Lab3 – SNP effects, GWAS, and Indirect Predictions (optional) in single-step using the BLUPF90 family

Prepared by D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica

The data for this lab was simulated using QMSim (Sargolzaei & Schenkel, 2009). A single trait animal model was simulated assuming heritability of 0.4. 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 0.40 and the residual variance was 0.60. The simulated phenotype was generated using the following model:

```
Phenotype = sex_effect + true_breeding_value + residual
```

Files are available in the folder lab4.

Description of files

data3.txt:
1: animal ID
2: generation
3: sex
4: phenotype
5: true breeding value (TBV)

snp3.2k:
1: animal ID
2: SNP genotype

ped3.txt:
1: animal ID
2: sire ID
3. dam ID

mrkmap.txt:
1: SNP ID
2: Chromosome
3: position
(Note: in this exercise, SNPs are sorted but this is not needed. The SNP ID has to match the order in the SNP file).

1. Run **`renumf90`** using `renum.par` parameter file to renumber the data.

2. Calculating SNP effects in ssGBLUP: given that SNP effects are calculated based on GEBV, run **`blupf90+`** to get GEBV. Before that, check below the options you will need to include in the parameter file (renf90.par), so **`blupf90+`** can provide all the files needed for the calculation of SNP effects and p-values (in exercise 3). Assume that the genotype quality control was already done.

   Options to include in renf90.par:
   OPTION saveGInverse
   OPTION saveA22Inverse
   OPTION snp_p_value
   OPTION no_quality_control
   OPTION use_yams

   Hint: you can save renf90.par with another name and modify the new parameter file.
   Example: `cp renf90.par blup.par`

3. Run `postGSf90` to compute p-values. Before that, check below the options you will need to include in the parameter file. Check the output files and the content of each column. For additional information, check

http://nce.ads.uga.edu/wiki/doku.php?id=readme.pregsf90. The the -log10(p-value) are in the file `chrsnp_pval`. The file named `Pft1e2.R` has an R script to obtain Manhattan plots. Run this script and look at `Pft1e2_manplot.pdf`. The typical threshold for detection uses the Bonferroni correction, which is 0.05/number_of_SNP. Are there any significant SNP?

Options to include in renf90.par:
OPTION readGInverse
OPTION readA22Inverse
OPTION snp_p_value
OPTION no_quality_control
OPTION use_yams

Hint: you can save renf90.par with another name and modify the new parameter file. Example: `cp renf90.par post.par`

4. Indirect predictions for young individuals: **`postGSf90`** creates a file snp_pred with information about the random effects (number of traits + correlated effects), the gene frequencies and the solutions of SNP effects. This is the file used by **`predf90`** to provide indirect predictions for young genotyped animals as **Z**a, where **Z** is a matrix of SNP content and a is a vector of SNP effects. Run **`predf90`** to get indirect predictions for young individuals, that were not included in the **`blupf90+`** and **`postGSf90`** runs. Genotypes for young animals are in a file called `new_animals`.

5. OPTIONAL: Compute reliability of indirect predictions obtained in exercise 4.