



UNIVERSITY OF  
**GEORGIA**

# Single-step genomic REML with APY and reduced pedigree



**Daniela Lourenco**

**Vinicius Junqueira, Yutaka Masuda, and Ignacy Misztal**

EAAP - Dubrovnik, Croatia - 05/30/2018

# AI REML in Animal Breeding and Genetics

- Most popular for variance components estimation
- Converges in few rounds
- Fast operations with medium-sized data
- Efficient computations based on sparse matrix techniques (AIREMLf90)
  - FSPAK (Pérez-Enciso et al., 1994)
  - YAMS (Masuda et al., 2014) - 10x faster

# Computing cost in AI REML

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

$$\theta_{n+1} = \theta_n - \mathbf{H}^{-1}(\theta_n)\mathbf{d}(\theta_n)$$

$$-2\mathbf{d}(\theta) = \begin{bmatrix} \frac{N_a}{\sigma_a^2} - \frac{\text{tr}(\mathbf{A}^{-1}\mathbf{C}^{uu})}{(\sigma_a^2)^2} - \frac{\hat{\mathbf{u}}'\mathbf{A}^{-1}\hat{\mathbf{u}}}{(\sigma_a^2)^2} \\ \frac{N - \text{rank}(\mathbf{X})}{\sigma_e^2} - \frac{1}{\sigma_e^2} \left[ N_a - \frac{\text{tr}(\mathbf{A}^{-1}\mathbf{C}^{uu})}{\sigma_a^2} \right] - \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{(\sigma_e^2)^2} \end{bmatrix}$$

↓

$$\text{tr}(\mathbf{A}^{-1}\mathbf{C}^{uu})$$

Steps for obtaining  $\mathbf{C}^{uu}$

- 1) Ordering
- 2) Factorization
- 3) Inversion

# Computing cost in AI ssGREML

- If genomic information is available for a fraction of pedigree animals
  - ssGREML

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{H}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$



