

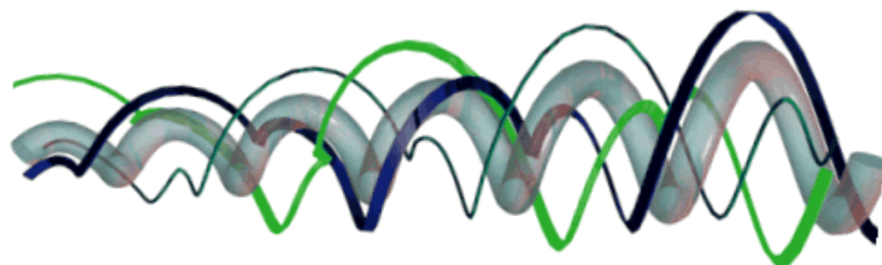
## **MCQTL : A reference manual**

**Brigitte Mangin**

**Rémy Cathelin, David Delannoy, Bertrand Escalière,  
Stéphane Lambert, Jacques Marcel, Baba Ngom,  
Marie-Françoise Jourjon, Abdellaziz Rahmani  
Damien Lerous , Sylvain Jasson**

RAPPORT UBIA TOULOUSE N° 2014/1 - JUIN 2014

DÉPARTEMENT DE MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES



The aim of MCQTL software package is to perform QTL mapping in multi-cross design. It allows to study the usual populations derived from inbred lines and proposes to link the families by assuming that the QTL locations are the same in all them. Moreover, a connected modelling of the QTL genotypic effect is allowed in multiple related families.

The implemented model is a linear regression model in which the probabilities of QTL genotypes given marker information are used as regressors (Haley and Knott, 1992). A composite interval mapping (Zeng, 1993) and an iterative QTL mapping are implemented (Charcosset et al., 2000) to deal with multiple QTL model. Marker cofactor selections by forward or backward stepwise methods are implemented as well as computation of threshold test value by permutation (Churchill and Doerge, 1994).

Finally, the model parameters are estimated. More precisely, Pvalues of the Fisher test at the QTL locations, LOD support intervals, QTL effect and residual parameter estimates, model and QTL percentage of explained variance are provided.

MCQTL package is composed of three software applications. The first component reads the data from MAPMAKER files. The second component, ProbaPop, computes QTL genotype probabilities at specific chromosome location given multiple marker data in each family and stores them in XML formatted files. The last component, Multipop, builds the joint model and fits the observations on genotype probabilities.

MCQTL runs on UNIX platforms.

MCQTL\_LD is an extension of MCQTL that incorporates the use of linkage disequilibrium in MCQTL. At each point along the genome, parent lines of the multicross design are clustered, using a similarity measure that reflects locally the probability of sharing the same ancestral allele. Hence, the number of parameters of the MCQTL model, originally links to the number of parent lines, is lessened and a significant increase in the detection accuracy is achieved.

The similarity measure is based on an extended Li and Jiang's similarity score (Bionformatics, 2005). The Li and Jiang's similarity score between two parent lines is a weighted sum of the number of shared alleles and of the longest common chromosome segment in a sliding window. We extended this score by taking into account prior knowledge on parent relatedness. The implemented clustering method is a filtering and transitive closure of the graph of all the multicross parent lines, weighted by these similarity scores. We proposed two methods to compute by simulation an efficient threshold value for the graph filtering purpose. Our assumption is that the parent lines belonging to the same class transmit the same ancestral allele. So their putative QTL allelic effects can be modeled with the same parameter, leading to a parsimonious model, that is plugged in MCQTL. The name of this MCQTL plugin is clushaplo (Leroux et al., 2012). Clushaplo runs on the R environment of statistical computing and graphics. It provides the user with friendly graphical outputs and the link with MCQTL is totally transparent.

**Important change** : since the local clustering of parent lines leads to varying degrees of freedom for the QTL effects along the genome, the QTL detection method is based, now, on  **$-\log_{10}$  of the Fisher test Pvalue** instead of the Fisher test value.

The great advantage of this new test criterion is that its threshold value does not really vary with the type of design, the number of parents or the number of families, contrarily to Fisher or LOD tests.

## Table des matières

<b>1</b>	<b>Methods</b>	<b>4</b>
1.1	Models . . . . .	4
1.1.1	Intra-family model for a single QTL . . . . .	4
1.1.2	Intra-family model for QTLs in epistasis . . . . .	4
1.1.3	MCQTL_LD model . . . . .	5
1.2	Tests and estimates . . . . .	5
1.2.1	Analysis of a single trait . . . . .	5
1.2.2	Multivariate analysis . . . . .	6
1.3	Detection method . . . . .	7
1.4	Threshold computation . . . . .	7
<b>2</b>	<b>Architecture and command lines</b>	<b>11</b>
<b>3</b>	<b>Data entry files</b>	<b>12</b>
3.1	.gen file . . . . .	12
3.2	.phen file . . . . .	13
3.3	.map file . . . . .	14
3.4	.inf file . . . . .	14
<b>4</b>	<b>XML parameter file for MultiPop</b>	<b>14</b>
4.1	Data and model . . . . .	15
4.2	MCQTL_LD model . . . . .	16
4.3	Threshold computation . . . . .	16
4.4	QTL detection . . . . .	17
4.5	Model estimation . . . . .	19
4.6	Tag inactivation . . . . .	19
4.7	Examples . . . . .	20
<b>5</b>	<b>Internal files</b>	<b>23</b>
<b>6</b>	<b>Output files</b>	<b>24</b>
<b>7</b>	<b>User interface</b>	<b>27</b>
<b>8</b>	<b>References</b>	<b>27</b>
<b>A</b>	<b>List of errors of TranslateData</b>	<b>28</b>

# 1 Methods

## 1.1 Models

Intra-family models implemented in MCQTL permit to analyse a single QTL with only additive effect or with both additive and dominance effects. To deal with interaction between QTLs an epistasis model is also allowed.

In order to link multiple families, the same intra-family model is assumed for all families but genetic parameters, that are parameters of QTL and genetic cofactor effects, can be supposed to be either identical or not over the families.

### 1.1.1 Intra-family model for a single QTL

The intra-family model is a usual regression model with genetic cofactor and a single QTL. Let  $c$  denotes the cross between two parent lines  $i, j$ , the phenotypic value  $Y_{ck}$  of the  $k$ th individual is modeled by

$$Y_{ck} = \mu_c + \sum_{l=1}^L \sum_{ij} p_{ck,ij}^l \theta_{c,ij}^l + \epsilon_{ck}$$

where  $\mu_c$  is the global mean in the cross  $c$ ,  $L - 1$  is the number of genetic cofactors,  $p_{ck,ij}^l$  is the probability of the  $k$ th individual having genotype  $ij$  at the QTL or cofactor locus  $l$  given the marker information,  $\theta_{c,ij}^l$  is the mean of the  $ij$  genotype at locus  $l$  in cross  $c$  and  $\epsilon_{ck}$  the residual error.

Moreover, the QTL or genetic cofactor effect is decomposed in two effects, the additive and the dominance ones

$$\theta_{c,ij}^l = \alpha_{c,i}^l + \alpha_{c,j}^l + \delta_{c,ij}^l$$

In an additive model, the dominance parameters are all assumed to be equal to zero.

When genetic parameters are assumed to be identical over the families, the equality of QTL and cofactor parameters are added in the connected model, leading to

$$\alpha_{c,i}^l = \alpha_i^l \quad \forall c \quad \forall i$$

$$\delta_{c,ij}^l = \delta_{ij}^l \quad \forall c \quad \forall ij$$

Note, that only genetic parameters can be chosen to be identical or not over the families, so, as a consequence of linear model postulates, the intra-family residual variances are all assumed to be identical over the families.

