

Computing strategies for national dairy cattle evaluations

M. Bermann, D. Lourenco, I. Misztal

06/2022

2022 ADSA National 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

Dairy cattle in the US

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



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)

13M Lactation, Calving, Breeding and Health records added for each triannual evaluation

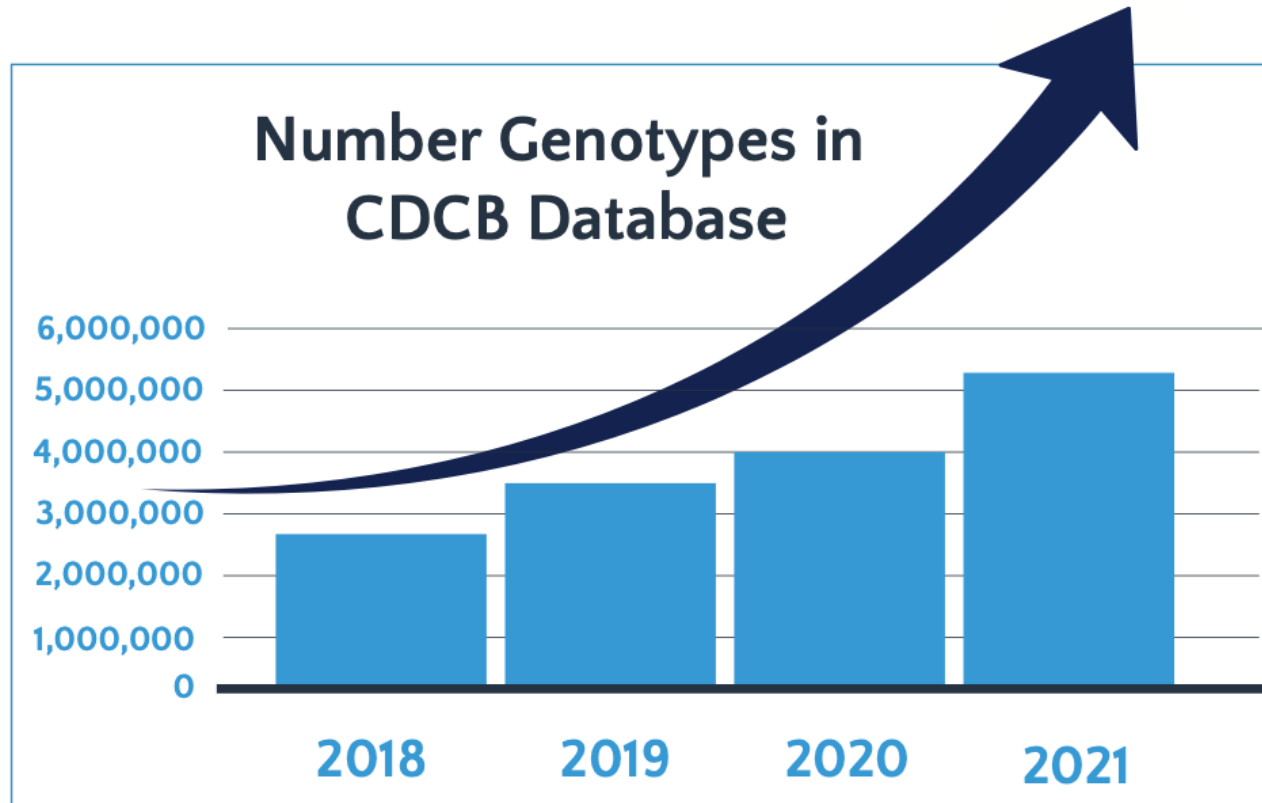
6.7M DHI records in CDCB health evaluations Holstein

885k 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. Blending the GRM
4. Increased number of rounds for PCG
5. Calculating accuracy

