Documentation for **OptiMAS**:

a decision support tool for marker-assisted assembly of diverse alleles

Version 1.4.1

F. Valente, F. Gauthier, N. Bardol, G. Blanc, J. Joets,
A. Charcosset & L. Moreau

Code by Fabio Valente and Franck Gauthier with contribution from Guylaine Blanc

Project started in December 2009

Emails: fvalente@moulon.inra.fr, moreau@moulon.inra.fr, charcos@moulon.inra.fr

Website: http://moulon.inra.fr/optimas

Address for correspondence:

UMR de Génétique Végétale
INRA - Univ Paris-Sud - CNRS - AgroParisTech
Ferme du Moulon, F-91190 Gif-sur-Yvette, France

This program was funded by the Generation Challenge Program.

Contents

1 Introduction

1.1 Aims

With the increasing use of markers in breeding programs, it is important to develop decision support tools to help breeders in implementing their Marker-Assisted Selection (MAS) project. OptiMAS has been developed with the possibility to consider a multi-allelic context, which opens new prospects to further accelerate genetic gain by assembling favorable alleles issued from diverse parents.

1.2 Principles

Algorithms have been deployed to trace parental QTL alleles identified as favorable throughout selection generations, using information given by markers located in the vicinity of the estimated QTL positions. Using these results, probabilities of allele transmission are computed in different MAS schemes and mating designs (intercrossing, selfing, backcrossing, double haploids, RIL) with the possibility of considering generations without genotypic information. Then, strategies are proposed to select the best plants and to efficiently intermate them based on the expected value of their progenies.

1.3 Functions

OptiMAS includes in a Graphical User Interface (GUI) three different modules, corresponding to the different steps of a selection program (see Fig. 1):

- Step 1: Computation of genotypic probabilities - Estimation of genetic values
The tool provides for each candidate individual the probabilities of being homozygous or heterozygous for parental alleles at each QTL. Based on the classification of parental alleles into favorable and unfavorable categories, a molecular score (expected probability of favorable allele) is computed for each QTL. Individual molecular scores are then combined into a global genetic value by assigning identical or different weights to QTL. A colored view of the molecular score table is displayed to identify more easily QTL for which a given individual is already fixed or not. Graphs are generated to show the distribution of several indicators (QTL molecular scores at individual QTL, global genetic values…) and their evolution over the different cycles of selection.

- Step 2: Selection of individuals
Different options are available to select candidates. Truncation selection can be performed based on (i) the above described genetic value, or (ii) a utility criterion which considers the probabilities of obtaining superior progenies following gametic segregation. QTL complementation selection (Hospital *et al.*, 2000) can be performed in order to prevent the loss of favorable allele(s). Different lists of selected plants can be compared via graphs showing the distribution of above mentioned indicators. All lists can be adjusted manually. A visualization tool of the pedigree of the selected plants is also provided.

- Step 3: Identification of crosses to be made among selected individuals
We implemented three simple cases: (i) half-diallel between selected candidates, (ii) "better-half" strategy (Bernardo *et al.*, 2006) which consists of avoiding crosses between selected individuals with the lowest scores or (iii) factorial design between two lists of selected plants. Constraints on the contribution of parents or on the maximum number of crosses to be done

can be applied. In each case, OptiMAS computes the expected molecular score of the progeny. A graph is automatically generated showing a view of the crosses to be done.

