Challenges in dairy breeding under genomic selection

Conversations with the University of Georgia and others.

T.J. Lawlor, I. Misztal, Y. Masuda and S. Tsuruta

U.S. Registered Holsteins
FOR MAXIMUM PROFIT

THE UNIVERSITY OF GEORGIA

Why do we want to do this?

Genomic selection

- Accurate, unbiased genetic evaluations.
- Faster genetic improvement.

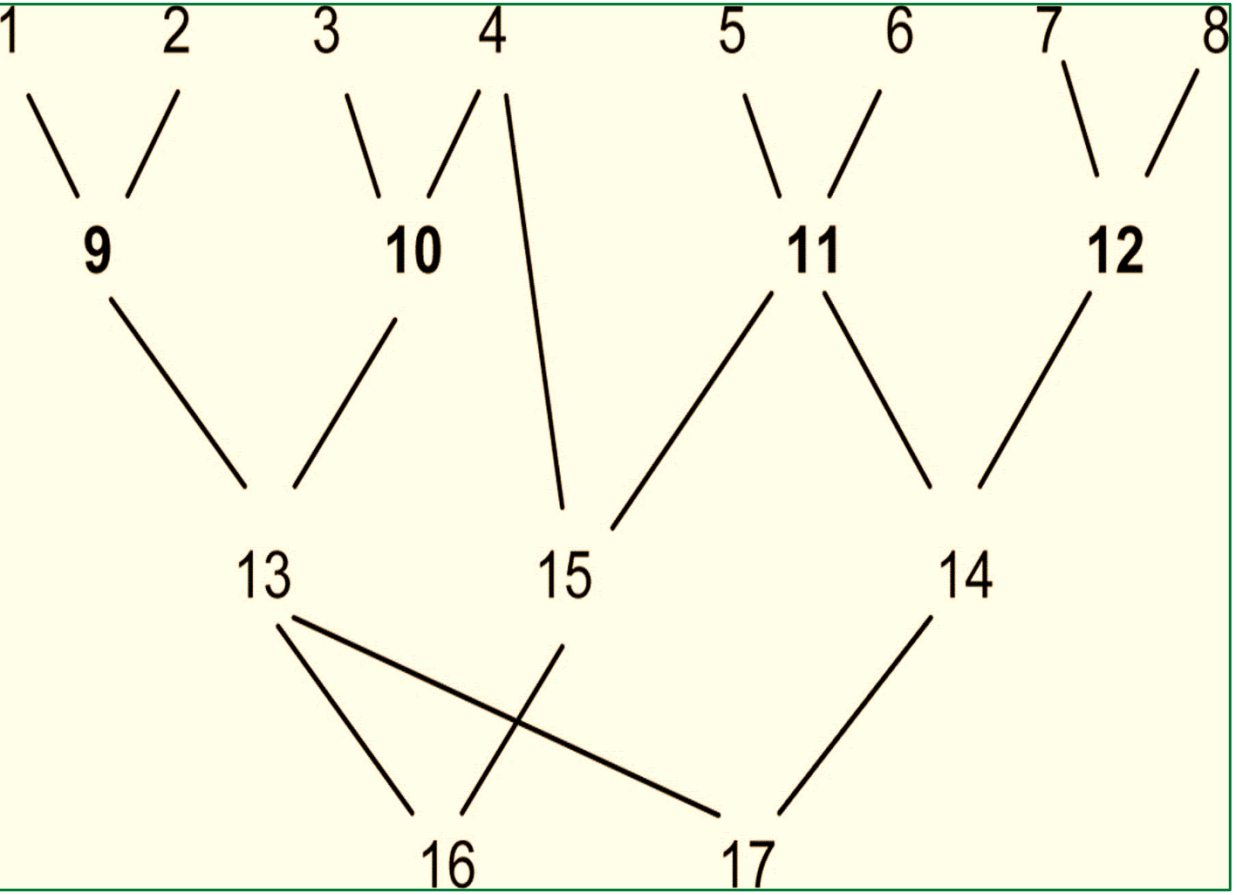
ssGBLUP

- A more unified approach – tie the adjustments together
- More complete conceptual picture

  Genetic explanation of the phenotypes that we're observing and the analysis being applied.

