

DATA SIMULATION (including genomics)

Zulma Vitezica,
INRA, UMR 1289, 31326 Castanet-Tolosan, France
zulma.vitezica@ensat.fr

The objective of these exercises is to simulate genomic data sets using software QMSim (Sargolzaei & Schenkel, 2009, Bioinformatics 25:680-681).

This software can be found in <http://www.aps.uoguelph.ca/~msargol/qmsim/> where the manual can be found.

Copy directory /home/ads-guest3/zulma/

Go to QMSim_UGA. Take a look at files.

For running QMSim :

```
$ echo ex.prm | ./QMSim
```

Exercise 1

1. Run the QMSim program. An example of the parameter file is 'ex01.prm'. 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 'ex01.prm' 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 ?
3. Using the script 'after_qmsim.sh', create the pedigree, the phenotype, and the genotype files for BLUPf90 from the QMSim simulated data.
Note that `$./after_qmsim.sh` uses the directory `r_ex01/`
4. Using AWK, check the number of animals in each file:

```
$ awk 'END {print NR}' pedigree.txt
```
5. Using 'freq_allele.awk', estimate the allele frequency of markers.

```
./freq_allele.awk r_ex01/p1_mrk_001.txt > frequencies
```

Exercise 2

1. Run the QMSim program with the parameter file: 'ex02.prm'. Note the population structure involves 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 ?
2. Write a parameter file for simulating a backcross between F1 and line 1