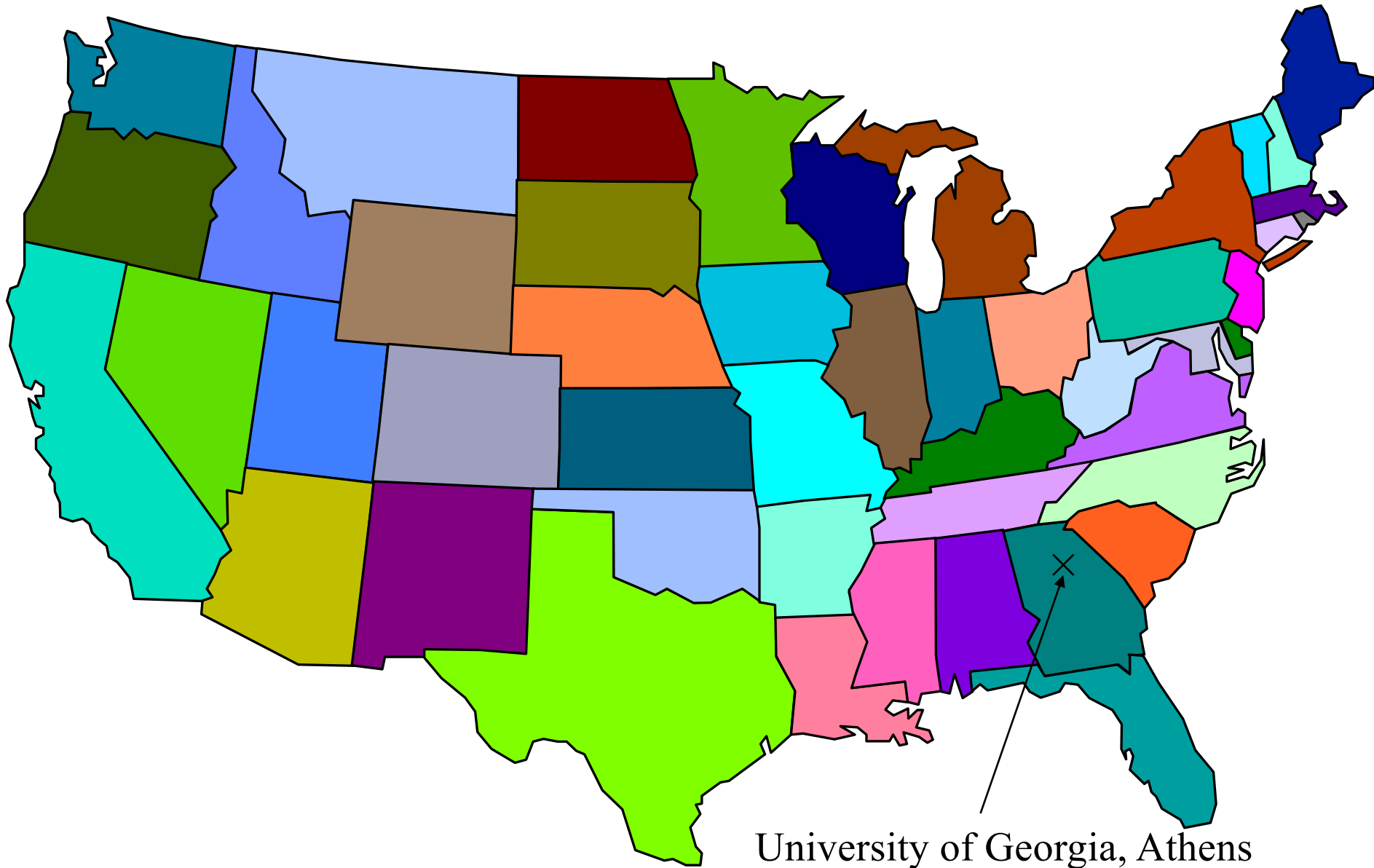


Animal breeding projects at UGA

Ignacy Misztal and Daniela
Lourenco

University of Georgia



University of Georgia, Athens

UGA projects in Animal Breeding -
Uppsala 2019



THE UNIVERSITY OF GEORGIA



Law School

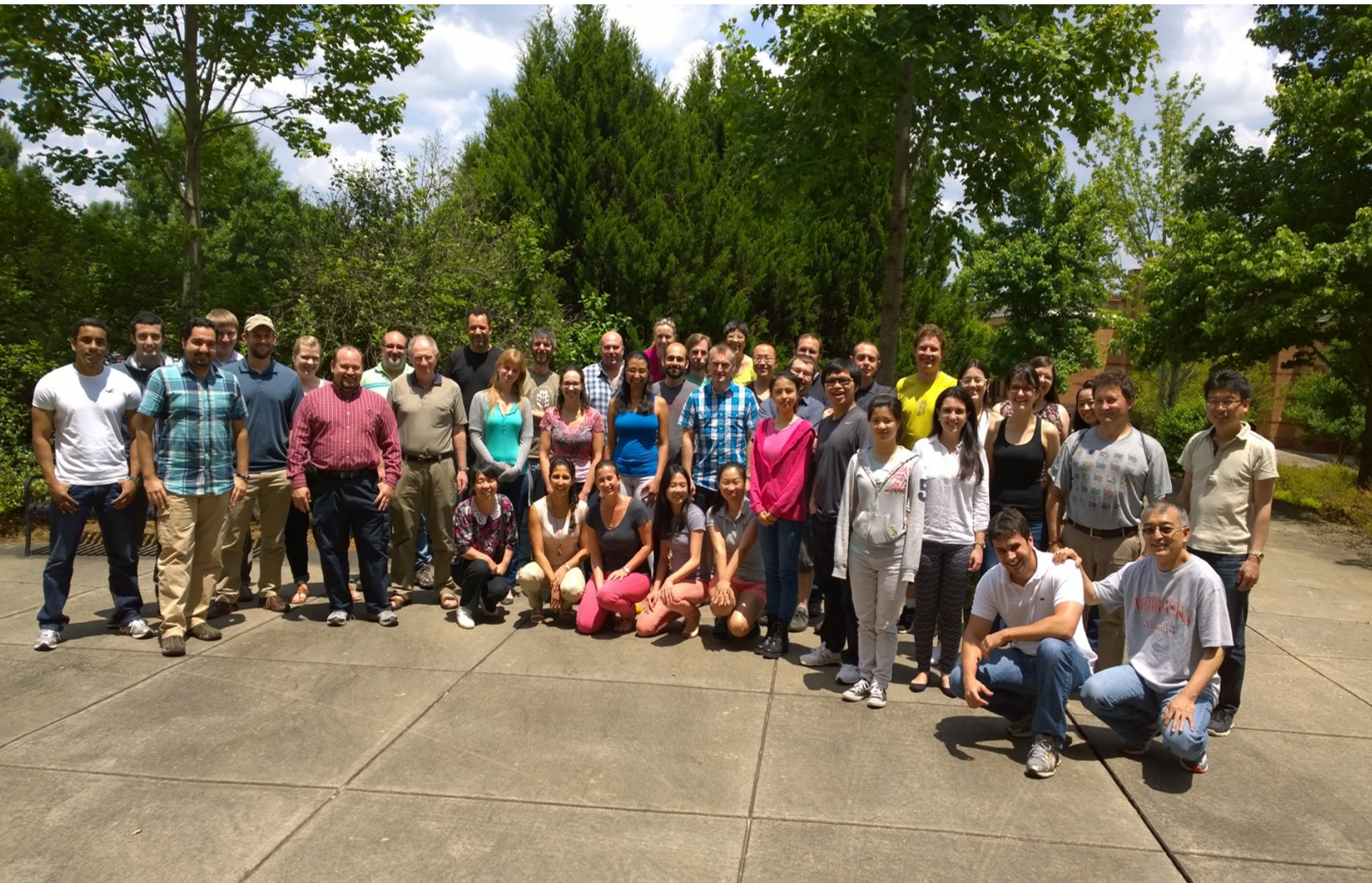


Georgia Museum of Art



Edgar L. Rhodes
Animal Science Center

UGA projects in Animal Breeding -
Uppsala 2019



UGA projects in Animal Breeding -
Uppsala 2019

Research in Breeding & Genetics lab at UGA

- 15-25 people (Postdocs + grad students+ visitors)
- Focus on genetic evaluation - 0.99 projects avoided
- Sponsors across species
 - Holsteins Assoc.
 - Zoetis (dairy and beef)
 - Angus Assoc +
 - Major pig companies (PIC, Smithfield, Maschoffs, DNA Genetics)
 - Cobb (broiler chicken)
 - USDA (Dairy, beef and fish)
- Comprehensive access to large data across species

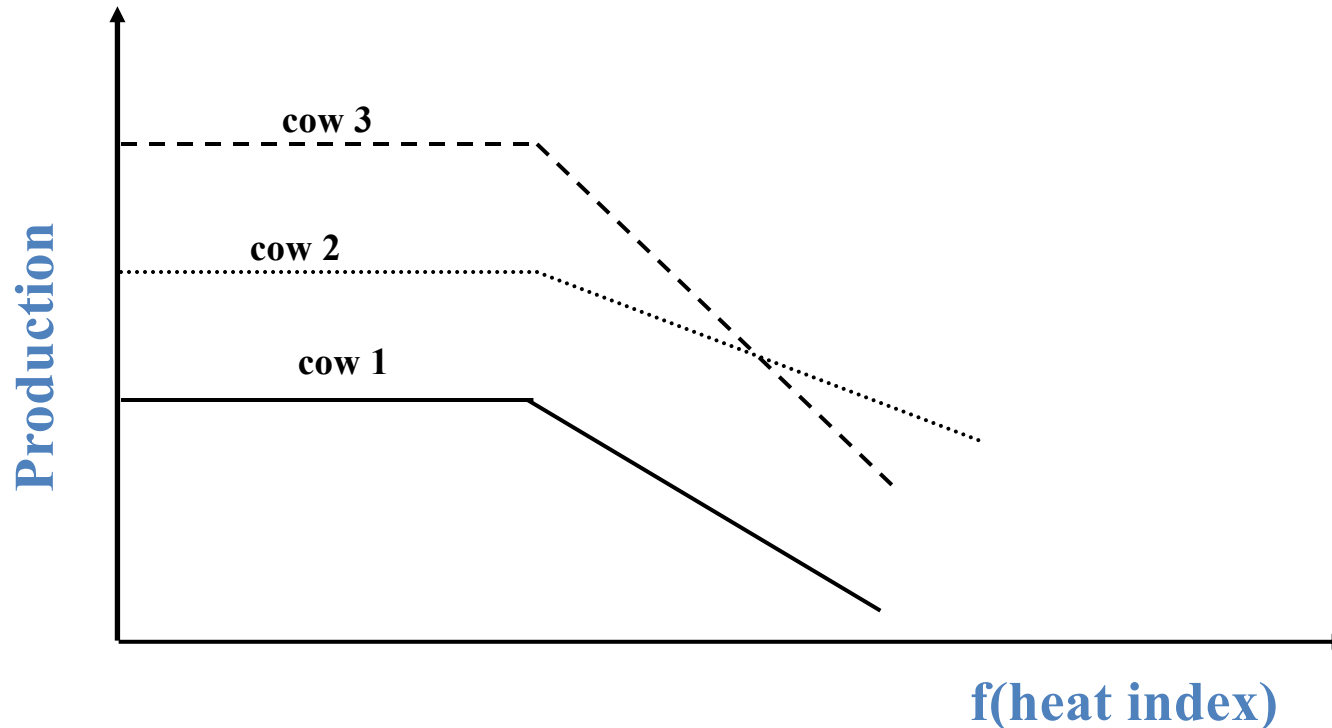
Projects across species

- Computing methods for genomic evaluation
- Mortality, morbidity and fertility
- GxE across managements, regions and countries
- Purebred and crossbred performance
- Changes in genetic parameters over time
- Modeling unknown parents for complex pedigree structure
- Heat stress
- Genomic preselection
- Competition effects
- Sexual dimorphism

Some questions

- Modeling heat stress
- Impact of low dimensionality of genomic information in farm species
- Selection as optimization - what are the losers?

Model for heat stress



Breeding value: $BV = a + f(\text{THI}) * v$

a – regular breeding value v – heat-tolerance breeding value

$f(\text{THI})$ – function of temperature humidity index

Heat stress

National evaluation by random-regression test-day model



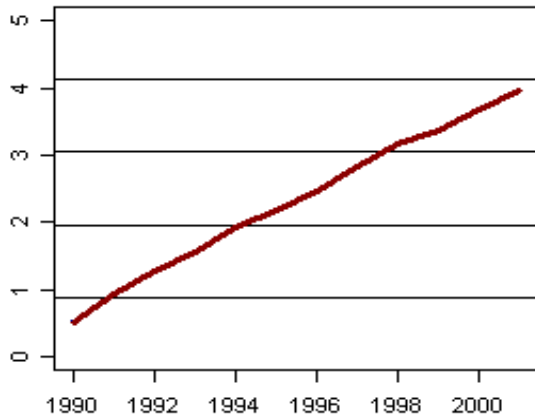
Ignacio AGUILAR*^{1,2}, Ignacy MISZTAL¹ and Shogo TSURUTA¹

¹ Animal and Dairy Science Department, University of Georgia

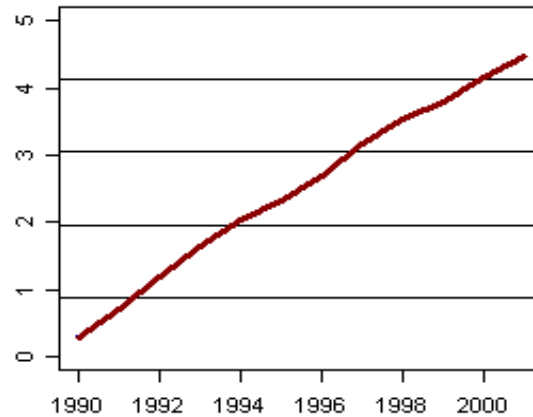
² Instituto Nacional de Investigación Agropecuaria, Las Brujas, Uruguay

