QTLMap 0.8

User's guide

08/11/10

Table des matières

1. Introduction

QTLMap is a software dedicated to the detection of QTL from **experimental designs** in **outbred population**. QTLMap software is developed at INRA (French National Institute for Agronomical Research). The statistical techniques used are linkage analysis (LA) and linkage disequilibrium linkage analysis (LDLA) using **interval mapping**. Different versions of the LA are proposed from a quasi Maximum Likelihood approach to a fully linear (regression) model. The LDLA is a regression approach (Legarra and Fernando, 2009). The population may be sets of **half-sib families** or **mixture of full- and half- sib families**. The computations of **Phase and Transmission probabilities** are optimized to be rapid and as exact as possible. QTLMap is able to deal with large numbers of markers (SNP) and traits (eQTL).

The aim of QTLMap developers is to propose various genetic models depending on 1) the number of QTL alleles segregating (biallelic in crosses between monomorphic breeds, biallelic without hypothesis on the origin, multiallelic, haplotype identity), 2) the number of QTL segregating (one, two linked, several unlinked), 3) the number of traits under the QTL influence. The trait determinism may vary depending on 1) the trait distribution (gaussian trait, survival trait or threshold distribution), 2) the interactions between the QTL and fixed effects or other loci, 3) the residual variance structure (homo- or heteroskedasticity for half-sib families). Due to differences with the asymptotical conditions from the chi2 theory, the test statistic significance are evaluated either through numerical approximations, or through empirical calculations obtained from permutations or simulations under the null hypothesis.

QTLmap is written in fortran and either uses the NAG or SLATEC libraries.

Up to now, the following functionnalities have been implemented :

- QTL detection in half-sib families or mixture of full- and half-sib families
- One or two linked QTL segregating in the population
- Single trait or multiple trait analyses
- Nuisance parameters (e.g. sex, batch, weight...) and their interactions with QTL can be included in the analysis
- Gaussian, *discrete or survival (Cox model) data*
- Familial heterogeneity of variances (heteroscedasticity)
- Can handle eQTL analyses
- Computation of transmission and phase probabilities adapted to high throughput genotyping (SNP)
- Empirical thresholds are estimated using simulations under the null hypothesis or *permutations* of trait values
- Computation of power and accuracy of your design or any simulated design

2. Contributors

Pascale Le Roy, UMR GARen, Rennes, France
Jean-Michel Elsen, SAGA, INRA, Toulouse, France

Hélène Gilbert, GABI, INRA, Jouy-en-Josas, France
Carole Moreno, SAGA, INRA, Toulouse, France
Andres Legarra,SAGA, INRA, Toulouse, France
Olivier Filangi, UMR GARen, Rennes

3. Support

Subsribe and post any message/question to the qtlmap-users list :

mailto:qtlmap-users@listes.inra.fr

4. Setting up QTLMap

4.1. Runtime environment with GNU software component

Pre-requisites

- The GNU compiler collection : gfortran 4.4, gcc
- Cmake 2.6.4 ,cross-platform, open-source build system.

Compilation

```
>cd ${QTLMAP_DIR}
>mkdir build
>cd build
>cmake -DCMAKE_BUILD_TYPE=Release ..
>cmake -DCMAKE_Fortran_COMPILER=gfortran ..
>make
```

The binary qtlmap is created in the `${QTLMAP_DIR}/build/src` directory.

To install the qtlmap binary in the bin directory `${QTLMAP_DIR}/bin` :

```
>make install
```

OpenMP support

supports multi-platform shared-memory parallel programming

```
To define the number of threads :
>export OMP_NUM_THREADS=8
```

5. Input files

To carry on an analysis, you need
4 data files :

Marker map
Pedigree

Marker genotypes
Quantitative traits values

1 file describing the performance :

Model

5.1. Dataset format

Pedigree file

The file contains pedigree information for the 2 last generations of a design which comprises 3 generations, i.e. parents and progeny. It must not contain the grand parental pedigree information. Each line is made of an alphanumeric ID triplet (individual, sire, dam). A fourth information gives the generation number : « 1 » for the parental generation ; « 2 » for the progeny generation. An animal missing one or both parents ID has not to be included in the file. The missing value code (given in the parameterization of the analyses, see 6.2) cannot be used in the pedigree file. The file must be sorted by generation, sire ID and dam ID

```
922961 911287 902206 1
944547 924758 911714 1
944985 924758 912892 1
961924 922961 944547 2
961925 922961 944547 2
961926 922961 944547 2
963187 922961 944985 2
963188 922961 944985 2
963189 922961 944985 2
963190 922961 944985 2
```

Texte 1: Example of a pedigree file

means that the pedigree includes 7 progeny born from 1 sire and 2 dams. Sire 922961 is the son of sire 911287 and dam 902206 etc...

constraint

The file must be sorted by generation, sire ID and dam ID

Marker map file

This file gives the locations of the markers on the chromosome(s). Each line corresponds to a single marker, and gives (order to be followed) :

- marker name (alphanumerique) ;
- name of the chromosome carrying the marker (alphanumerique) ;
- marker position of the marker on the average map (in Morgan) ;
- marker position of the marker on the male map (in Morgan) ;
- marker position of the marker on the female map (in Morgan) ;

- inclusion key (=1 if the marker has to be included in the analysis, 0 if not)

```
SW552   1  0.08 0.05 0.09 1
SW64    1  0.24 0.24 0.25 0
CGA     1  0.49 0.45 0.55 1
S0088   15 0.50 0.37 0.59 1
SWR1002 15 0.58 0.49 0.63 1
```

Texte 2: Example of a marker map file

means that marker SW552 is on chromosome 1, at position 0.08 on the average map, 0.05 on male map and 0.09 on the female map, and will be included in the analysis of chromosome 1, etc...

The marker genotypes file

This file contains the animals phenotypes at the markers. The first line gives the marker names, the markers must belong to the marker map file. For each animal, a line gives its ID (as decribed in the pedigree file) followed by the markers phenotypes, ranked following in the first line order . Each phenotype is made of 2 alleles, unordered. When an animal has no phenotype for a marker, both alleles must be given the missing value code as given in the parametrisation of the analysis (see 6.2).

```
mark1 mark2 mark3
911714 2 5 3 1 4 13
912892 8 2 6 5 4 13
924758 2 5 6 1 12 5
922961 2 2 3 1 12 13
944547 2 5 1 3 12 4
944985 2 8 1 5 12 4
961924 2 5 0 0 13 4
961925 * * 0 0 13 4
961926 2 5 0 0 0 0
963187 2 8 0 0 12 4
963188 2 2 3 1 13 4
963189 2 2 1 1 12 4
963190 2 8 1 5 12 4
```

Texte 3: Example of a marker genotypes file

means that, amongst the 5 grand parents, 3 were genotyped (911714, 912892 et 924758). For instance, grand dam 911714 is heterozygous « 2 5 » at marker SW552, the individual 961925 has no genotype at marker mark1 …etc.

Quantitative trait values file

This file gives the phenotypes of the traits to be analysed.
The progeny performances only are considered in the analysis and must be given in the file.

For each animal, its ID (identical to the ID given in the pedigree file) is followed by information about nuisance effects (fixed effect levels, covariable value) and then by three information for each trait : the performance, an 0/1 variable IP which indicates if (IP=1) or not (IP=0) the trait was measured for this animal and must be included in the analysis, and 0/1 variable (IC) which indicates if (IC=0) it was censored or not (IC=1), this IC information being needed for survival analysis (by default IC=1).

```
961924 1 10.43   7.8   1 1   77.6  1 1
961925 2 5.34    0.0   0 1   90.   1 1
961926 1 12.34   11.3  1 1   103.  1 1
963187 2 9.45    12.7  1 1   98.   1 1
963188 1 11.10   13.5  1 1   0.0   0 1
963189 2 10.11   10.   1 1   94.8  1 1
963190 1 9.98    14.2  1 1   98.3  1 1
```

Texte 4: Example of a quantitative trait values file

This file describes 2 traits. For progeny 961924, the recorded information are : sexe 1 (fixed effect), body weight 10.43 (covariable), backfat thickness 7.8mm (trait 1) and fatening period of 77.6 days (trait 2) etc...

Expression quantitative trait values file

This file gives the phenotypes expression traits to be analysed.

The header line is the list of animals phenotyped. The following line are the fixed effects, covariates and finally the phenotype.

