Single- and multi-breed ssGBLUP evaluations with sequence data for over 200k pigs

Daniela Lourenco

S. Jang, R. Ros-Freixedes, J. Hickey, C.Y. Chen, W. Herring, I. Misztal

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

September 5, 2022

Affordability of sequence data

Euan Ashley (2022 @Nanopore Conference)

Largest pig sequence data

Line	Genotyped individuals	Sequenced individuals	Sequenced/Imputed
ML1	76k	1,365	76k
ML2	67k	1,491	67k
TL1	60k	731	60k
TL2	42k	760	42k
TL3	105k	1,865	105k
TL4	29k	381	29k

Total = 379k

Objectives

- Accuracy of GEBV with preselected SNP from sequence data
- Large-scale ssGBLUP
- Single-line and multi-line evaluations
- Compare ssGBLUP with BayesR from Roslin

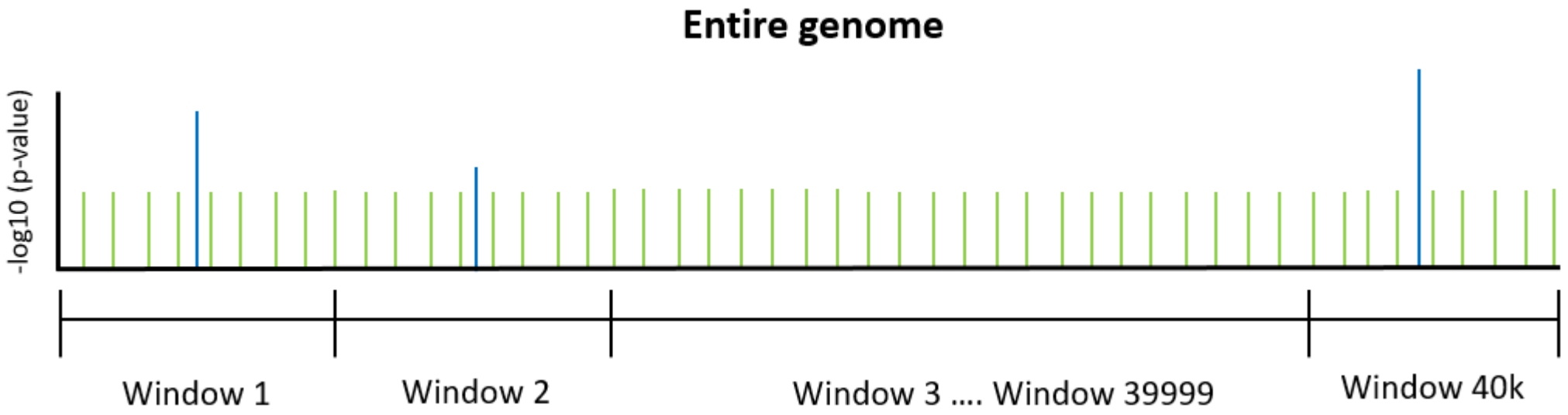
Data

Lines	ADFI	ADG	BF	ADGX	BFX	Animals in pedigree	Sequenced/ Imputed
TL1	35k	0.36M	0.34M	150k	149k	1.13M	60k
TL2	40k	0.30M	0.30M	158k	156k	0.84M	42k
TL3	64k	0.94M	0.86M	299k	247k	3.14M	105k
MLE	140k	1.60M	1.50M	578k	525k	> 5M	207k

