# Unknown-parent groups and incomplete pedigrees in single-step genomic evaluation

Ignacy Misztal, University of Georgia, Athens, GA, USA
Zulma Vitezica, Université de Toulouse, France
Andres Legarra, INRA, France
Ignacio Aguilar, INIA, Las Brujas, Uruguay
Andrew Swan, University of New England, Australia

# Introduction

- ssGBLUP - single-step genomic BLUP
  - Replaces $A^{-1}$ with $H^{-1}$ matrix
  - Any model
  - Efficient implementation (BLUPF90 package)
- Many applications at UGA or using UGA software
- In comparisons usually most accurate
  - No approximations with DYD or DP
  - Parental index included automatically
  - Simpler

# Problems

- Biases and convergence problems in some data sets using unknown parent groups (UPG)
- Convergence problems with complicated data sets – better with weight on $A_{22}^{-1}$

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - 0.7A_{22}^{-1} \end{bmatrix}$$

# Unknown parent groups (Westell et al., 1987)

$$\mathbf{y} = .. + \mathbf{ZQg} + \mathbf{Za} + \mathbf{e}$$

g – unknown parent group effect
a –animal effect
Q –matrix relating animals to UPG
Z – matrix relating records to animals

$$\begin{bmatrix} \mathbf{Z'Z+A^{-1}} & \mathbf{Z'ZQ} \\ \mathbf{-Q'Z'Z} & \mathbf{Z'Q'QZ} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z'y} \\ \mathbf{(ZQ)'y} \end{bmatrix}$$

BLUP with unknown parent groups after QP transformation (Quaas, 1988)

$$\mathbf{y} = .. + \mathbf{Zu} + \mathbf{e}, \qquad \mathbf{u} = \mathbf{Qg} + \mathbf{a}$$

$$\begin{bmatrix} Z'Z+A^{-1} & -A^{-1}Q \\ -Q'A^{-1} & Q'A^{-1}Q \end{bmatrix} \begin{bmatrix} \hat{u} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} Z'y \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} A^{-1} & -A^{-1}Q \\ -Q'A^{-1} & Q'A^{-1}Q \end{bmatrix}$$

easy to create from pedigree

# Unknown parent groups in single-step GBLUP

$$\begin{bmatrix} \mathbf{Z}'\mathbf{Z}+\mathbf{H}^{-1} & -\mathbf{H}^{-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{H}^{-1} & \mathbf{Q}'\mathbf{H}^{-1}\mathbf{Q} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'\mathbf{y} \\ \mathbf{0} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{Z}'\mathbf{Z}+\mathbf{H}^{-1} & -\left(\mathbf{A}^{-1}+\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1} \end{bmatrix}\right)\mathbf{Q} \\ -\mathbf{Q}'\left(\mathbf{A}^{-1}+\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1} \end{bmatrix}\right) & \mathbf{Q}'\left(\mathbf{A}^{-1}+\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1} \end{bmatrix}\right)\mathbf{Q} \end{bmatrix} \rightarrow$$

