Adaptation of BLUPF90 package for genomic computations

Ignacy Misztal, Shogo Tsuruta, University of Georgia, USA
Ignacio Aguilar, INIA, Uruguay
Andres Legarra, INRA, France

Combined pedigree-genomic relationship matrix

$$H = A + \begin{bmatrix} A_{12}A_{22}^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I \\ I \end{bmatrix} [G - A_{22}] \begin{bmatrix} I & I \end{bmatrix} \begin{bmatrix} A_{22}^{-1}A_{21} & 0 \\ 0 & I \end{bmatrix}$$

(Legarra et al., 2009)

G – genomic relationship matrix
1 – ungenotyped animals
2 – genotyped animals

Inverse of combined relationship matrix

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

(Aguilar et al., 2010)

Challenges
- Quality control of genotypes
- Compatibility between **G** and A_{22}
- Performance

Implementation of single-step GBLUP at UGA

- Module genomic in BLUPF90 package (Aguilar et al. 2011)
- Option SNP_File xxx in RENUMF90
  HOL234A4 0120122110501225...
- Lots of options with defaults
- Creation of G^{-1}: minutes for 10k genotypes, hours for 50k genotypes

Programs in BLUPF90 package

Renumbering
RENUMF90

Computing of "genomic" matrices
PreGSF90

BLUP in memory
BLUPF90

BLUP – iteration on data
BLUP90IOD2
CBLUP90IOD

Variance component estimation
REMLF90 AIREMLF90
GIBBS2F90 THRGIBBS1F90

Approximate accuracies
ACCF90

Sample analysis
POSTGIBBSF90

Predictions via SNP
PredGSF90

**GEBV to SNP conversions
GWAS**
PostGSF90

PreGSF90

- Quality control
- Creation of **G** and A_{22}
- Scaling of **G**
- Inversion
- Many options

Matrices

$$G_0 = \frac{(M - P)'D(M - P)}{q}$$

$$G = aG_0 + bA_{22} + gI + d$$

M – matrix of genotypes,
P – matrix of gene frequencies
D – matrix of weights (default: **I**)

Defaults: α=0.95, β=0.05, γ=δ=0

G adjusted to scale of A_{22} - different options
(default: same means of diagonal and off-diagonal elements

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & tG^{-1} - wA_{22}^{-1} \end{bmatrix}$$
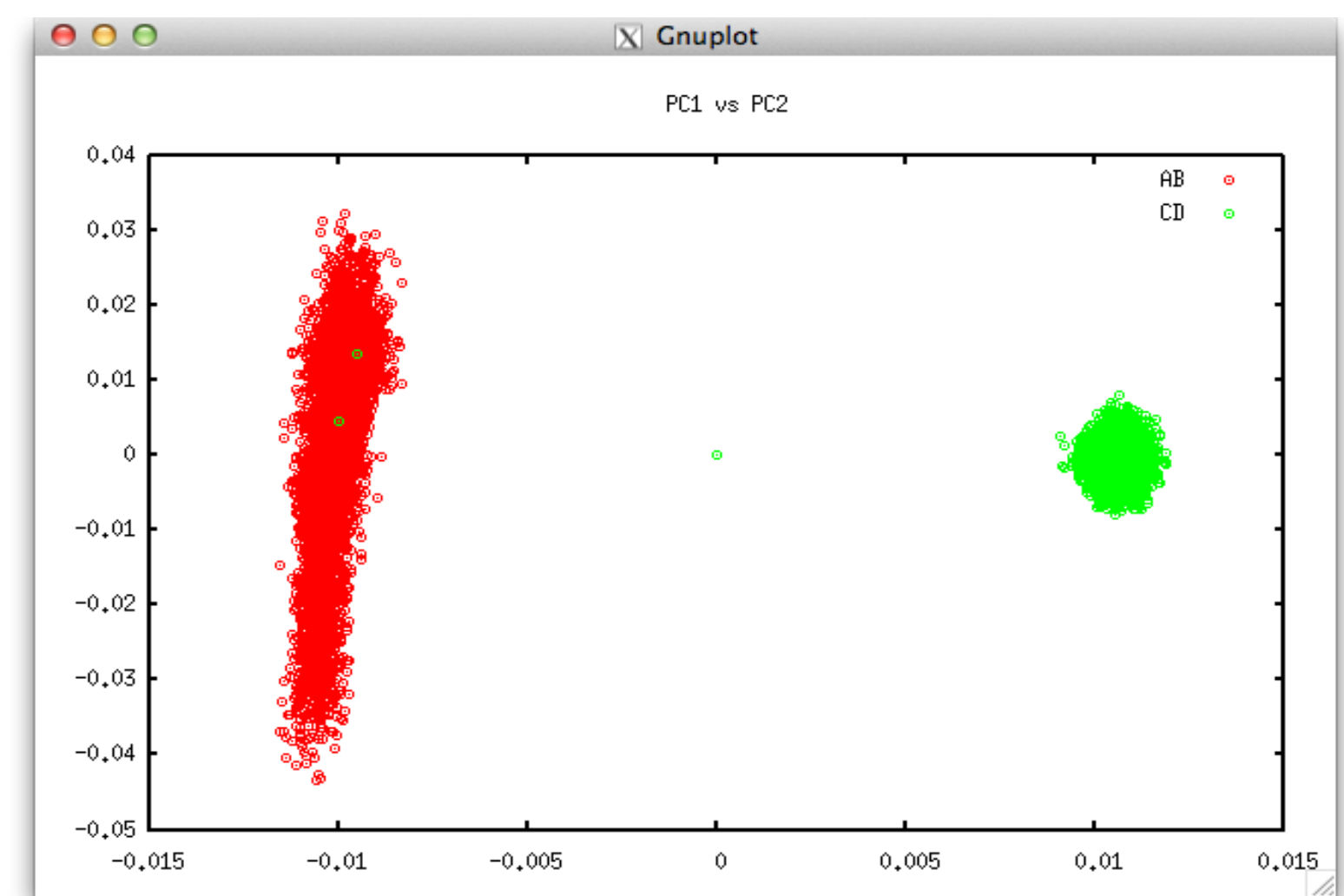
Adjustments for "real life"
(default: τ=ω=1)