### 1.1.2 Intra-family model for QTLs in epistasis

The intra-family epistasis modelling between loci  $l$  and locus  $l'$  is

$$Y_{ck} = \dots + \sum_{l \neq l'} \sum_{g, g'} p_{ck, g, g'}^{l, l'} (\theta_{c, g}^l + \theta_{c, g'}^{l'} + \phi_{c, g, g'}^{l, l'}) + \epsilon_{ck}$$

where

- ... stands for QTL and genetic cofactor effects without epistasis effect,
- $p_{ck,g,g'}^{l,l'}$  is the probability given the marker information of the  $k$ th individual in cross  $c$  having genotype  $g$  at locus  $l$  and  $g'$  at locus  $l'$ ,
- $\theta_{c,g}^l$  is the  $g$  genotypic effect at  $l$  and  $\theta_{c,g'}^{l'}$  the  $g'$  genotypic effect at  $l'$  in cross  $c$ ,
- $\phi_{c,g,g'}^{l,l'}$  is the epistasis effect between the  $g$  genotype at locus  $l$  and the  $g'$  genotype at  $l'$  in the cross  $c$ .

The probability given the marker information  $p_{ck,g,g'}^{l,l'}$  is computed as the product of  $p_{ck,g}^l$  and  $p_{ck,g'}^{l'}$  which are the probabilities of the  $k$ th individual in cross  $c$  having the genotype  $g$  at locus  $l$  and of the genotype  $g'$  at loci  $l'$ . This product gives an exact probability as soon as the genotypes of an individual at locus  $l$  and  $l'$  are independent given the marker information. This occurs if  $l$  and  $l'$  belong to different chromosomes or if there exists one completely informative marker for the individual between loci  $l$  and  $l'$ .

MCQTL is currently limited to the additive x additive epistasis, which means that  $\phi_{c,g,g'}^{l,l'} = 0$  if  $g$  or  $g'$  are not homozygous genotypes.

When genetic parameters are assumed to be identical over the families, the following equalities are added in the connected model

$$\phi_{c,g,g'}^{l,l'} = \phi_{g,g'}^{l,l'} \quad \forall c \quad \forall (g, g')$$

### 1.1.3 MCQTL\_LD model

In this version, MCQTL\_LD is limited to be used with parent inbred lines and connected model, i.e. the genetic effects (QTL or cofactors) are assumed to be identical over the families.

After the clusthaplo analysis, at each scanned locus  $t$ , each parent line  $i$  is assigned to a cluster  $cl$  that is assumed to be an ancestral allele. Let  $f^t$  be the function that describes the injection of the parent lines into their corresponding class at position  $t$ , i.e.  $f^t(i) = cl$ . The MCQTL\_LD model is obtained by plugging the function in the parameters of the connected model leading to

$$\theta_{ij}^l = \begin{cases} \alpha_{f^l(i)}^l + \alpha_{f^l(j)}^l + \delta_{f^l(i)f^l(j)}^l & \text{if } f^l(i) \neq f^l(j) \\ \alpha_{f^l(i)}^l + \alpha_{f^l(j)}^l & \text{if } f^l(i) = f^l(j) \end{cases}$$

## 1.2 Tests and estimates

### 1.2.1 Analysis of a single trait

Fisher tests of comparison of stacked models are computed to test QTL location. Those tests depend on the chosen intra-family model. For a single QTL with additive or both additive and dominance effects, the compared models are with and without the QTL location. For a couple of QTLs in epistasis, the compared models are with or without the epistasis effects. The degrees of freedom of the test depend on the number of individuals, the number of families and the chosen model.

From a Fisher test value  $F$ ,

- the  $-\log_{10}$  of Fisher test Pvalue is computed using  $-\log_{10}(P(\mathcal{F}(\text{ddl1}, \text{ddl2}) > F))$
- and an approximate LOD test value is obtained using  $\text{LOD} \simeq 0.217 \text{ddl1} \ F$

where  $\text{ddl1}$ ,  $\text{ddl2}$  are the numerator and denominator degree of freedom of the Fisher test and  $\mathcal{F}$  denotes the Fisher probability law.

Parameter estimates are obtained by the least square method with added constraints that depend on the chosen intra and inter family modelling.

### Constraints

- for additive effect, parameters dependent on families

$$\alpha_{c,i}^l + \alpha_{c,j}^l = 0 \ \forall c$$

- for additive effect, parameters indential over the families

$$\sum_{i \in S} \alpha_i^l = 0 \ \forall S$$

where  $S$  denotes a set of connected parent lines

- for dominance effect, there is no mandatory constraint
- for epistasis effect, for both choices of parameter dependency

$$\phi_{[c],ii,ii}^{l,l'} = \phi_{[c],jj,jj}^{l,l'} = -\phi_{[c],ii,jj}^{l,l'} = -\phi_{[c],jj,ii}^{l,l'} \ \forall c$$

- when an other parameter is not estimable, despite of the following constraints, it is set to zero. Percentages of explained variance ( $R^2$ ) are defined as the ratio of estimated residual variances of the two stacked models. For the percentage of variance explained by a single QTL, the compared models are without and with the QTL location. For the percentage of variance explained by the epistasis effect between two QTLs, the compared models are without and with the epistasis effect. For the global model percentage, they are without and with all QTL locations.

## 1.2.2 Multivariate analysis

A pleiotropic QTL test, following (Mangin et al., 1998) is implemented in MCQTL. It consists in estimating the variance-covariance residual matrix of the traits in a model with no genetic effect, computing the principal components of the traits for this estimated residual matrix, computing an approximate likelihood ratio test for each principal components and finally adding these tests at each locus. This procedure was proved to give a test asymptotically equivalent to the likelihood ratio test for pleiotropic QTL. It follows asymptotically a chi-squared distribution.

From the  $\chi^2$  test value, the  $-\log_{10}$  of the  $\chi^2$  test Pvalue is computed and an approximate LOD test value is obtained using  $\text{LOD} = 0.217 \ \chi^2$

Note that all the individuals that have one missing phenotypic value are discarded from the analysis.

Estimates of model parameters are the trait by trait estimates. Percentage of explained variance are not computed for pleiotropic QTL effects.

### 1.3 Detection method

We named CIM-, the easiest composite interval mapping method (Zeng, 1993) which consists in dropping out the genetic cofactors belonging to a chromosome while it is scanned.

iQTLm (Charcosset et al., 2000) is an iterative QTL mapping method that aims to find automatically a multiple QTL model. It works chromosome by chromosome and is composed of two steps, which algorithm is described in the following figure.

The first step estimates precisely the QTL positions, keeps in the model those with a Fisher test greater than a threshold and drops out of the model the others. The second step can add a QTL to the model. The whole process consists on a loop on chromosomes. During the whole process, the entry chromosome lists of positions are not updated. If for a chromosome, the entry list of positions is null iQTLm begins by the second step.

We observed with simulated data that a cycle (i.e. the entry and the output lists of positions exchange themselves from step to step) occurred in some rare situations. In that case, the method stops and the user is given a warning.

There is one iterative method named “simple” implemented in MCQTL to detect epistasis between loci. This iterative method is a forward method, however, to limit the computation time, the research of epistasis effects is not exhaustive in the genome x genome space. A non empty list of locations  $L$ , given by the user, which might be QTL positions detected by an additive model is necessary at the beginning. The model contains all the genotypic effects of loci in  $L$ . Epistasis Fisher test values are computed for all couples of loci belonging to  $L \times L$  union  $L \times \text{genome}$  (except for positions in the vicinity of locations in  $L$ ). The supremum of the Fisher test value is then compared to a threshold. If inferior, the procedure of detection is stopped. If superior the couple of loci is kept, its epistasis effect is added to the model, the list  $L$  is updated if necessary and a new genome scan begins.