Genetic trends of daily milk yield for 3 parities – regular effect

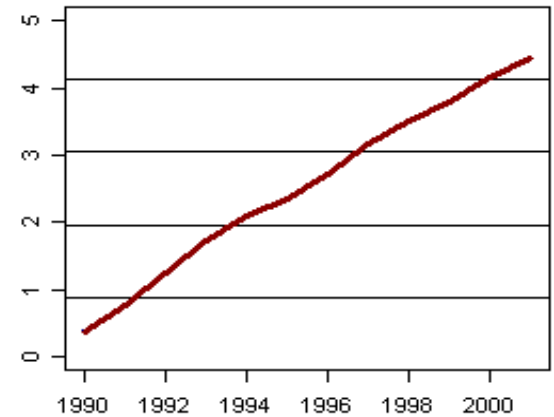
First



Second

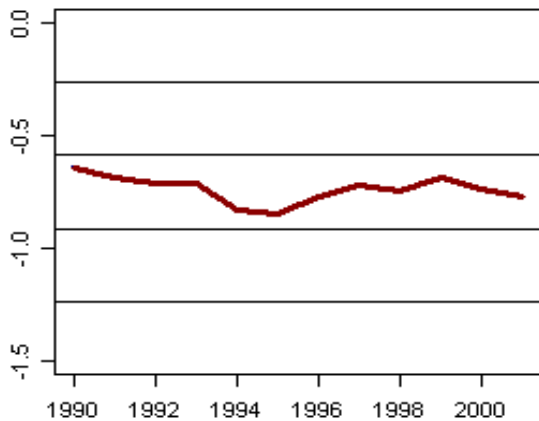


Third

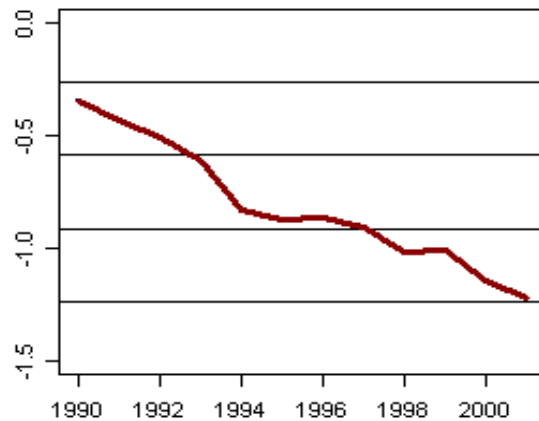


Genetic trends for heat stress effect at 5.5° C over the threshold

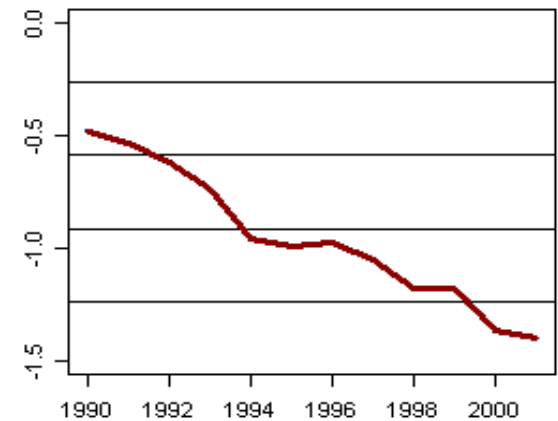
First



Second



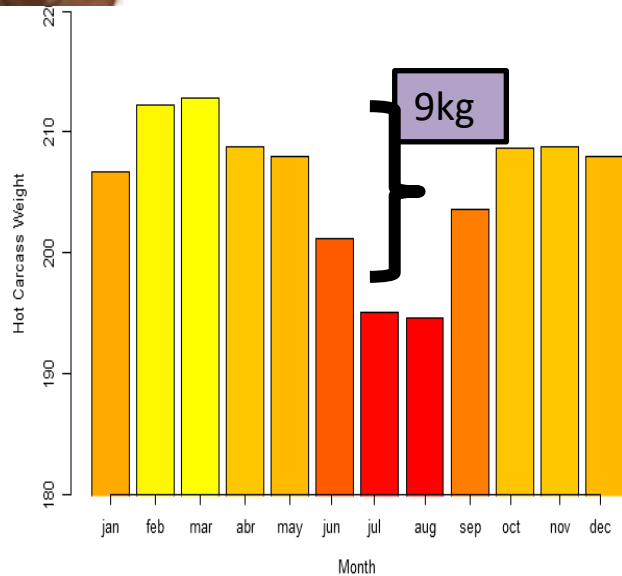
Third



Heat stress in purebred and crossbred pigs

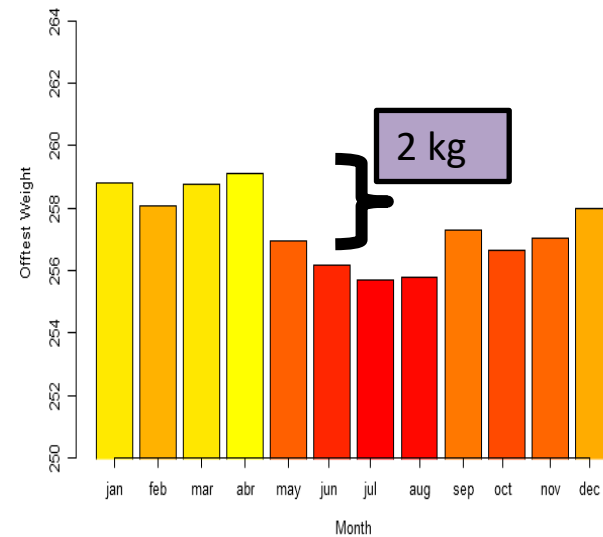


Crossbred



Fragomeni et al., 2016

Purebred



Better environment almost eliminates heat stress

Status of genomic selection

- Widespread genotyping
 - > 1.5 million in US Holsteins
 - > 0.5 million in Angus
 - > 30k/line in broilers and pigs
- Genomic evaluation
 - Single-step method in pigs, beef and broilers
 - Multistep method in most dairy evaluations
 - preselection bias
- Interest in sequence data
 - Identification of recessive genes
 - Identification of causative SNP
 - Targets for gene editing

Questions in genomic selection

- SNP are genes, markers or something else?
- Good accuracy at 30k SNP , standard 50-60k, a bit better at 700k
 - What is magic with 50K?
 - Why not more noise at 600K
 - Causative SNP?
- OK accuracy with few genotyped animals 1k-2k
 - Rise with extra genotypes slow
 - Discrepancy between simulation and field-data results

Inverse of matrix that combines pedigree and genomic relationships

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

Aguilar et al., 2010

$$\text{EBV}_{\text{young}} = w_1 \text{PA} + w_2 \text{DGV} - w_3 \text{PI}$$



J. Dairy Sci. 93:743–752

doi:10.3168/jds.2009-2730

© American Dairy Science Association®, 2010.

Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score¹

I. Aguilar,^{*†‡} I. Misztal,^{*} D. L. Johnson,[‡] A. Legarra,[§] S. Tsuruta,^{*} and T. J. Lawlor[#]

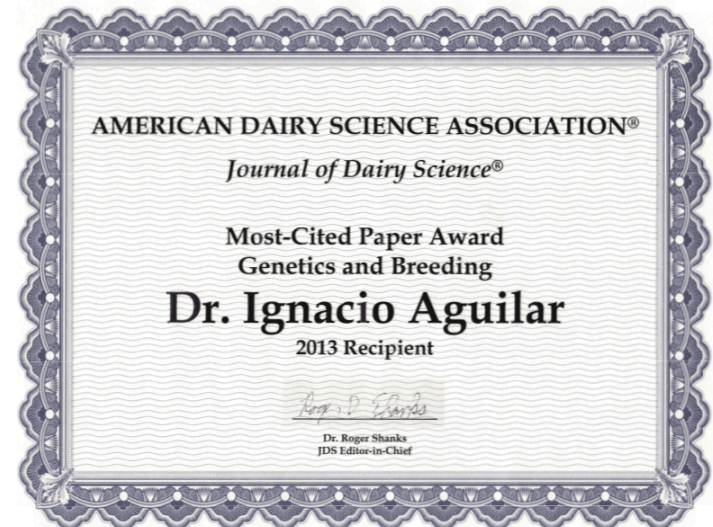
^{*}Animal and Dairy Science Department, University of Georgia, Athens 30602

[†]Instituto Nacional de Investigación Agropecuaria, Las Brujas 90200, Uruguay

[‡]Livestock Improvement Corp., Private Bag 3016, Hamilton 3240, New Zealand

[§]INRA, UR631 SAGA, BP 52627, 32326 Castanet-Tolosan, France

[#]Holstein Association USA Inc., Brattleboro, VT 05302-0808



Renumbering
RENUMF90

Computing of extra matrices
PreGSF90

BLUP in memory
BLUPF90

BLUP – iteration on data
BLUP90IODF
CBLUP90IOD

Variance component estimation
REMLF90 AIREMLF90
GIBBS2F90 THRGIBBS2F90

Approximate accuracies
ACCF90

Sample analysis
POSTGIBBSF90

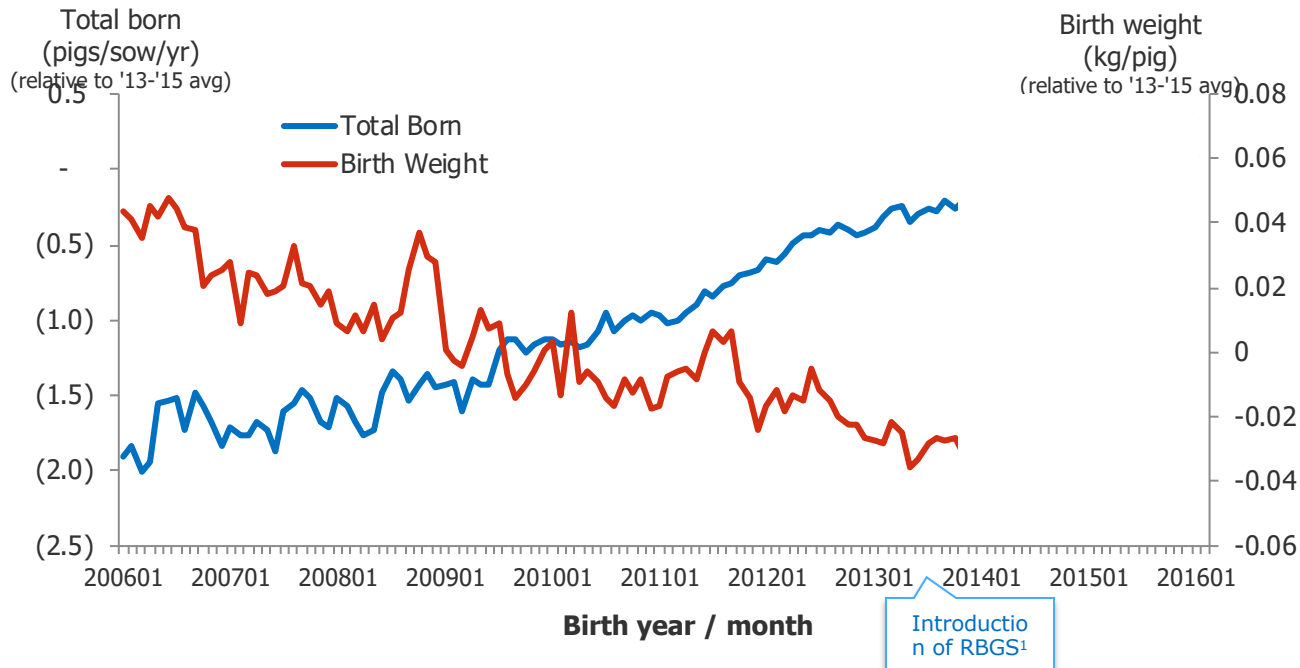
Predictions via SNP
PredGSF90

GEBV to SNP conversions
GWAS
PostGSF90



Improvement of total born and piglet birth weight

Trend: genetic improvement in birth weight and total born
(PIC Genetic Nucleus)



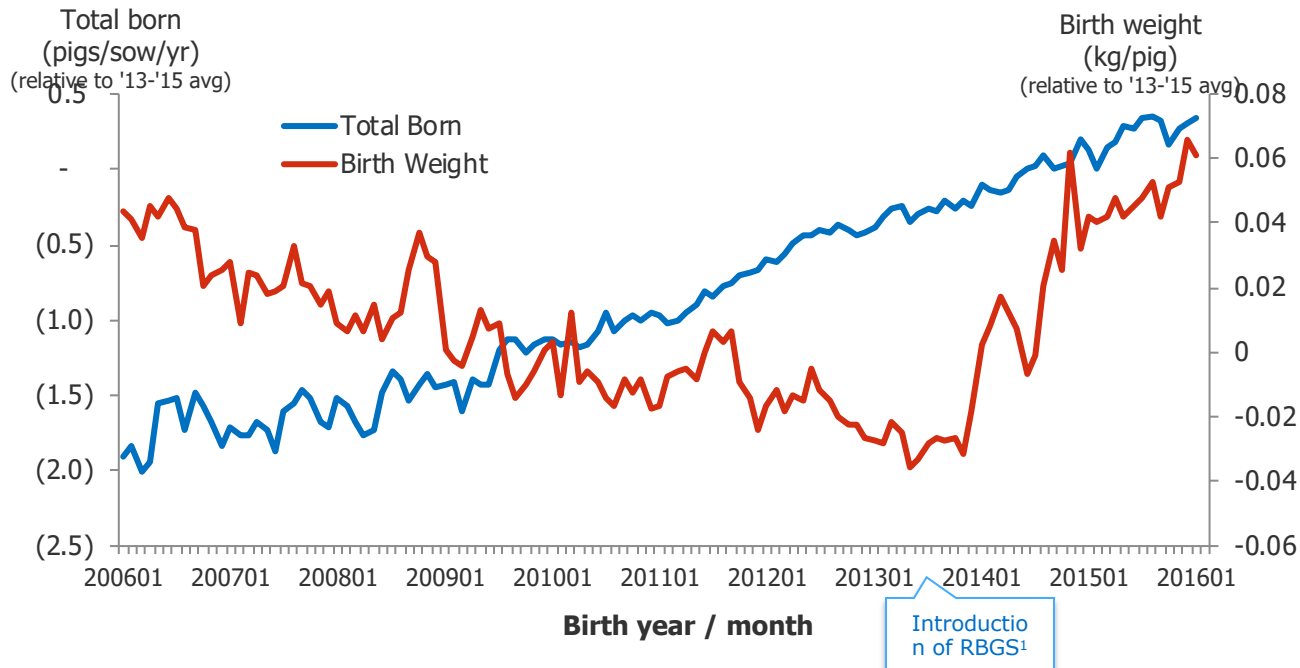
1. Relationship based genomic selection
Source: PIC L02, L03 pure lines (Camborough)





Improvement of total born and piglet birth weight

Trend: genetic improvement in birth weight and total born
(PIC Genetic Nucleus)



1. Relationship based genomic selection
Source: PIC L02, L03 pure lines (Camborough)



Large data challenge

- Computations easy up to 50k genotyped animals
- Holsteins > 3M, Angus >700k, broilers > 100k....
- Many options proposed, none worked well

Basis for genetic evaluation

Pedigree relationships (Henderson, 1976):

$$u_i = f(\textit{sire}, \textit{dam}) + \varphi$$

Hypothesis for genomic relationships:

$$u_i = f(\textit{relatives}) + \varphi$$

Faux et al., 2012

$$u_i = f(\textit{thousand animals}) + \varphi$$

Misztal et al., 2014

Algorithm for proven and young animals (APY)

For young animals

$$u_i \mid u_1, u_2, \dots, u_{i-1} = \sum_{j=\text{"proven"}} p_{ij} u_j + \sum_{j=\text{"young"}} p_{ij} u_j + \varepsilon_i$$

=0 in GBLUP

Misztal et al. (2014)

p=proven y=young; $G=ZZ'$

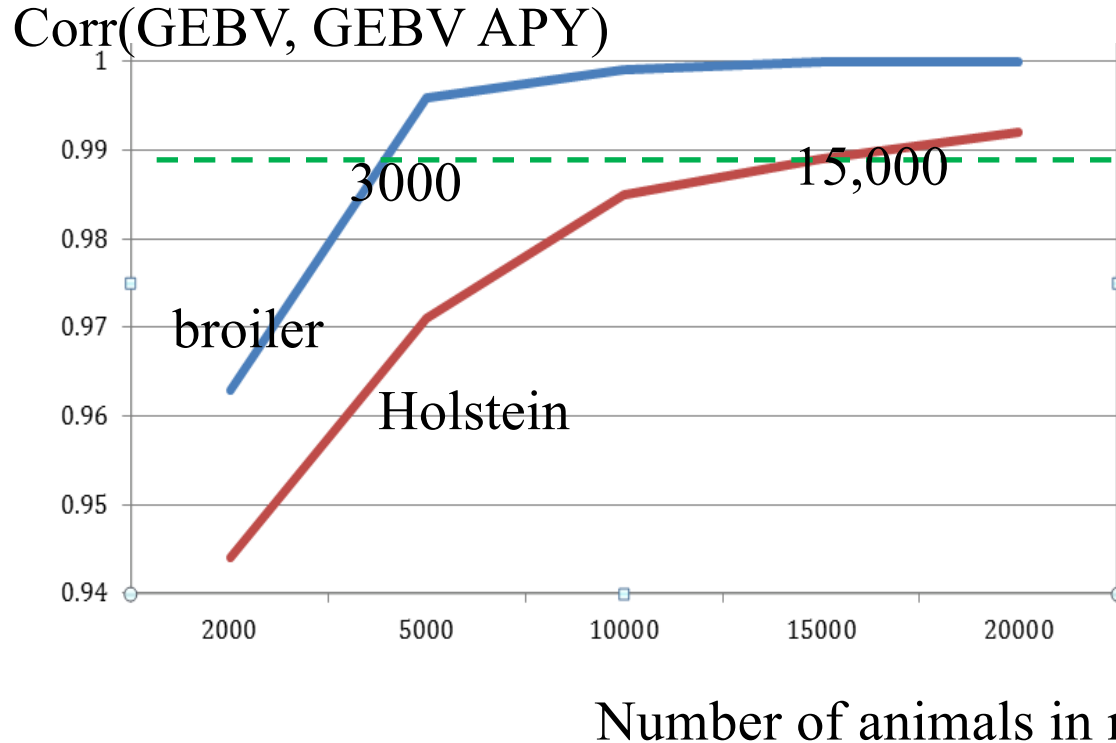
$$\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{G}_{pp}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{pp}^{-1} \mathbf{G}_{py} \\ \mathbf{I} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} \mathbf{G}_{yp} \mathbf{G}_{pp}^{-1} & \\ & \mathbf{I} \end{bmatrix}$$

\mathbf{Z}_p – genotypes for proven animals

\mathbf{Z}_y – genotypes for young animals

$$m_i = g_{ii} - \mathbf{z}_i' \mathbf{Z}_p' \mathbf{G}_{pp}^{-1} \mathbf{Z}_p \mathbf{z}_i$$

