**Lab4 Quality Control and Single-Step GBLUP**

Prepared by I. Aguilar and D. Lourenco

The data for this lab is based on a public pig dataset from PIC (Cleveland et al.

2012 - G3 Journal). Originally, this dataset was filtered for MAF and missing SNP were imputed, however some modifications were introduced to generate commons problems that are found in real datasets.

Files are available in the folder:

/home/course/courseadmin/course/lab4

**Description of files**

**phenotypes\_new.txt:**

1: Animal ID

2: Trait 1

3: Trait 2

4: Trait 3

5: Trait 4

6: Trait 5

7: Mean

**pedigree\_new.txt :**

1: animal ID

2: sire ID

3: dam ID

**genotypes\_new.txt :**

1: animal ID

2: marker information

1. Copy the full folder into your directory

**cp –rl /home/course/courseadmin/course/lab4 .**

Go to <http://nce.ads.uga.edu/wiki/doku.php?id=readme.pregsf90>

and check all options available for preGSf90

1. Run preGSf90

- Check number of SNPs, all statistics related to SNPs, number of duplicated genomic samples (possible clones), and check parent-progeny conflicts

- If there is any parent-progeny conflict, use seekparentf90 to discover the true parents. Check for options and output files from seekparent program in <http://nce.ads.uga.edu/wiki/doku.php?id=readme.seekparentf90>

Use the new pedigree file for the next exercise.

1. Run preGSf90 WITH quality control, removing duplicated animals and saving clean files. Check the number of animals and number of genotypes in the clean files.
2. With clean marker files, the option for no quality control is useful for saving time. Plot principal components and check population structure.
3. With clean marker file, explore options to save G and A22 matrices in text format and with original IDs.