Where we started

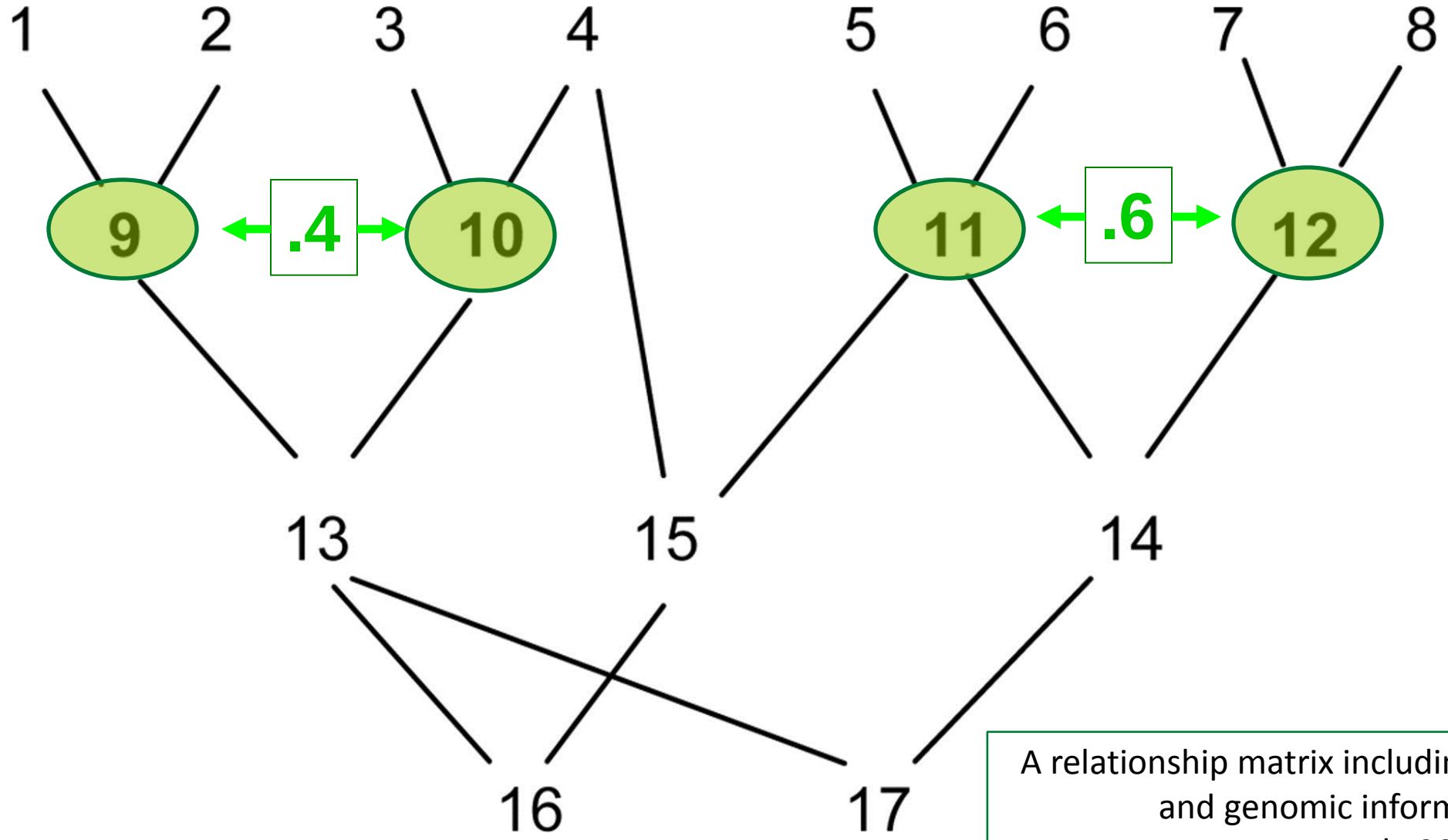
Pedigree relationships



1.00							0.50				0.25			0.13	0.13
	1.00						0.50				0.25			0.13	0.13
		1.00						0.50			0.25			0.13	0.13
			1.00					0.50			0.25		0.50	0.38	0.13
				1.00					0.50			0.25	0.25	0.13	0.13
					1.00				0.50			0.25	0.25	0.13	0.13
						1.00				0.50		0.25			0.13
							1.00			0.50		0.25			0.13
0.50	0.50							1.00			0.50			0.25	0.25
		0.50	0.50						1.00		0.50		0.25	0.38	0.25
				0.50	0.50					1.00		0.50	0.50	0.25	0.25
						0.50	0.50				1.00		0.50		0.25
0.25	0.25	0.25	0.25					0.50	0.50			1.00		0.13	0.56	0.50
				0.25	0.25	0.25	0.25			0.50	0.50		1.00	0.25	0.13	0.50
			0.50	0.25	0.25				0.25	0.50		0.13	0.25	1.00	0.56	0.19
0.13	0.13	0.13	0.38	0.13	0.13			0.25	0.38	0.25		0.56	0.13	0.56	1.06	0.34
0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.25	0.25	0.25	0.25	0.50	0.50	0.19	0.34	1.00

What's new
Molecular markers tells us more about the relationship of the animals

Genomic relationships



A relationship matrix including full pedigree and genomic information
Legarra et al., 2009

