

# Computing strategies for national dairy cattle evaluations

M. Bermann, A. Cesarani, D. Lourenco, I. Misztal

09/2022

2022 EAAP Meeting



UNIVERSITY OF  
**GEORGIA**

College of Agricultural &  
Environmental Sciences

*Animal Breeding and  
Genetics Group*

# Single-step genetic evaluations

- Move to single-step evaluations
- Less biased than multi-step methods
- Flexible for various models:
  - Repeatability
  - Multiple-trait
  - Maternal
  - Threshold
  - Random regression
- Software availability



J. Dairy Sci. 105:5141–5152  
<https://doi.org/10.3168/jds.2021-21505>

© 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,<sup>1\*</sup> D. Lourenco,<sup>1</sup> S. Tsuruta,<sup>1</sup> A. Legarra,<sup>2</sup> E. L. Nicolazzi,<sup>3</sup> P. M. VanRaden,<sup>4</sup> and I. Misztal<sup>1</sup>



J. Dairy Sci. 105:3306–3322  
<https://doi.org/10.3168/jds.2021-21145>

© 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/>).

## Comparison of a single-step with a multistep single nucleotide polymorphism best linear unbiased predictor model for genomic evaluation of conformation traits in German Holsteins

H. Alkhoder, Z. Liu,\* D. Segelke, and R. Reents

ORIGINAL ARTICLE

Journal of  
Animal Breeding and Genetics WILEY

## Single-step genomic evaluation of Russian dairy cattle using internal and external information

Andrei A. Kudinov<sup>1,2,3</sup> | Esa A. Mäntysaari<sup>1</sup> | Timo J. Pitkänen<sup>1</sup> |  
Ekaterina I. Saksa<sup>3</sup> | Gert P. Aamand<sup>4</sup> | Pekka Uimari<sup>2</sup> | Ismo Strandén<sup>1</sup>

# Dairy cattle in the US

**1.09 million** animals genotyped in the last 12 months (10.1.20 to 9.30.21)

**6** annual DHI summaries produced

**4.08M** cows in 11,691 herds enrolled in DHI test plans (2021)



**50 traits** calculated by CDCB

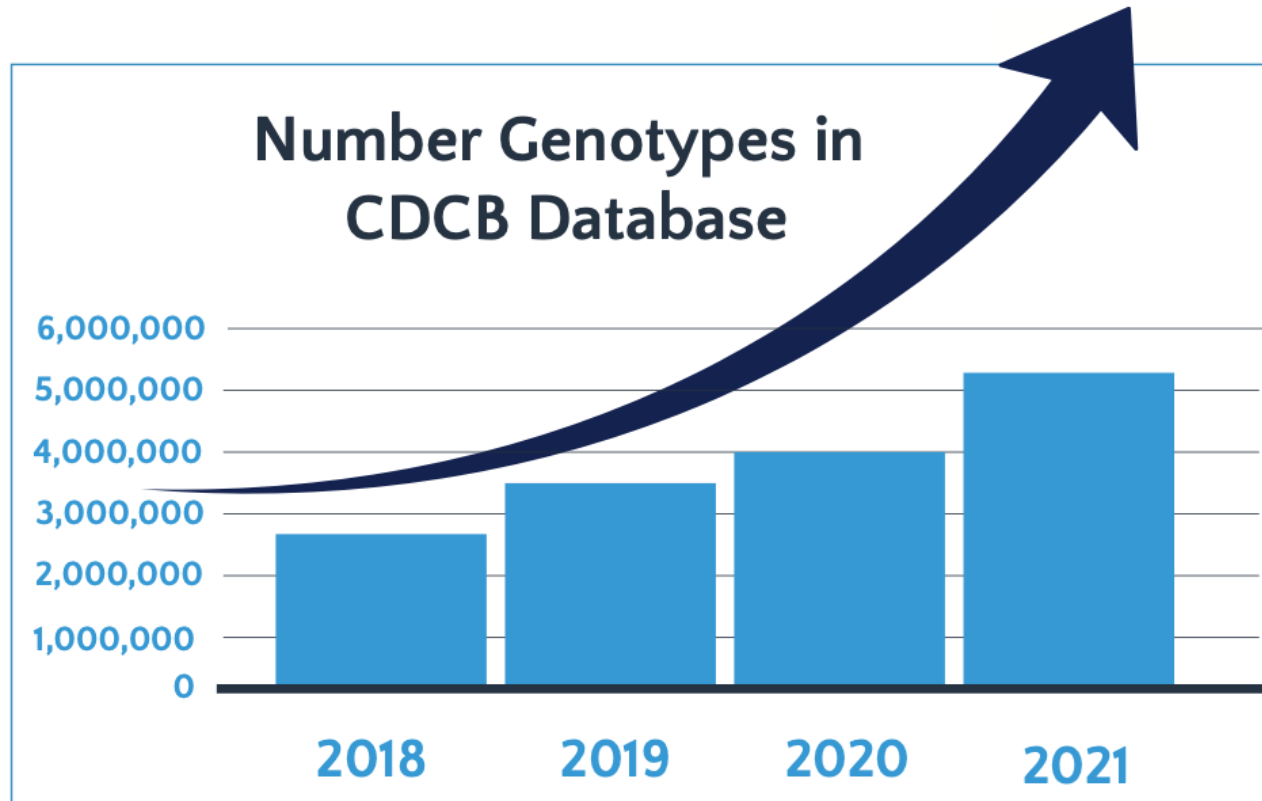
- 4 selection indexes
- 5 production traits
- 21 health, fertility & calving traits
- 22 conformation traits
- 25 official genetic conditions & haplotypes

