

Large-Scale Single- Step Genome Wide Association Studies with the Algorithm for Proven and Young

M. Bermann, D. Lourenco, I. Misztal

01/2023

2023 Plant and Animal Genome Meeting



UNIVERSITY OF
GEORGIA

College of Agricultural &
Environmental Sciences

*Animal Breeding and
Genetics Group*

Introduction

- Breeding value models (GBLUP) and marker-effects models (SNP-BLUP) are equivalent:
 - Obtain marker-effects from EBV and vice versa
 - Exploit computational advantages of both models
 - Number of SNP vs. Number of genotyped animals
 - Perform GWAS using breeding values
- Many genotyped individuals → computational challenges → inverse of GRM
- Ways to overcome challenges: efficient algorithms like APY
- Marker-effects models are not equivalent to GBLUP with APY
- **Objective:**
 - **Derive marker-effect models equivalent to APY**
 - **Show applications of these models**



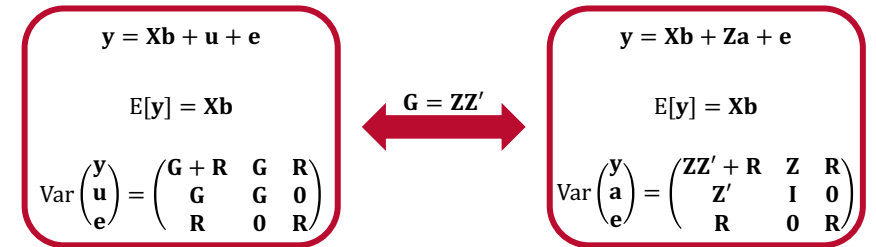
Outline

1. Consequences of the equivalence between BV and marker-effects models
2. Computational challenges in ssGBLUP
3. The Algorithm for Proven and Young (APY)
4. Equivalence between BV and marker-effects models with APY
5. Theoretical and practical consequences
6. Application example
7. Concluding remarks



BV and marker-effects models

- Under some assumptions GBLUP and SNP-BLUP are equivalent
- Consequences:
 - Estimable functions are equal
 - SNP-effects can be back-solved from EBV
 - Indirect predictions for young-animals
 - Perform GWAS from GBLUP
 - Choice of the most convenient model



Copyright © 2007 by the Genetics Society of America
DOI: 10.1534/genetics.107.081190

The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values

D. Habier,¹ R. L. Fernando and J. C. M. Dekkers

J. Dairy Sci. 92:2971–2975
doi:10.3168/jds.2008-1929
© American Dairy Science Association, 2009.

Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit

I. Strandén*¹ and D. J. Garrick†‡

Single-step genomic evaluations

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \sigma_g^{-2}\mathbf{H}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

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

Pedigree

Genomics

Single-step BV and marker-effects models

- ssGBLUP ↔ ssBR ↔ ssSNP-BLUP
- Consequences:
 - Estimable functions are equal
 - SNP-effects can be back-solved from GEBV
 - Indirect predictions for young-animals
 - Perform GWAS from ssGBLUP
 - Choice of the most convenient model?

J. Dairy Sci. 92:4656–4663
doi:10.3168/jds.2009-2061
© American Dairy Science Association, 2009.

A relationship matrix including full pedigree and genomic information

A. Legarra,^{*†} I. Aguilar,^{†‡} and I. Misztal[†]
^{*}INRA, UR631 SAGA, BP 52627, 32326 Castanet-Tolosan, France
[†]Department of Animal and Dairy Science, University of Georgia, Athens 30602
[‡]Instituto Nacional de Investigación Agropecuaria, Las Brujas, Uruguay

Fernando et al. *Genetics Selection Evolution* 2014, **46**:50
<http://www.gsejournal.org/content/46/50>



RESEARCH

Open Access

A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses

Rohan L Fernando^{1*}, Jack CM Dekkers¹ and Dorian J Garrick^{1,2}



J. Dairy Sci. 97:5833–5850
<http://dx.doi.org/10.3168/jds.2014-7924>
© American Dairy Science Association[®], 2014. Open access under [CC BY-NC-ND license](#).

