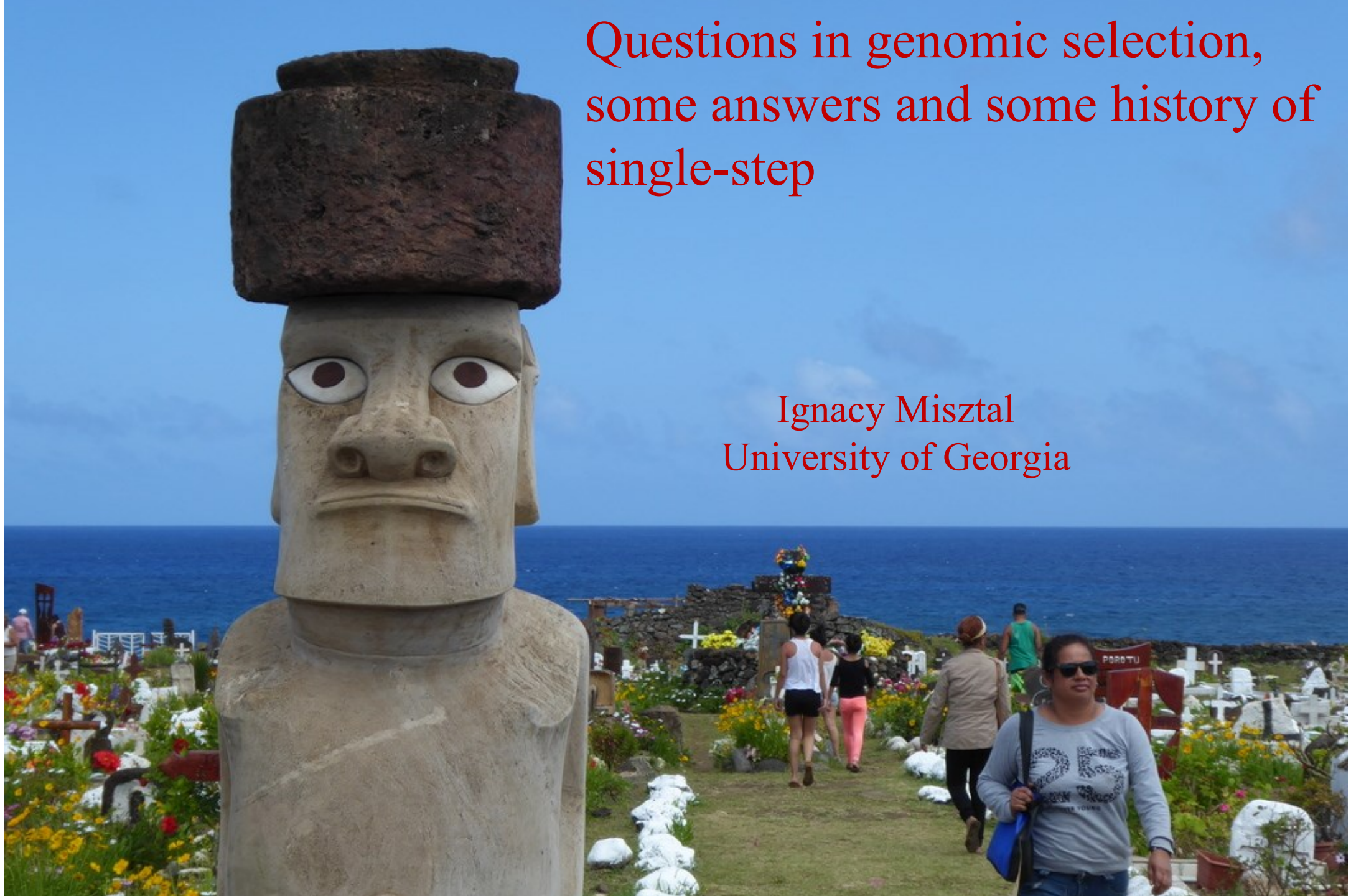
# Questions in genomic selection, some answers and some history of single-step



Ignacy Misztal
University of Georgia

# 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?
- Stability problems with GRM
  - At about 5k, usually blended with A
- OK accuracy with few genotyped animals 1k-2k
  - Good in farm
  - Rise with extra genotypes slow
  - Discrepancy between simulation and field-data results

# Inversion by recursion

$$u_i \mid u_1, u_2, .., u_{i-1} = \mathbf{p}_i' \mathbf{u} + \varphi_i$$ Generic recursion

$$\mathbf{u} = \mathbf{P}\mathbf{u} + \mathbf{\Phi}$$

$$\mathrm{var}(\mathbf{u})^{-1} = (\mathbf{I} - \mathbf{P})' \mathrm{var}(\mathbf{\Phi})^{-1} (\mathbf{I} - \mathbf{P})$$ Cost low only if P sparse

For pedigree relationships (Henderson, 1976):

$$u_i = 0.5 u_{si} + 0.5 u_{di} + \varphi_i$$ P very sparse

Is limited recursion applicable to genomic relationships?

# Algorithm for proven and young animals (APY)

For young animals

=0 in GBLUP

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

Misztal et al. (2014)

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

$\mathbf{Z_p}$ – genotypes for proven animals

$\mathbf{Z_y}$ – genotypes for young animals

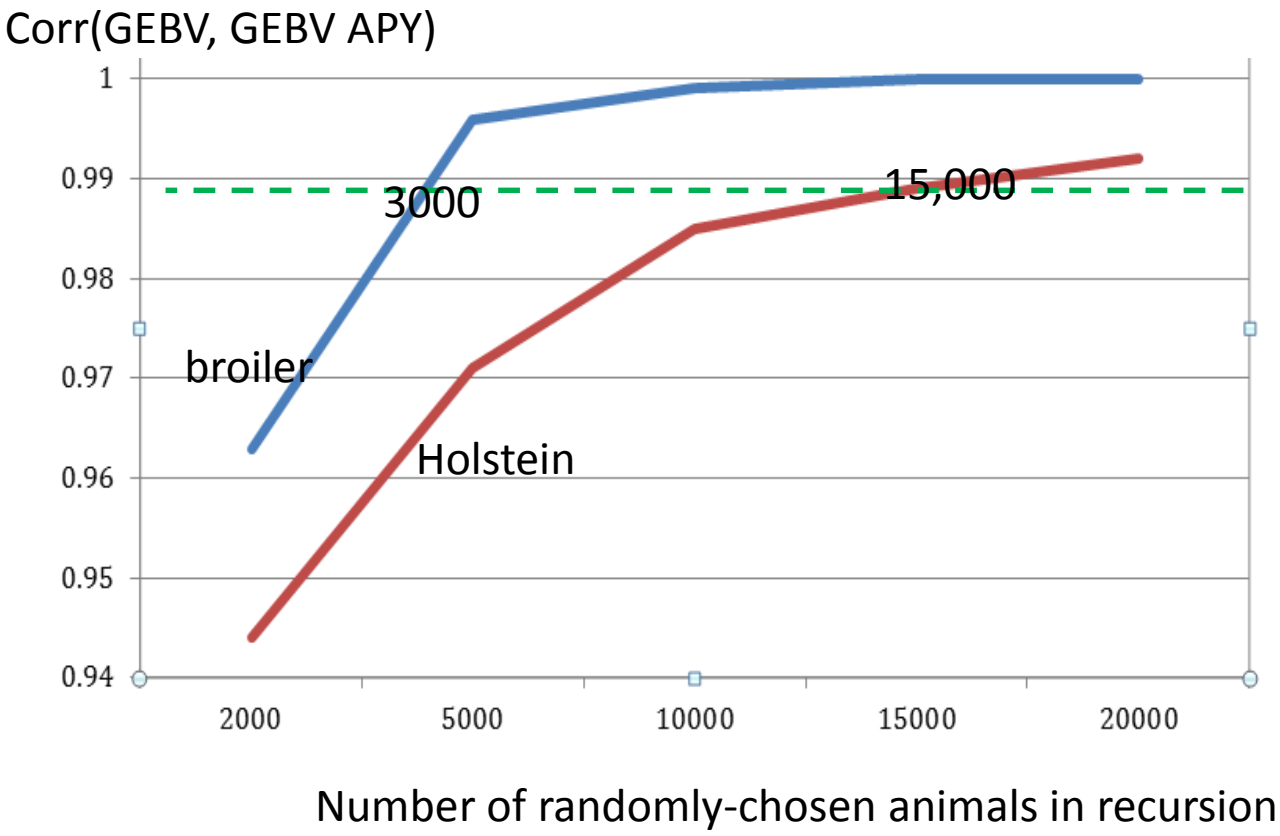
$m_i = g_{ii} - \mathbf{z_i'Z_p'G_{pp}^{-1}Z_p z_i}$

Linear cost for young animals

# Tests with Holsteins (Fragomeni et al., 2015)

| | G needed | $G^{-1}$ | Correlations of GEBV with regular inverse |
|---|---|---|---|
| Regular inverse | | | |
| APY inverse | | | |
| 23k bulls as core | | | > 0.99 |
| 17k cows as core | | | > 0.99 |
| 20k random animals as core | | | > 0.99 |

# Impact of recursion size in Holsteins and chicken



# Theory of junctions

Heterogenetic and homogenic tracts in genome (Stam, 1980)

……  ……

Called independent chromosome segments Me
(Goddard et al., 2009; Daetwyler et al., 2010)

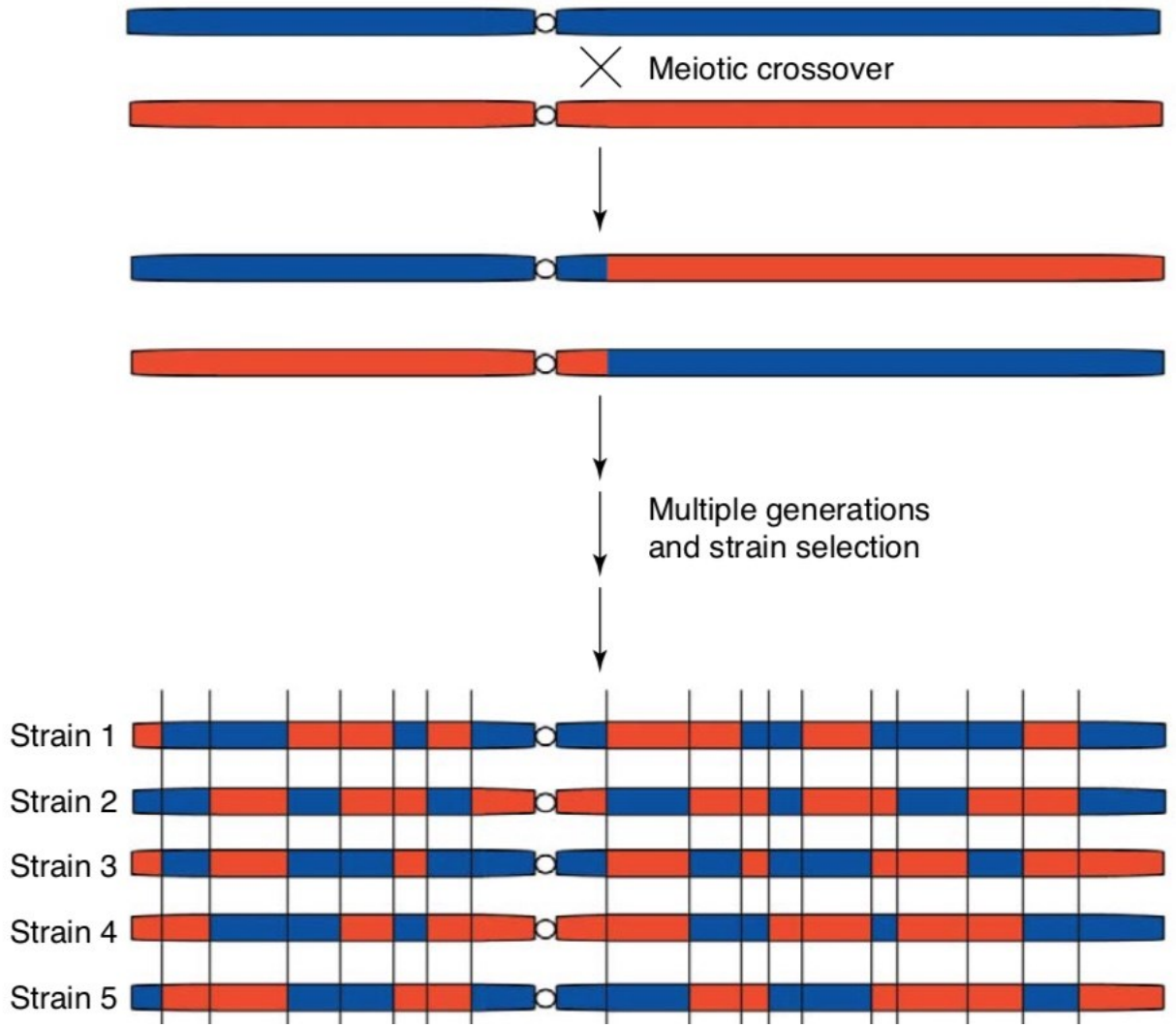
E(Me)=4NeL (Stam, 1980)

Ne – effective population size
L –length of genome in Morgans

Need 12 Me SNPs to detect 90% of junctions
(MacLeod et al., 2005)

# Haplotype blocks = Independent chromosome segments

- $E(Me) = 4N_eL$ Stam (1980)
  - $N_e$ – Effective population size
  - L – Length of genome in Morgans
- Me
  - $2N_eL$ Hayes *et al.* (2009)
  - $2N_eL/[\log(N_eL)]$ Goddard *et al.* (2011)
  - Many more Brard and Ricard (2015)




Cuppen *(2005)*

# Theory of APY based on segments

Breeding value chromosome segments

$$\mathbf{u} = \mathbf{Ts}$$

Choose core "**c**" and noncore "**n**" animals

$$\mathbf{s} = \mathbf{Q}\mathbf{u}_c + \boldsymbol{\varepsilon}_c$$

$$\mathbf{u}_n = \mathbf{P}_{nc}\mathbf{u}_c + \boldsymbol{\varepsilon}_n$$

small if number of core animals > number of segments

Choose core "**c**" and noncore "**n**" animals

$$\mathbf{u}_c = \mathbf{u}_c$$

BV of *noncore* animals linear function of *core* animals

$$\mathbf{u}_n = \mathbf{P}_{nc}\mathbf{u}_c + \varepsilon_n$$

Matrix notation

$$\begin{bmatrix} \mathbf{u}_c \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{P}_{nc} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{u}_c \\ \boldsymbol{\varepsilon}_n \end{bmatrix}$$

Var(u)

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{P}_{nc} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{G}_{cc} & \mathbf{0} \\ 0 & \mathbf{M}_{nn} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{P}_{cn} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

The inverse

$$\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{P}_{cn} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{G}_{cc}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{nn}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{P}_{nc} & \mathbf{I} \end{bmatrix}$$

Misztal&Legarra&Aguilar (2014)

Unknown matrices from conditional expectation

$$\mathbf{P}_{nc} = \mathbf{G}_{nc}\mathbf{G}_{cc}^{-1}, \quad \mathbf{M}_{\mathrm{nn}} = diag\{g_{i,i} - \mathbf{p}_{\mathrm{i,1:}i-1}\mathbf{g}'_{i,\mathrm{1:}i-1}\}$$