Genomic relationships for direct ancestors, descendants and all other relatives of genotyped animals

1	0	0.15	0.15	0	0	0	0	0.5	0.3	0	0	0.4	0	0.075	0.238	0.2
0	1	0.15	0.15	0	0	0	0	0.5	0.3	0	0	0.4	0	0.075	0.238	0.2
0.15	0.15	1	0	0	0	0	0	0.3	0.5	0	0	0.4	0	0	0.2	0.2
0.15	0.15	0	1	0	0	0	0	0.3	0.5	0	0	0.4	0	0.5	0.45	0.2
0	0	0	0	1	0	0.1	0.1	0	0	0.5	0.2	0	0.35	0.25	0.125	0.175
0	0	0	0	0	1	0.1	0.1	0	0	0.5	0.2	0	0.35	0.25	0.125	0.175
0	0	0	0	0.1	0.1	1	0	0	0	0.2	0.5	0	0.35	0.1	0.05	0.175
0	0	0	0	0.1	0.1	0	1	0	0	0.2	0.5	0	0.35	0.1	0.05	0.175
0.5	0.5	0.3	0.3	0	0	0	0	1	0.6	0	0	0.8	0	0.15	0.475	0.4
0.3	0.3	0.5	0.5	0	0	0	0	0.6	1	0	0	0.8	0	0.25	0.525	0.4
0	0	0	0	0.5	0.5	0.2	0.2	0	0	1	0.4	0	0.7	0.5	0.25	0.35
0	0	0	0	0.2	0.2	0.5	0.5	0	0	0.4	1	0	0.7	0.2	0.1	0.35
0.4	0.4	0.4	0.4	0	0	0	0	0.8	0.8	0	0	1.3	0	0.2	0.75	0.65
0	0	0	0	0.35	0.35	0.35	0.35	0	0	0.7	0.7	0	1.2	0.35	0.175	0.6
0.075	0.075	0	0.5	0.25	0.25	0.1	0.1	0.15	0.25	0.5	0.2	0.2	0.35	1	0.6	0.275
0.238	0.238	0.2	0.45	0.125	0.125	0.05	0.05	0.475	0.525	0.25	0.1	0.75	0.175	0.6	1.175	0.463
0.2	0.2	0.2	0.2	0.175	0.175	0.175	0.175	0.4	0.4	0.35	0.35	0.65	0.6	0.275	0.463	1.125