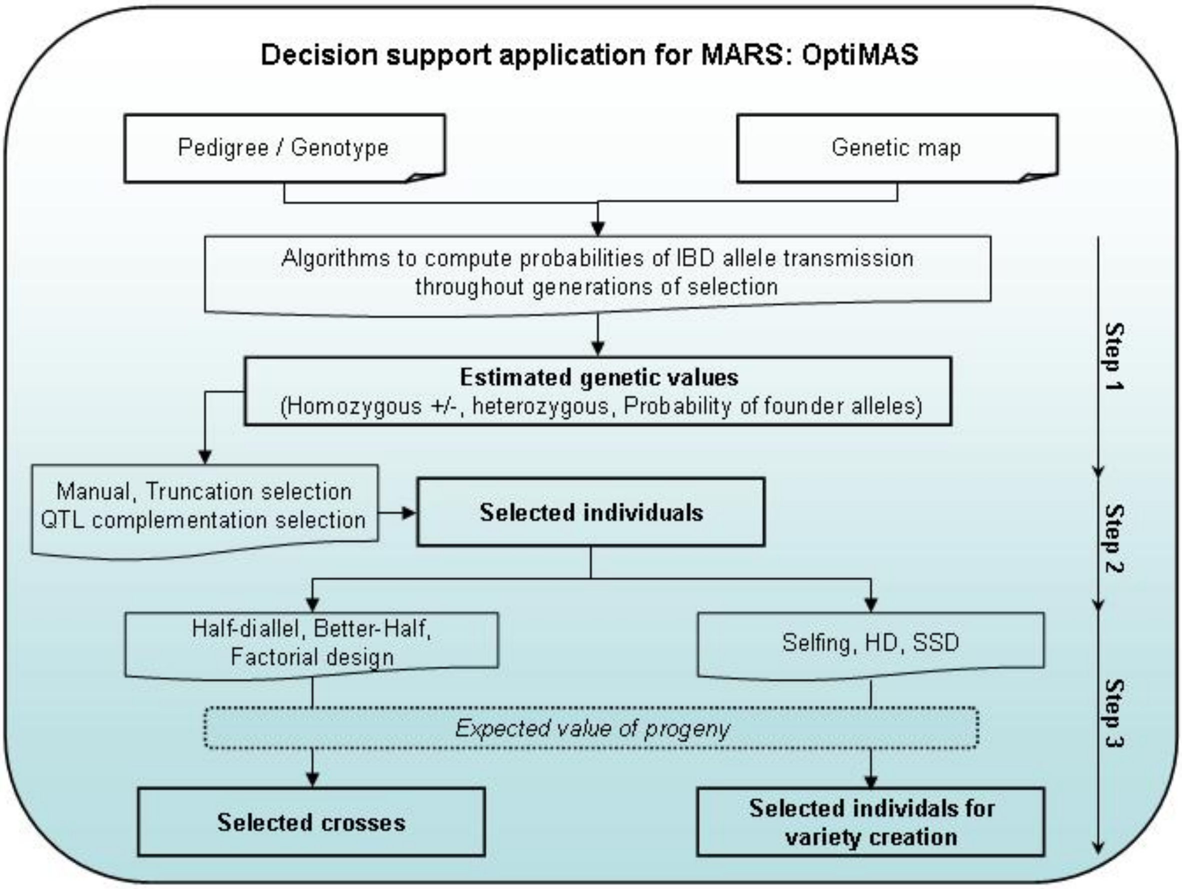

Decision support application for MARS: OptiMAS
Pedigree / Genotype
Genetic map
Algorithms to compute probabilities of IBD allele transmission throughout generations of selection
Estimated genetic values
(Homozygous +/-, heterozygous, Probability of founder alleles)
Manual, Truncation selection
QTL complementation selection
Selected individuals
Half-diallel, Better-Half, Factorial design
Selfing, HD, SSD
Expected value of progeny
Selected crosses
Selected individals for variety creation
Step 1
Step 2
Step 3

Figure 1: OptiMAS GUI functionalities

2 Installation procedure

OptiMAS installable versions are distributed to run under most modern **GNU/Linux**, **Windows (XP/7)** and **MacOSX** (10.5 or later with Intel processor) systems. Please note that extensive testing has only been done under Linux. Ready-to-use binaries installation packages and the source code, which you are welcome to attempt to compile on your favorite platform, are available via http://moulon.inra.fr/optimas.

Two versions of the tool have been developed. The first one, called "optimas" manages computationally intensive processes for step 1. It runs in command line and is written in C-ANSI language. The second version integrates the C program and additional functionalities to display results and facilitate breeding decisions within a Graphical User Interface named "optimas_gui" coded in C++ using Qt, Qwt & Graphviz libraries.

2.1 OptiMAS in command line

If you are dealing with huge amount of data or complex MARS schemes, it could be better for you to use OptiMAS in command line (on a server for example) and then reload the results folder via the GUI (see section 5.5).

2.1.1 Windows (32 bits)

The executable comes in a zipped file. Extract it with your favorite file archiver software (e.g. 7-zip). This will create a new directory called "optimas_cmd_win". Move to this directory via the terminal application (click on ***Start > Execute > type "cmd" > OK***) before attempting to run the program (for example):

```
> cd Desktop\optimas_cmd_win
```

To run the program, you must supply 4 input files. Instructions for how to prepare the input files are given below (see section 3). Input file examples (input directory) are supplied with the software. You can run the program on the test example data supplied by typing (see section 4.1 for more details about the parameters and options):

```
> optimas.exe input\blanc.dat input\blanc.map 0.000001 0.0 output\run_blanc 0 [verb]
```

Make sure that both "optimas.exe" and the "input" folder with the examples are in the current directory. The program will output a summary of the results in the folder output\run_blanc. Open the file "tab_scores.txt" to see the genotypic values calculated for the plants present in the dataset.

2.1.2 Linux (32/64 bits)

To build and install "optimas" in command line on your system, extract the zipped file "optimas_cmd_linux.zip" by typing (for example):

```
$ unzip optimas_cmd_linux.zip
```

This will create a new directory called "optimas_cmd_linux". Then, open a terminal, move to this directory before attempting to run the program and run the installation shell (bash) script:

```
$ ./install_optimas_on_linux.sh --no-gui
```

As **root** (or sudoer), it will perform an installation for all users in /usr/local/bin/optimas. The input/output file examples will be stored respectively in /usr/local/share/OptiMAS/input and /usr/local/share/OptiMAS/output.