Weekly genomic predictions for new genotyped animals	Monthly genomic evaluations	Triannual evaluations conventional, genomic & Interbull (in APR, AUG & DEC)
<b>13M</b> Lactation, Calving, Breeding and Health records added for each triannual evaluation	<b>6.7M</b> DHI records in CDCB health evaluations Holstein	<b>885k</b> DHI records in CDCB health evaluations Jersey



ACTIVITY REPORT OCT 2020 / SEP 2021

# Dairy cattle in the US



ACTIVITY REPORT OCT 2020 / SEP 2021

# Single-step genetic evaluations

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \sigma_g^{-2}\mathbf{H}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

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

Pedigree

Genomics

# Single-step with APY

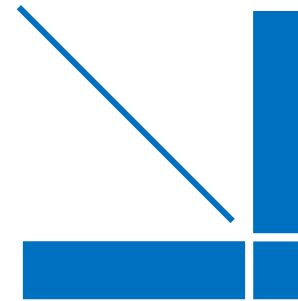
$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{pmatrix}$$



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



$\mathbf{G}^{-1}$



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

# From raw data to EBVs

- Read genotypes
- Quality control
- Set up single-step matrices

Genomic setup

- Read data and pedigree
- Solve the MME
- Approximate accuracies

Breeding value estimation

# Challenges with more genotyped animals

1. Read and store genotypes
2. Set up APY blocks
3. Adding a RPG
4. Increased number of rounds for PCG
5. Calculating accuracies



# Challenges with more genotyped animals

1. Read and store genotypes
2. Set up APY blocks
- 3. Adding a RPG**
4. Increased number of rounds for PCG
5. Calculating accuracies

# Challenges with more genotyped animals

1. Read and store genotypes
2. Set up APY blocks
- 3. Adding a RPG**
4. Increased number of rounds for PCG
- 5. Calculating accuracies**

