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

Yutaka Masuda, Ignacy Misztal, and Tom Lawlor

University of Georgia and Holstein Association USA Inc.

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 ($\mathbf{G}^{-1}_{APY}$)
  - Avoidance of direct computation of $\mathbf{A}^{-1}_{22}$
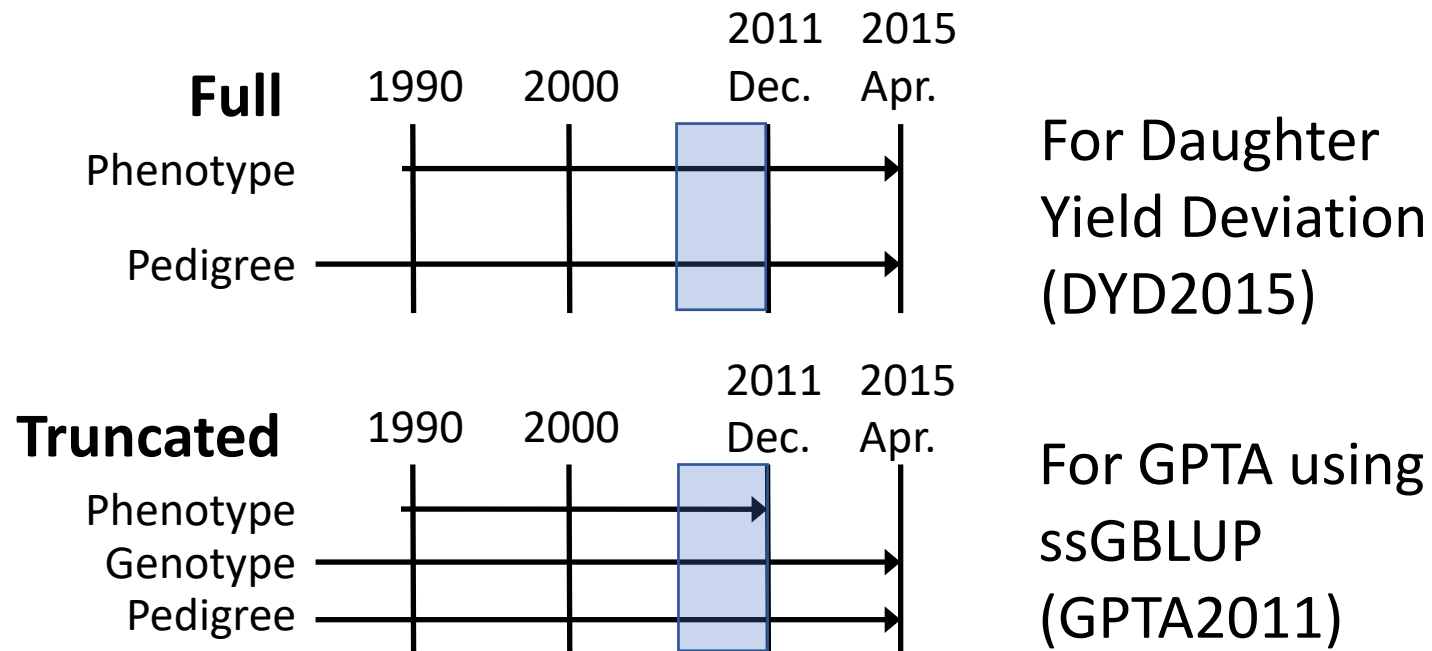- How does it work for production traits for US Holsteins?

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

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

Full

2011 2015
1990 2000 Dec. Apr.

Phenotype

Pedigree

For Daughter Yield Deviation (DYD2015)

Truncated

2011 2015
1990 2000 Dec. Apr.

Phenotype
Genotype
Pedigree

For GPTA using ssGBLUP (GPTA2011)

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

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

Core animals: Bulls with at least 1 tested daughter before 2011 (~13K)

Genotyped animals

Genotyped animals	# of genotyped animals
All	764,029
Known parents	656,313
Bulls	139,057
Proven bulls	19,622
Bulls + cows with record(s)	27,2029
Proven bulls + cows with record(s)	152,594

Core animals: Bulls with at least 1 tested daughter before 2011 (~13K)

Results

| Prediction	Method	Genotypes	R^2	b1 slope
Official PTA	BLUP	NA	0.27	0.71
Official GPTA*	BayesA-type	Reference(>20K)	0.51	0.81
Single-step	ssGBLUP	All (760K)	0.45	1.06