As a **common user**, it will perform a local install in user's personal directory $HOME/bin/optimas. The input/output file examples will be stored respectively in $HOME/OptiMAS/input and $HOME/OptiMAS/output.

You can now run optimas from a terminal by typing optimas or /usr/local/bin/optimas or $HOME/bin/optimas.

Note: it is not necessary to specify the complete path if the binary "optimas" is present in the PATH environment variable.

To run the program, you must supply 4 input files. Instructions for how to prepare the input files are given below (see section 3). Input file examples (input directory) are supplied with the software. You can run the program on the test data supplied by typing as an example (see section 4.1 for more details about the parameters and options):

```
$ optimas $HOME/OptiMAS/input/blanc.dat $HOME/OptiMAS/input/blanc.map 0.000001
0.0 $HOME/OptiMAS/output/run_blanc 0 [verb]
```

The program will output a summary of the results in the folder "run_blanc". Open the file "tab_scores.txt" to see the genotypic values calculated for the plants present in the dataset.

To uninstall OptiMAS, run the script uninstall_optimas.sh located in the same directory as optimas.

2.1.3 Mac OS X (32 bits)

The executable comes in a zipped file. Extract it with your favorite file archiver software (e.g. Archive Utility). This will create a new directory called "optimas_cmd_mac". Move to this directory via the terminal application (***Applications*** > ***Utilities*** > ***Terminal***) before attempting to run the program (for example):

```
$ cd Desktop/optimas_cmd_mac
```

To run the program, you must supply 4 input files. Instructions for how to prepare the input files are given below (see section 3). Input file examples (input directory) are supplied with the software. You can run the program on the test data supplied by typing (see section 4.1 for the details about the parameters and options):

```
$ optimas input/blanc.dat input/blanc.map 0.000001 0.0 output/run_blanc 0 [verb]
```

Make sure that both "optimas" and the "input" folder with the examples are in the current directory. The program will output a summary of the results in the folder output\run_blanc. Open the file "tab_scores.txt" to see the genotypic values calculated for the plants present in the dataset.

2.2 OptiMAS Graphical User Interface (GUI)

As ready-to-use binaries installation packages are provided for Windows and MacOSX platforms, we will only describe the steps for compiling OptiMAS binaries on GNU/Linux systems. More details on the building and installation instructions from the sources for all supported systems are described in the INSTALL file.

2.2.1 Windows (32 bits)

To set up OptiMAS on your Windows computer, double-click on the install file i.e. optimas_win_x86_v1_12_06_01.exe (i.e. version 1, 1st June 2012) and follow the Wizard procedure. By default, OptiMAS will be installed in "C:\Program Files\OptiMAS" for Windows XP or "C:\Program Files (x86)\OptiMAS" for Windows 7. A new folder named "OptiMAS" containing the two data set examples will be created in your home directory.

You can now launch OptiMAS interface via the "start menu" or the desktop shortcut.

2.2.2 Linux (32/64 bits)

The software has been tested on Debian, Ubuntu and Fedora. We strongly recommend to have g++, make, Qt, qwt and graphviz installed via the package manager of your GNU/Linux distribution (aptitude/apt on Debian and Ubuntu, or yum on Fedora).

Building prerequisite:

1. GNU compiler collections version 4.0.1 or later: http://gcc.gnu.org/
2. Qt development package (v4.4.3 or later, Qt5 not tested yet): http://qt-project.org/
3. qwt development package (v5.x, v6 not supported yet): http://sourceforge.net/projects/qwt
4. graphviz development package (version 2.20.2 or later): http://www.graphviz.org/

e.g.

- Debian/Ubuntu platforms: ***apt-get install g++ libqt4-dev libqwt5-qt4-dev libgraphviz-dev***.
- Redhat/Fedora/CentOS platforms: ***yum install gcc-c++ qt4-devel qwt-devel graphviz-devel***.

To build and install "optimas_gui" GUI on your system, extract the zipped file "optimas_gui_linux.zip" by typing (for example):

```
$ unzip optimas_gui_linux.zip
```

This will create a new directory called "optimas_gui_linux". Then, open a terminal, move to this directory before attempting to run the program and run the installation shell (bash) script:

```
$ ./install_optimas_on_linux.sh
```

As **root** (or sudoer), it will perform an installation for all users in /usr/local/bin/optimas_gui. The input/output file examples will be stored respectively in /usr/local/share/OptiMAS/input and /usr/local/share/OptiMAS/output.

As a **common user**, it will perform a local install in user's personal directory $HOME/bin/optimas_gui. The input/output file examples will be stored respectively in $HOME/OptiMAS/input and $HOME/OptiMAS/output.

Note: if the installation script fails in finding qwt and/or graphviz libraries paths on your system, you can specify them in the config.in file (see INSTALL and README files for more details).