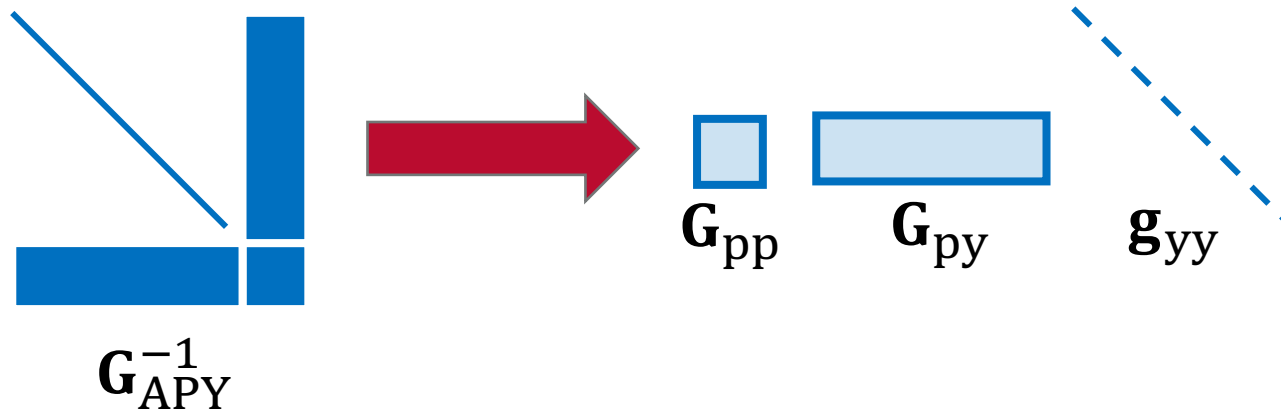
Genomic setup: read and store genotypes

Genotype	Int representation	4-byte integer	2-bit integer
AA	2	0000 0000 0000 0000 0000 0000 0000 0010	10
Aa	1	0000 0000 0000 0000 0000 0000 0000 0001	01
aa	0	0000 0000 0000 0000 0000 0000 0000 0000	00
Missing	5	0000 0000 0000 0000 0000 0000 0000 0101	11

- 1 million 50k genotypes
 - 200 gb -> 12.5 gb
 - ~2-5 times faster

The word 'zlib' is rendered in a 3D, green, blocky font with a slight shadow, giving it a three-dimensional appearance.The word 'gzip' is rendered in a 3D, blue, cursive font with a slight shadow, giving it a three-dimensional appearance.

Genomic setup: APY blocks



- Dense matrices
- Optimized libraries
- Parallel computing

Intel® oneAPI Math Kernel Library

OpenMP®

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

Y. Masuda,^{*1} I. Misztal,^{*} S. Tsuruta,^{*} A. Legarra,[†] I. Aguilar,[‡] D. A. L. Lourenco,^{*} B. O. Fragomeni,^{*} and T. J. Lawlor[§]

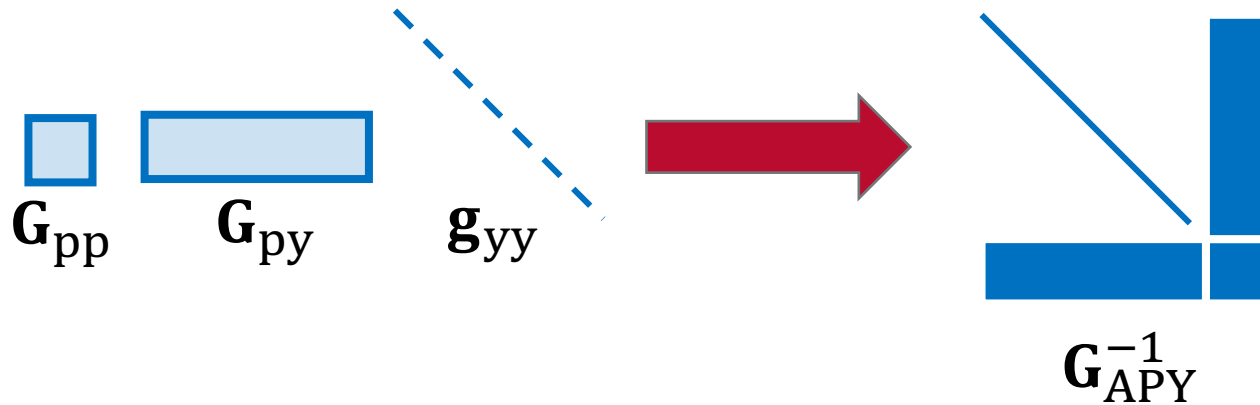
^{*}Department of Animal and Dairy Science, University of Georgia, Athens 30602