The format of the nuisances effects and phenotype line is :

```
<IDANIMAL> <VALUE_ANIMAL1><VALUE_ANIMAL2>...
```

For missing data, insert a character string which is not interpretable as a numeric(*e.g.* n/a).

```
4112 4114 6380 6553 4142 4120 6388 6537 6548 6536
sexe 1 1 1 1 1 1 1 1 1
cov1 0.3 0.4 0.3 0.5 0.5 0.6 0.3 0.2 0.6 0.9
1 0.0184170490684831 -0.143560443113406 -0.118137020630747 -0.06666521254513
0.0642879011796014 -0.255460347400393 -0.189477060869665 -0.25462868498086
-0.00530461929594204 -0.254172485884001
2 -0.127806826817031 -0.163876647400758 0.0184043832497863
-0.296146098377366 -0.112715209230912 -0.0684375510992924 -0.180990247175303
-0.182892021501701 -0.063104337861525 -0.0334596435779586
3 -0.259405679027549 -0.365184085691961         n/a       -0.104403755609133
-0.154653751085067 -0.213511162284327 -0.190633612968503 -0.344837877148359
0.154406432653772 0.328663903209088
4 0.151093991655429 0.10964888434473 0.15832262904679 0.284848089326391
0.0808434990010986 0.306550168430082 0.00906573426897184 0.10731093171816
0.390146267506709 0.0562950676047775
```

Texte 5: Example of a expression quantitative trait values file

In this previous example, the animal 6380 have a missing data for the gene 3.

6. Descriprtion of the dataset

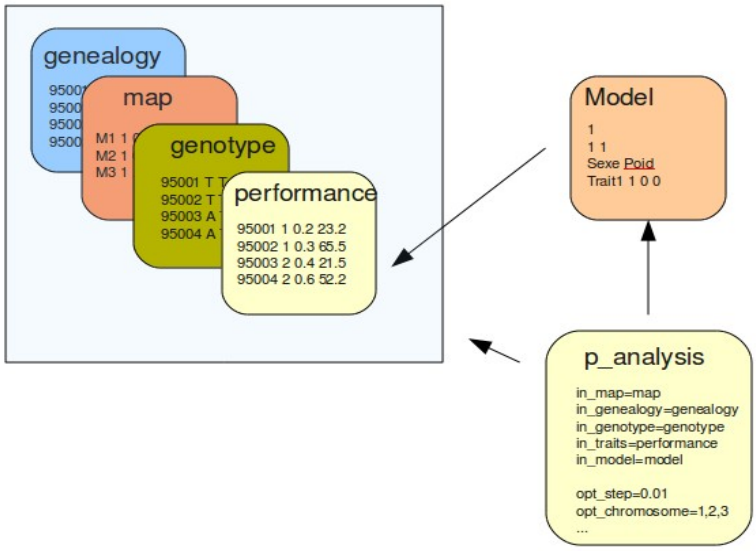


Illustration 1: set of needed files for the analysis

6.1. The model file

In this file the model analysis of each trait is described

- Number of traits
- Number of fixed effetcs (nf), Number of covariables (nc)
- Names of the fixed effects and covariables
- Name of the 1st trait, nature of trait ('r' for real value, 'i' discrete ordered data and 'c' categorial data) model for this trait symbolized by 0/1 indicators for each fixed effects (nf first indicators), each covariables (nc following) and each interactions beetwen the QTL and the fixed effects (nf last indicators). A fixed effect, covariable or interaction will be included in the analysis if its indicator is 1, will not be if it is 0.
- Name of the 2nd trait,...
- .....
- (Optional) The heritability h2, phenotypics and genotype correlation between traits (classical traits)
- A filter list of traits be kept in the analysis. This line is optional. If absent all traits described above will be analysed.

```
3              ! Number of traits
1 1            ! Number of  fixed effects and covariables
sexe poids     ! Names of the fixed effects and covariables
malade r 1 1 0 ! 1st trait, (nature : real value) model
malcor r 0 0 1 ! 2nd trait,(nature : real value) model
third  r 0 0 0 ! 3nd trait,(nature : real value) model

correlation_matrix

0.35 0.28 0.29
0.20 0.32 0.28
0.20 0.20 0.33
```

Texte 6: Example 1 of a model file

This model file describes the performance file where one fixed effect, one covariate and three performances are referenced for each animals.

The model for each performance is :

malade = μ + sexe + β . poids + ε

malcor =μ + QTL × sexe + ε

third =μ + ε

The correlation matrix are given according the following rules :

- The heritability (h2) are defined in the diagonal
- Phenotype correlations : the upper triangle matrix
- Genotype correlations : the lower triangle matrix

The following example gives a model file with a filter on the trait names third and malcor

```
5              ! Number of traits
1 1            ! Number of  fixed effects and covariables
sexe poids     ! Names of the fixed effects and covariables
malade r 1 1 0 ! 1st trait, (nature : real value) model
malcor r 0 0 1 ! 2nd trait,(nature : real value) model
third  r 0 0 0 ! 3nd trait,(nature : real value) model

correlation_matrix

0.35 0.28 0.29
0.20 0.32 0.28
0.20 0.20 0.33

third malcor
```

Texte 7: Example 2 of a model file

The key word « **all** » allows the use of the same model for all the traits (useful for eQTL detection).

```
10000          ! Number of traits
1 1            ! Number of  fixed effects and covariables
sexe cov1     ! Names of the fixed effects and covariables
all r 1 1 0 ! all is a word key: the model will be applied for all
            ! the 10000 expression trait
```

Texte 8: Example 3 of a model file

To apply a filter with the key word « all » the user have to give an index trait list (referenced in the phenotype file. Trait one → index 1, Trait two → 2).

```
10000          ! Number of traits
1 1            ! Number of  fixed effects and covariables
sexe cov1     ! Names of the fixed effects and covariables
all r 1 1 0
3 4 5 6 45 46
```

Texte 9: Example 4 of a model file

6.2. The parameter file

All information needed by an analysis is the parameter file *p_analyse.*:

- name of the dataset files : genealogy, map, genotypes and performances
- name of the model file describing the performances
- paths and names of the ouput files :
  - full information analysis result file
  - summary of the analysis
  - sire and dam family **likelihood ratio test (LRT)** along the linkage group
  - sire and dam **QTL effect estimations** along the linkage group (under hypothesis H1 = 1 QTL and H2 = 2 QTL)
  - grand parental **segment transmission** marginal and joint probabilities
- fixed options:
  - chromosomes explored
  - step length of the scan
  - minimum size of a full sib above which the dam effects (QTL and polygenic) are estimated

- minimal paternal and maternal phase probability
- missing genotype value

The parameter file use the format <key>=<value>. None of the characters after the character '#' are interpreted (useful to add comments).

several key may be defined :

input file keys :

in_map= <path file> the map file

in_genealogy = <path file> the genalogy file

in_genotype= <path file> the genotype file

in_traits= <path file> the traits file

in_model= <path file> the model files describing the performances

optionals keys

opt_step = <*real*> step length of the scan (Morgan)

opt_ndmin=<*real*> Minimal number of progeny by dam : offspring size above which the polygenic and QTL effects of the dam are estimated

opt_mindamphaseproba=<*real*> Minimal maternal phase probability : threshold above which the probable maternal phases will be considered in the analysis

opt_minsirephaseproba=<*real*> Minimal paternal phase probability : the analysis is interrupted if for a sire, none of its phases reach this threshold

opt_chromosome=<string,string,...> chromosomes to be analysed, as denoted in the marker map file

opt_unknown_char=<string> string code for missing value

main output file

out_output=<*path file*> *:* Full information about the results

output analysis files keys

out_summary=<*path file*> *:* Short information about the results

out_lrtsires=<*path file*> *:* Sire family likelihood ratio test file

out_lrtdams=<*path file*> *:* Dam family likelihood ratio test file

out_pateff=<*path file*> *:* Sire QTL effect estimations file under Hypothesis H1

out_mateff=<*path file*> *:* Dam QTL effect estimations file

out_phases=<*path file*> : Parental phases informations

out_freqall=<*path file*> : Alleles frequencies informations

out_grid2qtl=<*path file*> *:* Sire QTL effect estimations file under Hypothesis H2

out_pded=<*path file*> *: G*rand parental segment transmission marginal probabilities

out_pdedjoin=<*path file*> *: G*rand parental segment transmission joint probabilities

out_haplotypes=<*path file*> *:*

out_coeffda=<*path file*> *:*

input simulation file

in_paramsimul=<*path file*>

output simulation file

out_maxlrt=<*path file*>

```
#qtlmap --help-panalyse : for more information
##### USER FILES
in_map=carte
in_genealogy=genea
in_genotype=typage
in_traits=perf
in_model=model

##### ANALYSIS PARAMETERS
# analysis step :  in Morgan
opt_step = 0.1
# minimal number of progeny by dams
opt_ndmin=20
#Minimal paternal phase probability
opt_minsirephaseproba=0.80
# overload :
opt_minsirephaseproba=0.90
#Minimal maternal phase probability
opt_mindamphaseproba=0.10
# chromosome to analyse
opt_chromosome=7
#for several chromosomes
#opt_chromosome=7,8,Y
#missing phenotype marker value
opt_unknown_char=0
###### OUTPUT
out_output=./OUTPUT/result
out_summary=./OUTPUT/summary
out_lrtsires=./OUTPUT/sires
out_lrtdams=./OUTPUT/dams
out_pded=./OUTPUT/pded
out_pdedjoin=./OUTPUT/pdedjoin
out_pateff=./OUTPUT/pateff
out_mateff=./OUTPUT/mateff
out_phases=./OUTPUT/phases
out_haplotypes=./OUTPUT/haplotypes
```