You can now launch OptiMAS interface from a terminal: ***optimas_gui*** or /path/to/optimas_gui or double-click on optimas_gui in your file browser.
Note: it is not necessary to specify the complete path if the folder including the binary (optimas_gui executable) is present in the PATH environment variable.

To uninstall OptiMAS, run the script ***uninstall_optimas.sh*** located in the same directory as optimas_gui and optimas.

2.2.3 MacOSX (32 bits)

After downloading the application (optimas_gui.app), to install it, you just need to drag it in your “Applications” folder (/Applications). Then, launch OptiMAS by double-clicking on the optimas_gui icon present in your file browser. A new folder named "OptiMAS" containing the two data set examples will be created in your home directory.

2.3 Files and directories description

Organization and description of the files that have been installed or supplied within the software package:

```
-+- README: this file.
 |
 +- INSTALL: building and installation instructions.
 |
 +- COPYING: license.
 |
 +- AUTHORS: list of authors.
 |
 +- optimas/ --> sources code to build optimas command line executable.
 |
 +- optimas_gui/ --> sources and other files to build OptiMAS GUI.
 |   |
 |   +- doc/ --> user manual/tutorial in PDF format.
 |   |
 |   +- input/ --> input sample data (genetic map & genotype/pedigree).
 |   |   |
 |   |   + moreau.dat, moreau.map --> biparental input example (old format map).
 |   |   |
 |   |   + blanc.dat, blanc(.map .qtlpos .qtll) --> multiparental input example.
 |   |
 |   +- output/ --> results data obtained from a further analysis
 |   |   |            (can be reloaded in the GUI).
 |   |   |
 |   |   + moreau/ --> biparental example ready to be analyzed.
 |   |   |
 |   |   + blanc/ --> multiparental example ready to be analyzed.
 |   |
 |   +- optimas & optimas_gui --> OptiMAS command line & GUI executables.
 |   |
 |   +- website/ --> local html version of the documentation/tutorial.
 |
 +- install_optimas_on_linux.sh --> installation shell script for Linux system.
```

3 Data preparation

To run OptiMAS v1.4.1, you must supply data containing all information about your MAS design, i.e. genotype/pedigree data, a classic genetic map, and information on QTLs.

Four input files are needed: a genotypes/pedigree file, a classic map file with marker positions, a QTL positions file, and a file used to assign a list of marker to each QTL). Examples of these files (blanc.dat, blanc.map, blanc.qtlpos, and blanc.qtll) are supplied within the user’s home directory [/user_name/OptiMAS/input/] for multiparental designs. Note that all these files must be in plain-text, tab-delimited format and that the markers present in the map file must be ordered and match those in the genotypes/pedigree file. To analyze your own data, you must prepare the input files in the appropriate format (as described below). Note that OptiMAS does not check that your input files are strictly conform to their expected format. This may lead to errors not necessarily detectable by the software.

Note: The old map file format (gathering information on markers and QTLs within a single file) used in previous versions of OptiMAS is still supported by OptiMAS v1.4.1, but we now recommend using the new format. An additional example (following the old map format) is provided within the user’s home directory [/user_name/OptiMAS/input/] for biparental designs.

3.1 Genotypes / Pedigree file (.dat)

The genotypes/pedigree file, which name has to end with *.dat*, should contain individuals, pedigree information and genotypic data. It also requires a header line specifying the information in each column. The file organization will be exemplified below for the following MAS pedigree.

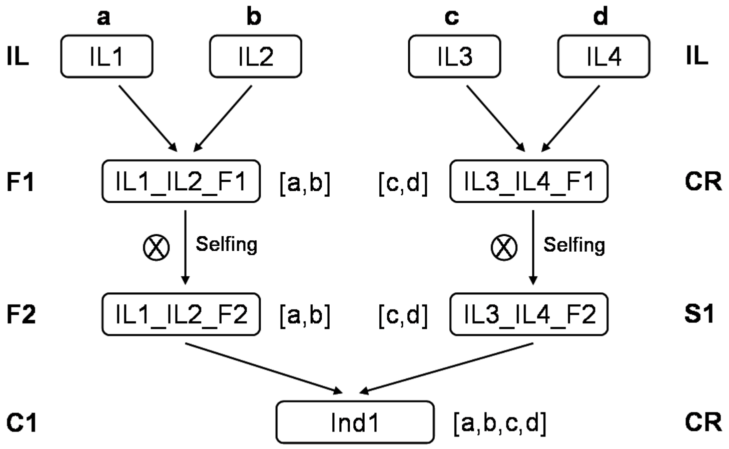


Figure 2: example of a multiparental design

The default structure for the input file (corresponding to Fig. 2 above) can be represented as follows:

Id	*P1*	*P2*	*Step*	*Cycle*	*Group*	*Mk1*	*Mk2*	*Mk3*	...	
IL1	a	a	IL	IL	-	T	C	A	...	**[1]**
IL2	b	b	IL	IL	-	C	C	G	...	
IL3	c	c	IL	IL	-	T	C	A	...	
IL4	d	d	IL	IL	-	T	C	G	...	
IL1_IL2_F1	IL1	IL2	CR	F1					...	**[2]**
IL3_IL4_F1	IL3	IL4	CR	F1						
IL1_IL2_F2	IL1_IL2_F1	IL1_IL2_F1	S1	F2	-	T	C	A	...	**[3]**
IL3_IL4_F2	IL3_IL4_F1	IL3_IL4_F1	S1	F2	-	T	C	G/A	...	
Ind1	IL1_IL2_F2	IL3_IL4_F2	CR	C1	Ear1	T	-	G/A	...	**[4]**
Ind2	**IL1_IL2_F2**	**IL3_IL4_F2**	**CR**	**C1**	**Ear2**				...	**[5]**

The genotype/pedigree file must be in plain-text, tab-delimited format (no space between fields). The header should not be changed even if the 2 optional columns (Cycle/Group) are left blank (or "-").

After the header line, each line contains the name (Id) of the current individual followed by its parents (P1/P2), the pedigree relationship linking the two generations (Step), assignation to cycle and group (if relevant) and the genotyping data (Mk1-Mkn). Note that this format is very close to the input format of the Flapjack software except that five additional columns (in red) must be added.

Columns of this file correspond to:

Id: corresponds to the name of each individual coded as a character string without any blanks and special characters. It must be unique.

Parent 1, Parent 2: correspond to the name(s) of the parent(s) of the individual (must exist as individuals above in the file except for founder parents). Individuals must be ranked according to generations (from oldest to most recent).

Founder parents of the program ([1] in table) are assumed to be homozygous lines with no residual heterozygosity[1]. Their pedigree is assumed to be unknown. Parent1, parent2 columns indicate in this case the allele that will be transmitted through generations. This allele is identified as a single character. OptiMAS can also handle selection schemes starting from heterozygous individuals (e.g. fruit tree breeding). In this case two virtual inbred lines must be defined for each heterozygous founder (described then as a Cross, see below).

Step: corresponds to the pedigree relationship between the individual and its parent(s):

CR (Cross): indicates that the individual results from a cross between its two parents.

Sn: indicates that the individual results from n (integer between 1 and 20) generations of selfing of its parent (in this case the two parents Id must be identical).

RIL: Recombinant Inbred Lines (assumes that the individual results from an infinite number of selfing generations from an initial F1 hybrid).

DH: Double Haploids (assumes that the individual results from haplo-diploïdisation from an initial F1 hybrid).

IL: indicates founder inbred status (see above).

Cycle: optional information regarding the generation in the program (e.g. first cycle, second cycle, F2, F4, etc).

Group: optional information regarding another classification criterion (e.g. subprograms, families, etc).

Mk1-Mkn: genotyping results. The software can deal with SNPs, microsatellites and any bi/multi allelic marker genotyping technique with either dominant or codominant scoring. Note that:

The markers present in the genotype/pedigree file must match those in the map file (same name and number of markers), but not necessarily be ordered.

Homozygous genotypes for an allele (e.g. A) can be scored either as A or A/A.

Heterozygotes are expected to be separated by a "/" (e.g. A/G). Heterozygous genotypes are assumed to be unphased (i.e. A/B equivalent to B/A).

Missing data at marker loci are allowed and must be entered as "-" (or can be left blank).

For dominant markers, assuming *A* dominant vs. *a* recessive, genotypes presenting allele *A* must be coded *A/-*.

Parental inbred lines should not contain missing data.

<u>Description of the genotypes/pedigree file example, in relationship with the multiparental MARS schema (see table above)</u>:

[1] Parental line where *a* is the name given to its alleles (same name for all loci).

[2] Example of a hybrid F1 obtained by intercrossing two parental lines. In this case, the genotype is inferred from parents and does not have to be declared.

[3] Plant issued from a selfing process. The name of the two parents is the same in this case. S1 means that this individual was obtained after one generation of selfing. The number of generation(s) can vary.

[1] Note that residual heterozygosity can be declared at markers (not QTL), however using this option should be discouraged due to loss of information.

[4] Plant issued from the cross between two individuals (in this case, two F2 coming from different parental lines) already declared and genotyped. In this case, the four parental alleles may have been transmitted to this individual.

[5] Plant, without genotyping data information, issued from the cross between two F2 individuals. All possible genotypes will be considered to evaluate the genetic value of this plant.

Given the **possibility to include non genotyped individuals** (e.g. Ind2), this makes it possible to analyze most common MAS schemes and mating designs. So, if several (non genotyped) steps were required to obtain a specific individual, we must generate virtual individuals in these intermediate steps.