[†]Institut National de la Recherche Agronomique, UMR1388 GenPhySE, 31326 Castanet Tolosan, France

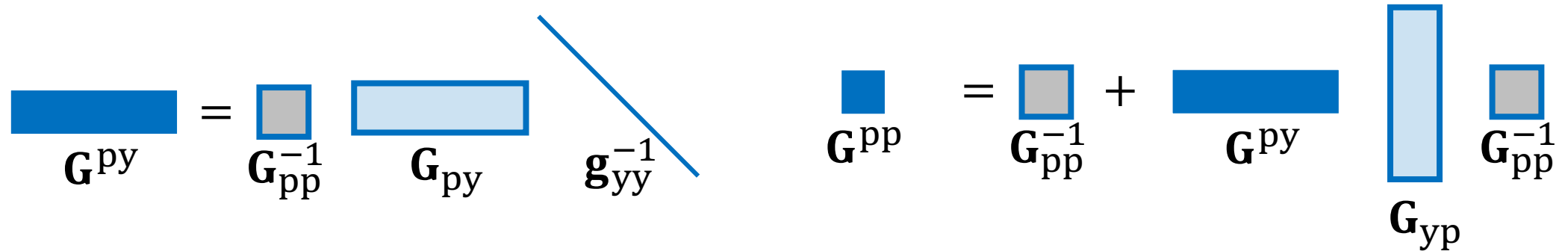
[‡]Instituto Nacional de Investigación Agropecuaria, Canelones, Uruguay 90200