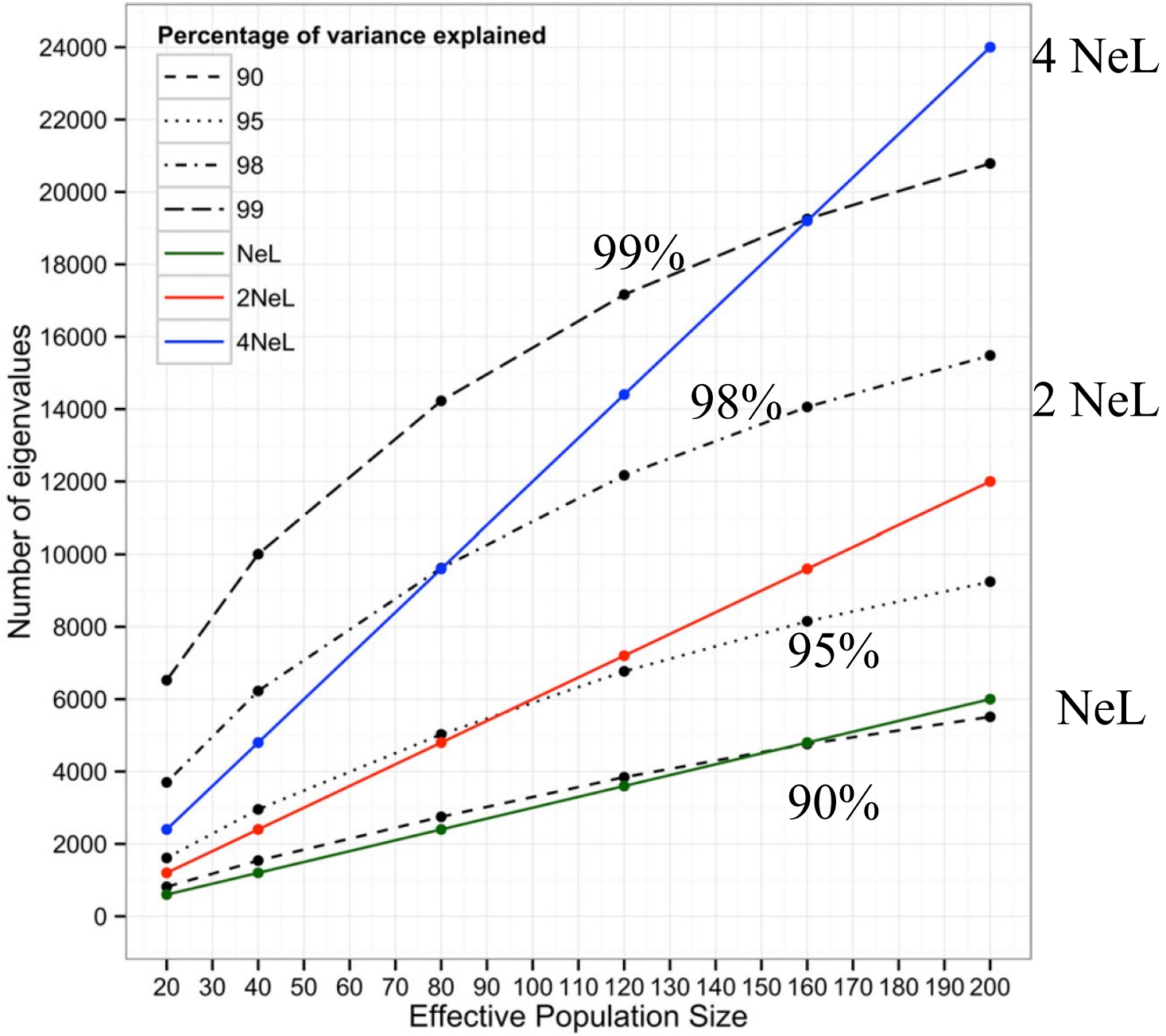
# Finding dimensionalities by eigenvalues

$$G = U \; D \; U' = U_s \; D_s \; U_s'$$

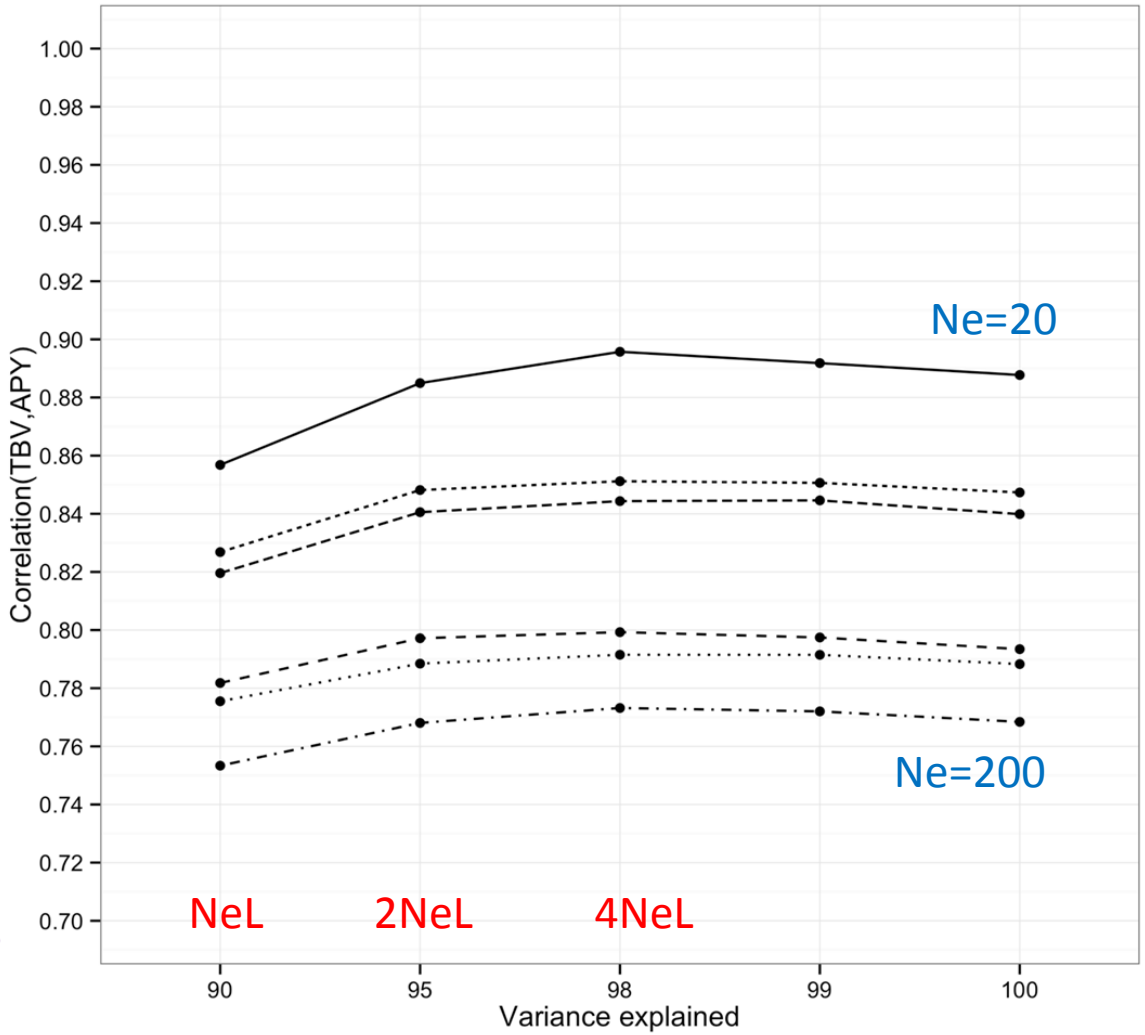
U – eigenvalues
D – eigenvectors

Eigenvalues sum to 100%

What % is useful, 95%? 98% 99%, 99.999%?


Percentage of variance explained
90
95
98
99
NeL
2NeL
4NeL
Number of eigenvalues
24000
22000
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
Effective Population Size
99%
98%
95%
90%
4 NeL
2 NeL
NeL

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

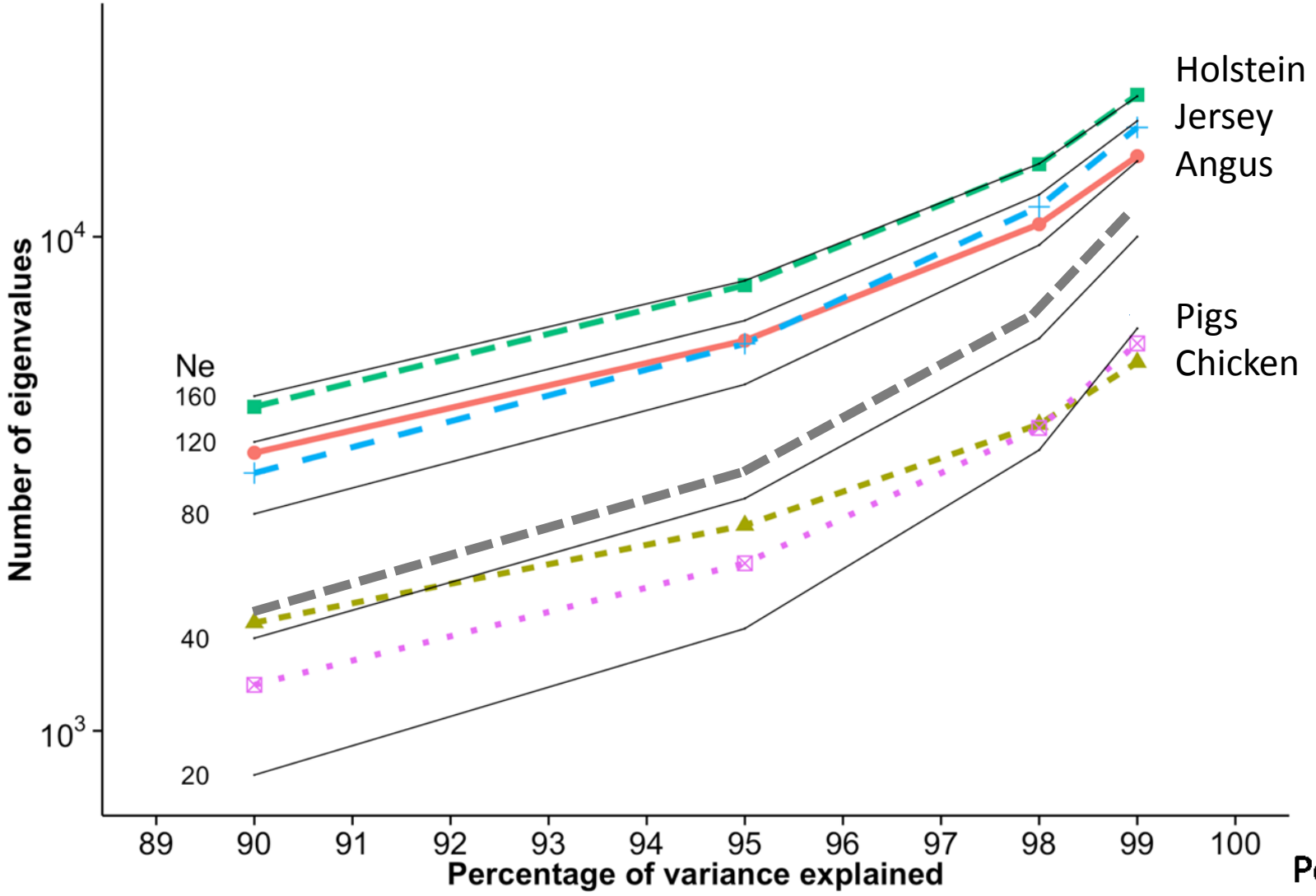
Pocrnic et al., 2016a

GENETICS | INVESTIGATION

# The Dimensionality of Genomic Information and Its Effect on Genomic Prediction

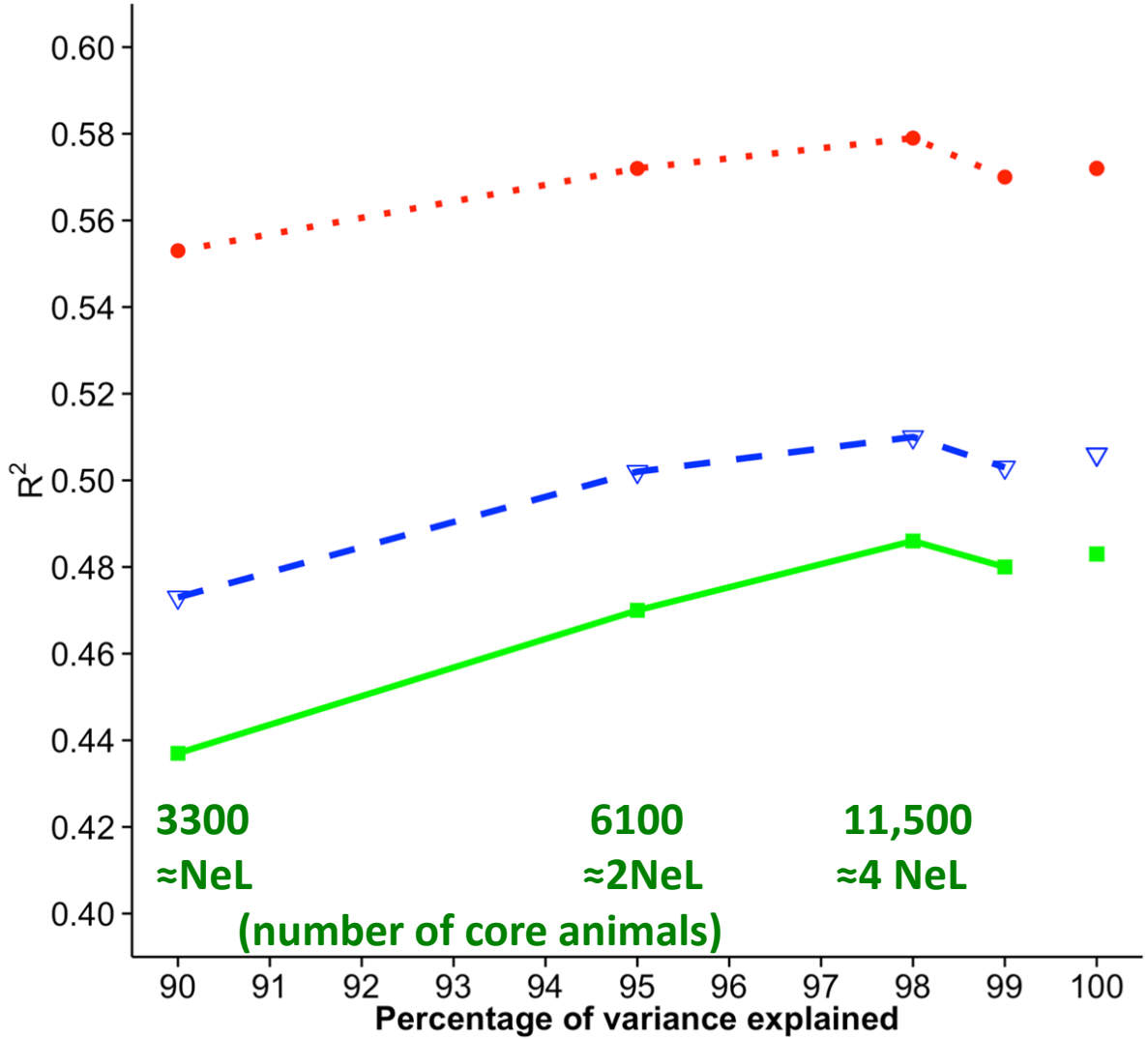
**Ivan Pocrnic,*,[1] Daniela A. L. Lourenco,* Yutaka Masuda,* Andres Legarra,[†] and Ignacy Misztal***
*Department of Animal and Dairy Science, University of Georgia, Athens, Georgia 30602, and [†]Institut National de la Recherche Agronomique, GenPhySE, F-31326 Castanet-Tolosan, France

# Number of eigenvalues in G to explain given fraction of variability




Pocrnic et al., 2016a

# Reliabilities – Jerseys (75k animals)




Pocrnic et al., 2016b

100% = full inverse ➔ lower accuracy

# Estimated dimensionality, effective population size and optimal number of SNP

| Specie | Approx Me (98%) | Effective population size (L=30M) | Optimal number of SNP (12 x Me) |
|---|---|---|---|
| Holsteins | 14k | 149 | 170k |
| Jerseys | 10k | 101 | 120k |
| Angus | 11k | 113 | 130k |
| Pigs | 4k | 43 (L=20M) | 50k |
| Chicken | 4k | 44 | 50k |

Pocrnic et al. (2016b)

