

Genomic evaluation with multibreed and crossbred data

Ignacy Misztal and Daniela Lourenco
University of Georgia



Why multibreed evaluation?

- Desire for
 - Evaluation for all animals
 - Simplicity – one evaluation instead of many
- Hopes:
 - High accuracy for purebreds
 - High accuracy for crossbreds
 - Small or no biases compared to single-breed runs



Case studies

- Dairy
 - USDA/CDCB
 - Kiwi crosses
 - Australian crosses
 - Scandinavian
- Beef
 - Irish crosses (> 40 breeds)
 - International Genetic Solutions (Simmental++, > 10 breeds)
- Sheep
- Pigs

Methods in multibreed/crossbred evaluations

- Use SNP effects from purebred evaluation by breed percentage
- All breeds altogether with same SNP effects or common genomic relationship matrix
- Each breed as different but correlated trait
- If low accuracy - use imputed putative QTN from sequence data

Predictions across breeds

- Predictions based on one breed poor for other breeds (Olson et al., 2012)
- With combined reference data, predictions not better than with single breeds (Olson et al., 2012)
- In beef, similar results (Kachman et al., 2013)
- Limited improvement with sequence data (Meuwissen et al., 2021)
- Simulations – poor across breed predictivity even if identical QTLs (Stein et al., 2020)

Prediction for crossbreds by breed proportions

Jersey
SNP_{Jersey}

Holstein
SNP_{Holsteins}

BrownSwiss
SNP_{Brownswiss}

Crossbred, 20% Jersey, 50% Holstein, 30% Brownswiss

$$\text{GEBV} = \text{genotype} * (0.2 \text{ SNP}_{\text{Jersey}} + 0.5 \text{ SNP}_{\text{Holsteins}} + 0.3 \text{ SNP}_{\text{Brownswiss}})$$

Implicit assumption:

SNP effects are markers to QTLs, and their solutions are breed specific

Prediction by breed composition at CDCB (VanRaden et al., 2021)

Trait	Squared correlation		
	Genomic PTA	Diff	Parent average
Milk yield	0.46	+0.05	0.41
Fat yield	0.23	+0.05	0.18
Protein yield	0.26	+0.04	0.22
Productive life	0.01	+0.00	0.01
SCS	0.07	+0.03	0.04
Daughter pregnancy rate	0.02	+0.01	0.01
Cow conception rate ²	0.13	+0.02	0.11
Heifer conception rate ²	0.12	+0.03	0.09

Reliability barely higher than parent average, low for low h2 traits



Kiwi crosses (Winkelman et al., 2015)

- Holstein, Jerseys and crosses
- Approximate single-step
- Large crossbred populations

Trait	Breed	Accuracy
Milk	Friesian	0.72
	Jersey	0.70
	Friesian-Jersey	0.83
Fat	Friesian	0.75
	Jersey	0.70
	FJ cross	0.70

High accuracy for crossbreds

Multibreed prediction from different reference generations - Zoetis

- 90k Holsteins, 9k Jerseys, 1.6k crossbreds with genotypes and phenotypes
- Predictions by single-step with difference reference populations – one additive effect for all

Reference population	Predictive ability		
	Jersey	Holsteins	Cross
Jersey	0.48	0.13	0.50
Holstein	0.09	0.45	0.47
Cross	0.24	0.26	0.50
Jersey + Holstein	0.45	0.44	0.50
Proportions			0.32

(Stein et al 2021)

Reduced accuracy for Holsteins and Jerseys with Holstein+Jerseys
High (too high) accuracy for crossbreds based on few crossbreds

Single and multibreed prediction of US dairy with national data

- Single-step with unknown parent groups
- Up to 5 breeds

Breed	Phenotypes	Genotypes	Reliability *100 for protein	
			Single breed	Multibreed
Holstein	26.5M	3.4M	74	74
Jersey	2.4M	0.4M	69	67
Brown Swiss	0.3M	0.05M	45	21

