Lab 2 – Simulation of genomic data

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

We will use the software QMSim (Sargolzaei & Schenkel, 2009, Bioinformatics 25:680-681). The software and its manual can be found in <http://www.aps.uoguelph.ca/~msargol/qmsim/>

The files for these exercises are in:/home/course/lab2/QMsim/

For running QMSim, use: echo ex1.par | ./QMsim16

**Exercise 1**

1. Run the QMSim program. An example of the parameter file is *ex1.par*. Note that historical population was generated by mutation and drift over 100 generations (t) with an effective population size of 100 (t = 1 to 95) and gradually expanded to 3,000 offspring (t = 100).
2. Now change *ex1.par* and simulate a base population of 200 males and 2,600 females, 5 generations of selection for a trait (i.e. live weight) with a phenotypic variance of 1.
   * 1. How many SNPs did you simulate?
     2. How many QTLs might potentially affect the phenotype?
     3. How many animals do you have in the recent population?
     4. Answer this question assuming a litter size equal to 12.
     5. Which is the mean of the TBVs after 5 generations?
     6. Use selection and culling based on EBVs, does the mean of the TBV change?
     7. Include positive assortative mating.
     8. Which is the value of the polygenic variance?

1. Look at the simulated file *p1\_data\_001.txt*, you will have the following columns:

1: animal id

2: sire id

3: dam id

4: sex

5: generation

6: number of males’ progenies

7: number of females’ progenies

8: inbreeding

9: homozygosity

10: phenotype

11: simulated residual (e)

12: individual true breeding value for polygene

13: individual true breeding value for direct effect (qtl)

14: EBV from QMSim internal BLUP

1. Check the file p1\_freq\_mrk\_001.txt. What does it show?
2. Take a look at script *edit\_data.sh.* Run it by typing bash edit\_data.sh

This script creates the pedigree, the phenotype, and the genotype files for BLUPf90 from the QMSim simulated data. Which are the pedigree file, the phenotype file, and the marker file?

Note that edit\_data.sh uses the directory r\_ex01b/

1. Using awk, check the number of animals in each file:

wc -l ped.txt

Do the same for the phenotype file

1. Before continuing the analysis, it’s important to check the “quality” of the files for some typical errors in the file. For example, are there duplicated animals in the pedigree? Check it using

awk '{print $1}' ped.txt | sort +0 -1 | uniq -c | awk '$1>1'

1. Which number of progeny of each sire?

awk '{print $2}' ped.txt | sort +0 -1 | uniq -c > sire.prog

1. How many genotyped animals are in the SNP file?

wc -l snp.txt

1. How many SNP genotyped?

awk '{print length($2)}' snp.txt

1. Does everyone have the same number of loci in the marker file?

awk '{print length($2)}' snp.txt | sort -u

1. Let’s extract the first 200 individuals *while keeping the format.* This is a one-line command, beware of simple and double quotes!

awk 'NR <= 200' snp.txt > anim200.temp

OR

awk 'NR <= 200 {printf("%10s %1s%" length($2) "s\n", $1, " ", $2) }' snp.txt > anim200b.temp

Note that the format is defined by "%10s%1s%" length($2) "s\n" which means “10 positions, 1 position (for the space in “ “ later), as many positions as SNPs we have for each individual (in %" length($2) "s ), and the line return in \n.

1. Extract SNP number 50 for all animals we’ll come back to this later:

awk ' {printf("%10s%1s%1s\n", $1, " ", substr($2,50,1)) }' snp.txt > snp50.temp

**Optional Exercise**

1. Run the QMSim program with the parameter file: **'ex02.par'**. Note that the population structure involves an F2 design produced from inbred lines with divergent phenotypes.
   1. How many SNPs do you simulate?
   2. How many animals do you have in the cross between line 1 and line 2 after 2 generations?
   3. Which are the values of inbreeding in lines 1 and 2 ?

1. Edit phenotypes, pedigree, and genotypes. Be aware you need tom combine data from lines 1, 2, and F1.
2. Write a parameter file to simulate a backcross between the F1 and the line 1.