Corrects pedigree relationships in view of genomic relationship

Parents

Genotyped Animals

Progeny

Grandprogeny

Extends genomic information to non-genotyped animals

Single step GBLUP
BLUP with H^{-1} replacing A^{-1}

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

Aguilar et al., 2010; Christensen and Lund, 2010

Overcoming the challenges

- Put A and G on the same scale.
- Account for polygenic variation – done differently in different studies.
- Fine tuning of parameters.
- The size of the problem – how do you invert the “growing” G matrix.
- Empirical evidence that it works.
- Delivery of genomic information. Weekly results, sharing and incorporating international information.

Putting A and G on the same scale.

- Deviation from base allele frequency
  - ➢ Deviation from 0.5

- Adjust G to match A
- ➢ **Adjust A to match G**

- Use ALL genotyped animals
  - ➢ Use bulls born since 1990

- Base animals actually share genes identical by descent, which shift relationships and inbreeding values up or down.
- Genomic and pedigree-based matrices should be compatible in scale.

Forni et al 2011; Vitezica et al., 2011; Christensen et al 2012

Genomic and pedigree relationships
All Holsteins born since 1990

Genomic relationships are similar to pedigree relationships but more accurate.
Diagonals and off-diagonals of $\boldsymbol{G}$ and $\boldsymbol{A}$ should be of the same size and scale

Three accuracy measures of $\boldsymbol{G}$ and $\boldsymbol{A}$ are:

Average inbreeding

Differences of off-diagonals:

standard deviation $\boldsymbol{G-A}$

correlation between $\mathbf{G}$ and $\mathbf{A}$

Parameters from U.S. Holsteins

Pedigree inbreeding = 5.91

Genomic inbreeding = 5.91

sd (G-A) = 0.03

Correlation (G,A) = 0.7

Are we done with correctly combining G and A?
No, but no major obstacles exist.

- Should we continue to use the 1990 base?
  - Allele frequencies continue to change over time. *Legarra e t al, 2014*
  - What about a shorter time span or use a more homogeneous base? *Lourenco et al, 2014*

- What about external populations or missing pedigrees (foreign genetics)?
  - *Unknown Parent Groups – changes inbreeding, drift and across-group relationships.*
    Misztal et al, 2013; Tsuruta et al, 2014

- *What about multiple breeds and crossbreds?*
  - *Ancestral relationships using Metafounders.* *Legarra et al, 2015*
  - *Correct for the underestimation of inbreeding in A, with respect to G.*
  - *Leads to a "coherent theoretical framework" of the base population.*

Choice of parameters
Putting ssGBLUP into practice

Theory, intuition and empirical results.

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

$$\mathbf{u}_2 \,|\, \mathbf{A}_{22}, \mathbf{G} \quad \sim \quad N(0, \frac{\mathbf{G}}{\tau}) \qquad N(0, \frac{\mathbf{A}_{22}}{1\text{-}\omega})$$

Single Trait Analysis

Final score - Aguilar et al., 2010

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau G^{-1} - \omega A_{22}^{-1} \end{bmatrix}$$

Prediction in 2004	DD2009	
	R^2	b-value
Parent Avg	24	0.76
Multistep (VanRaden)	40	0.86
Single-step		
G^{-1}-A_{22}^{-1}	41	0.76
$1.5G^{-1}$-$0.9A_{22}^{-1}$	42	0.87
$1.5G^{-1}$-$0.6A_{22}^{-1}$	41	0.96

