

# Practical problems and solutions using unknown parent groups in combined commercial pig sub-lines

Ivan Pocrnić<sup>1</sup>, D.A.L. Lourenco<sup>1</sup>, S. Tsuruta<sup>1</sup>, Y. Masuda<sup>1</sup>, C.Y. Chen<sup>2</sup>, I. Misztal<sup>1</sup>



The University of Georgia

<sup>1</sup>University of Georgia, USA

<sup>2</sup>Genus PIC, USA

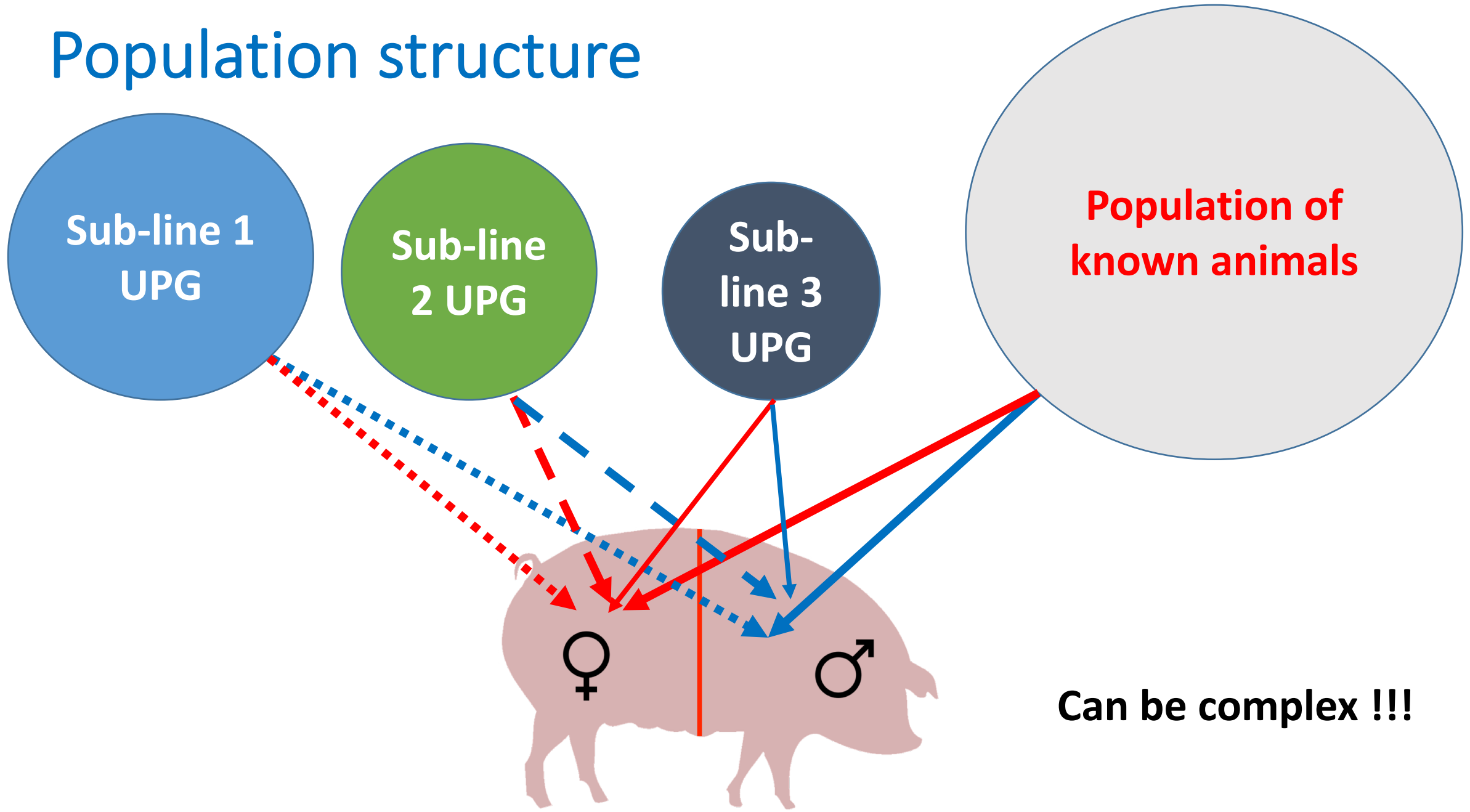
Vancouver, Canada, July 2018



# Introduction

- Unknown parent groups (**UPG**) are commonly used when different lines are present in genetic evaluations (heterogeneous population)
- Combining several lines from the same breed is common in pig breeding
- Can cause some problems, e.g.:
  - Confounding with other fixed effects
  - Convergence issues
  - Potential bias
  - Poorly assigned groups and/or small groups

