**Assignment Blupf90 - Chapingo**

* Quality control and construction of
* Estimation of Variance Components and Prediction of BV

The data for this lab was simulated using QMSim (Sargolzaei & Schenkel, 2009). A single-trait animal model was simulated, assuming a heritability of 0.40. All the genetic variance was explained by 500 QTL. Animals were genotyped for 45,000 SNP, and the average LD was 0.18. The simulated additive genetic variance was 0.4, and the residual variance was 0.60. The phenotype was generated using the following model:

*Phenotype = sex\_effect + true\_breeding\_value + residual*

**Description of files:**

**data3.txt:**

1: individual ID

2: generation

3: sex

4: phenotype

5: true breeding value (TBV)

**ped3.txt:**

1: individual ID

2: sire ID

3. dam ID

**snp3.2k:**

1: individual ID

2: SNP genotype

**mrkmap.txt**:

1: SNP ID

2: Chromosome

3: position

(Note: in this exercise, SNPs are sorted but this is not needed. The SNP ID must match the order in the SNP file).

1. Modify an existent renumf90 parameter file (or create a new one) according to the data file to fit the following model:

*y = sex + animal + e*

1. Run the renumf90 program to renumber the data.
2. Check the renf90.par, renf90.dat, and renaddxx.ped. From the renaddxx.ped file, identify genotyped animals and check the content of each column with the wiki (<http://nce.ads.uga.edu/wiki/doku.php?id=readme.renumf90>). What is the content of **snp3.2k\_XrefID**?
3. preGSf90 is a stand-alone program that encapsulates the genomic library, including reading pedigree and markers, quality control, and buildup of **G** and **A**22 and their inverses. Run preGSf90, including the option to save a clean SNP file after quality control. In addition, do include “OPTION msg 100” (to have more output on screen). Check the output. Which quality checks for both SNP and animals were done by default? Are there any duplicated genotypes? What is the correlation between **G** and **A**22? Check averages of **G** and **A**22.
4. Run the **blupf90+** program to estimate VC using EM-REML and AI-REML algorithms.
5. Run the **gibssf90+** program to estimate VC by Gibbs Sampling.
6. Run blupf90+ to compute breeding values without SNP information. Now run blupf90+ to compute breeding values using genomic information and compare CPU time and solutions.

Hint: use the following command to provide computing time and save outputs to a log file:

time blupf90+ renf90.par | tee blup1.log