

Current Status of Genomics applied to Beef Cattle Improvement in North America

Dorian Garrick
dorian@iastate.edu

American Angus Association

- Two Competing Commercial Products both derived from training using Bayesian regression models
 - Pfizer
 - Merial Igenity
- DGV included as correlated traits
- Zoetis recently retrained
 - Now a single product fitting DGV as correlated traits
- Prototyping single-step GBLUP
 - >50,000 genotyped animals

Current prediction accuracies: Bhoddireddy et al 2014 J ANIM SCI 92:485-497

USDA 2,000 Bulls Project

- USDA genotyped widely-used bulls representing all breeds with National Cattle Evaluation, each non-Angus breed represented roughly in proportion to its contribution to the US industry

Predicting Other Breeds from Angus

- Prediction of merit of 2,000 bulls showed that neither the Pfizer/Zoetis or Merial-Igenity predictions had no utility in other breeds
 - Each breed needed its own training populations
 - The 2,000 bull population provided 200-400 bulls in any particular breed that could form a foundation for a breed-specific training population

New Training

- Proceeded as each breed recognized it to be strategically important
- Used deregressed breeding values with parent average removed in weighted Bayesian multiple regression
 - Mostly BayesC with $\pi > 0.9$

K-fold Cross Validation

- Partition the dataset into k (say 3) groups

Training	G1	
	G2	✓
	G3	✓
Validation		G1



Derive g-EBV



Compute the correlation between predicted genetic merit from g-EBV and observed performance

3-fold Cross Validation

- Every animal is in exactly one validation set

Training	G1		✓	✓
	G2	✓		✓
	G3	✓	✓	
Validation		G1	G2	G3

Genetic relationship between training and validation data influences results!

Cross-fold Validation

- Form k-means groups based on a distance matrix that clusters like animals together
 - We use a distance matrix based on additive genetic correlations
 - Need (extended) pedigree of genotyped animals
- Bivariate AS-REML with pedigree-based var-cov matrix to get a single estimate of accuracy
 - Zero the cross-fold blocks of the **A** matrix
 - Approximation since folds can vary in accuracy

Canadian Genotyping Projects

- Genome Canada
 - Steve Moore, Steve Miller, Paul Stothard
 - Genotyped >800 individuals in 10 breeds
- Other Genome Canada and provincial funding
 - Genotyped additional animals to migrate parentage testing from microsatellite to SNP

50k Predictions in Beef Cattle Breeds

Trait	RedAngus (6,412)	Angus (3,500)	Hereford (2,980)	Simmental (2,800)	Limousin (2,400)	Gelbvieh (1,321)+
BirthWt	0.75	0.64	0.68	0.65	0.58	0.62
WeanWt	0.67	0.67	0.52	0.52	0.58	0.52
YlgWt	0.69	0.75	0.60	0.45	0.76	0.53
Milk	0.51	0.51	0.37	0.34	0.46	0.39
Fat	0.90	0.70	0.48	0.29		0.75
REA	0.75	0.75	0.49	0.59	0.63	0.61
Marbling	0.85	0.80	0.43	0.63	0.65	0.87
CED	0.60	0.69	0.68	0.45	0.52	0.47
CEM	0.32	0.73	0.51	0.32	0.51	0.62
SC		0.71	0.43		0.45	
Average	0.67	0.69	0.52	0.47	0.57	0.56