Multiple-Trait Genomic Prediction

Linear type traits - Tsuruta et al, 2011

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

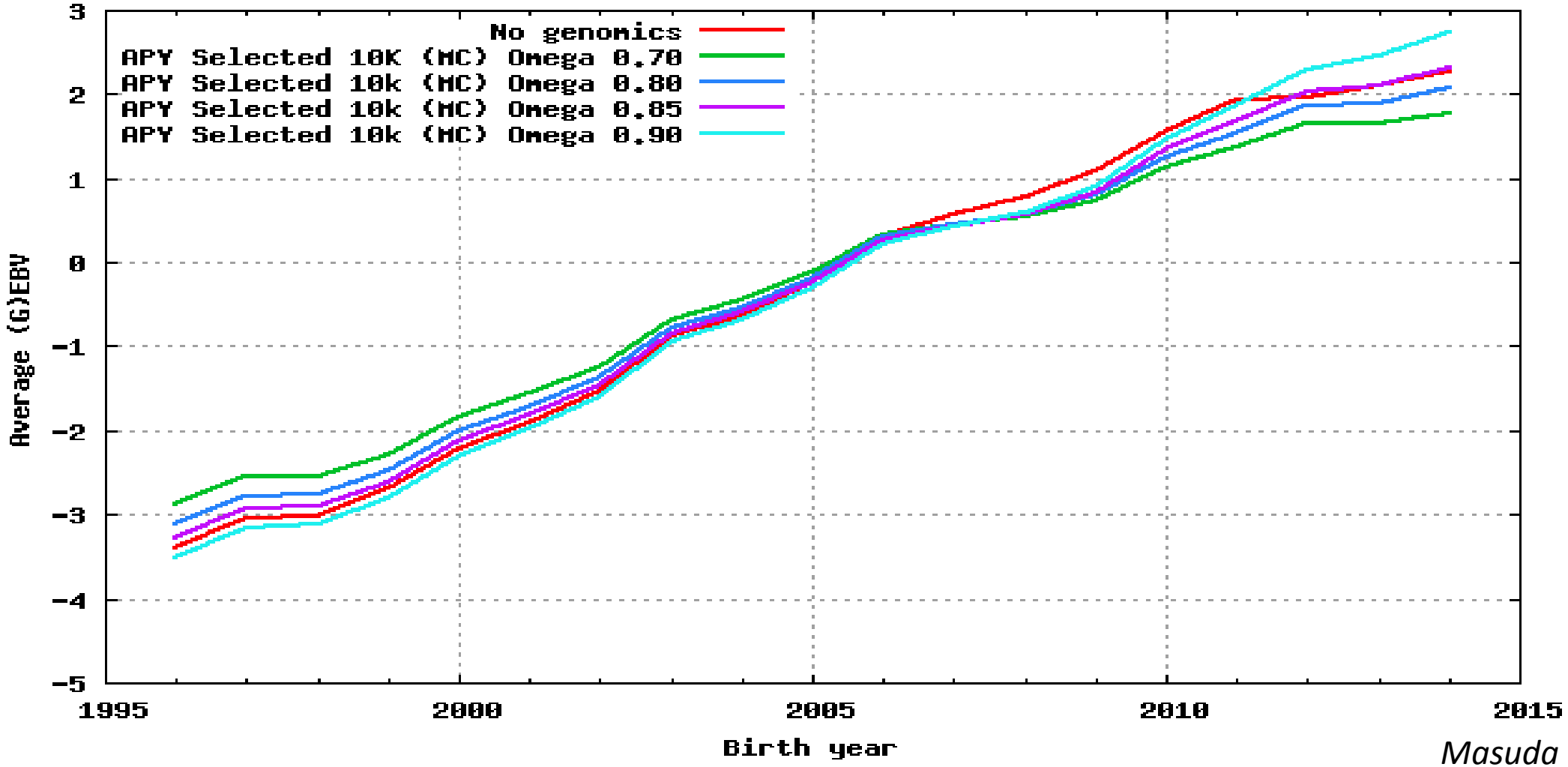
	R^2	Weight for A_{22}^{-1}	
		$w = 1$	$w = 0.7$
Traditional MT Parent Average	20.5	.77	.77
Genomic Single Trait	34.6	.79	.91
Genomic Multiple Trait	37.3	.80	**.93**

scaling A_{22}^{-1} more important than scaling G for controlling bias

Genetic Trend of bulls with Different Omega (ω)

smaller ω = lower genetic trend

changes heritability



Masuda et al, 2015

Single-step genomic evaluation using multitrait random regression model and test-day data. *Koivula et al, 2015*