Correlations between new and old algorithms



Same accuracy with any choice of animals in recursion

Heterogenetic and homogenic tracts in genome (Stam, 1980)



$$E(\#tracts) = 4N_eL \text{ (Stam, 1980)}$$

N_e – effective population size

L – length of genome in Morgans

Holsteins: $N_e \approx 100$ $L = 30$

$M_e = 12,000$



J. Dairy Sci. 97:3943–3952
<http://dx.doi.org/10.3168/jds.2013-7752>
© American Dairy Science Association®, 2014.

Using recursion to compute the inverse of the genomic relationship matrix

I. Misztal,^{*1} A. Legarra,[†] and I. Aguilar[‡]

^{*}Department of Animal and Dairy Science, University of Georgia, Athens 30602-2771

[†]INRA, UR631-SAGA, BP 52627, 31326 Castanet-Tolosan Cedex, France

[‡]Instituto Nacional de Investigación Agropecuaria, Las Brujas 90200, Uruguay



J. Dairy Sci. 98:1–5
<http://dx.doi.org/10.3168/jds.2014-9125>
© American Dairy Science Association®, 2015.

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

B. O. Fragomeni,^{*1} D. A. L. Lourenco,^{*} S. Tsuruta,^{*} Y. Masuda,^{*} I. Aguilar,[†] A. Legarra,[‡] T. J. Lawlor,[§] and I. Misztal^{*}

^{*}Department of Animal and Dairy Science, University of Georgia, Athens 30602

[†]Instituto Nacional de Investigación Agropecuaria, Canelones, 90200, Uruguay

[‡]INRA, UMR1388 GenePhySE, Castanet Tolosan, 31326, France

[§]Holstein Association USA Inc., Brattleboro, VT 05302

GENETICS | INVESTIGATION

Inexpensive Computation of the Inverse of the Genomic Relationship Matrix in Populations with Small Effective Population Size

Ignacy Misztal¹

Animal and Dairy Science, University of Georgia, Athens, Georgia 30602

GENETICS | INVESTIGATION

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, GenPhySE, F-31326 Castanet-Tolosan, France

Dimensionality of genomic information

genotypes

$$\mathbf{Z} = \mathbf{U} \mathbf{\Delta} \mathbf{V}$$

Singular value decomposition
 $\mathbf{U}'\mathbf{U}=\mathbf{I}$, $\mathbf{V}'\mathbf{V}=\mathbf{I}$, $\mathbf{\Delta}$

$$\mathbf{G} = \mathbf{U}\mathbf{\Delta}\mathbf{\Delta}\mathbf{U}' = \mathbf{U}\mathbf{D}\mathbf{U}' \quad \text{Genomic relationship matrix}$$
$$\text{Rank}(\mathbf{G}) \leq \min(\#\text{SNP}, \#\text{anim})$$

$$\mathbf{Z}'\mathbf{Z} = \mathbf{V}'\mathbf{\Delta}\mathbf{\Delta}\mathbf{V} \quad \text{SNP BLUP design matrix}$$
$$\text{Rank}(\mathbf{Z}'\mathbf{Z}) \leq \min(\#\text{SNP}, \#\text{anim})$$

Same dimensionality for genotypes, GRM and SNP BLUP

Dimensionality around 5-15k (VanRaden, 2008; Maciotta et al., 2013)

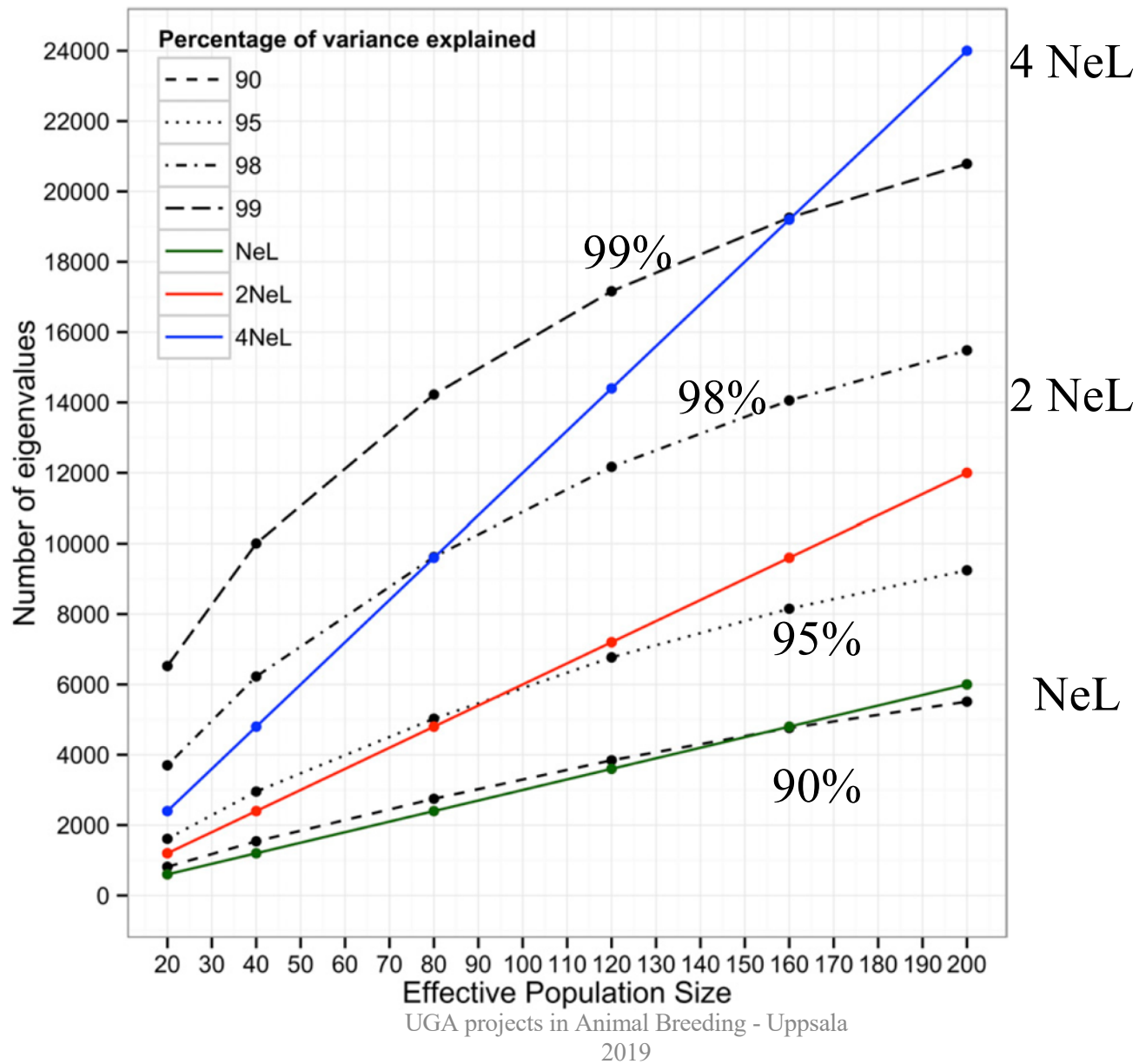


GENETICS | INVESTIGATION

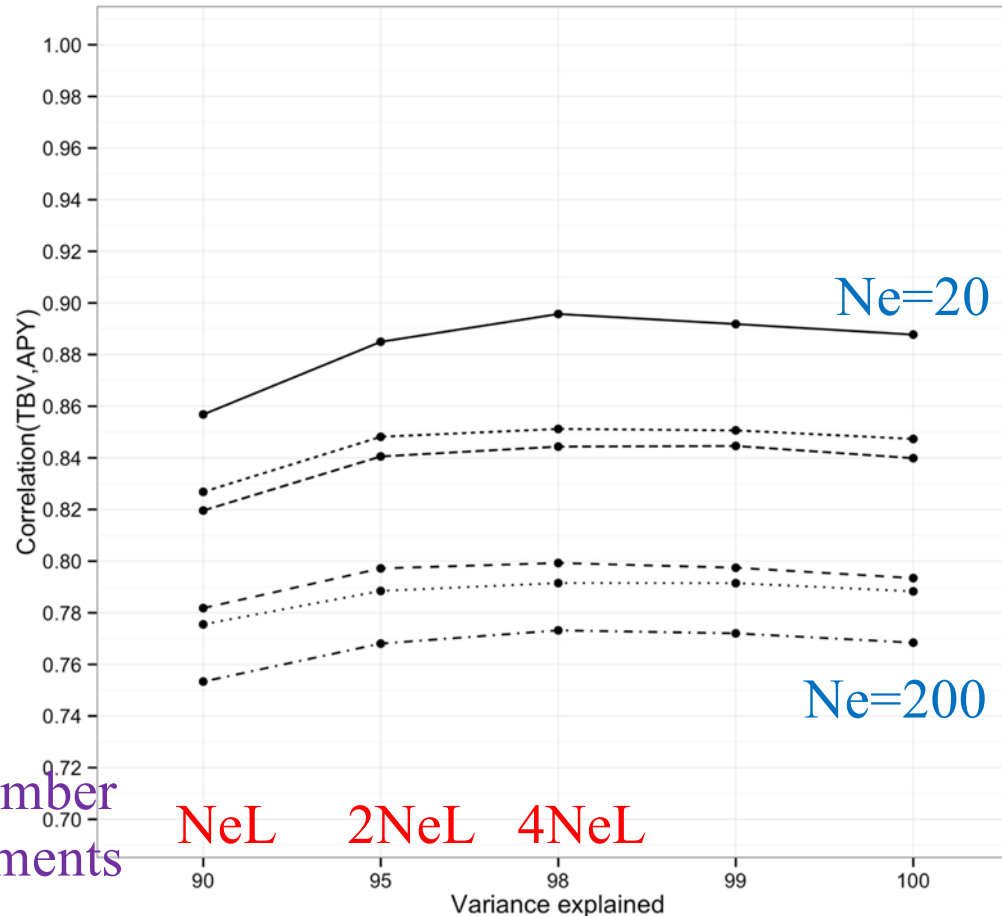
The Dimensionality of Genomic Information and Its Effect on Genomic Prediction

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, GenPhySE, F-31326 Castanet-Tolosan, France



True accuracies as function of number of eigenvalues corresponding to given explained variance in G



Approximate number of animals / segments

Accuracies maximized by 98% “information in G, 95% almost as good
Last 2% of information in G noise

Real populations study

(Pocrnic et al., 2016)

- Dairy, beef, pigs, broilers...



Pocrnic et al. *Genet Sel Evol* (2016) 48:82
DOI 10.1186/s12711-016-0261-6




RESEARCH ARTICLE

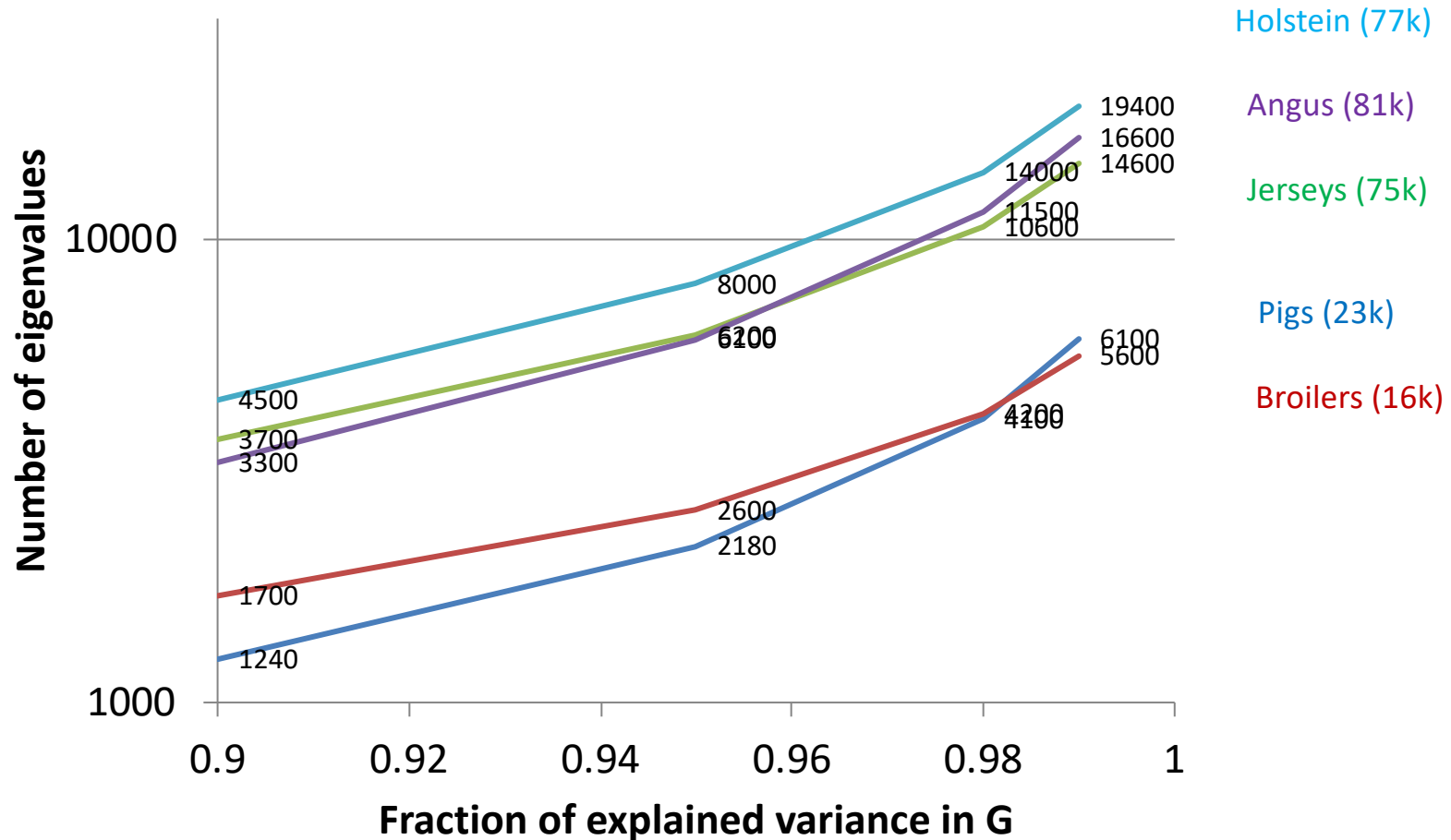
Open Access



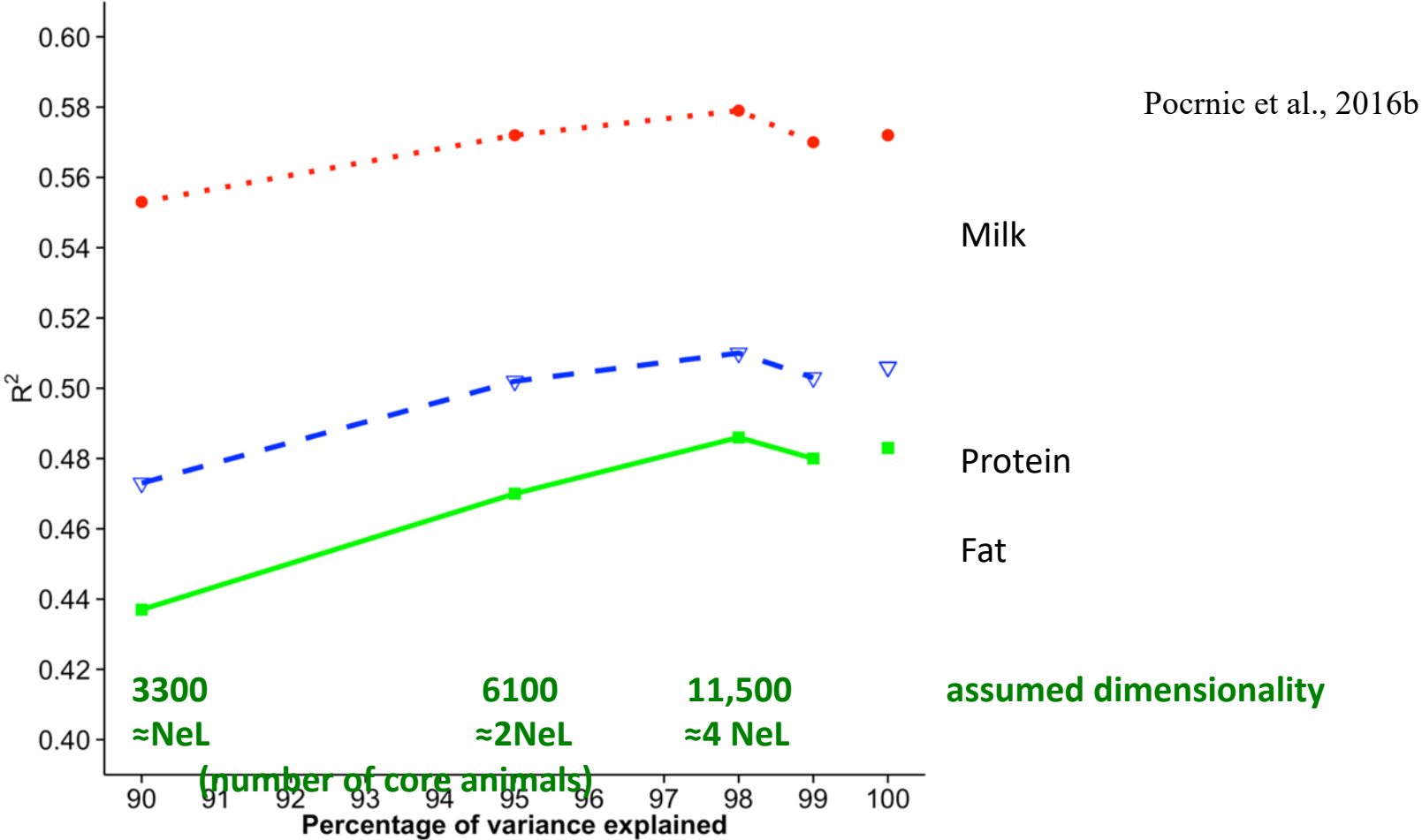
Dimensionality of genomic information and performance of the Algorithm for Proven and Young for different livestock species