- 15M to 20M variants
  - 9.9M segregated across lines

SNP preselection

- **Top 40k**

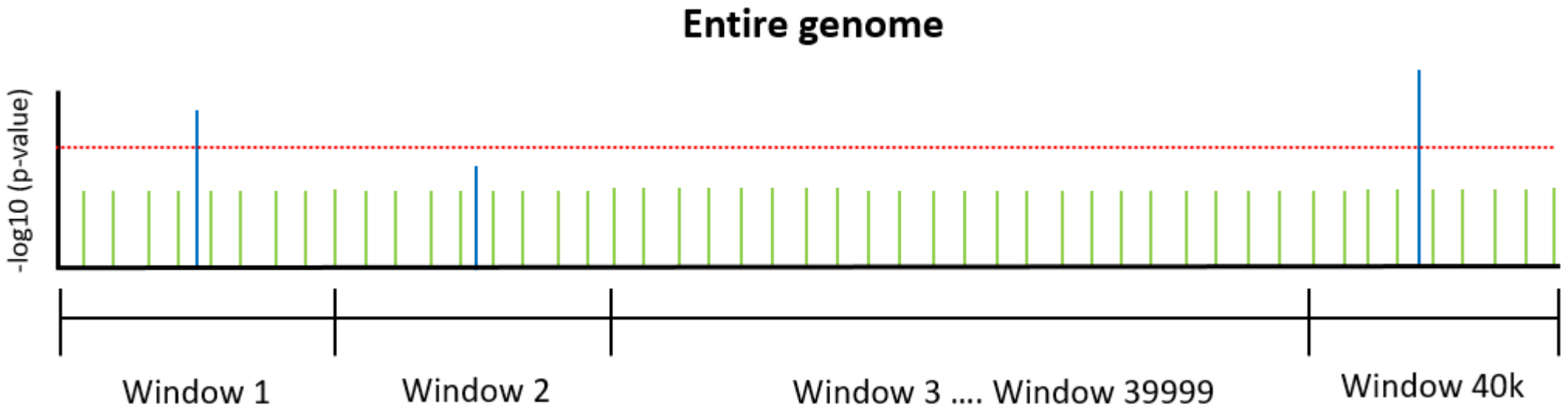


Extracting only 40k SNP: Similar number as the regular SNP chip (~40k)

Ros-Freixedes et al., 2022

SNP preselection

- **Chip+Sign**



Extracting only significant ones + 40k SNP chip

Ros-Freixedes et al., 2022

Analyses

- ssGBLUP
- Single-line: models by PIC
- Multi-line:
  - Line-specific fixed effects
  - Different genetic base

- QP-transformation for $\mathbf{H}^{-1}$ (Misztal et al., 2013)

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

- Altered QP-transformation for $\mathbf{H}^{-1}$ (Tsuruta et al., 2019)

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

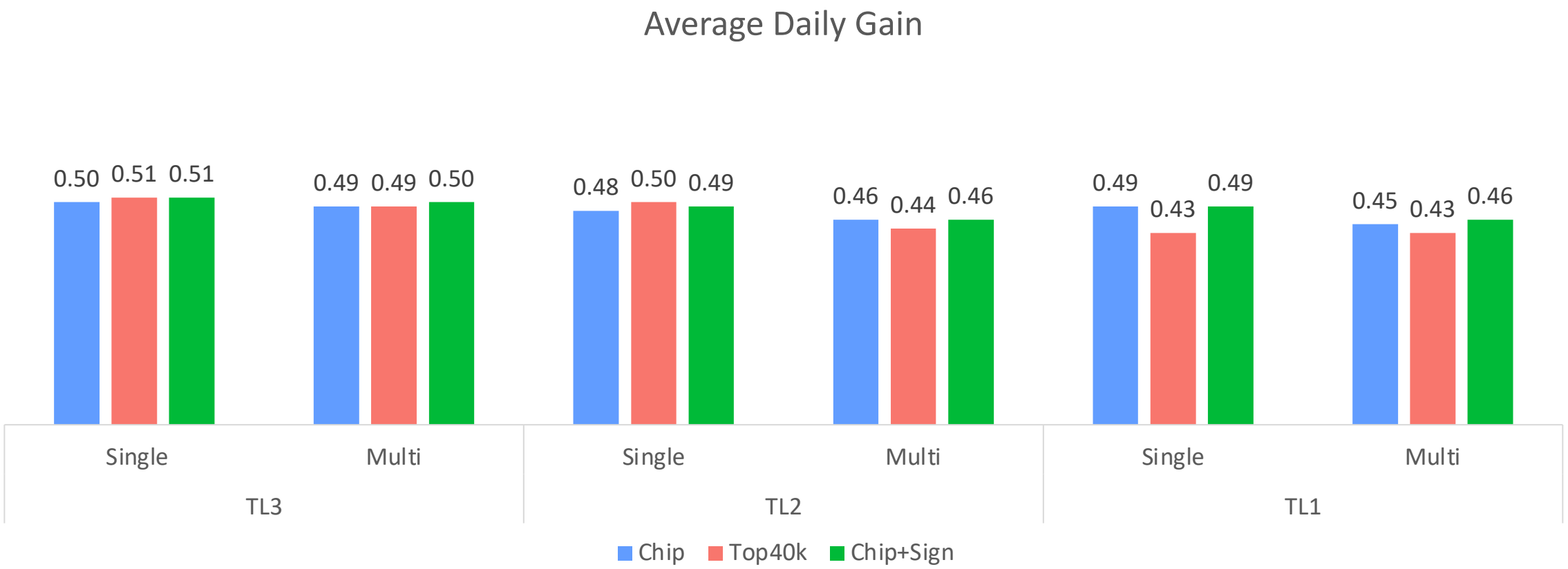
- Metafounders (Legarra et al. (2015)