Bull validation results - showing regression coefficients (b_1)

ssGBLUP[1]		Milk	Protein	Fat	$\mathbf{G}_w = (1 - w)\mathbf{G} + w\,\mathbf{A}_{22}$
w_{20}	80% on G	0.87	0.73	0.72	Weight on G had a small effect. Lower weight on G results in Less over-prediction
w_{15}	85% on G	0.86	0.72	0.72	
w_{10}	90% on G	0.84	0.71	0.71	
					$1.0\ G_w^{-1} + 0.5\ A_{22}^{-1}$
$\tau = 1.0$ $\omega = 0.5$ $w_{10} = .90G$	weight on $\mathbf{A}_{22}^{-1}$	1.09	0.92	0.87	Weight on $\mathbf{A}_{22}^{-1}$ had a larger effect. Lower heritability results in <u>Much</u> less over-prediction

w = proportion of polygenic variance; τ = weight for $\mathbf{G}^{-1}$; ω = weight for $\mathbf{A}_{22}^{-1}$ matrix

Adjusting for Polygenic Variance

Currently done in multiple places,

$G_w = G + A_{22}$ and in H^{-1} with $G_w^{-1} + A_{22}^{-1}$

- What do we see.
  - Changes the heritability, differs by trait, differs by de-regression procedure, influenced by amount of foreign data, length of data , etc..

What are these adjustments doing?

- Lowers heritability ---> genomic heritability
- Animals with longer pedigrees receive less weight ---> linkage decay

Modeling in the genomic era

- Scale **G** and $\mathbf{A}_{22}$ appropriately for ancestral relationships by (**α**).
- Fine tune $\mathbf{G}_w = (.xx\mathbf{G} + (1 - .xx)\mathbf{A}_{22})$, i.e., keep it positive definite.
- Account for reduced heritability in combined relationship matrix $\mathbf{A}_{22}^{-1}$

  genomic heritability is lower because of LD. (**τ** and **ω**)

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau(0.95\mathbf{G} + 0.05\mathbf{A}_{22} + \alpha)^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix}$$

Laurenco et al, 2014

Recent changes by USDA-CDCB to make the multi-step procedure more accurate.