# Population structure



# Data

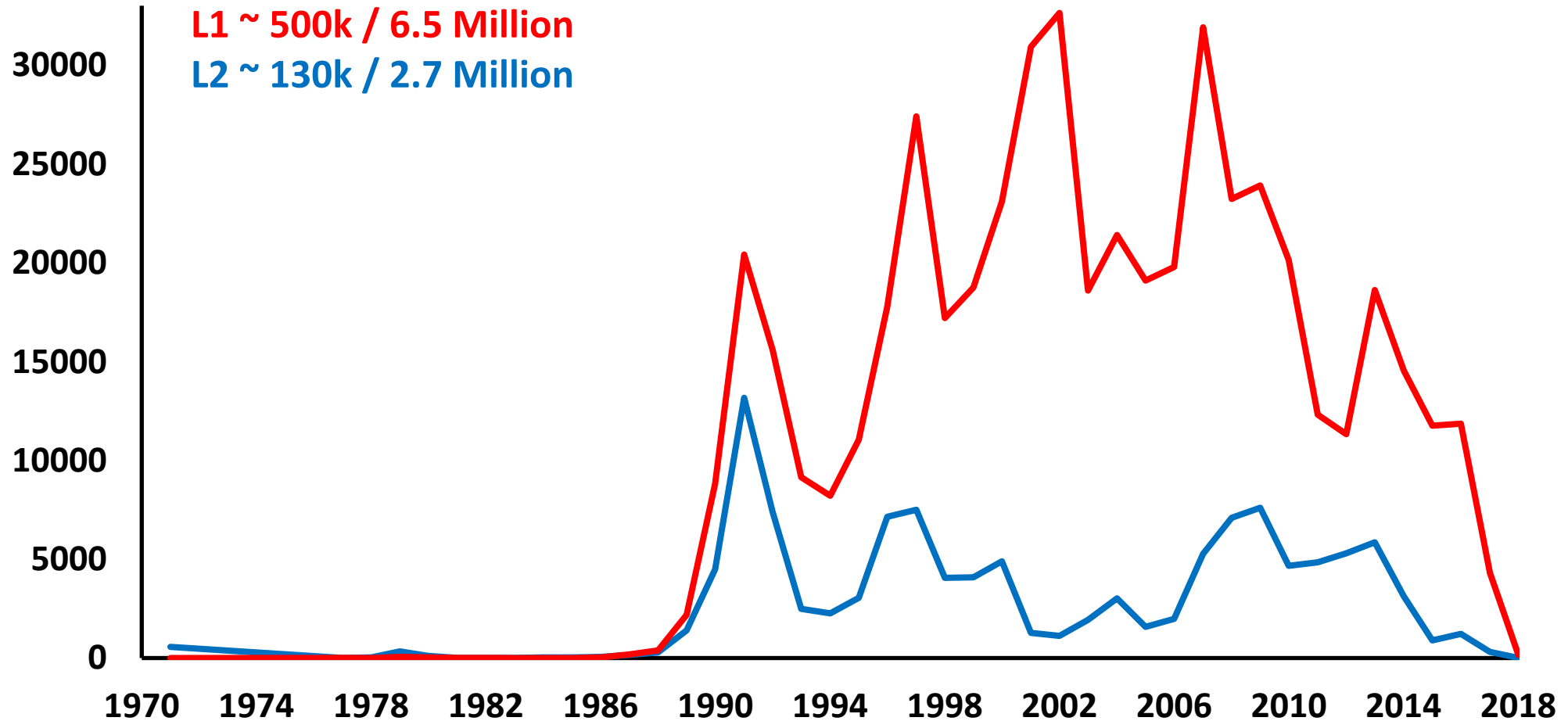
- **Line 1 (L1) – Case study I**

- Three-trait model
- Moderate heritabilities
  
- Up to 586,827 records
- 6.5 M pedigree
  
- 43,448 SNP markers
- 41,361 genotyped animals

- **Line 2 (L2) – Case study II**

- Single-trait model
- Low heritability
  
- 655,156 records
- 2.7 M pedigree
  
- 41,563 SNP markers
- 47,021 genotyped animals

# Animals with UPGs per year of birth



# Case study I: Line 1 problem

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

$$\mathbf{G} = \alpha \mathbf{G}_0 + \beta \mathbf{A}_{22}$$

- Small changes to  $\alpha$  &  $\beta$  caused instability of UPG and GEBV solutions
- Potential source of problem: weak connections between the sub-lines

# Case study I: Instability of UPG solutions

- Example: UPG2 – UPG1

| Alpha | TRAIT 1 | TRAIT 2 | TRAIT 3 |
|-------|---------|---------|---------|
| 0.95  | 10.5    | -5.2    | 7.5     |
| 0.9   | 8.9     | -4.9    | 6.9     |
| 0.8   | 9.4     | -5.3    | 7.1     |

