Dimensionality of genomic information and APY inverse of the genomic relationship matrix

I. Pocrnic[1], I. Misztal[1], D.A.L. Lourenco[1], Y. Masuda[1], A. Legarra[2]

[1]University of Georgia, USA

[2]INRA, France

EAAP, Belfast UK, Aug/Sep 2016

Introduction

- Single-step GBLUP

- Requires inverse of genomic relationship matrix (GRM)

- Proposed solution by Misztal et al. (2014)
  - GRM can be efficiently inverted by Algorithm for Proven and Young (APY) – recursion on small number of core animals
- Theory by Misztal (2016)
  - Limited dimensionality of genomic information
  - Limited number of effective SNP markers (ESM) or independent chromosome segments (Me)

Algorithm for Proven and Young (APY)

core animals (c)

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

noncore animals (n)
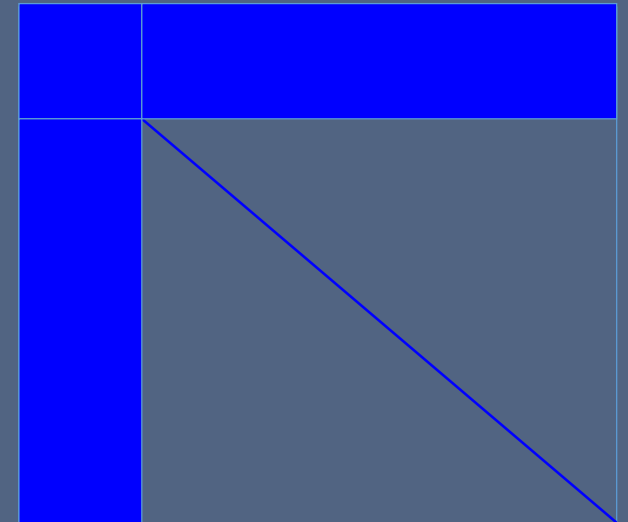
APY $\mathbf{G}^{-1}$

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

BV of noncore animals are linear function of BV of core animals

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

ssGBLUP and APY

- Fragomeni et al. (2015); 100,000 genotyped US Holstein

- Lourenco et al. (2015); 36,000-50,000 genotyped US Angus

- Masuda et al. (2016); 570,000 genotyped US Holstein (< 2h)

- Ostersen et al. (2016); 13,000-21,000 genotyped Danish pigs

- Pocrnic et al. (2016 - submitted); US Holstein, US Jersey, US Angus, broiler chicken, and pigs

Independent chromosome segments (Me)

- $E(Me) = 4N_eL$ Stam (1980)

  - N_e – Effective population size
  - L – Length of genome in Morgans

$$Me \begin{cases} 2N_eL & \text{Hayes et al. (2009)} \\ 2N_eL/[\log(N_eL)] & \text{Goddard et al. (2011)} \\ \text{Many more ...} & \text{Brard and Ricard (2015)} \end{cases}$$

Objectives

- Test the theory of APY inverse with simulated data

- What is optimal number of core animals?

- Are GEBVs by APY more accurate than by regular inverse?

- How to determine dimensionality of GRM?

Simulation

- 6 populations (Ne=20,40,80,120,160,200)

- 10 discrete generations taken from historical population

- Mating random / no selection

- Generations 8-10 genotyped (75,000 individuals)

- Validation: Gen. 10 (no phenotypes)

- 30 Chromosomes (100 cM each)

- 49,980 SNPs

- 4,980 QTLs

- Heritability = 0.3

- 5 replications
 - QMSim software (Sargolzaei and Schenkel, 2009)

Number of core animals

- Construct $\mathbf{G_0} = \dfrac{\mathbf{ZZ'}}{2\Sigma p_j(1-p_j)}$ (VanRaden, 2008)

- Find number of the largest eigenvalues in $\mathbf{G_0}$ that explained 90, 95, 98 and 99 percent of variance

- Core animals were randomly selected

Dimensionality of genomic information

- $\mathbf{G_0} = \dfrac{\mathbf{ZZ}'}{2\Sigma p_j(1-p_j)}$ (VanRaden, 2008)

  - Usually not full rank (usually blended)