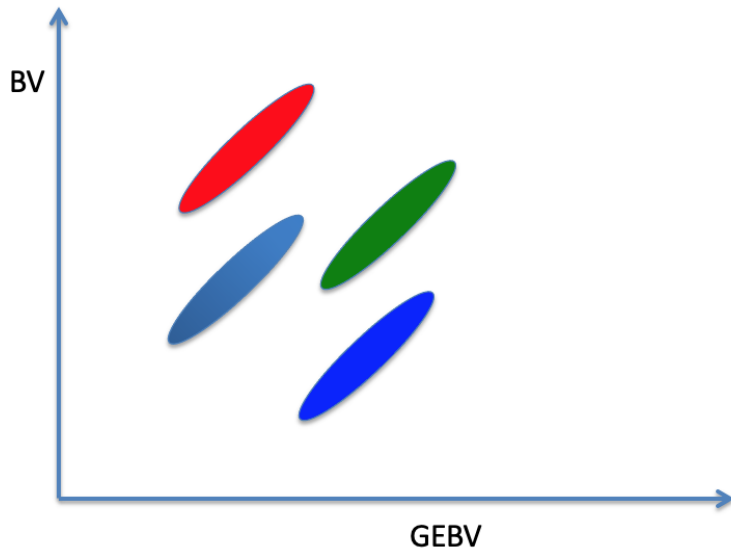
(Cesarani et al., 2021)

Reduction of accuracy for smaller breeds with multibreed data

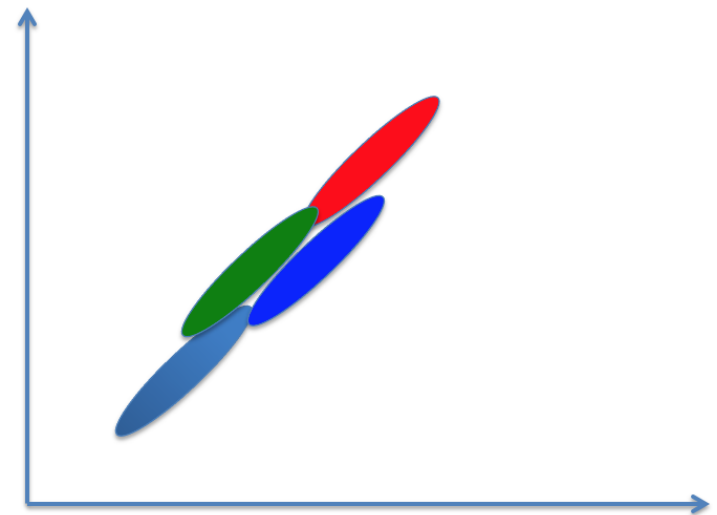
Issues with one genetic effect for all animals

- Different genetic levels per breed type
 - Fixed effects or unknown parent groups
 - Possibly heterosis and recombination loss
- Compatibility of genomic and pedigree information
 - Scaling of GRM
 - Special unknown parent groups (Masuda et al., 2021)
 - Metafounders (Legarra et al., 2017)

Distribution of BV and GEBV of multiple breeds



Plain analysis



Refined analysis

How many breeds per 50k SNP?

- Assume breeds independent
- If 10 breeds & 50k SNP, 5 k per breed - possible reduction in accuracy
- Simulation – 5 breeds with identical QTLs (Stein et al., 2019)
 - Accuracy within breed - ~ 0.70
 - Accuracy across breeds ~ 0.10
 - Reduction in accuracies compared to single-breeds with lower SNP density
- Irish beef evaluation - 47 breeds
- IGS evaluation - 12 breeds

Detailed accuracies for large multibreed systems not available

Predictability if breed types not in the model

- $\text{Pred} = \text{corr}(\text{GEBV}, y - Xb)$ $\text{Acc} = \text{Pred}/h$
- Correct model $y = Xb + g + BV + e$
g – breed type, e.g., F1, F2, reciprocal cross, for each breed combination
- If breed type not accounted for
 $y - xb \approx BV + g$
 $\text{Pred} \approx \text{corr}(\text{GEBV}, BV + g) \approx \text{corr}(\text{GEBV}, BV) + \text{corr}(\text{GEBV}, g)$
Possibly $\text{acc} = \text{pred}/h > 1$ (Bermann et al., 2021)

If breed type not in the model, predictivity may predict breed type, with poor accuracy within breed type

