

How ssGBLUP became suitable for national dairy cattle evaluations

Ignacy Misztal, D. Lourenco, S. Tsuruta, M. Bermann

University of Georgia, USA

Alberto Cesarani, **University of Sassari, Sardinia**

A. Legarra, **INRA, France**

Y. Masuda, **Rakuno University, Japan**

I. Aguilar, **INIA, Uruguay**

Genomic developments in dairy

- Multistep method (VanRaden, 2008) and software
 - Only option when ownership of phenotypes and genotypes separate
- Bias through genomic preselection (Patry and Ducrocq, 2011, Masuda et al., 2017)
- Potential double counting with male and female genotypes (Wiggans et al., 2012)
- Interbull and single step
 - 2017 – Workshop
 - 2018 - Implementations in Czechia and Belgium
 - 2022- test implementations in many countries – superior reliability

Why ssGBLUP not yet in dairy

- Free multistep software giving pretty good reliabilities
- Unavailability of free ssGBLUP software
- Desire for national solutions

- Biases and high computing costs

ssGBLUP in dairy at UGA

Trait	# animals	# genotypes	$b_1 * 100$	Validation	Reference
HO stature	7M	6k	68-82	DD 2009 / GEBV 2005	Aguilar et al., 2010
HO 18 type traits	7M	6k	53-93	DD 2009 / GEBV 2005	Tsuruta et al., 2011
HO stature	10M	570k	85	DD 2014/ GEBV 2009	Masuda et al., 2015
					Bradford et al
HO MFP	18M-30M	861K	0.97-1.03	DD2018 / GEBV 2014	Cesarani et al., 2021
MFP 5 breeds	19M	3.8M	97-103 HOL 92-106 JER	DD2020 GEBV 2016	Cesarani et al., 2022

Initial bias and convergence problems

- Partial fix by:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{0.7A}_{22}^{-1} \end{bmatrix}$$

- Why:
 - No inbreeding in A
 - Missing pedigrees

Pedigree length, scaling and convergence

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

$$\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$$

Good convergence and genotyped animals biased down

$$\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$$

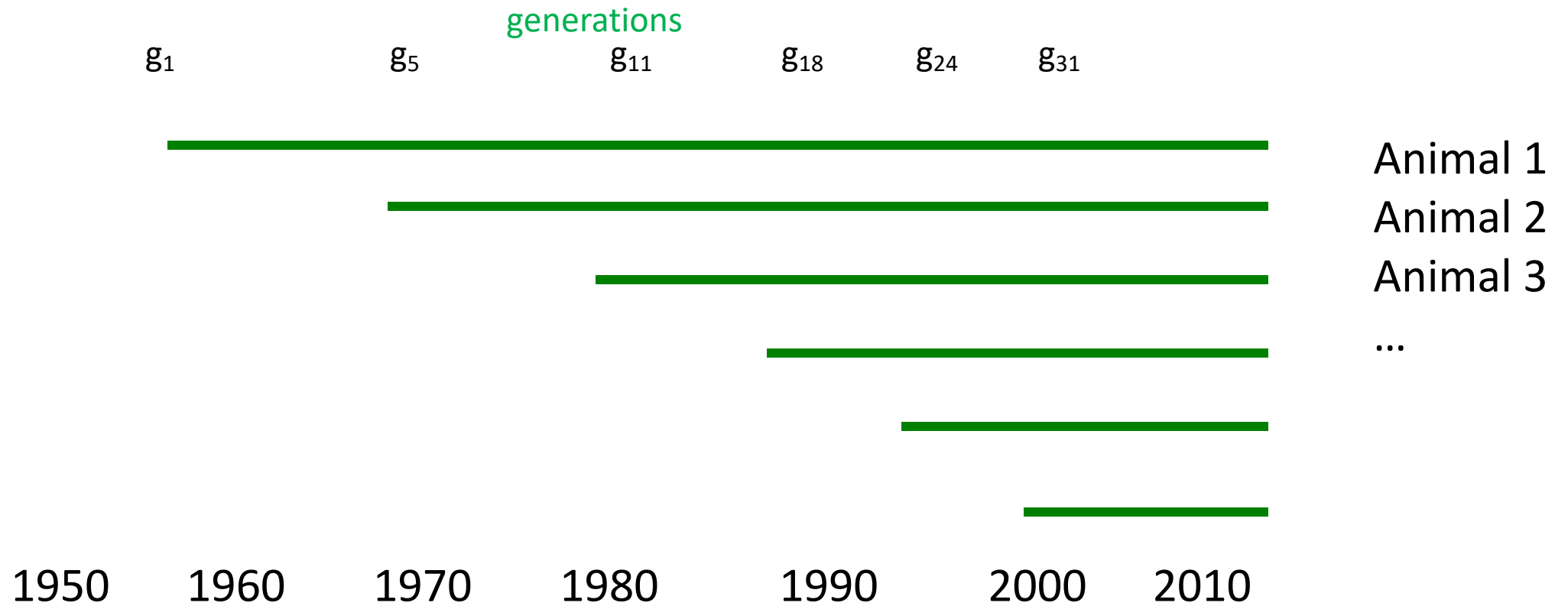
Bad convergence and genotyped animals biased up

