**Lab3 Single-Step GBLUP**

Prepared by D. Lourenco and I. Aguilar

**Animal model for single-step GBLUP**

The data for this lab was created using simulation for a single trait animal model. Simulation was done using QMSim (Sargolzaei, M. and F. S. Schenkel. 2009)

The simulated phenotype were generated using the following model:

*Phenotype = mean + true\_ebv + residual*

**Parameters**:

Mean = 1.0

True variances:

direct genetic = 0.25

residual =0.72

Simulated files are available in the folder:

/home/course/courseadmin/course/lab3

**Description of files**

**data.txt :**

1: Animal ID

2: Sire ID

3: Dam ID

4: Generation number

5: Sex code

6: Number of male progeny

7: Number of female progeny

8: Inbreeding coefficient

9: Homozogosity

10: Phenotype

11: Simulated residual

12: Polygenic effect

13: True EBV

14: Internal EBV from QMSim

15: Mean (column of ones to fit mean effect in BLUPF90)

**ped.txt :**

1: animal ID

2: sire ID

3: dam ID

**snps.txt :**

1: animal ID

2: marker information

1. Copy the full folder into your directory

cp –r /home/course/courseadmin/course/lab3 .

1. From raw data, modify renumf90 parameter file (renlab.par) according to the data file and to fit the following model for genomic evaluation:

*y = mean + animal + e*

1. Run renumf90 program to renumber data, pedigree file, and marker data.
2. Check the renf90.par, renf90.dat, and renaddxx.ped. From the renaddxx.ped file, identify genotyped animals, and check with wiki (<http://nce.ads.uga.edu/wiki/doku.php?id=readme.renumf90>) the content of each column. Print only information from genotyped animals into another file.
3. Estimate variance components considering and ignoring marker information.

From the airemlf90 output find the following statistics: number of genotyped animals, number of SNP markers

1. Run blupf90 without marker information using estimated variance components. Now run blupf90 using genomic information and compare cpu time and solutions. Obs: Check wiki to see how to read an external or pre-computed genomic matrix.
2. Validation on young candidates (individuals from 10th generation with no phenotypes).
3. Remove the phenotypic information from the 10th generation and obtain solutions from a model with marker information and with no marker information.

Hint: if generation information is in column 4, the new data can be created using the AWK Linux tool

awk ‘$4!=10’ renf90.dat > renf90.dat.pred

1. Compare correlations with true breeding values for genetic additive direct effect. Hint: have renumf90 passing to the data a column containing the generation information.
2. Matching G and A22.

Run blupf90 with the truncated data set (renf90.dat.pred) with the following option: OPTION tunedG 0

Compare statistics from G and correlations between G and A22 using OPTION tunedG 0 and the default (OPTION tunedG 2).

Check the average of EBV from both runs and compare them with the average of TEBV.

**OPTIONAL**

Exclude chromosomes with even numbers and repeat validation using truncated data and compare correlation with TEBV for reduced and full genome.