A single-step genomic model with direct estimation of marker effects

Z. Liu,^{*†} M. E. Goddard,[†] F. Reinhardt,^{*} and R. Reents^{*}
^{*}Vereinigte Informationssysteme Tierhaltung w.V. (VIT), Heideweg 1, D-27283 Verden, Germany
[†]Melbourne School of Land and Environment, University of Melbourne, Parkville, Victoria 3010, Australia



Computational challenges in single-step

- Depending on the model
 - Inversion of the GRM
 - Multiplication of large and dense matrices
 - Convergence and speed of MCMC
 - Convergence of solvers



J. Dairy Sci. 97:3943–3952
<http://dx.doi.org/10.3168/jds.2013-7752>
© American Dairy Science Association[®], 2014. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/)

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

Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals¹

E. A. Mäntysaari,^{*2} R. D. Evans,[†] and I. Strandén^{*}

RESEARCH ARTICLE

Open Access



Computational strategies for the preconditioned conjugate gradient method applied to ssSNPBLUP, with an application to a multivariate maternal model

Jeremie Vandenplas^{1*}, Herwin Eding², Maarten Bosmans³ and Mario P. L. Calus¹

RESEARCH ARTICLE

Open Access



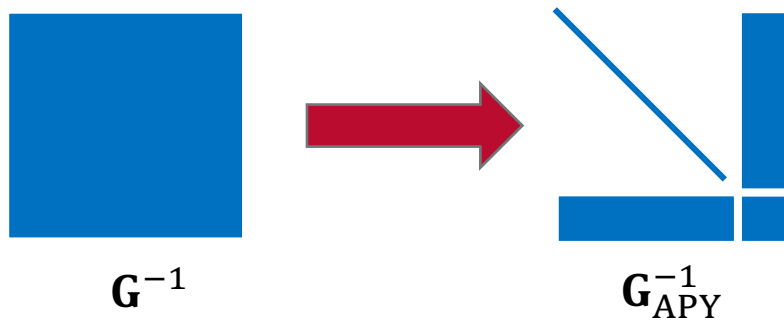
Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals

Rohan L. Fernando^{1*}, Hao Cheng¹, Bruce L. Golden² and Dorian J. Garrick^{1,3}



The Algorithm for Proven and Young (APY)

- Genomic information is redundant
- Most of the variation in GRM is explained by a set of core (proven) individuals
- BV of noncore individuals are a linear combination of BV of core individuals plus an error term
- The noncore (young) individuals are conditionally dependent on the core



$$A^{-1} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G^{-1} - A_{22}^{-1} \end{pmatrix} \longrightarrow A^{-1} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G_{APY}^{-1} - A_{22}^{-1} \end{pmatrix}$$

Computational advantages of APY

- Fast approximation of G^{-1}
- Memory savings
- Less time per iteration
- Efficient addition of RPG
- Closed-form sparse inverse
- Fast approximation of accuracies

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

ORCID ID: 0000-0002-0382-1897 (I.M.)

ORIGINAL ARTICLE

WILEY  *Journal of Animal Breeding and Genetics*

Solving efficiently large single-step genomic best linear unbiased prediction models

I. Strandén¹  | K. Matilainen¹ | G.P. Aamand² | E.A. Mäntysaari¹

Impact of blending the genomic relationship matrix with different levels of pedigree relationships or the identity matrix on genetic evaluations

Mary Kate Hollifield,*  Matias Bermann,  Daniela Lourenco,  and Ignacy Misztal 

Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the Algorithm for Proven and Young

Matias Bermann,¹  Daniela Lourenco, and Ignacy Misztal 

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

¹Corresponding author: mbermann@uga.edu



BV and marker-effects models with APY

- Back-solve marker effects from EBV calculated with APY
 - Calculate indirect predictions and accuracies
 - GWAS
- APY ssGBLUP is not equivalent to a marker-effects model
 - Statistical properties are lost
 - Numerical results might be sound