# Addition of RPG

- Explicit estimation
- GTBLUP
- Adding  $A_{22}$  to  $G$

**Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals**

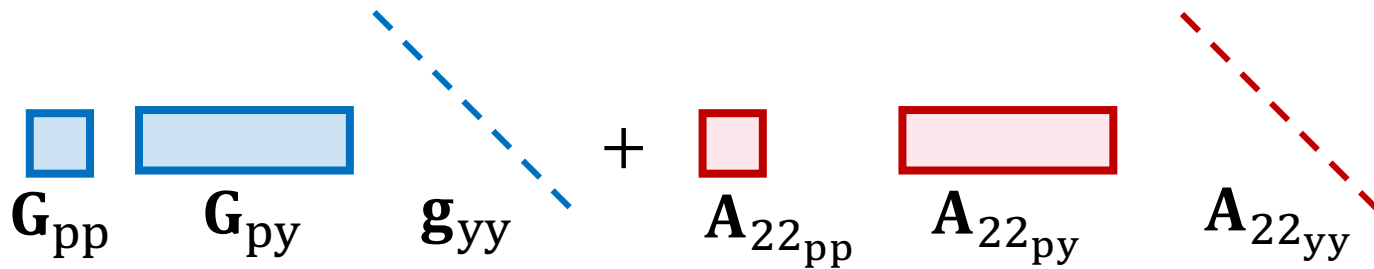
Y. Masuda,<sup>\*1</sup> I. Misztal,<sup>\*</sup> S. Tsuruta,<sup>\*</sup> A. Legarra,<sup>†</sup> I. Aguilar,<sup>‡</sup> D. A. L. Lourenco,<sup>\*</sup> B. O. Fragomeni,<sup>\*</sup> and T. J. Lawlor<sup>§</sup>

<sup>\*</sup>Department of Animal and Dairy Science, University of Georgia, Athens 30602

<sup>†</sup>Institut National de la Recherche Agronomique, UMR1388 GenPhySE, 31326 Castanet Tolosan, France

<sup>‡</sup>Instituto Nacional de Investigación Agropecuaria, Canelones, Uruguay 90200

<sup>§</sup>Holstein Association USA Inc., Brattleboro, VT 05301



# Addition of RPG

## Impact of blending the genomic relationship matrix with different levels of pedigree relationships or the identity matrix on genetic evaluations

Mary Kate Hollifield,\*  Matias Bermann,  Daniela Lourenco,  and Ignacy Misztal 

### Pseudocode for the old blending algorithm.

```
1: Input:  $\mathbf{G}_{cc}, \mathbf{G}_{cn}, \mathbf{g}_n, \alpha, \beta$ .
2: Output: blended  $\mathbf{G}_{cc}, \mathbf{G}_{cn}, \mathbf{g}_n$  and averages of diagonal and off-diagonal elements of  $\mathbf{A}_{22}$ .

3:  $i, j \leftarrow$  integers
4:  $n \leftarrow$  number of genotyped animals
5:  $\alpha, \beta \leftarrow$  blending parameters
6:  $\mathbf{M}(:, \cdot) \leftarrow$  row, column, or element of a matrix  $\mathbf{M}$ 
7:  $Avg_d \leftarrow$  average of the diagonal elements of  $\mathbf{A}_{22}$ 
8:  $Avg_o \leftarrow$  average of the off-diagonal elements of  $\mathbf{A}_{22}$ 
9:  $core(i) \leftarrow$  determines if the  $i^{th}$  animal is a core animal

10: Calculate inbreeding coefficients for genotyped animals and their ancestors
11: for  $i$  in  $1:n$  do
12:   Calculate  $\mathbf{A}_{22}(:, i)$  and Accumulate  $Avg_d$  and  $Avg_o$ 
13:   if  $core(i)$  then
14:     for  $j$  in  $1:n$  do
15:       if  $core(j)$  then
16:          $\mathbf{G}_{cc}(j, i) \leftarrow \alpha \mathbf{G}_{cc}(j, i) + \beta \mathbf{A}_{22}(j, i)$ 
17:       end if
18:     end for
19:   else
20:     for  $j$  in  $1:n$  do
21:       if  $core(j)$  then
22:          $\mathbf{G}_{cn}(j, i) \leftarrow \alpha \mathbf{G}_{cn}(j, i) + \beta \mathbf{A}_{22}(j, i)$ 
23:       else if  $i = j$  then
24:          $\mathbf{g}_n(i) \leftarrow \alpha \mathbf{g}_n(i) + \beta \mathbf{A}_{22}(i, i)$ 
25:       end if
26:     end for
27:   end if
28: end for
```