$$\mathbf{G}^{-1} - \mathbf{A}_{22,1}^{-1} - \mathbf{A}_{22,2}^{-1} - \mathbf{A}_{22,3}^{-1}$$

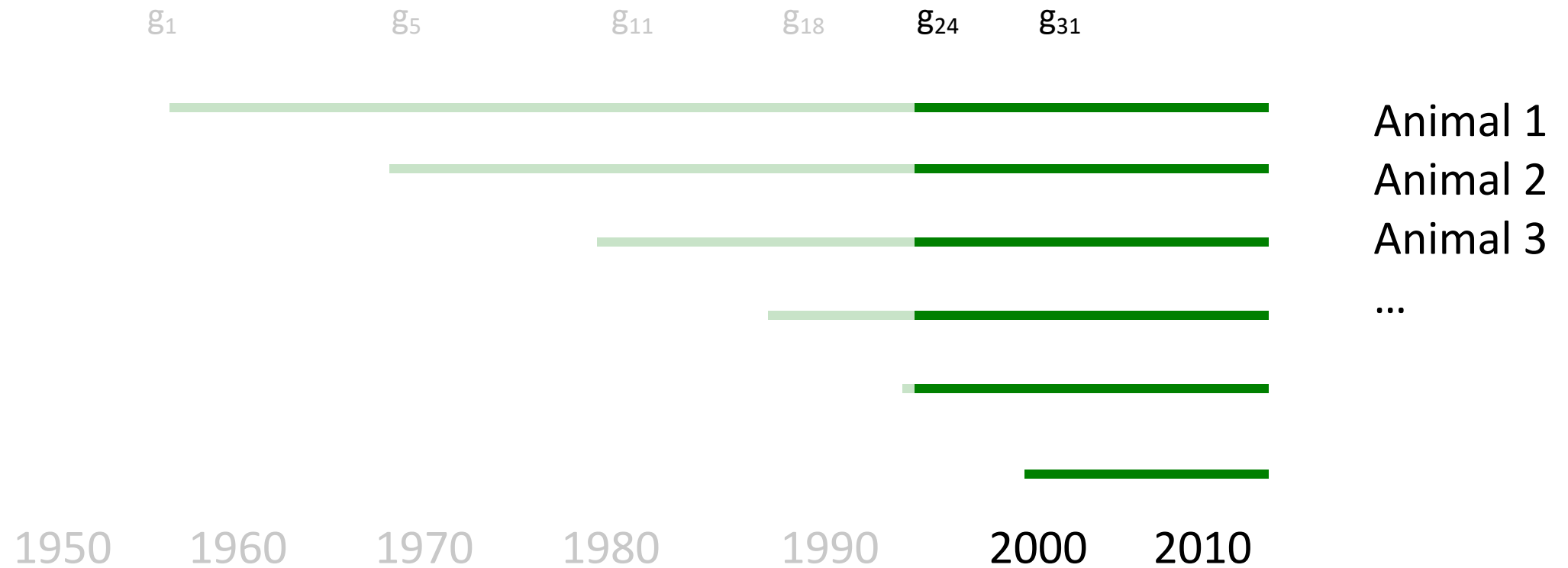
Long medium short
pedigrees

Bad convergence and genotyped animals biased down and up

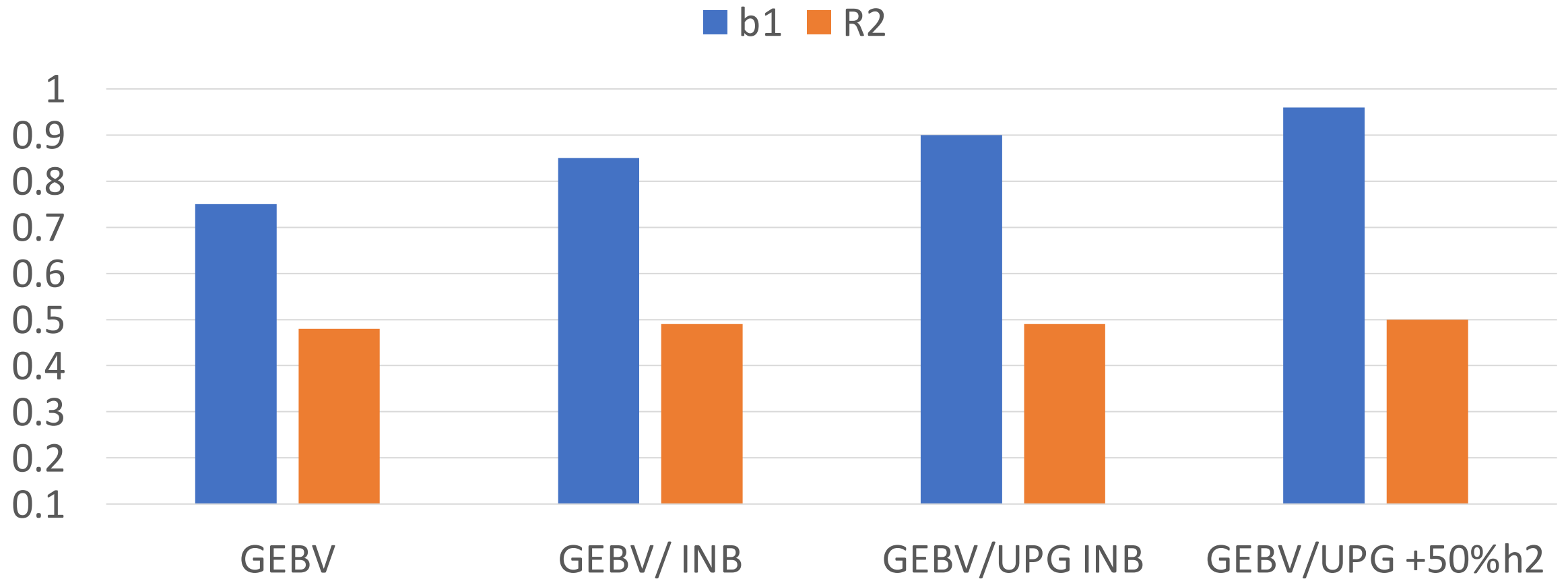
Pedigree depth for young animals



Cut pedigree and data?



Steps to increase b_1



UPG INB – inbreeding for unknown parents

Unknown parent group formulas

- $$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix}$$

QP-transformation for \mathbf{A}^{-1} (Westell et al., 1988)

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

QP-transformation for \mathbf{H}^{-1} (Misztal et al., 2013)

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

Altered QP-transformation for \mathbf{H}^{-1} (Tsuruta et al., 2019)
 Small number of groups (< 10)

- $$\mathbf{H}_F^{-1} = \mathbf{A}_F^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{05}^{-1} - \mathbf{A}_{F22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Metafounders (Legarra et al., 2015)