$$tr(\mathbf{H}^{-1}\mathbf{C}^{uu})$$

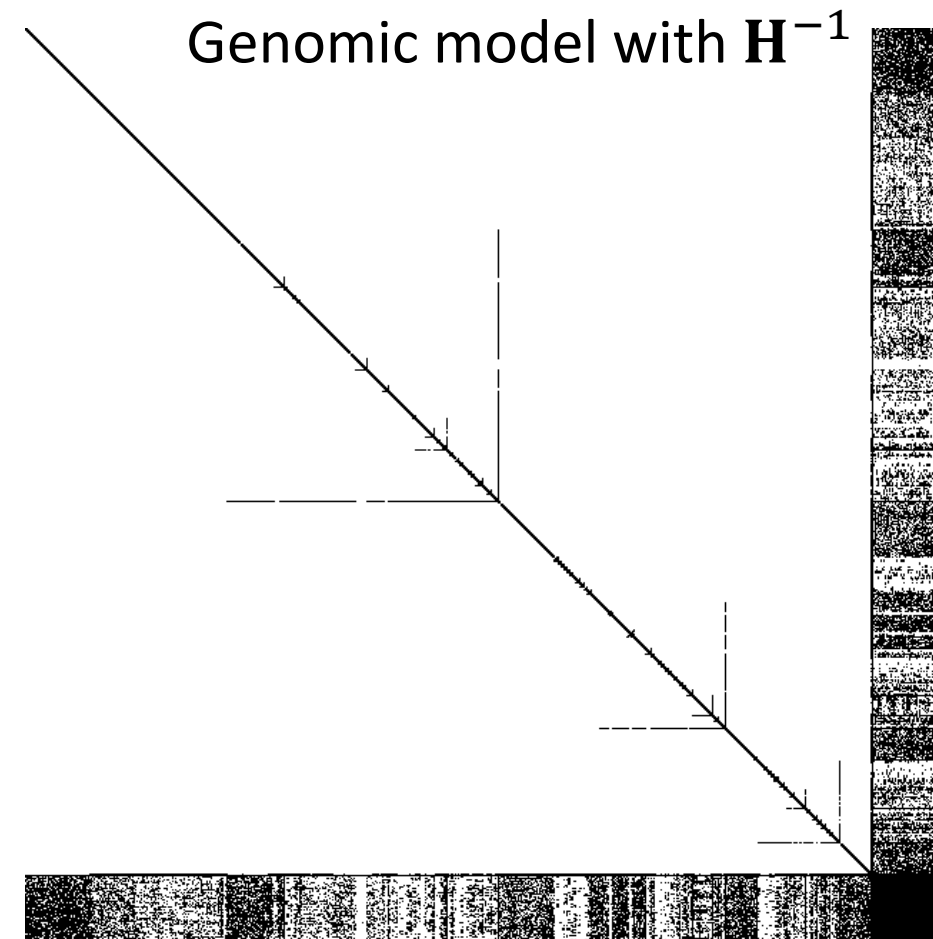
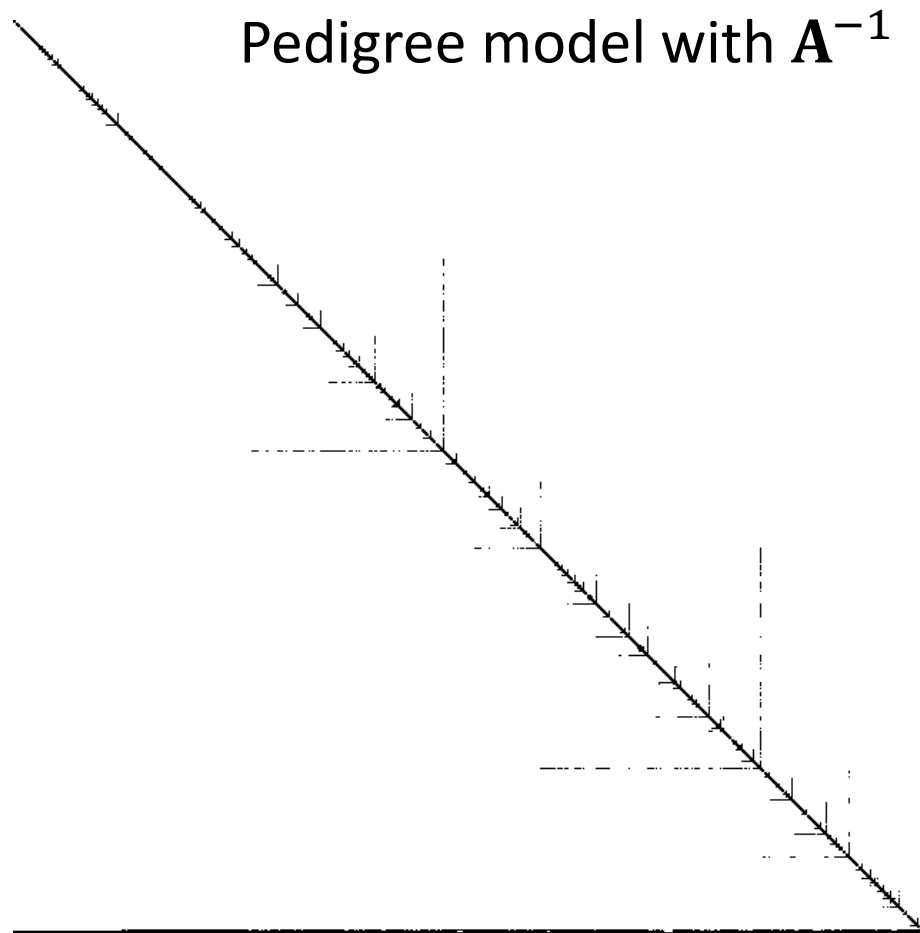
Steps for obtaining  $\mathbf{C}^{uu}$