Indirect predictions with a large number of genotyped animals using the algorithm for proven and young

Andre L.S. Garcia,^{†,1} Yutaka Masuda,[†] Shogo Tsuruta,[†] Stephen Miller,[†] Ignacy Misztal,[†] and Daniela Lourenco[†]

[†]Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, [‡]Angus Genetics Inc. St. Joseph, MO 64506

RESEARCH ARTICLE

Open Access

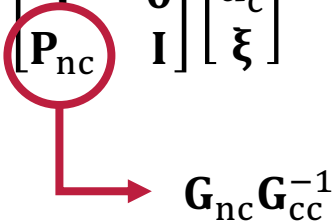
Theoretical accuracy for indirect predictions based on SNP effects from single-step GBLUP

Andre Garcia^{1*}, Ignacio Aguilar², Andres Legarra³, Shogo Tsuruta¹, Ignacy Misztal¹ and Daniela Lourenco¹



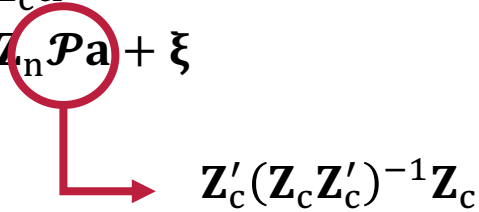
APY marker-effects models

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





$$\begin{aligned} \mathbf{u}_c &= \mathbf{Z}_c \mathbf{a} \\ \mathbf{u}_n &= \mathbf{Z}_n \mathcal{P} \mathbf{a} + \xi \end{aligned}$$



$$\text{Var} \begin{bmatrix} \mathbf{u}_c \\ \mathbf{u}_n \end{bmatrix} = \mathbf{G}_{\text{APY}} = \begin{bmatrix} \mathbf{G}_{cc} & \mathbf{G}_{cn} \\ \mathbf{G}_{nc} & \mathbf{M}_{nn} + \mathbf{G}_{nc} \mathbf{G}_{cc}^{-1} \mathbf{G}_{cn} \end{bmatrix}$$



$$\text{Var} \begin{bmatrix} \mathbf{u}_c \\ \mathbf{u}_n \end{bmatrix}^{-1} = \mathbf{G}_{\text{APY}}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} + \mathbf{P}_{cn} \mathbf{M}_{nn}^{-1} \mathbf{P}_{nc} & -\mathbf{P}_{cn} \mathbf{M}_{nn}^{-1} \\ -\mathbf{M}_{nn}^{-1} \mathbf{P}_{nc} & \mathbf{M}_{nn}^{-1} \end{bmatrix}$$



APY marker-effects models

$$\begin{bmatrix} \mathbf{y}_c \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_c \\ \mathbf{X}_n \end{bmatrix} \mathbf{b} + \begin{bmatrix} \mathbf{W}_c & 0 \\ 0 & \mathbf{W}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_c \\ \mathbf{u}_n \end{bmatrix} + \mathbf{e} \quad \longrightarrow \quad \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W}_c & \mathbf{X}'\mathbf{W}_n \\ \mathbf{W}'_c\mathbf{X}_c & \mathbf{W}'_c\mathbf{W}_c + \mathbf{G}_{APY}^{cc}\alpha & \mathbf{G}_{APY}^{cn}\alpha \\ \mathbf{W}'_n\mathbf{X}_n & \mathbf{G}_{APY}^{nc}\alpha & \mathbf{W}'_n\mathbf{W}_n + \mathbf{G}_{APY}^{nn}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_c \\ \hat{\mathbf{u}}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'_c\mathbf{y}_c \\ \mathbf{W}'_n\mathbf{y}_n \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{y}_c \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_c \\ \mathbf{X}_n \end{bmatrix} \mathbf{b} + \begin{bmatrix} \mathbf{W}_c & 0 \\ 0 & \mathbf{W}_n \end{bmatrix} \begin{bmatrix} \mathbf{Z}_c \\ \mathbf{Z}_n\mathcal{P} \end{bmatrix} \mathbf{a} + \begin{bmatrix} 0 \\ \mathbf{W}_n \end{bmatrix} \xi + \mathbf{e}$$