$$\mathbf{H}^{\Gamma-1} = \mathbf{A}^{\Gamma-1} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}_{05}^{-1} - \mathbf{A}_{22}^{\Gamma-1} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

No difference!

Results – SNP preselection scenarios

- Prediction accuracy = cor(DEBV, GEBV)

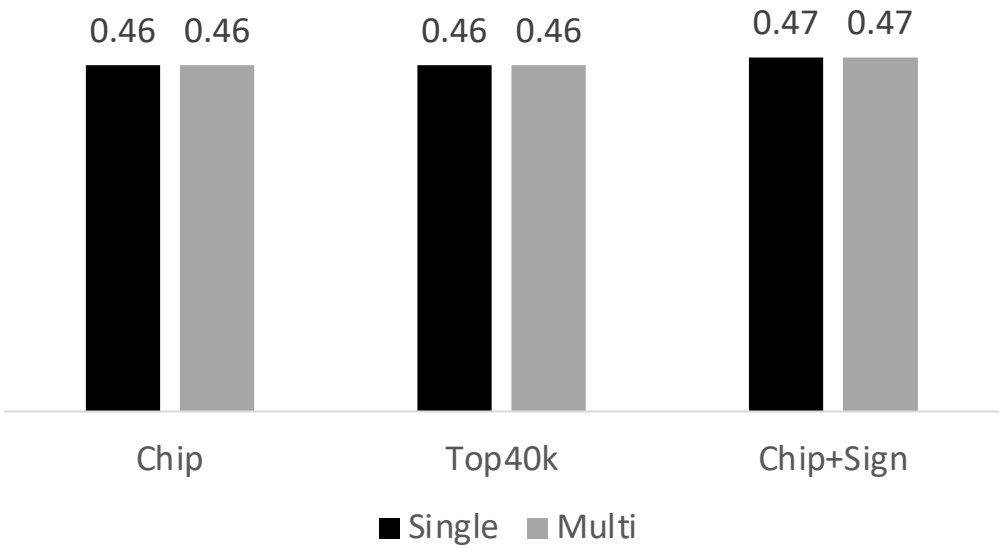


Results – Single vs. Multi-line all traits

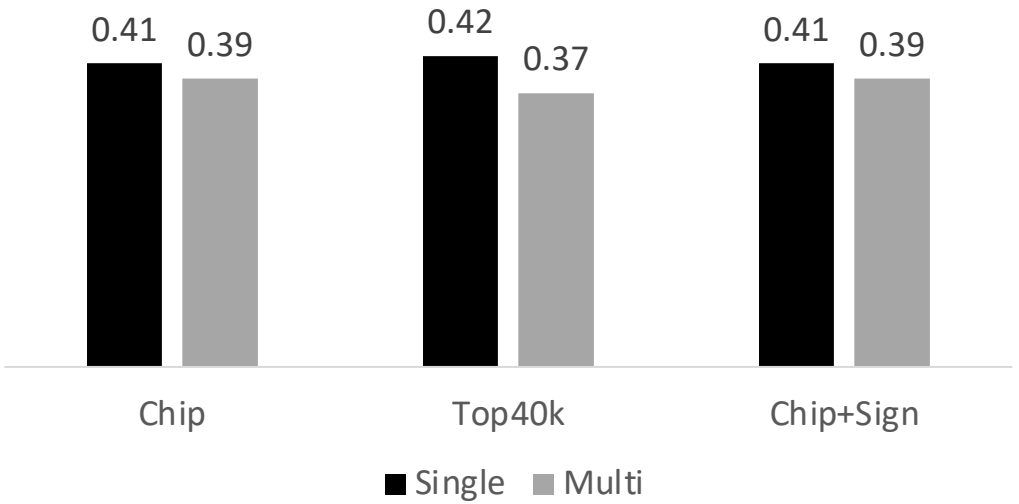
- Prediction accuracy = cor(DEBV, GEBV)