Texte 10: Example of a parameter analyse file

6.3. Principes

Mixture of half-sib and full sib families

The maximul likelihood methods implemented in QTLMap considers the population as being a mixture of half sib and full sib families. The sires and the dams are supposed unrelated. A sire (*resp.* a dam) may be mated to more than one dam (*resp.* sire) . Thus, two animals of the second generation may be unrelated, half sibs or full sibs. A polygenic and a QTL effect are estimated for each parent having a large enough family. To avoid numerical difficulties, these effects are not estimated for dams having too small offspring. In this case, the dam progeny are considered as sire half sibs only. A control of the structure is allowed through the option number of progeny *opt_ndmin* which is given in the parameter file.

	ndmin=5	ndmin=7
Families Full sib	910014-910001 910014-910002	910014-910002

You may overload the option opt_ndmin and consider all families as half-sib using the runtime option –family=1.

Minimal paternal and maternal phases probability

In the current release QTLMap considers only one phase for the sire. When the runtime option –haplotype=1,2,3 is used, the probabilities of all possible sire and dam phases are computed. If none of those probabilities for the sire exceed a given threshold (opt_minsirephaseproba in the parameter file) the process is aborted.

As the dams generally have a lower offspring size, all phases the probability of which exceeds a given threshold (opt_mindamphaseproba in the parameter file) are considered in the analysis.

7. Analyses

7.1. Available analysis

Calcul	Description	QTL	Type data
1	LA for a single trait with pre-corrected data	1,2	Real
2	LA for a single data with a model description	1	Real,Discrete
3	LA for a single data with a model description (likelihood linearised - homoscedatic)	1,n	Real
4	LA for a single data with a model description (likelihood linearised - heteroscedastic)	1,n	Real

5	LA for a set of traits with a multivariate analysis (based on a multi-normal penetrance function)	1	Real
6	LA for a set of traits (without missing data) with a discriminante analysis	1	Real
7	LA for a single survey trait with the cox model	1	Real with censored data
8	LD for a single data with a model description	1	Real
9	LDLA for a single data with a model description	1	Real
25	LD for a single data with a model description (likelihood linearised - homoscedatic)	1,n	Real
26	LD for a single data with a model description (likelihood linearised - heteroscedastic)	1,n	Real
27	LDLA for a single data with a model description (likelihood linearised - homoscedatic)	1,n	Real
28	LDLA for a single data with a model description (likelihood linearised - heteroscedastic)	1,n	Real
23	LA for a set of traits with a model description	1,n	Real

7.2. Single real trait with pre corrected data

A remplir

7.3. Single real or discrete trait with a model description

A remplir

7.4. Single real trait with a model description and a complete linearised likelihood

A remplir

7.5. Set of real traits with a multivariate analysis (based on a multi-normal penetrance function)

A remplir

7.6. Set of traits with a discriminante analysis

A remplir

7.7. Single survey trait with the cox model with a model description

In order to take into account censored data, a likelihood function using a semi parametric model (Cox model) was developed. This model allows to search QTL without any assumption about the data distribution. Consequently, this model is unable to give estimations of QTL and nuisance effects expressed in phenotypic standard deviation. The estimated effects are expressed as relative risks compared to a reference level (the first one). For more details, the equations of the approximated likelihood used in the QTLMAP software were presented in the annex part of the paper of Moreno et al. (2005).

7.8. Runtime options

Analyse

The calcul runtime option allows the choice between different types of modelling.

1) Analysis of a single real trait with pre corrected data

   ```
   >${QTLMAP_PATH}/qtlmap p_analyse --calcul=1
   ```

2) Analysis a single real or discrete trait with a model description

   ```
   >${QTLMAP_PATH}/qtlmap p_analyse --calcul=2
   ```

3) Analysis a single real trait with a model description and a complete linearised likelihood (homoscedastic and heteroscedastic)

   ```
   >${QTLMAP_PATH}/qtlmap p_analyse –calcul=3
   >${QTLMAP_PATH}/qtlmap p_analyse –calcul=4
   ```

4) Analysis a set of real traits (without missing data) with a multivariate analysis (based on a multi-normal penetrance function)

   ```
   >${QTLMAP_PATH}/qtlmap p_analyse –calcul=5
   ```

5) Analysis a set of traits (without missing data) with a discriminant analysis

   ```
   >${QTLMAP_PATH}/qtlmap p_analyse –calcul=6
   ```

6) Analyse a single survey trait with the cox model

   ```
   >${QTLMAP_PATH}/qtlmap p_analyse –calcul=7
   ```

7) Analyse a single survey trait with the LD

   ```
   >${QTLMAP_PATH}/qtlmap p_analyse –calcul=8
   ```

8) Analyse a single survey trait with the LDLA

   ```
   >${QTLMAP_PATH}/qtlmap p_analyse –calcul=9
   ```

9) Analysis a single real trait with a model description and a complete linearised likelihood

(homoscedastic and heteroscedastic) with the LD

```
>${QTLMAP_PATH}/qtlmap p_analyse –calcul=25
>${QTLMAP_PATH}/qtlmap p_analyse –calcul=26
```

10) Analysis a single real trait with a model description and a complete linearised likelihood (homoscedastic and heteroscedastic) with the LDLA

```
>${QTLMAP_PATH}/qtlmap p_analyse –calcul=27
>${QTLMAP_PATH}/qtlmap p_analyse –calcul=28
```

Haplotype

Changing the calculus of the parental phases and for all progeny, the grand parental segment transmission adapted for SNP.

```
>${QTLMAP_PATH}/qtlmap p_analyse –calcul=1 –snp
```

--haplotype=	Description
1	“Classical” approach by enumeration All possible phases are considered in turn and their probability computed Transmission probabilities are computed using all available information `Recommended for small number of markers`
2	Optimised approach for sparse maps All possible phases are considered in turn and their probability computed `Transmission probabilities are computed using local information`
3	Approximate phasing based on closest marker information Exact transmission probability minimising the computation `Recommended for dense maps`
4	

Optimisation

The `–optim` runtime option allows a control of the optimisation procedure. The following table describes the available methods.

--optim=	Description	DEPENDANCES

1	E04JYF NAG routine - quasi-Newton	NAGG
2	L-BFGS routine - **the Broyden–Fletcher–Goldfarb–Shanno** quasi-Newton	no
5,...,11	LUKSAN optimisation	no
12,...,47	NLOPT Optimisation	GCC

methods may be parametrized with the following options :

- *opt_optim_maxeval : maximum number of objective function*
- *opt_optim_maxtime : maximum time to find the solution of the objective function*
- *opt_optim_tolx : tolerance lower bound of a step*
- *opt_optim_tolf : stopping criteria lower bound of the objective function*
- *opt_optim_tolg : stopping criteria lower bound of the gradient*
- *opt_optim_h_precision : precision to obtain the gradient*

Console ouput mode

- To get the maximum information during the process, add –v (or --verbose) to the command

  ```
  >${QTLMAP_PATH}/qtlmap p_analyse –calcul=1 -v
  ```

- When debuging the software , add –d (or --debug) to the command

  ```
  >${QTLMAP_PATH}/qtlmap p_analyse –calcul=1 -d
  ```

- To avoid outpout, add –q (or --quiet) to the command

  ```
  >${QTLMAP_PATH}/qtlmap p_analyse –calcul=1 -q
  ```

Report output mode

When performing eQTL analysis (using –data-transcriptomic command) or simulation the output is minimised. To force the classical reporting format, use the runtime option –`print-all`.

Example :

```
>${QTLMAP_PATH}/qtlmap p_analyse –calcul=1 –data-transcriptomic --print-all
```

Number of qtl detection available

For most of the analyses (controlled by the runtime option --calcul), only 1 QTL is considered in the model. However, this number may be increased to 2 if calcul=1 to 2 or more if calcul = 3 or 4. The number of QTL is given by the --qtl runtime option.