[§]Holstein Association USA Inc., Brattleboro, VT 05301

Genomic setup: creating G_{APY}^{-1}



- Optimized libraries
- Parallel computing

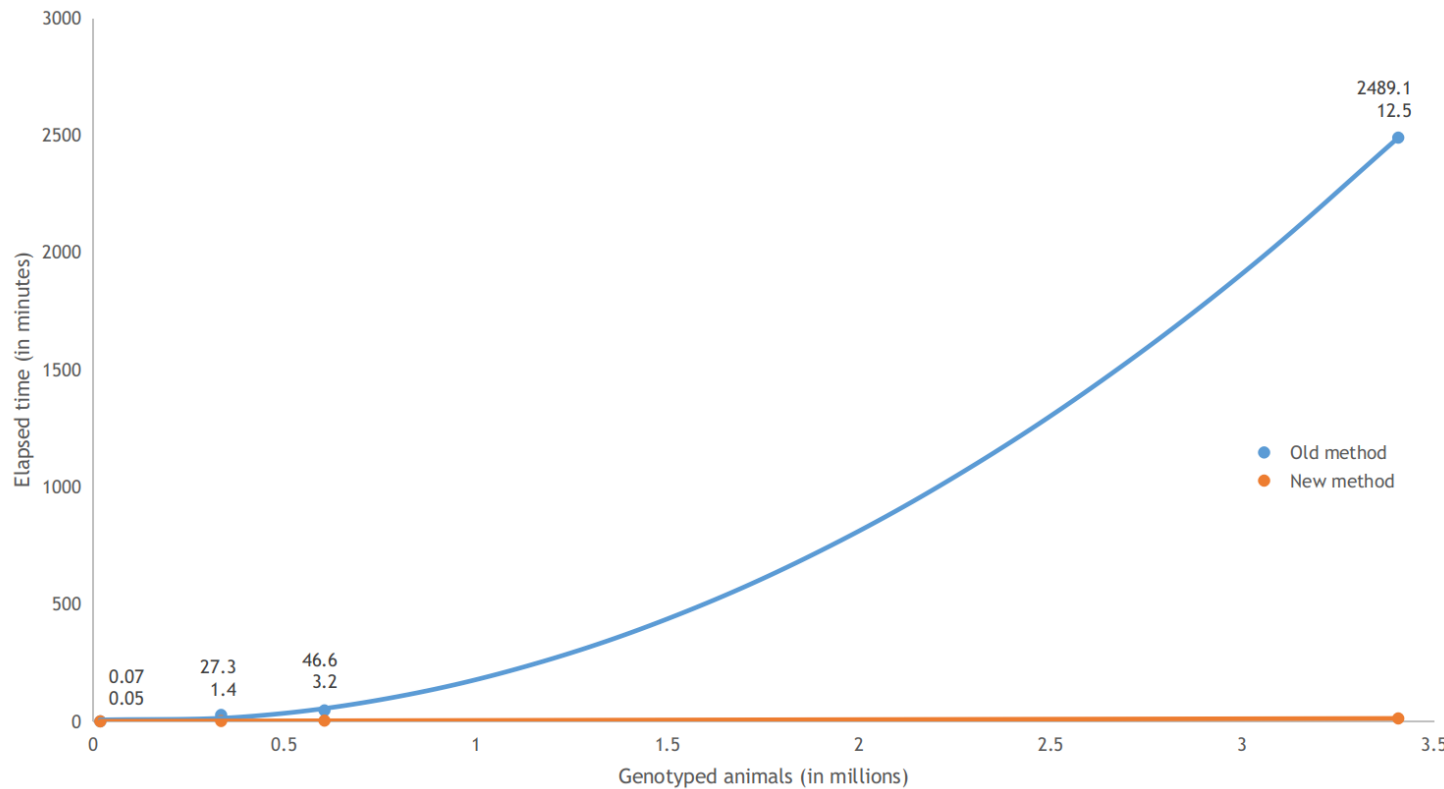


Genomic setup: blending

- Make GRM invertible and add a polygenic effect

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

Elapsed time in minutes for old and new method

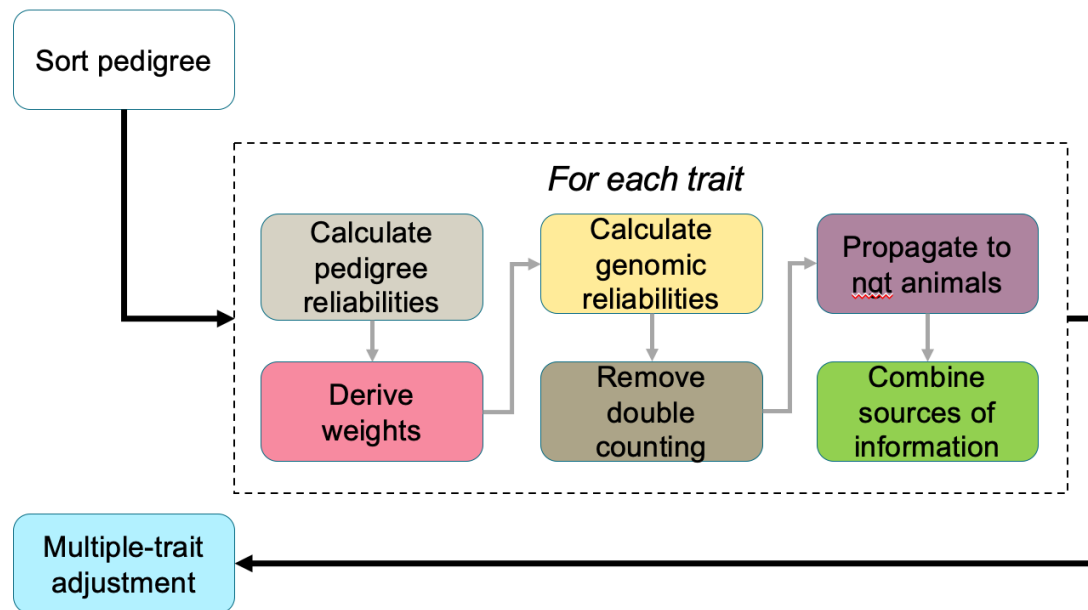


Impact of blending the genomic relationship matrix with different levels of pedigree relationships or the identity matrix on genetic evaluations