Multi-line GWAS and predictions dominated by TL3

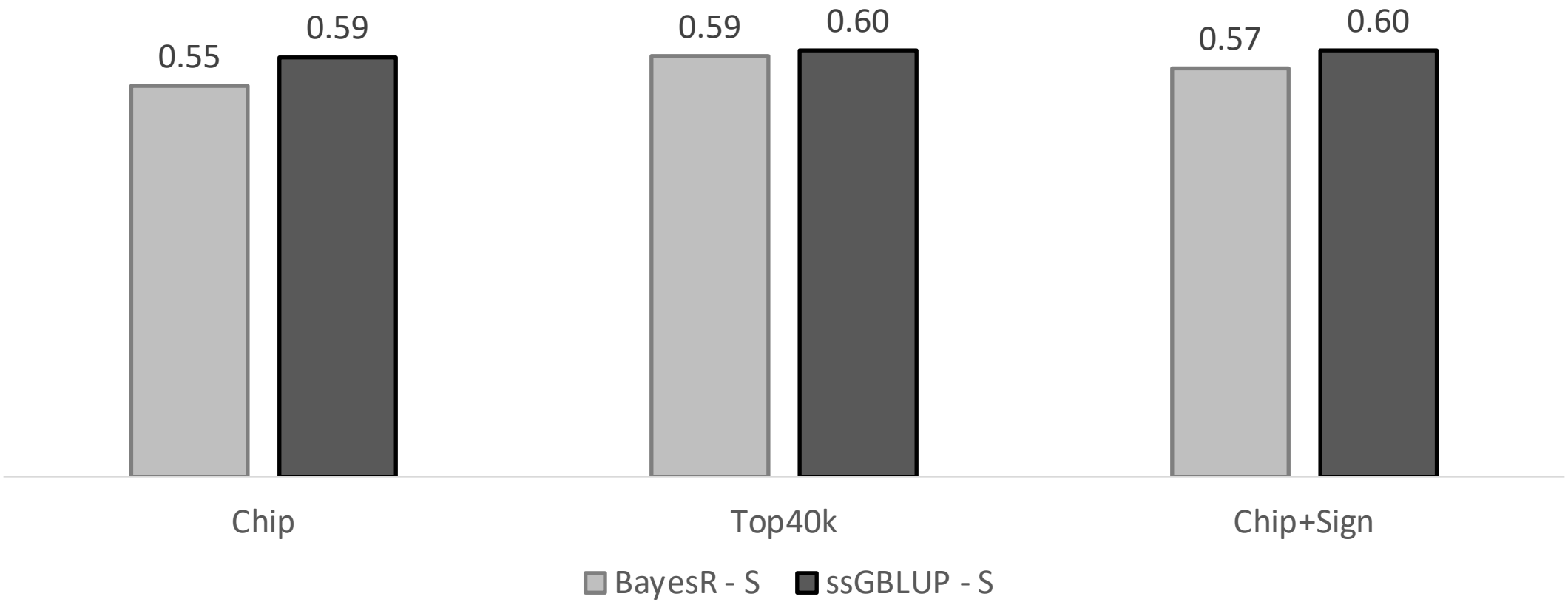
TL3 105k



TL2 42k



Single-line ssGBLUP vs. BayesR



BayesR from Ros-Freixedes et al., 2022

ssGBLUP vs. BayesA in Dairy

VanRaden et al. Genet Sel Evol (2017) 49:32
DOI 10.1186/s12711-017-0307-4

Genetics Selection Evolution

RESEARCH ARTICLE Open Access

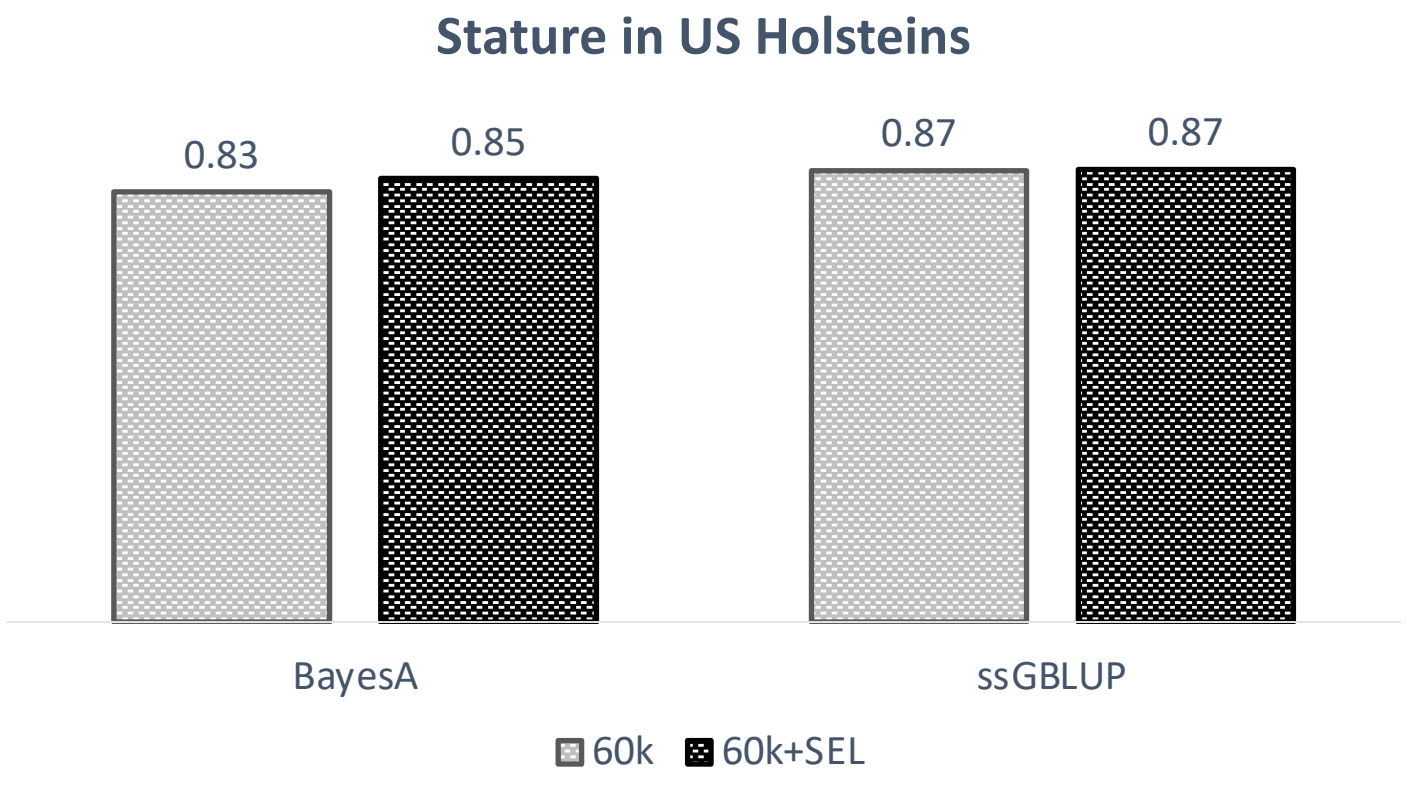
Selecting sequence variants to improve genomic predictions for dairy cattle

Paul M. VanRaden[1*], Melvin E. Tooker[1], Jeffrey R. O'Connell[2], John B. Cole[1] and Derek M. Bickhart[1]

J. Dairy Sci. 102:10012–10019
https://doi.org/10.3168/jds.2019-16262

Alternative SNP weighting for single-step genomic best linear unbiased predictor evaluation of stature in US Holsteins in the presence of selected sequence variants

B. O. Fragomeni,[1]* D. A. L. Lourenco,[2] A. Legarra,[3] P. M. VanRaden,[4] and I. Misztal[2]
[1]Department of Animal Science, University of Connecticut, Storrs-Mansfield 06269
[2]Department of Animal and Dairy Science, University of Georgia, Athens 30602
[3]Institut National de la Recherche Agronomique, UMR1388 GenPhySE, Castanet Tolosan, France 31326
[4]Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705



Why small gains in accuracy with sequence?

