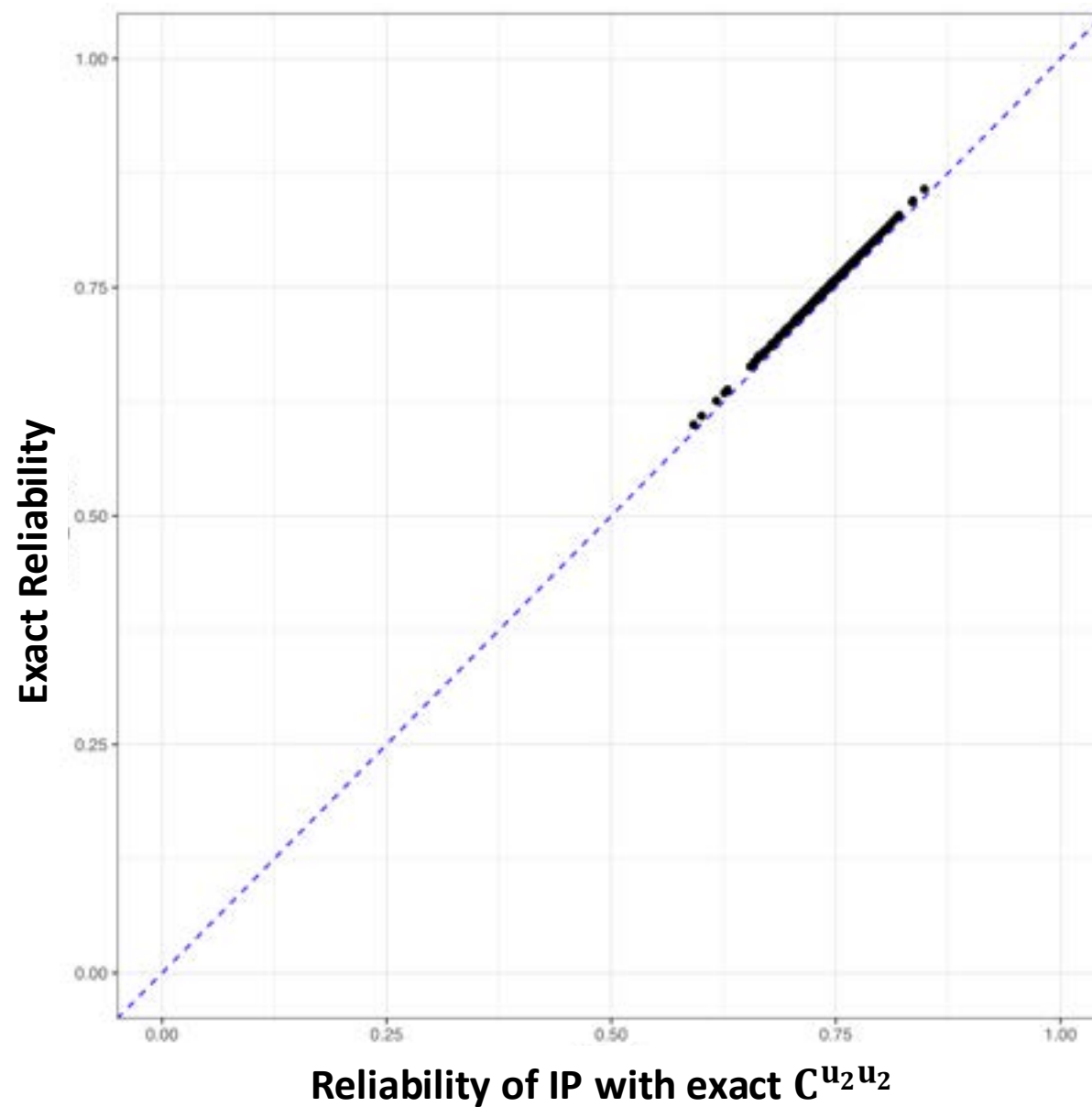
Validation (2017): 768 bulls



Tabet et al.
(In progress)

- **Exact reliabilities based on the inverse** (training + validation)
- **Reliabilities of IP** (for validation) **with exact $C^{u_2 u_2}$** (from training)
- **Reliabilities of IP** (for validation) **with approximated $C^{u_2 c u_2 c}$** (from training)