# Side effects of reduced dimensionality

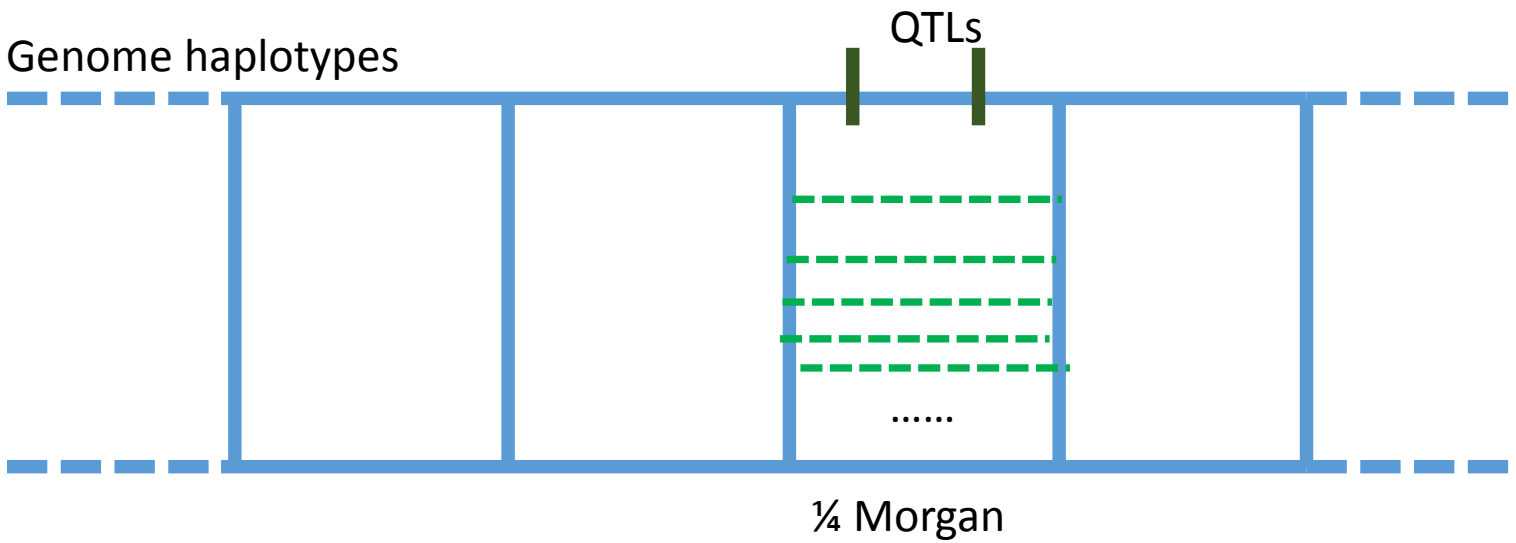
- Number of segments
  - 800k in humans
  - 5-15k in animals

- Impact on SNP selection and GWAS

# Theory of limited dimensionality

Number of haplotypes: 4 Ne L

Ne within each ¼ Morgan segment

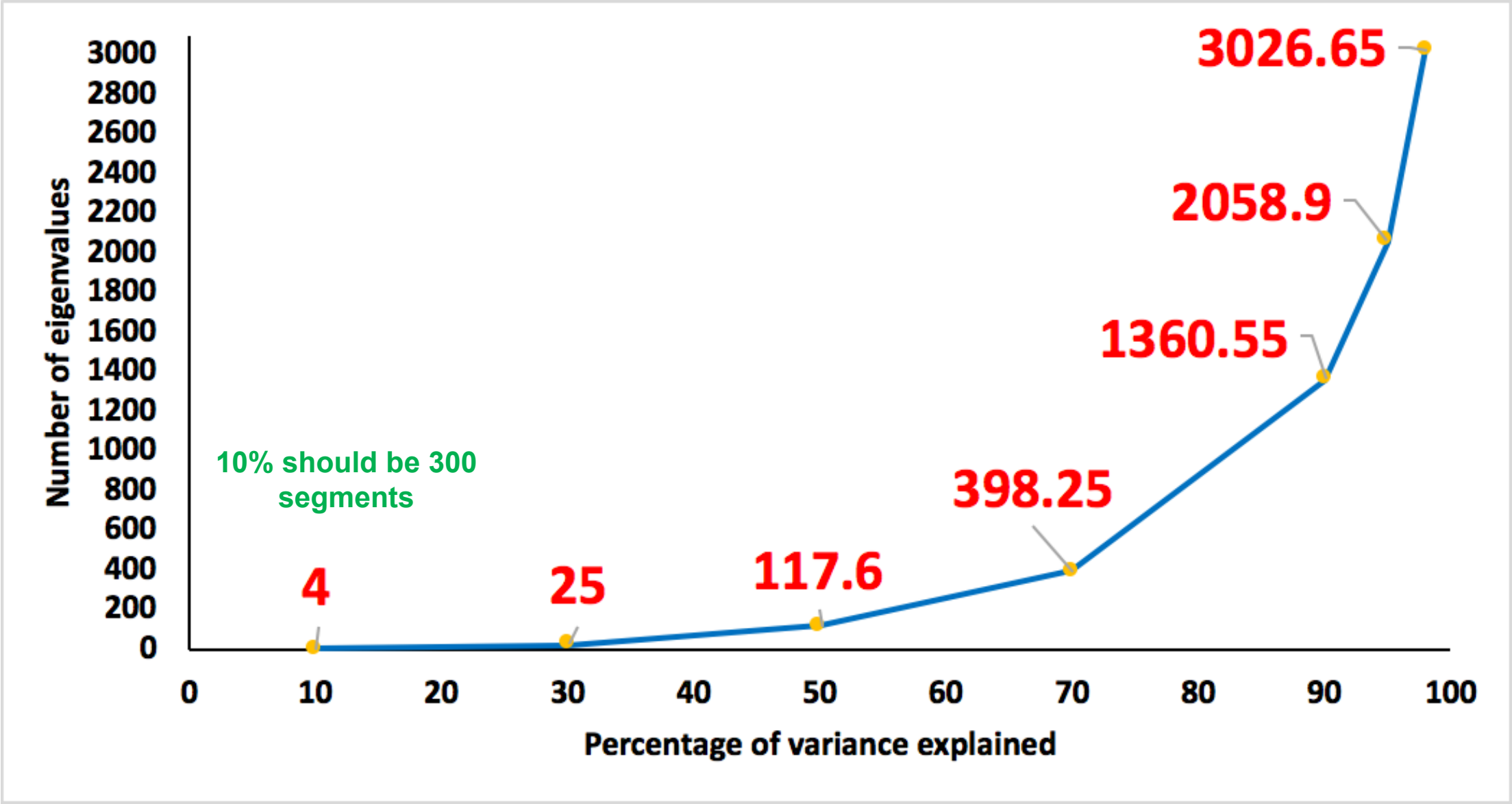


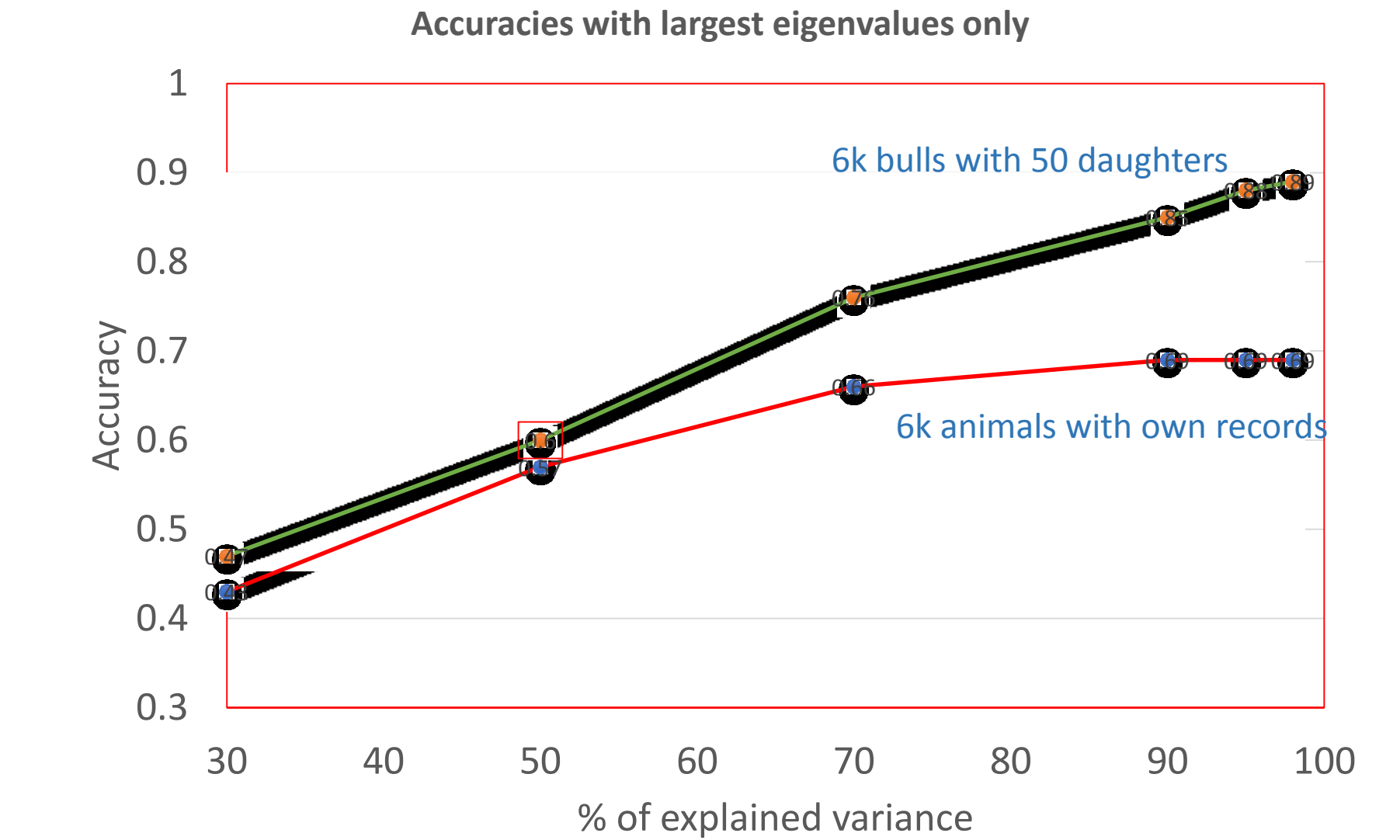

Ne haplotypes within each ¼ Morgan segment

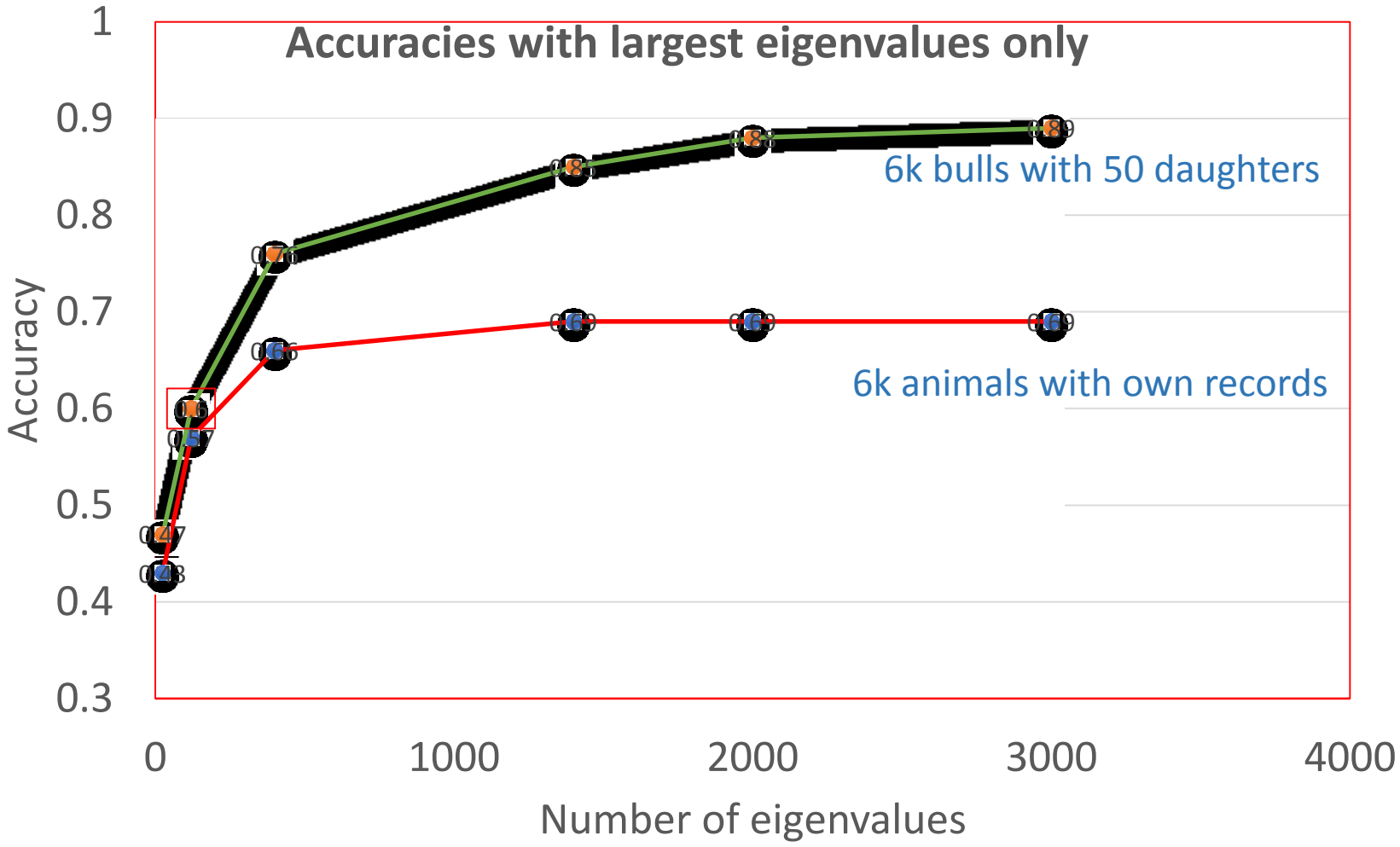
Dimensionality of ¼ Morgan case: Ne or number of identified QTLs

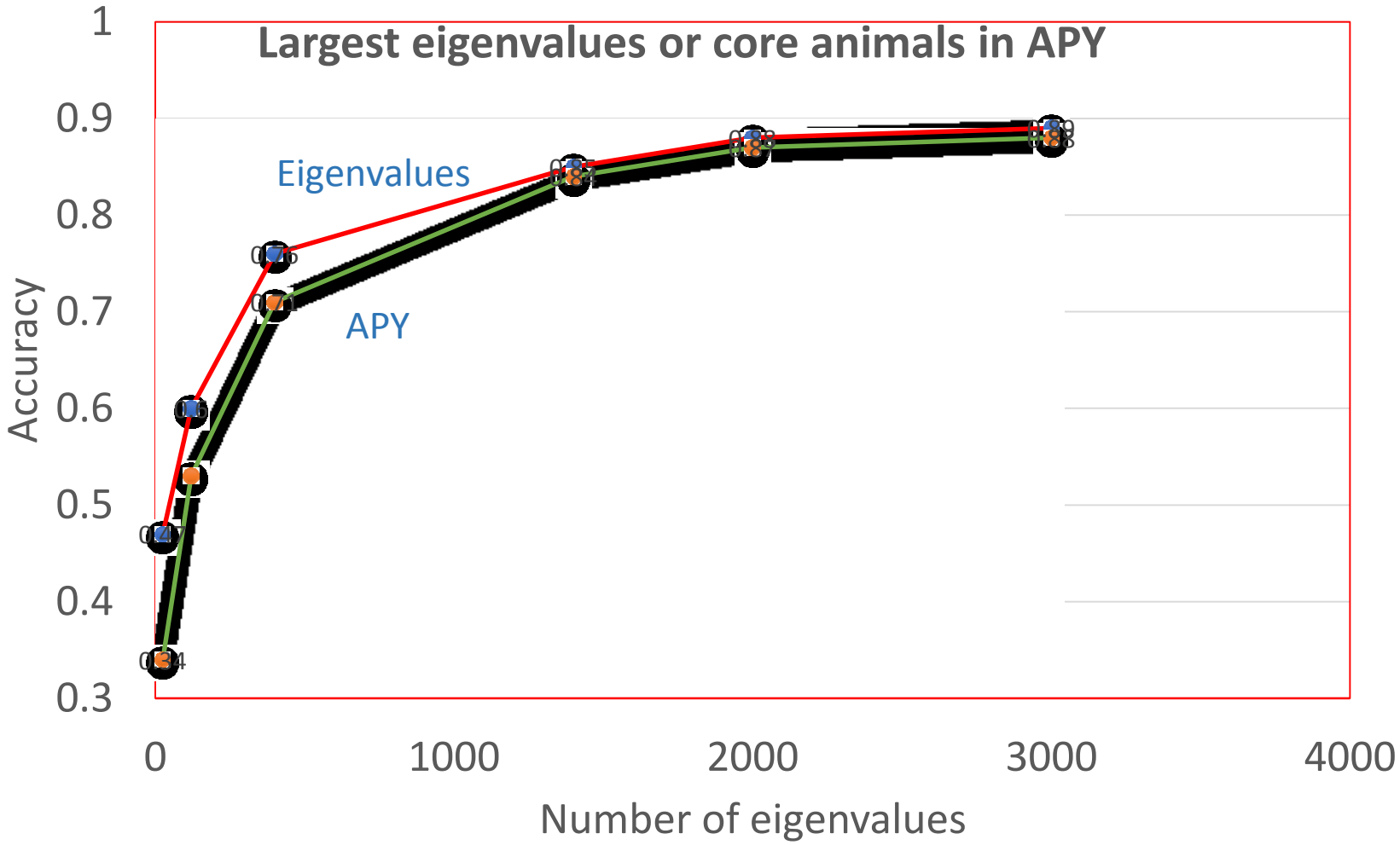
➔ Reduced dimensionality with weighted GRM