LHS

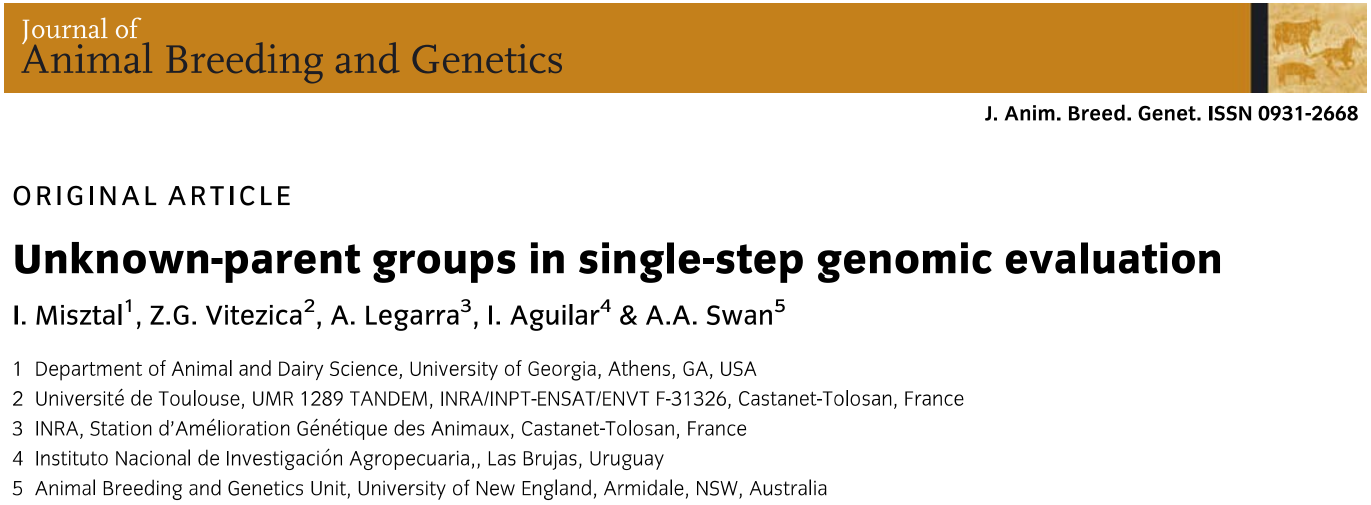
$$\begin{bmatrix} \mathbf{Z}'\mathbf{Z}+\mathbf{H}^{-1} & -\mathbf{A}^{-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'\mathbf{y} \\ \mathbf{0} \end{bmatrix}$$

Approximation

# Why unknown parent groups

- Different line
- Different breed (Harris and Johnson, 2012)

