

# Single-step GBLUP using APY inverse for protein yield in US Holstein with a large number of genotyped animals

Yutaka Masuda, Ignacy Misztal, Paul VanRaden, and Tom Lawlor

University of Georgia, USDA AGIL, and Holstein Association USA Inc.

# Acknowledgement

- U.S. Department of Agriculture's National Institute of Food and Agriculture (Agriculture and Food Research Initiative competitive grant 2015-67015-22936)
- George Wiggans, Melvin Tooker, and John Cole (Animal Genomics and Improvement Laboratory )
- Council on Dairy Cattle Breeding



# Background

- Desirable features in single-step GBLUP (ssGBLUP)
  - Simplicity, avoidance of double counting, and accountability of pre-selection bias
- Breakthroughs in computing difficulties
  - Simple  $\mathbf{G}^{-1}$  with the “APY” algorithm ( $\mathbf{G}_{APY}^{-1}$ )
  - Avoidance of direct computation of  $\mathbf{A}_{22}^{-1}$
- Large-scale application
  - How does it work for production traits for Holsteins?
  - How do we improve the reliability & bias in genomic predictions?

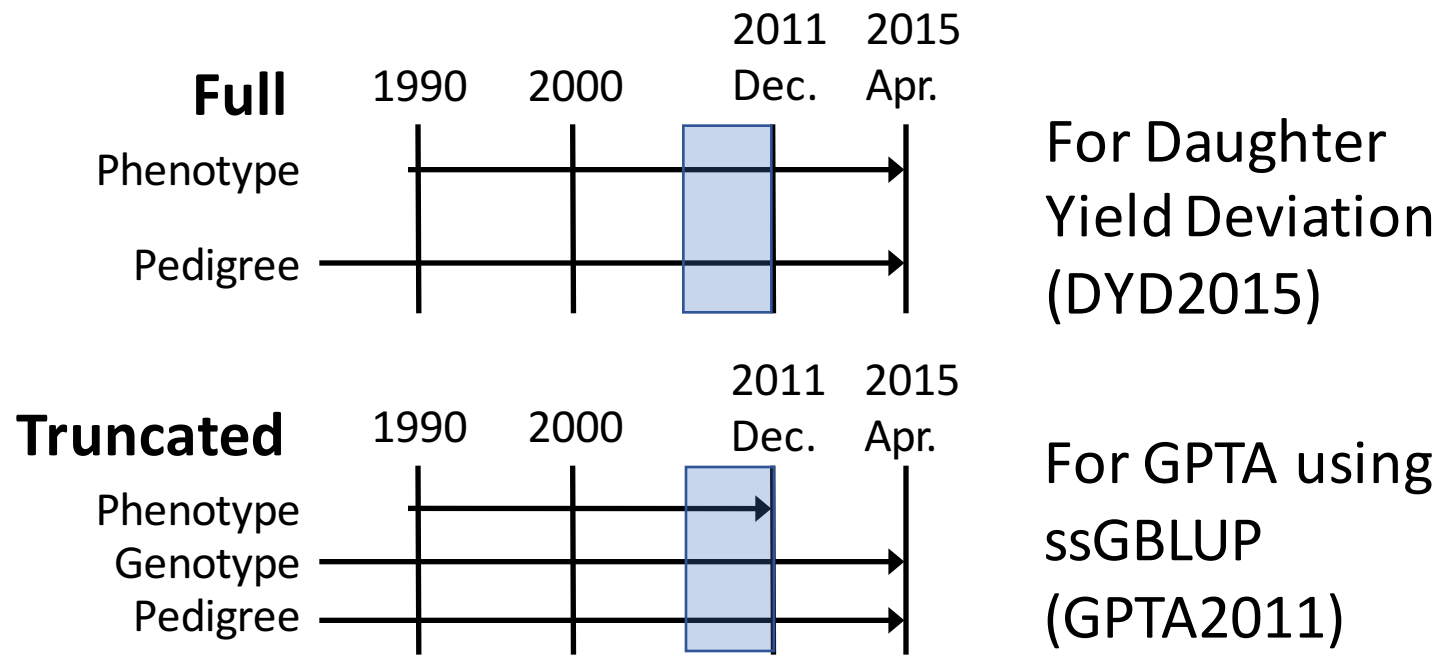
# Objectives

- To validate genomic predictions for young Holstein bulls
  - 305-d protein yield
  - Single-step GBLUP with  $\mathbf{G}_{APY}^{-1}$
  - Comparison with the official GPTA from a multi-step method
- To show some key factors for the greater reliability & less bias in ssGBLUP
  - Limitation of genotyped animals
  - Number of “core” animals
  - Incorporation of QTL effects