### 1.4 Threshold computation

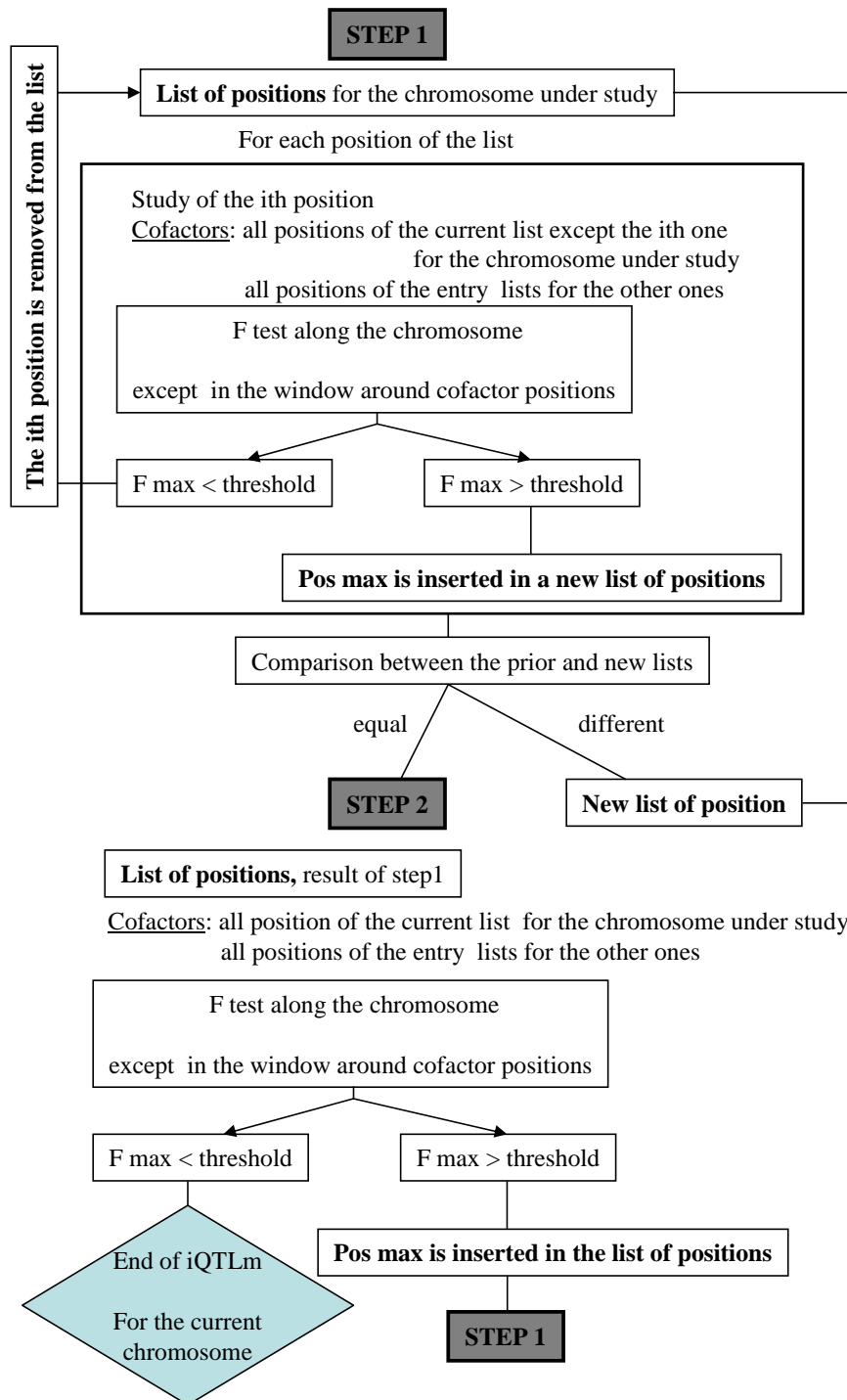
During a scan search of a putative QTL, i.e. when the location is moving along the genome and when the supremum over the locations is computed, it is necessary to have a threshold to declare that a QTL is detected. A resampling method by permutation of the trait data is implemented (Churchill and Doerge, 1994). The adaptation to multiple cross design is obtained by limiting permutations of the trait data to intra-family permutations.

The permutation procedure, proposed to approximate the threshold of a scan search for a putative epistasis effect, is a double stratified sampling. The first stratification is subject to the family index. The second stratification is subject to the genotype at a fixed locus, as described in Mangin et al. (1999). This second stratification permits to approximate the probability distribution of the Fisher test statistics given that a QTL is present at the fixed locus. This situation looks like the first iteration of the simple detection method when the list  $L$  is restricted to a single locus.

The following table presents a comparison of the threshold values obtained by the permutation

procedure in a additive model. It permits to show that the probability distribution function of the trait has a very small influence on the threshold values. Moreover, contrarily to threshold value of Fisher test or LOD test, the threshold value of the  $-\log_{10}$  of the Fisher test Pvalue depends only weakly on the multi-cross design.





*This figure illustrates the iQTLm algorithm.*

## Threshold ( $-\log_{10}(\text{Pvalue})$ ) value comparisons, additive model

Type I error	Map length	Marker density	Complete half-diallel 5 parent lines 10 F2 families		Star 5 parent lines 4 F2 families		Chain 5 parent lines 4 F2 families		Complete half-diallel 3 parent lines 3 families F2		2 parent lines 1 F2 family		Complete half-diallel 10 parent lines 45 families F2
			Gaussian	truncated	Gaussian	truncated	Gaussian	truncated	Gaussian	truncated	Gaussian	truncated	Gaussian
10 %	1 Morgan	20 cM	1.74	1.76	1.73	1.77	1.74	1.69	1.70	1.70	1.70	1.70	1.77
		5 cM	1.99	1.94	2.01	2.01	1.96	1.89	1.92	1.95	1.84	1.85	1.99
	2 Morgans	20 cM	2.03	2.08	2.01	2.11	2.06	2.00	2.01	2.02	2.02	1.98	2.01
		5 cM	2.29	2.30	2.25	2.38	2.29	2.26	2.25	2.30	2.25	2.22	2.26
5 %	1 Morgan	20 cM	2.10	2.08	2.05	2.12	2.09	2.03	2.06	2.08	2.06	2.01	2.12
		5 cM	2.32	2.30	2.33	2.37	2.29	2.22	2.30	2.33	2.14	2.19	2.30
	2 Morgans	20 cM	2.37	2.44	2.37	2.48	2.40	2.32	2.33	2.36	2.41	2.31	2.41
		5 cM	2.66	2.64	2.57	2.79	2.63	2.61	2.60	2.67	2.60	2.56	2.64
2.5 %	1 Morgan	20 cM	2.44	2.39	2.34	2.45	2.45	2.35	2.34	2.42	2.40	2.32	2.44
		5 cM	2.68	2.64	2.68	2.70	2.65	2.57	2.63	2.66	2.46	2.52	2.66
	2 Morgans	20 cM	2.66	2.72	2.68	2.81	2.78	2.66	2.70	2.67	2.75	2.65	2.75
		5 cM	2.95	2.96	2.93	3.17	2.95	2.95	2.93	3.03	2.84	2.94	3.00