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{W}\mathbf{Z}^\dagger\mathbf{a} + \mathbf{Q}\xi + \mathbf{e} \quad \longrightarrow \quad \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{Z}^\dagger & \mathbf{X}'\mathbf{W}_n \\ \mathbf{Z}^{\dagger'}\mathbf{W}'\mathbf{X} & \mathbf{Z}^{\dagger'}\mathbf{W}'\mathbf{W}\mathbf{Z}^\dagger + \mathbf{I}_\gamma & \mathcal{P}\mathbf{Z}'_n\mathbf{W}'_n\mathbf{W}_n \\ \mathbf{W}'_n\mathbf{X}_n & \mathbf{W}'_n\mathbf{W}_n\mathbf{Z}_n\mathcal{P} & \mathbf{W}'_n\mathbf{W}_n + \mathbf{M}_{nn}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\xi} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}^{\dagger'}\mathbf{W}'\mathbf{y} \\ \mathbf{W}'_n\mathbf{y}_n \end{bmatrix}$$

APY marker-effects model

- # core individuals \rightarrow # markers
 - $\xi \rightarrow \mathbf{0}$
 - $\mathcal{P} \rightarrow \mathbf{I}$
- APY-SNP-BLUP \rightarrow SNP-BLUP

$$y = Xb + WZ^\dagger a + Q\xi + e$$



$$y = Xb + WZa + e$$

$$\begin{bmatrix} X'X & X'WZ^\dagger & X'_n W_n \\ Z'^\dagger W'X & Z'^\dagger W'WZ^\dagger + I_\gamma & \mathcal{P}Z'_n W'_n W_n \\ W'_n X_n & W'_n W_n Z_n \mathcal{P} & W'_n W_n + M_{nn}^{-1} \alpha \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{\xi} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'^\dagger W'y \\ W'_n y_n \end{bmatrix}$$

$$\begin{bmatrix} X'X & X'W \\ W'X & Z'W'WZ + I_\gamma \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'W'y \end{bmatrix}$$

Distributional properties of marker effects

- Marker effects estimates depends only on core animals

$$\hat{\mathbf{a}}|\hat{\mathbf{u}} = E[\mathbf{a}|\mathbf{u} = \hat{\mathbf{u}}] = k\mathbf{Z}^+ \mathbf{G}_{APY}^{-1} \hat{\mathbf{u}} = k\mathbf{Z}'_c \mathbf{G}_{cc}^{-1} \hat{\mathbf{u}}_c$$

$$\hat{\mathbf{a}}|\hat{\mathbf{u}} = E[\mathbf{a}|\mathbf{u} = \hat{\mathbf{u}}] = k\mathbf{Z}'\mathbf{G}^{-1}\hat{\mathbf{u}}$$

- The variance of marker effect estimates to

$$\text{Var}(\hat{\mathbf{a}}|\hat{\mathbf{u}}) = k^2 \mathbf{Z}'_c \mathbf{G}_{cc}^{-1} (\mathbf{G}_{cc} - \text{PEV}_{\text{core}}) \mathbf{G}_{cc}^{-1} \mathbf{Z}_c$$

$$\text{Var}(\hat{\mathbf{a}}|\hat{\mathbf{u}}) = k\mathbf{Z}'\mathbf{G}^{-1}(\mathbf{G} - \text{PEV})\mathbf{G}^{-1}\mathbf{Z}k$$

- Efficient GWAS

$$p - \text{value} = 1 - \Phi\left(\frac{|\hat{a}|}{\text{Var}(\hat{a})}\right)$$

$$p - \text{value} = 1 - \Phi\left(\frac{|\hat{a}|}{\text{Var}(\hat{a})}\right)$$