Analysis --calcul	QTL test detection --qtl
1	1,2

2,7,8,9,10	1
3,4,25,26,27,28	>=1
5,6	1

Example:

```
>${QTLMAP_PATH}/qtlmap p_analyse –calcul=1 --qtl=1
```

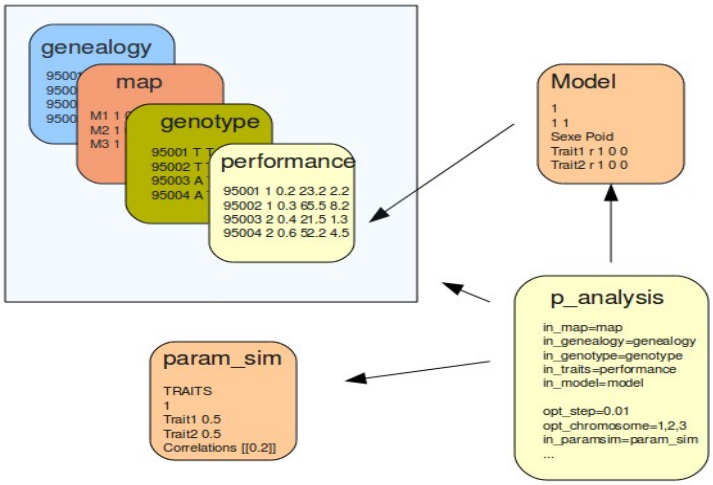
EQTL analysis

When looking for eQTL the number of traits to be analysed becomes very large. In this case, specific routines are needed, and ad hoc output are produced. To get this situation, the runtime option data-transcriptomic must be indicated

```
>${QTLMAP_PATH}/qtlmap p_analyse –calcul=1 –qtl=1 --data-transcriptomic
```

8. Estimation of the test statistic rejection thresholds

8.1. Estimation of the test statistic rejection thresholds with missing data



A specific file, *opt_paramsimul (param_sim)* must be provided by the user. This file contains the needed information about the simulation :

- QTLs informations
  - Number of QTLs (N)
  - N QTL positions in Morgan
  - N chromosomes where are localised QTLs
  - N QTL allele frequencies in the grand sire population
- Traits informations
  - Number of traits (M)
  - List of traits (M lines) corresponding to the model file
  - N QTLs effects for each M traits

If the simulations are made under the null hypothesis (No QTL on the linkage group) the user has only to give the second part (Trait) of the simulation parameter file.

In the case of simulations made under the hypothesis of N QTL, N≠0, (this case occurs when the aim is to get rejection thresholds for the test of H1 “only 1 QTL” *vs.* H2 “2 QTLs” segregating), the QTL is supposed to be biallelic Q1,Q2 and the genotypes frequencies in the parental population are Q1Q1 : f1.(1-f1), Q1Q2 : f1.f1+(1-f1).(1-f1), Q2Q2 : (1-f1).f1, where f1 is the frequency of the first allele if the grand sire population, the second allele in the grand dam population. To get for instance all parents heterozygous, the frequency f1 must be given the value 1. or 0.

Format of the simulation parameter file

```
QTL
<integer>
```

The specific “QTL” Label on the first line, followed by and the number of QTLs to be simulated

```
Position    <real> <real> ...
chromosome  <integer> <integer> ...
frequency    <real> <real> ...
```

The user defined for each QTL:

- its position
- the chromosome where it is located
- the frequency in grand sire population P1

```
TRAITS
<integer>
```

The specific TRAITS Label on a first line, then the number of traits to be simulated

```
<IDNAME>
```

For continuously distributed traits : the name of one of the traits as referenced in the model file

```
<IDNAME_DISCR_DATA> <int> <real> <real>
```

For discrete traits : the name of one of the discrete traits as referenced in the model file, with :

- its heritability
- the number of modalities
- the frequency of each modality

```
qtleffect <real> <real>...
```

Only if one or more QTL is defined :

- QTL 1 Effect on trait 1, QTL 1 Effect on trait 2,...,QTL 2 Effect on trait 1,QTL 2 Effect on trait 2,....

On the whole, the *opt_paramsimul* is the following :

The entirely format

```
QTL
<integer>
Position    <real> <real> ...
chromosome  <integer> <integer> ...
frequency    <real> <real> ...

TRAITS
<integer>
<IDNAME> | <IDNAME_DISCR_DATA> <int> <real> <real>
( qtleffect <real> <real>...) 0/1 (*)
```

(*) : The qtleffect line is defined if at least one QTL are simulated.

Example of a parameter file for the estimation of the rejection thresholds for the test « There are one qtl on the linkage group» against « there are no QTL »

```
TRAITS
2
imf
bardiere
```

Texte 11: Parameter simulation file

```
2
0 0
nofix nocov
imf r 0 0 0
bardiere r 0 0 0
```

Texte 12: Model file

Example of a parameter file for the estimation of the rejection thresholds for the test « There are two qtl on the linkage group» against « there are one QTL at the position 0.6 Morgan on the first chromosome on the linkage group»

In this example, the QTL simulated have an effect 0.4 on the first trait and 0.5 on the second traits. The QTL have a frequence of 100%...

```
QTL
1
position 0.6
chromosome 1
frequency 1.0

TRAITS
2
imf
bardiere

qtleffect 0.4 0.5
```

Texte 13: Parameter simulation file

Addition keys in the parameter file

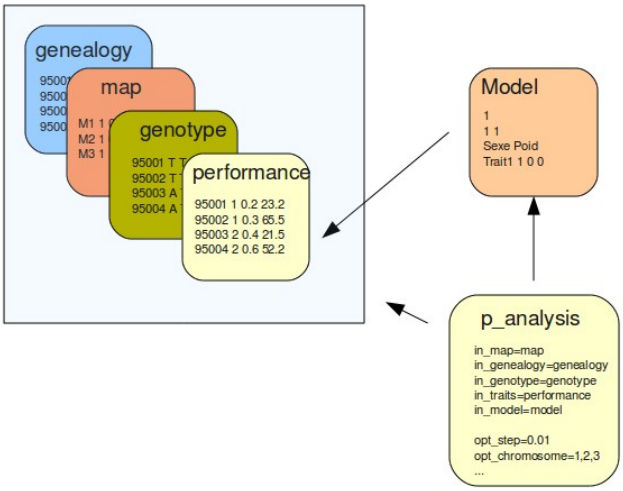
The parameters simulation file is given in the parameter analyse file with the key *in_paramsimul*.

A second key (optional) *out_maxlrt* specifies the name of a file reporting the maximum likelihood ratio test values found in the simulations.

```
#qtlmap --help-panalyse : for more information
##### USER FILES
in_map=carte
in_genealogy=genea
in_genotype=typage
in_traits=perf
in_model=model
in_paramsimul=param_sim_simul

##### ANALYSIS PARAMETERS
# analysis step :  in Morgan
#minimum : 0.000001
opt_step = 0.1
# minimal number of progeny by dams
opt_ndmin=20
#Minimal paternal phase probability
opt_minsirephaseproba=0.80
# overload :
opt_minsirephaseproba=0.90
#Minimal maternal phase probability
opt_mindamphaseproba=0.10
# chromosome to analyse
opt_chromosome=7
#for several chromosomes
#opt_chromosome=7,8,Y
#missing phenotype marker value
opt_unknown_char=0
###### OUTPUT
out_output=./OUTPUT/result
out_summary=./OUTPUT/summary
out_maxlrt=./OUTPUTSIM/simul
```

Texte 14: Example of a parameter file to estimate the rejections thresholds with missing data

8.2. Permutations



In QTLMAP software, the permutation option allows to permute the nuisance effects and phenotypes between genotyped animals within full and/or half sib families to calculate the distribution of likelihood under null hypothesis. The permutation procedure, proposed by Churchill and Doerge (1994) is an intuitive method for estimating threshold which accurately reflects the specificities of an experimental situation". However, when the permutation groups becomes small, the number of permutation possibilities decrease and then the user is advertised that the simulation method is more adapted to estimate the likelihood distribution under H0. In order to forbidden unsuited calculus, an arbitrary threshold for family sizes was fixed to 10. Different permutation situations were considered:

- When the full sib family size is higher than nd_min key (or 10 if nd_min_key<10), genotyped animals are permuted within the sire full sib family
- When the full sib family is smaller than nd_min_key (or 10 if nd_min_key<10), the permutation is performed within half sib family.
- When a full sib family is smaller than 10, no permutation is performed and an error message was printed.

In case of a multitrait analysis (multivariate or discriminante), only phenotyped animals are permuted. In case of a unitrait analysis, animal with at least one phenotyped trait among the traits of performance file are permuted.