100 descendants per F2 family

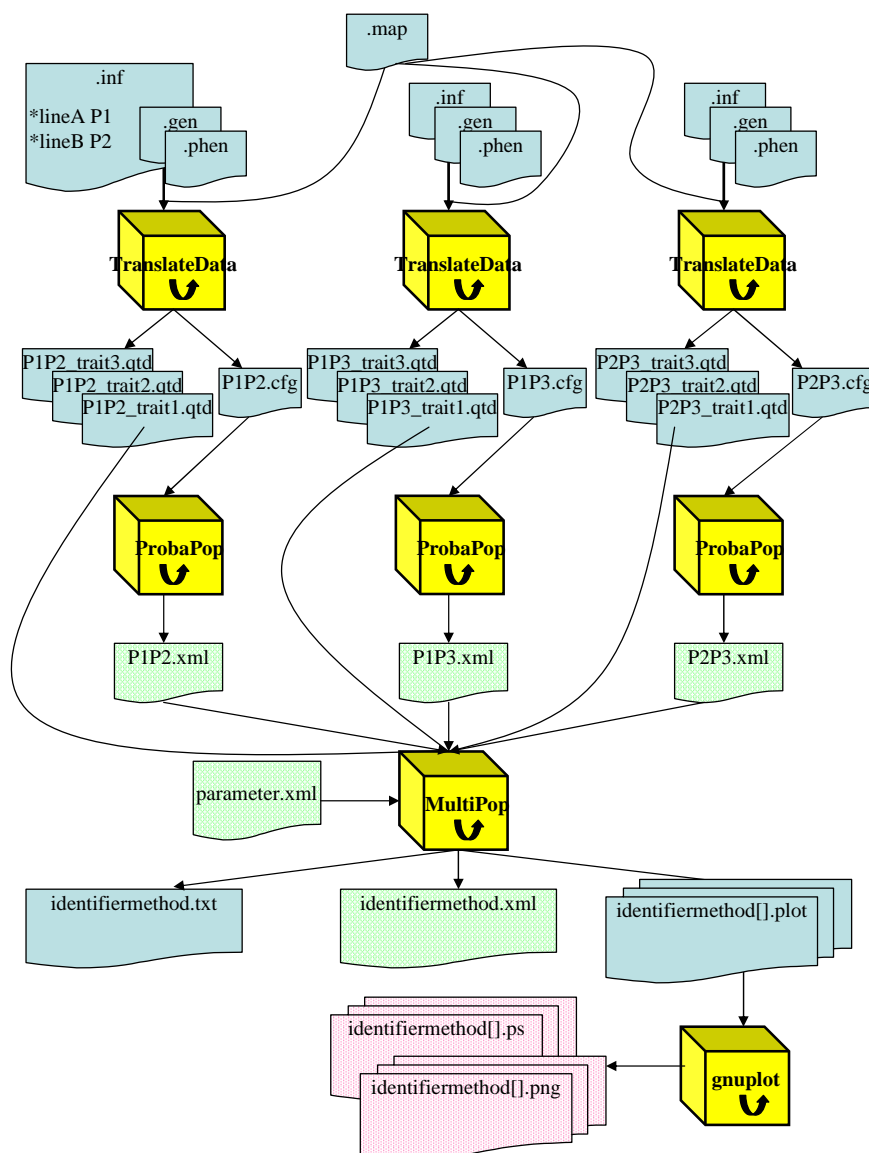
5000 permutations

Truncated Gaussian means that negative values were set to 0 in order to simulate a non-symmetrical and non-exponentially decreased trait.

**Conclusion:** Features that have a clear impact on threshold values are type I error, map length and marker density. Other features as non Gaussian probability distributions, **the type of designs, the number of parent lines and the number of families** have only a very **low impact**.

## 2 Architecture and command lines

MCQTL package is composed by three software applications. The first component, Translate-Data reads the data from files in format similar to MAPMAKER format. The second component, ProbaPop or F3toF7Proba computes QTL genotype probabilities given marker information at each chromosome location for each family and stores them in XML formatted files. The last component, Multipop builds the pooled model, fits the observations on genotype probabilities and produces outputs.



*This figure illustrates the architecture of MCQTL package for a half-diallel design with parent lines named P1, P2 and P3.*

The line commands are :

```
> TranslateData file1.map file2.phen file3.gen file4.inf
["/home/dir/"] [--strict]
```

The order of the .map, .phen, .gen, .inf entry files is mandatory. By default the output files are in the current directory. They can be redirected using a absolute path like "/home/dir/". If you add the "strict" option, TranslateData doesn't accept separation between individual genotypic data in the .gen file.

```
> ProbaPop cross.cfg x or > F3toF7Proba cross.cfg x
```

The entry file must be a .cfg file created by TranslateData. x is the progression step in centiMorgans for the scan between two markers. This value must be the same for all families involved together in MultiPop. Despite two different programs to compute conditional probabilities, F3 to F7 families can be involved in a joint analysis with back-cross, double-haploids, F2 or RIL.

```
> Multipop parameter.xml identifier
```

The entry file must be a .xml formatted file created by the user. It must be followed by an alphanumeric string, used to identify the output files.

### 3 Data entry files

Four ASCII files with mandatory extension .gen .phen .map and .inf are necessary to run the TranslateData program. The first three files are quite similar to parts of a MAPMAKER .raw file (Lincoln et al. 1993). They contain the marker information data and the type of cross (.gen), the quantitative trait data (.phen) and a consensus map (.map). In the last file (.inf), names of the parent lines are provided.

△ In all entry files when a name is expected, it must be a string beginning by an alphabetic character using alphabetic, numeric or underscore characters. Separator and punctuation characters are forbidden as for example space, carriage return, tabulator comma and point. You should keep in mind that upper and lower characters lead to a distinction.

#### 3.1 .gen file

The first line of this file begins by data type, a space and one of the following strings :

bc	for back-cross
dh	for double haploid
f2	for F <sub>2</sub>
ril	for recombinant inbred line
f3	for F <sub>3</sub>
f4	for F <sub>4</sub>
f5	for F <sub>5</sub>
f6	for F <sub>6</sub>
f7	for F <sub>7</sub>

The second line begins with 2 integer numbers separate by a space. They stand for, in the following order, the number of individuals and markers. This line can contain an alternative genotypic code.

Next, the marker information follow in a simple format. An asterisk (\*), a name, a space and genotypic data. "White space characters" (e.g. tabulation, carriage return, formfeed...) if they exist in genotypic data are ignored. The implemented genotypic code used A, B, H, C, D, and minus sign with the following meanings :

```

A homozygote for A allele
B homozygote for B allele
H heterozygote
C not a homozygote for A allele
D not a homozygote for B allele
- missing information

```

To specify an alternative genotypic code, a regular expression as X=A Y=B 0=- can be added at the end of the second line.

⚠ For F<sub>3</sub> to F<sub>7</sub> families the marker information "not homozygote for A/B allele" are rejected. Note that you can use C=- and/or D=- in order to analyse a file that contains C and/or D. The corresponding values are then treated as missing values.

⚠ Heterozygous genotypes are forbidden in RIL families. Note that you can use H=- in order to analyse a RIL genotype file that contains this code.

⚠ The number and the name of the markers must be the same as those given in the .map file. If a marker is not polymorphic in a family, all individuals must be scored with the missing information code.

#### example of a .gen file

```

data type f2
10 3 m=-
*mark1 ABHmH HBAAm
*mark2 mmmmm mmmmm
*mark3 AHBBm ABHmA

```

## 3.2 .phen file

For each trait data, the format is : an asterisk (\*), a trait name, a space and as many information as the number of individuals given in the .gen file. These individual information are either a real, an integer or a minus sign (for missing data). They are separated by a space.

#### example of a .phen file

```

*trait1 1.2 3 - -2 0.2 -1.56 2. - 2.1 -3.0
*trait2 0.2 3.2 4 2.0 1.4 - 0 0.45 0.998 1

```

### 3.3 .map file

This file contains a consensus map for families that are going to be analysed together with MultiPop. The format for a linkage group begins with an asterisk(\*), a group name, a space, the number of markers, a space. Then come alternatively, separated by a space, a marker name and the distance between the consecutive markers. Each linkage group is given in a single line.

△ The inter-marker distance must be a Haldane distance given in centiMorgans.

example of a .map file

```
*chrom1 4 mark1 20.2 mark2 15 mark3 17.98 mark4
*chrom2 6 mark5 3.4 mark6 4.6 mark7 9 mark8 3.3 mark9 0.0 mark10
```

### 3.4 .inf file

The first 2 lines contain the names of the parent lines. The third line is specific of a back-cross population and contains the name of the recurrent parent.