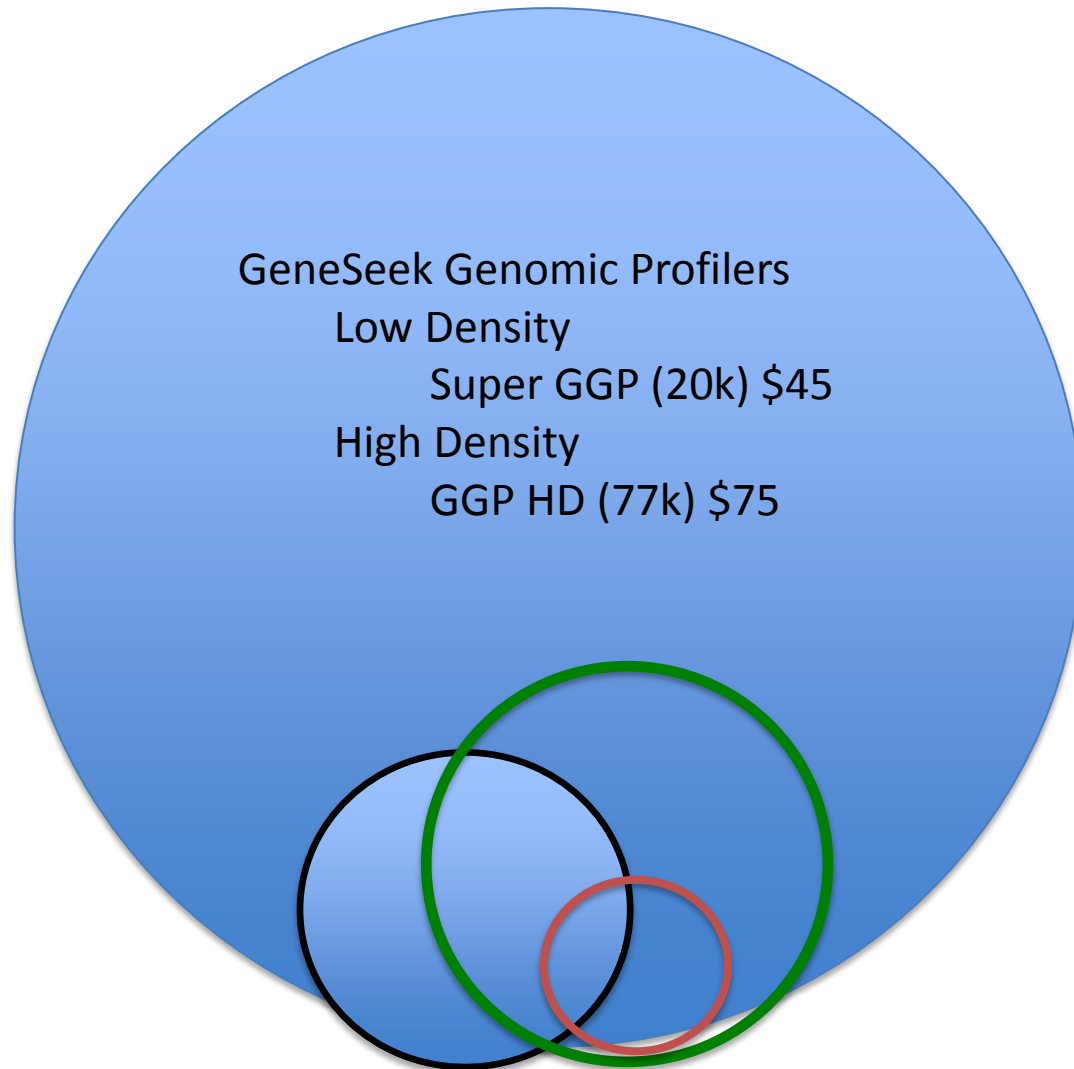
Genetic correlations from k-fold validation Saatchi et al (GSE, 2011; 2012; J Anim Sc, 2013)

Early 2014 Genotype Counts

Breed	9k	GGP-LD	50k	GGP-HD	BOS-1	700k HD	TOTAL
AAN		911	13,409	787		947	16,054
HER			7,064	1,887	471	850	10,272
BSH			325			136	461
CHA			1,617			525	2,142
GVH	186	209	1,643	371	414	430	3,253
LIM		429	3,420	8	461	675	4,993
RAN			1,931	1,183	226		3,340
RDP			1,394				1,394
SIM	5,223	7,026	6,501	1,347	1,601	674	22,372
BRG			1,128	173		243	1,544
NEL						2,571	2,571
TOTALS	5,409	8,575	38,432	5,756	3,173	7,051	68,396