The rejection thresholds may be obtained with permutations on performances. This option is available with the runtime option ***--permute***

```
>${QTLMAP_PATH}/qtlmap p_analyse –calcul=1 –nsim=100 –permute
```

Information about the permutation process

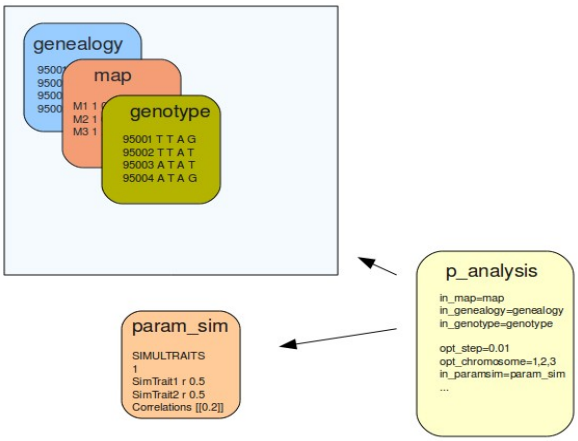
The permutation option concerns the phenotypes and all nuisances effects attached to the phenotypes.

The performances are permuted within the full sib family. However, if the number of progeny for a dam is less than the minimum between opt_ndmin key value (building full sib family) and 10 (this figure was chosen by the developers of QTLMap and will be controlled by advanced users soon), the permutation is realized within half sib family.

In multi-trait analysis (multi-variate or discriminant), only phenotyped animals are permuted.

In successive uni-trait analysis, animal without any phenotype are not included in the permutation.

8.3. Estimate of the test statistic rejection thresholds without missing data



The user have the possibility to estimate thresholds rejections for dummy traits, assuming there is no missing data. In this case, the parameter file does not need the keys *in_model nor in_trait*.

The parameter simulation file will have a specific head section for simulation trait : SIMULTRAITS.

This section is identical to the TRAIT section but an additional information abouit the nature of the

trait as described for the model file. This information is given next the IDNAME of trait :

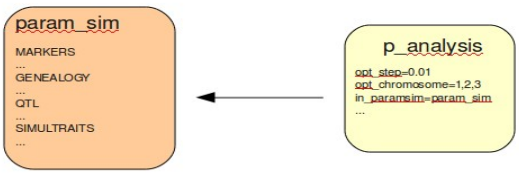
- « r » for real data
- « i » for integer (ordered discrete data)

```
QTL
<integer>
Position    <real> <real> ...
chromosome  <integer> <integer> ...
frequency    <real> <real> …

SIMULTRAITS
<integer>
<IDNAME> r <real> | <IDNAME_DISCR_DATA> i <real> <int> <real> <real>
( correlation [ [ a ] [ b c] [ d e f] ... ] ) 0/1 (*)
( qtleffect <real> <real>...) 0/1 (**)
```

```
#qtlmap --help-panalyse : for more information
##### USER FILES
in_map=carte
in_genealogy=genea
in_genotype=typage
in_paramsimul=param_sim_simul

##### ANALYSIS PARAMETERS
# analysis step :  in Morgan
#minimum : 0.000001
opt_step = 0.1
# minimal number of progeny by dams
opt_ndmin=20
#Minimal paternal phase probability
opt_minsirephaseproba=0.80
# overload :
opt_minsirephaseproba=0.90
#Minimal maternal phase probability
opt_mindamphaseproba=0.10
# chromosome to analyse
opt_chromosome=7
#for several chromosomes
#opt_chromosome=7,8,Y
#missing phenotype marker value
opt_unknown_char=0
###### OUTPUT
out_output=./OUTPUT/result
out_summary=./OUTPUT/summary
out_maxlrt=./OUTPUTSIM/simul
```

Texte 15: Example of a parameter file to estimate the rejections thresholds without missing data

9. Simulate and design a new protocol



QTLMap offers you the possibilty of simulating all the data (markers, genealogy, traits) in order to plan a new experiment. You will get in the output file (named by the out_maxlrt=./OUTPUTSIM/simul option in the following example) the value of the LRT resulting from the simulation, allowing an estimation of designs power.

To perform those simulations, two specific section must be created in the param_sim file :
The first, with the head section MARKERS, must give on a single line
Marker density (M), number alleles/marker, map size (Morgan)

The second, with the head section GENEALOGY, followed by the key word F2, BC or OUTBRED depending on the type of population, and a line giving the number of sires, of dam/sire and of progeny / dam

```
MARKERS
<real> <integer> <integer> <integer> <character>

GENEALOGY
<F2|BC|OUTBRED>
<integer><integer><integer>

QTL
<integer>
Position    <real> <real> ...
chromosome  <integer> <integer> ...
frequency    <real> <real> ...

SIMULTRAITS
<integer>
<IDNAME> r <real> | <IDNAME_DISCR_DATA> i <real> <int> <real> <real>
( correlation [ [ a ] [ b c] [ d e f] ... ] ) 0/1 (*)
( qtleffect <real> <real>...) 0/1 (**)
```

```
in_paramsimul=param_sim_optim

##### ANALYSIS PARAMETERS
# analysis step :  in Morgan
#minimum : 0.000001
opt_step = 0.1
# minimal number of progeny by dams
opt_ndmin=20
# chromosome to analyse
opt_chromosome=7

###### OUTPUT
out_output=./OUTPUT/result
out_summary=./OUTPUT/summary
out_maxlrt=./OUTPUTSIM/simul
```

Texte 16: Example of a parameter file to design a new protocol

10. Output files

A set of files is proposed to the user as the result of an analysis or a simulation :

- The main output (analyse report, simulation report)
- A summary

Additional files (optional) in analysis case :

- Likelihood ratio test profile (per Sire, per Dam, global)
- QTL effect estimation at each tested position (Sire and dam)
- Parental phases report
- Alleles frequencies informations
- Haplotypes assigned from parents
- Grand parental segment transmission marginal probabilities
- Grand parental segment transmission joint probabilities

Specifics files :

- Coefficients of the discriminant analysis among the linkage group

Additional file (optional) in a simulation/permutation case :

- Maximum likelihood Ratio Test and optimal positions reached for each simulations/permutations

10.1.Analysis report

The **first part** describes the data as given by the user

The name of the corresponding file is given by the user with the key ***out_output*** in the parameter file

Configuration defined by the user

The list of option keys used by the application (runtime environment) is given (All keys are described at the end of this document).

Description of the genealogy

Number of parents, grand-parents and progenies

Description of the markers

Number of animal genotyped

Number and names of the genetic markers, of alleles by marker and allele frequencies

Warning about the equilibrium of marker transmission within each family

Description of the traits

Names of the quantitative traits, for each trait :

- number of animals measured
- number of animals measured for both performance traits and marker genotypes
- mean, variance, minimum and maximum
- Names of fixed effect, if any, with the list of levels
- Names of the covariates, if any, with their mean, variance, minimum and maximum

The **second part** describes the result of the phase building

Parental phases

A part of the **most probable phases of the reproducers**, built from available marker and pedigree information, are listed. The full information is found in the specific file.
A control is given to the user with the keys
opt_minsirephaseproba and ***opt_mindamphaseproba*** (Minimal sire and dam phase probability)

In the **third part**, results of the genome scan are given for each traits. Details depends on tests and models.

Genome scan

Section \ calcul	1	2	3	4	5	6	7	8
Possible confusions between QTL and other effects		x	x	x				
Residual variances and estimation of the main effects (polygenic,QTL)	x	x	x	x	x	x		x
LRT for the nuisance effects		x	x	x			x	
Risk Factor estimation							x	
Precision of the parameter estimation		x	x	x				
General Mean estimation		x	x	x				
Nuisances effects estimations		x	x	x			x	
Interactions between QTL and fixed effects		x	x	x			x	
Traits residual correlations					x			

Tableau 1: Output availables according to the analysis

Confusion between QTL effects and all other effects

As the design may be poorly balanced, leading to strong colinearity between QTL and some other effects in the model, a warning is provided if this situation occurs. The confusion is measured by the correlation between the columns of the incidence matrix in an equivalent fully linear model at the starting position of the scan (a warning is edited if this correlation exceeds ***opt_eps_confusion***).

A second test of confusion between the QTL and other estimable effects finally kept in the model is edited.