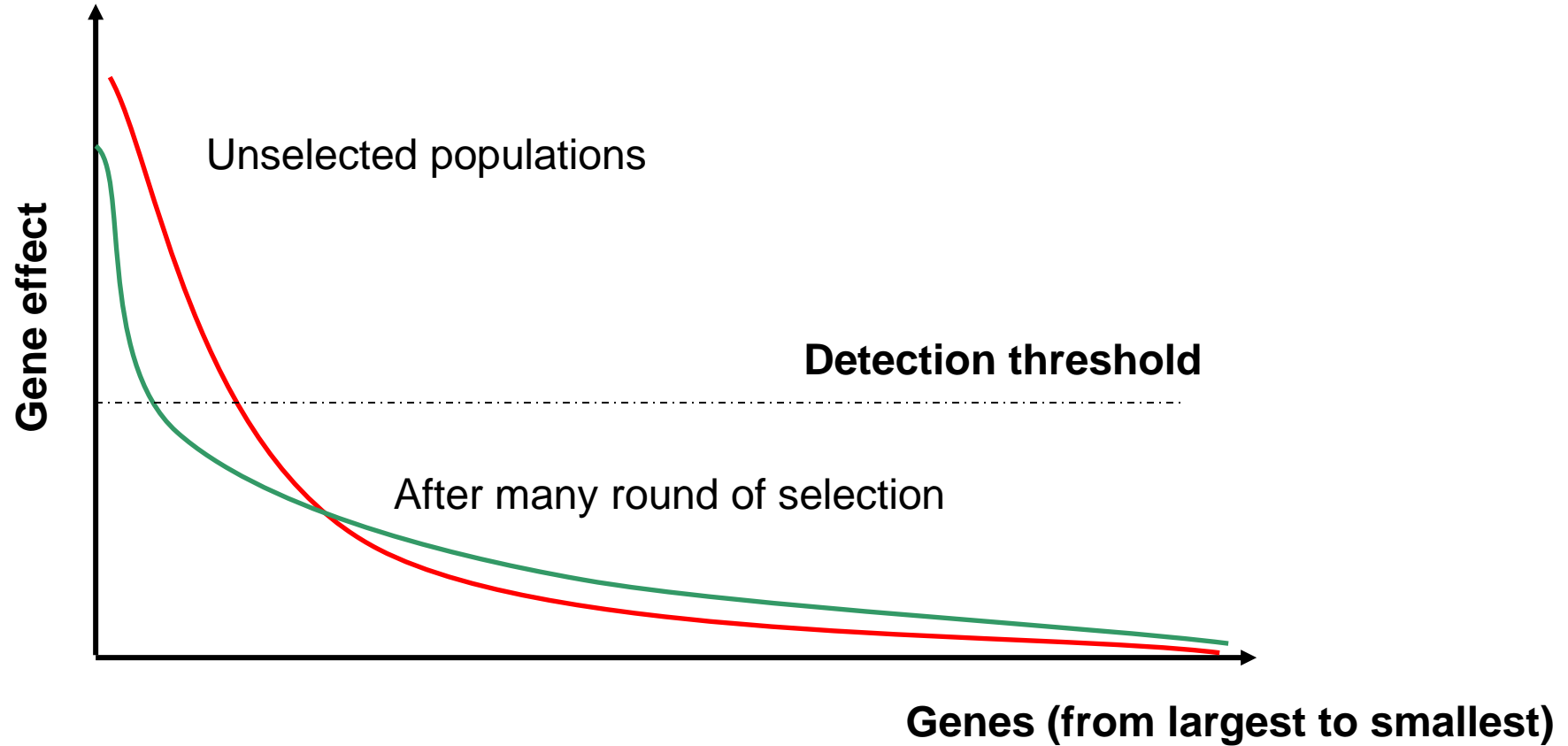
Gains in accuracy from including sequence data

- If same QTLs with same known substitution values across breeds - 100% reliability across breeds
- If same QTLs with unknown but similar substitution value
 - little benefit for GBLUP because of shrinkage
 - Potential higher reliability if QTN assigned higher variance, e.g., by BayesX
- Study using potential QTNs in Australia (Meuwisen et al, 2021)