- Genomic selection acts in segments and not individual SNP
  - *Me* = 4NeL Stam (1980)
  - *Me*= 4NeL = Eig98

GENETICS | GENOMIC SELECTION

The Dimensionality of Genomic Information and Its Effect on Genomic Prediction

Ivan Pocrnic,*,[1] Daniela A. L. Lourenco,* Yutaka Masuda,* Andres Legarra,[†] and Ignacy Misztal*
*Department of Animal and Dairy Science, University of Georgia, Athens, Georgia 30602, and [†]Institut National de la Recherche Agronomique, Génétique, Physiologie et Systèmes d'Elevage, F-31326 Castanet-Tolosan, France

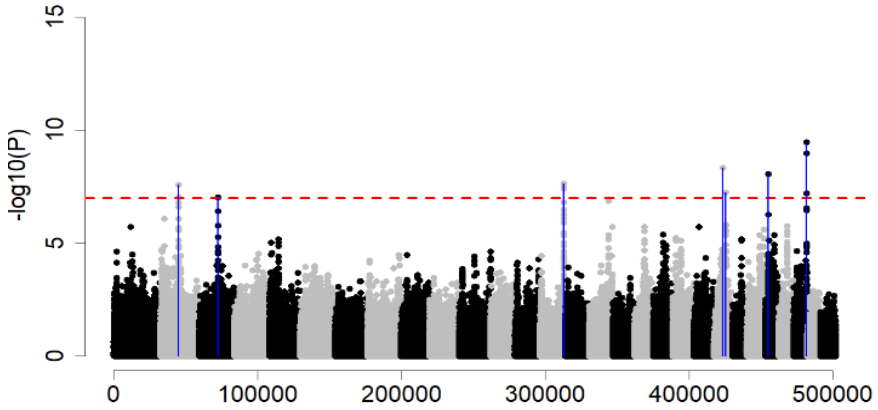
Species/breed	Ne	Segments
Pigs	48	4.1k
Chicken	44	4.2k
Jersey	101	11.5k
Angus	113	10.6k
Holstein	149	14k

Why small gains in accuracy with sequence?

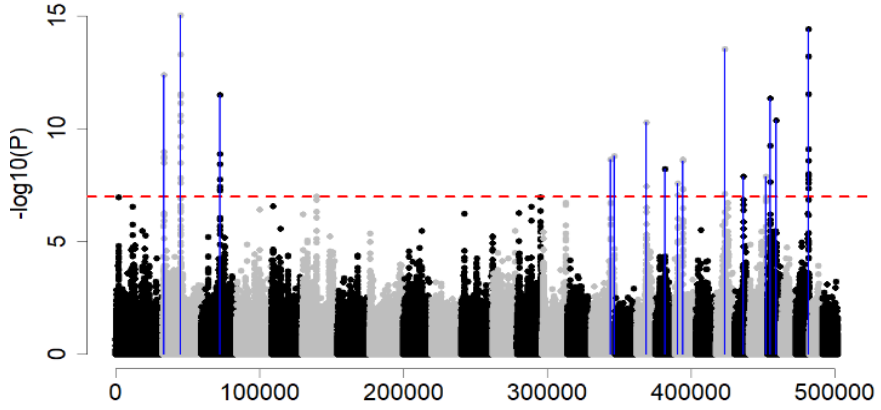
- Amount of information to identify causative variants