The first line begins with an asterisk (\*), `lineA`, a space and the name of the parent that has transmitted the A allele.

The second line is identical except that information is given for the B allele.

The third line begins with an asterisk (\*), `recurrent_line`, a space and the name of the recurrent parent. It must be one of the parent names.

△ Parent names are crucial for Multipop. Indeed, phenotypes of crosses that share a parent name are modelled with equal genetic parameters when the genetic parameters are assumed to be identical over the families.

example of a .inf file

```
*lineA parent1
*lineB parent2
*recurrent_line parent1
```

## 4 XML parameter file for MultiPop

*A quick look of XML format :*

*An XML formatted file is a set of tag, which permits to exchange data.*

*A tag is written between bracket as <TAG>. A start tag <TAG> must have a closing tag </TAG>, or it can be close in the same line as <TAG/>.*

*Syntax of tag is sensitive, <TAG> is different from < TAG> or from <Tag>.*

*A tag can have attributes as <TAG attribute-name="attribute-value" />. The value of an attribute needs to be quoted, even if it contains only numbers.*

*A tag can have children tags that permit to describe hierarchical data, <TAG> <CHILDREN-TAG/> </TAG>.*

*An XML formatted file begins by a root tag and its closing tag ends the file.*

MultiPop provides threshold computation, QTL detection methods and model estimation. These three functionalities are the possible values of the **OPERATION** tag. Some information are common and necessary for the three functionalities, they are described in the data and model section. Information specific to a single functionality are described in the corresponding section. File examples are given in the last section.

## 4.1 Data and model

Data information necessary to run MultiPop are :

- the name of the parent lines (one name by parent) which is given by  
`<LINE name=" . . . " />`  
 This name must be one of the parent names of .inf files.  
 ▲ Keep in mind that MCQTL does not correctly manage names of parent lines when a parent name is a multi-copy of another parent line, like P1 and P1P1.
- the name of XML formatted files created by ProbaPop and/or F3toF7Proba (one file by cross) which is given by  
`<XMLFILE name=" . . . " />`
- the name of the .qtd files created by TranslateData (one file by cross and by trait) which is given by  
`<QTDFILE name=" . . . " />`
- optionnally, if the LD part of MCQTL is to be used, the name of the .xml files created by the R package clusthaplo (one file by parent line) which is given by
- the name, as read in the entry map, of the studied chromosomes  
`<CHROMOSOME name=" . . . " study="yes" >`  
`</CHROMOSOME >` is the closing tag. It is necessary to set the limits of hierarchical information brought by **BONE**, **COFACTOR** and **QTL** children tags.
- the model, described using **MODEL** tags. A single **MODEL** tag is sufficient for single QTL search, two **MODEL** tags, one for the single QTL model and one for the epistasis model, are mandatory in case of epistasis effect search. Theses tags are detailed below.

The **MODEL** tag has got two mandatory attributes, **type** and **interpop**.

`<MODEL type="additive" interpop=" . . . " />`

For all chromosome locations, dominance parameters are set to 0.

`<MODEL type="dominance" interpop=" . . . " />`

For all chromosome locations, both additive and dominance parameters are involved in the model.

`<MODEL type="epistasy" interpop=" . . . " />`

For all chromosome locations, epistasis effects are involved in the model.

`<MODEL type=" . . . " interpop= "connected" />`

For all chromosome locations, QTL and genetic cofactor parameters are indential over the families.

`<MODEL type=" . . . " interpop= "disconnected" />`

For all chromosome locations, QTL and genetic cofactor parameters are intra-family parameters.

△ In this version only “connected” value for the attribute “interpop” is allowed when there are, in the model, epistasis effects.

<DESIGN > and </DESIGN > are the start and the closing root of the parameter files.

## 4.2 MCQTL\_LD model

To analyse your data with a MCQTL\_LD model, it is sufficient to add in the XML parameter file for MultiPop, one following tag per parent line to specify the name of the XML file created by clusthapo.

<LD\_FILE name=" . . . " >

△ The order of the parents within the LINE tags must be identical to the order within the LD\_file tags. Indeed, MCQTL attributes to the first parent of the LINE tags the first file of the LD\_file tags, to the second parent of the line tags the second file of the LD\_file tags, ... without any check.

## 4.3 Threshold computation

Two permutation methods are implemented in MCQTL to compute an empirical threshold value for the test detection, the following tag permits to run a threshold computation.

< OPERATION value=" **threshold** " replication=" . . . " method=" . . . " study=" . . . " >

The specific and necessary information for this operation is the desired quantile associated to a  $x$  chosen type I error ( $0 \leq x \leq 1$ ) which is specified by

<QUANTILE value=" **x** " >.

More than one QUANTILE tags are permitted.

<OPERATION value=" threshold " replication=" **x** " method=" . . . " study=" . . . " >

The real value  $x$  stands for the number of replications of the resampling process.

<OPERATION value=" threshold " replication=" . . . " method=" **genowide** " study=" . . . " >

A single putative QTL is analysed along all the chromosomes. A singleThe order of the parents within the LINE tags must be identical to the order within the LD\_file tags. Indeed, MCQTL attributes to the first parent of the LINE tags the first file of the LD\_file tags, to the second parent of the line tags the second file of the LD\_file tags, ... without any check. threshold value is computed for the whole genome.

<OPERATION value=" threshold " replication=" . . . " method=" **bychrom** " study=" . . . " >

A single putative QTL is analysed along each chromosome. The process is repeated for each chromosome and the results are one threshold value per chromosome.

<OPERATION value=" threshold " replication=" . . . " method=" **genowideforepistasy** " study=" . . . " >



This method is specific to the model with epistasis effects. Only, the simplest model is permitted. So there is no genetic cofactor and a single epistasis effect between a known QTL and other loci. The position of the known QTL, mandatory for the epistasis search, is given using the QTL tag, children tag of the CHROMOSOME tag (see below the model estimation section).

<OPERATION value="threshold" replication=" . . . " method="..." **study="by variable"** >  
Trait by trait threshold computations are performed. The study attribute is not mandatory, by default a trait by trait analysis is performed.

<OPERATION value="threshold" replication=" . . . " method="..." **study="multivariable"** >  
Threshold computations are performed for a pleiotropic test using all the traits.

#### 4.4 QTL detection

<OPERATION value="QTL detection" method=" . . . " threshold=" . . . " window=" . . . " study=" . . . " > is the tag which permits to run QTL detection analysis. This analysis is automatically followed by a model estimation and can be preceded by a cofactor search which tag is named **COFACTOR\_SELECTION**. By default, the cofactor selection operates on all the marker positions. However, it is possible to limit the search to some markers or positions, the **SKELETON** tag permits this pre-selection.

<COFACTOR\_SELECTION method="forward" threshold=" . . . " />

A forward stepwise method with all marker positions or all skeleton positions as regressors is performed.

<COFACTOR\_SELECTION method="backward" threshold=" . . . " />

For each chromosome, a backward method with all marker positions or all skeleton positions of the chromosome as regressors is performed.

<COFACTOR\_SELECTION method=" . . . " threshold="x" />

The real number  $x$  is the minimum test value necessary to select a regressor in the stepwise method.

$x$  should be inferior to the threshold value of the detection test.

<SKELETON value="automatic" distance="x" />

From the first marker on each chromosome, markers automatically chosen to build the skeleton have got a minimal inter distance of  $x$  centiMorgans.

<SKELETON value="manual" />

Users specify manually the positions that belong to the skeleton, using the **BONE** tag, children of the **CHROMOSOME** tag. If no bone is specified then all the markers of the chromosome become bones, by default.