No longer using Illumina 50k

Panel Comparison



Orange = GGP-Super LD 19k

Green = GGP-HD (taurus) 70k

Black = Illumina 50K

GGP also include custom SNP

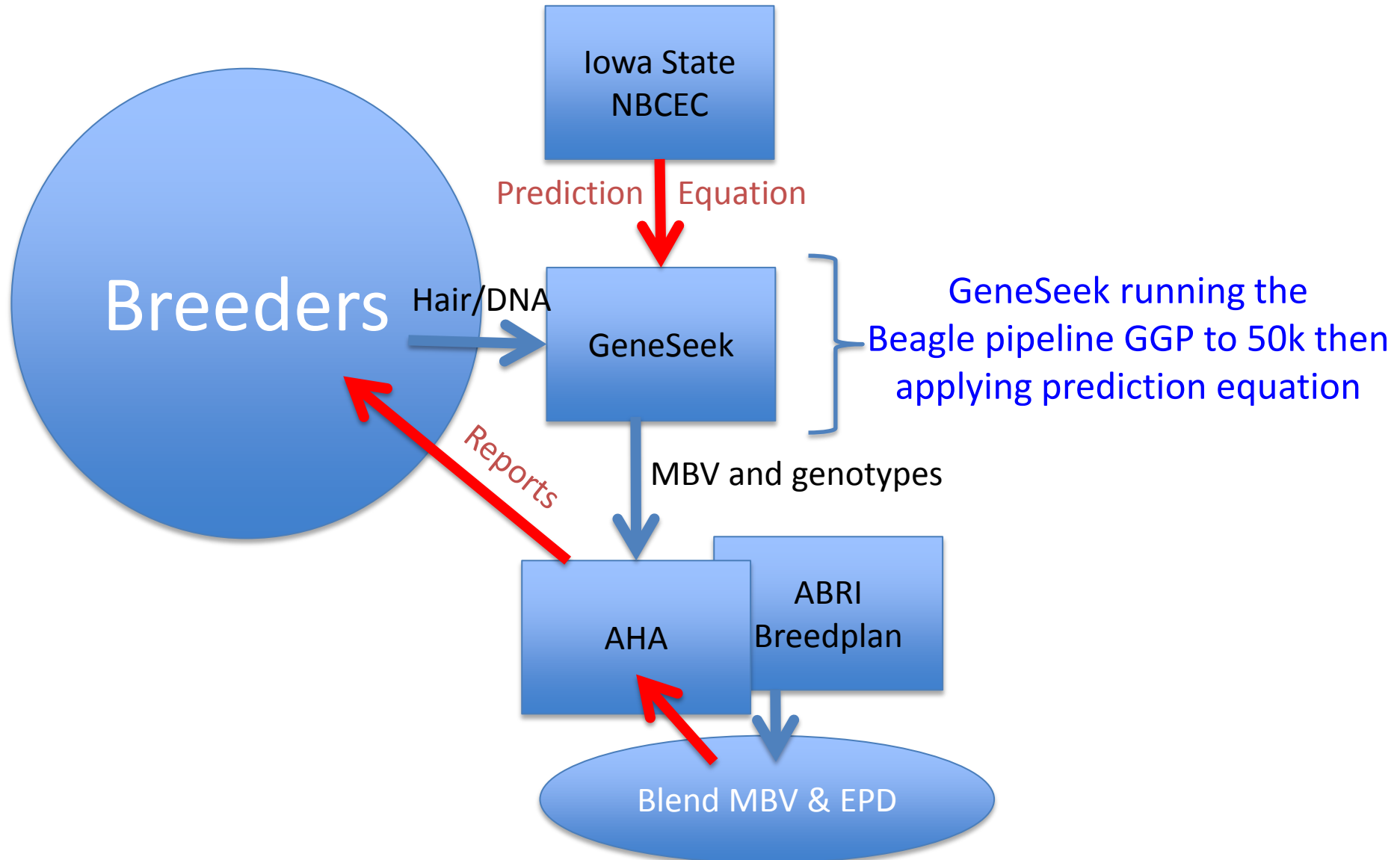
50k and GGP-HD share 28K

50k and GGP-Super LD share 8k

Also a separate GGP-HD-I (Indicus)

There are multiple minor variants of all these panels!