# Case study I: Methods

- Traditional UPG model:  $y = \dots + \mathbf{ZQg} + \mathbf{Za} + e$

- QP transformed UPG model:  
(Quaas & Pollak 1981)  $y = \dots + \mathbf{Zu} + e$   
 $u = \mathbf{Qg} + a$

- Adding approximate UPG to ssGBLUP:

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

$\dots$  - other fixed effects  
 $g$  - fixed UPG effect  
 $a$  - random animal effect  
 $e$  - random residual  
 $\mathbf{Z}$  - connecting animals with records  
 $\mathbf{Q}$  - connecting animals with UPGs

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

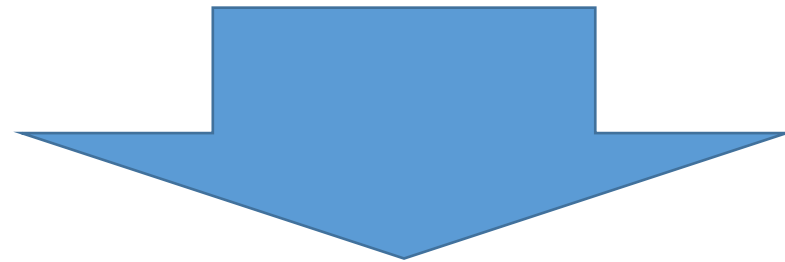
$$\lambda = \frac{\sigma_e^2}{\sigma_a^2}$$



# Case study I: Solutions

- Adding random UPG to ssGBLUP:

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



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

# Case study I: Results using random UPG

- Example: UPG2 – UPG1

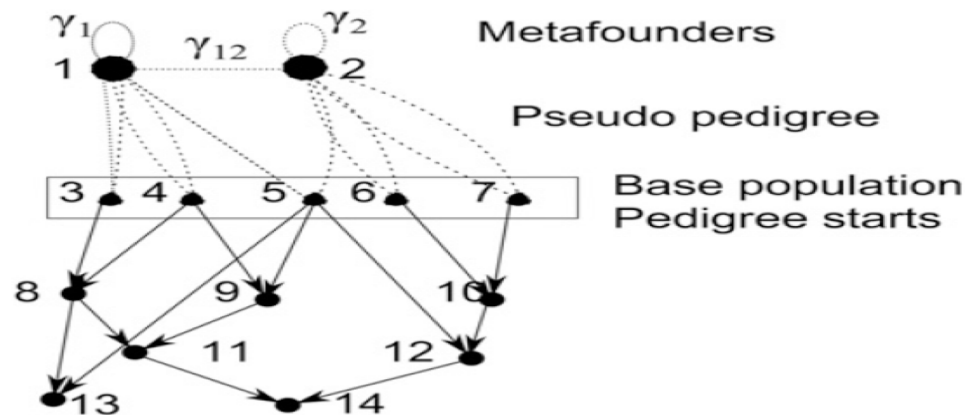
| Alpha | TRAIT 1 | TRAIT 2 | TRAIT 3 |
|-------|---------|---------|---------|
| 0.95  | -2.0    | -1.1    | 2.6     |
| 0.9   | -2.0    | -1.2    | 2.3     |
| 0.8   | -1.9    | -1.2    | 2.3     |

# Case study I: Alternatives

- Expand model to exact UPG (Misztal et al. 2013):

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

- Use metafounders instead UPG (Legarra et al. 2015):