1) Ordering

2) Factorization

3) Inversion

# Sparse pattern in the Cholesky factor



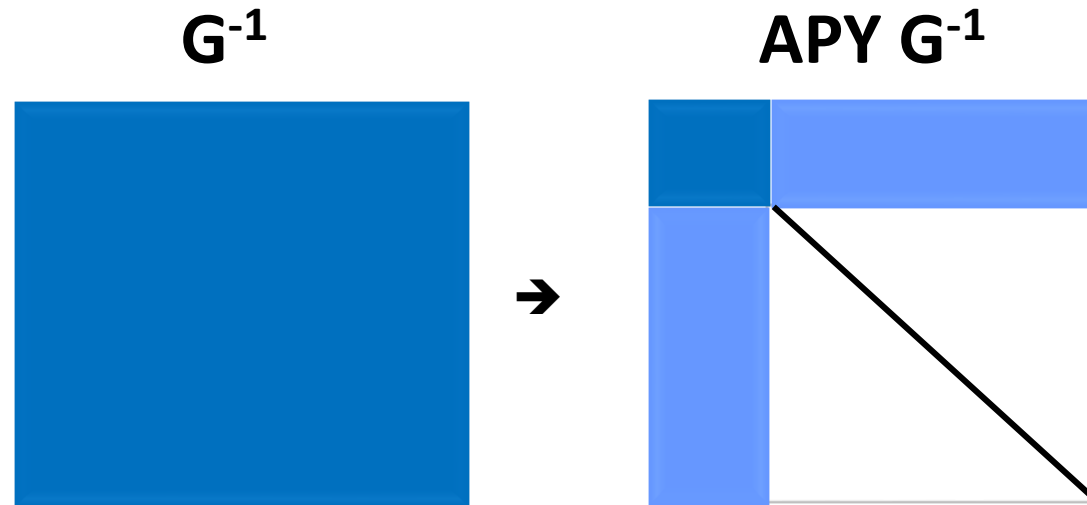
# Single-step with large number of genotypes

- $\mathbf{H}^{-1}$  makes the system denser because of  $\mathbf{G}^{-1}$
- High computing cost with more genotypes (denser)
- Can we increase sparsity of  $\mathbf{H}^{-1}$ ?

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

# Single-step with large number of genotypes

- APY  $\mathbf{G}^{-1}$  is sparser than  $\mathbf{G}^{-1}$  (Misztal et al., 2014)
  - Limited dimensionality of genomic info



# Single-step with large number of genotypes

- Computing advantages of APY  $\mathbf{G}^{-1}$  in ssGBLUP
  - Angus data with 82k genotyped animals (Lourenco et al., 2015)
    - Reduced computing time by 8-fold
    - Reduced memory requirements by 14-fold



# Objectives

1) Feasibility of AI ssGREML using APY  $\mathbf{G}^{-1}$

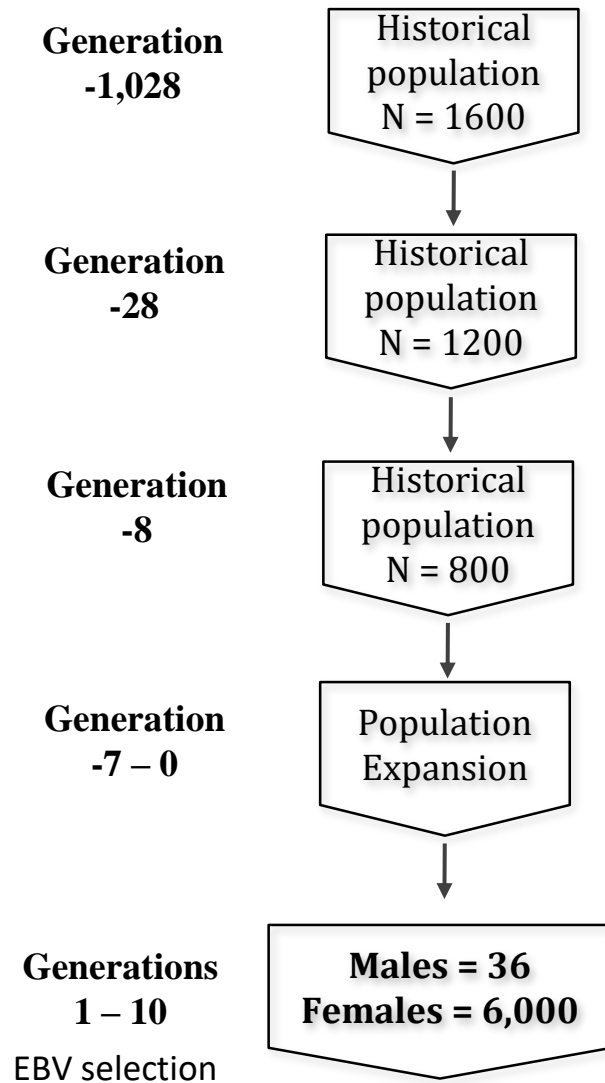
2) Benefits of increasing sparsity in  $\mathbf{A}_{22}^{-1}$  given that APY  $\mathbf{G}^{-1}$  is sparse

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

- $\mathbf{A}^{-1}$  and  $\mathbf{A}_{22}^{-1}$  are sparser if pedigree is shorter