Variances and estimation of main effects

Within sire residual variance estimations are printed under all tested hypotheses (no QTL, one QTL, two QTL,...). MThe maximum likelihood solutions for the parameters are given, with an indication about their precision (available only for calcul =2, 3, 4), estimated by the diagonal element of the incidence matrix in an equivalent fully linear model: the lower the better :

- *global mean*
- sire QTL effects

- dam QTL effects
- sire polygenic effects
- dam polygenic effects
- *covariables*
- *fixed effects*

The two following example give difference report according to the calcul option.

```
---------------------------------------------------------------
 Estimation of parameters under H0
 ---------------------------------------------------------------

 Within sire standard deviation
  ** Trait bardiere **
 sire 910001  s.d. :     0.551
 sire 910045  s.d. :     0.578
 sire 910081  s.d. :     0.659
 sire 910088  s.d. :     0.663



  parameter            estimable ?    value     precision

 Mean Sire

Sire 910001                yes      6.902     0.000
Sire 910045                yes      7.091     0.000
Sire 910081                yes      7.220     0.000
Sire 910088                yes      7.441     0.000

 Mean dam

Dam 910014 [Sire 910001]   yes      0.040     0.000
Dam 910002 [Sire 910081]   yes      0.000     0.000
Dam 910010 [Sire 910081]   yes      0.000     0.000
Dam 910074 [Sire 910088]   yes      0.000     0.000
...
```

Texte 17: Estimation of variances and polygenic effect under hypothesis null with the calcul=1

Note that with calcul=1, the precision is not computed and is arbitrary given the vaue 0.0

```
-----------------------------------------------------------------
 Estimation of parameters under H0
 -----------------------------------------------------------------

 Within sire standard deviation
  ** Trait bardiere **
 sire 910001  s.d. :     0.550
 sire 910045  s.d. :     0.579
 sire 910081  s.d. :     0.658
 sire 910088  s.d. :     0.654



  parameter            estimable ?    value     precision

 General Mean

                           yes      7.539     0.033

 Sire polygenic effects

Sire 910001                yes     -0.666     0.067
Sire 910045                yes     -0.448     0.058
Sire 910081                yes     -0.264     0.065
Sire 910088                no

 Dam polygenic effects

Dam 910014 [Sire 910001]   yes      0.061     0.069
Dam 910002 [Sire 910081]   yes     -0.052     0.073
Dam 910010 [Sire 910081]   yes     -0.129     0.068
Dam 910074 [Sire 910088]   yes     -0.221     0.075

  NOTE: known allelic origin means QTL effect =  maternal - paternal allele effects
...
```

Texte 18: Estimation of variances, general mean and polygenic effect under hypothesis null with the calcul=2

Interactions between QTL and fixed effects

When interactions between the QTL and m fixed effects are considered in the model, the dam and sire qtl effects are estimated for each level of the composite interacting fixed effect (if n_1, n_2.. n_m are the number of levels for effect 1, 2...m, a total of $n_1.n_2...n_m$ qtl effects are estimated for each parents)

Testing nuisances effects

For each of the nuisance effect, a LRT is reported with the value and significance of the likelihood ratio when comparing a model with or without this effect. The significance is the probability for the LRT to be higher than the observed value under H0 (no effect). When this probability exceeds the standard threshold corresponding to the 5, 1 or 0.1 Pent level, the effect should be removed from the model.

```
*******************************************************************************
test of the effets of the model


Tested effect     df.    Likelihood     p-value
                         ratio

f1          (direct effect)    23    100.823   1.000
f2          (direct effect)    10    121.576   1.000
sex         (direct effect)     2     11.146   1.000
...
```

Texte 19: Test of the nuisances effects

Risks factor estimation

Traits residual correlations

10.2. EQTL analysis report

A special format presents the report analysis for each gene expression (depends the dynamic flag ***--data-transcriptomic***). Only calculus 1,2,3,4 manage this format (single trait analysis).

For each hypothesis , the report gives :

- The header of the following array
- Array with :
  - first column : gene name
  - others column : estimation of each parameters given in the header

note :
The values 0.0 means that the parameter is not estimable.

```
 Hypothesis :0
 Given parameters are respectively :
 Gene position on the array, [ *std dev *1940][General Mean][Sire polygenic effects]

  note : 0.0 mean not estimable

                   1  0.132 -0.106  0.000
                   2  0.116 -0.114  0.000
                   3  0.165 -0.140  0.000
                   4  0.097  0.174  0.000
                   5  0.135 -0.147  0.000
                   6  0.259 -0.059  0.000
...
```

Texte 20: EQTL report under hypothesis 0

```
Hypothesis :1
 Given parameters are respectively :
 Gene position on the array, Chromosome 1, QTL Position 1,H0/H1,[ *std dev *1940][General Mean][Sire QTL effects [1]][Sire
polygenic effects]

  note : 0.0 mean not estimable

                   1  1.000   0.930   2.301   0.128  -0.106   0.033   0.000
                   2  1.000   0.830   0.653   0.115  -0.114  -0.017   0.000
                   3  1.000   1.430   4.446   0.157  -0.139  -0.055   0.000
                   4  1.000   1.430   2.248   0.095   0.174  -0.023   0.000
                   5  1.000   1.230   0.247   0.134  -0.147  -0.010   0.000
                   6  1.000   1.430   2.007   0.254  -0.057  -0.059   0.000
...
```

Texte 21: EQTL report under hypothesis 1

```
 Hypothesis :2
 Given parameters are respectively :
 Gene position on the array, Chromosome 1, QTL Position 1,Chromosome 2, QTL Position 2,H0/H2,H1/H2,[ *std dev *1940][General
Mean][Sire QTL effects [1]][Sire QTL effects [2]][Sire polygenic effects]

  note : 0.0 mean not estimable

                  1  1.000   1.130   1.000   1.430   4.933   2.632   0.125  -0.105   0.084  -0.071   0.000
                  2  1.000   1.530   1.000   1.730   1.104   0.451   0.114  -0.113  -0.030   0.026   0.000
                  3  1.000   0.930   1.000   1.030   9.842   5.396   0.148  -0.142   0.371  -0.365   0.000
                  4  1.000   1.030   1.000   1.330   2.963   0.715   0.094   0.174   0.019  -0.037   0.000
                  5  1.000   1.530   1.000   1.730   1.095   0.848   0.133  -0.146  -0.032   0.034   0.000
                  6  1.000   0.830   1.000   1.530   2.245   0.237   0.253  -0.057  -0.029  -0.045   0.000
...
```

Texte 22: EQTL report under hypothesis 2

10.3. Analyse summary

In the file SUMMARY (parameter file key ***out_summary***), several chapters are given summarising the analysis under all hypothesis.
For each hypothesis (H0 : 0 qtl, H1 : 1 qtl, H2 : 2qtl, ...)
for each analysed variable (by lines)

- Number of genotyped progeny with phenotypes for the trait
- Maximum likelihood ratio
- QTL most likely positions
- for each sire
  - Estimations of the QTL effect
  - Within sire family standard deviation
  - Significance of the QTL effect (based on a Student test). ‘sign’ = significant; ‘ns’= not significant; ‘na’=not available.

```
 ********************************************************************************
 Summary 0 QTL versus 1 QTL
Variable  N        Max Lik        Pos (M)    Sire      910001     910045       910081         910088
              0/1QTL    Chr 1    Pos1 eff1 SD  sig1 eff1 SD  sig1 eff1 SD  sig1 eff1 SD  sig1
bardiere 236   45.2      1    0.7        -0.089  0.511  sign -0.118  0.560  sign -0.162  0.572  sign -0.167  0.598  sign
     imf 236   43.7      1    0.7         0.156  0.338  sign  0.187  0.426  sign  0.133  0.355  sign  0.051  0.339   ns

 ********************************************************************************
 Summary 0 QTL versus 2 QTL,1 QTL versus 2 QTL
Variable  N        Max Lik        Pos (M)    Sire    910001       910045       910081         910088
              0/2QTL 1/2QTL    Chr 1    Pos1   Chr 2    Pos2 eff1 eff2 SD  sig1 sig2 eff1 eff2 SD  sig1 sig2 eff1 eff2 SD  sig1
sig2 eff1 eff2 SD  sig1 sig2
bardiere 236  57.0 11.9   1  1  0.7  1.1 -0.148  0.082  0.481  sign sign -0.226  0.160  0.543  sign sign -0.182  0.030  0.570
sign  ns  -0.239  0.122  0.589  sign sign
     imf 236   49.3    5.6      1      1    0.9    1.0         0.405 -0.245  0.335  sign sign  0.415 -0.227  0.427  sign sign
0.348 -0.227  0.351  sign sign  0.265 -0.214  0.329  sign sign

 ********************************************************************************
 Summary 0 QTL versus 3 QTL,1 QTL versus 3 QTL,2 QTL versus 3 QTL
Variable  N        Max Lik        Pos (M)    Sire    910001        910045       910081       910088
              0/3QTL 1/3QTL 2/3QTL    Chr 1    Pos1   Chr 2    Pos2   Chr 3    Pos3 eff1 eff2 eff3 SD  sig1 sig2 sig3 eff1 eff2
eff3 SD  sig1 sig2 sig3 eff1 eff2 eff3 SD  sig1 sig2 sig3 eff1 eff2 eff3 SD          sig1 sig2 sig3
bardiere 236   63.9   18.8    6.9      1      1      1    0.7    0.8    1.1        -0.340  0.266  0.006  0.480  sign sign  ns
0.211 -0.528  0.271  0.533  sign sign sign -0.701  0.676 -0.145  0.561  sign sign sign -0.838  0.819 -0.133  0.575  sign sign
sign
imf 236   60.6   16.9   11.3      1      1      1    0.1    0.3    0.7        -0.123  0.092  0.132  0.324  sign sign sign -0.439
0.540  0.072  0.408  sign sign  ns   0.010 -0.042  0.145  0.351   ns   ns  sign  0.097 -0.151  0.083  0.319  sign sign sign
```