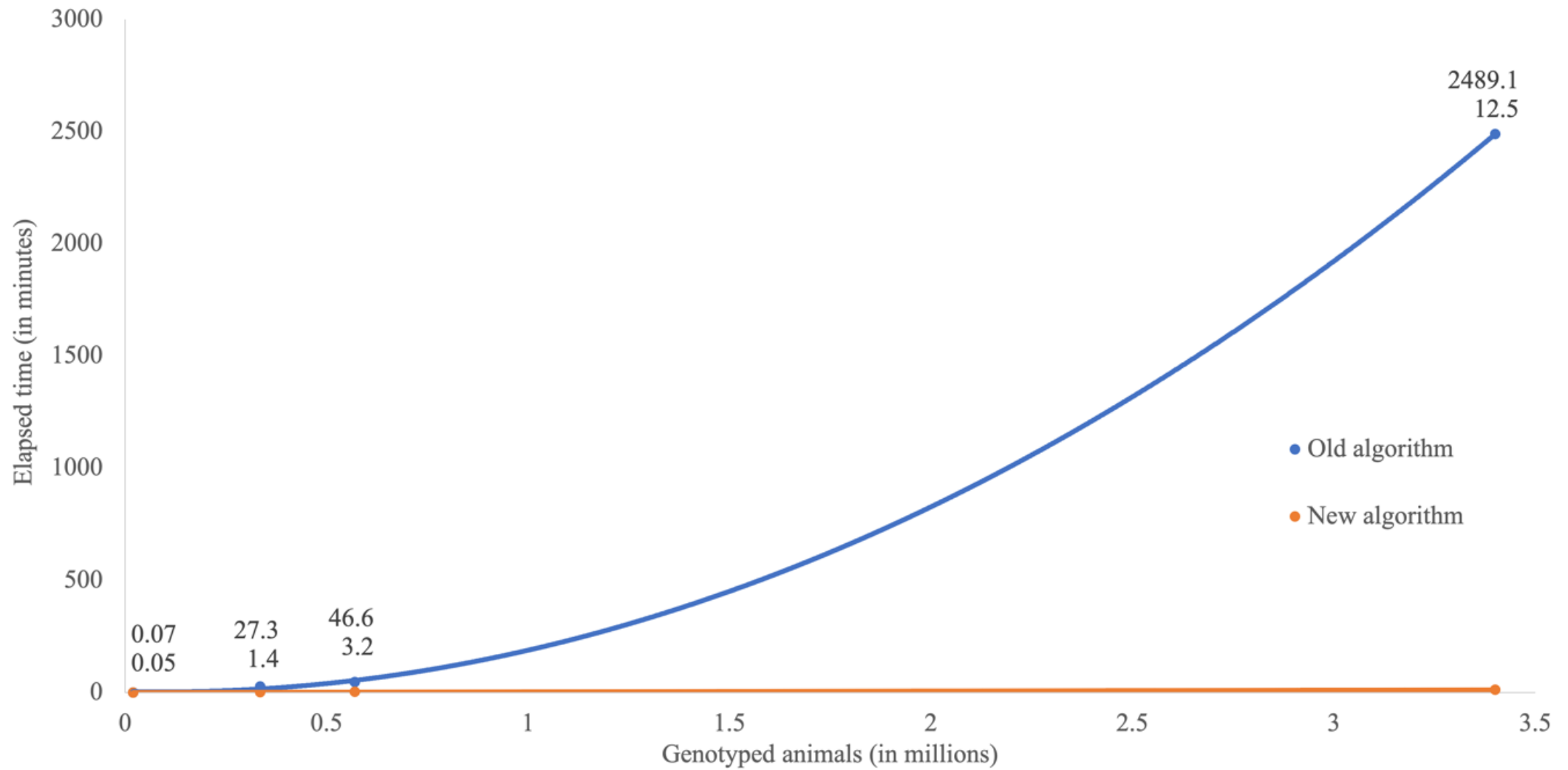
### Pseudocode for the new blending algorithm.

```
1: Input:  $\mathbf{G}_{cc}, \mathbf{G}_{cn}, \mathbf{g}_n, \alpha, \beta$ .
2: Output: blended  $\mathbf{G}_{cc}, \mathbf{G}_{cn}, \mathbf{g}_n$  and averages of diagonal and off-diagonal elements of  $\mathbf{A}_{22}$ .

3:  $i, j \leftarrow$  integers
4:  $n, n_c, n_n \leftarrow$  number of genotyped animals, core, and non-core animals
5:  $\mathbf{f}, \mathbf{f}_c, \mathbf{f}_n \leftarrow$  vector of inbreeding for all, core, and non-core animals
6:  $\mathbf{1} \leftarrow$  vector of ones
7:  $\alpha, \beta \leftarrow$  blending parameters
8:  $\mathbf{M}(:, \cdot) \leftarrow$  row, column, or element of a matrix  $\mathbf{M}$ 
9:  $Avg_d \leftarrow$  average of the diagonal elements of  $\mathbf{A}_{22}$ 