Quality control

Calling rate for SNP (default 0.9)
Calling rate for animals (default 0.9)
Inconsistencies
- Mendelian
- Parent-progeny
Corr(g_{ii}, $a_{22,ii}$) > threshold   (default > 0.5)
Graphs
- Diagonals of **G**
- 2 largest eigenvalues

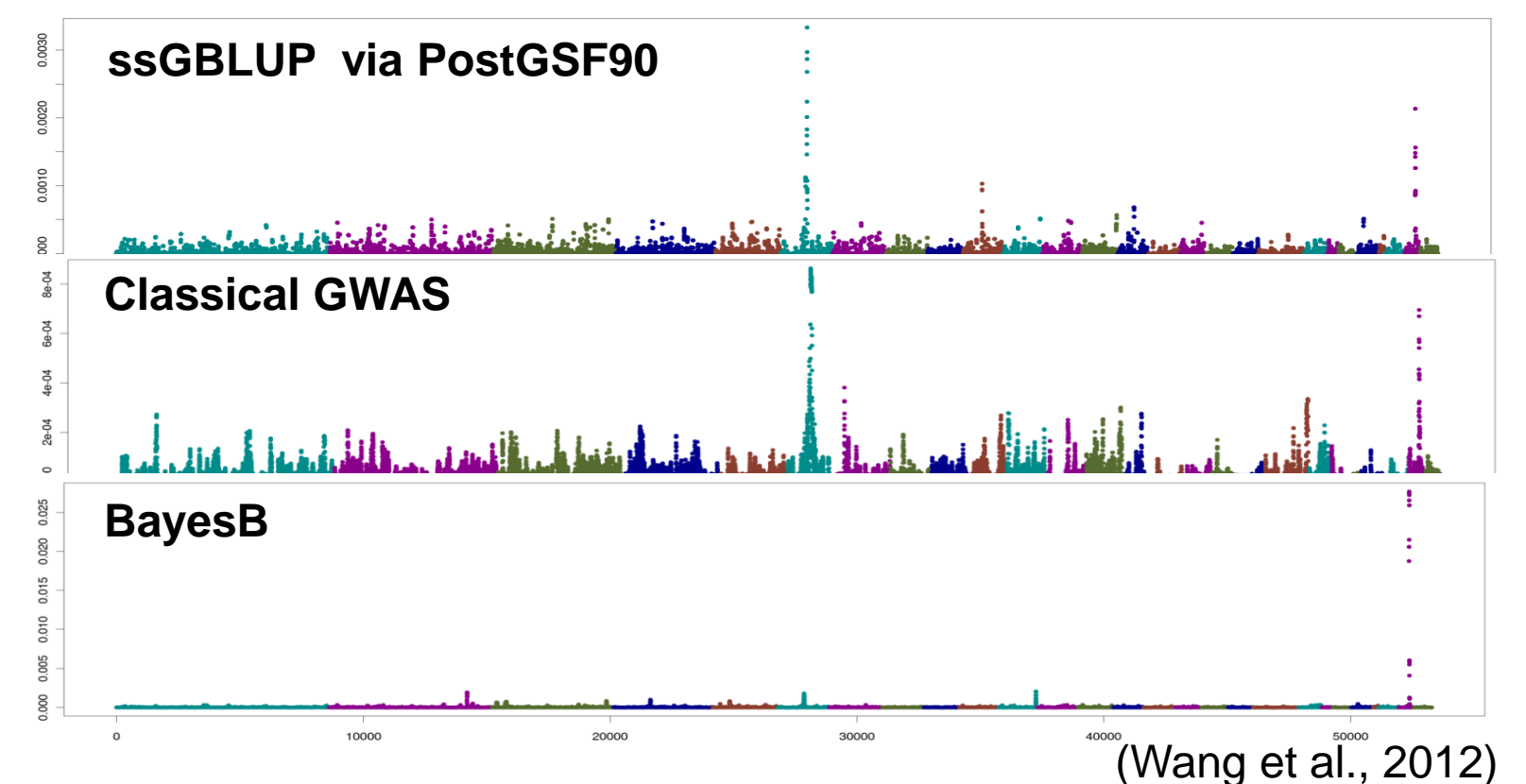## Distribution of diagonal elements of G