Ivan Pocrnic^{*} , Daniela A. L. Lourenco, Yutaka Masuda and Ignacy Misztal

Number of eigenvalues in G to explain given fraction of variability



Reliabilities – Jerseys (75k animals)



100% = full inverse → lower accuracy

Estimated dimensionality, effective population size and optimal number of SNP

Specie	Range of Me (95-99%)	Effective population size (L=30M)	Optimal number of SNP (12 x Me)
Holsteins	8k-14k	149	100-180k
Jerseys	6k-12k	101	70k-150k
Angus	6k-11k	113	70k-130k
Pigs	2k-6k	43 (L=20M)	24k-72k
Chicken	3k-6k	44	36K-72k

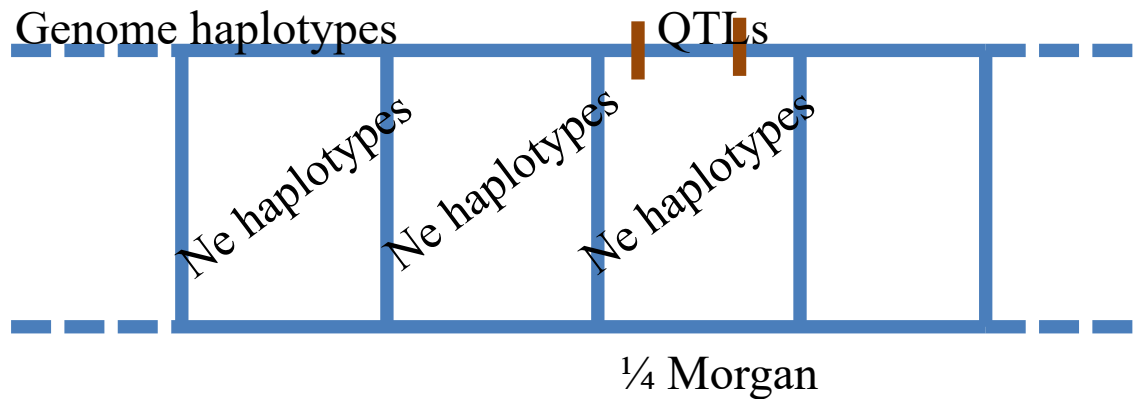
Pocrnic et al. (2016b)

Side effects of reduced dimensionality

- We estimate effects of SNP blocks
 - 800k in humans
 - 5-15k in animals
- Impact on SNP selection and GWAS

Understanding of limited dimensionality (II)

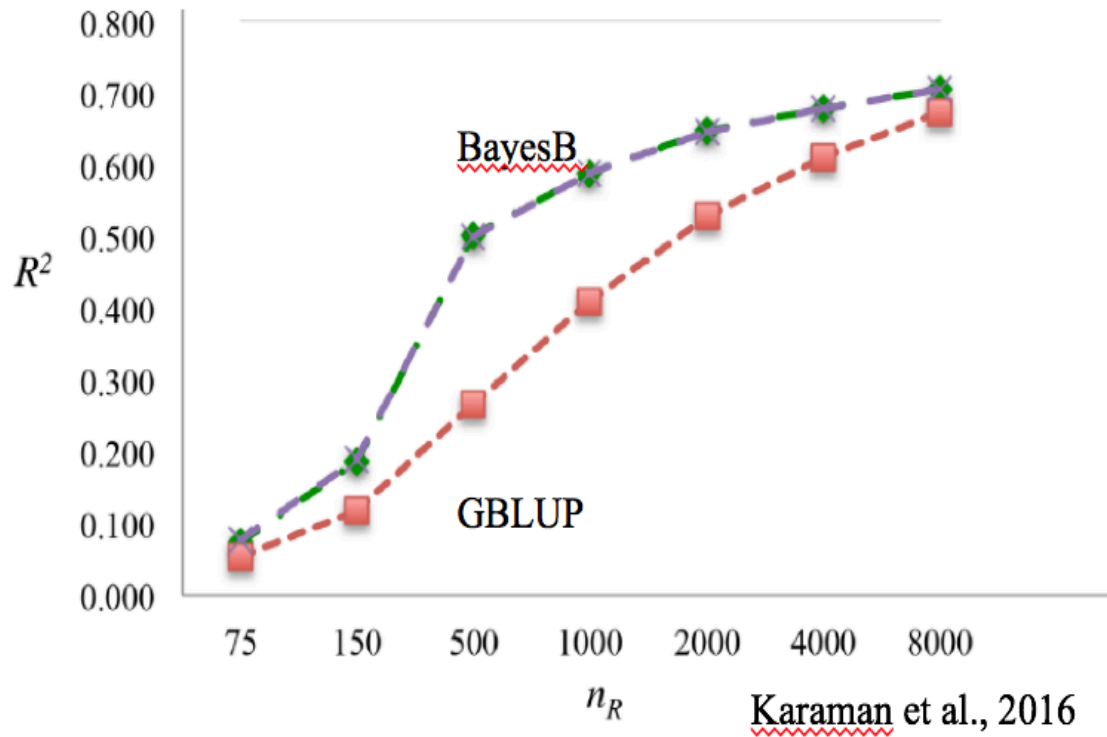
Number of haplotypes: $4 N_e L$
Ne within each $\frac{1}{4}$ Morgan segment



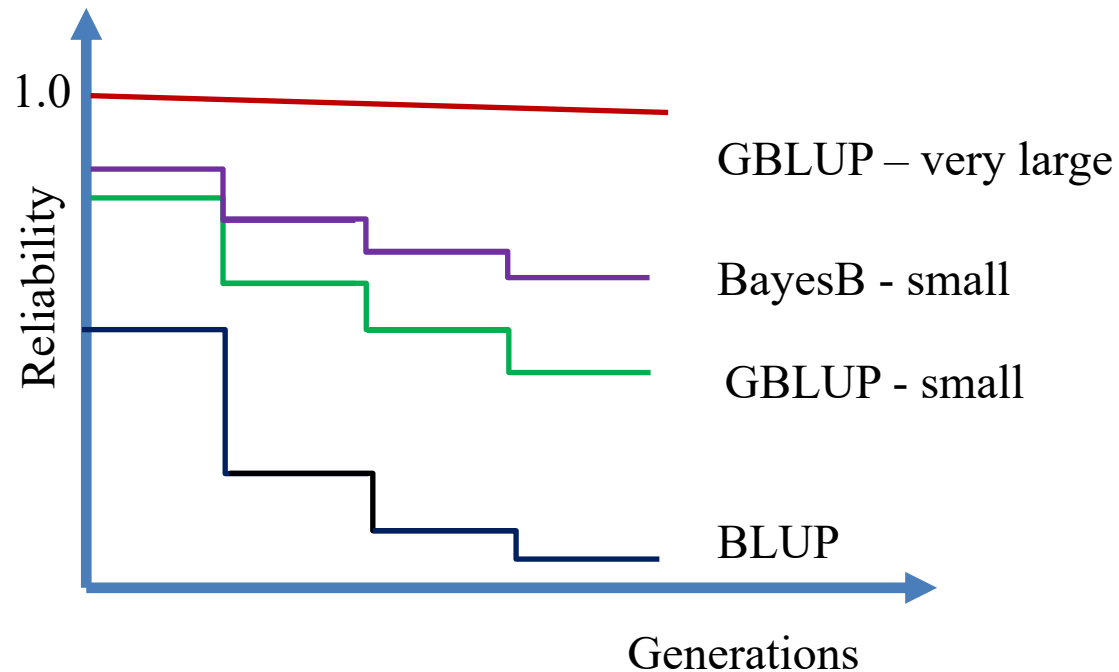
Dimensionality of $\frac{1}{4}$ Morgan case: N_e

→ Reduced dimensionality with weighted GRM

Advantage of SNP selection and size of data

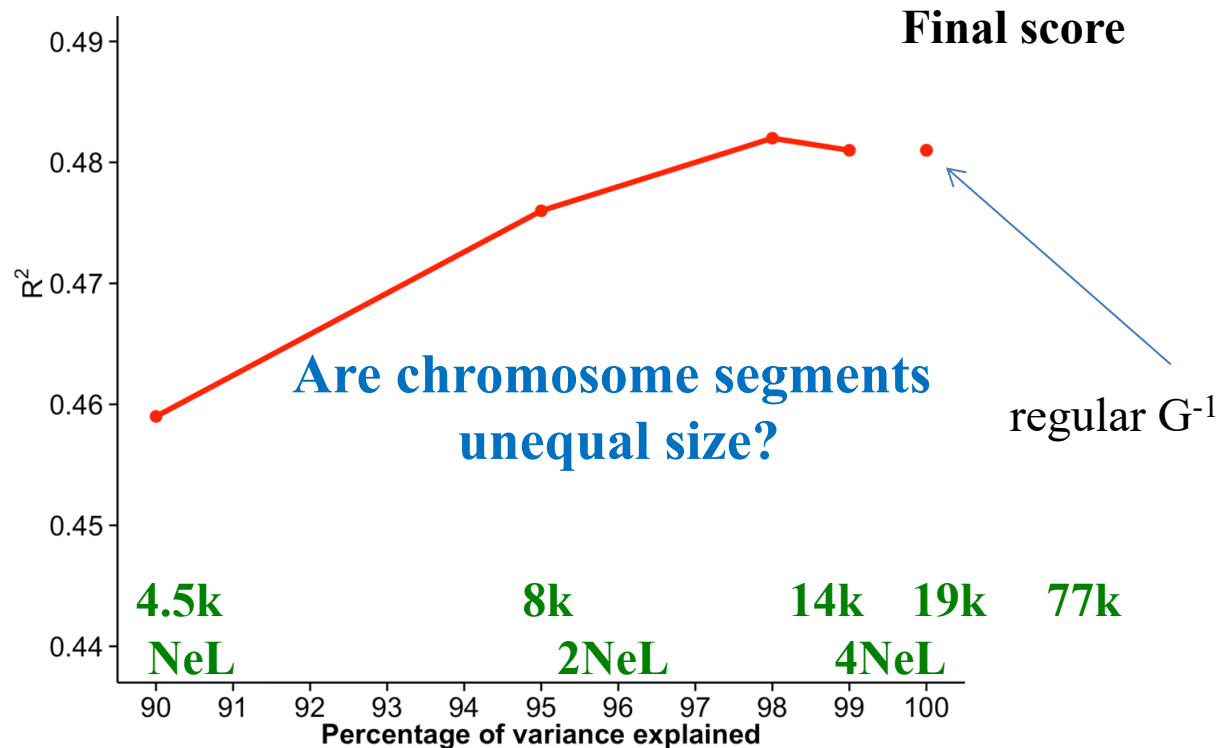


Persistence over generations with different sizes of reference populations



Very large – equivalent to 4NeL animals with 99% accuracy
Are SNP effects from Holstein national populations converging?

Reliabilities assuming different dimensionality with APY inverse – Holsteins

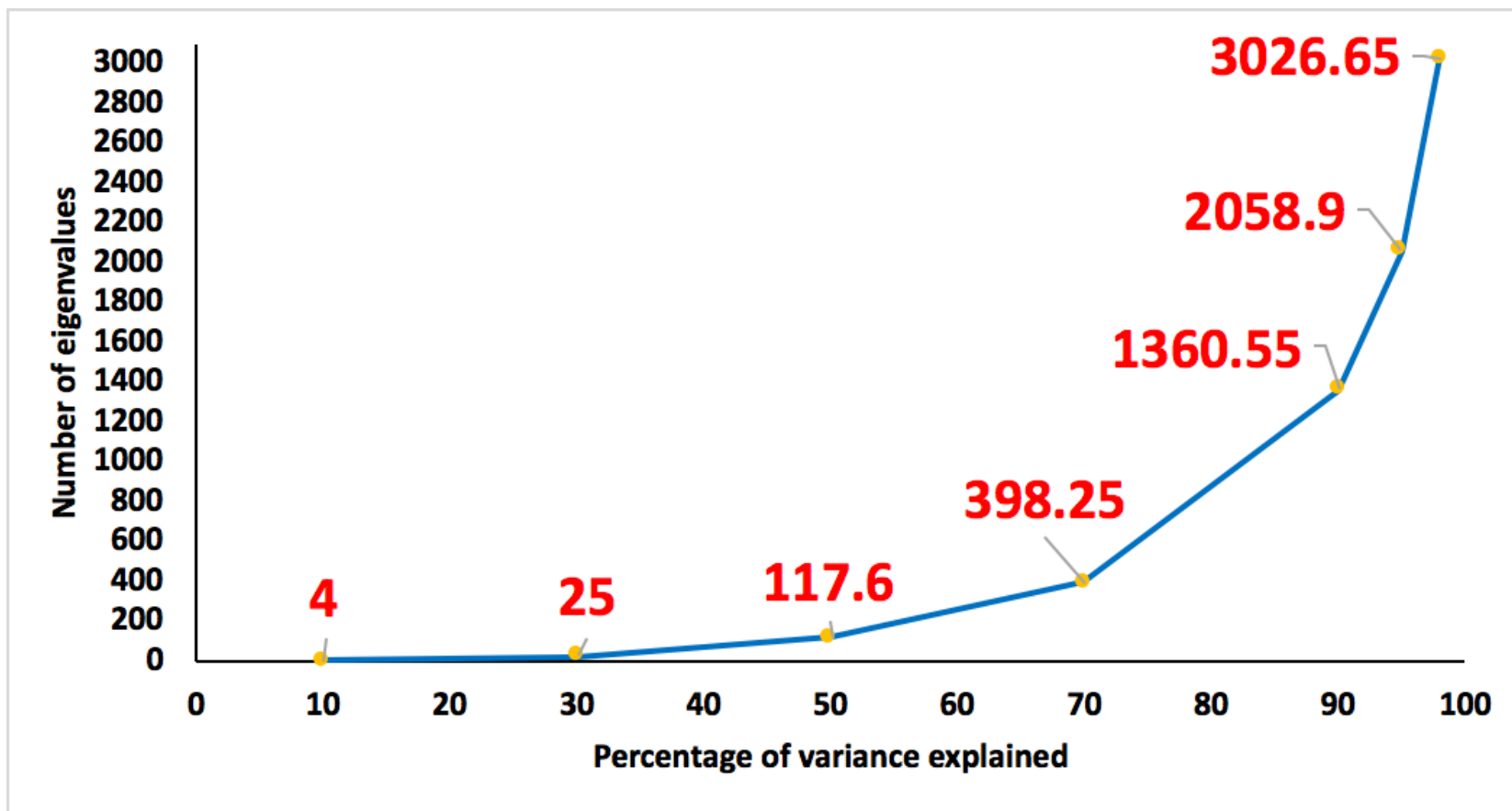


Pocrnic et al., 2016b

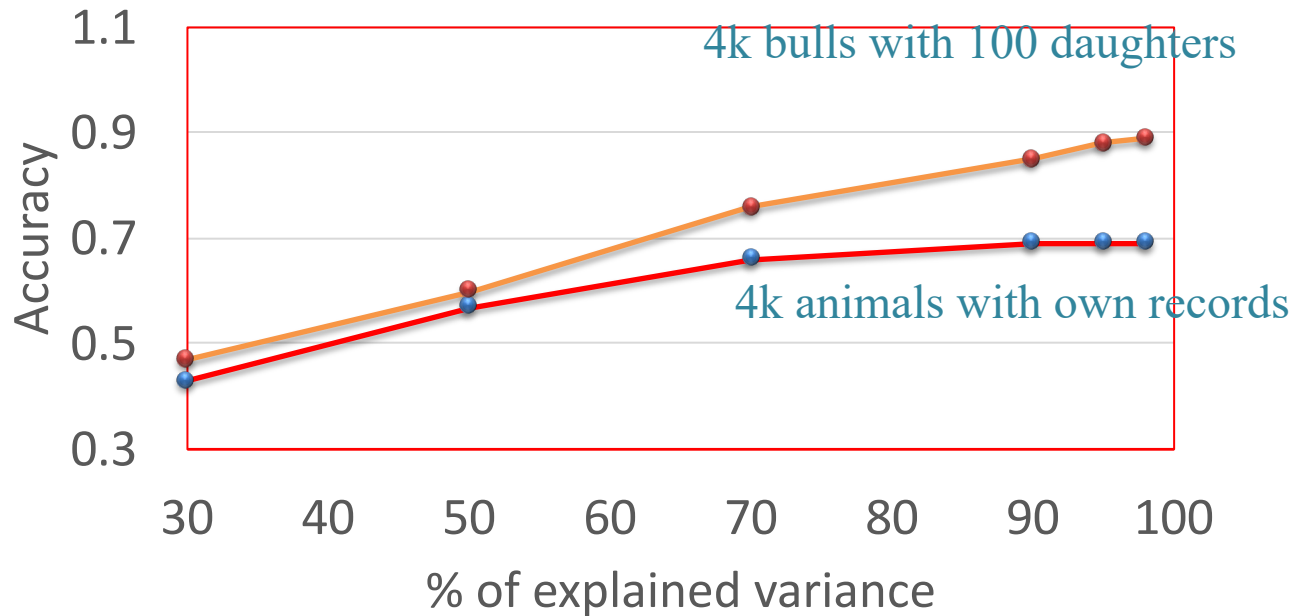
Is genomic selection on chromosome segments or chromosome clusters ?