Journal:	<i>JDS Communications</i>
Manuscript ID:	JDSC.2022-0229.R2
Article Type:	Short Communication
Date Submitted by the Author:	06-Jun-2022
Complete List of Authors:	Hollifield, Mary; University of Georgia, Animal and Dairy Science Bermann, Matias; University of Georgia, Department of Animal and Dairy Science Lourenco, Daniela; University of Georgia, Animal and Dairy Science Misztal, Ignacy; University of Georgia
Key Words:	large-scale genetic evaluations, prediction accuracy, residual polygenic effect, ssGBLUP

BV estimation: approximation of accuracies

- Combine different sources of information
 - Records
 - Pedigree
 - Genomics



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

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 Misztal¹

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

¹Corresponding author: mbermann@uga.edu

Applications

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. Nicolazzi,³ P. M. VanRaden,⁴ and I. Misztal¹

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

²INRA, UMR1388 GenPhySE, Castanet-Tolosan, France, 31320

³Council on Dairy Cattle Breeding, Bowie, MD 20716

⁴Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705

Table 7. Computational costs for all considered scenarios in terms of number of rounds and seconds per round needed to reach a convergence criterion of 10^{-15}

Item ¹	Rounds		s/round		Time ^{2,3}	
	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
AY_BS_GU	592	1,534	3.01	0.81	~3 min	~20 min
AY_BS_GU_30k		1,581		4.11		~1.4 h
AY_BS_GU_direct		1,529		2.22		~50 min
5_BREEDS	643	1,142	27.01	64.84	4.8 h	~20 h
5_BREEDS_45k		1,763		130.68		~64 h
HO_JE	479	1,403	40.57	81.71	5.4 h	~32 h

¹AY = Ayrshire; BS = Brown Swiss; GU = Guernsey; HO = Holstein; JE = Jersey; 5_BREEDS = all 5 breeds together.

²Time to reach the convergence (i.e., only time for the preconditioning conjugate gradient with iteration on data). ssGBLUP = single-step genomic BLUP.

³Computations were carried out on a Linux server (x86_64) equipped with Intel Xeon E5-2683 v4 2.10 GHz processor with 32 cores. Parallel threads used between 12 and 20 cores.

Challenges: more evaluations and information

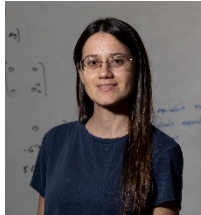
- Convergence
 - Seconds per round
 - Code
 - Mathematical operations
 - Number of rounds
 - Preconditioning
 - Removing old data
- Complicated models
 - Multiple-trait
 - Threshold
 - Multi/cross breed
- New information
- Variance components



Conclusions

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

The UGA team



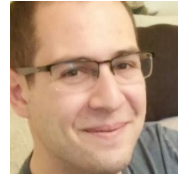
Daniela
Lourenco



Shogo
Tsuruta



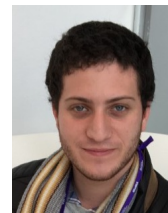
Ignacy
Misztal



Alberto
Cesarani



Jorge
Hidalgo



Matias
Bermann



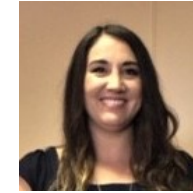
Natalia
Galoro



Sungbong
Jang



Jennifer
Richter



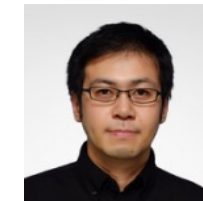
Taylor
Mcwhorther



Mary Kate
Hollifield



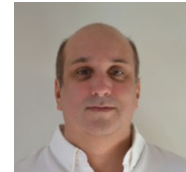
Yvette
Steyn



Yutaka
Masuda



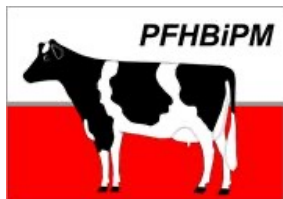
Andres
Legarra



Ignacio
Aguilar



Zulma
Vitezica



Questions?

mbermann@uga.edu