# Eigenvalue profile




Accuracies with largest eigenvalues only
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
Accuracy
6k bulls with 50 daughters
6k animals with own records
30
40
50
60
70
80
90
100
% of explained variance
Eigenvalues 25 120 400 1400 2000 3000


Accuracies with largest eigenvalues only
6k bulls with 50 daughters
6k animals with own records
Accuracy
Number of eigenvalues
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
1000
2000
3000
4000


Largest eigenvalues or core animals in APY
Eigenvalues
APY
Accuracy
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
1000
2000
3000
4000
Number of eigenvalues

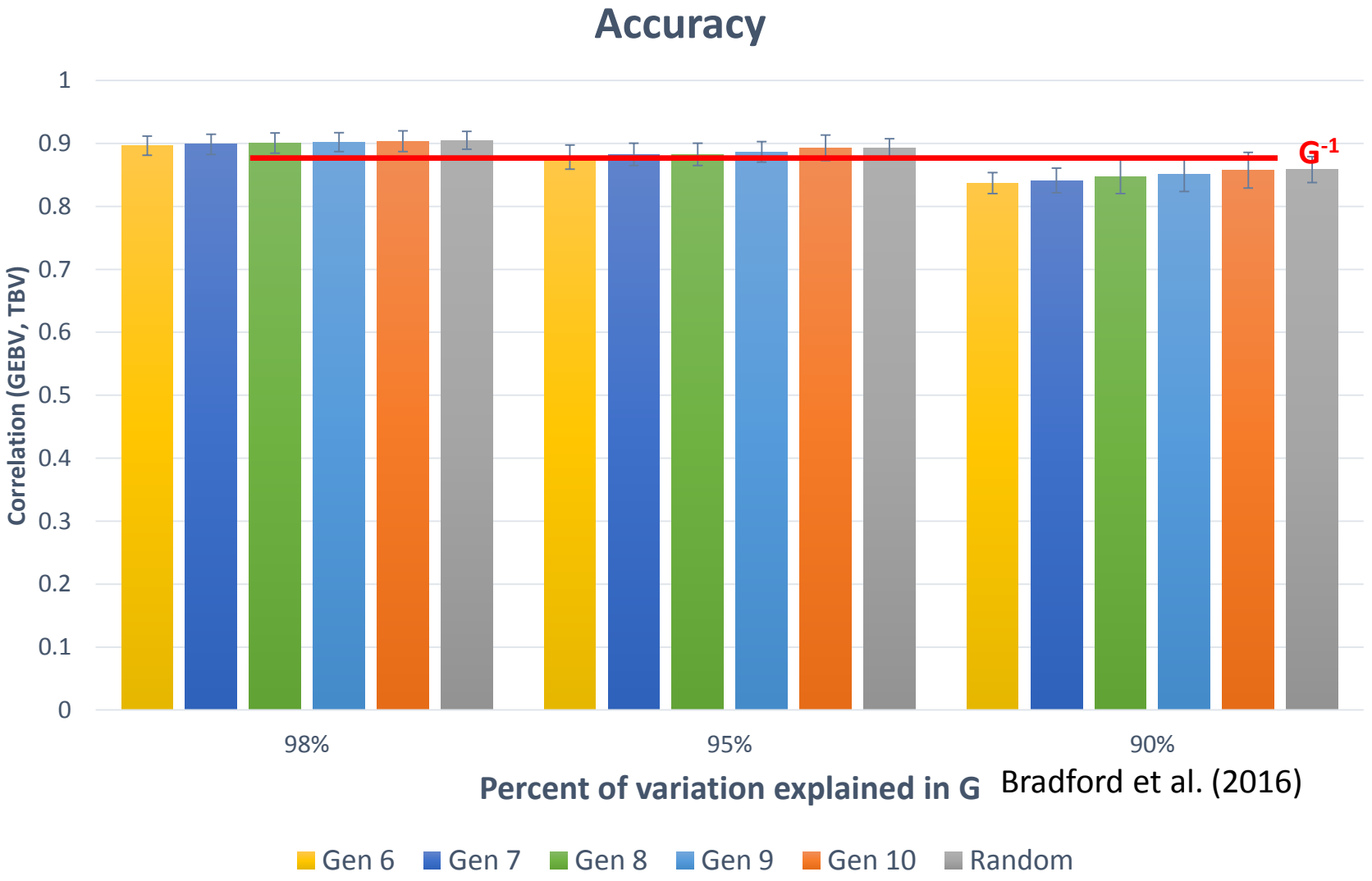
# Which core animals in APY?

Bradford et al. (2017)

- Simulated populations (QMSim; Sargolzaei and Schenkel, 2009)
- Ne = 40
- #genotyped animals = 50,000

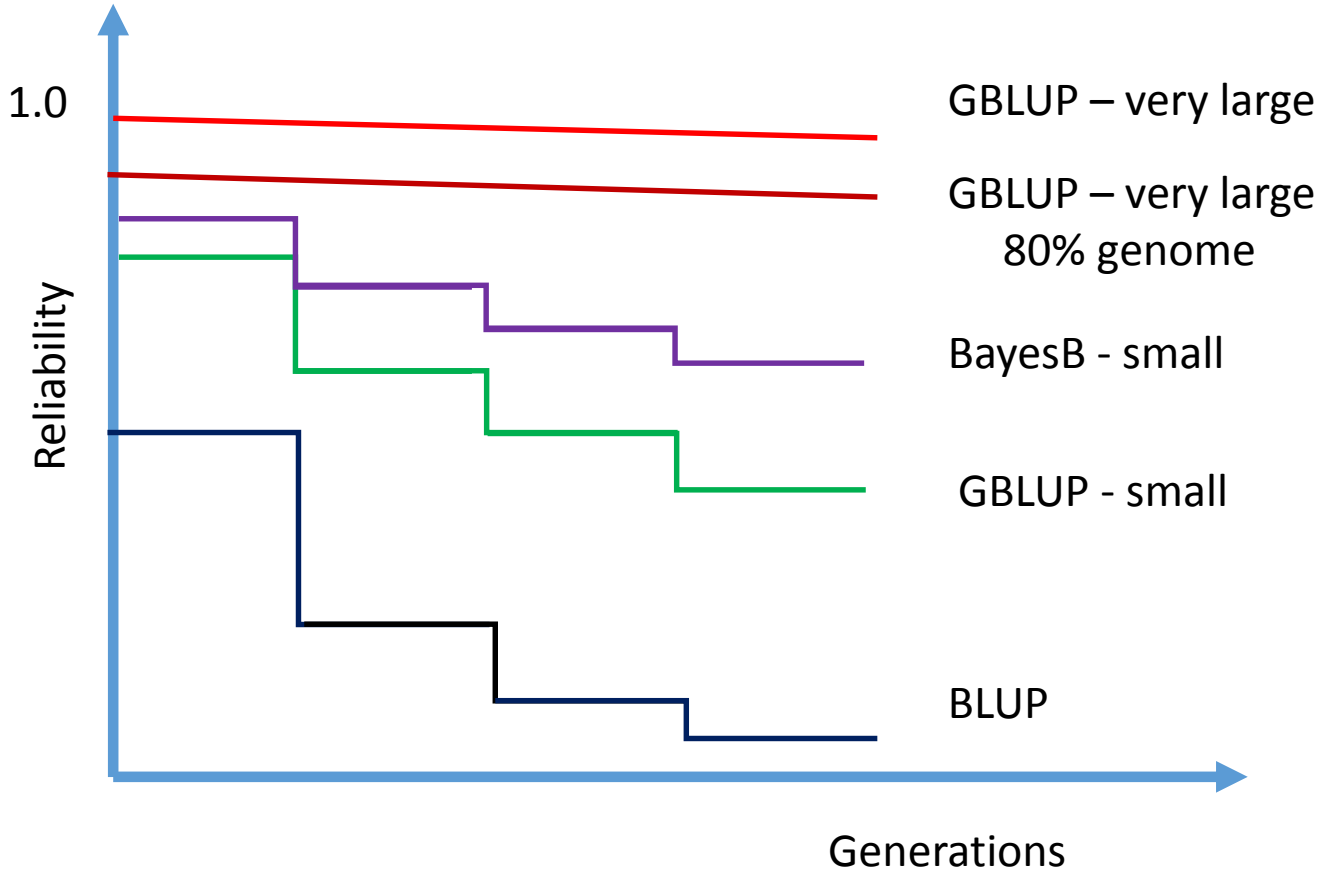
- Core animals:
  - Random gen 6 || gen 7 || gen8 || gen9 || gen 10 (y)
  - Random all generations
  - Incomplete pedigree
  - Genotypes in gen 9 and 10 imputed with 98% accuracy

# Which core animals in APY?




Bradford et al. (2016)

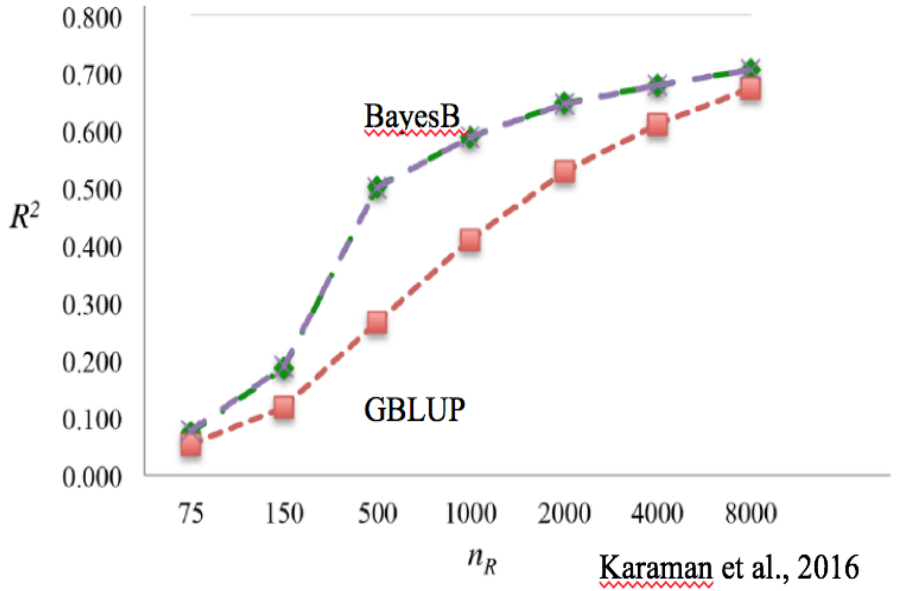
# Persistence over generations




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

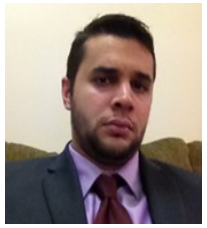
# Multitrait ssGBLUP: Is SNP selection important? Causative SNPs?

- SNP selection/weighting (BayesB, etc.)
  - Large impact with few genotypes
  - Little or no impact with many

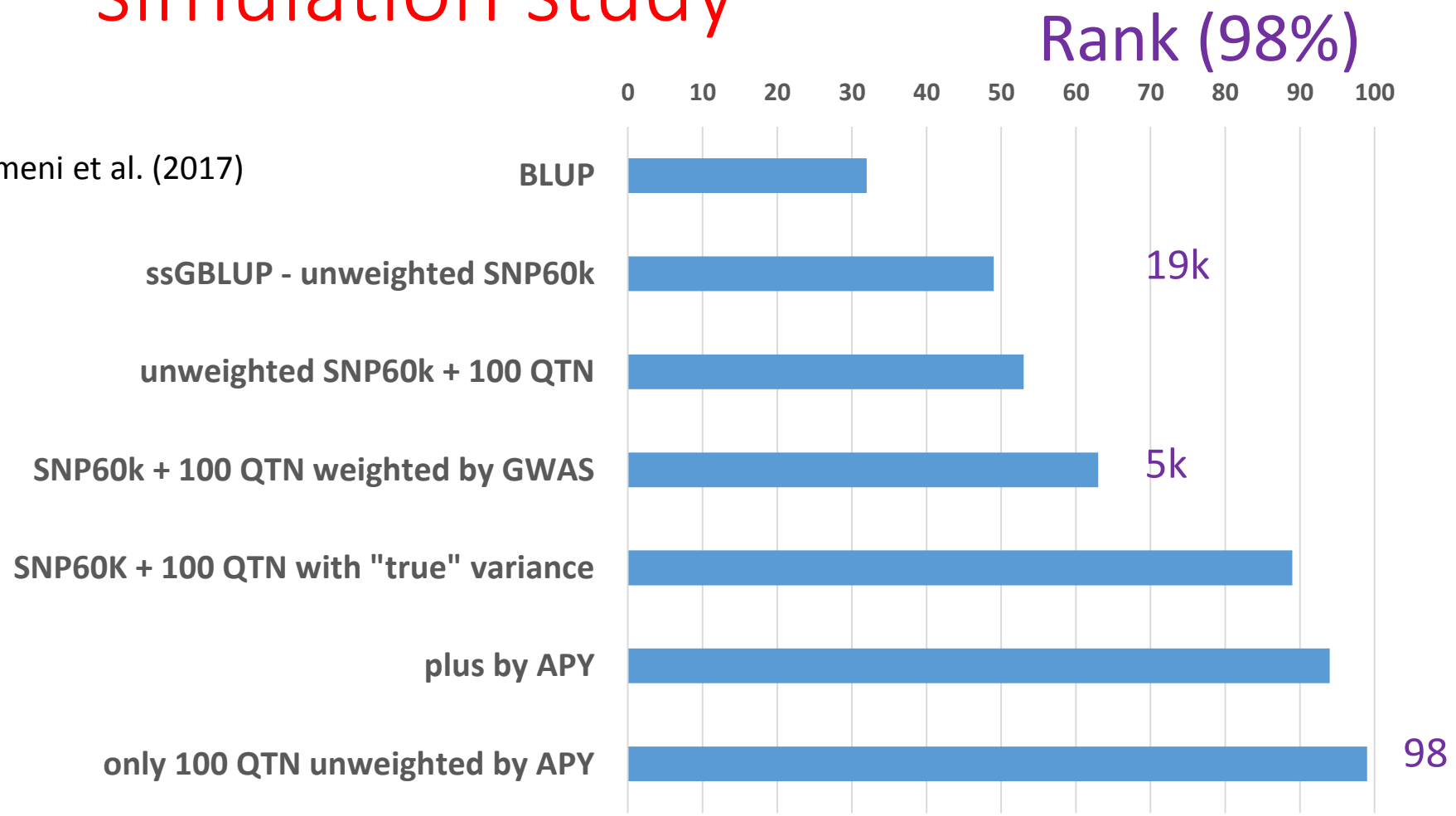



Karaman et al., 2016

# ssGBLUP accuracies using SNP60K and 100 QTNs – simulation study

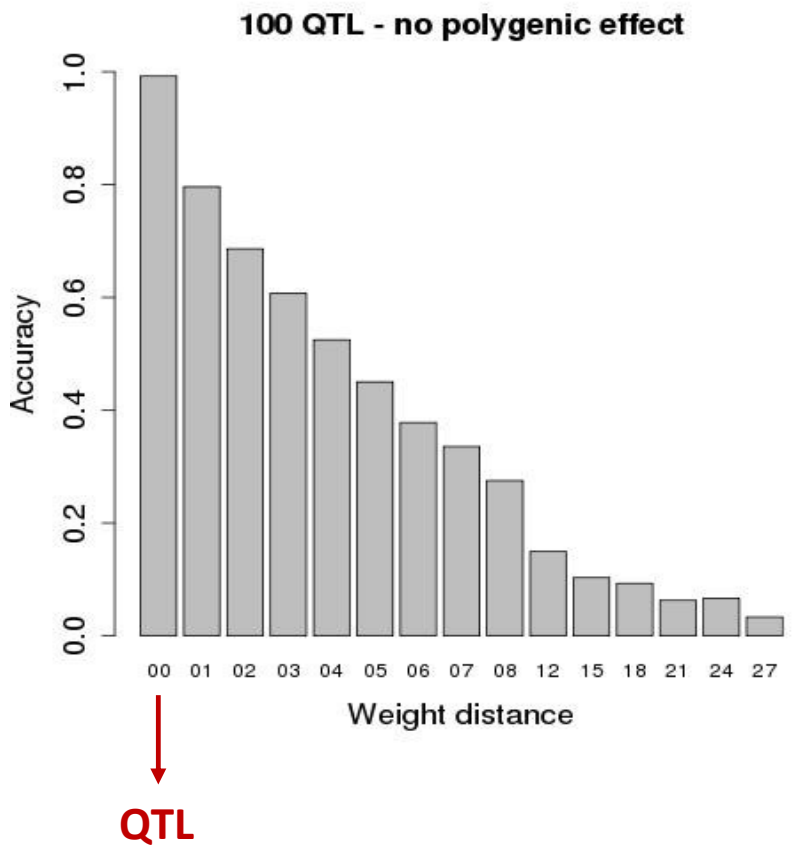


Fragomeni et al. (2017)