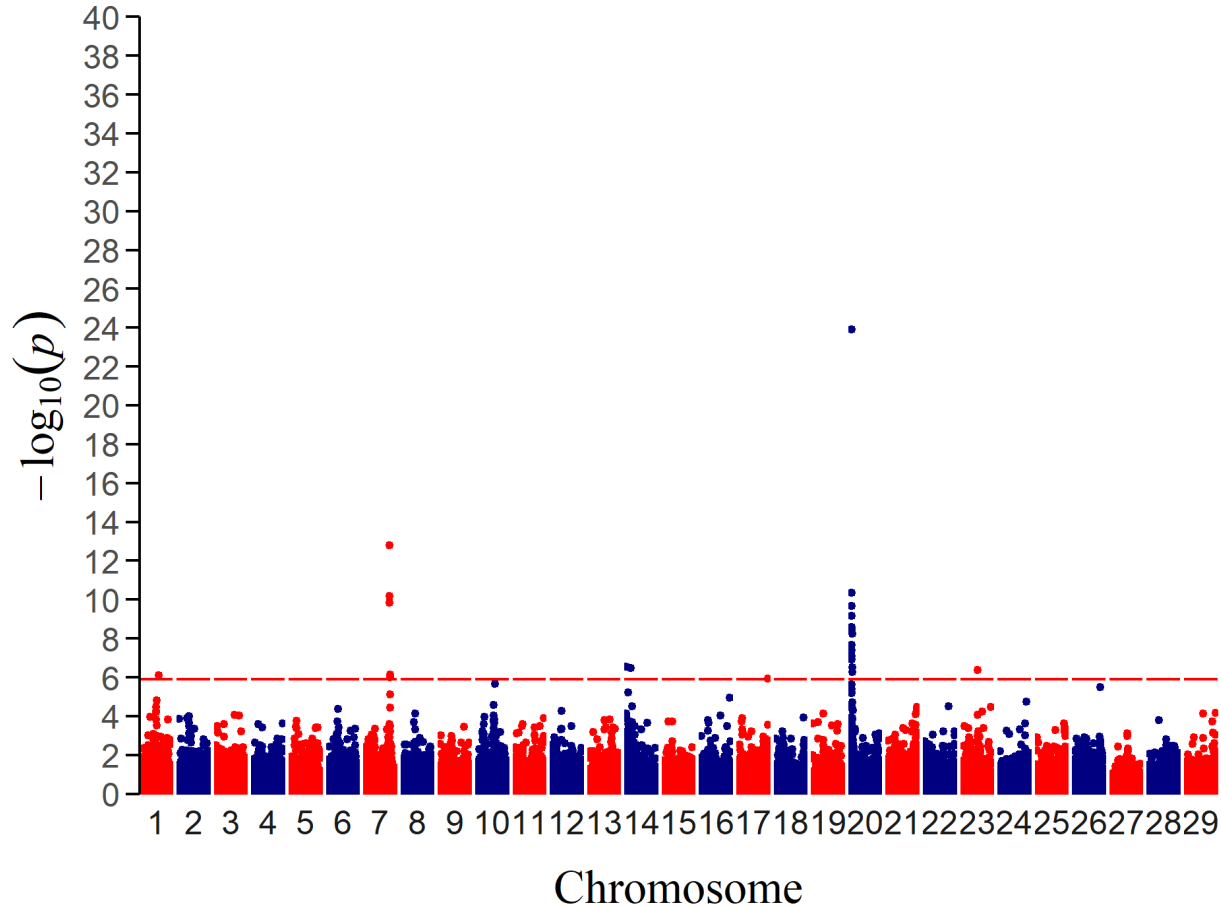
Application example

- Post-weaning gain in American Angus
- 845,000 phenotypes
- 450,000 genotypes
- 1,570,000 animals in the pedigree
- ssGBLUP (50k genotyped animals) vs. APY-ssGBLUP (450k genotyped animals)
- We expect:
 - Higher power
 - Less noise
 - Less false-positives