- Simulation
 - 6k animals with 50 k SNP
 - $N_e \approx 50$, $L = 10M$
- GBLUP
 - Use GRM with limited number of eigenvalues (corresponding to 10 to 99% variation)
 - 4k animals in reference population, 2k in validation

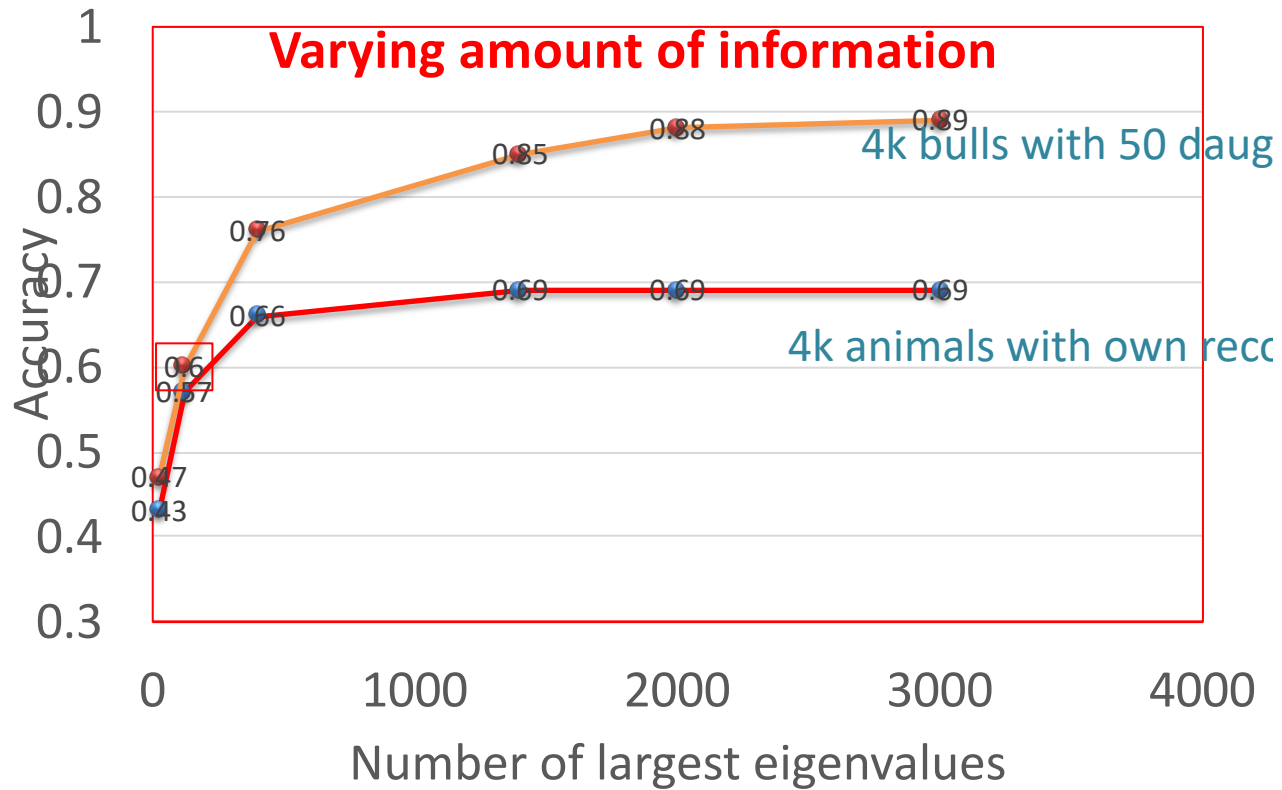
Eigenvalue profile of GRM



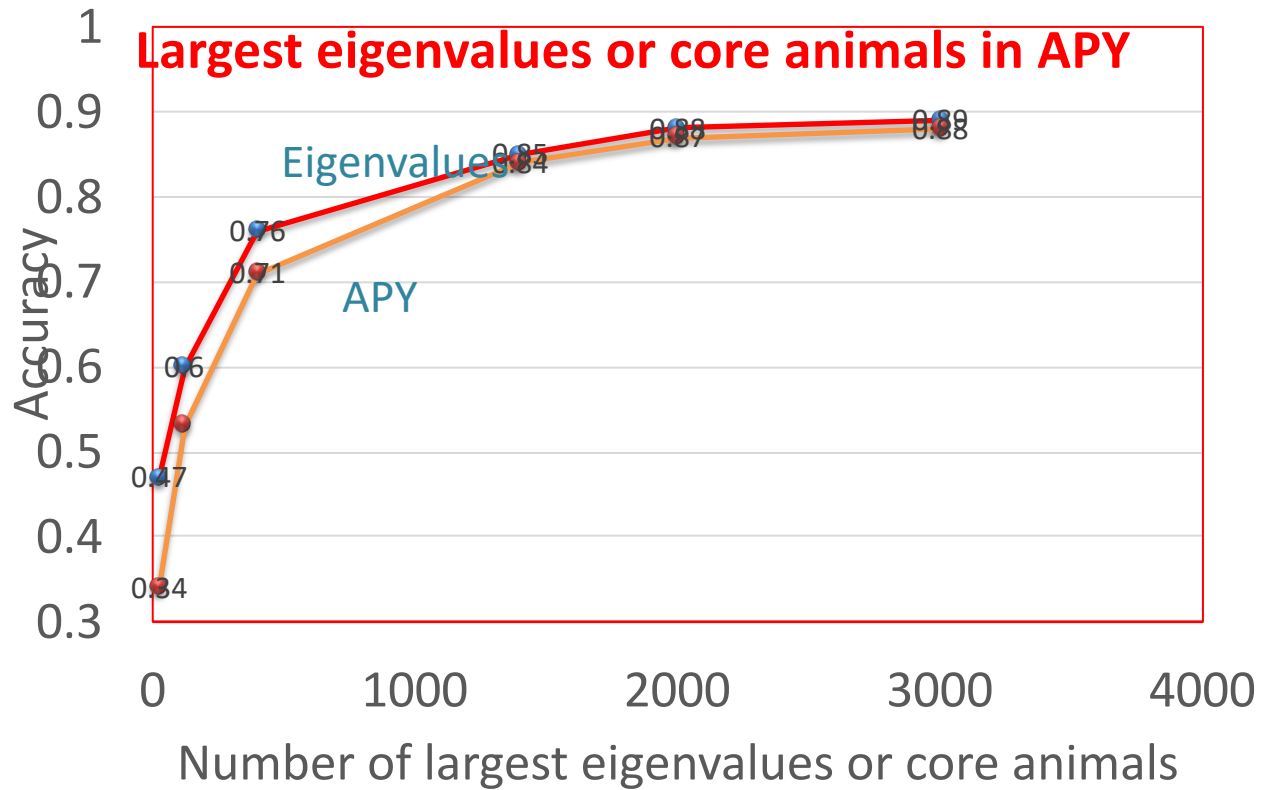
Accuracies of GBLUP using GRM with largest eigenvalues only



Eigenvalues 25 117 398 1368 2058 3026



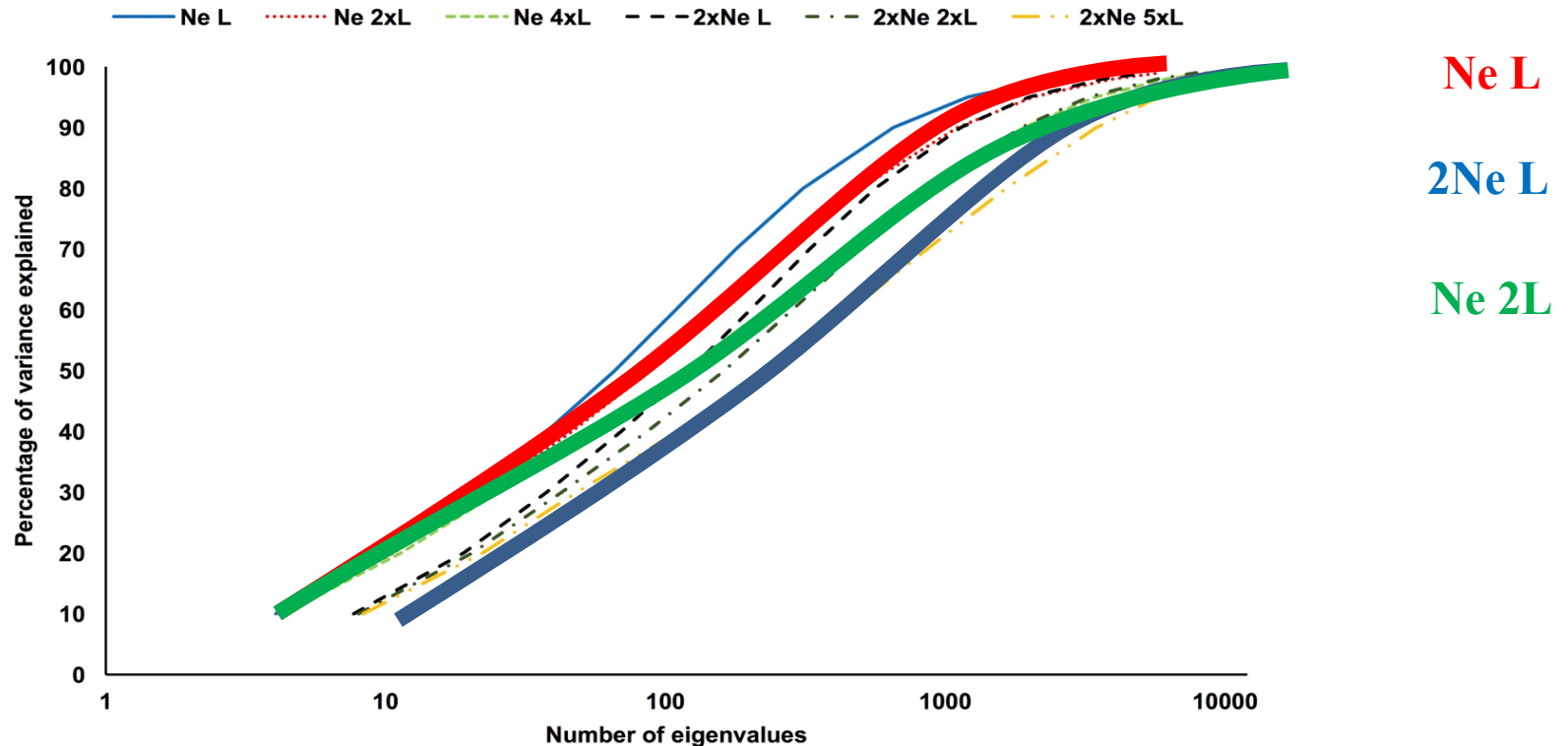
Does APY algorithm for inversion of GRM work on segments or eigenvalues



Selection on largest eigenvalues – important ancestors – reduced N_e
If largest eigenvalues excluded- increased diversity?

How are eigenvalues influenced by effective population size and genome length?

Graph of dimensionality as % for different Ne and L



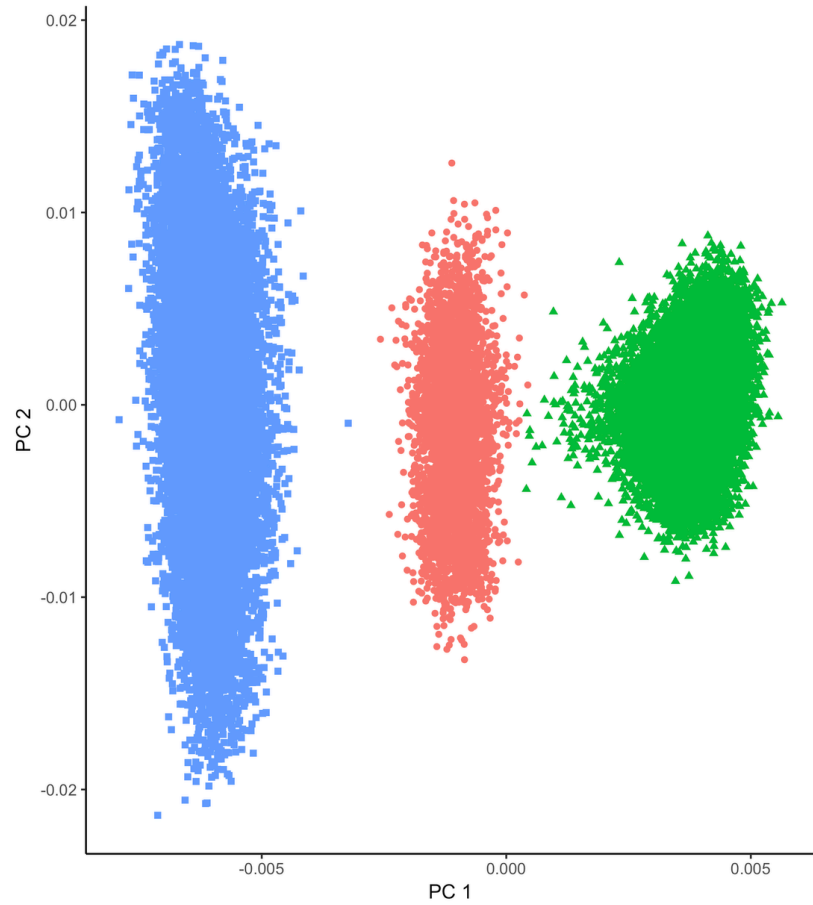
Largest eigenvalues do not depend on genome size - cluster haplotypes across all genome

Can one predict Ne and L from small populations ?