Exact vs. IP reliabilities

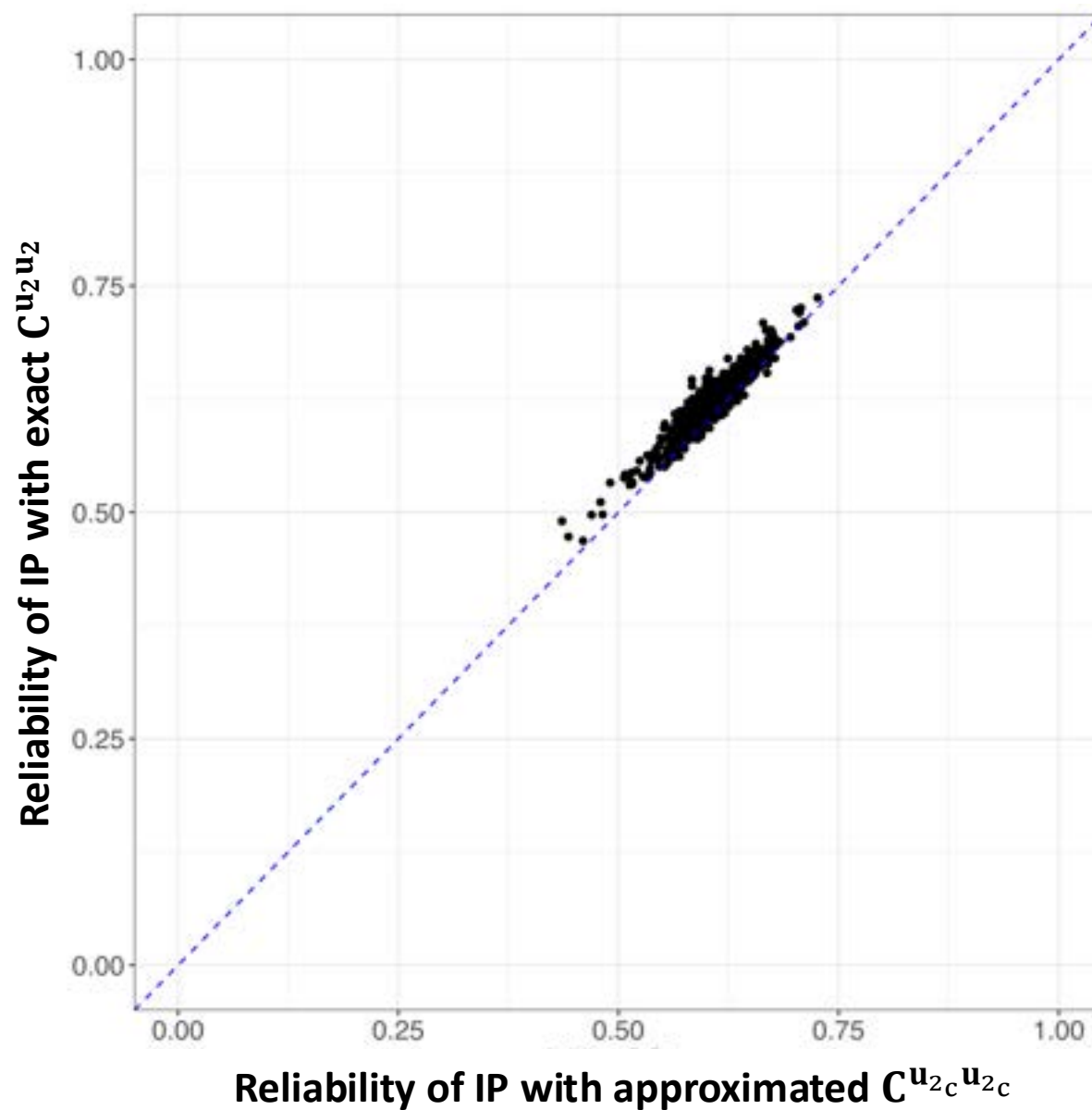


$$b_0 = 0.01$$

$$b_1 = 0.99$$

$$\text{Cor} = 0.99$$

IP reliabilities with exact vs. approx. $C^{u_2c}u_2c$



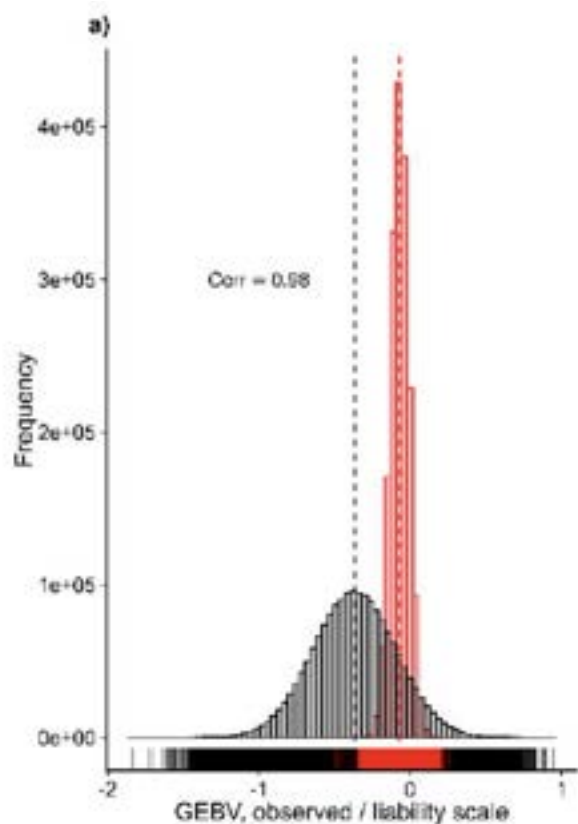
$$b_0 = 0.07$$

$$b_1 = 0.90$$

$$\text{Cor} = 0.94$$

Genomic predictions – binary/categorical

- Threshold models - Binary or categorical traits
 - > 10x more time to reach convergence
 - Liability solutions into probabilities
 - Linear solutions into probabilities (???)