3) Impact on VC and properties of inverse of coefficient matrix

# Dataset



- **Genome**

- Genome size: 30M
- 30 chromosomes
- 50,000 SNP
- 5000 QTL

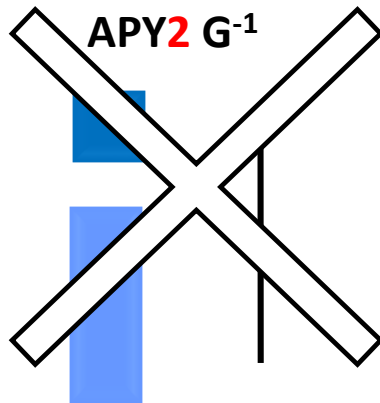
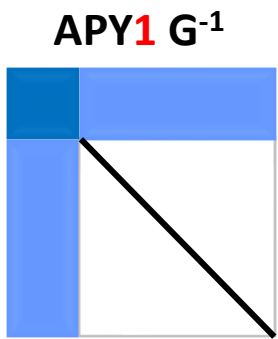
- **Population**

- 66,036 in pedigree
- 18,000 genotyped (G8, G9, G10)
- Phenotypes for all but last
- $h^2 = 0.3$

(QMSim; Sargolzaei and Schenkel, 2009)

# Scenarios

- AI ssGREML
- AI ssGREML + APY
  - 5k core animals

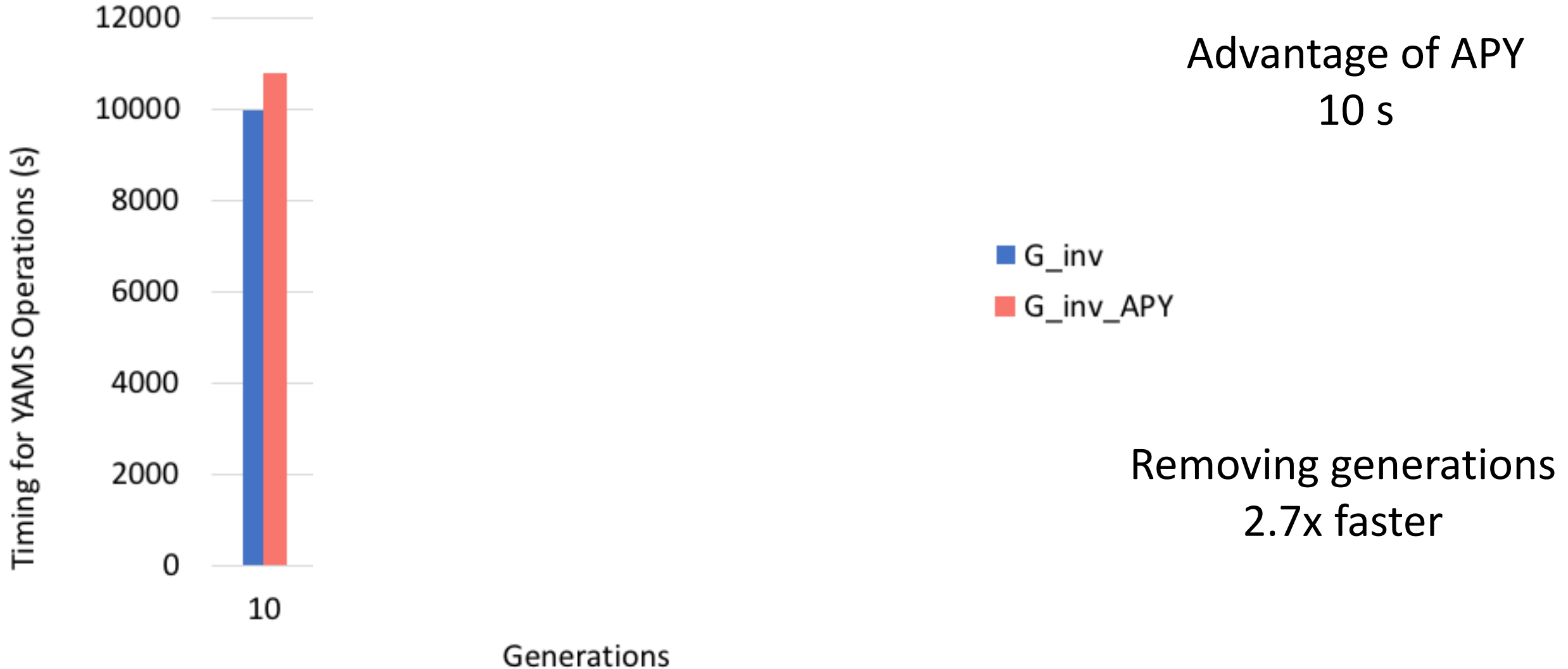


- Software
  - AIREMLF90 (Misztal et al., 2002)
    - + YAMS (Masuda et al., 2015)
    - + APY (Masuda et al., 2016)