10:  $Avg_o \leftarrow$  average of the off-diagonal elements of  $\mathbf{A}_{22}$ 
11:  $core(i) \leftarrow$  determines if the  $i^{th}$  animal is a core animal

12: Calculate inbreeding coefficients for genotyped animals and their ancestors
13: for  $i$  in  $1:n_c$  do
14:   Calculate  $\mathbf{A}_{22}(:, i)$  and permute it
15:    $\mathbf{G}_{cc}(i, \cdot) \leftarrow \alpha \mathbf{G}_{cc}(i, \cdot) + \beta \mathbf{A}_{22}(i, 1 : n_c)$ 
16:    $\mathbf{G}_{cn}(i, \cdot) \leftarrow \alpha \mathbf{G}_{cn}(i, \cdot) + \beta \mathbf{A}_{22}(i, 1 + n_c : n)$ 
17: end for
18:  $\mathbf{g}_n \leftarrow \alpha \mathbf{g}_n + \beta (\mathbf{1} + \mathbf{f}_n)$ 
19:  $Avg_d \leftarrow 1 + n^{-1} \sum_{i=1}^n \mathbf{f}_i$ 
20:  $Avg_o \leftarrow (\mathbf{1}' \mathbf{A}_{22} \mathbf{1} - n Avg_d) / (n^2 - n)$ 
```