- Unrecorded parents across generations



Journal of
Animal Breeding and Genetics

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

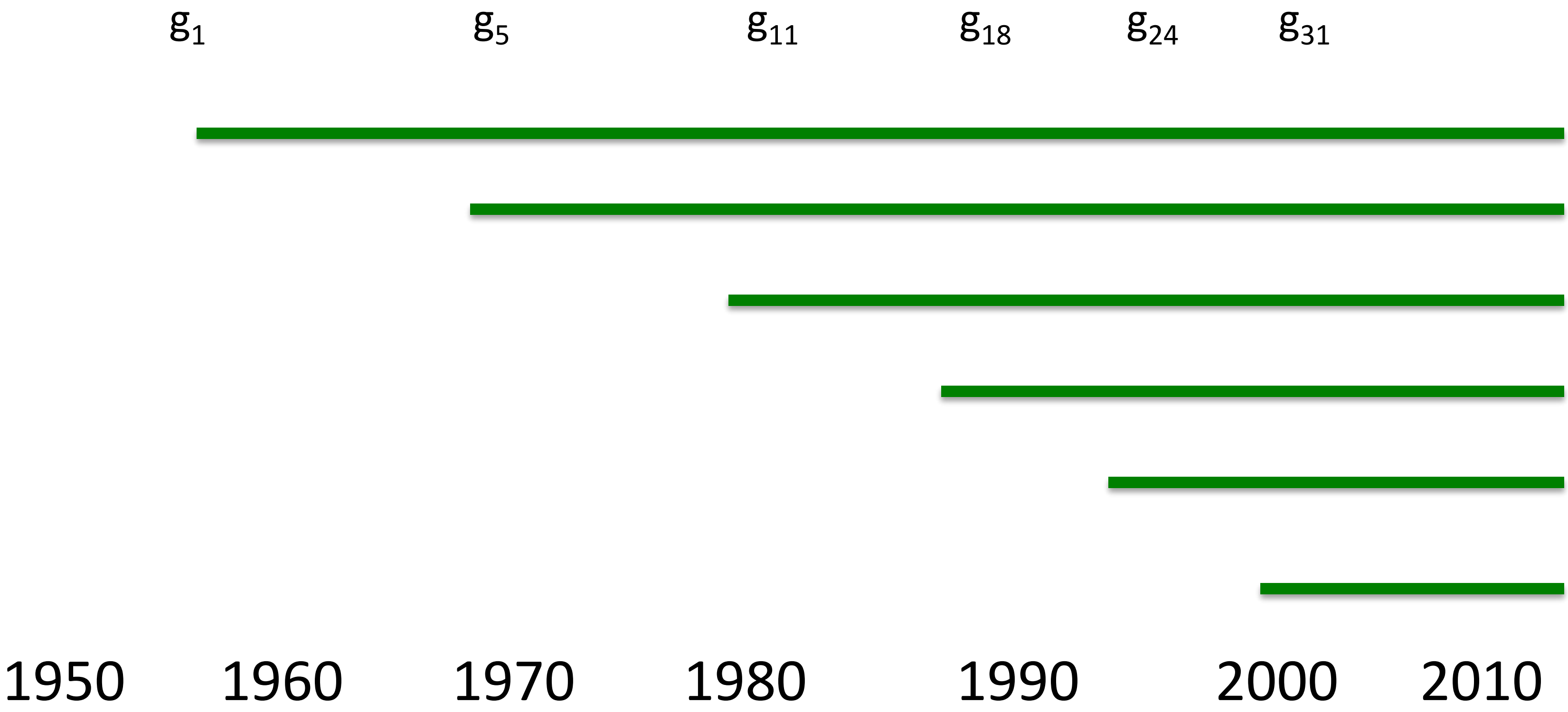
ORIGINAL ARTICLE

**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

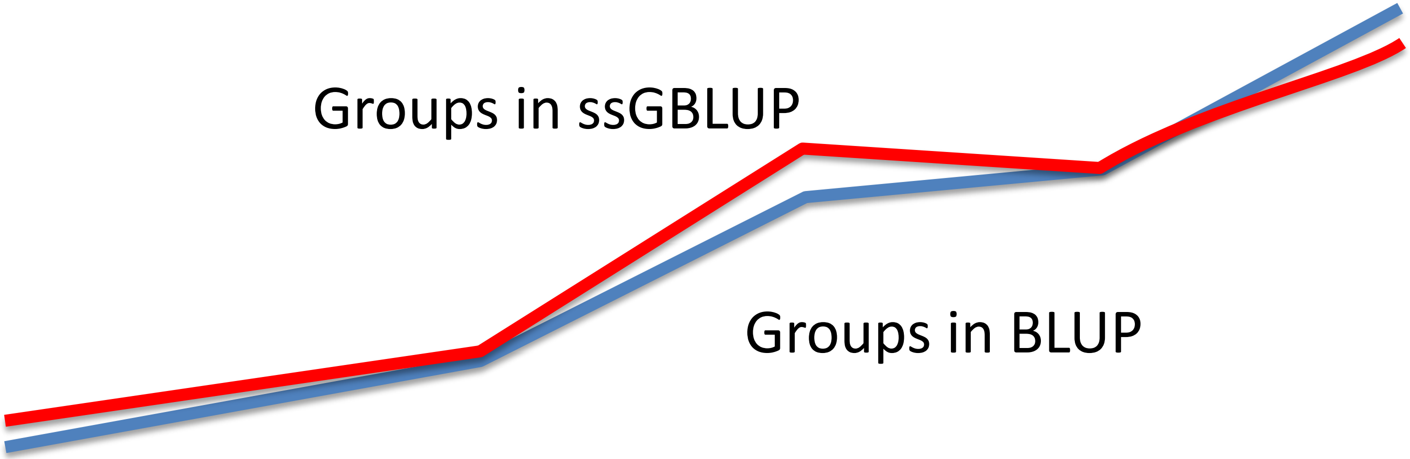


# Choices – add equations

$$\begin{bmatrix} 0 & -\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1} \end{bmatrix}\mathbf{Q} \\ -\mathbf{Q}'\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1} \end{bmatrix} & \mathbf{Q}'\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1} \end{bmatrix}\mathbf{Q} \end{bmatrix}$$