$$GEBV_{lia} \approx \frac{GEBV_{lin}}{\sqrt{\sigma_{e_{lin}}^2 * \left(1 - \frac{h_{lin}^2}{h_{lia}^2}\right)}}$$

$$P_i = 1 - \Phi\left(\frac{t - \mu_u - u_i}{\sigma_e}\right)$$

Mastitis
 $h^2 = 0.09$
 $\% = 27$



J. Dairy Sci. TBC
<https://doi.org/10.3168/jds.2024-24767>

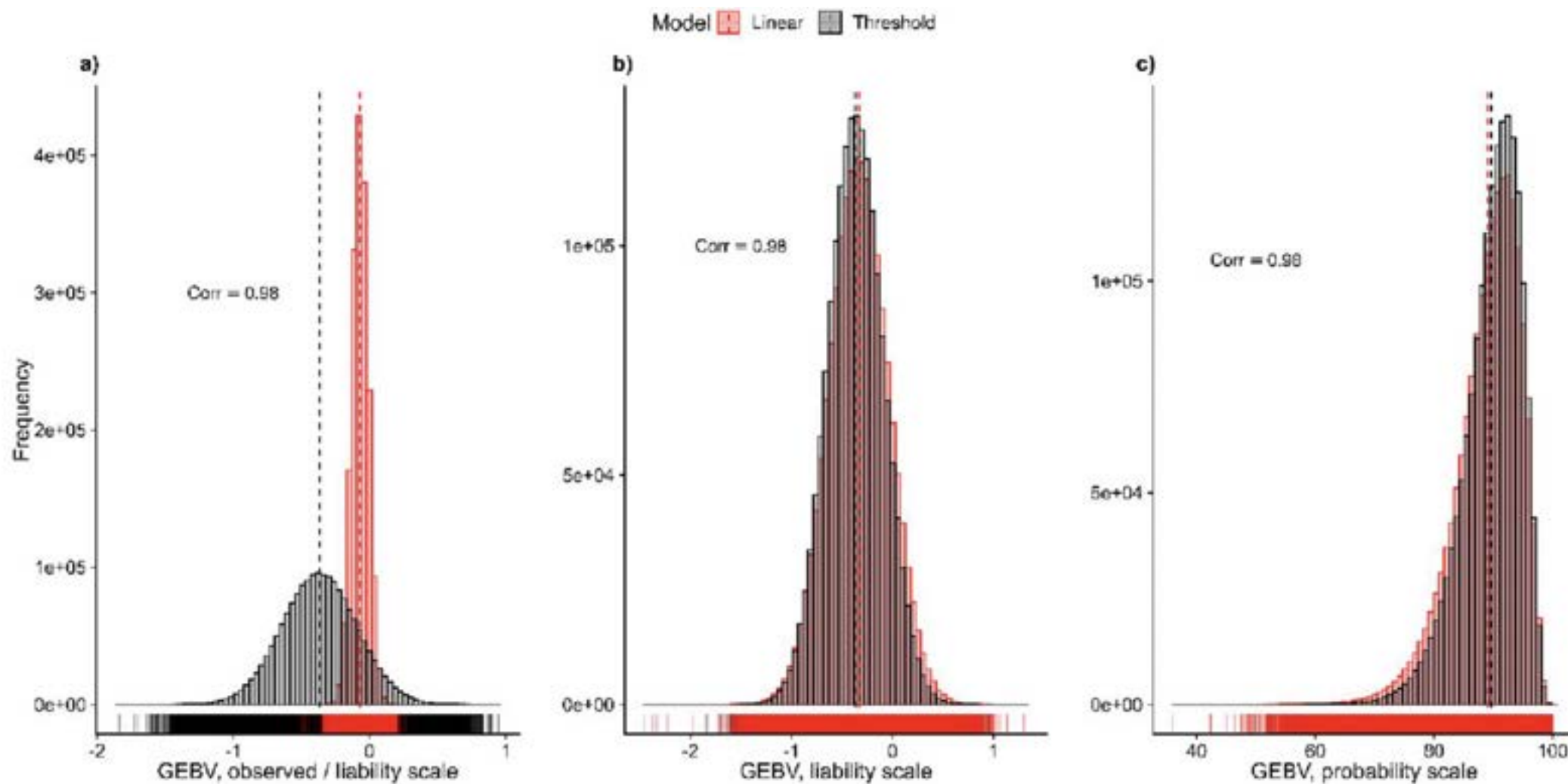
© TBC. The Authors. Published by Elsevier Inc. on behalf of the American Dairy Science Association[®].
 This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Converting estimated breeding values from the observed to probability scale for health traits