# Accuracy and distance from markers to QTL

Fragomeni et al. (2017)



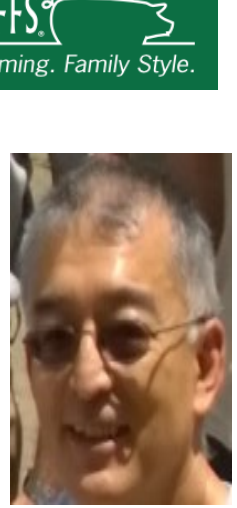
Nothing can be more fatal to progress than a too confident reliance on mathematical symbols; for the student is only too apt to take the easier course, and consider the *formula* not the *fact* as the physical reality.”

Kelvin

# Group and sponsors

United States Department of Agriculture

National Institute of Food and Agriculture



Shogo Tsuruta

Ignacio Aguilar

Breno Fragomeni

Ivan Pocrnic

Daniela Lourenco

Yutaka Masuda

Andres Legarra

Heather Bradford

J. Anim. Breed. Genet. ISSN 0931-2668

EDITORIAL

# Shortage of quantitative geneticists in animal breeding

More and more I receive phone calls from various breeding companies looking to hire a PhD in quantitative genetics. They inquire if I know of a graduate versed in quantitative genetics and mixed models, with some programming skills, who can speak and write passable English, has a general understanding of markers and molecular genetics, can run and troubleshoot a genetic evaluation, and in general be a problem solver. I do not know of anyone available, I reply. There were many of them 10–15 years ago, but now they are rare. If they show up, they usually have very good offers well before graduation. My colleagues outside the USA are telling me of similar problems, although the severity of the PhD shortage is country dependent.

Why are PhDs in animal breeding with quantitative skills rare in the USA as well as in many other countries? Some 15 years ago there was a shift in governmental funding away from animal breeding and quantitative genetics to almost exclusively

Great hopes were put into finding markers for major genes (QTL) that could help solve the new challenges. Based on many association studies, there is growing consensus that few markers/QTLs can be detected, those that were detected had their estimated effects inflated, and that the benefits of using markers are limited. Of all markers found, very few were for low-heritability traits.

The new trend in animal breeding is genomic selection using SNP chips. In this methodology, one estimates effects of individual haplotypes, and genomic EBV (GEBV) is estimated as a sum of those effects. No effort is made to identify QTLs. The genomic selection is based on an assumption opposite from the previous effort in markers but the same as in 'black box' genetics: that a large number of genes are responsible for a trait.

When only a small fraction of the population is genotyped, the estimates of haplotype effects will be derived via EBV obtained through classical methods

# Development of the combined matrix

## Initial (Misztal et al., 2009)

$$H = A + \begin{bmatrix} 0 & 0 \\ 0 & G - A_{22} \end{bmatrix}$$

1-ungenotyped animals
2-genotyped animals

## Comprehensive (Legarra ,2009)

$$\mathbf{H} = \mathbf{A} + \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix} [\mathbf{G} - \mathbf{A}_{22}] [\mathbf{I} \quad \mathbf{I}] \begin{bmatrix} \mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

## Inverse of Comprehensive (Aguilar et al., 2010)

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

Christensen and Lund, 2010
Boemcke et al., 2010

# Implementation at UGA

- Module genomic in BLUPF90 package (Aguilar et al. 2011)
- Option SNP_File xxx in RENUMF90
- Lots of options with defaults
- Creation of $\mathbf{G}^{-1}$: minutes for 10k genotypes, hours for 50k genotypes

Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

**Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation**

I. Aguilar[1,2], I. Misztal[1], A. Legarra[3] & S. Tsuruta[1]

1 Animal and Dairy Science Department, University of Georgia, Athens, GA, USA
2 Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, Uruguay
3 INRA, UR631 SAGA, BP 52627, 32326 Castanet-Tolosan, France

# Predictions for US final scores in Holsteins (Aguilar et al., 2010)

| Prediction in 2004 | DD2009 | |
|---|---|---|
| | $R^2$ (%) | Inflation (%) |
| Parent Avg | 24 | 31 |
| Multistep (VanRaden) | +16 | 16 |
| Single-step | | |
| Regular $G^{-1}$-$A_{22}^{-1}$ | +17 | 31 |
| Refined $1.5G^{-1}$-$0.6A_{22}^{-1}$ | +17 | 4 |

**Ranking 11 out of 118**
Food Science and Technology

JDS's 5-Year impact factor is 2.942 and it ranks 1st in Agriculture, Dairy and Animal Science and 12th in Food Science and Technology.

[ADSA Member News]

[Interpretive summaries]

t Cited | ScienceDirect Top 25

**The effect of injectable butaphosphan and cyanocobalamin on postpartum serum β-hydroxybutyrate, calcium, and phosphorus concentrations in dairy cattle**
March 2010 (Vol. 93 | No. 3 | Pages 978-987)
E. Rollin, R.D. Berghaus, P. Rapnicki, S.M. Godden, M.W. Overton
Abstract | Full Text | PDF (166 KB)

**On the use of physical activity monitoring for estrus detection in dairy cows**
January 2010 (Vol. 93 | No. 1 | Pages 249-259)

do so. You will create an account and registered, proceed to activate (or cla online access to the journal.

** Please note that your previous Hig username and password, your "Nam "ADSA Member Number", or your ADS password will not provide you with ac must complete the registration and c

**About JDS** | About ADSA

The official journal of the American Dairy Science Association®, Journal of Dairy Science® (JDS) is the leading general dairy research journal in the world. JDS readers represent education, industry, and government agencies in more than 70 countries with interests in biochemistry, breeding, economics, engineering, environment, food science, genetics, microbiology, nutrition, pathology, physiology, processing, public health, quality assurance, and sanitation.

More

# Multitrait national genomic evaluation for type (Tsuruta et al., 2010)

- US Holsteins (10 million animals)
- 18 traits
- Almost 50,000 genotypes of bulls and cows
- 2 days computing

J. Dairy Sci. 94:4198–4204
doi:10.3168/jds.2011-4256


**Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins**

S. Tsuruta,*[1] I. Misztal,* I. Aguilar,† and T. J. Lawlor‡
*Animal and Dairy Science Department, University of Georgia, Athens 30602
†Instituto Nacional de Investigación Agropecuaria, Las Piedras, Canelones 90200, Uruguay
‡Holstein Association USA Inc., Brattleboro, VT 05301

# Genomic evaluations of broiler chicken (Chen et al., 2010)

- 180k broiler chicken
- 3 k genotyped with SNP60k chip

- 3 methods
  - BLUP- full data
  - BayesA – genotyped subset
  - Single step – subset and full data set

**Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens**

C. Y. Chen,*[1,2] I. Misztal,* I. Aguilar,*† S. Tsuruta,* T. H. E. Meuwissen,‡ S. E. Aggrey,§ T. Wing,# and W. M. Muir‖

*Department of Animal and Dairy Science, University of Georgia, Athens 30602-2771; †Instituto Nacional de Investigación Agropecuaria, Las Brujas 90200, Uruguay; ‡Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, NO-1432 As, Norway; §Department of Poultry Science, University of Georgia, Athens 30602-2772; #Cobb-Vantress Inc., PO Box 1030, Siloam Springs, AR 72761-1030; and ‖Department of Animal Science, Purdue University, West Lafayette, IN 47907-1151

# Accuracies for broiler chickens

| Trait | Accuracy*100 | | | |
|---|---|---|---|---|
| | BLUP | BayesA Subset | Single-step Subset | Single-step Full |
| Body Weight | 56 | +4 | +11 | +12 |
| Breast Meat | 35 | +1 | +0 | +6 |
| Leg Score | 29 | -20 | -23 | +7 |

BayesA – days of computing + errors

Single-step – 2 minutes

Next cycle of selection

Multiple trait

| | | | | | |
|---|---|---|---|---|---|
| Body Weight | 38 | | +13 | +22 | = |
| Breast Meat | 39 | | +10 | +26 | +29 |
| Leg Score | 28 | | -21 | +6 | = |

Forni *et al. Genetics Selection Evolution* 2011, **43**:1
http://www.gsejournal.org/content/43/1/1

GSE Genetics Selection Evolution

RESEARCH Open Access

# Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information

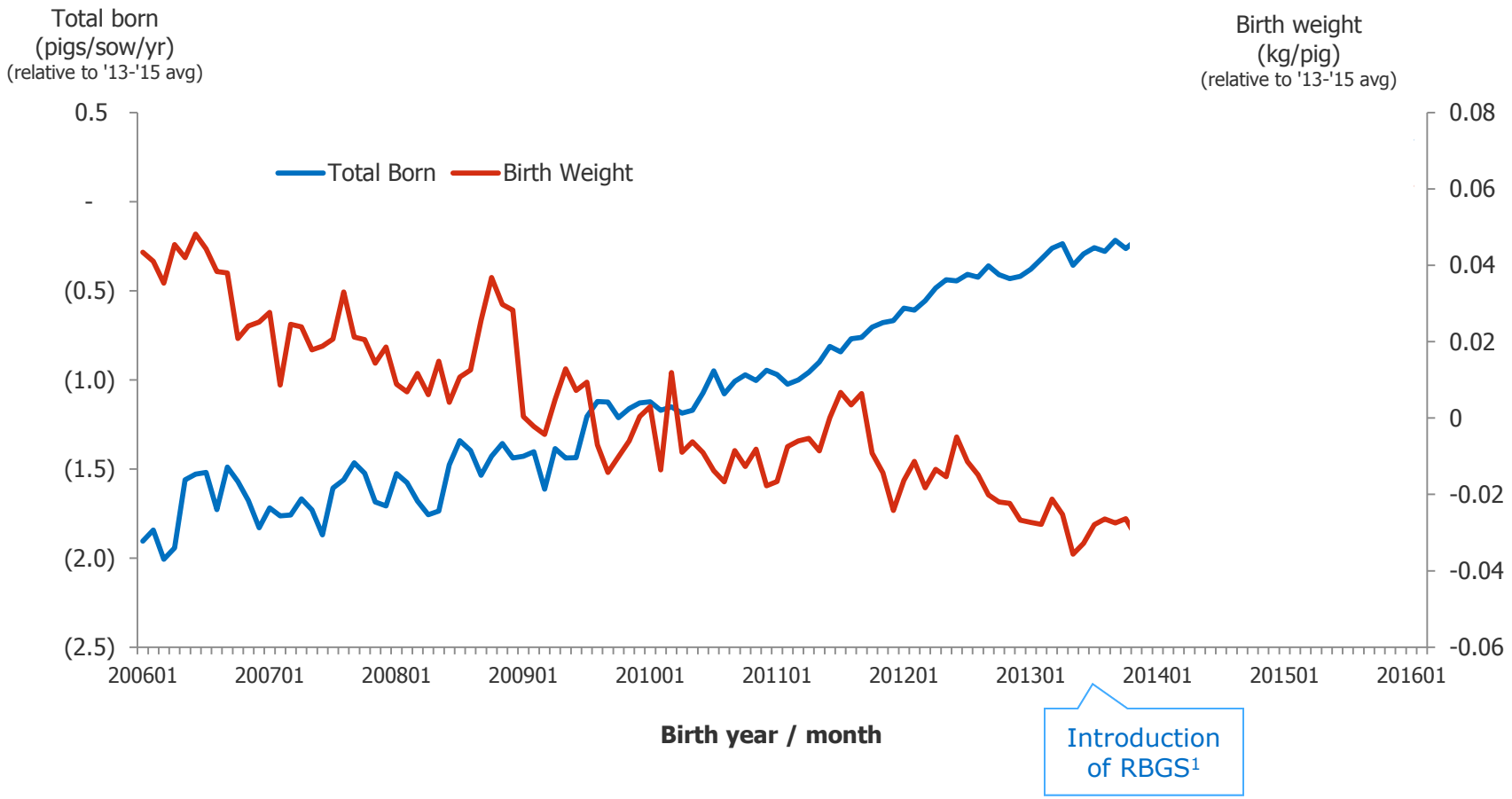
Selma Forni[1*], Ignacio Aguilar[2,3], Ignacy Misztal[3]

# Effect of different genomic relationship matrices on accuracy and scale[1]

**C. Y. Chen,*[2,3] I. Misztal,* I. Aguilar,*† A. Legarra,‡ and W. M. Muir§**

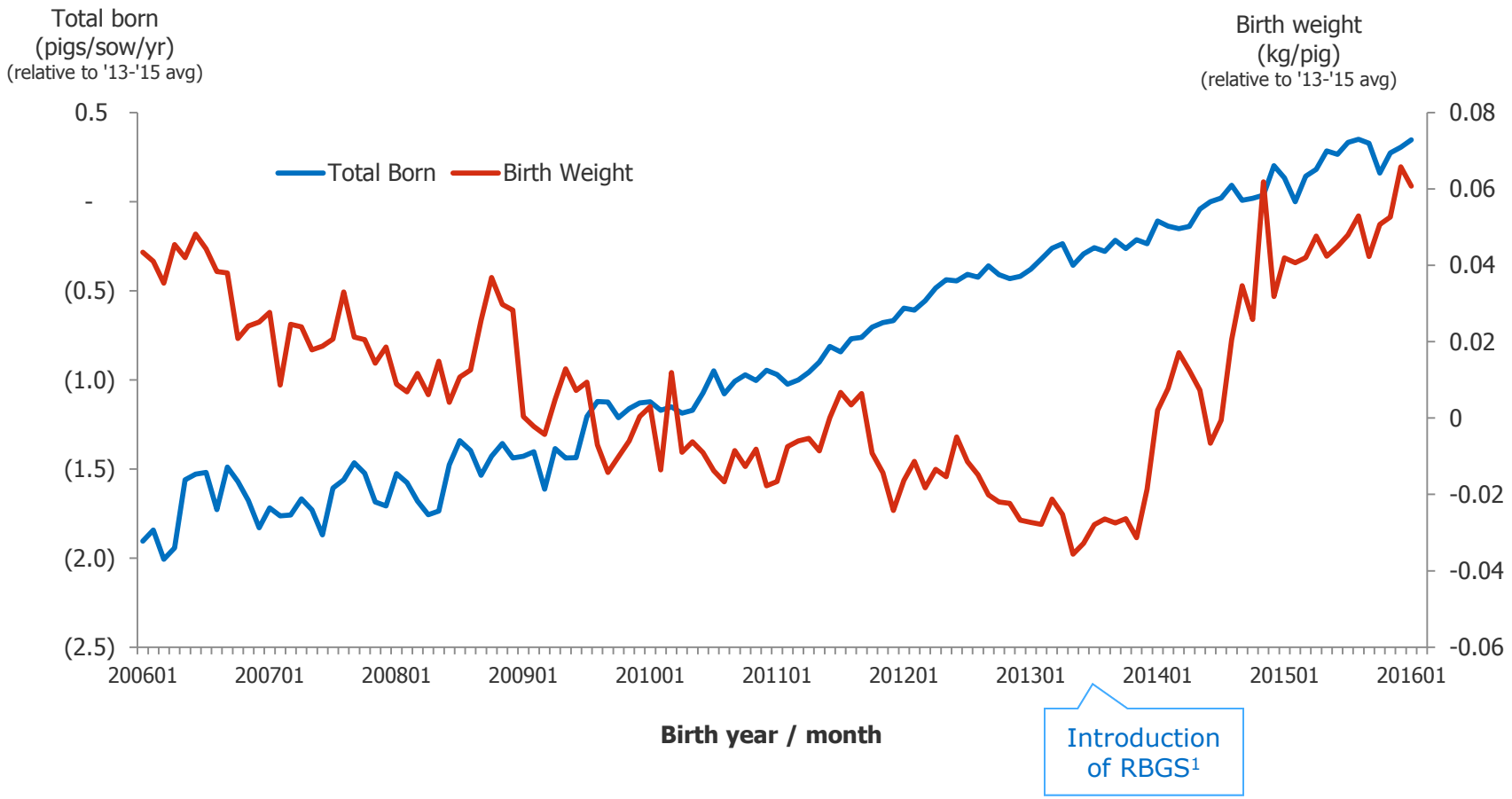
*Department of Animal and Dairy Science, University of Georgia, Athens 30602-2771;
†Instituto Nacional de Investigación Agropecuaria, Las Brujas 90200, Uruguay;
‡INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), BP 52627, 32326 Castanet-Tolosan, France; and §Department of Animal Science, Purdue University, West Lafayette, IN 47907-1151

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




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

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




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

Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

EDITORIAL

# FAQ for genomic selection

Genomic selection has been practiced in many species and in many organizations. In some cases, the results have been spectacular, and in some not. When the results fall short of expectations, questions remain as to whether they were because of inadequate statistics, too small chip size, problems with quality control or basic issues. In the end, one wonders what the limits of genomic selection are, and what will follow it. Based on published and unpublished results on genomic selection, one can prepare a FAQ sheet. Here it is. While looking at it, remember that FAQs change over time.