Complicated

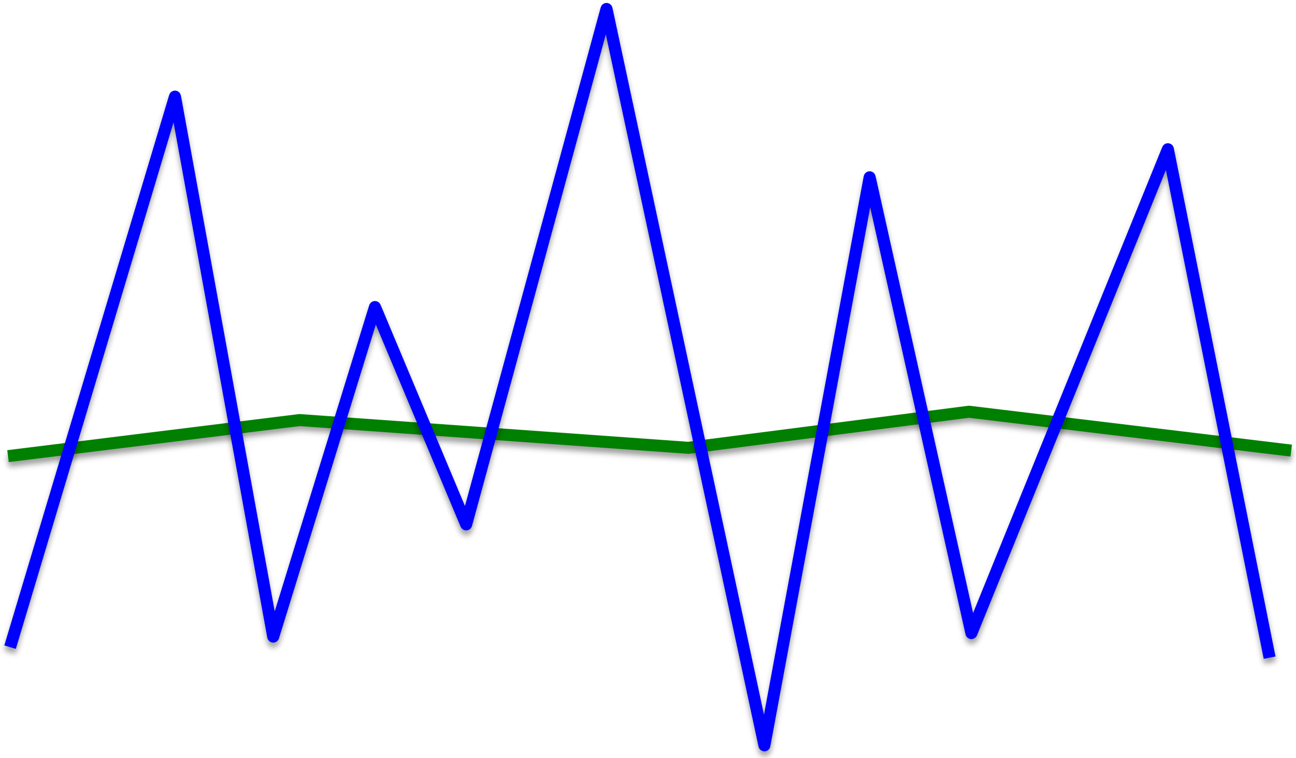
# Choices – do nothing



Groups in ssGBLUP

Groups in BLUP

# Choices – remove UPG from model

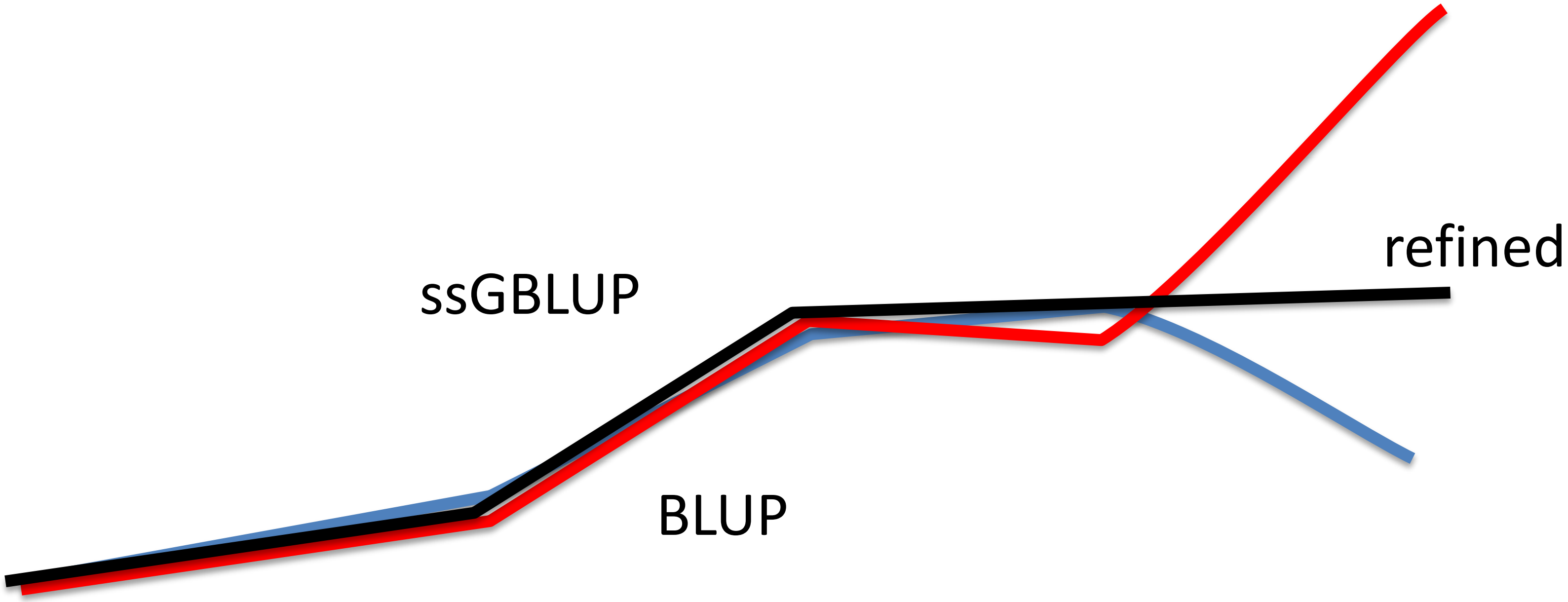


Trend

UPG in BLUP or ssGBLUP

No trend and large fluctuations in UPG