Jorge Hidalgo,^{1*} Shogo Tsuruta,¹ Dianelys Gonzalez,² Gerson de Oliveira,² Miguel Sanchez,² Asmita Kulkarni,² Cory Przybyla,² Giovana Vargas,² Natascha Vukasinovic,² Ignacy Misztal,¹ and Daniela Lourenco¹
¹Department of Animal and Dairy Science, University of Georgia, Athens, GA, 30602, USA
²Zoetis Genetics and Precision Animal Health, Kalamazoo, MI, 49007, USA



Genomic predictions – binary/categorical



Mastitis
 $h^2 = 0.09$
 % = 27

Genomic predictions – binary/categorical



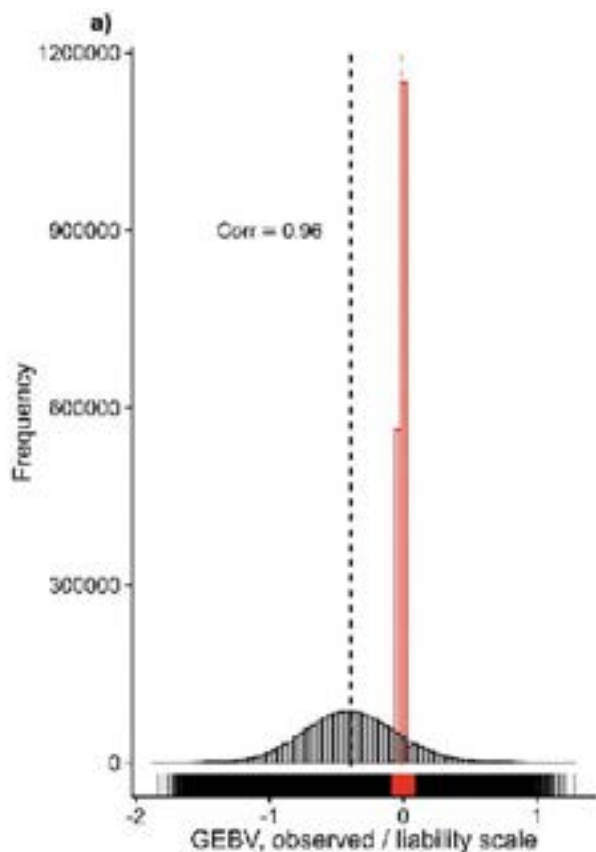
J. Dairy Sci. TBC
<https://doi.org/10.3168/jds.2024-24767>

© TBC. The Authors. Published by Elsevier Inc. on behalf of the American Dairy Science Association[®]
 This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



Converting estimated breeding values from the observed to probability scale for health traits

Jorge Hidalgo,^{1*} Shogo Tsuruta,¹ Dianelys Gonzalez,² Gerson de Oliveira,² Miguel Sanchez,² Asmita Kulkarni,² Cory Przybyla,² Giovana Vargas,² Natascha Vukasinovic,² Ignacy Misztal,¹ and Daniela Lourenco¹
¹Department of Animal and Dairy Science, University of Georgia, Athens, GA, 30602, USA
²Zoetis Genetics and Precision Animal Health, Kalamazoo, MI, 49007, USA



