

Assignment 4 - Single-step GBLUP and quality control of genomic data

Download and extract **pregsQC.zip** file from wiki.

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 common problems that are found in datasets.

Files are available in the folder pregsQC

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. Go to <http://nce.ads.uga.edu/wiki/doku.php?id=readme.pregsf90> and check all options available for preGSf90
2. Change parameter file for renumf90 as necessary and 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.

3. 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. Check also files *_removed.
4. With clean marker files, the option for no quality control is useful for saving time. Plot principal components and check population structure.

5. With clean marker file, explore options to save G and A22 matrices in text format and with original IDs.
6. Matching G and A22:
 - Run blupf90 with the with the OPTION tunedG 0
 - Compare statistics from G and correlations between G and A22 using OPTION tunedG 0 and the default (OPTION tunedG 2).
7. Run preGSf90 and save H^{-1} . Run blupf90 using the option to read external files to include H^{-1} in the MME. Compare solutions with blupf90 when H^{-1} is constructed internally and with blupf90 when reading GimA22i.