# Choices – redefine groups



See Tsuruta et al., 2013

# Choices – use explicit UPG

$$\mathbf{y} = .. + \mathbf{ZQg} + \mathbf{Za} + \mathbf{e}$$

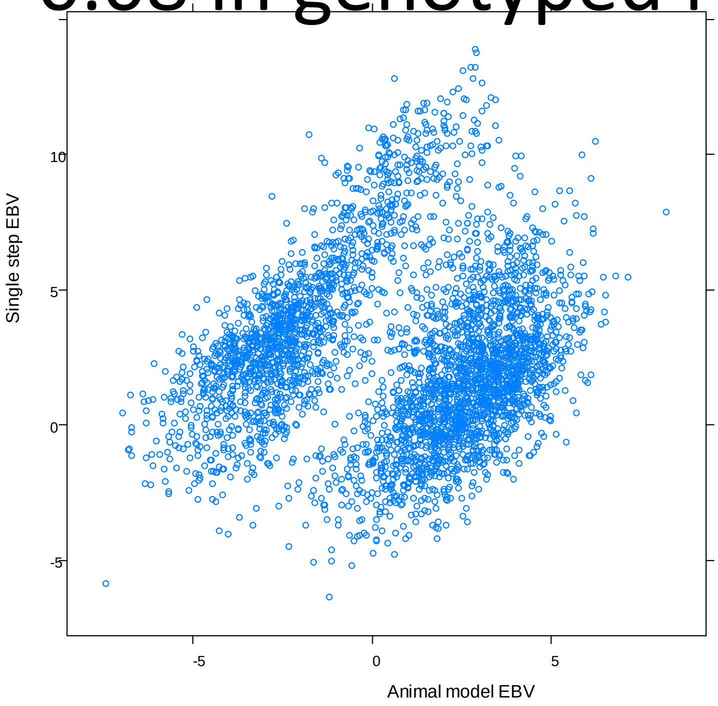
Q difficult to create if many groups
Adds many fixed effects