# Full data

	Description	Number of records/animals
Phenotype	Protein yield (305-d basis) for US Holstein cows recorded between Jan. 1990 and Apr. 2015	37,259,427
	Cows with phenotype(s)	15,891,366
Pedigree	Animals born in Apr. 2015 or earlier (3-gen. back from phenotyped cows)	22,963,255
Genotype	Animals born in Apr. 2015 or earlier	764,029
	SNP loci	60,671

# Validation study



Validation Bulls:  
Genotyped young bulls  
with no tested daughters  
in 2011 but with at least  
50 tested daughters in  
2015 (N=3,797)

$$DYD2015 = b_1 \times GPTA2011 + b_0$$

- $R^2$  value: validation reliability
- Slope ( $b_1$ ): Bias of prediction

# Model

- Single-trait repeatability model
  - Similar to the official model ( $h^2 = 0.20$  and  $rep. = 0.55$ )

- Single-step GBLUP

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

- “Core” and “noncore” animals in  $\mathbf{G}_{APY}^{-1}$ :

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

$\mathbf{M}^{-1}$ : diagonal matrix

# Genotyped animals

	Description	# of genotyped animals	# of core animals
All (Core 20K)	All available genotypes	764,029	20,000¶
All (Core 13K)	All available genotypes	764,029	12,913*
Higher-density	Animals with more than 38K SNP markers with known parents	149,941	12,882*
Bulls only	All genotyped bulls	139,057	12,895*

