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`.

OPTIONAL

5. Using metafounders to set the base populations. Files are available in the folder day3_metafounders: The data for this lab was simulated by D. Lourenco (Lourenco et al., 2016) using QMSim (Sargolzaei & Schenkel, 2009). A single trait animal model was simulated assuming heritability of 0.30. All the genetic variance was explained by 400 QTL. Two lines (1 and 2) under 9 generations of selection were simulated. Pure and F1 (12) progeny were generated in generation 10. Animals were genotyped for 40,000 SNP. The simulated additive genetic variance was 0.3 and the residual variance was 0.70. The phenotype was generated using the following model:

   Phenotype = general_mean + true_breeding_value + residual

Files are available in the website. Use curl to download it to your Linux or Mac device:

```
curl
http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=lab2mf_ufv.zip -o
lab2mf.zip
```

Description of files:

newdata.txt:
1: animal ID
2: sire ID
3: dam ID
4: generation
5: phenotype
6: true breeding value (TBV)
7: line code
8: mean incidence

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

snp_file.txt:
1: animal ID
2: SNP genotype

a) No metafounders: Run renumf90 and ssGBLUP using blupf90+.

b) With metafounders: In a separate folder, replace missing parents with -1 in line 1 and -2 in line 2 (MF coding). Run renumf90 with the modified pedigree that contains metafounders, estimate Gamma using gammaf90, and run ssGBLUP using blupf90test. Do not forget to replace to rename gamma.txt and change the random type in the parameter file (check the slides).

c) Compare solutions from a) and b). One way to compare the methods is to perform a validation for young individuals (e.g., accuracy, level bias – b_0, and dispersion bias – b_1). For that, remove phenotypes for individuals in generation 10. Run ssGBLUP with and without metafounders. Correlate solutions with TBV (column 6 in data3mf.txt). Do not forget that solutions are with renumbered IDs (different in ssGBLUP and ssGBLUP with MF) and TBV are with original IDs. Use `OPTION origID` to get solutions with original IDs