UGA projects in Animal Breeding - Uppsala

2019

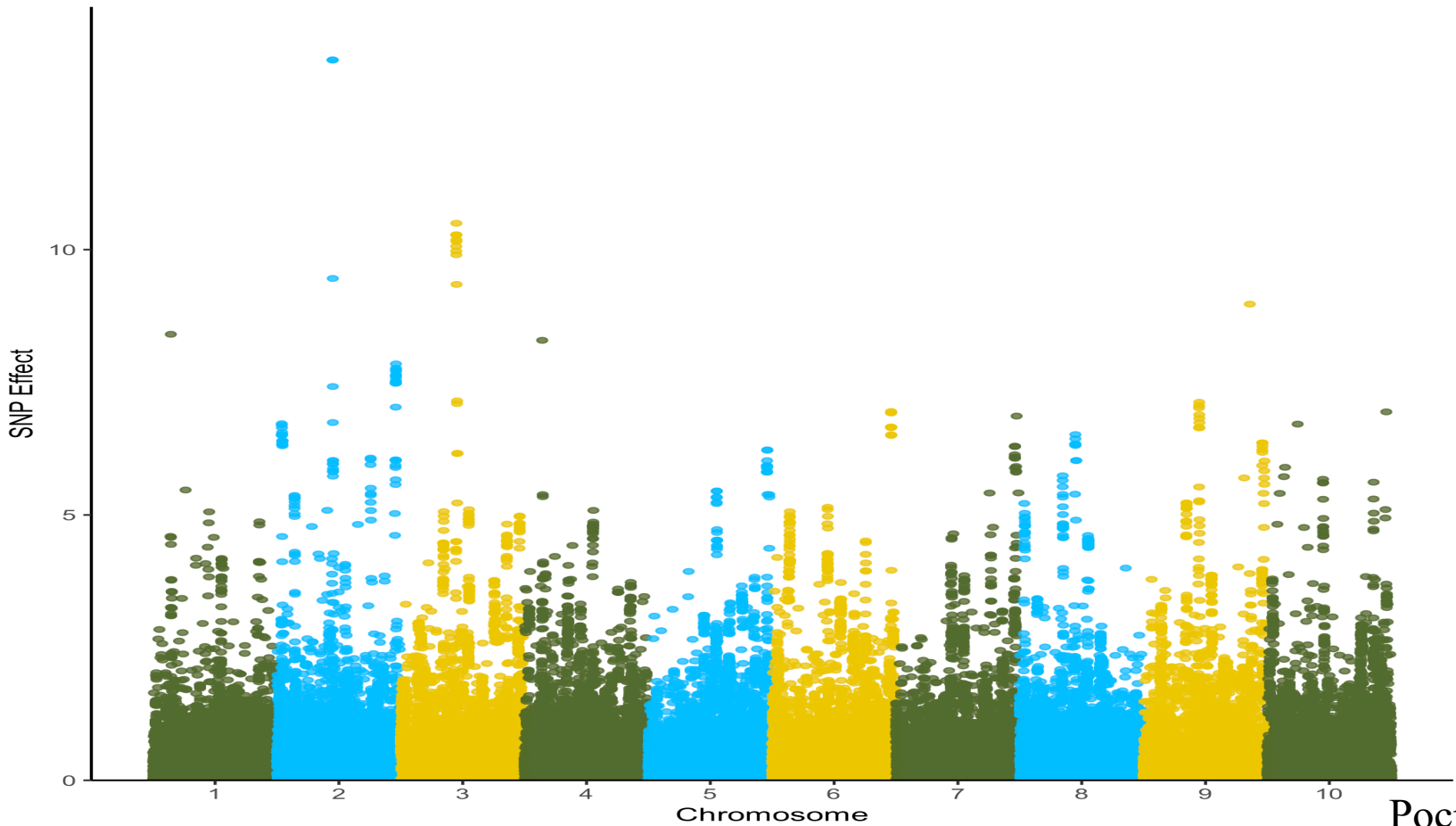
PCA Plot



PC1 and PC2 pool segments across genome

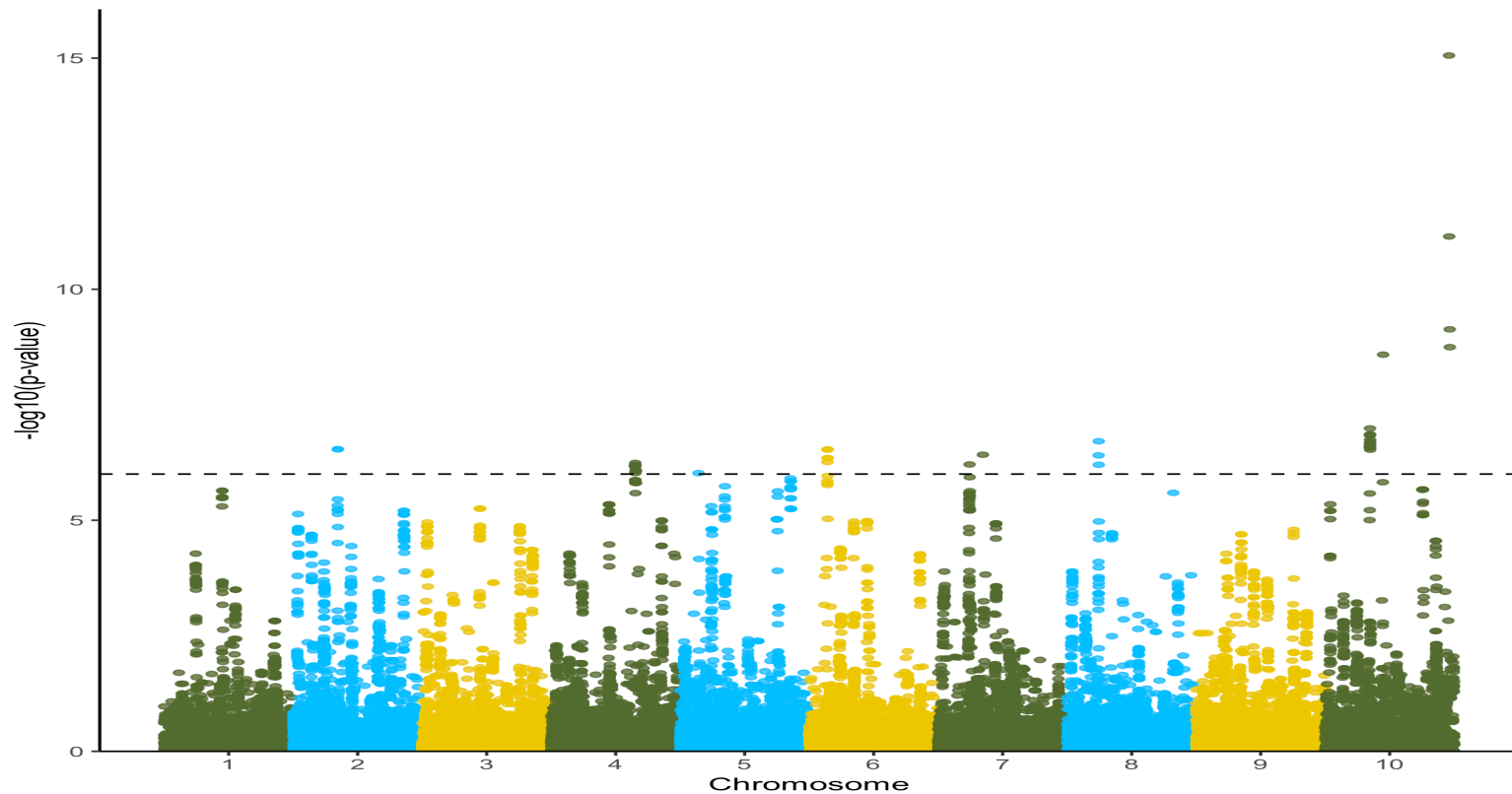
How easy is to find causative SNP?

Manhattan plot with 100 equidistant QTL with equal effect

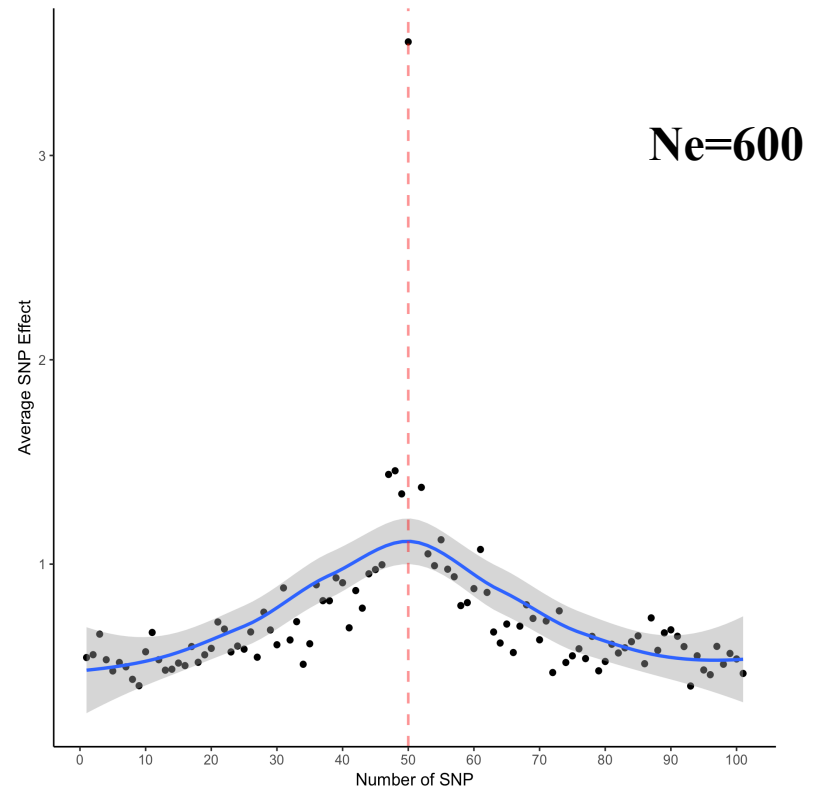
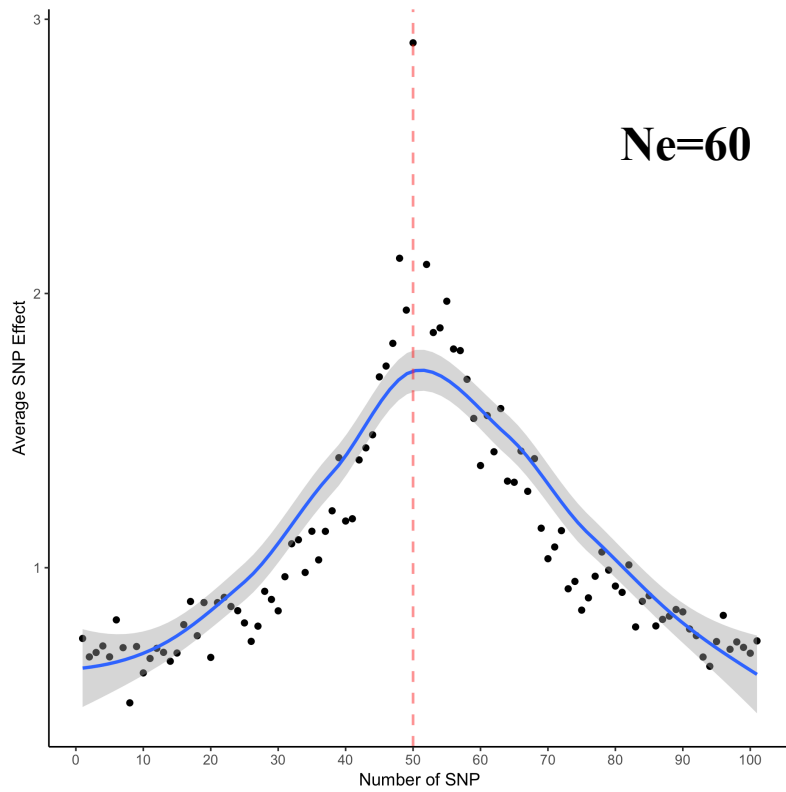


Pocrnic et al., 20

Manhattan plot with p-values



Pooled effect of 100 equidistant QTL



Questions with limited dimensionality

- Are segments physical?
- Can they be traced to ancestors?
- Can their number be predicted from small data?
- ...

Why not single-step in dairy

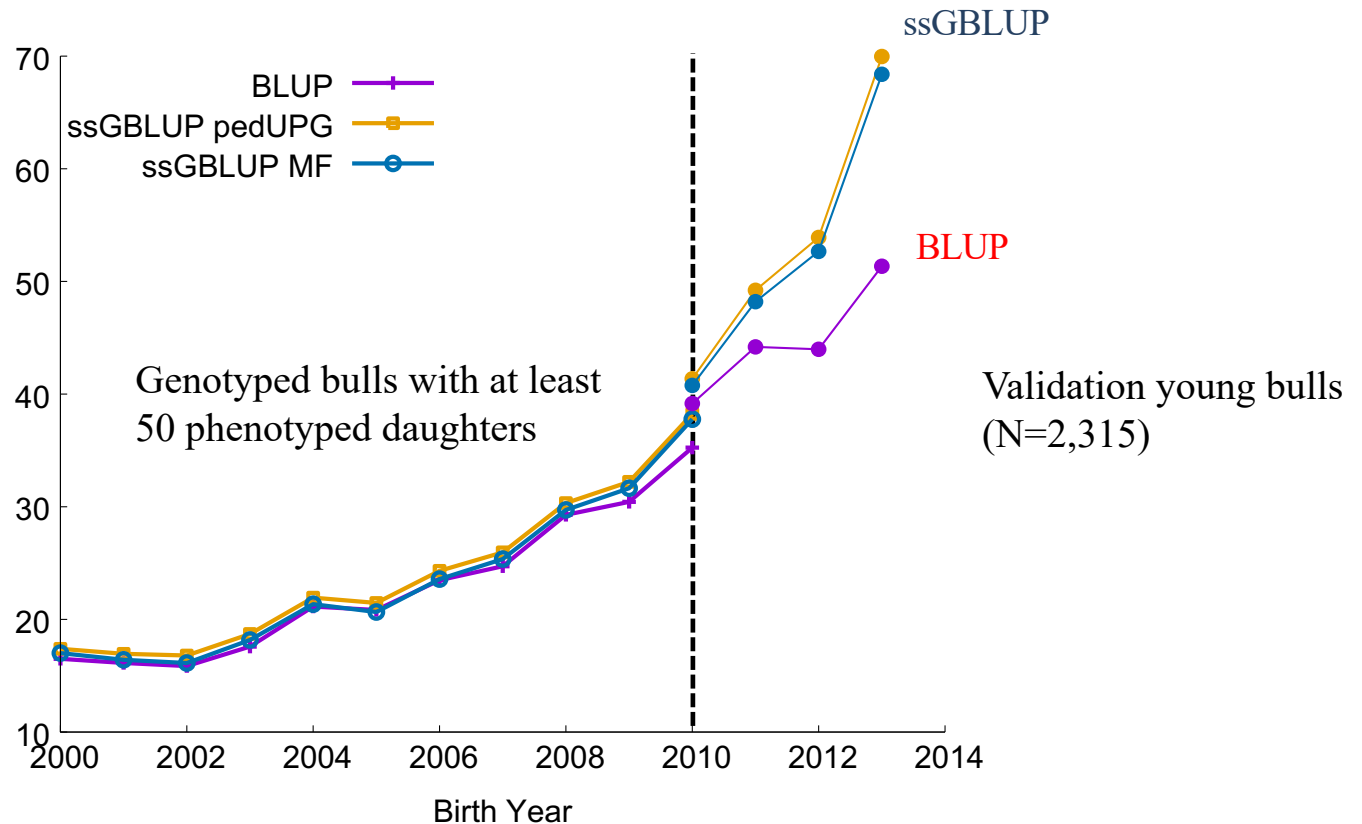
- High number of genotypes
- Biases everywhere
- No free software
- Old system OK after corrections