3.2 Genetic Map (.map)

The map file, which name has to end with *.map* is supplied by the user to specify the information regarding markers, and is very similar to the Flapjack format.

mrk	Chr	pos
Marker1	1	42.2
Marker2	1	64.0
Marker3	1	72.5
Marker4	1	90.8
Marker5	2	37.1
Marker6	2	52.2
Marker7	2	54.0
Marker8	2	59.5
Marker9	2	74.8

mrk: name of markers, without blank in character chain.
chr: index (numerical value) of the chromosome where the marker is located.
pos: relative chromosomal position of loci (cM). Positions of the different loci must be obtained using the Haldane's mapping function (i.e. by assuming no interference).

3.3 QTLs information files

The aim is to create a "target genotype" (ideotype) with all the favorable alleles at the QTL positions. So, before running OptiMAS it is necessary to define the parental alleles to assemble (see table below).
In addition, as the QTL position is rarely located at a marker, QTL alleles are unknown and must be inferred from flanking markers. Thus, it is very important to **select a subset of markers as informative as possible (especially in multi-parental context) to follow the favorable parental alleles (based on haplotypes)**. The number of markers selected per QTL should be in the range of 2-6 in order to avoid intensive computation time (more progress in this area will be made in the next version).

Two QTL files are supplied by the user to (i) specify the information regarding the QTL position and identification of favorable alleles, and (ii) define the QTL region, meaning affiliate a set of marker that will be used to compute the allele transmission.

In the first file, **.qtlpos file** (see tab below), each QTL is characterized by its estimated position in cM (**pos**) on a chromosome (**chr**), and the identification of the parent carrying the favorable allele (**All+**). The information on the confidence interval, i.e. the interval which is likely to include the QTL position (CI min and CI max), will be considered in a future version and therefore can be left empty.

QTL	chr	Pos	CI min	CI max	all+
qtl1	1	70.0			A
qtl2	2	55.0			b/c

QTL: name of the QTL, without blank in character chain, The QTL names and the marker names have to match those in the qtlpos file and in the map file respectively.must match
chr: index (numerical value) of the chromosome where the QTL is located.
pos: estimated QTL position coming from the QTL detection results.
All+: identification of the parental allele(s) considered as being favorable. For QTL 1, the favorable allele "a" refers to the parental line named "IL1" (see columns "P1"/"P2" of the genotype/pedigree file). For QTL 2, "b/c" refers to parental lines (IL2 and IL3) which can be considered as favorable relatively to other parental lines.

In a second file, **either .qtll or .qtln or .qtlw** (described in the tables below), the QTL region is defined respectively as, an explicit list of marker name, a number of flanking markers, or a window defined on either side of the QTL position.

.qtll file: A list of markers explicitly assigned to each QTL.

QTL	mrk_list				
qtl1	Marker1	Marker2	Marker3	Marker4	
qtl2	Marker5	Marker6	Marke7	Marker8	Marker9

.qtln file: Number of flanking markers. Marker closest to the QTL position are taken from the map file. Note: This imply that the resulting set of marker might not include both side of the QTL position.

QTL	mrk_nb
qtl1	4
qtl2	5

.qtlw file: genetic distance defined on either side of the QTL position set in the qtlpos file. The marker set consists of the markers from the map file included in the resulting window.

QTL	window
qtl1	30
qtl2	20.5

QTL: names of the QTLs, without blank in character chain, and have to appear in the same order than those in the qtlpos file.
mrk_list: list of marker names that have to match those in the map file.
mrk_nb: integer value indicating for each QTL the number of flanking markers.
window: genetic distance in cM

Note:

Never change the header field names (in the map file and QTL information files).

It is recommended to code parental alleles ("all+" column) by a single character (e.g. a, A, b, 1 ...).

Check that **decimals are 0.00 and not 0,00** for the marker/QTL positions column ("pos").

Every file must be in **plain-text, tab-delimited format. So, use tabulation and not spaces between fields** even if they are empty (e.g. *marker1{tab}1{tab}42.2* in the map file or *qtl1{tab}1{tab}70.0{tab}{tab}{tab}a* in the qtlpos file).

The markers present in the map file must be ordered and match those in the genotype/pedigree file (same number of markers).

Both QTL information files have to share the same base name .(e.g. *maize.qtlpos* and *maize.qtll*).

4 OptiMAS in command line

OptiMAS command line version manages computationally intensive processes corresponding only to the step 1 operation (see Fig. 1). The results and output files produced at the end of the run (exemplified below) can then be reloaded via the GUI (see section 5.5).

4.1 Running OptiMAS in command line

For the three different operating systems you will need to specify a list of 6 mandatory arguments to run optimas command line executable (for example):

```
$ optimas input.dat[1] input.map[2] 0.000001[3] 0.0[4] output_folder[5] 0[6] [verb][7]
```

[1] - Input file: path to the genotype/pedigree file (.dat).

[2] - Input file: path to the genetic map input file (.map).