# Comparisons

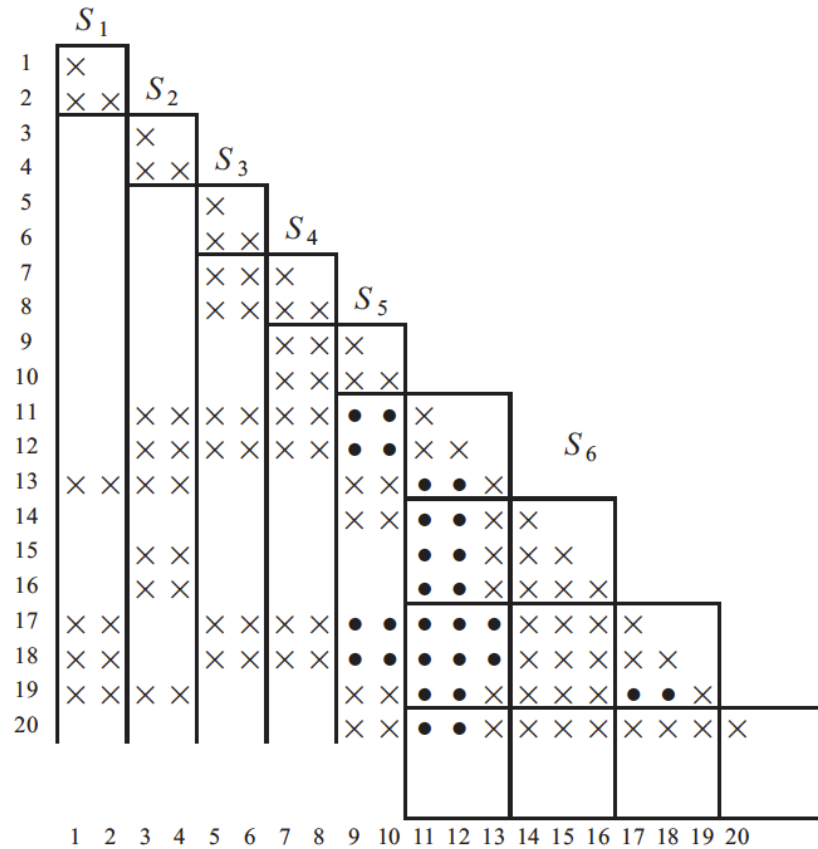
- Computing time
- Number of nonzero elements
  