**I have heard that with 1000 animals genotyped and phenotyped I will have accurate predictions for many generations. Is this true?** *Not really. One needs more genotypes and the genomic association number of recent ancestors in the reference population. If that number is high and the populations are strongly linked, the accuracy may be decent. If that number is low, the accuracy will be close to 0. In the extreme, the genomic prediction for a different population, while ignoring the parent average, may be less accurate than the regular EBV.*

**Are prediction equations developed with one breed useful for other breeds?** *They are not. They would be if SNP effects were gene effects that are similar across breeds. However, SNP effects point mostly to common haplotypes of recent ancestors, or in other words, we are getting 'better' additive relationships.*

**If this is the case, what fraction of the additive variability is explained by genes or closely**

# ssGBLUP for Genome Wide Association Studies

- Large research interest in GWAS
- Limitations if Bayesian methods
  - Simple models
  - Single trait
  - Complicated if not all animals genotyped

Can ssGBLUP be used for GWAS?

*Genet. Res., Camb.* (2012), **94**, *pp.* 73–83. 
doi:10.1017/S0016672312000274

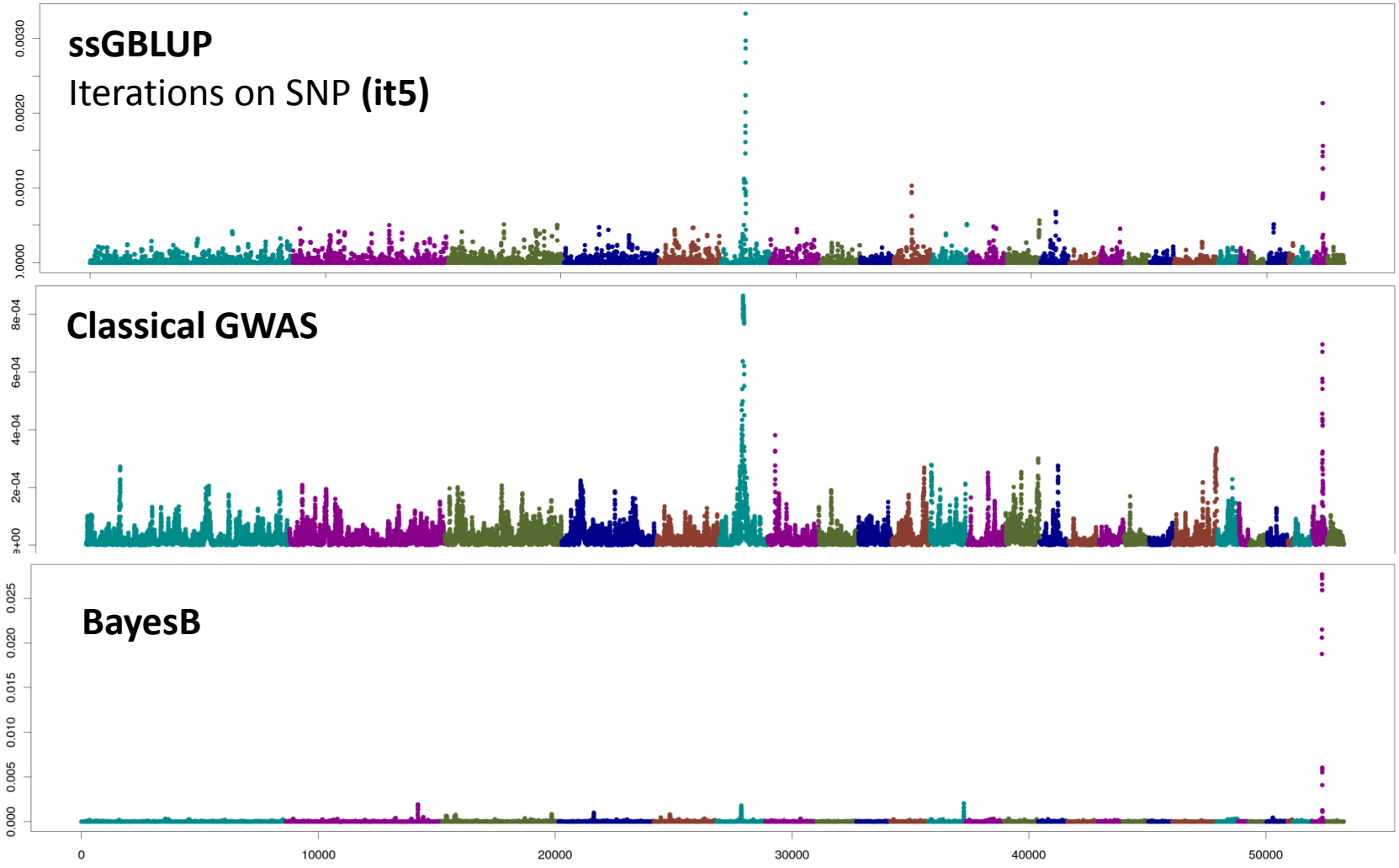


## Genome-wide association mapping including phenotypes from relatives without genotypes

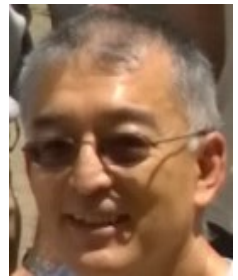
H. WANG[1]*, I. MISZTAL[1], I. AGUILAR[2], A. LEGARRA[3] AND W. M. MUIR[4]
[1] *Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602-2771, USA*
[2] *Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, 90200 Canelones, Uruguay*
[3] *INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), BP 52627, 32326 Castanet-Tolosan, France*
[4] *Department 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*)

# Three Methods for GWAS – chicken



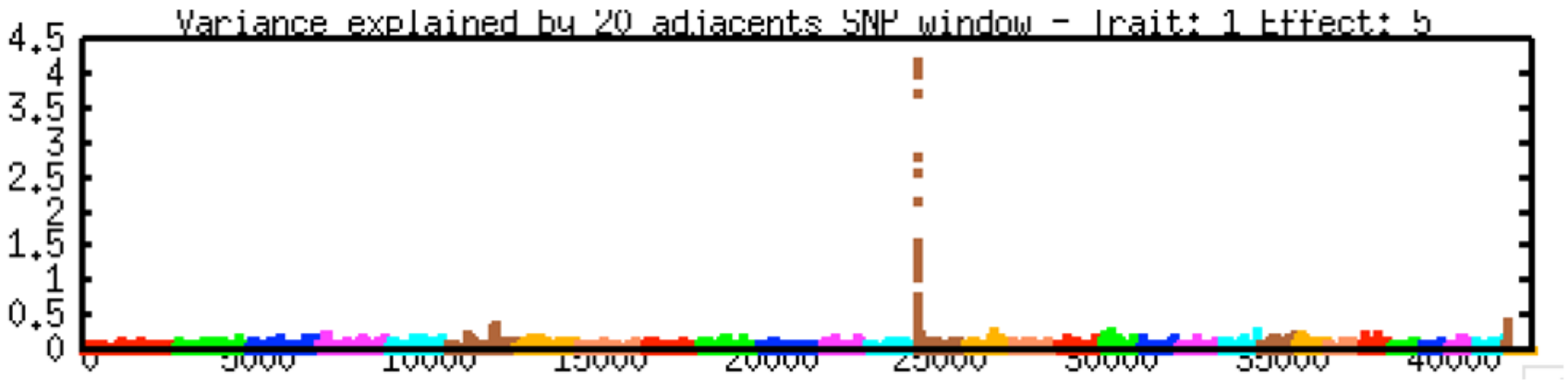
# Can large QTL exist despite  selection?



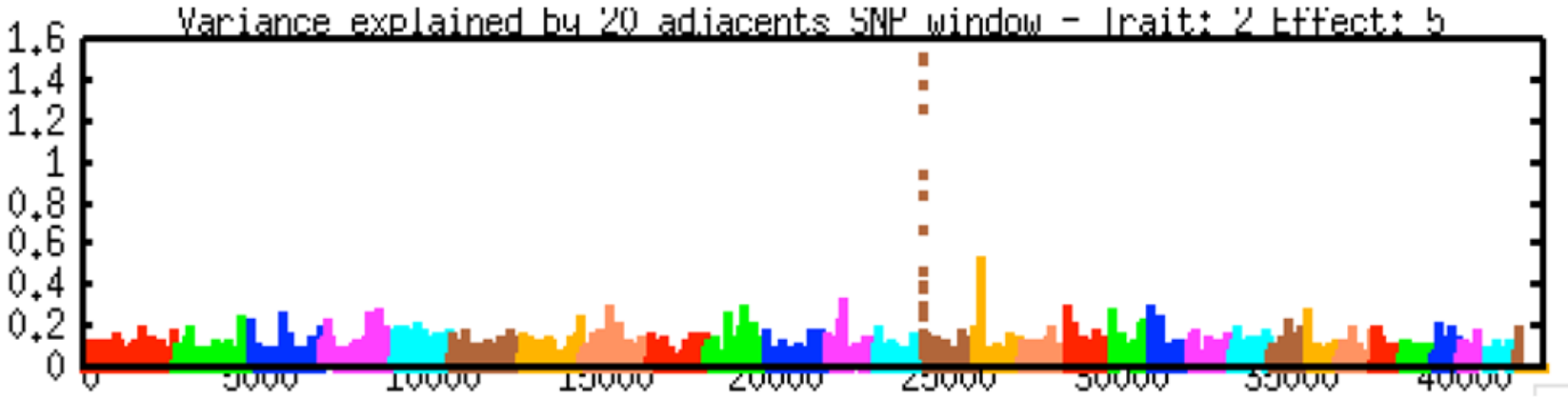
- Genetics and genomics of mortality in US Holsteins
- (Tokuhisa et al, 2014; Tsuruta et al., 2014)

- 6M records, SNP50k genotypes of 35k bulls

## Milk – first parity

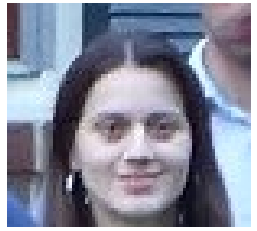


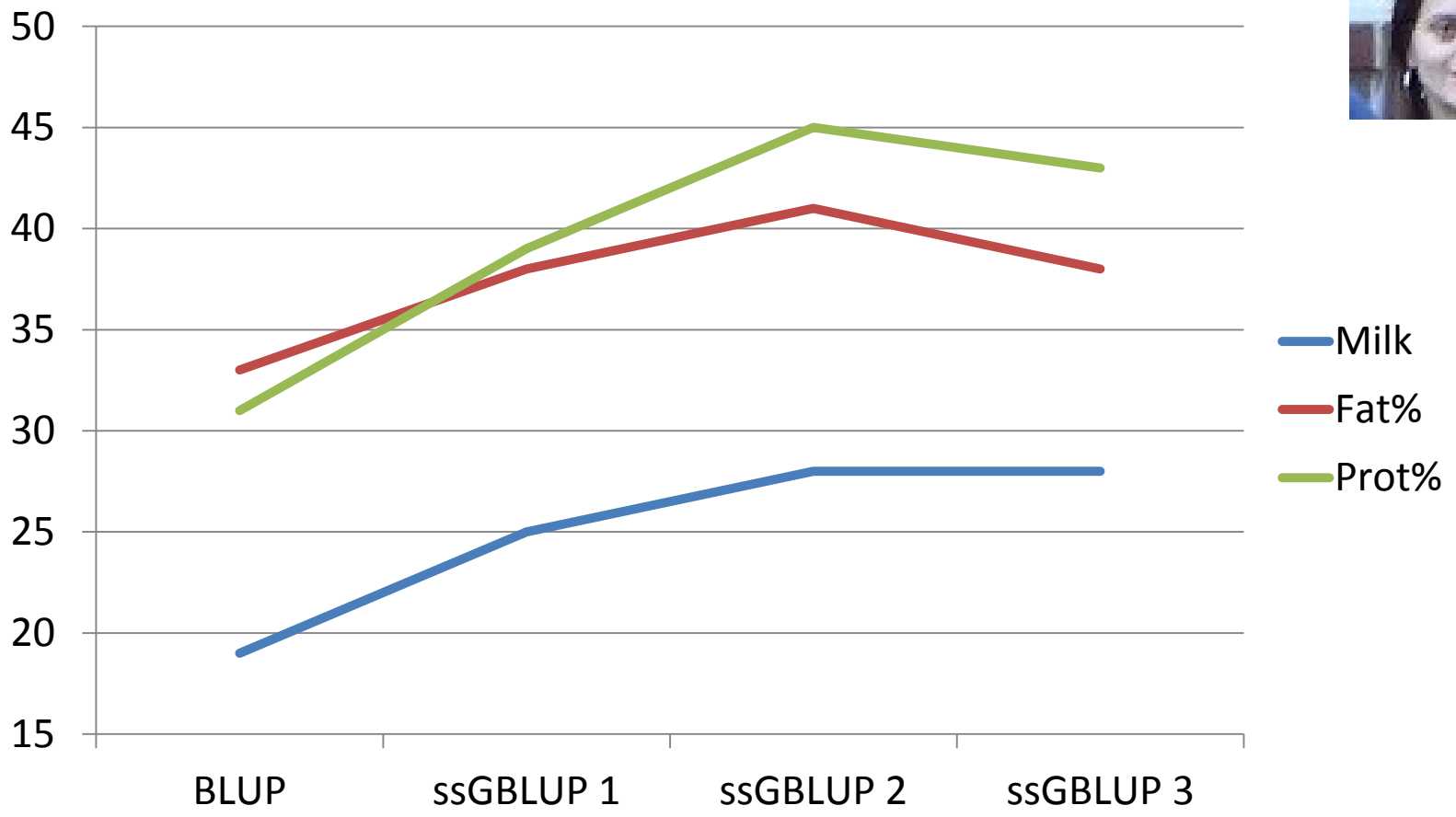

## Mortality – first parity



# $R^2$ in Israeli dairy – 1400 genotypes

(Lino et al., 2012)