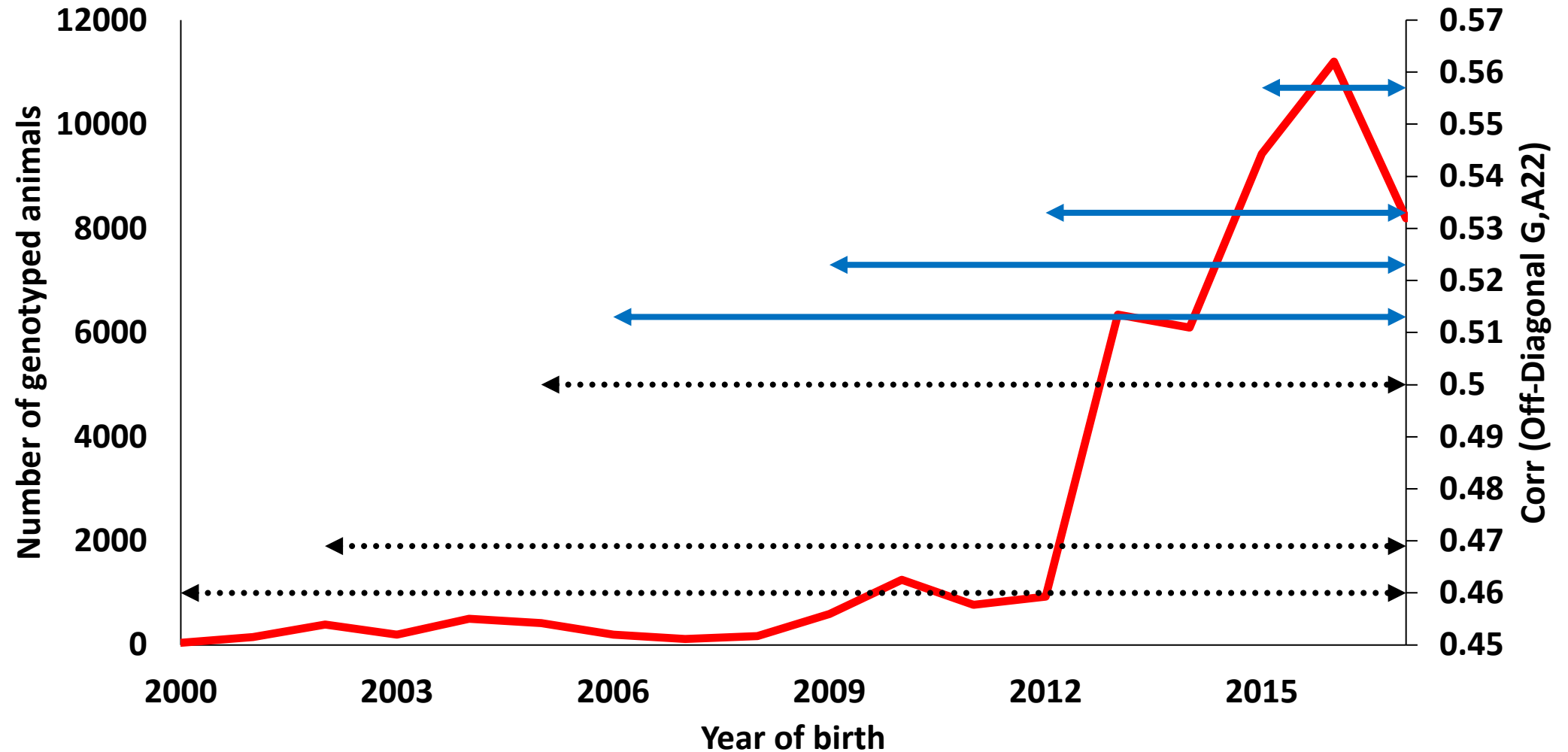
# Case study II: Line 2 problem

- Analysis was automatically stopped by program
- Source of the problem: Low correlations for off-diagonals in **G** and **A**<sub>22</sub>
- Default threshold in BLUPf90 is 0.3 to STOP / 0.5 to ALERT
  
- Typically  $0.6 \pm 0.15$
- Depend on pedigree depth
  
- Especially low in pig datasets:
  - Disconnected lines - connected through **G**, but not through **A**
  - A lot of full-sibs (e.g. 0.5 in **A**, but large deviation in **G**)

# Case study II: Methods

- ssGBLUP using several options:
  - Change  $\alpha$  &  $\beta$  ; ( $\alpha = 0.95, 0.9, 0.8$ ;  $\beta = 1 - \alpha$ )
  - Use random UPG
  - Truncate genotypes (based on year of birth)
  - Adjust threshold in BLUPf90 for off-diagonal correlations between **G** and **A<sub>22</sub>**
- Validation:
  - Changes to the mean GEBV for selection candidates
  - Changes to the UPG solutions

# Case study II: Truncation of genotypes



# Case study II: Solutions

1. Lower the threshold for correlation of off-diagonal elements (**G** and **A<sub>22</sub>**) in BLUPf90 – use default values
  2. Remove older genotypes
- Previous studies at UGA showed that reduction of data/pedigree doesn't influence accuracy, but helps convergence and stability (e.g. Lourenco et al. 2014; Pocrnic et al. 2017)

## Case study II: G-matrix elements statistics

| Allele Frequencies | r [off-diagonal (G,A)] | r [inbreeding (G,A)] |
|--------------------|------------------------|----------------------|
| <b>Current</b>     | 0.46                   | -0.17                |
|                    |                        |                      |
| <b>Equal (0.5)</b> | 0.78                   | 0.57                 |

- Highest values using the base allele frequencies (VanRaden, 2008)



# General problems & solutions

1. UPG problems - not estimable functions and dependent on population structure – weak connections between the sub-lines
    - Simple: Model UPG as random effects
  2. Old genotypes problems – due to quality of genotyping and/or multiple imputations
    - Simple: Remove old genotypes
- Solutions on several levels:
    - Data manipulation
    - Model assessment
    - Software fine adjustment

Thank you !!!