- Heritability
- Accuracy of GEBV from PEV

# Computing time (CPU)



# Computing time

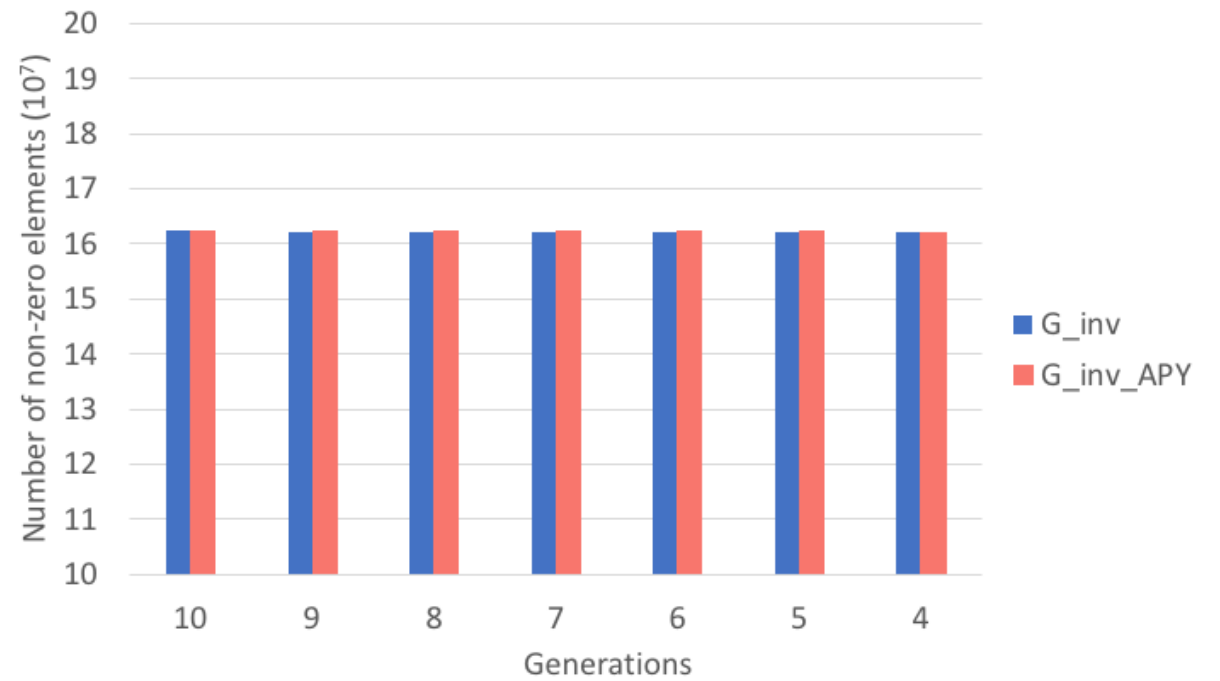
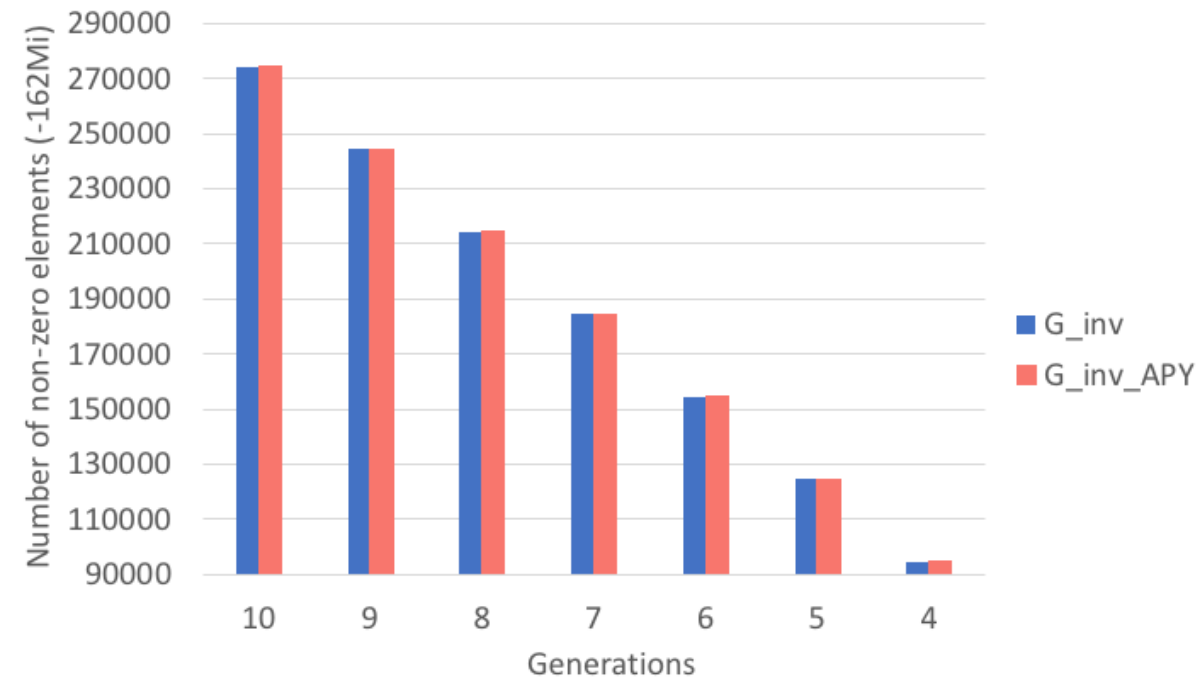
- Why no big computing advantages with APY  $\mathbf{G}^{-1}$  ?