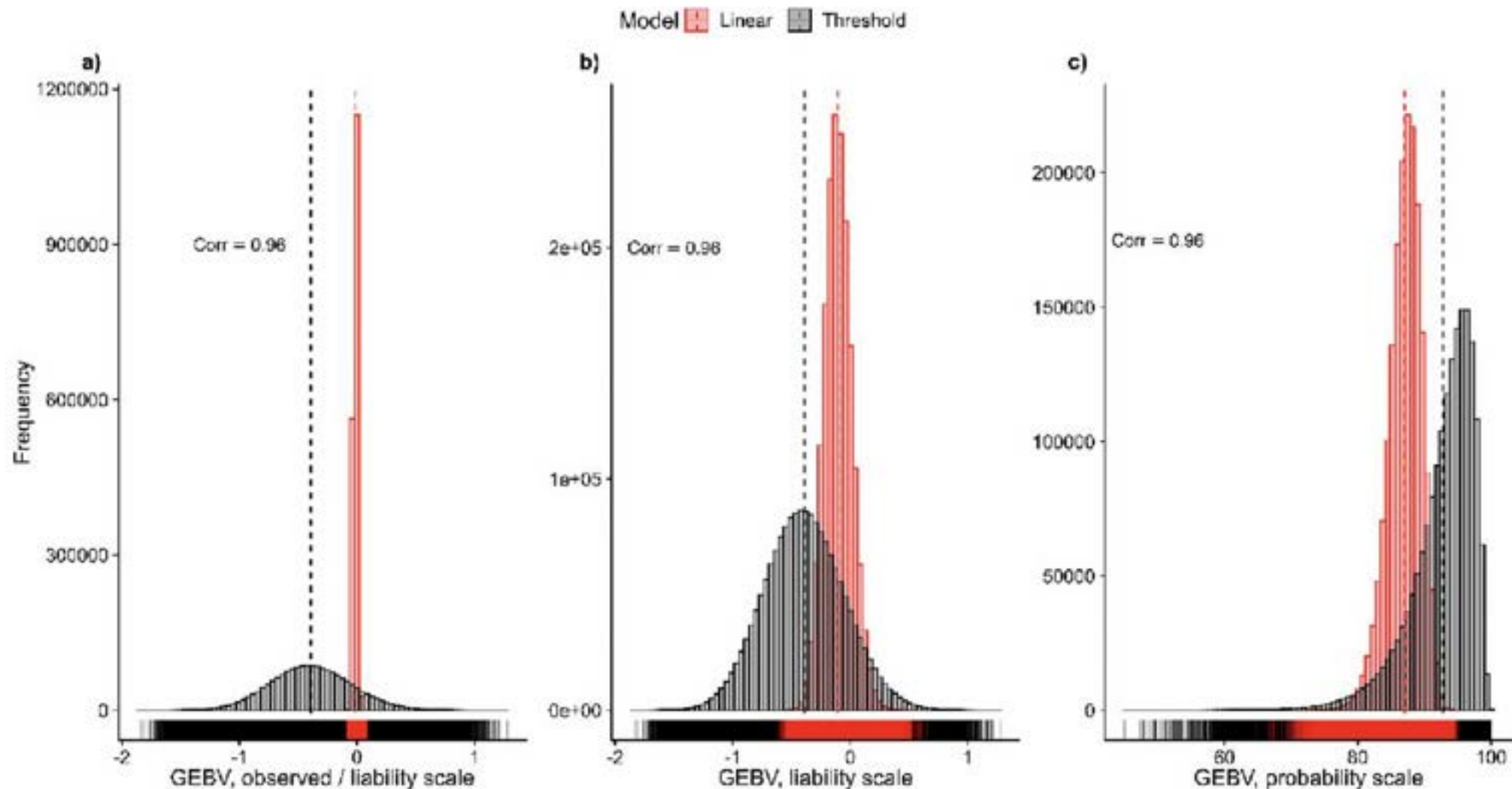
$$GEBV_{lia} \approx \frac{GEBV_{lin}}{\sqrt{\sigma_{e_{lin}}^2 * \left(1 - \frac{h_{lin}^2}{h_{lia}^2}\right)}}$$

$$P_i = 1 - \Phi\left(\frac{t - \mu_u - u_i}{\sigma_e}\right)$$

Displaced
 abomasum
 $h^2 = 0.09$
 $\% = 2$

Genomic predictions – binary/categorical

Hidalgo et al.: Estimated breeding values in probability scale



Displaced
 abomasum
 $h^2 = 0.09$
 $\% = 2$

Genomic predictions – binary/categorical

- Second approach for categorical/binary traits:
 - 1) Linear model until convergence (BLUP)
 - 2) Compute pseudo-phenotypes based on residuals (EM approach; Quaas, 1994)
 - 3) Iterate back to 1 using pseudo-phenotypes
 - 4) Do it until pseudo-phenotypes “do not change anymore”
- CATEGF90 as a wrapper for BLUP90IOD3
 - Benefit: allows for multiple categorical/binary traits



Jennifer
Richter



Andres
Legarra



Fernando
Bussiman

Genomic predictions – binary/categorical

Trait	# of Records	1	2	3 to 7
Ascites (AC)	163,971	161,950 (98.8%)	2,021 (1.2%)	-
Tibial Dyschondroplasia (TD)	59,124	57,995 (98.1%)	1,129 (1.9%)	-
Mortality (MT)	180,998	167,389 (92.4%)	13,609 (7.5%)	-
Femoral Head Necrosis (FN)	16,870	13,112 (77.7%)	2,295 (13.6%)	1,463 (8.7%)