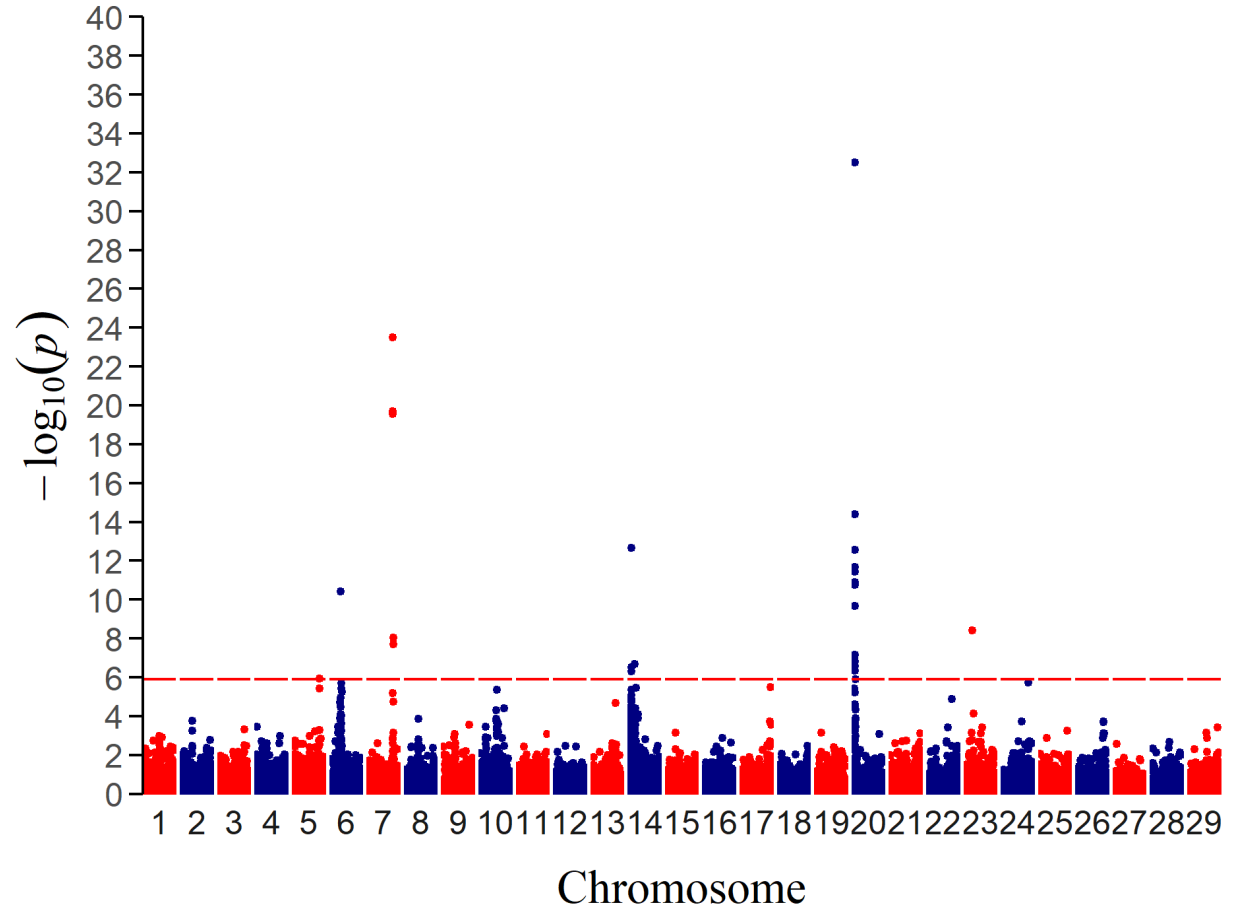


Leite et al.
(in progress)

50k genotyped animals



500k genotyped animals



Indirect predictions

- SNP effects can be used to predict GEBV for young individuals
- Indirect predictions depend only on core individuals

$$\hat{\mathbf{u}}_{ip} | \hat{\mathbf{a}} = \mathbf{Z}_{ip} \mathcal{P} \hat{\mathbf{a}} = \mathbf{G}_{ip,c} \mathbf{G}_{cc}^{-1} \hat{\mathbf{u}}_c$$

$$\hat{\mathbf{u}}_{ip} | \hat{\mathbf{a}} = \mathbf{Z}_{ip} \hat{\mathbf{a}} = \mathbf{G}_{ip,ref} \mathbf{G}^{-1} \hat{\mathbf{u}}$$