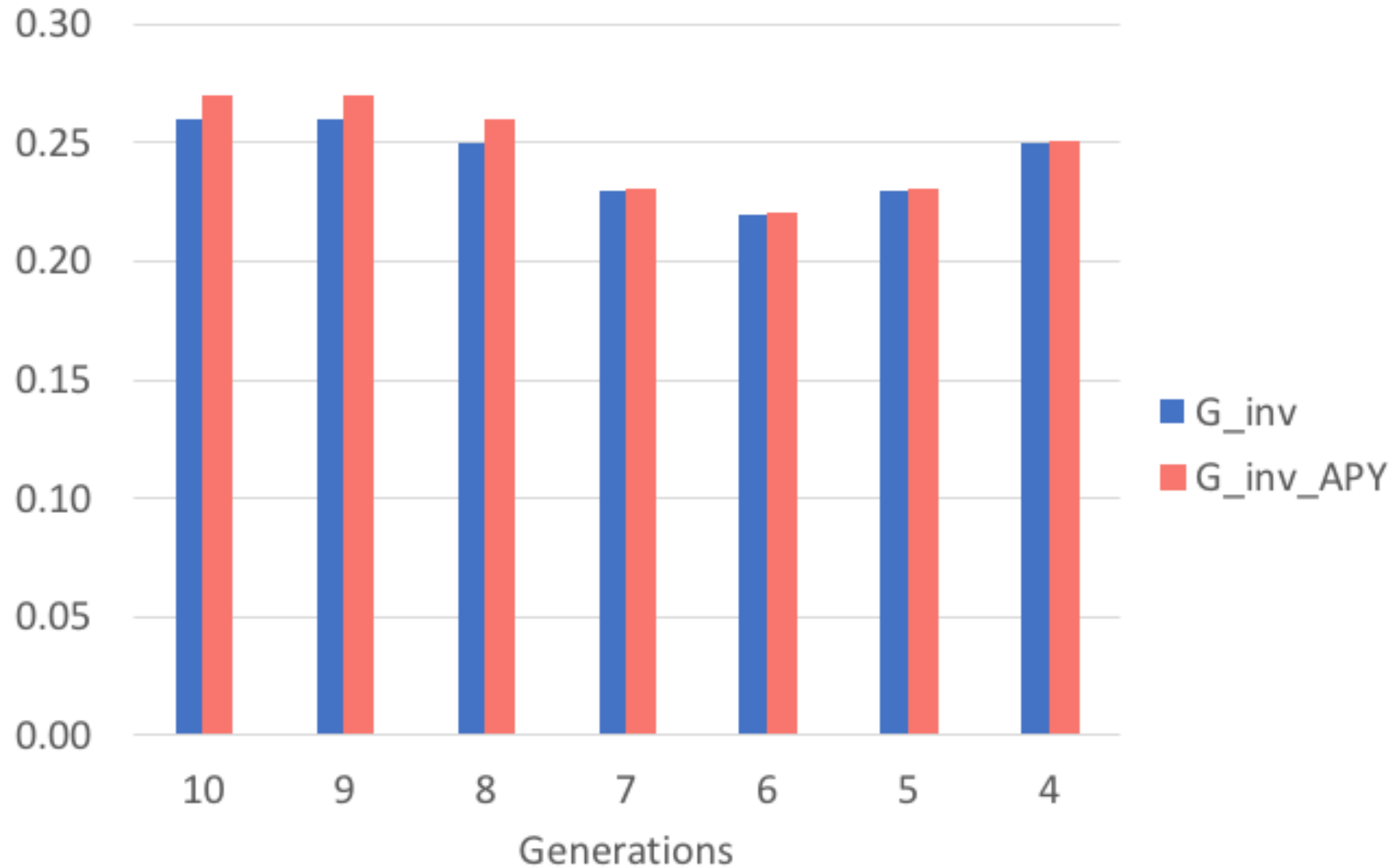
*Fill-in* effect: zero element in original and non-zero in the factor

**Figure 2** An example for a storage of nonzero elements in  $\mathbf{L}$ ; a triangular and a rectangular part are separately saved in  $S_i$  ( $i = 1-5$ ); the last supernode  $S_6$  is split into 10 squared submatrices (panels) with the order of 3.