	GIBBS	CBLUP	CATEGF90	GIBBS	CBLUP	CATEGF90
	AC			TD		
GIBBS	1	0.99	0.99	1	0.99	0.99
CBLUP		1	1		1	1
CATEGF90			1			1
	MT			FN		
GIBBS	1	1	1	1	1	1
CBLUP		1	1		1	1
CATEGF90			1			1



Jennifer
Richter



Andres
Legarra



Fernando
Bussiman

Genomic predictions – binary/categorical

Trait	# of Records	1	2	3 to 7
Ascites (AC)	1,092,037	1,083,836 (99.2%)	8,201 (0.8%)	-
Tibial Dyschondroplasia (TD)	365,676	354,143 (96.8%)	11,533 (3.2%)	-
Mortality (MT)	1,191,175	1,093,992 (91.8%)	97,183 (8.2%)	-
Femoral Head Necrosis (FN)	88,012	58,817 (66.8%)	16,935 (19.2%)	12,260 (13.9%)



Jennifer
Richter



Andres
Legarra



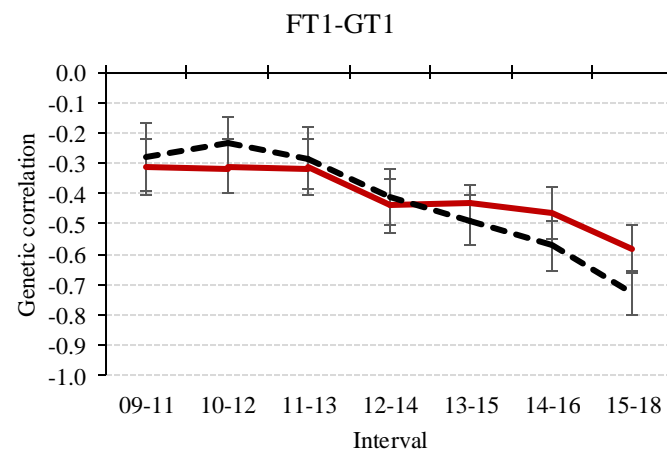
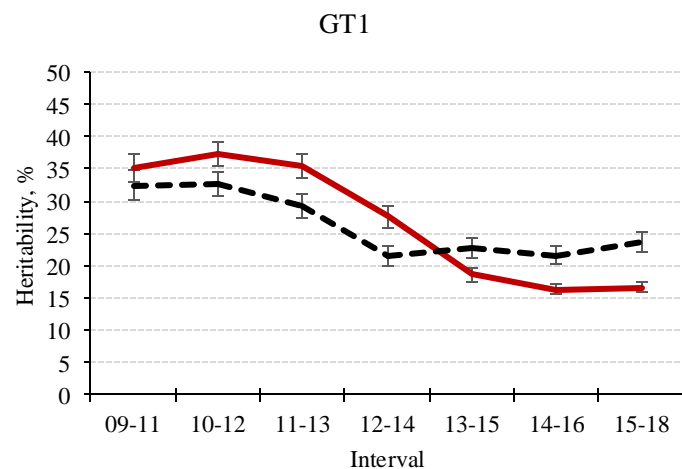
Fernando
Bussiman

	CBLUP	CATEGF90	CBLUP	CATEGF90
	AC		TD	
CBLUP	1	1	1	1
CATEGF90		1		1
	MT		FN	
CBLUP	1	1	1	1
CATEGF90		1		1

CATEGF90 iter = 515
 BLUP90IOD3 iter = 461,546
 Wall-clock time (min) = 15,897

Limitations - VCE

- Faster changes with genomic selection
- Different genetic parameters with and without genomics



- Hard to estimate VC with many genotyped individuals
 - Software optimization
 - New methods



JOURNAL ARTICLE

Changes in genetic parameters for fitness and growth traits in pigs under genomic selection

Jorge Hidalgo, Shogo Tsuruta, Daniela Lourenco, Yutaka Masuda, Yijian Huang, Kent A Gray, Ignacy Misztal

