Lab2 – SNP effects, Indirect Predictions, GWAS, and Metafounders 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 day2. Copy the entire folder using the following command:
```
curl http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=lab2_ufv.zip -o
lab2.zip
```

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` program 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 solutions. Before that, check the options you will need to include in the parameter file, so `blupf90+` can provide all the files needed for the calculation of SNP effects and p-values (in exercise 3). Assume that the quality control was already done (add `OPTION no_quality_cotrol`).

3. Add an option to read a map file (mrkmap.txt) and run `postGSf90` to compute p-values (the appropriate option can be found in the "`preGSf90`" part of the wiki). Check the output files and the content of each column (http://nce.ads.uga.edu/wiki/doku.php?id=readme.pregsf90). The p-values (actually, -log10(p-value) are in file `chrsnp_pval`. Plot Manhattan plots based on p-values and variance explained by SNP. The typical threshold for detection uses the Bonferroni corrrection, which is 0.05/number_of_markers. Are there any significant markers?

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 Za, 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`.