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

Results

Prediction	Method	Genotypes	R^2	b1 slope
Official PTA	BLUP	NA	0.27	0.71
Official GPTA*	BayesA-type	Reference(>20K)	0.51	0.81
Single-step	ssGBLUP	All (760K)	0.45	1.06
		Known parents (660K)	0.47	1.09
		Bulls (140K)	0.50	0.98
		Proven bulls (20K)	0.51	0.84
		Bulls + cows (270K)	0.50	0.96
		Proven bulls + cows (150K)	0.50	0.98

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

Genotyped animals

Phenotypes removed in validation

https://www.cdcb.us/Genotype/cur_density.html
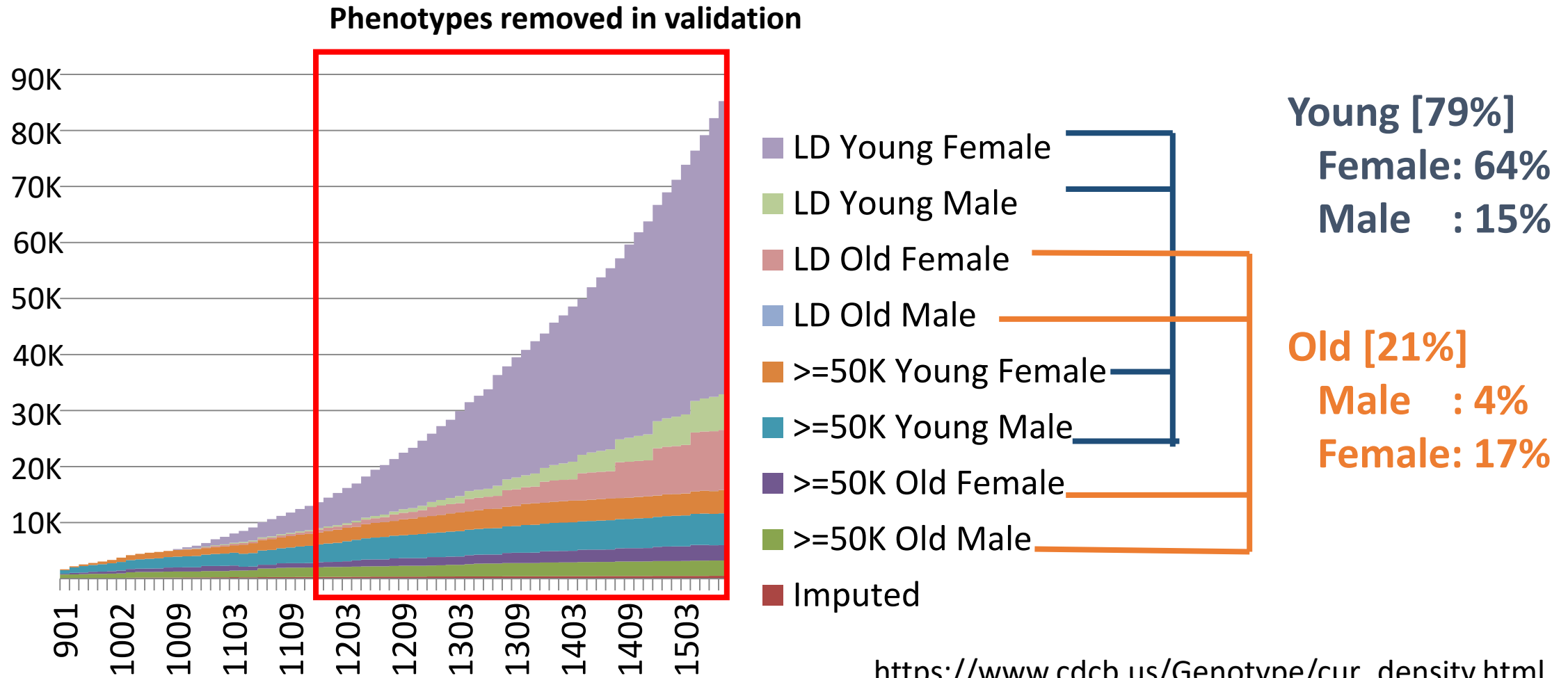
QTL effect of DGAT1

- A DGAT1 marker captured with commercial SNP panels
- Estimation of substitution effect (β)
  - Using 14,376 bulls with DYD reliability >= 72.5% (i.e. DE >= 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	GPTA		Blended GPTA	
		R^2	b1	R^2	b1
Official PTA	NA	0.27	0.71		
Official GPTA*	Reference(>20K)	0.51	0.81		
Single-step	All (760K)	0.46	0.97	0.47	0.93
	Bulls (140K)	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.

- Inclusion of major gene effect can slightly contribute to R^2 with little inflation in GPTA.

Acknowledgement