Large number of genotypes

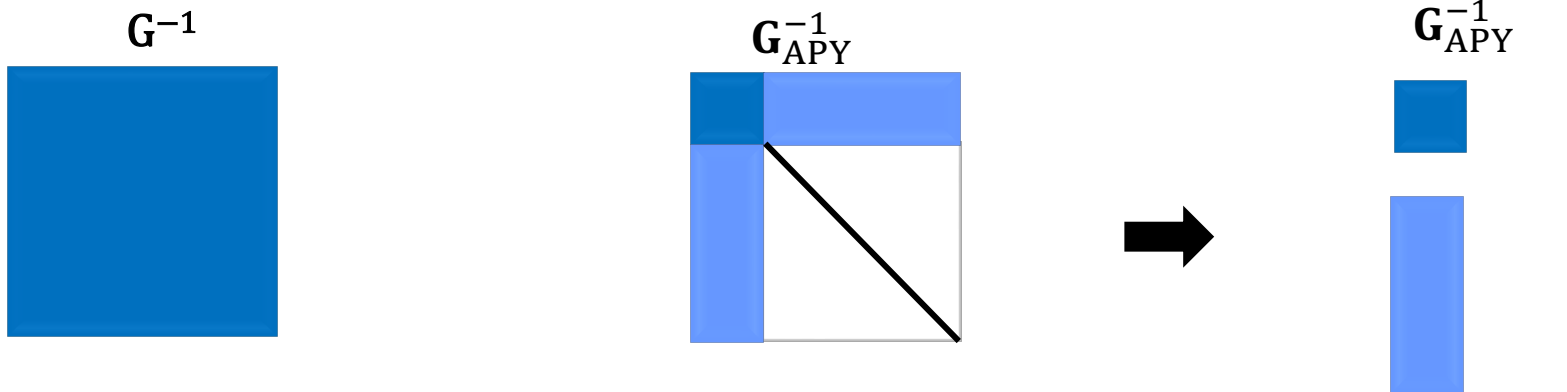
- Rank of G limited : $< \# \text{ animals}, < \# \text{ SNP}, < \# \text{ chromosome segments}$
- Approaches:
 - SNP for genotyped animals
 - SNP for all animals
 - inversion of G based on rank of SNP (Woodbury)
 - inversion of G based on rank of independent chromosome segments (APY)
- Number of chromosome segments $\sim 15\text{k}$ cattle,

APY algorithm

- Selected animals (=number of segments) = core animals in APY
- Other animals (remainder) = noncore animals in APY

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

- Genomic relationships in APY $u_i | u_{c1} + u_{c2} + u_{c3}, \dots, u_{ci} = \sum_{j=1}^c p_{ij} u_j + \varepsilon_i$



- G_{APY}^{-1} sparse
- Efficient computations

Masuda et al. (2016)

Easy implementation in existing software
Easy implementation of scaling and UPG
Convergence rate similar to BLUP

What data truncation?

- Data cutoff 1980, 1990, 2000
- Pedigree cutoff 2, 3 generations past phenotypes
- Same R^2 , $b_1=1\pm 0.05$
- If old bulls need to be included, include only genotypes



J. Dairy Sci. 104

<https://doi.org/10.3168/jds.2020-19789>

© 2021, 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-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Genomic predictions for yield traits in US Holsteins with unknown parent groups

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

¹Department of Animal and Dairy Science, University of Georgia, Athens 30602

²Council on Dairy Cattle Breeding, Bowie, MD 20716

³Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

Value of more data

	Data up to 2018	Data up to 2020
Genotypes	862k	3.4M + data
Validation bulls	1,529	1,529
Reliability		
Milk	0.72	0.81
Fat	0.69	0.80
Protein	0.67	0.77
b1		
Milk	1.06	1.01
Fat	1.02	1.00
Protein	1.05	1.01
Computing cost		
Genomic	6 hours	27 h
Rounds	404	473

Should all genotypes be used?

- Use only contributing genotypes
 - Bulls + bulldams + cows with phenotypes
- Evaluations of other animals indirectly
- Truncate old genotypes
- If high computing cost, reduce # core animals in APY
 - 0.01 reliability penalty for 2 times less computing

Steps for unbiased evaluation in dairy with US national data

- Inbreeding in A and A_{22}
 - Appropriate unknown parent group formulas, few groups
 - APY for inversion of GRM
 - Truncation of phenotypes and pedigrees
 - Larger reference data
 - Efficient computing algorithms
-
- Accomplished at UGA in 12 years