< BONE position="x" />

The position must be one of the positions computed by ProbaPop.  $x$  is a real number that measures the distance in centiMorgans from the first marker of the chromosome.

Instead of using an automatic method of cofactor selection, users can declare manually the

position of genetic cofactors using the COFACTOR tag.

```
<COFACTOR position="x" />
```

The position must be one of the positions computed by ProbaPop. x is a real number that measures the distance in centiMorgans from the first marker of the chromosome.

```
<OPERATION value="QTL detection" method="CIM-" threshold="..." window="..." study="...">
```

A single QTL model by chromosome, using declared cofactors on the other chromosomes is tested.

At the end of the process, a model is built with a single QTL position by chromosome and the model parameters are estimated. For each chromosome, the QTL position corresponds to the chromosome maximum test value provided that this value exceeds the threshold.

```
<OPERATION value="QTL detection" method="iQTLm" threshold="..." window="..." study="...">
```

Chromosome by chromosome, the iterative QTL method of detection as described in the section method is performed.

At the end of the process, a model is built with the detected QTL positions and the model parameters are estimated.

```
<OPERATION value="QTL detection" method="simple" threshold="..." window="..." study="..." >
```

An iterative genomewide search for epistasis effects, as described in the section model, is performed. The mandatory list of QTL positions to begin the epistasis search is described using QTL tag, children of the CHROMOSOME tag (see below the model estimation session).

At the end of the process, a model is built with detected epistasis effects and the model parameters are estimated.

```
<OPERATION value="QTL detection" method="..." threshold="x" window="..." study="..." >
```

The real number x is the minimal test value to detect a QTL or an epistasis effect between QTLs.

```
<OPERATION value="QTL detection" method="..." threshold="..." window="x" study="..." >
```

The real number x is a distance in centiMorgans. It is the left or right length surrounding a QTL or genetic cofactor positions where the search of another QTL is forbidden. It permits to avoid any problem with closely linked QTLs. So, the window length is 2 x centiMorgans.

```
<OPERATION value="QTL detection" method="..." threshold="..." window="..." study="by variable" >
```

Trait by trait analyses are performed. The study attribute is not mandatory, by default a trait by trait analysis is performed.

```
<OPERATION value="QTL detection" method="..." threshold="..." window="..." study="multivariable" >
```

A pleiotropic test is performed using all the traits.

## 4.5 Model estimation

<OPERATION value="model estimation" study=" . . . " > is the tag which permits to estimate the model. The model is described using the QTL tag, children of the CHROMOSOME tag and the EPISTASY tag.

<QTL position=" x " />

The position must be one of the positions computed by ProbaPop. x is a real number that measures the distance in centiMorgans from the first marker of the chromosome.

<EPISTASY firstposition=" x1 " firstchromosome=" c1 " secondposition=" x2 " secondchromosome=" c2 " />

C1 and c2 are chromosome names, previously given by CHROMOSOME tag. Positions must be one of the positions computed by ProbaPop. x1 and x2 are real numbers that measure the distance in centiMorgans from the first marker of the chromosome.

By default the LOD support interval is built using the value 1 for back-cross and double haploid, 1.5 for F<sub>2</sub> population and 2 for recombinant inbred lines. This default value can be change by using the following tag.

<LOD\_SUPPORT value=" x " />

x is the value of LOD decrease that defines the LOD support interval around the location of the test supremum.

This tag can be used with QTL detection methods, as well.

## 4.6 Tag inactivation

The following tags can be disabled.

<CHROMOSOME name=" . . . " study=" no " >

The chromosome is dropped out of the map.

<COFACTOR\_SELECTION method=" no " threshold=" . . . " />

<SKELETON value=" no " distance=" . . . " />

## 4.7 Examples

[threshold.xml](#) illustrates a threshold computation for a half diallel design.

```
<DESIGN >
  <LINE name="P1" />
  <LINE name="P2" />
  <LINE name="P3" />
  <XMLFILE name="P1P2.xml" />
  <XMLFILE name="P1P3.xml" />
  <XMLFILE name="P2P3.xml" />
  <QTDFILE name="P1P2_trait1.qtd" />
  <QTDFILE name="P1P3_trait1.qtd" />
  <QTDFILE name="P2P3_trait1.qtd" />
  <MODEL type="additive" interpop="connected" />
  <CHROMOSOME name="chrom1" study="yes" >
  </CHROMOSOME >
  <CHROMOSOME name="chrom2" study="yes" >
  </CHROMOSOME >
  <OPERATION value="threshold" replication="1000" method="genomewide" >
    <QUANTILE value="0.05" />
    <QUANTILE value="0.01" />
    <QUANTILE value="0.005" />
  </OPERATION >
</DESIGN >
```

[detection.xml](#) illustrates a QTL detection using *iQTLm* method for a half diallel design.

```
<DESIGN >
  <LINE name="P1" />
  <LINE name="P2" />
  <LINE name="P3" />
  <XMLFILE name="P1P2.xml" />
  <XMLFILE name="P1P3.xml" />
  <XMLFILE name="P2P3.xml" />
  <QTDFILE name="P1P2_trait1.qtd" />
  <QTDFILE name="P1P3_trait1.qtd" />
  <QTDFILE name="P2P3_trait1.qtd" />
  <MODEL type="additive" interpop="connected" />
  <SKELETON value="manual" />
  <CHROMOSOME name="chrom1" study="yes" >
    <BONE position="0" />
    <BONE position="15.8" />
  </CHROMOSOME >
</DESIGN >
```

```

    <BONE position="30.33" />
    <BONE position="50.2" />
    <BONE position="75.341" />
  </CHROMOSOME >
  <CHROMOSOME name="chrom2" study="yes" >
  </CHROMOSOME >
  <COFACTOR_SELECTION method="forward" threshold="2" />
  <OPERATION value="QTL detection" method="iQTLm" threshold="2.7"
    window="5" >
  </OPERATION >
</DESIGN >

```

estimation.xml illustrates a model estimation with 3 detected QTL for a half diallel design.

```

<DESIGN >
  <LINE name="P1" />
  <LINE name="P2" />
  <LINE name="P3" />
  <XMLFILE name="P1P2.xml" />
  <XMLFILE name="P1P3.xml" />
  <XMLFILE name="P2P3.xml" />
  <QTDFILE name="P1P2_trait1.qtd" />
  <QTDFILE name="P1P3_trait1.qtd" />
  <QTDFILE name="P2P3_trait1.qtd" />
  <LOD_SUPPORT value="3" />
  <MODEL type="additive" interpop="connected" />
  <CHROMOSOME name="chrom1" study="yes" >
    <QTL position="17.3" />
    <QTL position="78" />
  </CHROMOSOME>
  <CHROMOSOME name="chrom2" study="yes" >
    <QTL position="86.341" />
  </CHROMOSOME>
  <OPERATION value="model estimation" >
  </OPERATION >
</DESIGN >

```

threshold\_LD.xml illustrates a threshold computation for a half diallel design with MCQTL\_LD model. LD files were created by clusthaplo.