Journal of Animal Science, Volume 98, Issue 2, February 2020, skaa032,

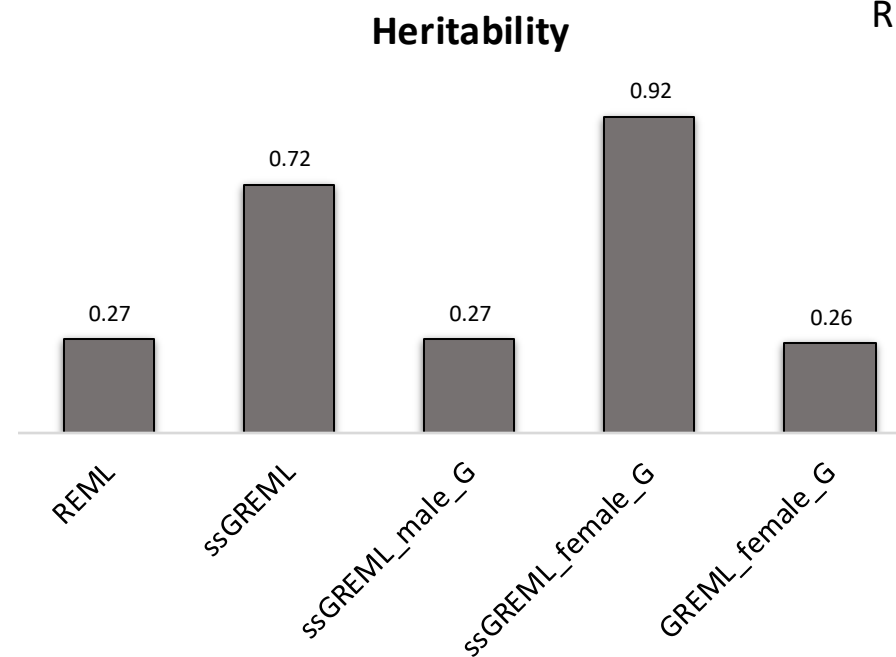
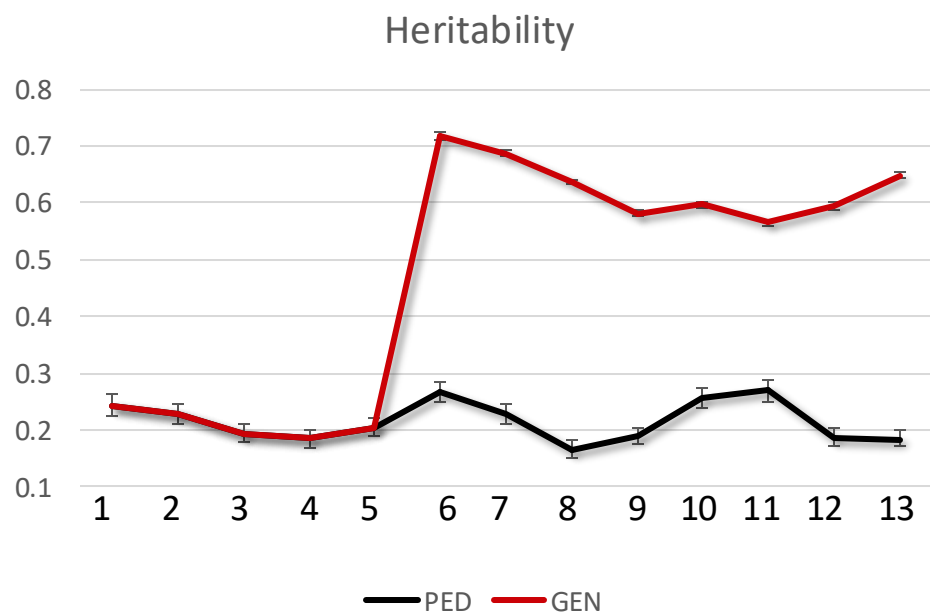
Tsuruta et al. (in progress)
 Gowane et al. (in progress)