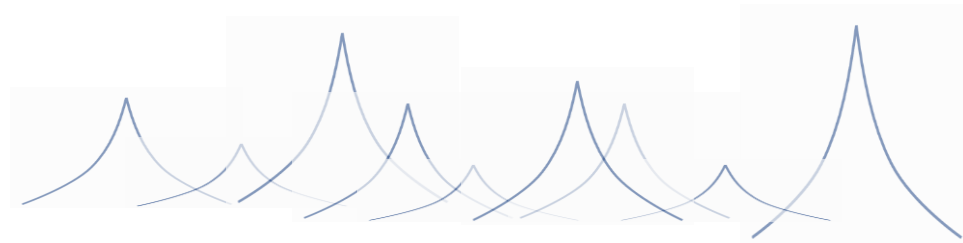
Breed	Reliability for Milk		
	GBLUP	BayesR	BayesGC
Holstein	0.71	+0.00	+0.02
Jersey	0.64	+0.01	+0.04
Australian Red Cow	0.23	+0.04	+0.03

Probably most QTN below detection limit

Distribution of QTL effects



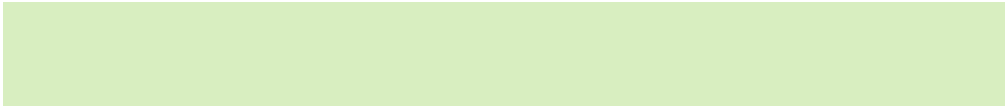
What is Manhattan plot composed of?



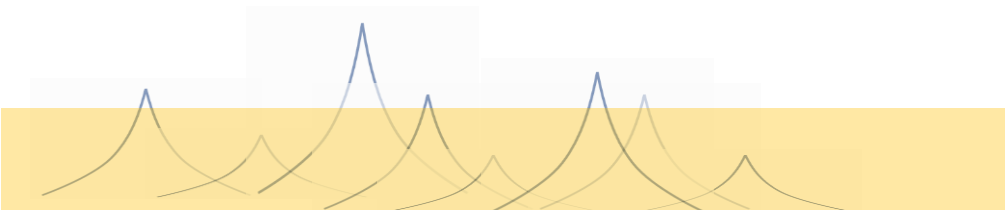
Signals from QTN



Relationships



Noise



Composite plot

Summary

- Little predictivity from breed to breed
- Small accuracy for crossbreds based on purebreds
- Sizeable influence of prediction of crossbreds by crossbreds
- Limited value of potential causative SNP



Carl Gustaf
1911-1972

Why

- Prediction primarily based on clusters of chromosome segments – not QTLs
- Good prediction of crossbreds require crossbred data
 - Careful with validation- include crossbred groups – $f_1 \neq f_2$
- Efficient use of QTN would require their exact identification + their variance – most QTNs most likely below detection limit
- With large data, GBLUP accounts for QTLs nearly completely

“Cornell” multibreed model for US beef - around 2000

J. Anim. Breed. Genet.

ORIGINAL ARTICLE

Multi-breed genetic evaluation in a Gelbvieh population

A. Legarra¹, J.K. Bertrand², T. Strabel³, R.L. Sapp⁴, J.P. Sánchez^{2,5} & I. Misztal²

1 INRA-Station d'Amélioration Génétique des Animaux, Castanet Tolosan Cedex, France

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

3 Department of Genetics and Animal Breeding, August Cieszkowski Agricultural University of Poznan, Poznan, Poland

4 Animal Genomics and Breeding Department, Monsanto Company, St Louis, MO, USA

5 Departamento de Producción Animal, Universidad de León, León, Spain

- Effects:
 - Heterosis + recombination loss
 - Unknown parent groups
- Poor estimates -priors on heterosis, recombination loss and breed average
- Poor trends- smoothing of unknown parent groups
- Too many small breeds – breeds replaced by 4 breed types
- Nearly same accuracy whether priors strong or weak

Conclusions

- Genomic prediction with crossbreds still a research topic
- High accuracies for group type probably require reference population of that group type
- Potential undesirable effects of putting many breeds together

Acknowledgements