# Why unknown parent groups

- Different lines or breeds (Harris and Johnson, 2012)
- Unrecorded parents across generations

Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

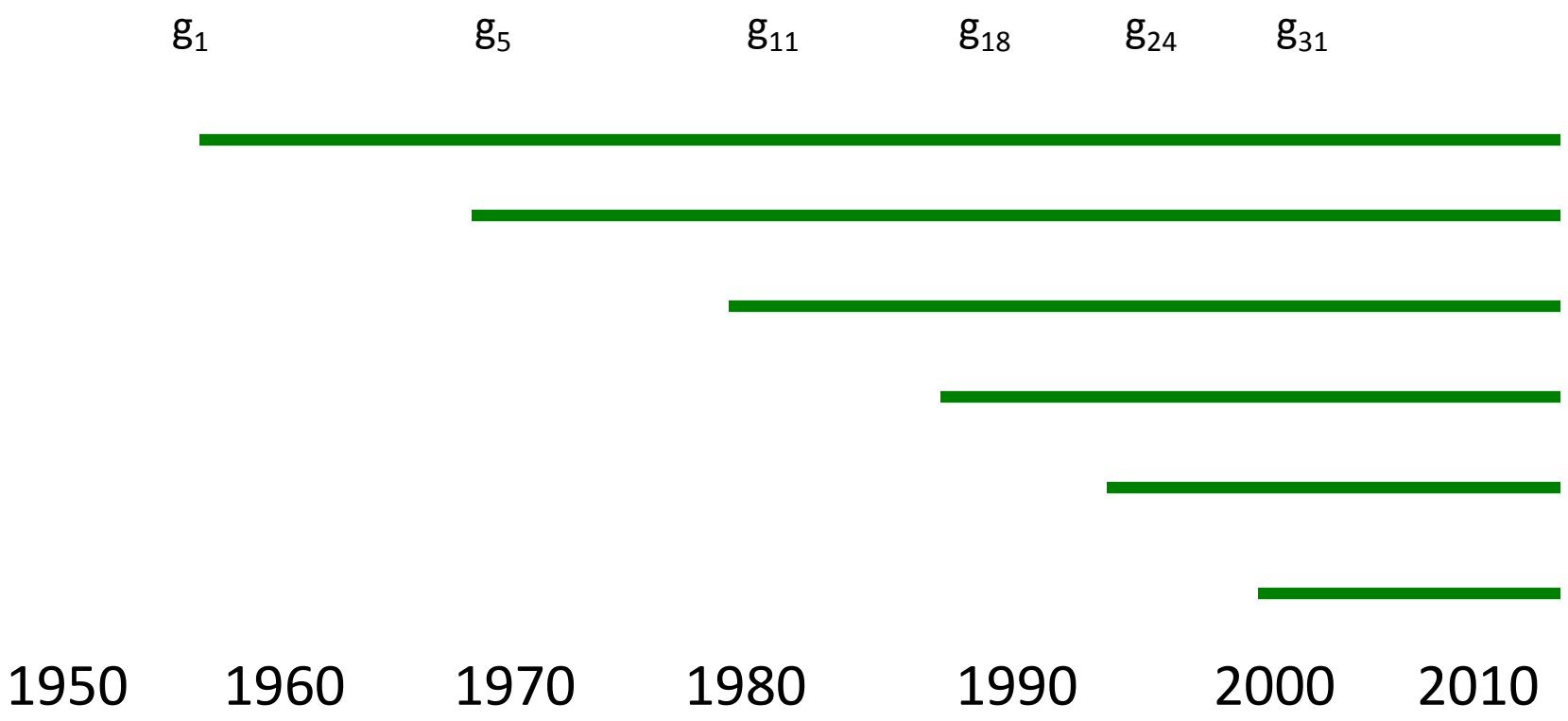
ORIGINAL ARTICLE

**Unknown-parent groups in single-step genomic evaluation**

I. Misztal[1], Z.G. Vitezica[2], A. Legarra[3], I. Aguilar[4] & A.A. Swan[5]

1 Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA
2 Université de Toulouse, UMR 1289 TANDEM, INRA/INPT-ENSAT/ENVT F-31326, Castanet-Tolosan, France
3 INRA, Station d'Amélioration Génétique des Animaux, Castanet-Tolosan, France
4 Instituto Nacional de Investigación Agropecuaria,, Las Brujas, Uruguay
5 Animal Breeding and Genetics Unit, University of New England, Armidale, NSW, Australia

# Pedigree depth for young animals



# Pedigree length and convergence

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

Big $A_{22}$ makes H less PD,
Reduces convergence rate

$$\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$$

Good convergence and genotyped animals biased down

$$\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$$

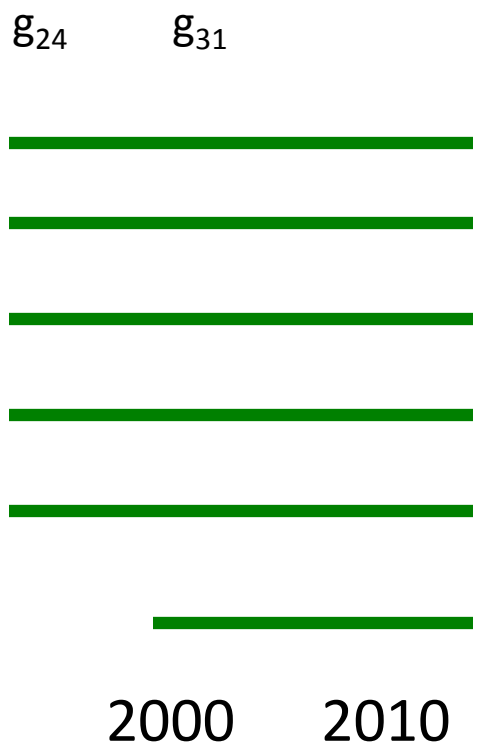
Bad convergence and genotyped animals biased up

$$\mathbf{G}^{-1} - \mathbf{A}_{22,1}^{-1} - \mathbf{A}_{22,2}^{-1} - \mathbf{A}_{22,3}^{-1}$$

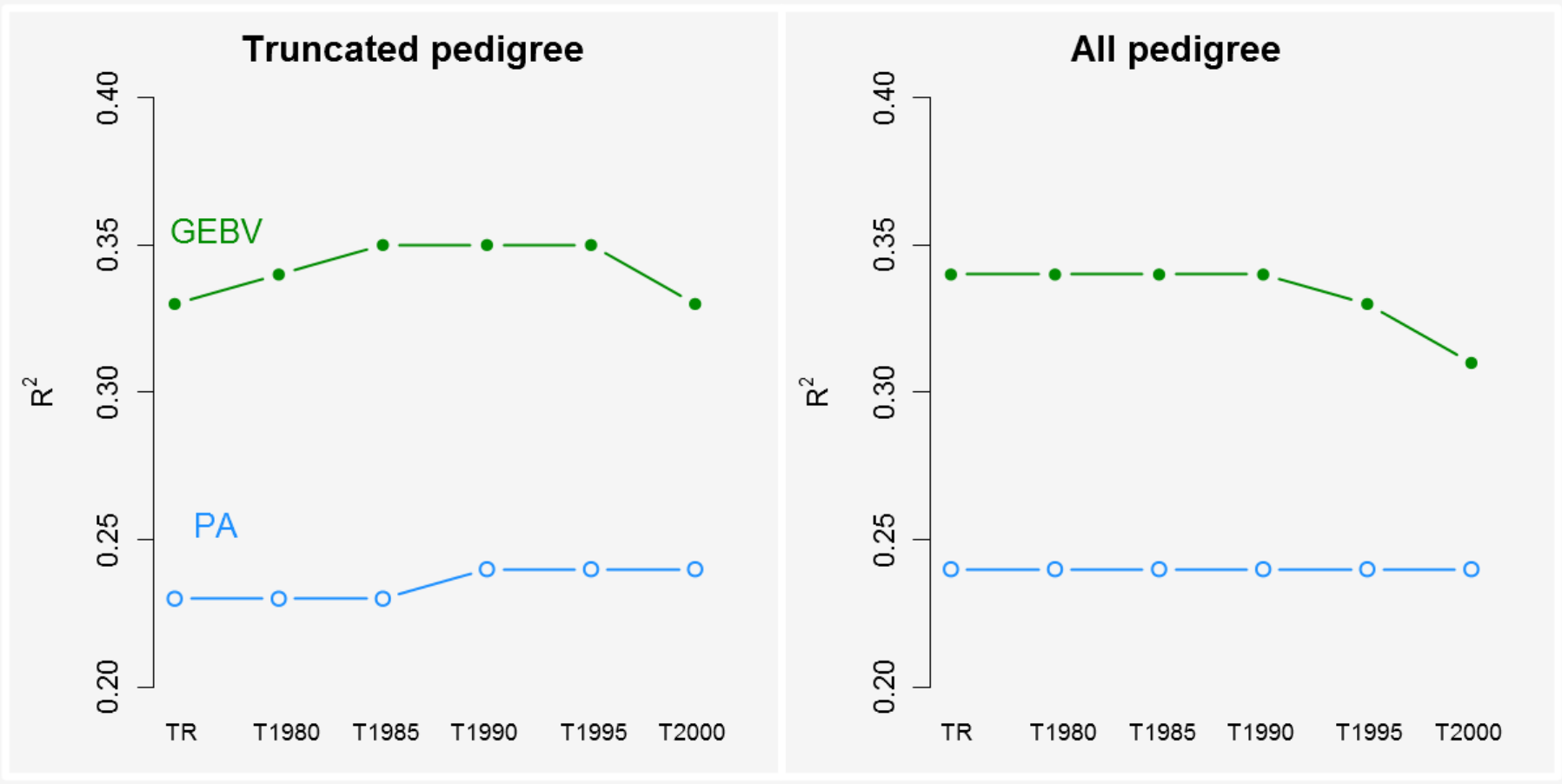
Long medium short pedigrees

Bad convergence and genotyped animals biased down and up

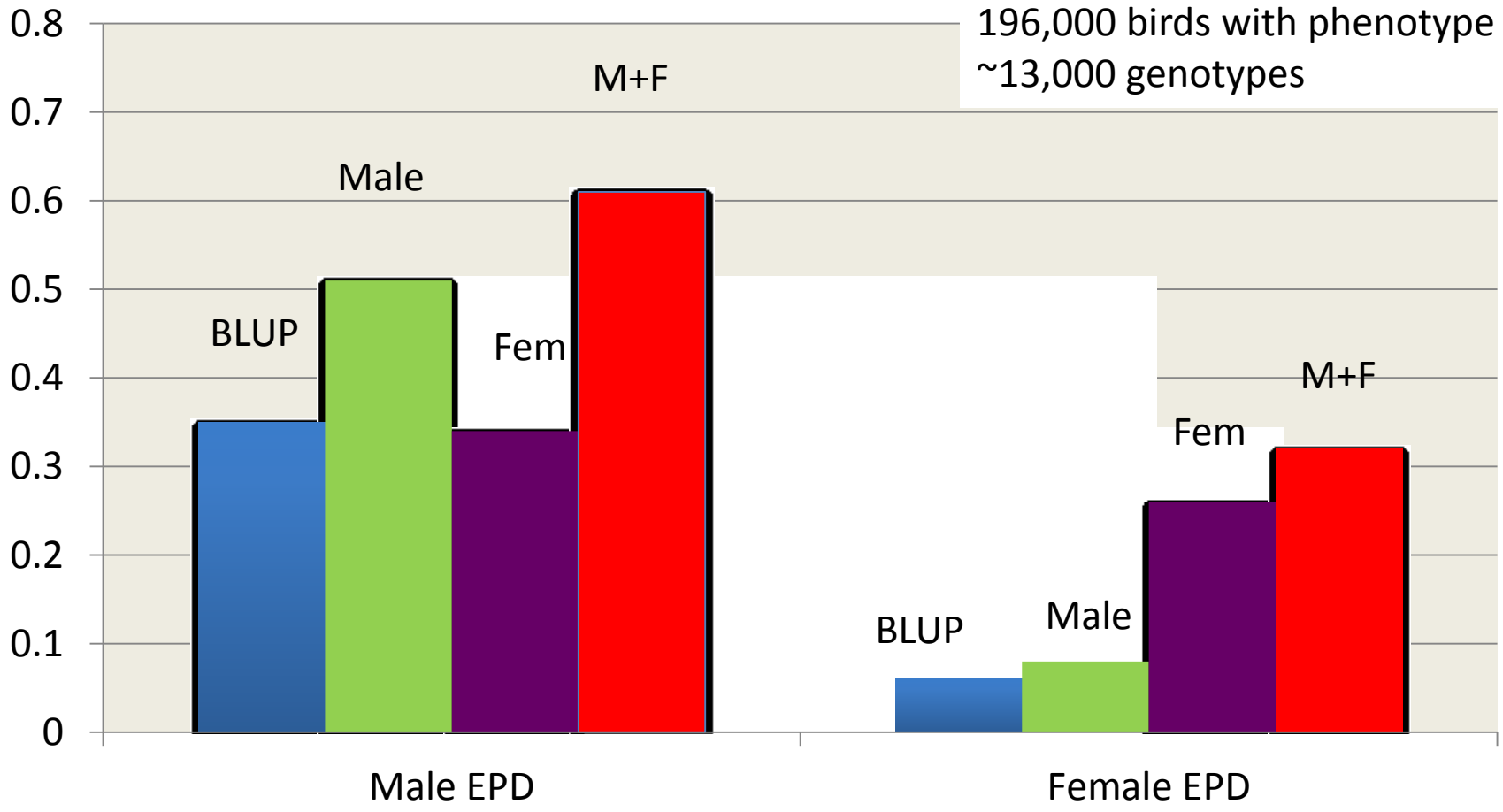
# Cut pedigree and data?