- Accuracies depend on PEV of core individuals

$$acc_i = \sqrt{1 - \frac{\mathbf{m}_{ii} + \mathbf{g}_{i,c} \mathbf{G}_{cc}^{-1} PEV_{core} \mathbf{G}_{cc}^{-1} \mathbf{g}_{c,i}}{\mathbf{g}_{ii}}}$$

$$acc_i = \sqrt{1 - \frac{\mathbf{g}_{ii} + \mathbf{g}_{i,ref} \mathbf{G}^{-1} PEV \mathbf{G}^{-1} \mathbf{g}_{ref,i}}{\mathbf{g}_{ii}}}$$

APY marker-effects models

- Equivalence holds when:
 - RPG is included
 - Single-step
- Computational advantages
 - Dimension = number of core individuals
 - Exact and efficient PEV for core individuals
 - PEV obtained while approximating accuracies
- Quality of approximation depends on EBV

RESEARCH ARTICLE

Open Access

On the equivalence between marker effect models and breeding value models and direct genomic values with the Algorithm for Proven and Young



Matias Bermann^{1*}, Daniela Lourenco¹, Natalia S. Forneris^{2,3}, Andres Legarra⁴ and Ignacy Misztal¹

Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the Algorithm for Proven and Young

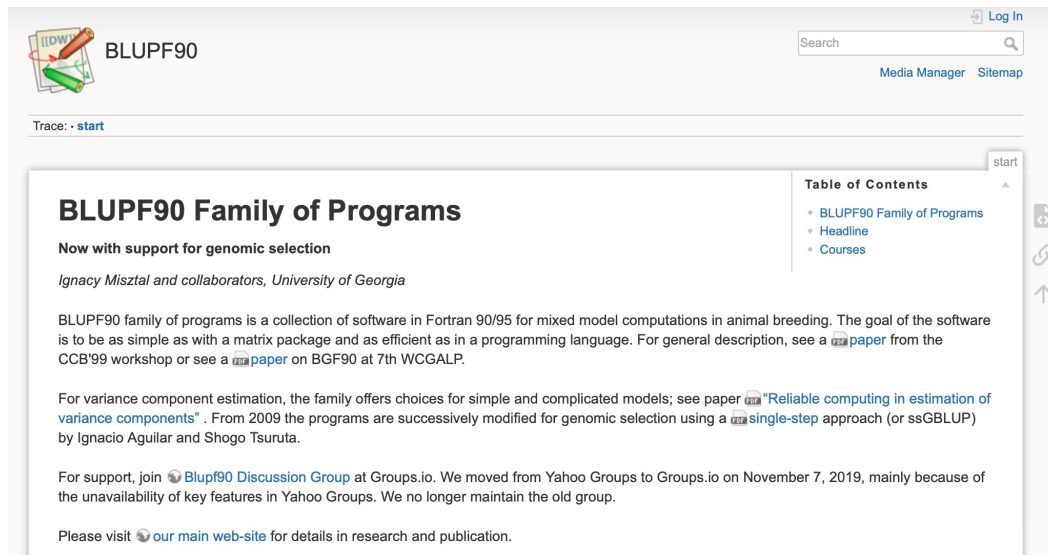
Matias Bermann¹, Daniela Lourenco, and Ignacy Misztal

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

*Corresponding author: mbermann@uga.edu

Software resources

- BLUPF90 software suite
- Efficient for very large datasets
- Continuous development



The screenshot shows the BLUPF90 website interface. At the top left is the BLUPF90 logo. A search bar and 'Log In' link are at the top right. Below the logo, there are links for 'Media Manager' and 'Sitemap'. The main content area features a 'Table of Contents' sidebar with links to 'BLUPF90 Family of Programs', 'Headline', and 'Courses'. The main text includes the heading 'BLUPF90 Family of Programs', a sub-heading 'Now with support for genomic selection', and a paragraph mentioning 'Ignacy Misztal and collaborators, University of Georgia'. It describes the software as a collection of Fortran 90/95 programs for mixed model computations in animal breeding. Further text discusses variance component estimation and provides contact information for support via a discussion group.