# Number of non-zero elements

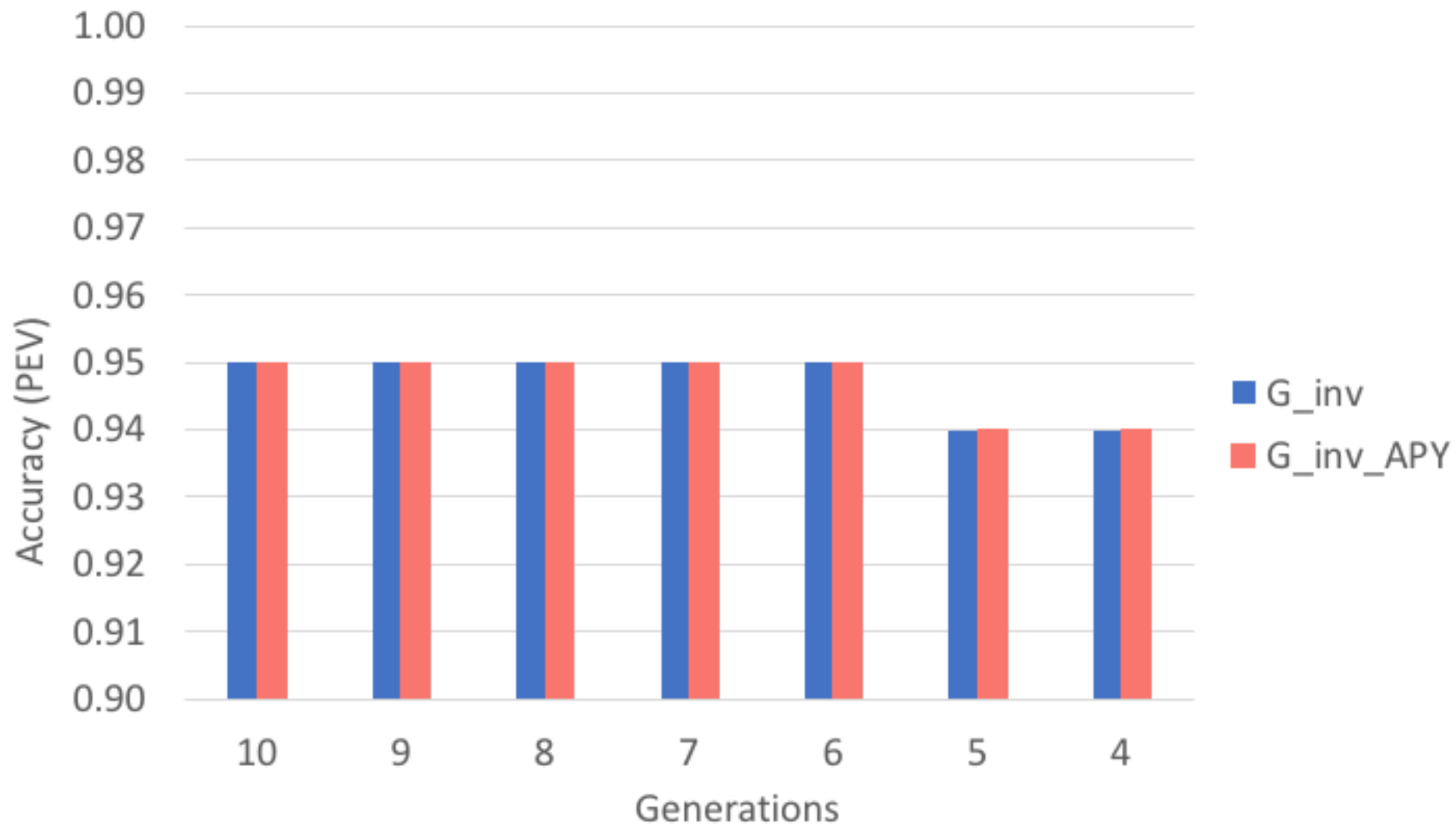


# Heritability





# Accuracy based on PEV



# Conclusion

- APY  $\mathbf{G}^{-1}$  can be used for variance components estimation
  - Similar heritabilities
  - No reduction in computing time and memory usage
  - Increasing sparsity in  $\mathbf{A}^{-1}$  reduces computing time
    - $\mathbf{G}^{-1}$  and APY  $\mathbf{G}^{-1}$
  - Benefits with more genotyped animals?
- PEV can be accurately estimated based on APY  $\mathbf{G}^{-1}$