Texte 23: Summary with --qtl=3 option

10.4. The family likelihood

The user have to define the following key to obtains the likelihood ratio test among the linkage group under hypothesis one : ***out_lrtsires , out_lrtdam,*** **and/or the grid of t**he likelihood ratio test under hypothesis two : ***out_grid2qtl***.

LRT Sires files

For each tested position, the file contains
Chromosome, Position, global LRT, Sire 1 LRT, Sire 2 LRT ….

```
Chr Pos    GlobalLRT     910001   910045   910081   910088
   1   0.010    8.63    4.93    0.91    2.47    0.33
   1   0.020    8.62    4.82    1.03    2.47    0.30
   1   0.030    8.56    4.66    1.14    2.45    0.31
   1   0.040    8.45    4.47    1.23    2.41    0.35
   1   0.050    8.29    4.24    1.28    2.34    0.42
   1   0.060    8.35    4.21    1.35    2.31    0.48
...
  Chr1  Chr2   Pos1      Pos2      GlobalLRT     910001   910045   910081   910088
...
     1     1    0.02    0.65    3.78    2.72   -0.15   -1.11    2.32
     1     1    0.02    0.66    4.70    3.05    0.12   -0.38    1.92
     1     1    0.02    0.67    5.38    3.31    0.40    0.26    1.41
     1     1    0.02    0.68    5.80    3.51    0.70    0.79    0.80
     1     1    0.02    0.69    5.96    3.65    1.01    1.19    0.11
     1     1    0.02    0.70    5.86    3.71    1.32    1.46   -0.63
...
```

Texte 24: Sire likelihood file

LRT Dams file

For each tested position, the file contains
Chromosome, Position, Dam 1 LRT, Dam 2 LRT ….
Note: when the offspring size of a dam is below the threshold for the search of the phase, the LRT is fixed at 0.000 (see ***opt_ndmin*** option).

LRT grid 2 QTL

The file presents two tables:

The first part of the output concerns the comparison between the 1 and 2 QTL hypotheses
The fist line gives possible 1st QTL position
The following lines give a possible 2nd QTL position, followed by the LRT (1 vs.2 QTL) for each couple of positions

The second part of the output concerns the comparison between the 0 and 2 QTL hypotheses
The fist line gives possible 1st QTL position
The following lines give a possible 2nd QTL position, followed by the LRT (0 vs.2 QTL) for each couple of positions

```
++++++++++++++++ TEST 1QTL / 2QTL ++++++++++++++++++++

        .01     .02     .03     .04     .05     .06
  .01     .00    3.67    8.42   10.30   11.66   12.80
  .02     .00     .00    3.74    8.43   10.30   11.68
  .03     .00     .00     .00    3.81    8.43   10.31
  .04     .00     .00     .00     .00    3.87    8.44
  .05     .00     .00     .00     .00     .00    3.91
[…]

  +++++++++++++++++ TEST 0QTL / 2QTL ++++++++++++++++++++

        .01     .02     .03     .04     .05     .06
  .01     .00   27.46   32.21   34.09   35.45   36.59
  .02     .00     .00   27.53   32.22   34.09   35.47
  .03     .00     .00     .00   27.60   32.22   34.10
  .04     .00     .00     .00     .00   27.66   32.23
  .05     .00     .00     .00     .00     .00   27.70
...
```

Texte 25: Likelihood Grid 2 QTL file

10.5. QTL effects estimations files

The user have to define the following key to obtains the QTL estimations among the linkage group under hypothesis one : ***out_pateff, out_mateff.***

QTL Paternal effects

For each tested position, the file contains
Chromosome, Position, Sire 1 QTL effect estimation, Sire 2 QTL effect estimation …

```
*****************************************
 This file is unvalide if interaction qtl case
 *****************************************
  Chr  Pos     910001   910045   910081   910088
   1   0.010   -0.24   -0.14   -0.13    0.02
   1   0.020   -0.24   -0.15   -0.14    0.01
   1   0.030   -0.24   -0.15   -0.14   -0.01
   1   0.040   -0.23   -0.16   -0.15   -0.03
   1   0.050   -0.22   -0.16   -0.15   -0.05
   1   0.060   -0.23   -0.16   -0.15   -0.06
   1   0.070   -0.23   -0.17   -0.16   -0.08
   1   0.080   -0.23   -0.17   -0.16   -0.09
...
Chr1    Chr2   Pos1     Pos2     910001/Qtl[1]    910001/Qtl[2]    910045/Qtl[1]    910045/Qtl[2]    910081/Qtl[1]
910081/Qtl[2]    910088/Qtl[1]    910088/Qtl[2]
     1     1    0.010    0.020    0.57    0.04    0.57    0.04    0.57    0.04    0.57    0.04
     1     1    0.010    0.030    0.24    0.04    0.24    0.04    0.24    0.04    0.24    0.04
     1     1    0.010    0.040    0.17    0.04    0.17    0.04    0.17    0.04    0.17    0.04
     1     1    0.010    0.050    0.14    0.04    0.14    0.04    0.14    0.04    0.14    0.04
     1     1    0.010    0.060    0.14    0.04    0.14    0.04    0.14    0.04    0.14    0.04
     1     1    0.010    0.070    0.14    0.03    0.14    0.03    0.14    0.03    0.14    0.03
     1     1    0.010    0.080    0.13    0.03    0.13    0.03    0.13    0.03    0.13    0.03
     1     1    0.010    0.090    0.12    0.02    0.12    0.02    0.12    0.02    0.12    0.02
...
```

Texte 26: Paternal qtl effect file

QTL Maternal effect

For each position, the file contains
Chromosome, Position, Dam 1 QTL effect estimation, Dam 2 QTL effect estimation …
Note: the QTL effect are given only for dams the offspring size of which is over the threshold given by ***opt_ndmin***

10.6. Parents phase report

10.7. Offspring phases

Offspring phases are determined assuming the known phases of sires (the most likely phases calculated by QTLMAP software). At this stage, the marker information coming from dams are not considered at all. For each marker, the origin of the phase (dam or sire) is searched.

1) When only one phase origin is possible, the marker genotype coming from the sire and the dam are noted in the first and second lines respectively for each animal.
2) When the phase is not fund, it is predicted using the closest flanking markers with known phase. If the probability of the phases is upper the threshold, the most likely marker genotype coming from the sire and the dam are noted followed by p (for predicted) in the offspring phase file.
3) When they are no likely phases for a marker, the phases are assumed missing for the marker.

out_phases_offspring= Name of the file where the predicted haplotypes of offspring are printed. There are two lines by animal: the first line for the phase coming from sire and the second line for the phase coming from the dam. By default, the software gives the offspring phases for all markers of the first chromosome. But the user is able to choose the size of the chromosomal region given the names of the first and the last markers, using the following options:
opt_phases_offspring_marker_start= Name of the first marker of the printed haplotypes of offspring
opt_phases_offspring_marker_end= Name of the last marker of the printed haplotypes of offspring

10.8. Grand parental segment transmission marginal probabilities

Each line gives for a tested QTL position x

- The sire ID
- The dam ID
- The dam phase number in the order of the main results file
- The progeny ID
- The probability that the progeny inherited the 2nd sire allele (in the order of the main result file) at position x given the dam phase
- The probability that the progeny inherited the 2nd dam allele (in the order of the main result file) at position x given the dam phase

```
Position   Sire    Dam Dam_Phase Animal   p(2nd sire allele)   p(2nd dam allele)
    1.      910001      910014     1      944217                1.000              0.000
    2.      910001      910014     1      944217                0.999              0.001
    3.      910001      910014     1      944217                0.999              0.001
    4.      910001      910014     1      944217                0.999              0.001
    5.      910001      910014     1      944217                0.999              0.001
...
```