# US Holsteins – final scores

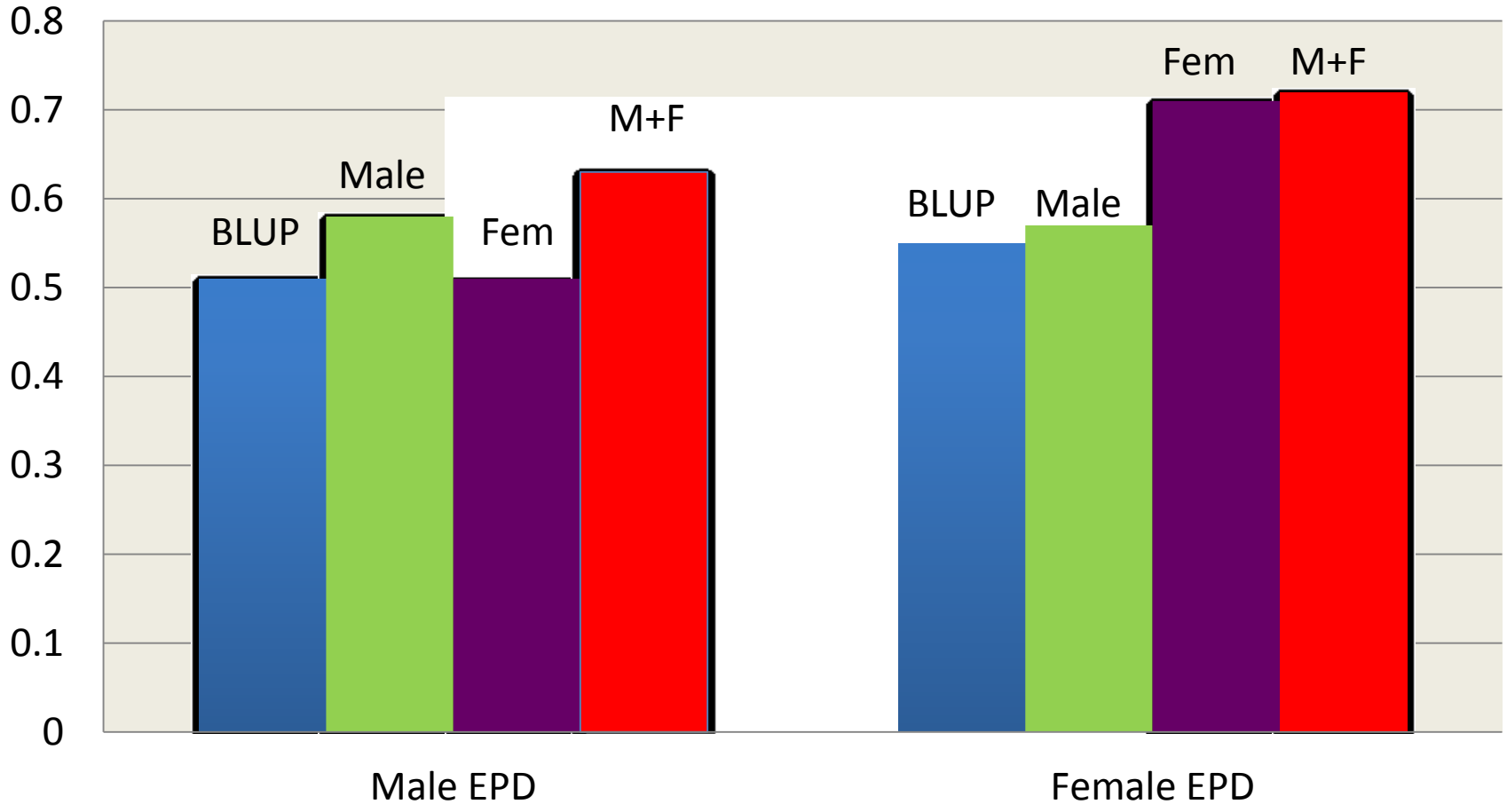


# Realized broiler accuracies with male, female or both genotypes – Trait A



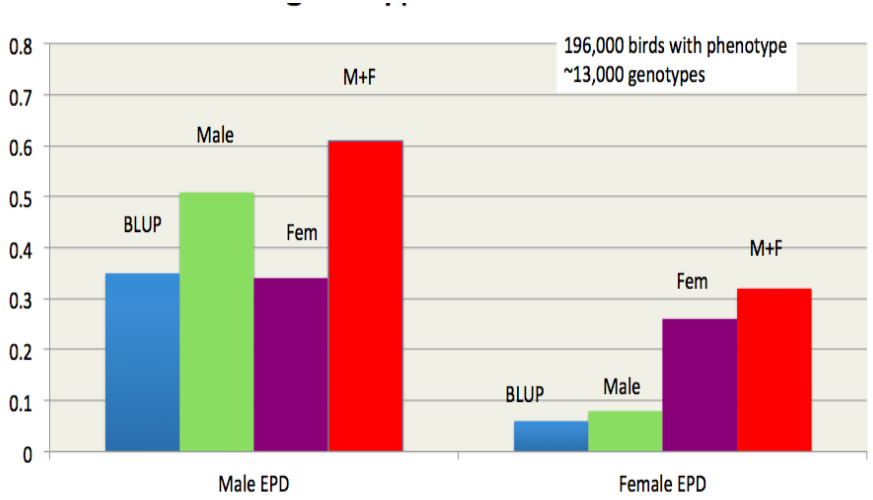

(Lourenco et al., submitted)

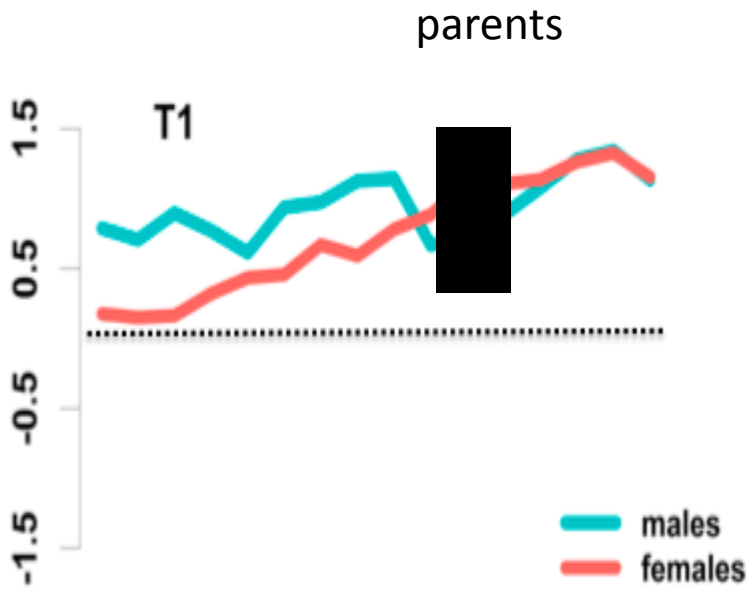
# Realized accuracies with male, female or both genotypes – Trait D




(Lourenco et al., in prep)

# Why realized accuracies differ by sex?



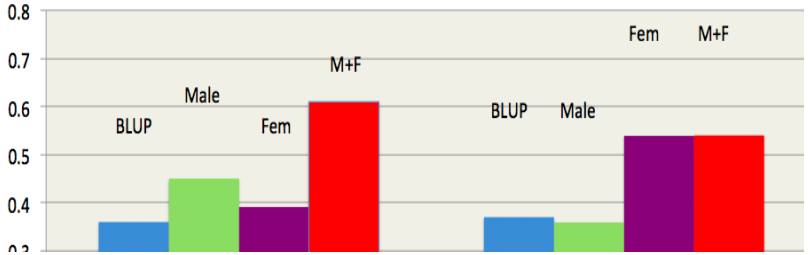




Bigger selection pressure on females

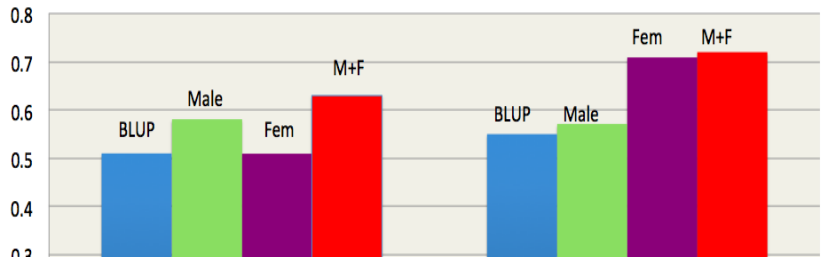
Selection graph for GEBV; possibly more differential selection EBV from BLUP

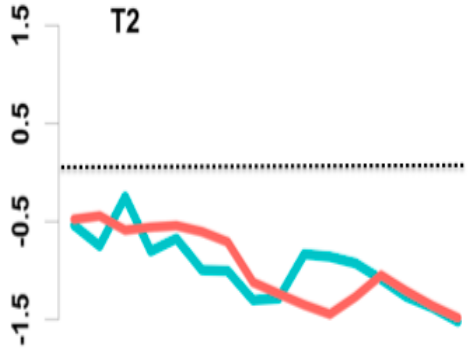
# Why realized accuracies differ by traits for similar $h^2$

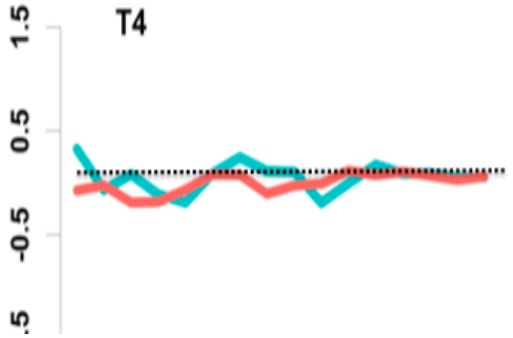
$h^2$=0.25




$h^2$=0.22



# Decomposition of GEBV in Single-step

$$\left\{ Z'MZ + \alpha A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \right\} \hat{u} = Z'My$$

$$GEBV = w_1 CD + w_2 PA + w_3 PC + w_4 DGV + w_5 PI$$

$$\square\ w_1 = 1$$

CD – contemporary deviation
PA – Parent average
PC – Progeny Contribution

DGV – direct genomic value
PI – Parental Index

No genotype, no extra accuracy

# GEBV for young animals

Complete $GEBV = w_2 PA + w_4 DGV + w_5 PI$

If genotype via SNP only $GEBV = DGV$

If no genotype $GEBV = PA$

Little improvement with genomics if animal not genotyped

J. Anim. Breed. Genet. ISSN 0931-2668

EDITORIAL

# Is genomic selection now a mature technology?

A couple of years ago, I wrote 'FAQ for genomic selection' in JABG (Volume 8, 245–246), and statements there are still intact IMHO. With many new studies, many murky points became clear and new puzzles appeared.

The genetic evaluation by BLUP became mature technology after the discovery of inexpensive inverse of the numerator relationship matrix (Henderson) and computing methodologies by iteration on data (Schaeffer). Then, the largest evaluations could be conducted by BLUP. Refinements continued, but the main steps were done. One can wonder whether now the genomic selection is also a mature technology.

validation continues. Properties of a particular validation are clearer by looking at the decomposition of GEBV into five components: parent average, yield deviation, progeny contribution, direct genomic value and pedigree index [e.g. Lourenco *et al.*, (2015) *Genet. Sel. Evol.*, **47**:56]. One big plus of a genomic validation is that it usually includes BLUP validation, often exposing problems in BLUP models such as excessive complexity. Good BLUP models are important as bad EBVs usually mean bad GEBVs. Realized accuracies may be very low due to strong selection [Bijma (2012) *J Anim Breed Genet.*, **129**:345–358].

Single-step GBLUP (ssGBLUP) became a universally

# New studies

- Unbiased evaluations of US Holstein with > 2 M genotypes of varying quality
- Helping Interbull survive
  - Unbiased pseudo-observations for bulls
  - GBLUP MACE
- Crossbreeding evaluation without reduction of accuracy
- Resilience and genomic selection

# Programming/methodology

- Better approximations of accuracy
- Better GWAS
- GUI?

# Applied studies

- Pigs
  - Mortality, survival, changing correlations
- Chickens
  - Sexual dimorphism,…
- Dairy
- Beef
  - Altitude, GxE
- Fish

- Heat stress
- GxE
- Resilience
- Theory

# Trends

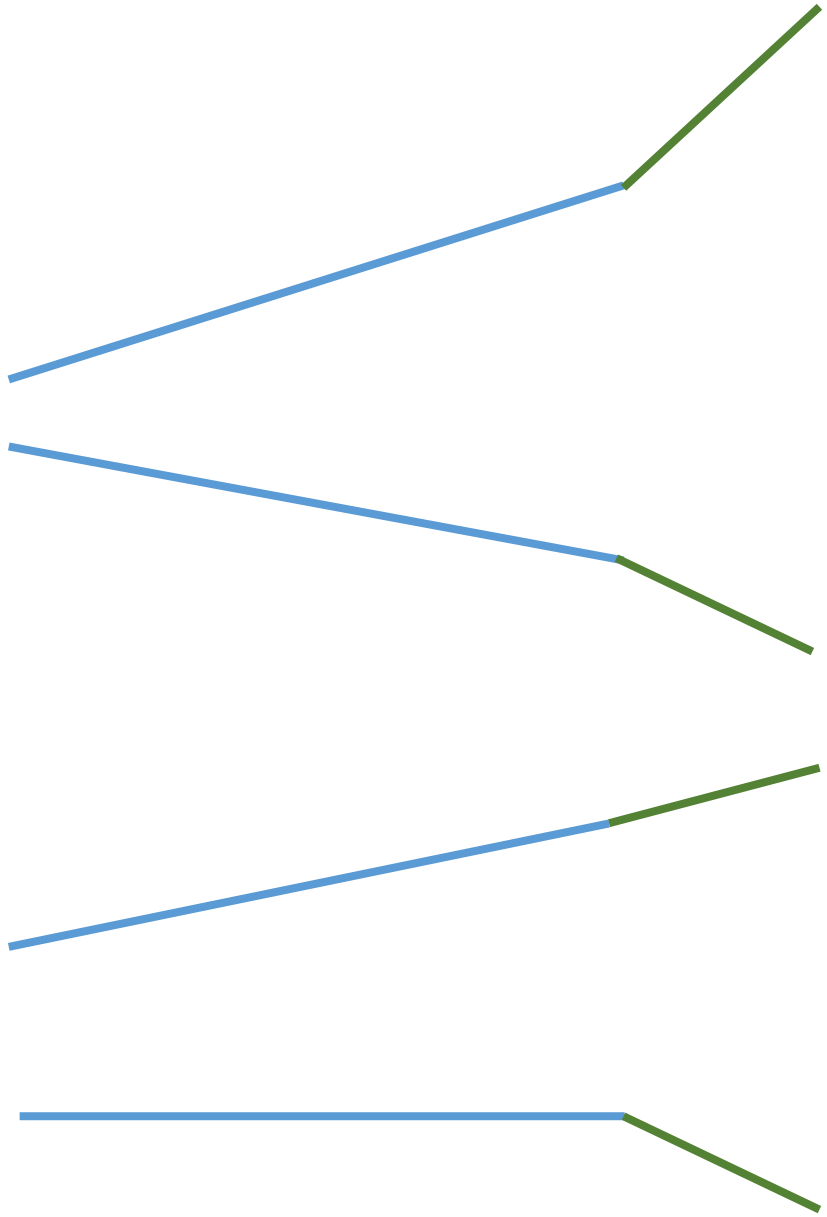
Genomic selection

Production (high $h^2$)

Raw fitness (low $h^2$)

Management

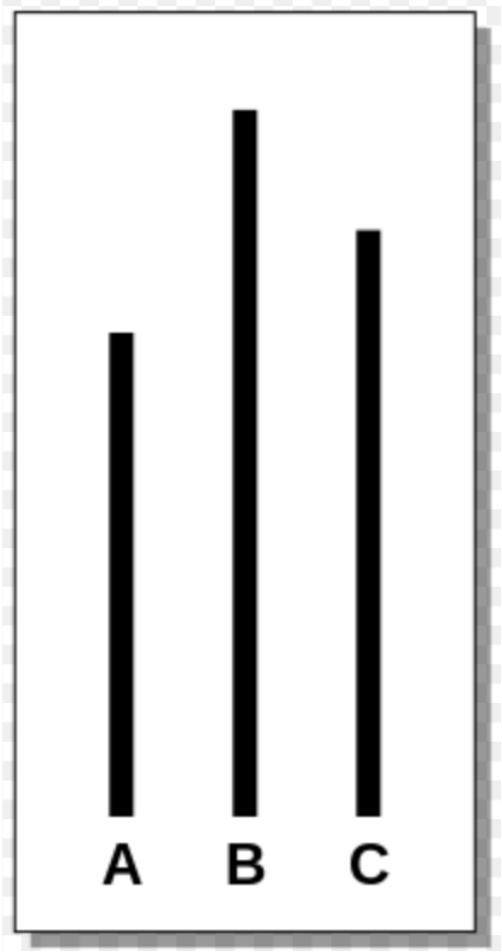
Realized fitness

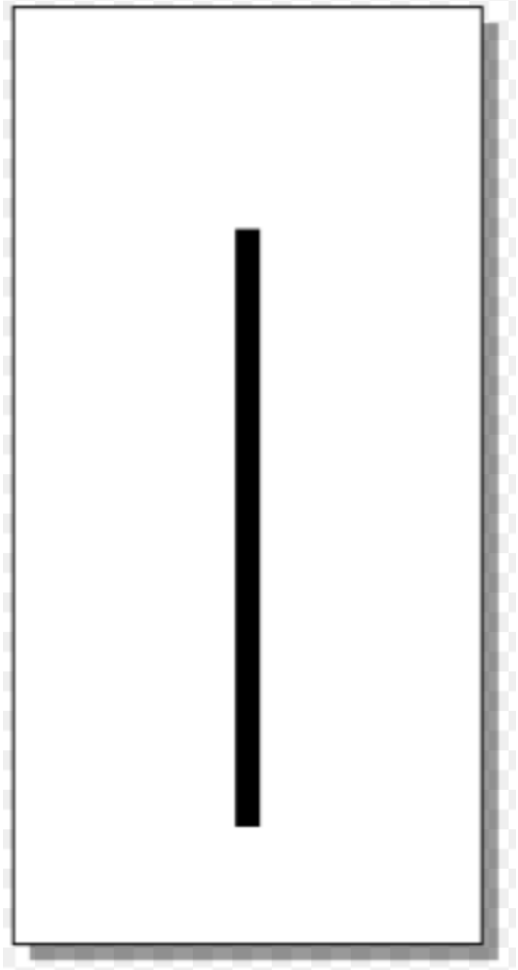


# Is UGA a good place to come?

- Good place for Science
- Improving South
- Politics small at universities
- Funding available
- Interesting projects
- Data from biggest animal institutions across species

# Asch (1951) experiments



A, B or C?