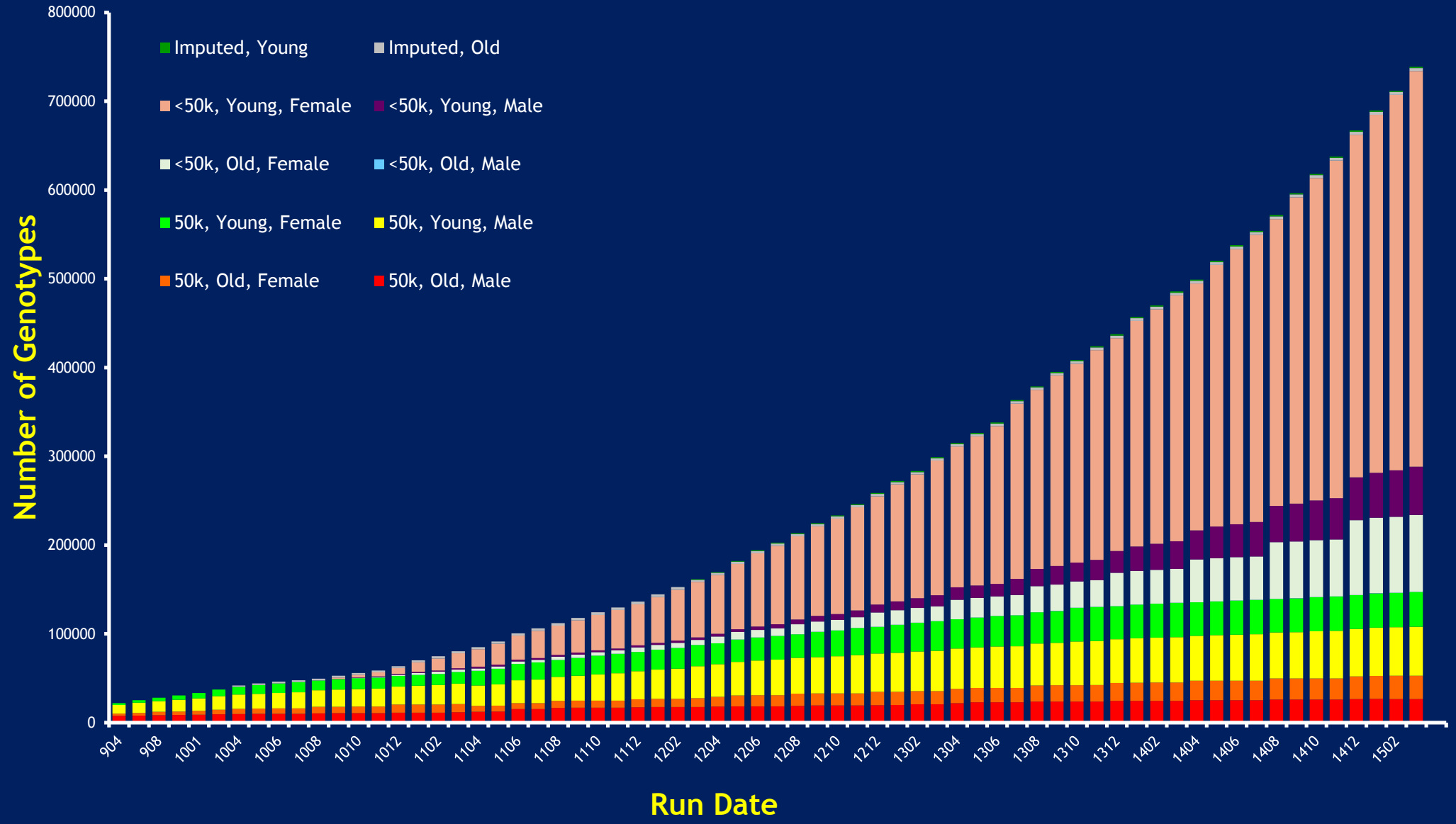
Polygenic effect was added (2010), the variance of the cow PTAs were reduced (2010), weight on Direct Genomic Value was reduced (2012, 2013), correlations in multi-trait Productive Life were reduced (2012)

Heritability of yield traits reduced (2014)

Source: VanRaden, 2015

Size of the problem

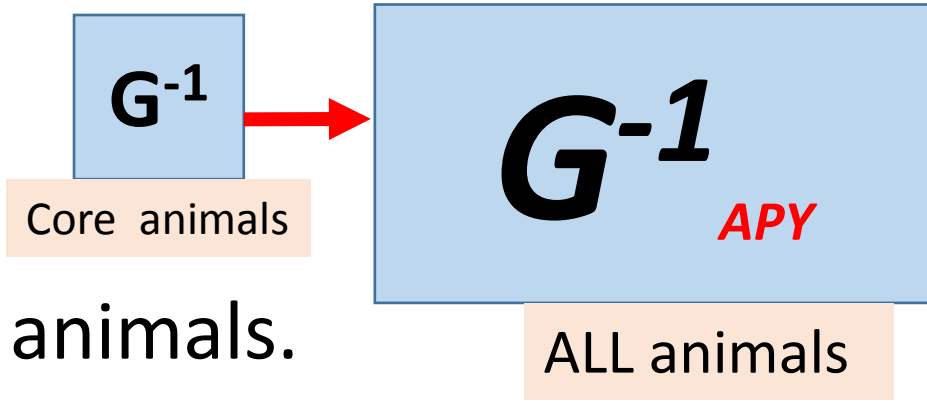
Genomic testing continues to grow



Large G matrix
inversion by genomic recursion

- Henderson, 1976: A^{-1} by recursion, animals ordered: oldest to youngest
- Misztal et al, 2014 - G^{-1} by APY.