VCE with and without genomics

- Variance components and genetic parameters



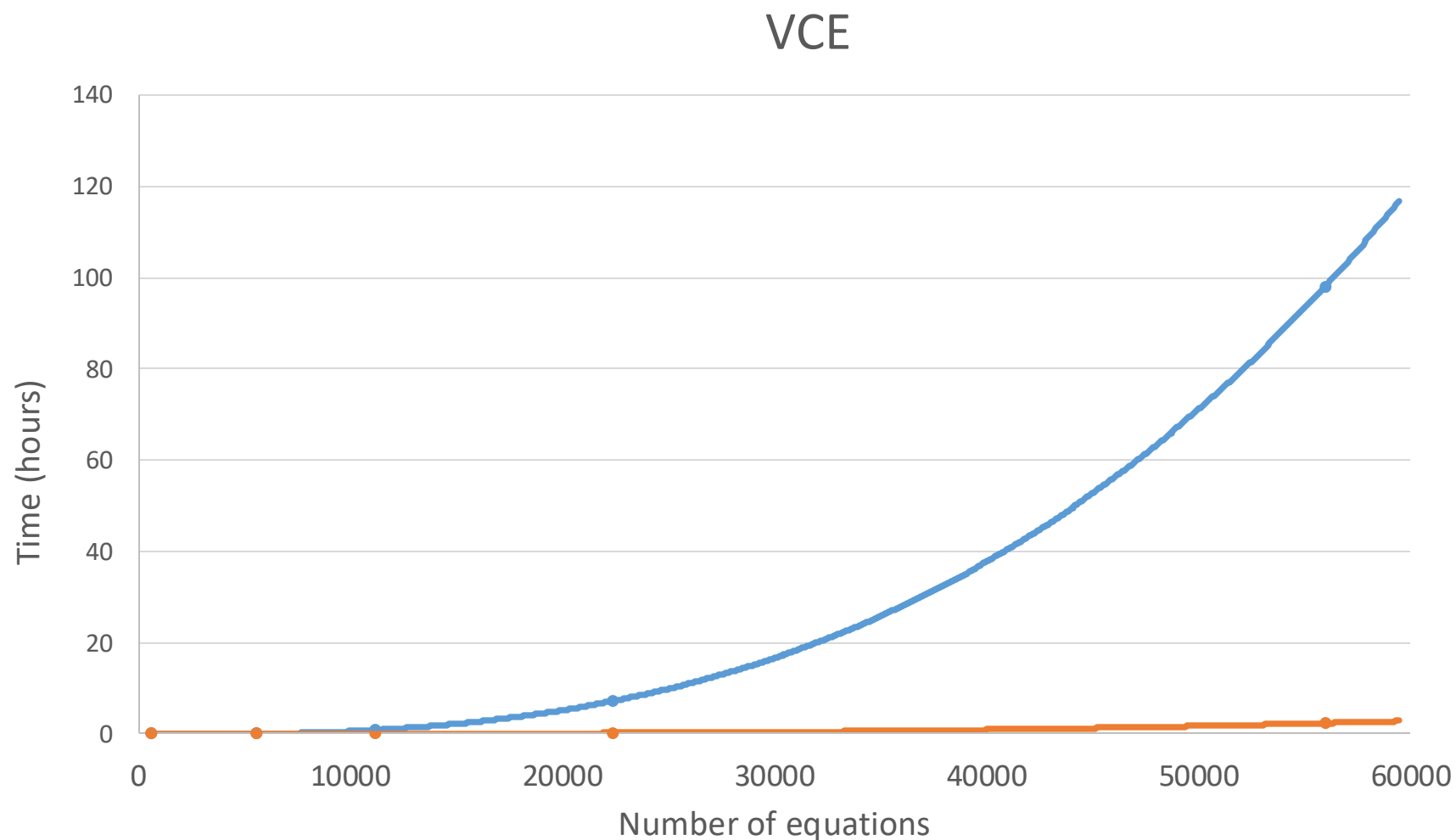
Richter et al.
(2024)



- Selection of females to genotype

Group	SD
Non-genotyped females	1.00
Genotyped females	0.63
Non-genotyped males	1.00
Genotyped males	0.89

Efficient VCE – software optimization



Bermann et al.
(unpublished)

Efficient VCE – New methods

Formulas for estimating heritability



$$\widehat{h^2}: \sqrt{\frac{Nh^2}{Nh^2 + M_e}} = \text{corr}(y - Xb, \hat{u})/h$$

$$\widehat{h^2} = \frac{c^2 + \sqrt{c^4 + 4c^2 M_e/N}}{2}, c = \text{corr}(y - Xb, \hat{u})$$

$$SE(\widehat{h^2}) \approx \frac{1}{\sqrt{N_{\text{val}}}} \left[c + \frac{2c^2 + \frac{4M_e}{N}}{\sqrt{c^2 + \frac{4M_e}{N}}} \right] \widehat{h^2} \approx \frac{3c}{\sqrt{N_{\text{val}}}}$$

N – # animals in reference N_{val} – number of animals in validation

How to estimate genetic correlations?

Predictivity for trait i

$$\text{corr}(y_i - Xb_i, \hat{u}_i) = \text{acc}_i h_i$$

What is predictivity from trait i to trait j ?

$$\text{corr}(y_i - Xb_i, \hat{u}_j) = ?$$

.....

$$\text{corr}(y_i - Xb_i, \hat{u}_j) = \text{acc}_j \text{corr}_{ij} h_i$$

$$\text{corr}_{ij} = \frac{\text{corr}(y_i - Xb_i, \hat{u}_j)}{h_i \text{acc}_j}$$

$$SD(\text{corr}_{ij}) \approx \frac{1}{h_i \text{acc}_j \sqrt{N_{\text{val}}}}$$

UGA AB&G team