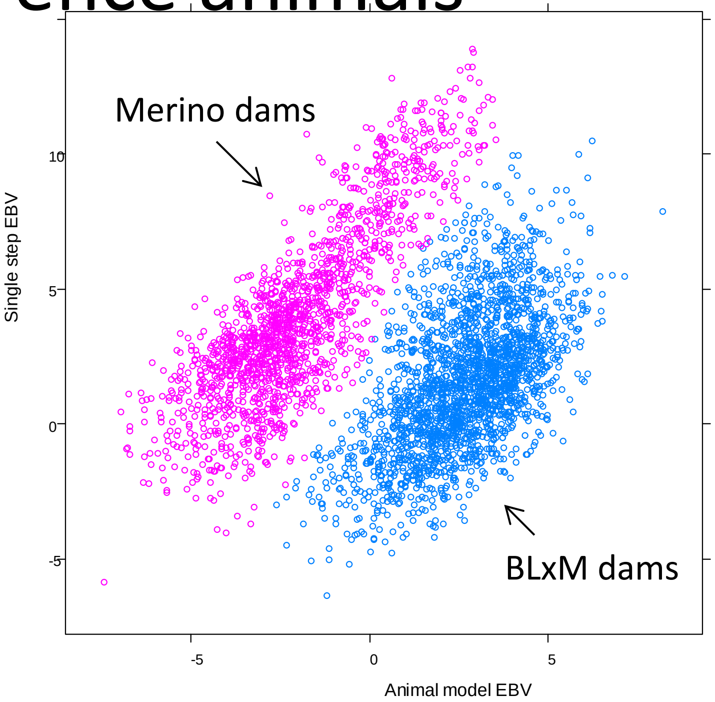
Choice in multiple-trait models if information very unequal per trait

Choice if parents have few phenotypes

# Australian sheep

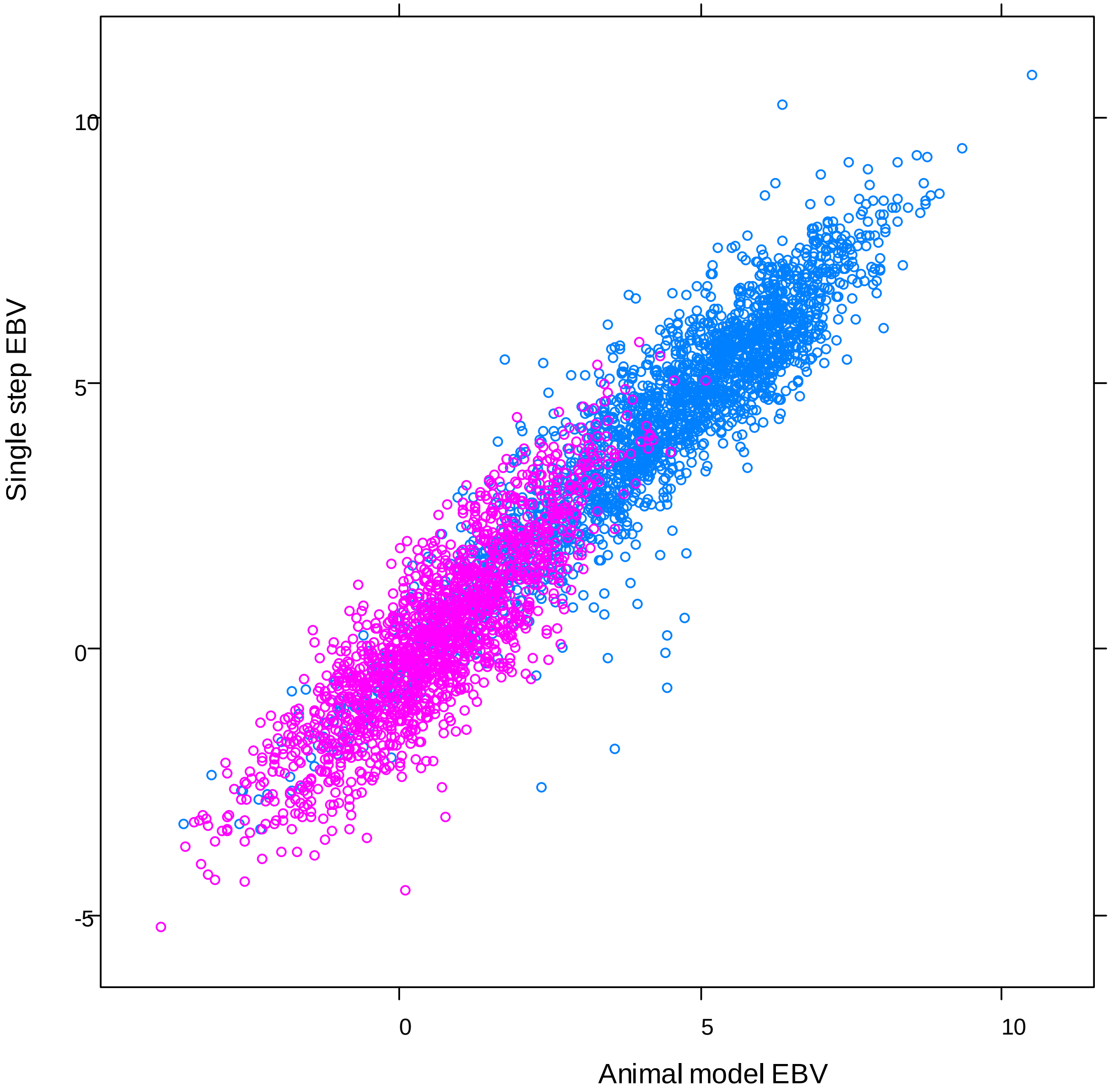
- Correlation between ssBLUP and BLUP EBVs = 0.08 in genotyped reference animals