- Dimensionality of $\mathbf{G_0} \leq \min(N_{SNP}, N_{IND}, Me)$

  - N_{SNP} – number of SNPs
  - N_{IND} – number of genotyped animals
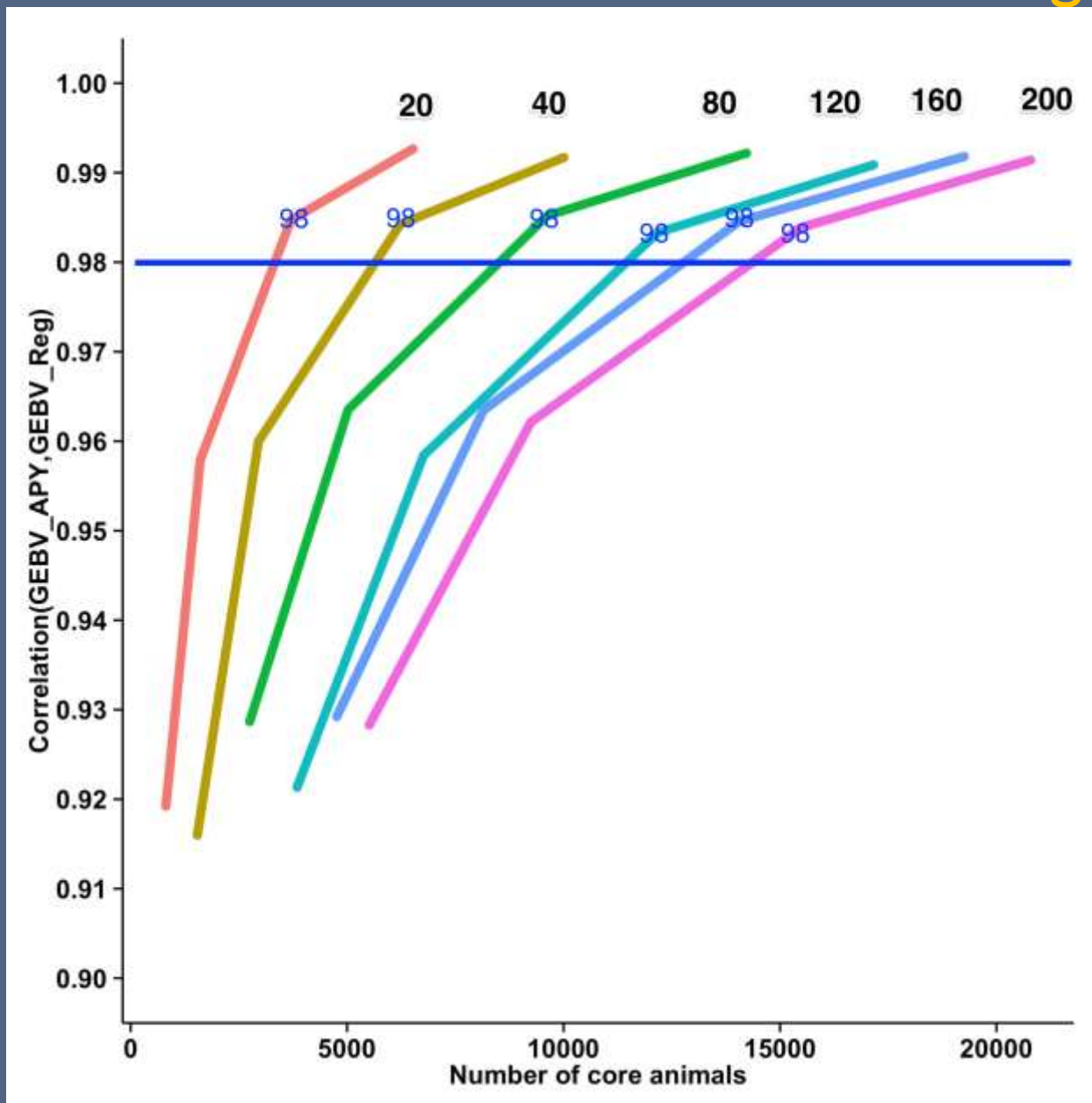 - Hypothesis: Me – number of independent chromosome segments