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<DESIGN>
  <LINE name="P1" />

```

```

<LINE name="P2" />
<LINE name="P3" />
<LINE name="P4" />
  <MODEL interpop="connected" type="additive"/>
<XMLFILE name="P1P2.xml" />
<XMLFILE name="P1P3.xml" />
<XMLFILE name="P1P4.xml" />
<XMLFILE name="P2P3.xml" />
<XMLFILE name="P2P4.xml" />
<XMLFILE name="P3P4.xml" />
<QTDFILE name="P1P2_trait01.qtd"/>
<QTDFILE name="P1P3_trait01.qtd"/>
<QTDFILE name="P1P4_trait01.qtd"/>
<QTDFILE name="P2P3_trait01.qtd"/>
<QTDFILE name="P2P4_trait01.qtd"/>
<QTDFILE name="P3P4_trait01.qtd"/>
<LD_FILE name="LD_file_P1.xml" />
<LD_FILE name="LD_file_P2.xml" />
<LD_FILE name="LD_file_P3.xml" />
<LD_FILE name="LD_file_P4.xml" />
<CHROMOSOME name="ch1" study="yes" />
<CHROMOSOME name="ch2" study="yes" />
<CHROMOSOME name="ch3" study="yes" />
<CHROMOSOME name="ch4" study="yes" />
<CHROMOSOME name="ch5" study="yes" />
<CHROMOSOME name="ch6" study="yes" />
<CHROMOSOME name="ch7" study="yes" />
<CHROMOSOME name="ch8" study="yes" />
<CHROMOSOME name="ch9" study="yes" />
<CHROMOSOME name="ch10" study="yes" />
<OPERATION value="threshold" replication="1000" method="genowide" />
<QUANTILE value="0.1" />
<QUANTILE value="0.05" />
</DESIGN>

```

*threshold\_LD.xml illustrates a QTL detection with epistasis effect for a half diallel design with MCQTL\_LD model. Two model tags are mandatory. An additive QTL (mandatory) was previously detected by the iQTLm detection method in chromosome 1 at position 38.*

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<DESIGN>
  <LINE name="P1" />
  <LINE name="P2" />
  <LINE name="P3" />

```

```

<LINE name="P4" />
  <MODEL interpop="connected" type="additive"/>
  <MODEL interpop="connected" type="epistasy"/>
<XMLFILE name="P1P2.xml" />
<XMLFILE name="P1P3.xml" />
<XMLFILE name="P1P4.xml" />
<XMLFILE name="P2P3.xml" />
<XMLFILE name="P2P4.xml" />
<XMLFILE name="P3P4.xml" />
<QTDFILE name="P1P2_trait01.qtd"/>
<QTDFILE name="P1P3_trait01.qtd"/>
<QTDFILE name="P1P4_trait01.qtd"/>
<QTDFILE name="P2P3_trait01.qtd"/>
<QTDFILE name="P2P4_trait01.qtd"/>
<QTDFILE name="P3P4_trait01.qtd"/>
<CHROMOSOME name="chrom1" study="yes">
<QTL position="38"/>
</CHROMOSOME>
<CHROMOSOME name="ch2" study="yes"/>
<CHROMOSOME name="ch3" study="yes"/>
<CHROMOSOME name="ch4" study="yes"/>
<CHROMOSOME name="ch5" study="yes"/>
<CHROMOSOME name="ch6" study="yes"/>
<CHROMOSOME name="ch7" study="yes"/>
<CHROMOSOME name="ch8" study="yes"/>
<CHROMOSOME name="ch9" study="yes"/>
<CHROMOSOME name="ch10" study="yes"/>
<OPERATION value="simple" threshold="2.7"
  window="5" />
</DESIGN>

```

## 5 Internal files

TranslateData creates 2 types of internal files. One of them, with extension .cfg, contains family information and genotype data. Other files are related to the trait data, their extension is .qtd. There is one .qtd file by quantitative trait.

ProbaPop creates one file. Its extension is .xml and it is an XML formatted file. It contains genotype probabilities of a putative QTL given marker information at each genome location.

The name of each internal file begins with the cross name which is the concatenation of parent names given by `lineA` and `lineB` in .inf entry file. For .qtd files, the cross name is followed by an underscore character and the trait name read in the .phen entry file.

## 6 Output files

Output files are different according to the analyses named threshold, backward, forward, CIM-, iQTLm, simple and estimation. They correspond respectively to the following tags :

```
<OPERATION value="threshold" replication="..." method="..."
study="..."~>
<COFACTOR_SELECTION method="backward" threshold="..." />
<COFACTOR_SELECTION method="forward" threshold="..." />
<OPERATION value="QTL detection" method="CIM-" threshold="..."
window="..." study="..." >
<OPERATION value="QTL detection" method="iQTLm" threshold="..."
window="..." study="..." >
<OPERATION value="QTL detection" method="simple" threshold="..."
window="..." study="..." >
<OPERATION value="model estimation" study="..." >
```

Extensions for output files are .txt, .xml, .plot, .ps and .png. They all begin by the identifier given in MultiPop line command.

The .txt file contains information that depends on the analysis. This file is produced by threshold computations, backward and forward cofactor selections or CIM- and iQTLm detection methods.

In the following table, the information printed in the .txt files is described.

method	.txt
threshold	estimated quantiles
backward	marker positions selected in the final model
forward	marker positions selected in the final model
CIM-	per chromosome, position of the test supremum if > threshold
iQTLm	per chromosome, list of successive models
simple	positions in epistasis selected in the final model

The .xml .plot, .png and .ps files are produced by the model estimation procedure which follows automatically detection methods. The .xml file contains the results of a model estimation replaced in the original map. The .plot files contain a chromosome or a part of chromosome scan, .ps and .png are the corresponding graphic files obtained by gnuplot.

For each chromosome, MultiPop produce as many chromosome scans as QTL locations in the chromosome. Indeed, each location in turn is dropped out of the model and a test is computed around the dropped location. In the corresponding .ps and .png graphic files, these scans are drawn all together. However, if there is no QTL position in a chromosome, a chromosome scan for a single putative QTL is produced.

A .plot file contains, in two columns, the test position following by the test value. There are two files per chromosome and single QTL in the model. One for the Fisher or  $\chi^2$  test and the other



for the approximate LOD test. The name of this plot is composed with the identifier, the name of the trait, the chromosome name and the QTL index.

The .xml output file contains parameter estimates of the model. This model is the manually specified model when the value of the OPERATION tag is "model estimation". After CIM-detection, the model contains no more than one QTL per chromosome. For each chromosome, it is located at the position of the Fisher test supremum provided that this test value is superior to the threshold. After iQTLm detection, the model is built with the convergent positions for each chromosome. After the simple method for epistasis search, the model contains the QTLs declared in the initial mandatory list and all the QTLs detected in epistasis.

The file is composed of two parts. The first one, delimited by GENETIC\_MAP tag, concerns the genetic consensus map in which QTL are included between markers. The second one, delimited by ESTIMATION tag, contains model parameter estimates.

#### result.xml

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<TRAIT name="trait1">
  <GENETIC_MAP name="default">
    <CHROMOSOME name="chrom1">
      <MARKER name="mark1" position="0" />
      <MARKER name="mark2" position="20.2" />
      <QTL position="30.1">
        <FTEST>17.864</FTEST>
        <R2>0.156889</R2>
        <NUMERATOR_dd1>2</NUMERATOR_dd1>
        <DENOMINATOR_dd1>192</DENOMINATOR_dd1>
        <LOD_SUPPORT local_max_position="30.1">
          <INFERIOR_POSITION value="23.5572" />
          <SUPERIOR_POSITION value="35.5978" />
        </LOD_SUPPORT>
      </QTL>
      <MARKER name="marker3" position="35.2" />
      .
      .
      .
    </CHROMOSOME>
    .
    .
    .
  </GENETIC_MAP>
  <ESTIMATION>
    <RESIDUAL_VARIANCE>0.98989</RESIDUAL_VARIANCE>
    <R2_GLOBAL>0.371195</R2_GLOBAL>
```

```

    <CROSS_EFFECT>
      <EFFECTS genetic_background="P1P2">-0.153123</EFFECTS>
      .
      .
    </CROSS_EFFECT>
    <QTL_EFFECT Chromosome_Origin="chrom1" Position="30.1">
      <EFFECTS additive="P1_P1P2P3">-0.411043</EFFECTS>
      .
      .
    </QTL_EFFECT>
    .
    .
  </ESTIMATION>
</TRAIT>

```

#### result\_LD.xml

The section dedicated to model parameter is slightly changed when using MCQTL\_LD. The EFFECTS tag gives the estimates of ancestral alleles and the PREDICTED tag gives the estimates of parent line alleles. The following example concerned a QTL detected on a map position where parent line P1 and P2 shared the same ancestral allele.