[3] - Algorithm parameter: cut-off (float number), genotypic probability below which a rare phased genotype (diplotype see definition below) is removed and no more considered in subsequent computations.

[4] - Algorithm parameter: cut-off (float number, 0.0 by default) for gametic probability. It corresponds to the probability that the number of crossovers expected in the region between flanking markers exceeds a given value. Thus, unlikely gametes with number of crossovers over this value are removed and no more considered in subsequent computations. Use of this option with values up to 0.01 is recommended in case many flanking markers per QTL lead to high computation time with default option.

[5] - Output folder: path to the output folder where the results will be stored (see section 4.2 for output description) the name of this directory must be changed from one run to another.

[6] - Algorithm option: 0 (by default), all the QTL present in the input files will be analyzed. 1-n (integer number), if you want to run the computations for a specific QTL.

[7] - Algorithm option (optional): verb for verbose mode. Verbose mode creates two files per QTL position, reporting respectively gamete and diplotype probabilities for all individuals.
With large/complex data, both files may take a lot of disk space, it is then recommended to disable this option.

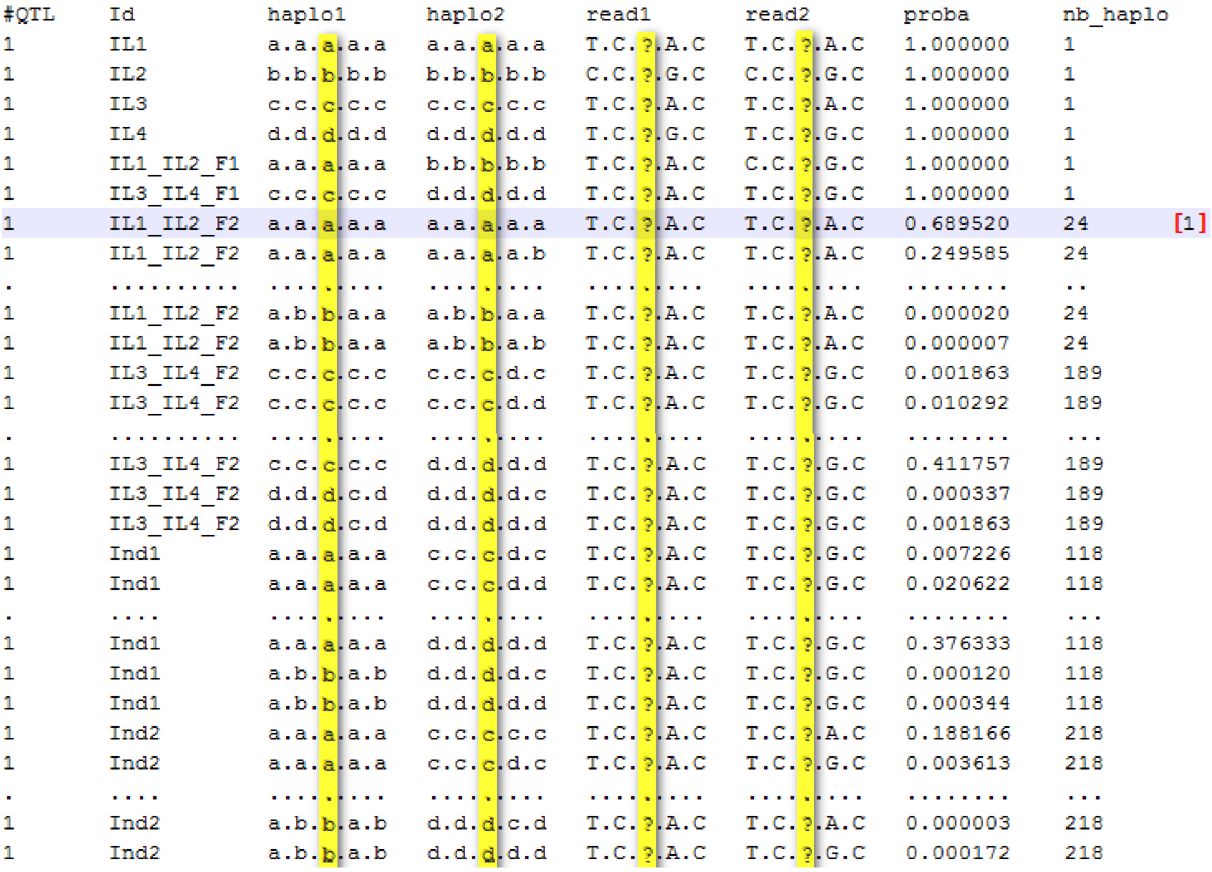
4.2 Results and output files interpretation

As the program continues to run, it keeps you informed of progress. At the end of a run, questionable results (likely genotyping error regarding pedigree) may be displayed in the file [output_folder/date/events_summary.log]. It is recommended to always look at it before any interpretation and/or breeding decision (see section 4.2.7). OptiMAS produces sets of files described below.