Genomic evaluation of US Holsteins

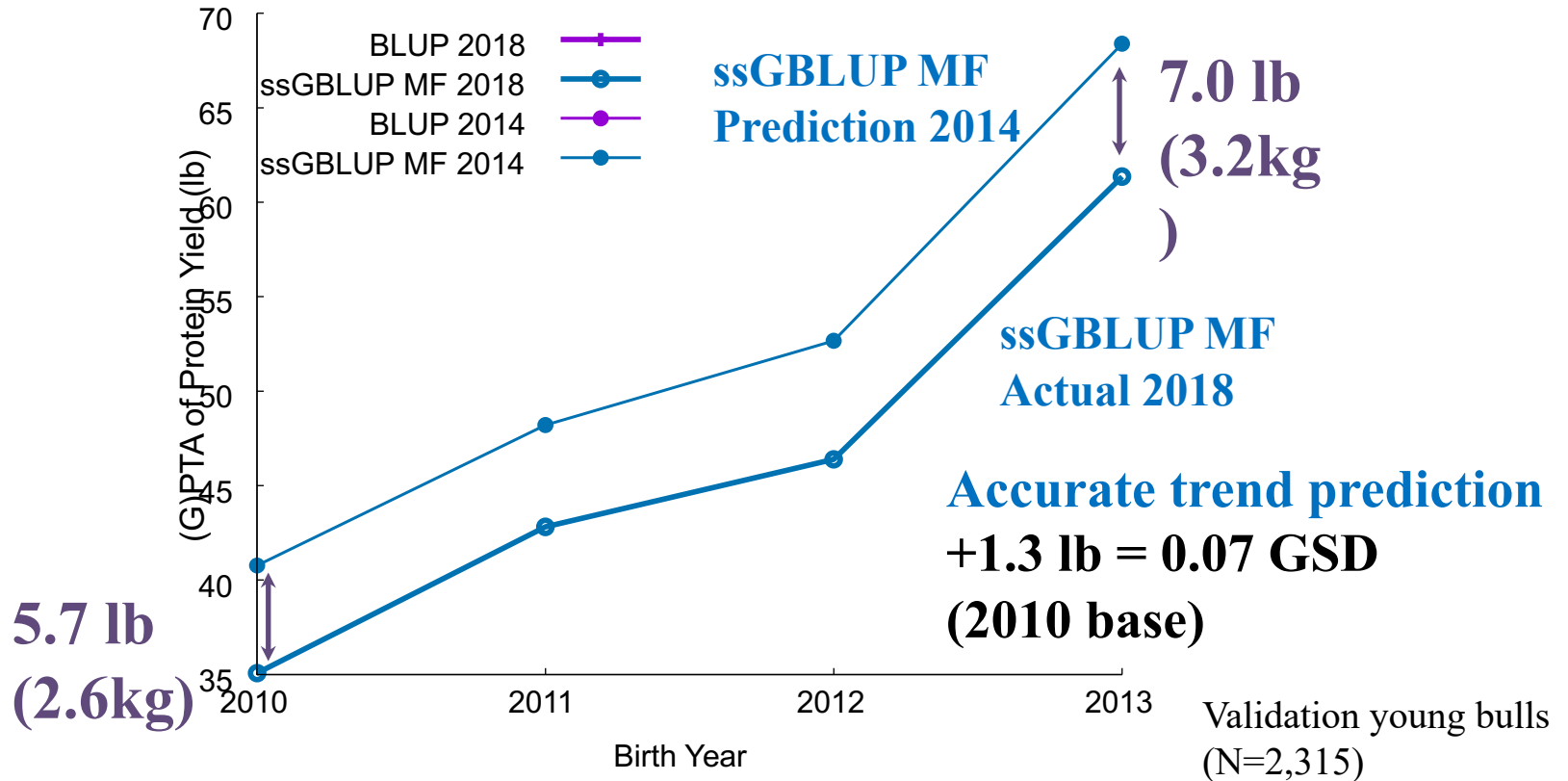
- ssGBLUP for 35 million US Holsteins (Masuda et al., 2017)
 - Protein
 - 2.3M genotypes of SNP60K
 - Computing time 6.5 h
- Unknown parent groups
 - Genomic
 - Metafounders

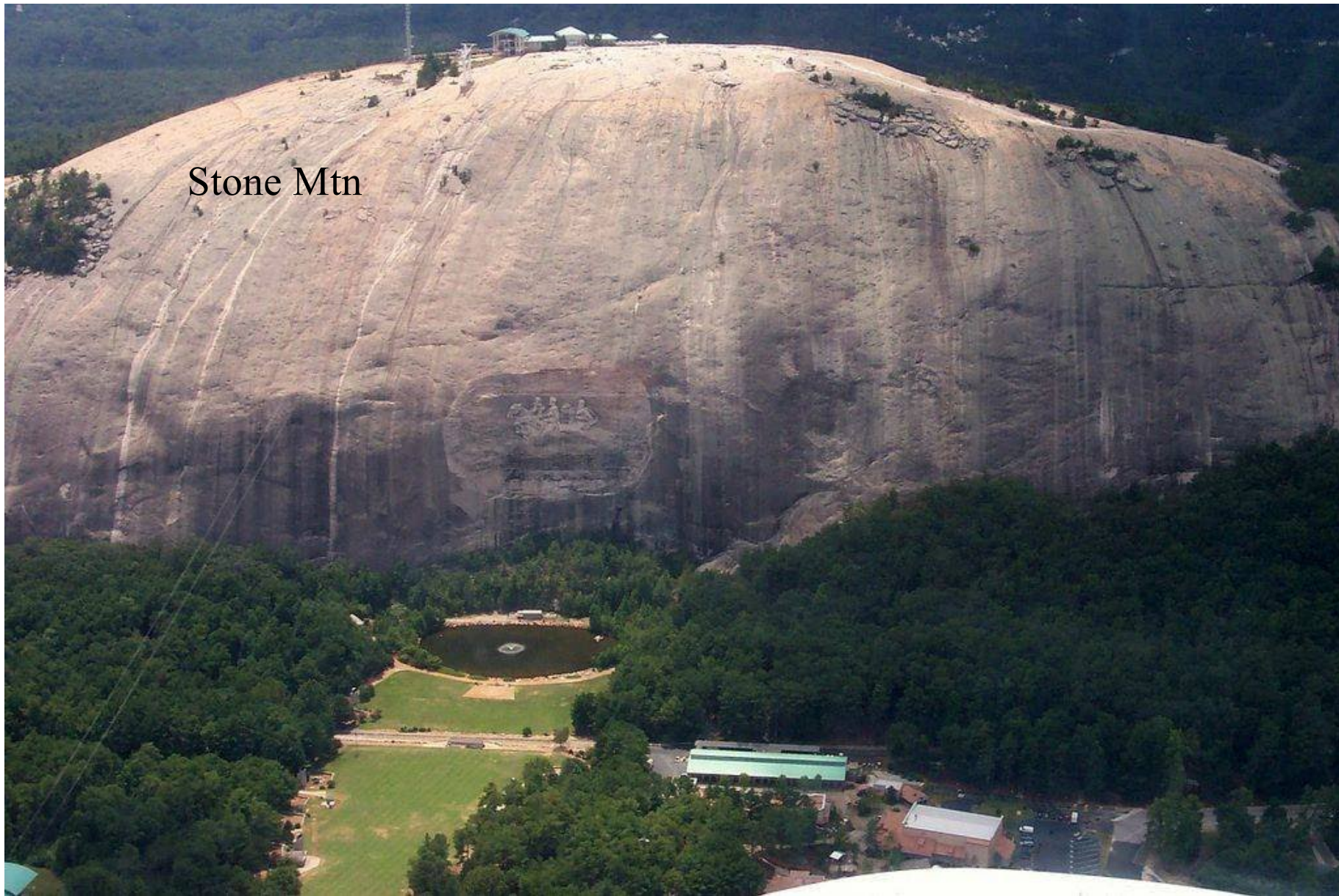


Genetic trend for genotyped bulls in 2014



Predicted and actual trends for bulls





Stone Mtn

UGA projects in Animal Breeding - Uppsala
2019



UGA projects in Animal Breeding - Uppsala
2019

Coca Cola Center



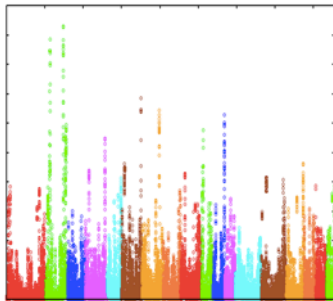
COA projects in Animal Breeding - Uppsala

ssGBLUP for Genome Wide Association Studies

- Large research interest in GWAS
- Limitations of Bayesian methods



$G=ZZ'$ unweighted genomic relationships
 $G=ZDZ'$ weighted G



Genet. Res., Camb. (2012), 94, pp. 73–83. © Cambridge University Press 2012
doi:10.1017/S0016672312000274

73

Genome-wide association mapping including phenotypes from relatives without genotypes

H. WANG^{1*}, I. MISZTAL¹, I. AGUILAR², A. LEGARRA³ AND W. M. MUIR⁴

¹Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602-2771, USA

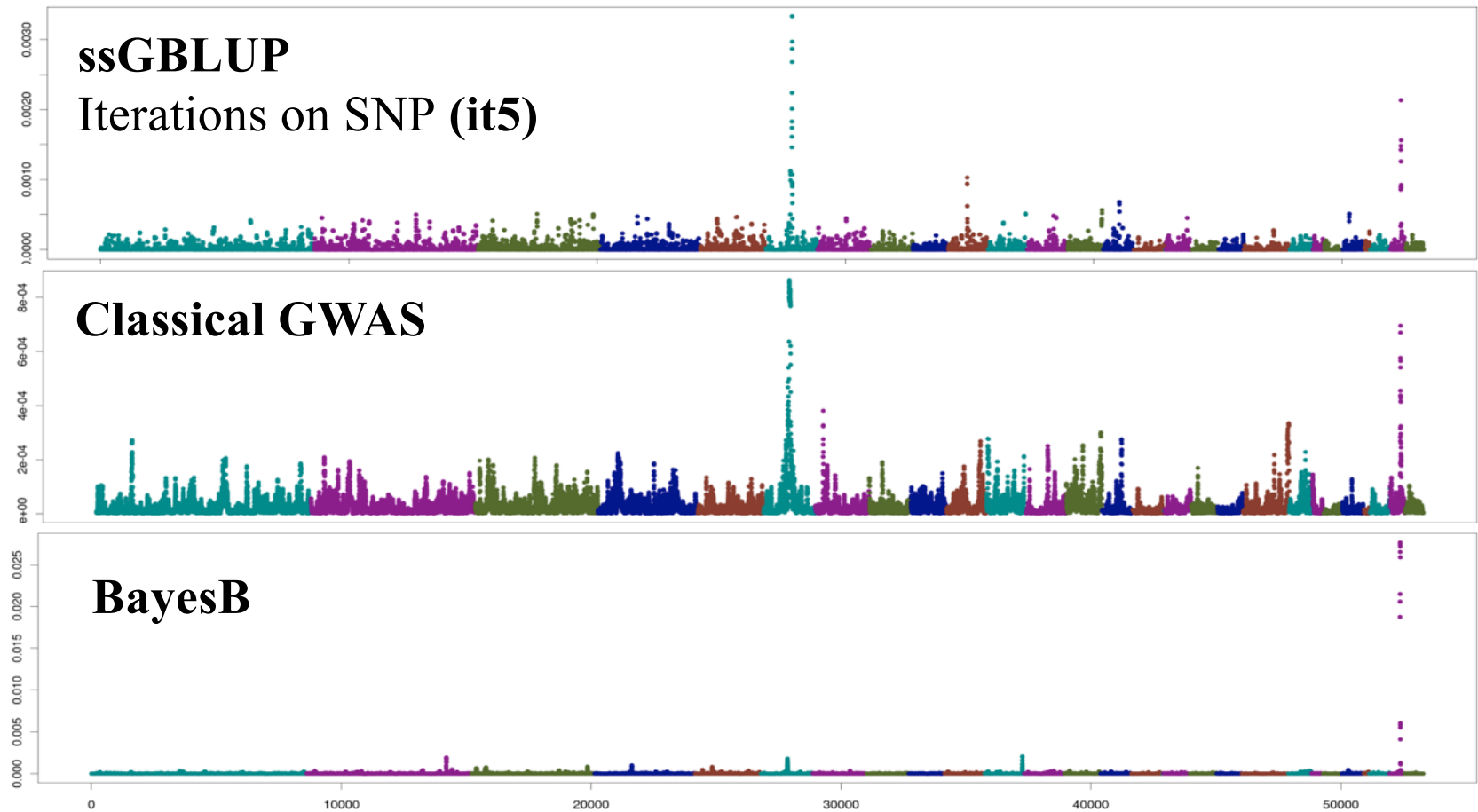
²Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, 90200 Canelones, Uruguay

³INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), BP 52627, 32326 Castanet-Tolosan, France

⁴Department of Animal Science, Purdue University, West Lafayette, IN 47907-1151, USA

(Received 19 September 2011; revised 8 December 2011, and 9 March 2012; accepted 13 March 2012)

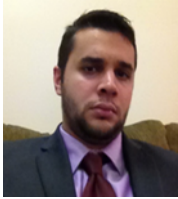
Comparison of Three Methods:



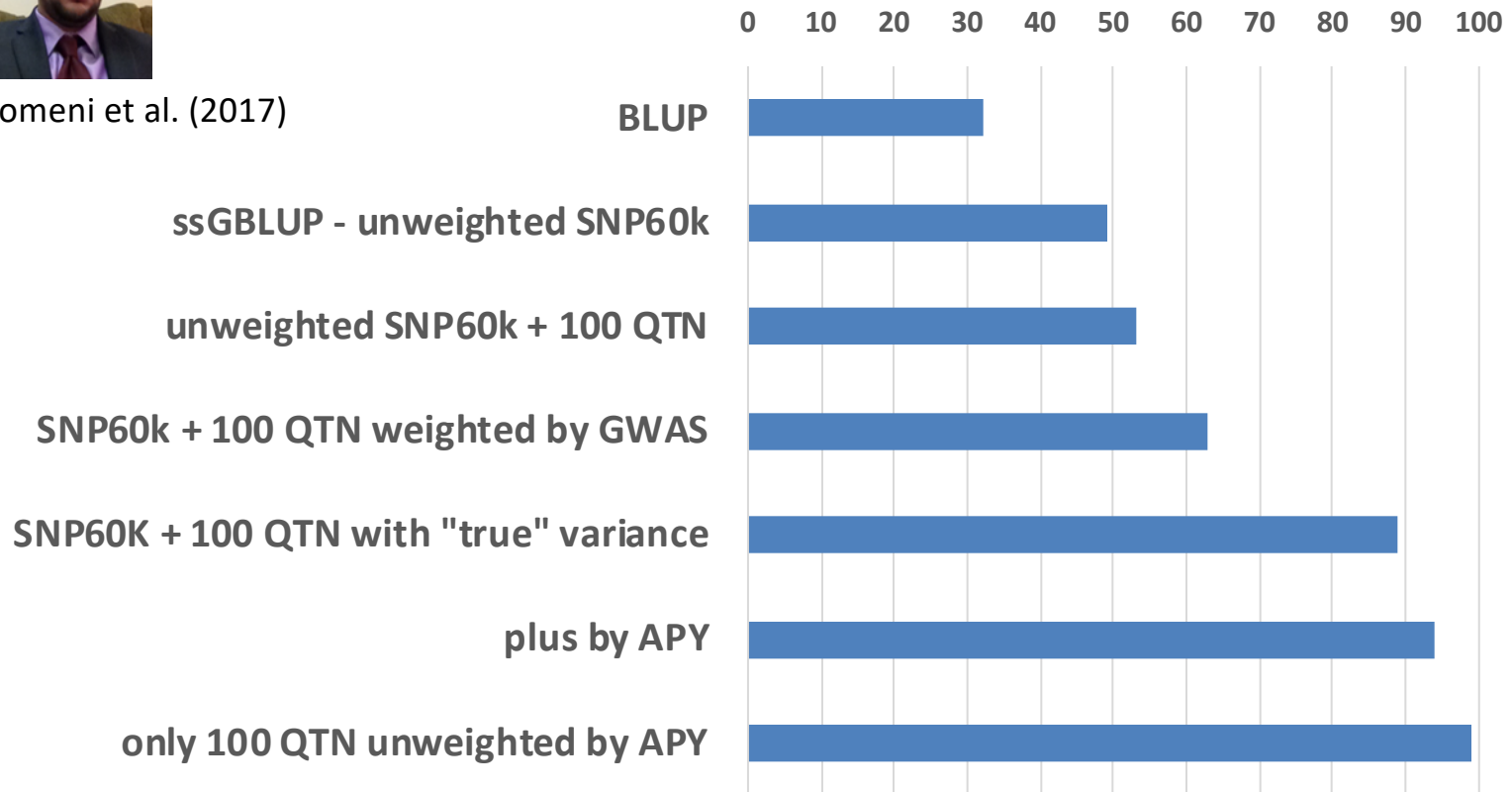
Inclusion of causative SNP information

- In dairy, results varied
 - No improvement in Dutch (Binsbergen, et al., 2015) or German Holsteins (Erbe et al., 2016)
 - Up to 5% improvement in Nordic/French (Brøndum et al.)
 - Up to 5% improvement in US Holsteins (Vanraden et al. (2017))
- In SNP BLUP, SNP effects regressed towards zero
- To include causative SNP:
 - Need location
 - need variance

ssGBLUP accuracies using SNP60K and 100 QTNs – simulation study



Fragomeni et al. (2017)