```

      .
      .
    </CROSS_EFFECT>
    <QTL_EFFECT Chromosome_Origin="chrom1" Position="30.1">
      <EFFECTS additive="P1_P1P2P3">-0.411043</EFFECTS>
      <EFFECTS additive="P2_P1P2P3">0</EFFECTS>
      .
      .
      .
      <PREDICTED additive="P1">-0.411043</PREDICTED>
      <PREDICTED additive="P2">-0.411043</PREDICTED>
      .
      .
      .

```

Tags have been chosen to be as explicit as they can. R2 tag stands for the percentage of variance explained by a QTL and R2\_GLOBAL tag for the percentage of variance explained by the model. The local\_max\_position attribute of a LOD\_SUPPORT tag is the position around the

corresponding QTL location where the Fisher test supremum is found. Ideally, this position should be the QTL location but it could be slightly different. The LOD support interval is computed around this local\_max\_position.

△ Note that the estimated values of additive effects are the estimated values of the additive allelic effects. The name of an additive allelic parameter is composed by the parent name which transmits the allele (P1 in the previous example), underscored by names of parent connected to P1.

## 7 User interface

A java user interface is provided with MCQTL package. The command line to run this interface is MCQTL\_Interface. Its aim is to help users throw the main steps of MCQTL package and especially to create correct xml files for Multipop component. However, this user interface does not allowed all the new features of the MCQTL version V4.0.

A quick tutorial for this user interface can be found in the MCQTLTutorial.pdf file.

## 8 References

Charcosset A., Mangin B., Moreau L., Combes L., Jourjon M.-F., Gallais A. 2001. Heterosis in maize investigated using connected RIL populations. *Quantitative genetics and breeding methods : the way ahead*. Les colloques n° 96, INRA Editions, Paris.

Churchill G. A., Doerge R.W. 1994. Empirical threshold values for quantitative trait mapping. *Genetics*, **138**, 963-971.

Haley C.S., Knott S.A. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69**, 315-324.

Li J., Jiang T. 2005. Haplotype-based linkage disequilibrium mapping via direct data mining. *Bioinformatics*, **21**, 4384-4393.

Leroux D., Rahmani A., Jasson S., Ventelon-Debout M., Louis F., Moreau L., Mangin B. 2014 Clusthaplo : a plugin for MCQTL to enhance QTL detection using ancestral alleles in multi-cross design. *TAG*, **127**, 921-933.

Lincoln S.E., Daly M.J., Lander E.S. 1993 Constructing Genetic Linkage Maps with MAPMAKER/EXP Version 3.0. A Tutorial and Reference Manual. *A Whitehead Institute for Biomedical Research Technical Report Third Edition* <http://linkage.rockefeller.edu/soft/mapmaker>

Mangin B., Thoquet P., Grimsley N.H., 1998. Pleiotropic QTL analysis. *Biometrics*, **54**, 88-99.

Mangin B., Thoquet P., Olivier J., Grimsley N.H., (1999). Temporal and multiple quantitative trait loci analyses of resistance to bacterail wilt in tomato permit the resolution of linked loci. *Genetics*, **151** : 1165-1172.

Zeng Z.-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. U.S.A.*, **136**, 1457-1468.

## A List of errors of TranslateData

### MISSING ARGUMENTS

USAGE : TranslateData file.map file.phen file.gen file.inf

ALTERNATE USAGE : TranslateData file.map file.phen file.gen file.inf Output-Path

### TOO MANY ARGUMENTS

USAGE : TranslateData file.map file.phen file.gen file.inf

ALTERNATE USAGE : TranslateData file.map file.phen file.gen file.inf Output-Path

ERROR ON FILE ORDER (right order : .map, .phen, .gen, .inf)

ERROR : UNKNOWN KIND OF FILES

ERROR : CANNOT CREATE OUTPUT FILE please verified UNIX file permissions and disk quota

ERROR : CANNOT WRITE OUTPUT FILE please verified UNIX file permissions and disk quota

ERROR ON DATA TYPE POPULATION, LINE 1 IN <...>.gen

ERROR ON NUMBER OF INDIVIDUALS, LINE 2 IN <...>.gen

ERROR ON NUMBER OF MARKERS, LINE 2 IN <...>.gen

ERROR ON ALTERNATE GENOTYPIC CODE, LINE 2 IN <...>.gen

ERROR : MISSING INFORMATION IN <...>.gen probably due to a missing carriage return

ERROR ON NUMBER OF MARKER GENOTYPIC DATA <marker name> IN <...>.gen probably due to invisible control character

ERROR : DIFFERENCE BETWEEN THE NUMBER OF MARKERS AND THE NUMBER OF MARKER LINES IN <...>.gen

ERROR IN <...>.gen GENOTYPIC CODE : <...> IS NOT DEFINED

ERROR : DIFFERENCE BETWEEN THE NUMBER OF INDIVIDUALS AND THE NUMBER OF TRAIT DATA IN <...>.phen

ERROR : THERE IS NO INFORMATION IN <...>.map OR IT ISN'T A RIGHT MAPMAKER FILE

ERROR : MISSING MARKER NAMES ON CHROMOSOME <chromosome number> IN <...>.map probably due to a marker name that does not begin by a letter

ERROR ON NUMBER OF MARKERS IN THE <chromosome number> CHROMOSOME IN <...>.map probably due to a marker name that does not begin by a letter

ERROR : MISSING INFORMATION IN <...>.map probably due to a missing carriage return

ERROR : MISSING INTERMARKER DISTANCE FOR CHROMOSOME <chromosome number> IN <...>.map

ERROR ON NUMBER OF DISTANCES FOR CHROMOSOME <chromosome number> IN <...>.map

ERROR : RECURRENT LINE FOR BACK-CROSS POPULATION MISSING IN .inf FILE

ERROR : MARKER <marker name> NOT EXIST IN <...>.gen

INTERNAL ERROR : WRITE ERROR 1 please report bug...

INTERNAL ERROR : WRITE ERROR 2 please report bug...

INTERNAL ERROR : BAD REGEXP please report bug ...

L'unité de Biométrie et Intelligence Artificielle de Toulouse a pour mission de développer et d'appliquer des méthodes de statistiques et d'intelligence artificielle dans le cadre des grands axes de recherche de l'INRA. L'unité travaille en particulier sur l'analyse par modélisation, simulation et optimisation de systèmes en agronomie, écologie, épidémiologie, gestion forestière, et dans le domaine de la biologie sur la localisation et l'identification d'éléments fonctionnels dans les génomes des bactéries, plantes et animaux, aux niveaux génétique, moléculaire et de l'expression de gènes.

Ces recherches s'accompagnent d'une activité de production de logiciels pour leur valorisation et d'une activité de formation pour leur diffusion.

*The Biometry and Artificial Intelligence laboratory at Toulouse is part of INRA (French National Institute for Agricultural Research). Its mission is to develop and apply statistical and artificial intelligence methodologies to the specific research domains of INRA. The laboratory develops methods for modeling, simulating and optimizing systems in agronomy, ecology, epidemiology, forest management. In biology, models and methods are developed that aim at locating and characterizing functional elements inside the genomes of bacteria, plants and animals, at the genetic, sequence and expression data levels. This also includes teaching and software development.*

INRA - UBIA, BP 52627 Auzeville F-31326 Castanet Tolosan cedex  
Tél : +33(0)5.61.28.52.75 - Fax : +33(0)5.61.28.53.35 - <http://mia.toulouse.inra.fr>

Image de synthèse : Benjamin Auriol  
Publication : Jérôme Juille (INRA - SDAR)

