Single-step genomic evaluation using the CDCB data

Ignacy Misztal, D. Lourenco, S. Tsuruta, M. Bermann

University of Georgia

Alberto Cesarani **University of Sassari**

A. Legarra **INRA**

E.L. Nicolazzi **CDCB**

P. M. VanRaden **USDA**

Introduction

- CDCB plans to implement single-step evaluation
  - Preselection
  - Ability for more complex models
  - Simplifications

- Analyzes
  - ssGBLUP for yield traits in US Holsteins with unknown parent groups
  - Multibreed ssGBLUP using purebred animals
  - Multibreed ssGBLUP using crossbred animals
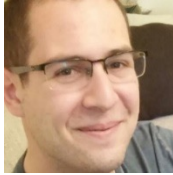 - Accuracy for genomic predictions

Steps for success

- Data and pedigree truncation useful
- UPG-exact (for A and A22) with few groups

$$\mathbf{A}^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}_{APY}^{-1} - \mathbf{A}_{22}^{-1} & -(-\mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ 0 & -\mathbf{Q}_2'(-\mathbf{A}_{22}^{-1}) & \mathbf{Q}_2'(-\mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix}$$

- Inversion of G by APY (~15k core)

ssGBLUP evaluation for Holsteins

- US Holstein data up to December 2018

Alberto Cesarani

Phenotype cut-off scenario	Records		Genotypes	Animals in pedigree	
	N	Cows		Depth=3	Depth=2
Pheno1980	77.8 M	31.5 M	862 K	40.5 M	39.9 M

What truncation?

A. Cesarani,[1]* Y. Masuda,[1] S. Tsuruta,[1] E. L. Nicolazzi,[2] P. M. VanRaden,[3] D. Lourenco,[1] and I. Misztal[1]
[1]Department of Animal and Dairy Science, University of Georgia, Athens 30602
[2]Council on Dairy Cattle Breeding, Bowie, MD 20716
[3]Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

- Data cutoff 1980, 1990, 2000

- Pedigree cutoff 2, 3 generations past phenotypes

- Same R^2, $b_1 = 1 \pm 0.05$

- If old bulls need to be included, include only genotypes

Multibreed evaluation
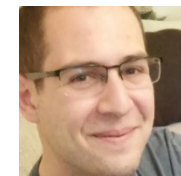
Breed	Phenotypes		Animals	
	N	**Cows**	**Genotypes**	**Total**
Ayrshire	116k	47k	9.2k	94k
Brown Swiss	328k	138k	47k	292k
Guernsey	129k	58k	5k	100k
Holstein	40.3M	17.5M	3.4M	26.6M
Jersey	4.1M	1.7M	427k	2.5M
MultiBreed	**45M**	**19.4M**	**3.9M**	**29.5M**

Alberto
Cesarani

About 2 days computing

Results for multiple breed evaluation

- Poor reliability for smaller breeds if random core

- Good reliability if 5k core per small breed

- No reduction of accuracy compared to single-breed analyses
  - Increase for Jerseys

A. Cesarani,[1]* D. Lourenco,[1] S. Tsuruta,[1] A. Legarra,[2] E. L. Nicolazzi,[3] P. M. VanRaden,[4] and I. Misztal[1]
[1]Department of Animal and Dairy Science, University of Georgia, Athens 30602
[2]INRA, UMR1388 GenPhySE, Castanet-Tolosan, France, 31320
[3]Council on Dairy Cattle Breeding, Bowie, MD 20716
[4]Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705

Comparison with official CDCB results

	Multistep	ssGBLUP
Interbull data	yes	no
SNP weighting	yes	no
Realized Reliability $\text{Corr}(DYD_{2021}, PTA_{2017})^2$	0.60	0.63
After age adjustment	0.69	0.68
Stability $\text{Corr}(PTA_{2021}, PTA_{2017})^2$	0.74	0.81
After age adjustment	0.80	0.85

- Multistep requires age adjustments
- ssGBLUP more accurate with less data
- ssGBLUP more stable
- DYD may not be good benchmark

Comparison of single-step and multi-step evaluations for U.S. milk, fat, and protein

R.R. Mota[1]*, A. Cesarani[2], and P.M. VanRaden[3]

[1] Council on Dairy Cattle Breeding, 4201 Northview Dr Ste 302, Bowie, MD, 20716, USA; *rodrigo.mota@uscdcb.com; [2] University of Georgia, 425 River Rd, Athens, GA, 30602, USA; [3] USDA Animal Genomics and Improvement Lab, Bldg 5 BARC-West, Beltsville, MD, 20705, USA

Multibreed ssGBLUP using crossbred animals

- Holstein, Jersey and their crosses

- 47M records from 20.3M of cows, 1.4M genotypes

- 4 groups of validation cows

  - HO purebred = HO animals with both sire and dam HO

  - JE purebred = JE animals with both sire and dam JE

  - Crosses = HO or JE animals with at least one parent of the opposite breed

  - Crosses 100 het = crosses with 100 heterosis

Group	Trait	N	ssGBLUP		
			Pred[1]	Stab[2]	b1[3]
	MY		0.55	0.95	1.04
HO pure	FY	688,985	0.56	0.96	0.98
	PY		0.53	0.95	1.01
	MY		0.52	0.94	1.08
JE pure	FY	119,743	0.46	0.91	0.90
	PY		0.51	0.93	1.04
	MY		0.68	0.98	1.03
Crosses	FY	1,857	0.51	0.93	1.03
	PY		0.59	0.96	1.09
	MY		0.66	0.98	1.02
Cross 100 het	FY	1,378	0.51	0.95	1.00
	PY		0.58	0.96	1.06

Why higher predictivity for crossbreds than for purebreds?

We model crossbreds as: PTA

Pred=corr(y-.. , PTA)

Different types of crossbreds (F1, F2, reciprocal HHJ and JJH,…)

Perhaps we could model as: breed_type + PTA

Pred=corr(y-..-breed_type, PTA-breed_type)

Predictivity inflated if breed effect missing (Bermann et al., 2021)

Information on crosses in data cryptic

Timings for crossbred studies

- Genomic preprocessing (pregsf90) 10h
- ssGBLUP (blup90iod2) 17h
- Approx. accuracies (accf90gs2) 3h

Genomic accuracy by accf90gs2 for Holsteins

Validation bulls (N=3407)

Correlation = 0.84

	ssGBLUP	CDCB
Min	82.9	88.3
Mean±SD	95.5±1.7	95.0±2.6
Max	99.9	99.9

Genomic accuracy: true vs approximated

Purebred Ayrshire

- 116,674 records

- 94,500 animals

- 9,202 genotypes

True accuracy = inversion of MME

Approximated accuracy = accf90GS2

Category	Correlation	b0	b1
Ungenotyped			
MY			0.98
FY	0.99	0.00	0.96
PY			0.99
Genotyped			
MY			0.96
FY	0.99	0.02	0.96
PY			0.97

Conclusions

- Removing phenotypes before 2000 does not diminish accuracy
- Multibreed analyses do not reduce accuracy if appropriate choice of core animals
- High predictability for crossbreds
  - Useful in practice
  - Theoretical issues
- Higher stability of SS than multistep PTA
- Any results with BLUP DYD suspect
- Reasonable computing time with 4M genotypes