Time comparison between the old and new blending algorithms.



# Challenges with more genotyped animals

1. Read and store genotypes
2. Set up APY blocks
3. Adding a RPG
4. Increased number of rounds for PCG
- 5. Calculating accuracies**

# Calculating accuracies

- Combine different sources of information
  - Records
  - Pedigree
  - Genomics
- Computationally efficient for any model

## **Approximating Genomic Reliabilities for National Genomic Evaluation**

*Z. Liu<sup>1</sup>, P. M. VanRaden<sup>2</sup>, M.H. Lidauer<sup>3</sup>, M. P. Calus<sup>4</sup>, H. Benhajali<sup>5</sup>, H. Jorjani<sup>5</sup>  
and V. Ducrocq<sup>6</sup>*

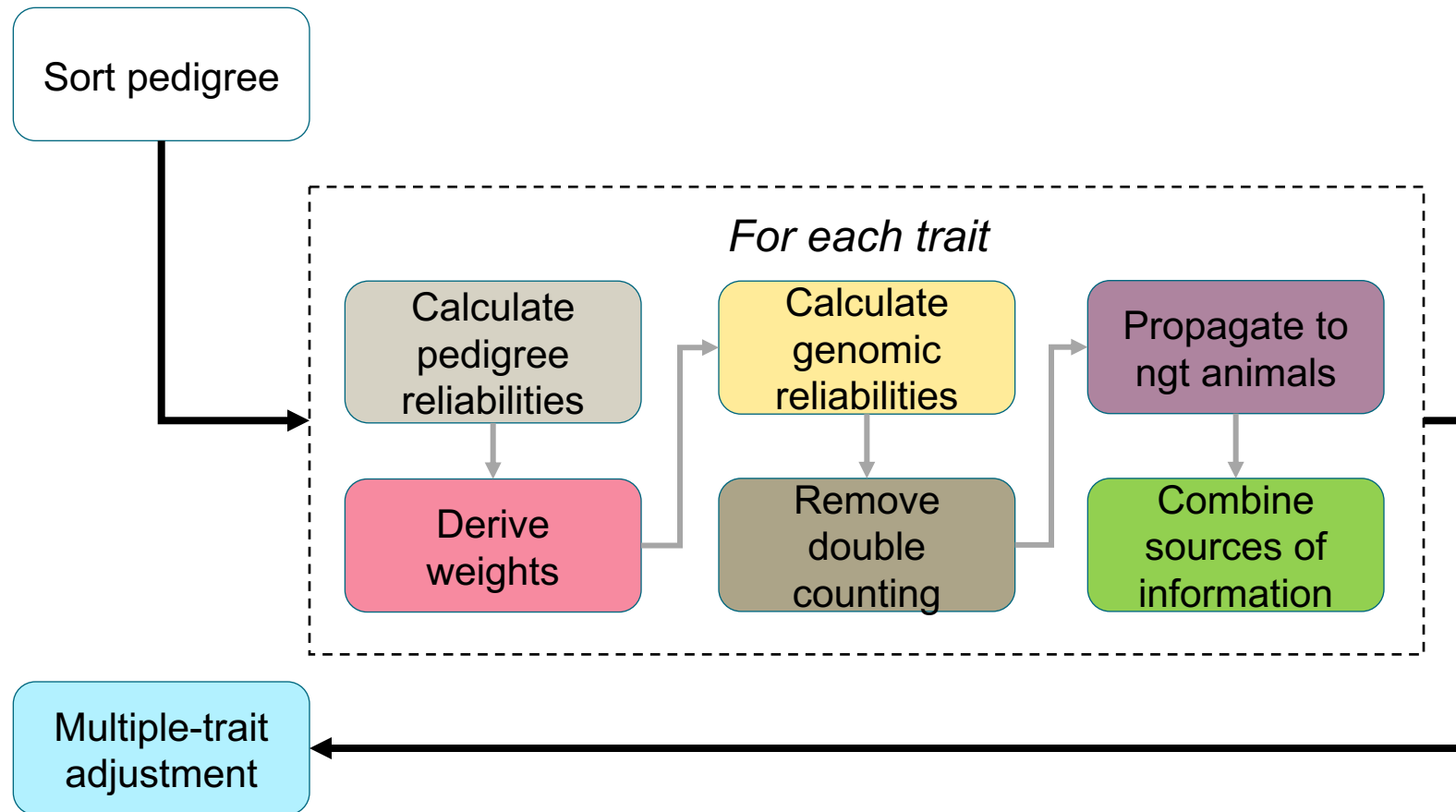
## **Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the Algorithm for Proven and Young**

**Matias Bermann,<sup>1</sup> Daniela Lourenco, and Ignacy Misztal**

Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA

<sup>1</sup>Corresponding author: [mbermann@uga.edu](mailto:mbermann@uga.edu)