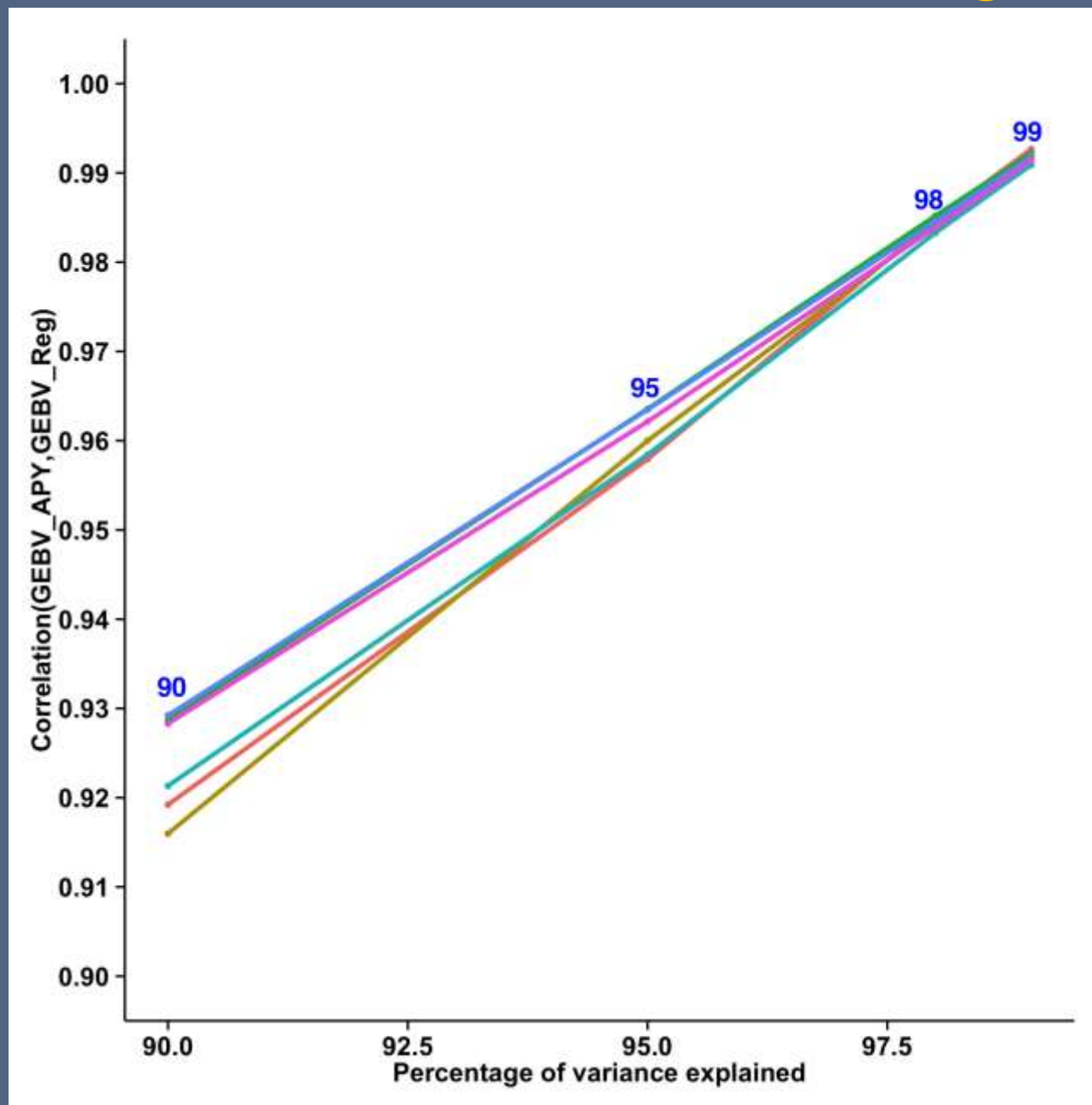# Number of largest eigenvalues to account for a given variance
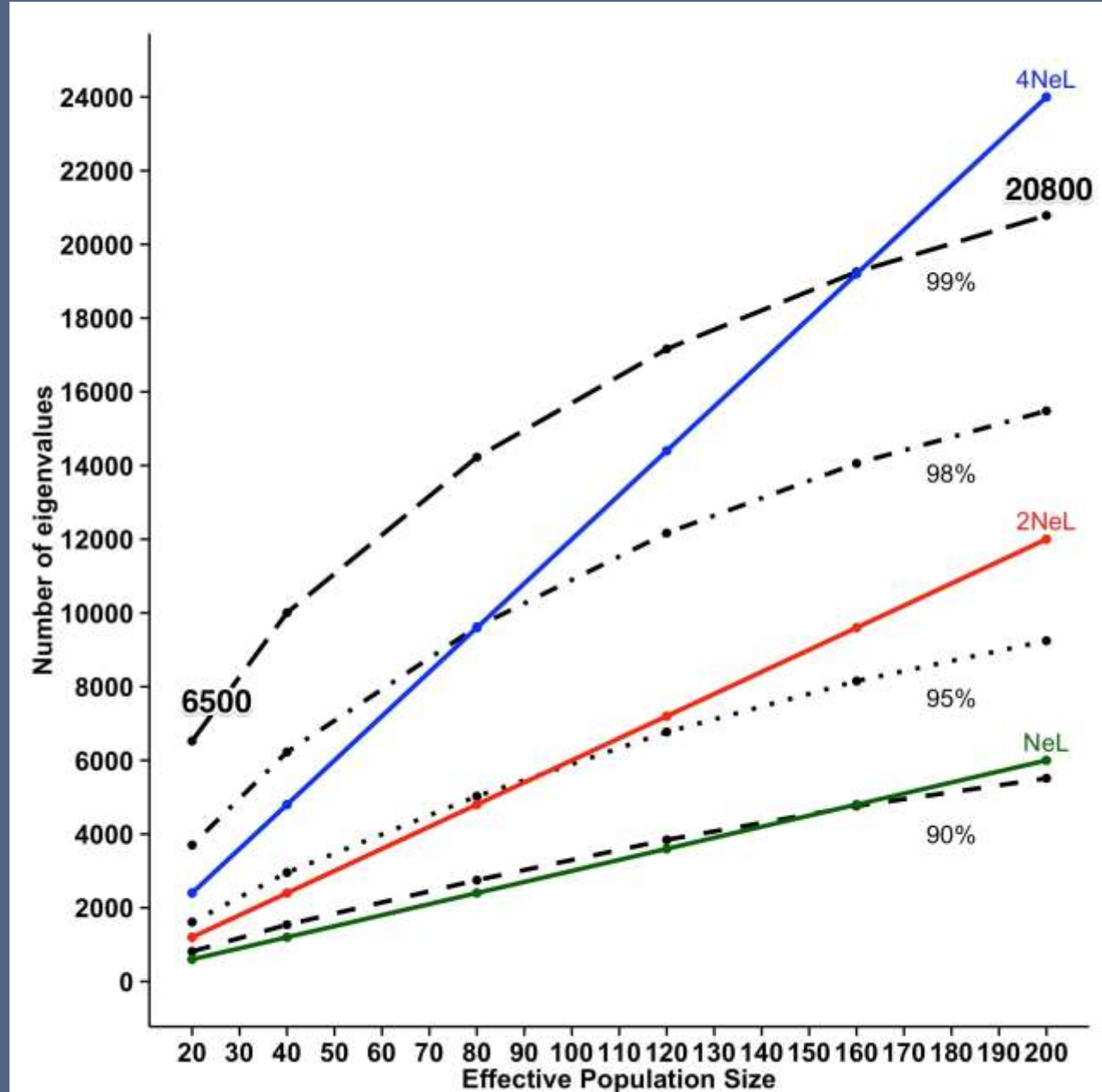
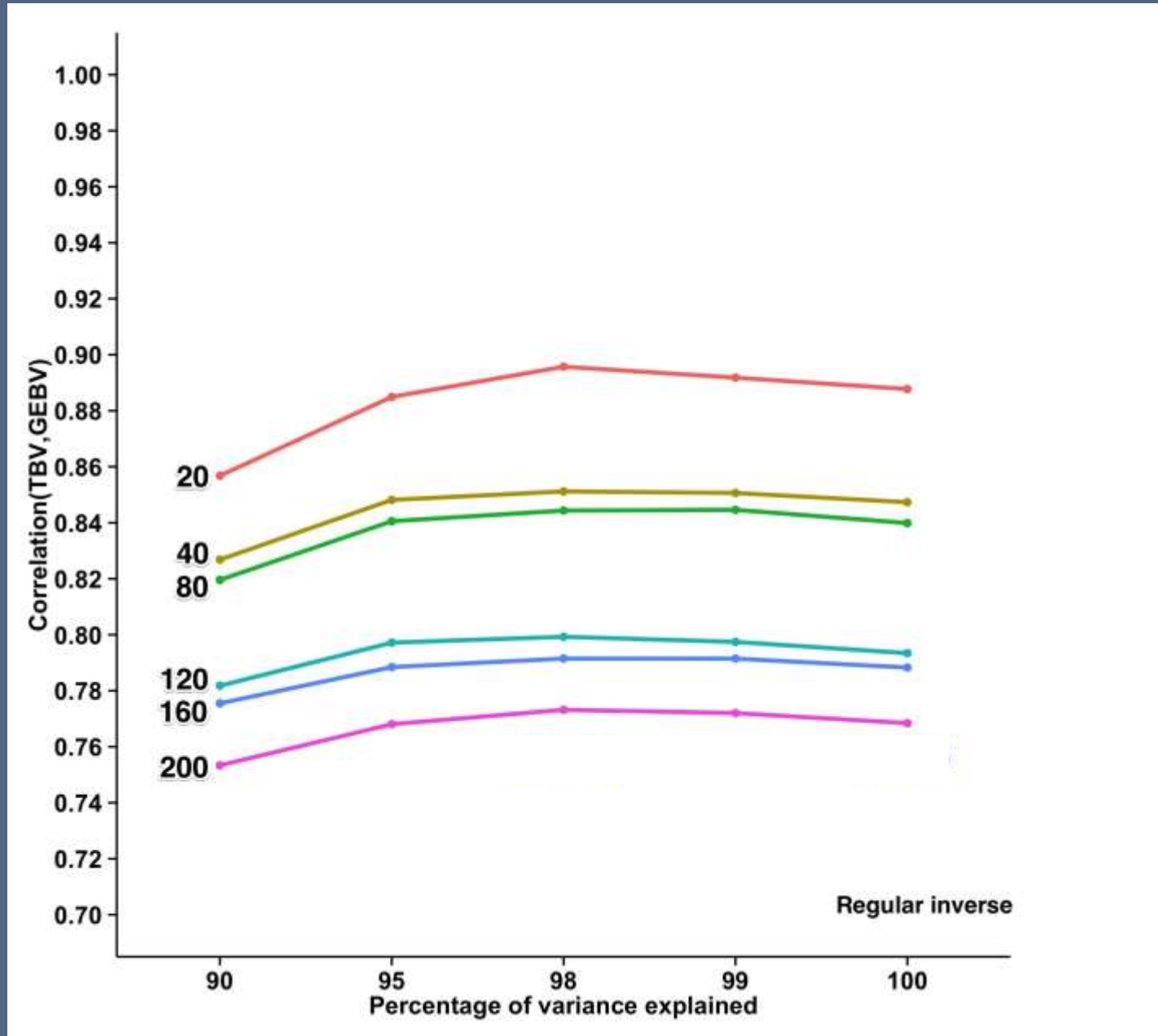# Number of largest eigenvalues standardized to Ne=20
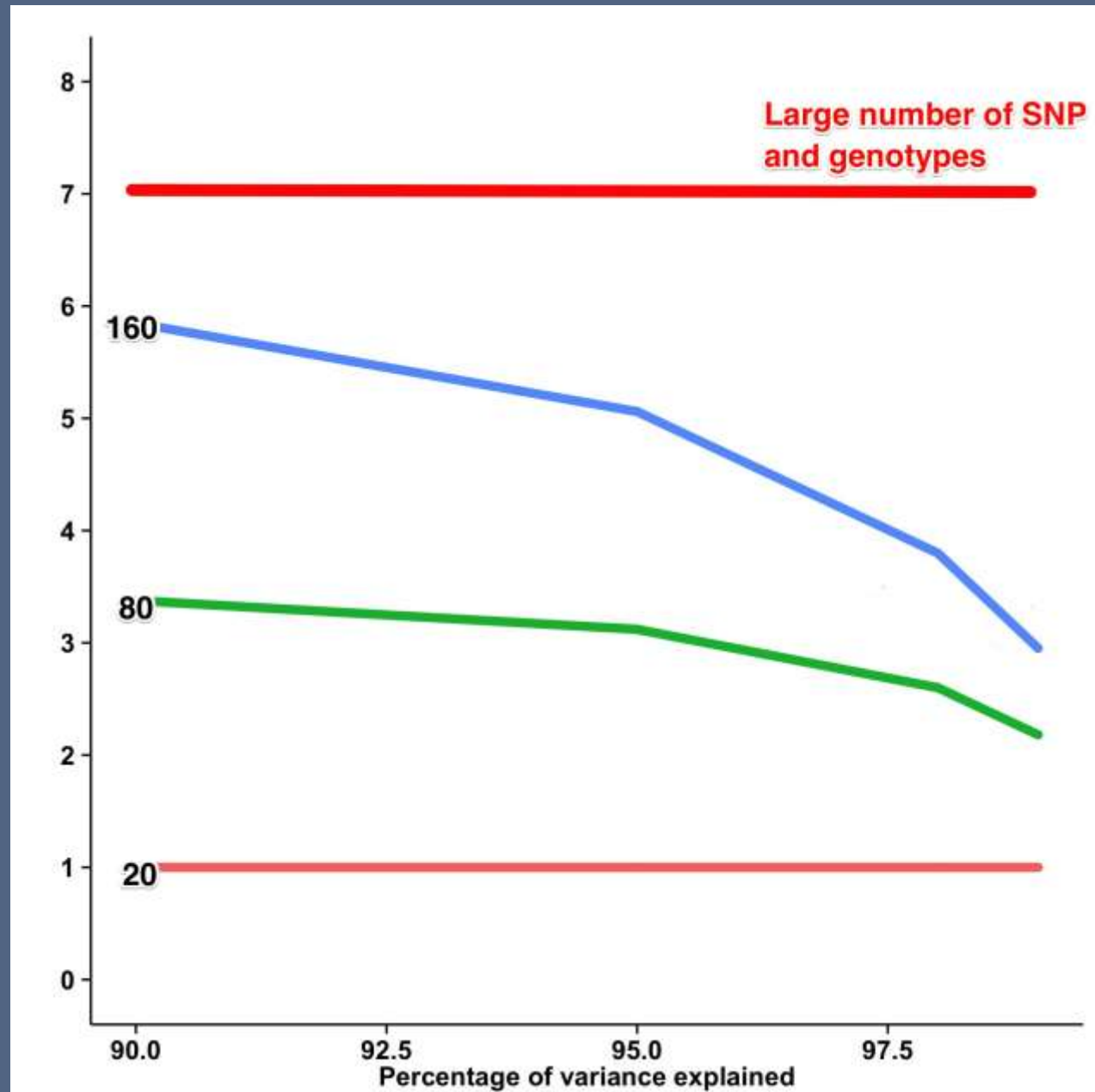
Correlations between GEBVs obtained with APY and regular inverse

Correlations between GEBVs obtained with APY and regular inverse

Accuracies as function of number of core animals

Conclusions

- Dimensionality of GRM function of effective population size
  - Affected by number of SNPs and genotyped animals
- Optimal number of core animals can be derived from eigenvalue analysis of GRM
- Accuracy is maximized when number of core animals corresponding to the number of eigenvalues explaining 98% of GRM variance
- Accuracy using APY is slightly greater than with regular inverse

Acknowledgements

The University of Georgia

Conclusions

- Dimensionality of GRM function of effective population size
  - Affected by number of SNPs and genotyped animals
- Optimal number of core animals can be derived from eigenvalue analysis of GRM
- Accuracy is maximized when number of core animals corresponding to the number of eigenvalues explaining 98% of GRM variance
- Accuracy using APY is slightly greater than with regular inverse

Thank you !!!