Swan, 2012

# With explicit UPG model



# Choices – two animal effects

Polygenic effect with UPG

Genomic effect

$$\mathbf{y} = .. + \mathbf{Zu} + \mathbf{Za}_2 + \mathbf{e}$$

$$\mathbf{u}_1 \sim N(\mathbf{Qg}, \eta \mathbf{A} \sigma_a^2)$$

$$\mathbf{a}_2 \sim N(0, (1-\eta) \mathbf{H} \sigma_a^2)$$

Slows down convergence,
Some approximation

# Pedigree length and convergence

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

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

$$G^{-1} - A_{22}^{-1}$$

Good convergence and genotyped animals biased down

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

Bad convergence and genotyped animals biased up

$$G^{-1} - \underset{\text{Long}}{A_{22,1}^{-1}} - \underset{\substack{\text{medium}\\ \text{pedigrees}}}{A_{22,2}^{-1}} - \underset{\text{short}}{A_{22,3}^{-1}}$$

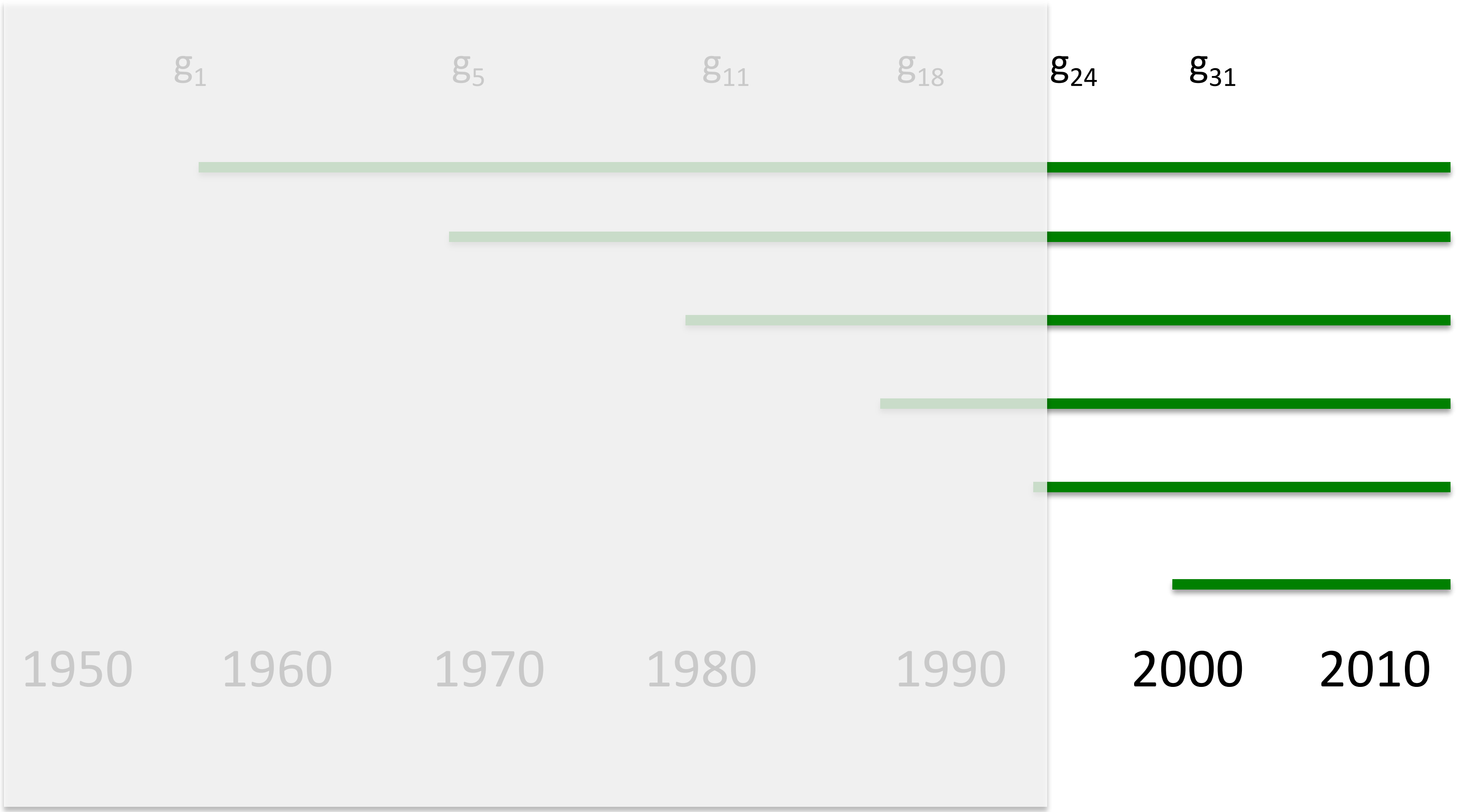
Bad convergence and genotyped animals biased down and up

# Why coefficient in $A_{22}^{-1}$ improves convergence rate?

$$\mathbf{G}^{-1} - 0.7\mathbf{A}_{22}^{-1} = \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$$

$$\mathbf{G}^{-1} - 0.7(\mathbf{A}_{22,1}^{-1} - \mathbf{A}_{22,2}^{-1} - \mathbf{A}_{22,3}^{-1}) = \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} - \mathbf{A}_{22,2}^{-1} - \mathbf{A}_{22,3}^{-1}$$

# Cut pedigree and data?



# Is very old data useful?

- When some parents cut – convergence improved 10 times, same accuracy (Forni et al., 2012)
- Data cut from 8 to 2 generations → same accuracy and reduced inbreeding (Mehrabani-Yeganeh et al., 1999)
- Little information beyond 3 generations (Luan et al., 2012; Lino et al., 2013)

# Conclusions

- UPG can cause problems in BLUP and ssGBLUP
- Many solutions – choice data driven
- Data beyond 3-4 generations possibly not useful
- Validation procedures of genomic selection useful for model refinement

# Acknowledgements

- AFRI Grants no. 2009-65205-05665 and 2010-65205-20366 from USDA NIFA
- Holstein Association
- PIC
- BARD Project IS-4394-11 R
- INPT, Toulouse, France
- AGBU Armidale & Bruce Tier