Genomic Prediction Pipeline



Problems with training/validation on DEBV

Brangus BW		Validation	
		DEBV	
Training	DEBV	0.6	
	PA+DEBV	0.65	



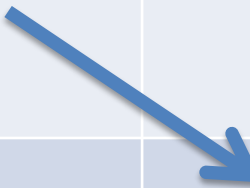
Problems with training/validation on DEBV

Brangus BW		Validation	
		DEBV	PA+DEBV
Training	DEBV	0.6	0.46
	PA+DEBV	0.65	0.51

However, you validate, PA+DYG is better than DYG for training

Problems with training/validation on DEBV

Brangus BW		Validation	
		DEBV	PA+DEBV
Training	DEBV	0.6	0.46
	PA+DEBV	0.65	0.51



However, you validate, PA+DYG is better than DYD for training

But training and validating on DEBV had higher r than PA+DYG

Problems with training/validation on DEBV

Brangus BW		Validation	
		DEBV	PA+DEBV
Training	DEBV	0.6	0.46
	PA+DEBV	0.65	0.51
ASREML variance	$\sigma_p^2=60$ $h^2=0.42$	$\sigma_{\text{"g"}}^2=10-11$ (too low)	$\sigma_{\text{"g"}}^2=38-39$ (too high)

And neither approach gives the expected estimate of “genetic” variance

Hereford BW DGV from PA+DEBV

	Saatchi et al		Using PA		Using Phenotype	
$\sigma_g^2=26$ $\sigma_e^2=34$	DEBV	DGV	PA+DEBV	DGV	AdjPhen	DGV
Validn Data	16 (>13)	0.65	47 (>>26)	0.52	26 (ok)	0.5
DGV	11	19	15	18	11	18

Also need to ensure the regression of phenotype on DGV is near 1

Results by fold – new BW

Fold	#MBV	#phen	h^2_{MBV}	h^2_{phen}	B_phen/MBV	rg
1	5,348	11,624	0.98	0.41	0.64	0.67
2	5,413	12,244	0.95	0.43	0.60	0.62
3	5,532	11,883	0.94	0.41	0.62	0.67
4	5,498	12,090	0.92	0.41	0.59	0.62
ALL	21,791	47,841	0.95	0.42	0.61	0.64

h^2 MBV near 1 as we would expect

h^2 phenotype adjusted for breed as expected

regressions of phenotype on MBV all slightly biased

genetic correlations indicate good across-breed predictive ability
but these include predictions of breed effects

Results by breed – new BW

Breed	#MBV	#phen	h2_MBV	h2_phen	B_phen/MBV	rg
AAN	1,770	1,151	1.00	0.41	0.69	0.66
GVH	1,788	1,508	1.00	0.51	0.53	0.64
RAN	2,251	3,206	1.00	0.56	0.65	0.64
RDP	585	380	1.00	0.77	0.42	0.44
SIM1	1,162	4,195	0.79	0.40	0.63	0.59
SIM2	4,128	11,209	0.89	0.44	0.69	0.63
SIM3	4,071	13,752	0.91	0.40	0.65	0.64
SIM4	2,711	13,117	0.90	0.41	0.67	0.68
SIM5	1,357	4,401	1.00	0.39	0.67	0.74
SIM6	1,968	4,817	1.00	0.39	0.65	0.65

