Laboratory of Recombination and Segregation analysis
Institute of Cytology and Genetics SD RAS
Novosibirsk, 630090, Russia

Mga-simulate

tutorial

Please address correspondance to: Yurii Aulchenko, Institute of Cytotlogy and Genetics, Novosibirsk, 630090, Russia. e-mail: yurii@bionet.nsc.ru

1 PREAPRING INPUT FILES:

You have to prepare 3 files to run MGA-SIMULATE. These are file containing pedigree structure, genetic map and allelic frequencies data and mode of inheritance for QT.

1.1 PEDIGREE DATA FILE:

The first line of the file could be any, though it is recommended to keep a description of fields here. id — an unique identification number for pedigree member, fa — father's is, mo — mother's id, pop — a population where individual comes from. For any non-founder should be 0 while for founders is 1 or 2. sex — geneder. 1 for male, 2 for female.

The second line is a number of individuals considered.

The third line is record for the 1st pedigree member.

The example pedigree data file could be found in example/twopop/ directory of the distribution, named 'testped.dat':

```
id fa mo pop sex
  8
1  0  0  1  1
2  0  0  1  1
10  0  0  2  2
11  0  0  2  2
100  1  10  0  1
101  2  11  0  2
200  100  101  0  1
201  100  101  0  2
```

Sires 1 and 2 come from breed 1, dams 10 and 11 come from breed 2. They are intercrossed (1 and 10; 2 and 11) giving 'F-1' consisting of sire 100 and dam 101. These are intercrossed again, giving 'F-2' consisting of two sibs 200 and 201.

1.2 GENETIC MAP/ALLELIC FREQUENCIES DATA FILE:

lines 1-8 — a comment

line 9 — fortran format for reading "number of loci" variable

line 10 — fortran format for reading periodic data line 1

line 11 — fortran format for reading periodic data line 2

line 12 — a comment

line 13 — number of loci

line 14 and 15 — record for locus 1

...

Record for a locus:

```
locus name         / --- N ofalleles
  |      position |    / --- freq. of allele 1 in pop. 1
  |      |        |    |   / --- freq. of allele 2 in pop. 1
d1mar17  33.2     4    0.01 0.00 0.98 --- freq. of allele 3 in pop. 1
                       0.50 0.44 0.02 --- freq. of allele 3 in pop. 2
                         |
```

```
                          \ --- freq. of allele 1 in pop. 2
```

Example file named 'chrom.cnt' could be found in example/twopop/ directory of the distribution of MGA-SIMULATE:

```
# locus name         / --- N ofalleles
#   |      position |    / --- freq. of allele 1 in pop. 1
#   |      |        |    |   / --- freq. of allele 2 in pop. 1
# d1mar17  33.2     4   0.01 0.0 0.975 --- freq. of allele 3 in pop. 1
#                       0.5 0.444 0.02 --- freq. of allele 3 in pop. 2
#                       |
#                        \ --- freq. of allele 1 in pop. 2
# formats:
(i5)
(a12,i3,f8.2,100f5.2)
(23x,100f5.2)
#number of loci:
  11
marker1      10    0.00 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
                        0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
marker2      10   10.00 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
                        0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
major-qtl     2   17.00 1.00
                        0.00
marker3      10   21.00 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
                        0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
marker4       2   33.00 0.80
                        0.30
marker5      10   50.00 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
                        0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
marker6      10   80.00 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
                        0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
marker7      10  100.00 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
                        0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10 0.10
minor-qtl1    2  500.00 0.30
                        0.70
minor-qtl2    2  900.00 0.70
                        0.30
minor-qtl3    2 1300.00 0.30
                        0.70
```

Here 11 loci are described. 8 of them are in the same sintheny group, while latter 3 are unlinked. Note, that if distance from one locus to another is more then 300 cM these loci are treated as completely unlinked (i.e. recombination frequency set to 0.5). Of 11 loci, 4 have substring 'qtl' in the name and therefore they are treated as QTLs. Other are treated as marker loci. Markers 1, 2, 3, 5, 6 and 7 have 10 alleles each and are fully informative in both populations. Marker 4 is biallelic; the frequency of allele '1' is 0.8 in population 1 and 0.3 in population 2. For major-qtl, in population 1, the only allele presented is 1, while in the population 2 — the allele 2. Minor unlinked QTLs are polymorphic in both breeds.

1.3 MODE OF INHERITANCE DATA FILE:

Lines 1-8 — a comment

Line 9 — number of QT loci [(i5) fortran fixed width format]

Line 10 — additive effect of 1st locus.

Line 11 — additive effect of 2nd locus...

...

Last line — square root of enviromental variance

Effect of some locus is written in (a12,i5,4f9.3) format fields are:

```
 / --- locus name
|               / --- number if alleles
minor-qtl1      2   -0.100    0.000    0.000    0.100
                      |        |        |         \ --- mean for genotype 2 2
                      |        |         \ --- mean for genotype 2 1
                      |         \ --- mean for genotype 1 2
                       \ --- mean for genotype 1 1
```

The example mode of inheritance data file 'qt-model.dat' could be found in example/twopop/ directory of the distribution:

```
#  / --- locus name
# |               / --- number if alleles
# minor-qtl1      2   -0.100    0.000    0.000    0.100
#                       |        |        |         \ --- mean for genotype 2 2
#                       |        |         \ --- mean for genotype 2 1
#                       |         \ --- mean for genotype 1 2
#                        \ --- mean for genotype 1 1
# number of loci:
    4
major-qtl       2   -1.370    0.000    0.000    1.370
minor-qtl1      2   -0.240    0.000    0.000    0.240
minor-qtl2      2   -0.240    0.000    0.000    0.240
minor-qtl3      2   -0.240    0.000    0.000    0.240
sqrt(env.var.)        0.600
```

Here 4 QTLs are listed. An effect of a genotype at individual locus is set directly (e.g. major-qtl: genotype 11 — effect -1.37, 12 and 21 — zero effect, 22 — +1.37). Effects from individual genotypes are added to each other to produce overall effect of genotype. The effect of environment is then generated by adding an random number from normal with mean zero and SD 0.6.

2 RUNNING MGA-SIMULATE:

Given 3 data files described above, you could be able to run MGA-SIMULATE using a bash script named mga-simulate. Copy data files from examples/twopop to the root of the distribution and run mga-simulate.

3 OUTPUT:

These are 4 files containing genetic map, pedigree structure, marker phenotypes and quantitative phenotypes. The output files are ready to be analysed by SOLAR package of programs. See SOLAR documentation for more details.

PEDIGREE FILE: first line — fields in pedigree, separated by comma: id,fa,mo,sex. id — personal identification number (id). Unique intger. fa — father's id mo — mather's id sex — gender. 1 if male, 2 if female.

MAP: first line: "10" (chromosome number). Then each line consist of

(marker-name) (position from the beginning of chromosome)

MARKER PHENOTYPES: first line: id,locus-name-1, locus-name-2, ... second line: marker genotypes for 1st p[Cerson. Genotypes on different loci are separated by comma, while alleles of the same genotype are separated by slash.

QT PHENOTYPES: Essentialy the same as MARKER PHENOTYPES file format.

4 MORE EXTENSIVE EXAMPLE:

Here is example of how data could be simulated and then analysed using SOLAR:

1. Copy data files from examples/large to root directory of distribution
2. Run mga-simulate script
3. Copy *.sol output files to some directory (e.g. 1/)
4. Copy examples/large/example.tcl to the same directory
5. Run SOLAR in the directory
6. Type 'example' and press return