  Algorithm based on decomposition for **P**roven and **Y**oung animals.
  - Recursion of Young animals on Proven (phenotypes or progeny) animals.
  - Order of animals – not critical.



- Recursion on <u>any core group</u> of about 10,000 animals.

  Limited number of independent chromosome segments. *(Stam,1980)*
- G^{-1} is approximate as covariances among young animals is ignored.

Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes.

Fragomeni et al, 2015

Genomic EBV (GEBV) were calculated with a regular inverse of **G**, and with the **G** inverse approximated by APY.

	Description	Number of records/animals
Phenotype	Final score for US Holstein cows classified in 2014 or earlier	11,102,702
	Cows classified	6,943,618
Genotyped Sires	23,000 (16,500 with > 5 progeny)	23,000
Genotyped Cows	Cows with records	27,000
Young Animals	Random Samples of genotyped animals	50,000

Genomic EBV (GEBV) were calculated with a regular inverse of **G**, and with the **G** inverse approximated by APY.

Core animals	Correlation	
100,000 ALL – Direct G^{-1}	1.00	
50,389 Sires and Cows	0.995	
23,174 Sires	0.994	
27,215 Cows	0.992	
Random 15,000 Proven	.989 - .990	Replicated 4 times
Random 20,000 Proven	.992 - .993	
Random 20,000 **Young**	.989 - .990	

Fragomeni et al, 2015

Single-step genomic evaluations with 570K genotyped animals in US Holsteins

Y. Masuda et al., 2015

2009 data predicting 2014 data --- 2,948 bulls with 30 daughters in 2014

Core animals	b_1	R^2	comment
9,406 bulls with at least 1 daughter	.96		ω = .7
9,406 bulls with at least 1 daughter	.82		ω = .9 *not optimized*
9,406 bulls and 7,422 classified cows	.83	.45	ω *not optimized*
5,000 random animals	.83	.39	ω *not optimized, 3 reps.*
10,000 random animals	.83	.44	ω *not optimized, 3 reps.*
15,000 random animals	.83	.44	ω *not optimized, 3 reps.*
20,000 random animals	.82	.44	ω *not optimized, 3 reps.*
30,000 random animals	.82	.44	ω *not optimized, 3 reps.*

Findings from APY study

- Genomic prediction with all genotypes and all available data is now possible.
- Genomic predictions for bulls were similar regardless of the definition of core animals.
- A larger number of core animals slowed down the rate of convergence.
- The parameter ω affected genomic evaluations :
  - Smaller omega, more accurate predictions.
  - Smaller omega, faster convergence.

Single-step GBLUP – other findings

- VanRaden (2012) applied the algorithm of Legarra-Ducrocq (2012) to USA Jersey yield traits and obtained good results.
- However, he was unable to obtain convergence when he applied the same procedure to the Holstein yield traits.
- For comparison purposes, Yutaka Masuda (2015) applied APY algorithm (G^{-1}_{APY}) to USA Jersey yield traits.
- New results are very similar to the results from the Legarra-Ducrocq (2012) algorithm.
- All indications are that the APY algorithm (G^{-1}_{APY}) will work for the U.S. Holstein yield data. USDA-CDCB expect good CPU time, but will need more memory

Single-step GBLUP – works

- Major challenges have been addressed.
- Final testing is underway.
- Data flow and coordination with others is being addressed.
- Now, suitable for a National Genetic Evaluation.

Any questions ?