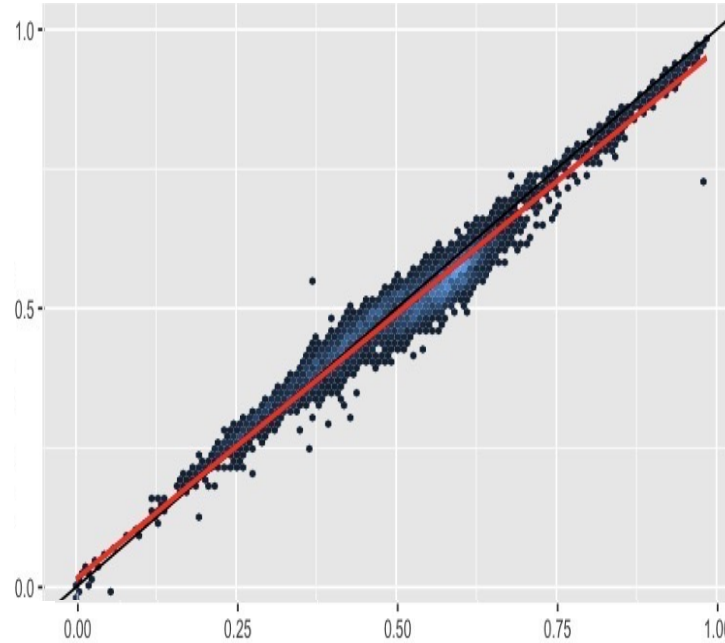
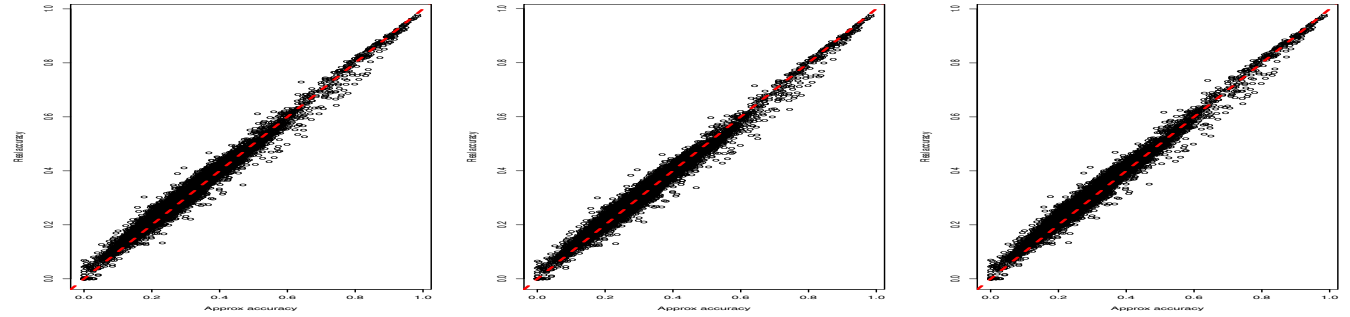
# ACCF90GS2 flowchart





# ACCFG2 in dairy cattle

- 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: more evaluations and information

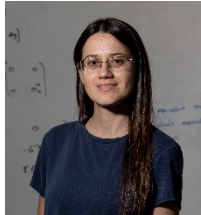
- Convergence
  - Seconds per round
    - Code
    - Mathematical operations
  - Number of rounds
    - Preconditioning
    - Removing old data
- Complicated models
  - Multiple-trait
  - Threshold
  - Multi/cross breed
- New information
- Variance components



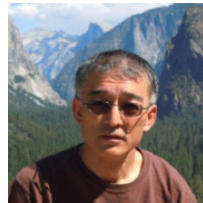
# Conclusions

- Different computing strategies allow weekly single-step evaluations
- There is room for improving computational efficiency
- Anticipate next challenges and bottlenecks

# The UGA team



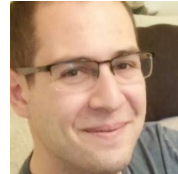
Daniela  
Lourenco



Shogo  
Tsuruta



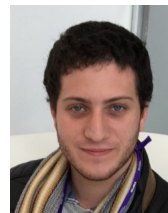
Ignacy  
Misztal



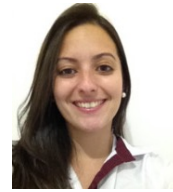
Alberto  
Cesarani



Jorge  
Hidalgo



Matias  
Bermann



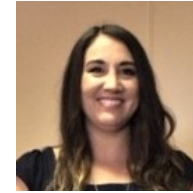
Natalia  
Galoro



Sungbong  
Jang



Jennifer  
Richter



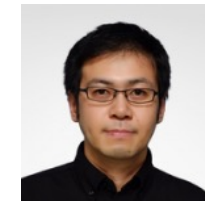
Taylor  
Mcwhorther



Mary Kate  
Hollifield



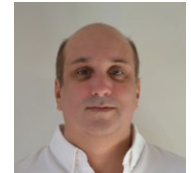
Yvette  
Steyn



Yutaka  
Masuda



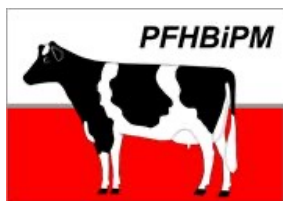
Andres  
Legarra



Ignacio  
Aguilar



Zulma  
Vitezica



# ***Questions?***

*mbermann@uga.edu*