Ne=200 QTN=2000

Eig98 N=15,200



All N=30,000



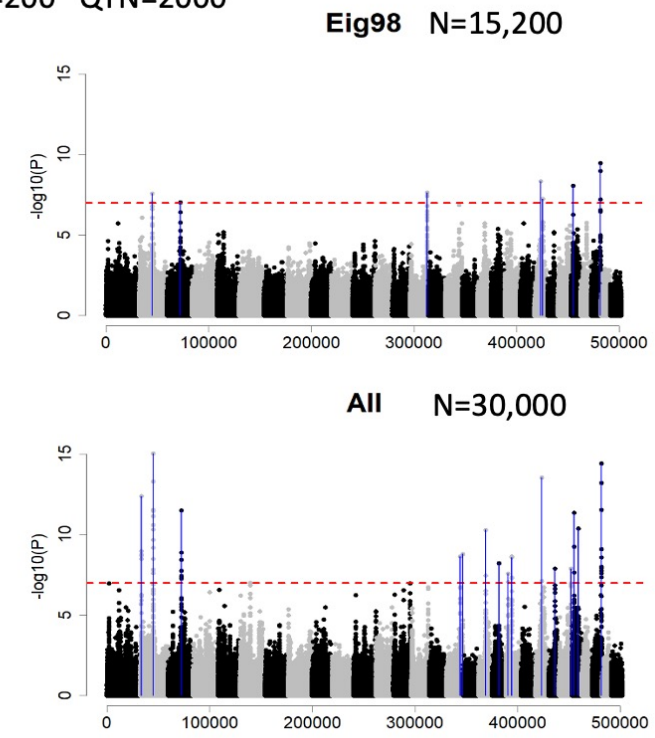
Jang et al.
(in preparation)

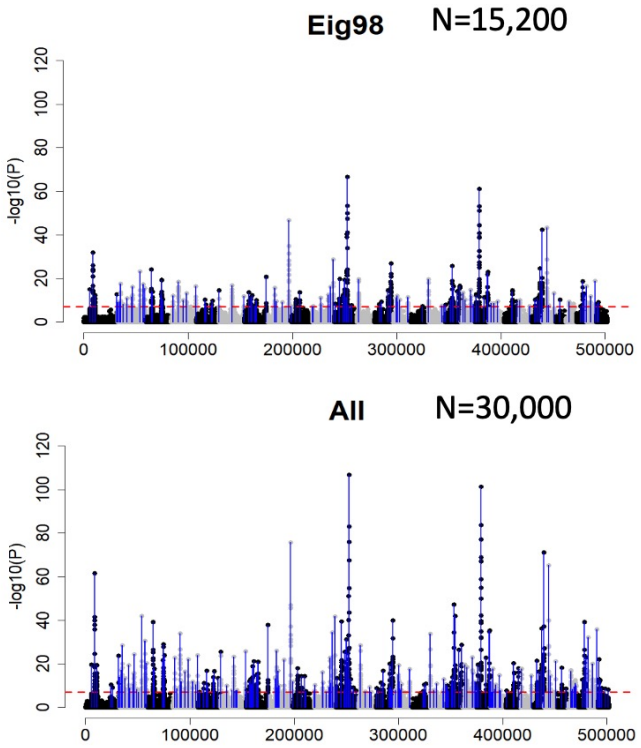
Why small gains in accuracy with sequence?

- Amount of information to identify causative variants
  - Animal with lots of information
    - GEBV accuracy ~ 0.99
  - GEBV backsolved to SNP effects
  - GWAS resolution with sample size = *Me* = Eig98 animals with almost perfect accuracy

Jang et al.
(in preparation)

Ne=200 QTN=2000





Lots of records for each genotyped animal

Conclusions

- Amount and sources of data not enough
  - Poor identification of true causative variants in GWAS
- Variants in the chip are covering the chromosome segments
- Small to no increase in accuracy with sequence data
  - Over 200k sequenced/imputed animals
  - Multi-line (multi-breed) evaluations
- Smaller gains with ssGBLUP compared to BayesX
  - Amount of data in ssGBLUP overwhelms any prior information
  - ssGBLUP accuracy is higher

http://nce.ads.uga.edu

Genus

PIC

Animal Breeding and Genetics Group

College of Agricultural & Environmental Sciences

UNIVERSITY OF GEORGIA

- Postdocs
- Assistant Professor