Training and validation is soon to be repeated with LIM and BSH included

Single Step

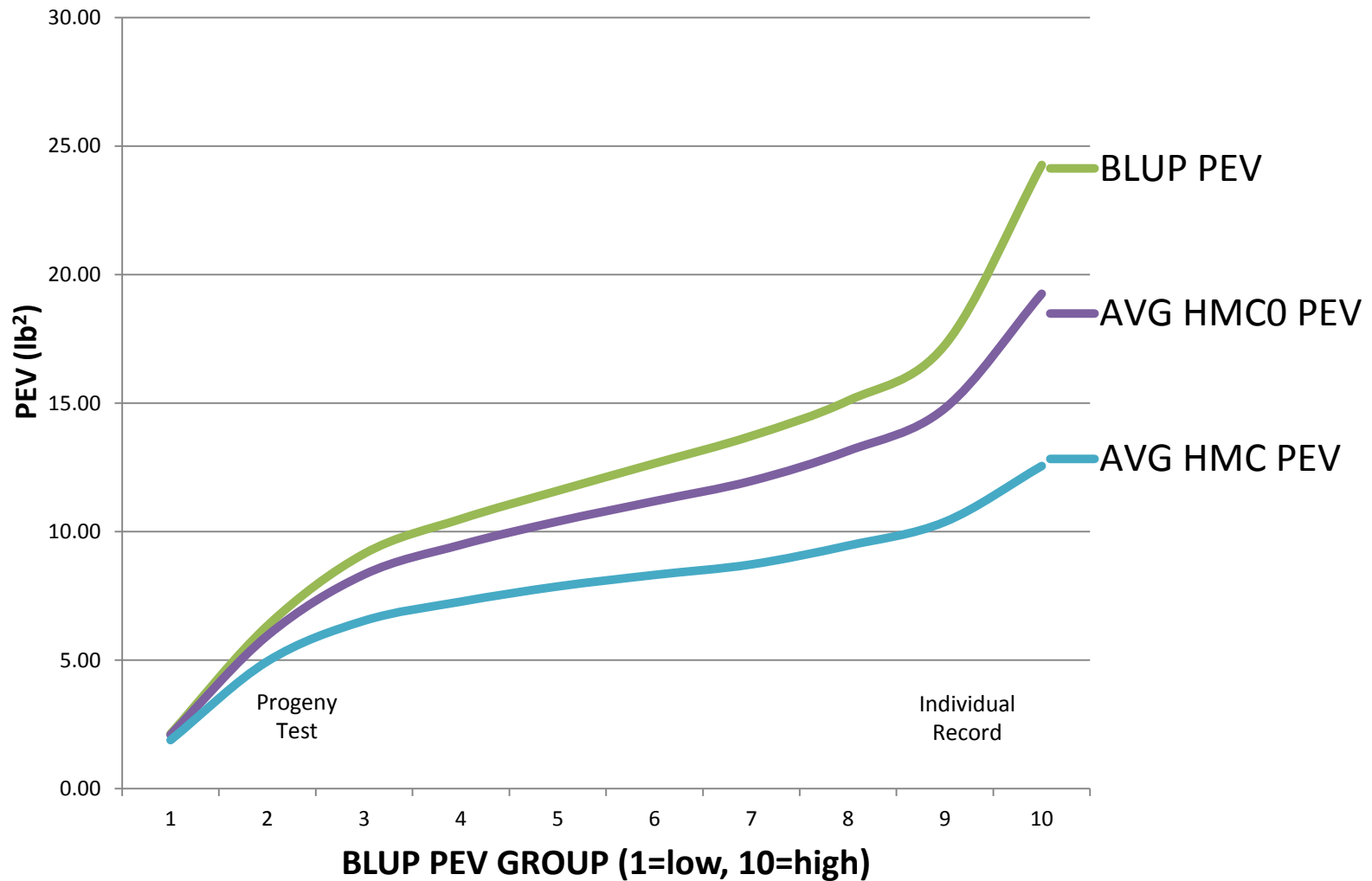
- Also want to use MCMC for PEV (and R^2)
 - Also want Prediction Error Covariance
- And to use mixture models

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}\mathbf{M} & \mathbf{X}'\mathbf{Z}_1 \\ \mathbf{M}'\mathbf{Z} & \mathbf{M}'\mathbf{Z}'\mathbf{Z}\mathbf{M} + \phi & \mathbf{M}'_1\mathbf{Z}'_1\mathbf{Z}_1 \\ \mathbf{Z}'_1\mathbf{X} & \mathbf{Z}'_1\mathbf{Z}_1\mathbf{M}_1 & \mathbf{Z}'_1\mathbf{Z}_1 + \mathbf{A}^{11}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\alpha} \\ \hat{\epsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{M}'\mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'_1\mathbf{y} \end{bmatrix}$$

Multibreed ASA Example

Number Animals total:	2,593,580
Number Genotyped	13,867
Number Imputed	2,579,713
Number Observations	1,959,890

PEV of Genotyped Animals



Genetic variance= 35 lb^2

Summary

- Genomic prediction is more accurate than parent average
 - Most accurate when close relatives in training
- Genomic prediction is an immature technology but is rapidly evolving
- It is now becoming routinely used by some North American beef cattle breed associations and this will provide the data required to speed up the evolution