(Simeone et al., 2011)
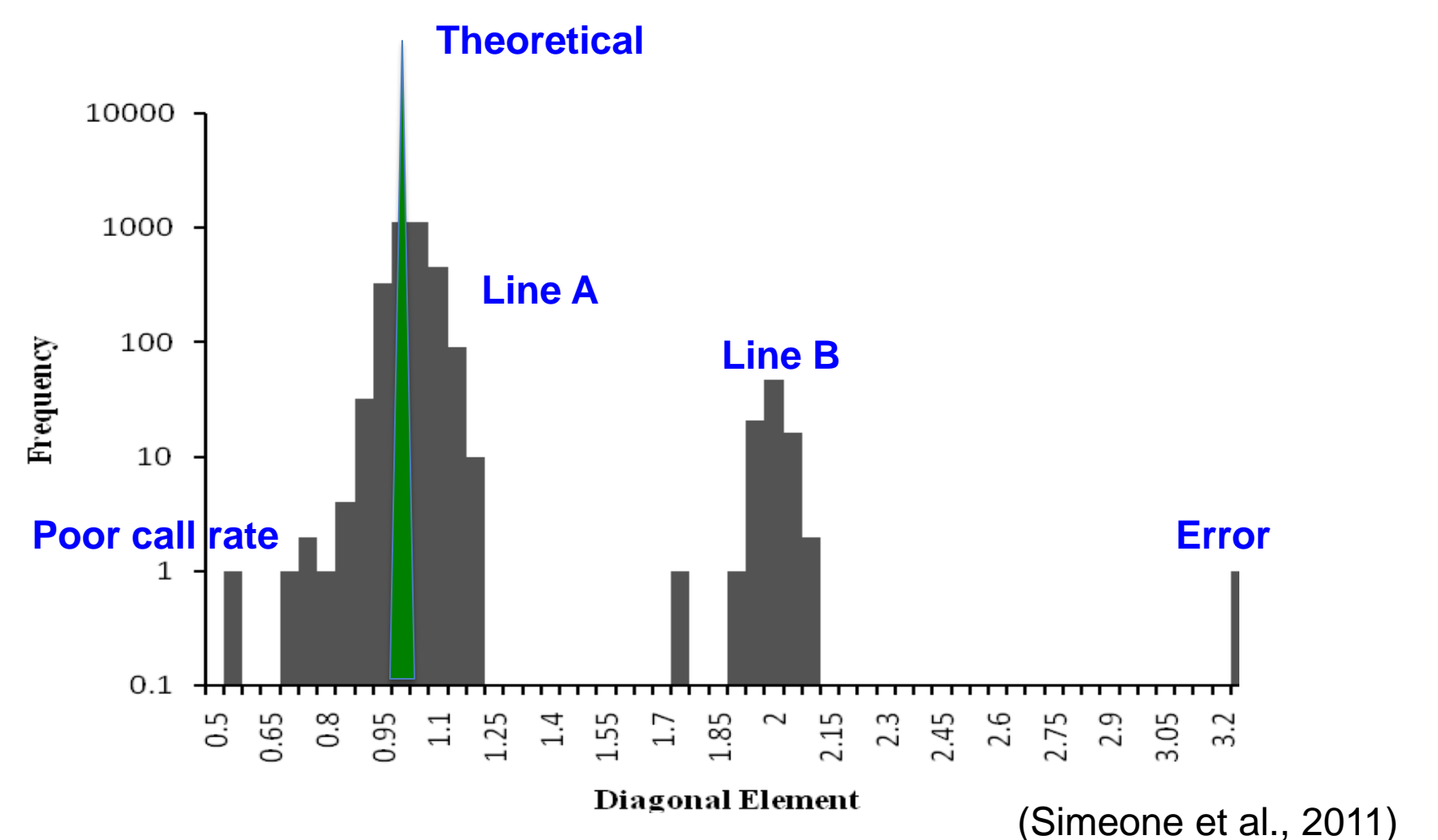
Distribution of 2 largest eigenvalues

GWAS via PostGSF90 - chicken data

ssGBLUP via PostGSF90
Classical GWAS
BayesB

(Wang et al., 2012)

Workshop on genomic selection using single-step methodology
Athens, May 26-June 1
Exercises and documentation on nce.ads.uga.edu

Summary

- BLUPF90 programs updated for genomic analyses
  — Variance component estimation
  — Genetic evaluation
  — GWAS – applicable to complex modes
  — Prediction equations
- Genomic analyzes simple, fast and accurate
- Availability mostly online at nce.ads.uga.edu
  — Large data programs with research contract or for research at UGA
- Development continues
  — Correction of complex genotyping mistakes
  — Adjustments for admixed populations
  — Unlimited number of genotypes