Manual for

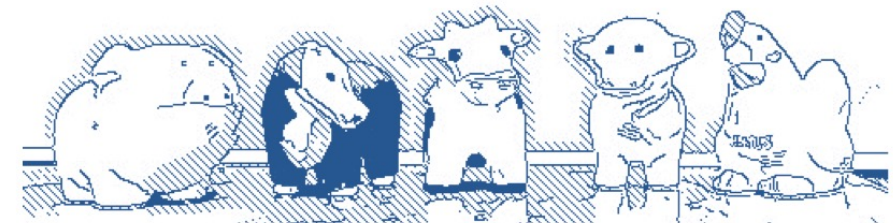
BLUPF90 family of programs

Ignacy Misztal (ignacy@uga.edu), Shogo Tsuruta (shogo@uga.edu),
Daniela Lourenco (danilino@uga.edu), Yutaka Masuda (yutaka@uga.edu)
University of Georgia, USA

Ignacio Aguilar (iaguilar@inia.org.uy)
INIA, Uruguay

Andres Legarra (andres.legarra@toulouse.inra.fr)
INRA Toulouse, France

Zulma Vitezica (zulma.vitezica@ensat.fr)
ENSAT, France



APY single-step + marker effects + p-values

1. Routine genetic evaluation

1. Calculate EBV (BLUPF90+ / BLUP90IOD3)
2. Approximate accuracies → PEV for core individuals (ACCF90GS2)

2. Indirect predictions for young individuals

1. Back-solve marker effects from EBV (POSTGSF90)
2. Obtain indirect predictions (PREDF90)
3. Calculate accuracies (PREDF90)

3. GWAS

1. Use marker effects from (2)
2. Calculate SD for marker effects' estimates (POSTGSF90)

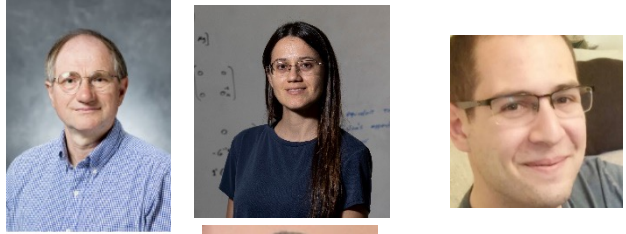


Conclusions

- APY models are equivalent to marker-effect models
 - Linear “imputation” for non-core individuals
 - Error term for non-core individuals
 - More core individuals -> APY converges to GBLUP
- Efficient computations:
 - Marker-effects estimates
 - Indirect predictions and accuracies
 - P-values
- Single-step analysis with larger datasets
 - Higher accuracy for predictions
 - Higher statistical power for GWAS



UGA Team



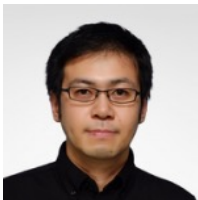
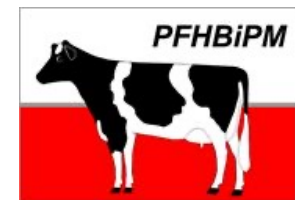
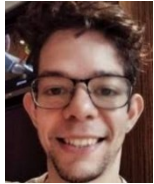
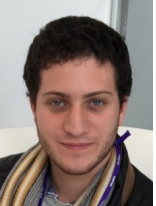
USDA United States Department of Agriculture
Agricultural Research Service

Warmwater Aquaculture Research Unit



USDA Agricultural Research Service
U.S. DEPARTMENT OF AGRICULTURE

Cool and Cold Water Aquaculture Research



Questions?

mbermann@uga.edu