Texte 27: Grand parental segment transmission marginal probabilities file

10.9. Grand parental segment transmission joint probabilities

Each line gives for a tested QTL position x

- Position
- Sire ID
- Dam ID

- Dam phase number in the order of the main results file
- Progeny ID
- Probability that the progeny inherited the 1st sire and 1st dam alleles (in the order of the main result file) at position x given the dam phase
- The probability that the progeny inherited the 1st sire and 2nd dam alleles (in the order of the main result file) at position x given the dam phase
- Probability that the progeny inherited the 2nd sire and 1st dam alleles (in the order of the main result file) at position x given the dam phase
- Probability that the progeny inherited the 2nd sire and 2nd dam alleles (in the order of the main result file) at position x given the dam phase

Position	Sire	Dam	Dam_Phase	Animal	p(Hs1/Hd1)	p(Hs1/Hd2)	p(Hs2/Hd1)	p(Hs2/Hd2)
1.	910001	910014	1	944217	0.000	0.000	1.000	0.000
2.	910001	910014	1	944217	0.001	0.000	0.999	0.001
3.	910001	910014	1	944217	0.001	0.000	0.998	0.001
4.	910001	910014	1	944217	0.001	0.001	0.998	0.001
5.	910001	910014	1	944217	0.000	0.001	0.999	0.000
6.	910001	910014	1	944217	0.001	0.001	0.941	0.056
7.	910001	910014	1	944217	0.003	0.001	0.884	0.112

...

Texte 28: Grand parental segment transmission marginal probabilities file

10.10. Simulation report

```
*---------------------------------------*
      Variable traitsimul1
*---------------------------------------*
          Test 0vs1Q
*---------------------------------------*

  Test statistic distribution  :
     Number of simulations :    100
     Mean                  :     14.24685
     Standard deviation    :      4.07168
     Skewness              :      0.70693
     Kurtosis              :      1.05302
     Minimum               :      6.62047
     Maximum               :     28.64581

 *---------------------------------------*
 | chromosome | genome      |  Threshold |
 |          level           |            |
 |----------------------------------------|
 |   0.1000   |             |     19.39  |
 |   0.0500   |             |     21.39  |
 |   0.0100   | chrom_level |     27.40  |
 |   0.0050   |      *      |     28.18  |
 |   0.0027   |   nb_chrom  |     28.44  |
 |   0.0010   |             |     28.58  |
 |   0.0005   |             |     28.61  |
 |   0.0001   |             |     28.64  |
 *---------------------------------------*
```

For each analysed variable, a single line gives the empirical thresholds at 5, 1 and 0.1 % at the chromosome and the genome level. The genome level corresponds to a genome scan of 18 autosomes in pigs. For any other species, the genome level is obtained easily multiplying the chromosome level by the number of chromosomes. In such cases, see the RESULT file for low chromosome wide quantile estimations.

```
    0vs1Q                 p_value at
Trait                chromosome level                 genome level
              5%        1%        0.1%          5%        1%        0.1%

traitsim     21.39     27.40     28.58         28.44     28.61     28.64
```

10.11. Report simulations result

This file give the maximum LRT reached with its associated position (and the linkage group) under the N hypothesis for each simulation/permutation.
For each analysed variable :

- a header to explain the following line to the user
- for each simulation :
  - The Maximum likelihood ratio test
  - Position and linkage group of the first QTL
  - Position and linkage group of the second QTL
  - ...

```
# Trait [traitsimul1] LRTMAX H0/H1 , Position CHR, Position DX
   12.7928    1    0.4100
   18.5180    1    0.1100
   17.0331    1    1.2100
 # Trait [traitsimul2] LRTMAX H0/H1 , Position CHR, Position DX
    8.9628    1    0.7100
    9.3228    1    1.0000
   16.6090    1    0.7100
```

Texte 29: The simulation report file H1

```
 # Trait [traitsimul1] LRTMAX H0/H1 , Position CHR, Position DX LRTMAX H1/H2 ,  Position1 CHR, Position1 DX
Position2 CHR, Position1 DX2
   12.7928    1    0.4100    9.6459    1    0.4100    1    1.2100
   18.5180    1    0.1100   14.2922    1    0.1100    1    1.0100
   17.0331    1    1.2100   15.4039    1    0.3100    1    1.2100
 # Trait [traitsimul2] LRTMAX H0/H1 , Position CHR, Position DX LRTMAX H1/H2 ,  Position1 CHR, Position1 DX
Position2 CHR, Position1 DX2
    8.9628    1    0.7100   12.8711    1    1.5100    1    1.6100
    9.3228    1    1.0000    8.4281    1    0.0100    1    0.3100
   16.6090    1    0.7100    9.5829    1    0.3100    1    0.4100
```

Texte 30: The simulation report file H2

11. Reference

Legarra A, Fernando RL, 2009. Linear models for joint association and linkage QTL mapping. Genet Sel Evol., 41:43.
Elsen JM, Filangi O, Gilbert H, Le Roy P, Moreno C, 2009. A fast algorithm for estimating transmission probabilities in QTL detection designs with dense maps. Genet Sel Evol., 41:50.
Gilbert H., Le Roy P., Moreno C., Robelin D., Elsen J. M., 2008. QTLMAP, a software for QTL detection in outbred population. Annals of Human Genetics, 72(5): 694.
Gilbert H, Le Roy P., 2007. Methods for the detection of multiple linked QTL applied to a mixture of full and half sib families. Genet Sel Evol., 39(2):139-58.
Moreno C.R., Elsen J.M., Le Roy P., Ducrocq V., 2005. Interval mapping methods for detecting QTL affecting survival and time–to–event phenotypes. Genet. Res. Camb., 85 : 139-149.
Goffinet B, Le Roy P, Boichard D, Elsen JM, Mangin B, 1999. Alternative models for QTL detection in livestock. III. Heteroskedastic model and models corresponding to several distributions of the QTL effect.. Genet. Sel. Evol., 31, 341-350.
Mangin B, Goffinet B, Le Roy P, Boichard D, Elsen JM, 1999. Alternative models for QTL detection in livestock. II. Likelihood approximations and sire marker genotype estimations. Genet. Sel. Evol., 31, 225-237.
Elsen JM, Mangin B, Goffinet B, Boichard D, Le Roy P, 1999. Alternative models for QTL detection in livestock. I. General introduction. Genet. Sel. Evol., 31, 213-224

12. Appendix

12.1. Parameter file Option Keys

Key	Description	Default
in_map	Input map file	
in_genealogy	Input genealogy file	
in_genotype	Input genotype file	
in_traits	Input traits file	
in_model	Input model description of traits	
in_paramsimul	Input simulation parameters	
opt_step	Chromosomic segment exploration steps in Morgan	0.05
opt_ndmin	Minimal number of progeny by dam	
opt_minsirephaseproba	Minimal sire phase probability	0.90
opt_mindamphaseproba	Minimal dam phase probability	0.10
opt_unknown_char	Unknown genotype value	'0'
opt_eps_cholesky	coeff cholesky decomposition	0.5
opt_chromosome	Linkage group	
out_output	Main report file	
out_summary	Output summary file	
out_lrtsires	Output file paternal effects	
out_lrtdams	Output file maternal effects	
out_pded	Grand parental segment transmission marginal probabilities	
out_pdedjoin	Grand parental segment transmission joint probabilities	
out_phases	Parental phases file	
out_freqall	Allele frequency file	
out_haplotypes	Haplotype file	
out_pateff	Sire QTL effect estimations	
out_mateff	Dam QTL effect estimations	
out_maxlrt	Simulation report(Position and max LRT)	
opt_eps_confusion	Threshold to test confusion betwwen level inside a contingence matrix	0.70
opt_eps_hwe	Threshold to check the equilibrium of	0.001

	marker transmission within each family	
opt_eps_linear_heteroscedastic	Threshold for convergence in the linear mode heteroscedastic	0.5
opt_max_iteration_linear_heteroscedastic	Maximum iteration in the linear mode heteroscedastic to avoid infinity loop	5
opt_eps_recomb		0.5
opt_nb_haplo_prior		200
opt_pro_haplo_min		
opt_long_min_ibs		
opt_longhap		
opt_optim_maxeval		
opt_optim_maxtime		
opt_optim_tolx		
opt_optim_tolf		
opt_optim_tolg		
opt_optim_h_precision		
out_phases_offspring		
opt_phases_offspring_marker_start		
opt_phases_offspring_marker_end		