Use of causative variants and SNP weighting in a single-step GBLUP context

Fragomeni BO¹, Lourenco DAL¹, Legarra A², Tooker ME³, VanRaden PM³, Misztal I¹

¹University of Georgia, Athens, USA

²INRA, Castanet-Tolosan, France

³AGIL ARS-USDA, Beltsville, USA



**UNIVERSITY OF
GEORGIA**



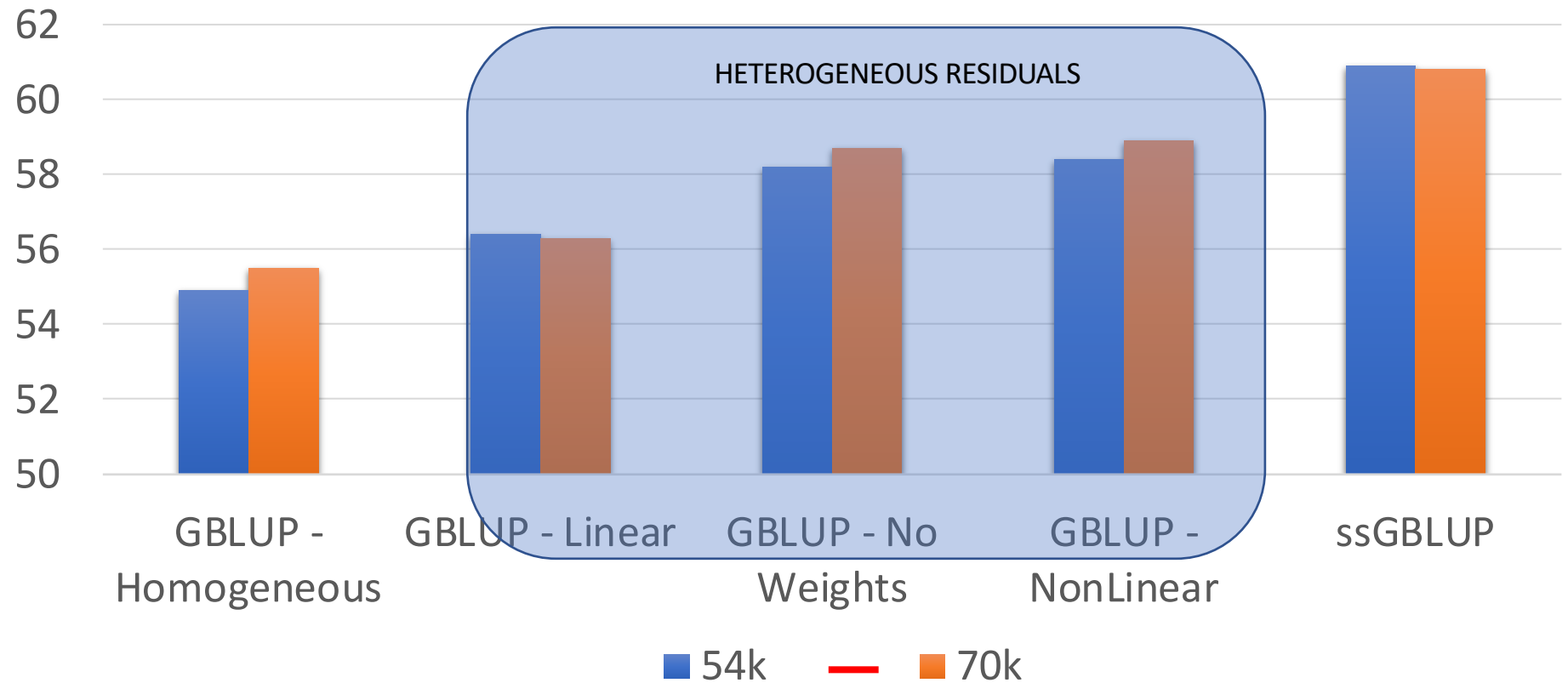
A O T E A C E N T R E , A U C K L A N D , N E W Z E A L A N D

US Holstein data

- 4M records for Stature
- 3M Cows
- 4.6M Animals in pedigree
- 27k Genotyped Sires

- 54k SNP
- 54k SNP + 17k Causative Variants (VanRaden et al., 2017)

Including causative variants



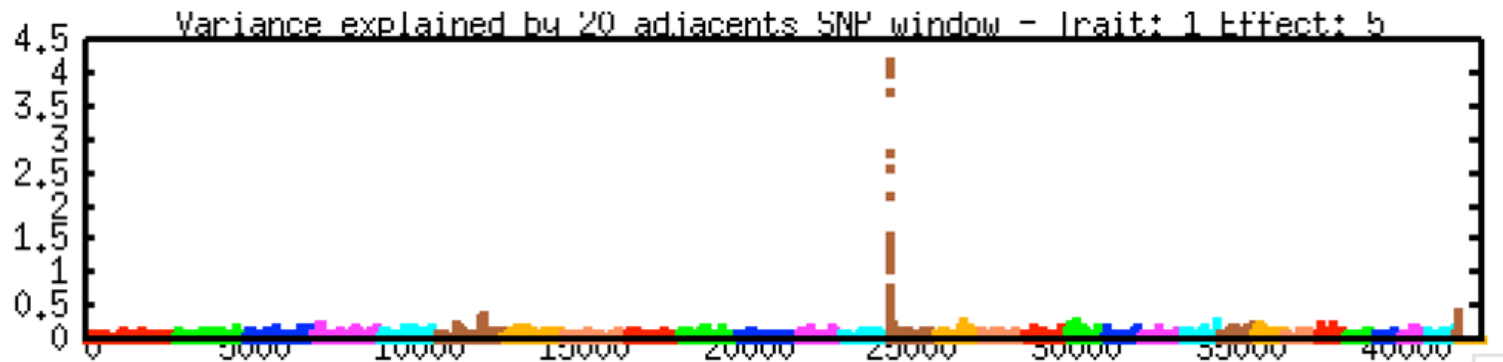
SNP weighting/selection
possibly artifact of
inadequate modeling

Can large QTL exist despite selection?

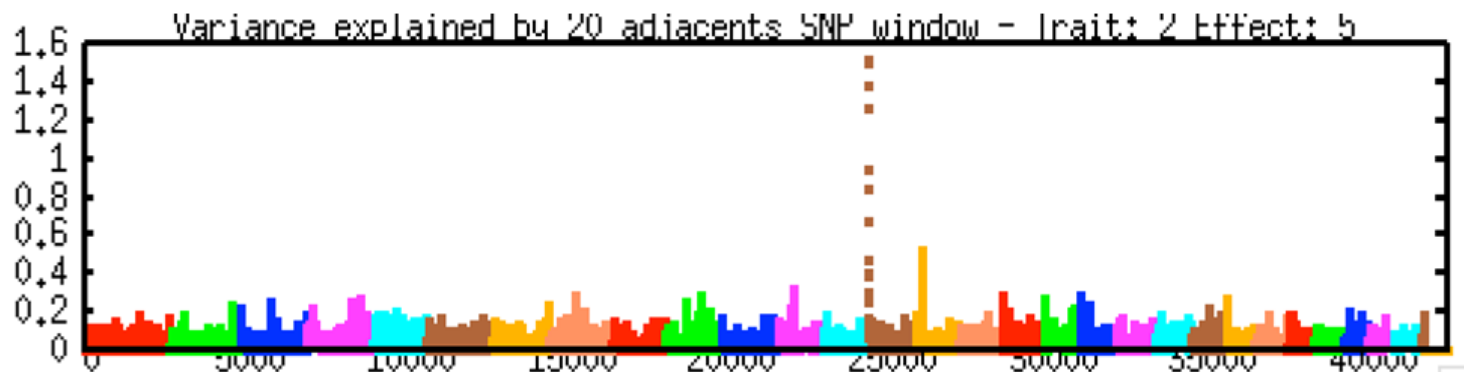


- Genetics and genomics of mortality in US Holsteins
- (Tokuhisa et al, 2014; Tsuruta et al., 2014)
- 6M records, SNP50k genotypes of 35k bulls

Milk – first parity

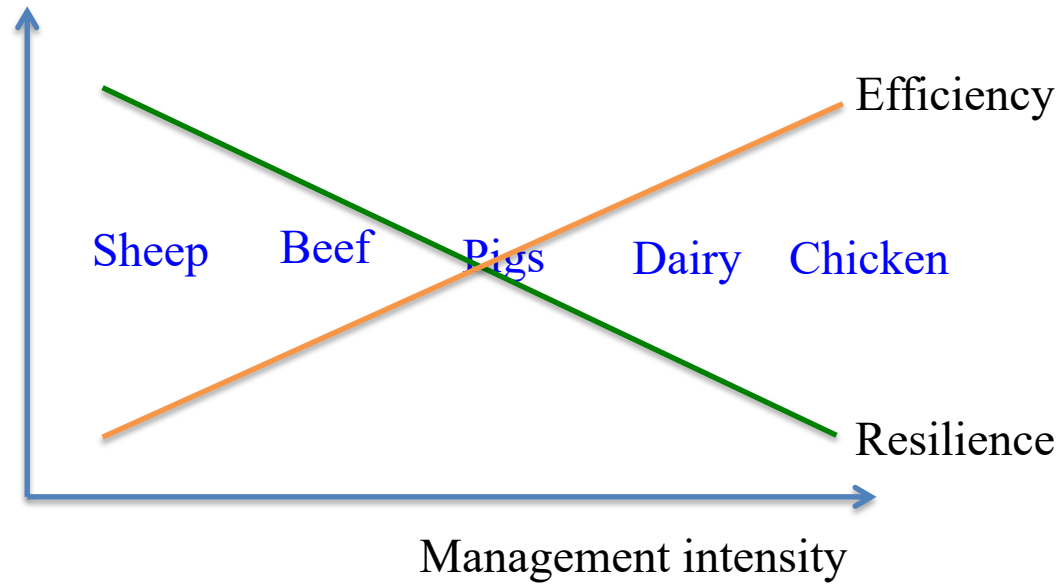


Mortality – first parity



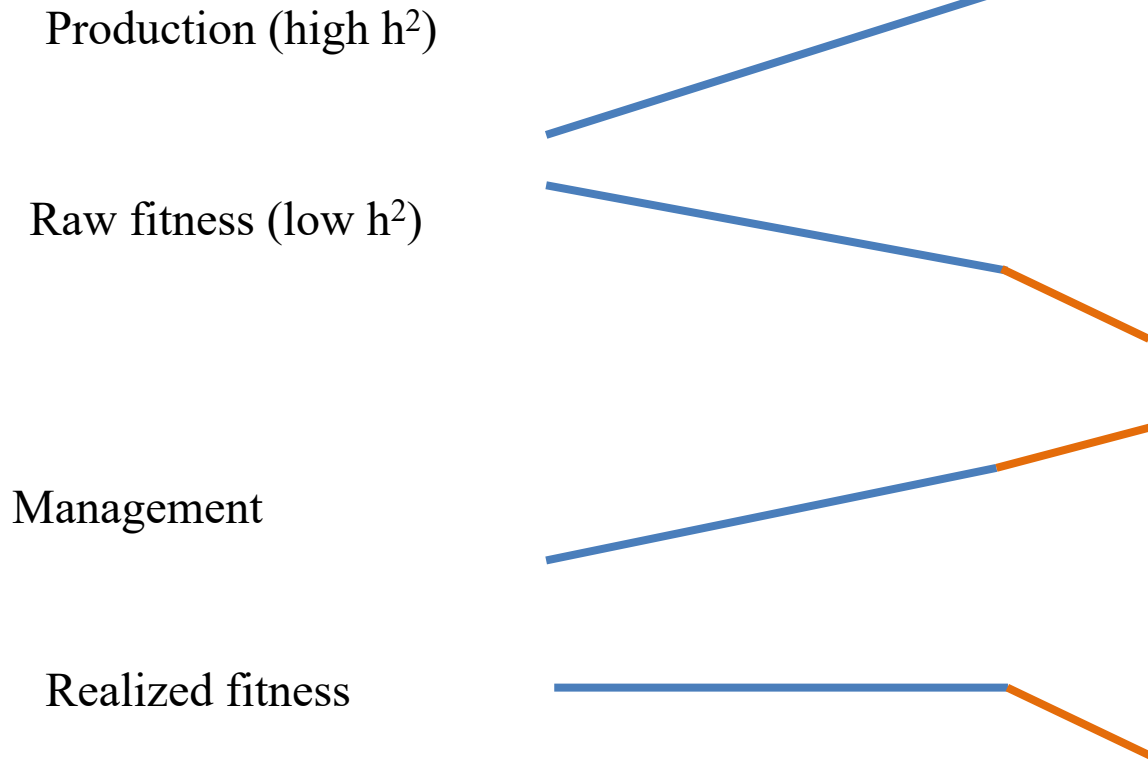


Resilience/efficiency and management intensity



Trends

Genomic selection



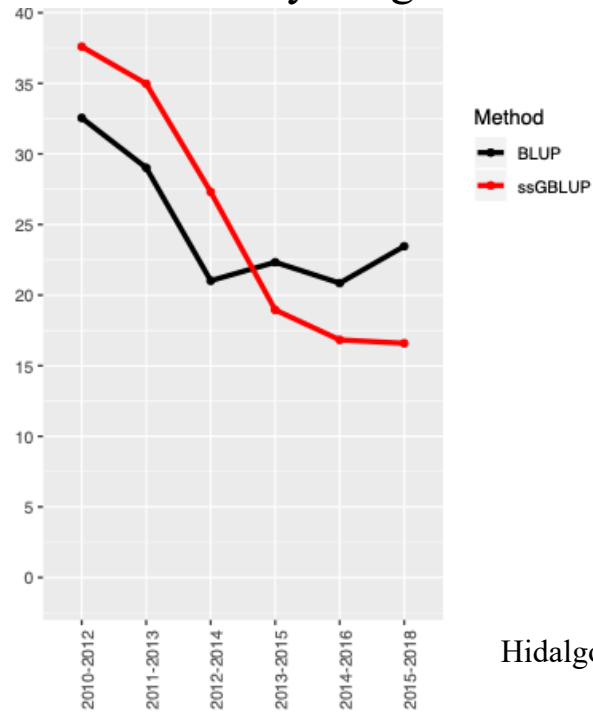
Parameter changes with genomic selection

- Better accuracy assuming 50% heritability for US Holsteins
- Bulmer effect?
- With changes, benefits below expectations
- Problem with estimating changes
 - Biases without genomic information
 - High cost with genomic evaluation

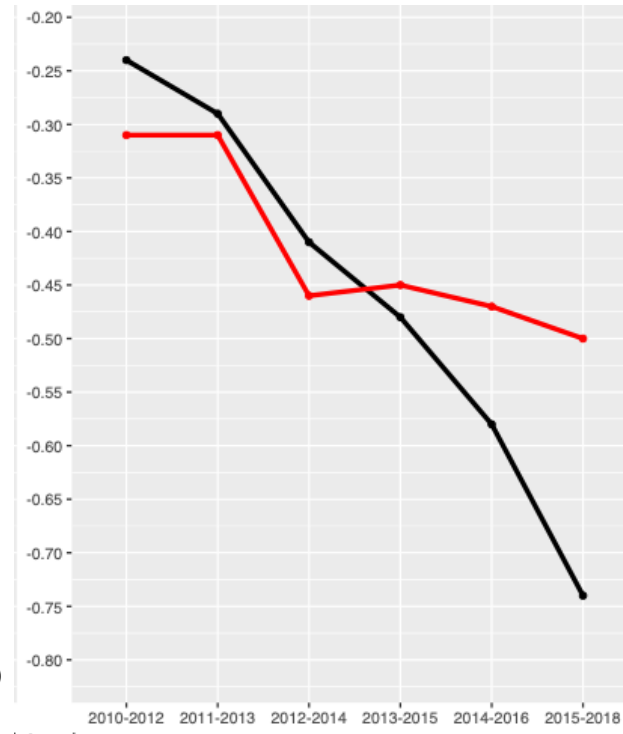
Changes in (co)variances in pigs due to genomic selection



Heritability for growth



Genetic correlation with reproduction



Hidalgo et al. (2019)

Possibilities and challenges ahead

- Possibilities to create perfect genotype for each environment
- 30,000 genes and changing environment
- Selection on gene networks?
- Will selection on many genes create epistasis and nonlinearities?
- Perfect animal or unbalanced animal?

Conclusions

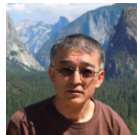
- Genomic selection relies on small effective population size
- Under genomic selection BLUP biased
- Single-step GBLUP standard and suitable for millions of genotypes
- Different results with small and large data
- UGA has access to data - open to visitors

Animal Breeding and Genetics Group

<http://nce.ads.uga.edu>



Ignacy
Misztal



Shogo
Tsuruta



Ivan
Pocnic



Yvette
Steyn



Jorge
Hidalgo



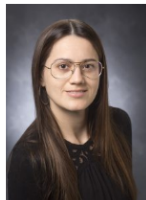
Matias
Berman



Ignacio
Aguilar



Andres
Legarra



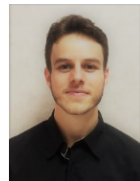
Daniela
Lourenc
o



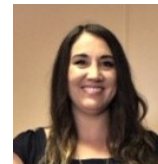
Yutaka
Masud
a



Breno
Fragomen
i



Andre
Garcia



Taylor
Mcwhorthe
r



Diogo
Silva



Zulma
Vitezic
a

Acknowledgements



United States
Department of
Agriculture

National Institute
of Food and
Agriculture