¶ Bulls\* + their dams + random cows with phenotypes

\* Bulls born before 2011; with at least 1 tested daughter before 2011

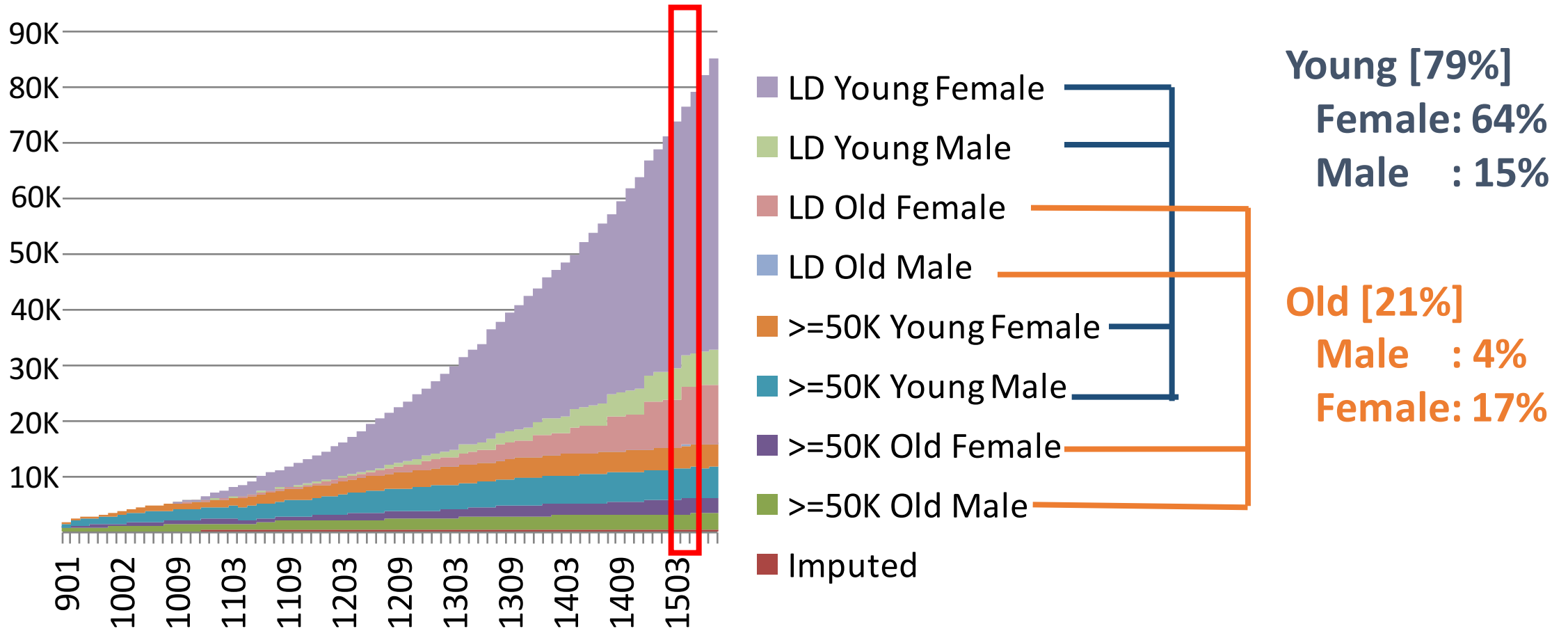


# Results

Prediction	Method	Genotypes	Core	R <sup>2</sup>	b1 slope
Official PTA	BLUP	NA	NA	0.27	0.71
Official GPTA*	BayesA-type	Reference(>20K)	NA	0.51	0.81
Single-step	ssGBLUP	All (760K)	Bulls+Cows (20K)	0.45	1.06
		All (760K)	Bulls (13K)	0.46	0.97
		HigherD (150K)	Bulls (13K)	0.49	0.97
		Bulls (140K)	Bulls (13K)	0.50	0.98

\* Included internationally-evaluated foreign bulls without daughters in the US.

# Genotyped animals



# Results

Prediction	Method	Genotypes	Core	R <sup>2</sup>	b1 slope
Official PTA	BLUP	NA	NA	0.27	0.71
Official GPTA*	BayesA-type	Reference(>20K)	NA	0.51	0.81
Single-step	ssGBLUP	All (760K)	Bulls+Cows (20K)	0.45	1.06
		All (760K)	Bulls (13K)	0.46	0.97
		HigherD (150K)	Bulls (13K)	0.49	0.97
		Bulls (140K)	Bulls (13K)	0.50	0.98

\* Included internationally-evaluated foreign bulls without daughters in the US.

# QTL effect of DGAT1

- A DGAT1 marker captured with commercial SNP panels
- Estimation of substitution effect ( $\beta$ )
  - Using 14,376 bulls with DYD reliability  $\geq 72.5\%$  (i.e. DE  $\geq 50$ )
  - $DYD = \mu + \beta * GeneContent + u + e$
  - $\beta = 8.69 (\pm 0.51)$  for protein yield
- Blended GPTA
  - $GPTA^* = GPTA + 0.5\beta * GeneContent$

# Results (Blended GPTA)

Model	Genotypes	Core	GPTA		Blended GPTA	
			R <sup>2</sup>	b1	R <sup>2</sup>	b1
Official PTA	NA		0.27	0.71		
Official GPTA*	Reference(>20K)		0.51	0.81		
Single-step	All (760K)	Bulls+Cows (20K)	0.45	1.06	0.47	1.00
	All (760K)	Bulls (13K)	0.46	0.97	0.47	0.93
	HigherD (150K)	Bulls (13K)	0.49	0.97	0.51	0.91
	Bulls (140K)	Bulls (13K)	0.50	0.98	0.51	0.92

\* Included internationally-evaluated foreign bulls without daughters in the US.

# Summary

- Single-step GBLUP provides GPTA with much less bias and similar accuracy compared to the official prediction.
- Extra genotyped animals don't improve the accuracy of prediction.
- More “core” animals don't contribute to better prediction.
- Inclusion of major gene effect can slightly contribute to  $R^2$  with little inflation in GPTA.