4.2.1 _diplotypes_set: probabilities of phased genotypes

Taking into account all information available (pedigree, distance between loci, molecular markers), OptiMAS computes for each QTL the probability of all possible phased genotypes (diplotypes). A diplotype is defined as the union of a pair of unambiguous haplotypes corresponding to parental gametes.

Results for each QTL (designated as *x*) are stored in a specific folder named "qtl*x*". Files [output_folder/date/each_qtl/qtl*x*/qtl*x*_diplotypes_set.txt] contain probabilities for diplotypes according to the following structure:



```
#QTL    Id          haplo1     haplo2     read1      read2      proba      nb_haplo
1       IL1         a.a.a.a.a  a.a.a.a.a  T.C.?.A.C  T.C.?.A.C  1.000000   1
1       IL2         b.b.b.b.b  b.b.b.b.b  C.C.?.G.C  C.C.?.G.C  1.000000   1
1       IL3         c.c.c.c.c  c.c.c.c.c  T.C.?.A.C  T.C.?.A.C  1.000000   1
1       IL4         d.d.d.d.d  d.d.d.d.d  T.C.?.G.C  T.C.?.G.C  1.000000   1
1       IL1_IL2_F1  a.a.a.a.a  b.b.b.b.b  T.C.?.A.C  C.C.?.G.C  1.000000   1
1       IL3_IL4_F1  c.c.c.c.c  d.d.d.d.d  T.C.?.A.C  T.C.?.G.C  1.000000   1
1       IL1_IL2_F2  a.a.a.a.a  a.a.a.a.a  T.C.?.A.C  T.C.?.A.C  0.689520   24          [1]
1       IL1_IL2_F2  a.a.a.a.a  a.a.a.a.b  T.C.?.A.C  T.C.?.A.C  0.249585   24
.       ..........  .........  .........  .........  .........  ........   ..
1       IL1_IL2_F2  a.b.b.a.a  a.b.b.a.a  T.C.?.A.C  T.C.?.A.C  0.000020   24
1       IL1_IL2_F2  a.b.b.a.a  a.b.b.a.b  T.C.?.A.C  T.C.?.A.C  0.000007   24
1       IL3_IL4_F2  c.c.c.c.c  c.c.c.d.c  T.C.?.A.C  T.C.?.G.C  0.001863   189
1       IL3_IL4_F2  c.c.c.c.c  c.c.c.d.d  T.C.?.A.C  T.C.?.G.C  0.010292   189
.       ..........  .........  .........  .........  .........  ........   ...
1       IL3_IL4_F2  c.c.c.c.c  d.d.d.d.d  T.C.?.A.C  T.C.?.G.C  0.411757   189
1       IL3_IL4_F2  d.d.d.c.d  d.d.d.d.c  T.C.?.A.C  T.C.?.G.C  0.000337   189
1       IL3_IL4_F2  d.d.d.c.d  d.d.d.d.d  T.C.?.A.C  T.C.?.G.C  0.001863   189
1       Ind1        a.a.a.a.a  c.c.c.d.c  T.C.?.A.C  T.C.?.G.C  0.007226   118
1       Ind1        a.a.a.a.a  c.c.c.d.d  T.C.?.A.C  T.C.?.G.C  0.020622   118
.       ....        .........  .........  .........  .........  ........   ...
1       Ind1        a.a.a.a.a  d.d.d.d.d  T.C.?.A.C  T.C.?.G.C  0.376333   118
1       Ind1        a.b.b.a.b  d.d.d.d.c  T.C.?.A.C  T.C.?.G.C  0.000120   118
1       Ind1        a.b.b.a.b  d.d.d.d.d  T.C.?.A.C  T.C.?.G.C  0.000344   118
1       Ind2        a.a.a.a.a  c.c.c.c.c  T.C.?.A.C  T.C.?.A.C  0.188166   218
1       Ind2        a.a.a.a.a  c.c.c.d.c  T.C.?.A.C  T.C.?.G.C  0.003613   218
.       ....        .........  .........  .........  .........  ........   ...
1       Ind2        a.b.b.a.b  d.d.d.c.d  T.C.?.A.C  T.C.?.A.C  0.000003   218
1       Ind2        a.b.b.a.b  d.d.d.d.d  T.C.?.A.C  T.C.?.G.C  0.000172   218
```

Figure 3: probabilities of phased genotypes (diplotypes) for each indivual at QTL1

Columns of this file correspond to:

#QTL: index of the QTL.

Id: corresponds to the name of each individual.

haplo1, haplo2: possible pair of haplotypes (ie. phased genotype, also called diplotype) defined according to parental origin.

read1, read2: translation of haplo1 and haplo 2 in terms of observed marker alleles. Note that a given (read1, read2) combination may correspond to several (haplo1, haplo2) combinations.

proba: probability of this specific possible phased genotype (diplotype).

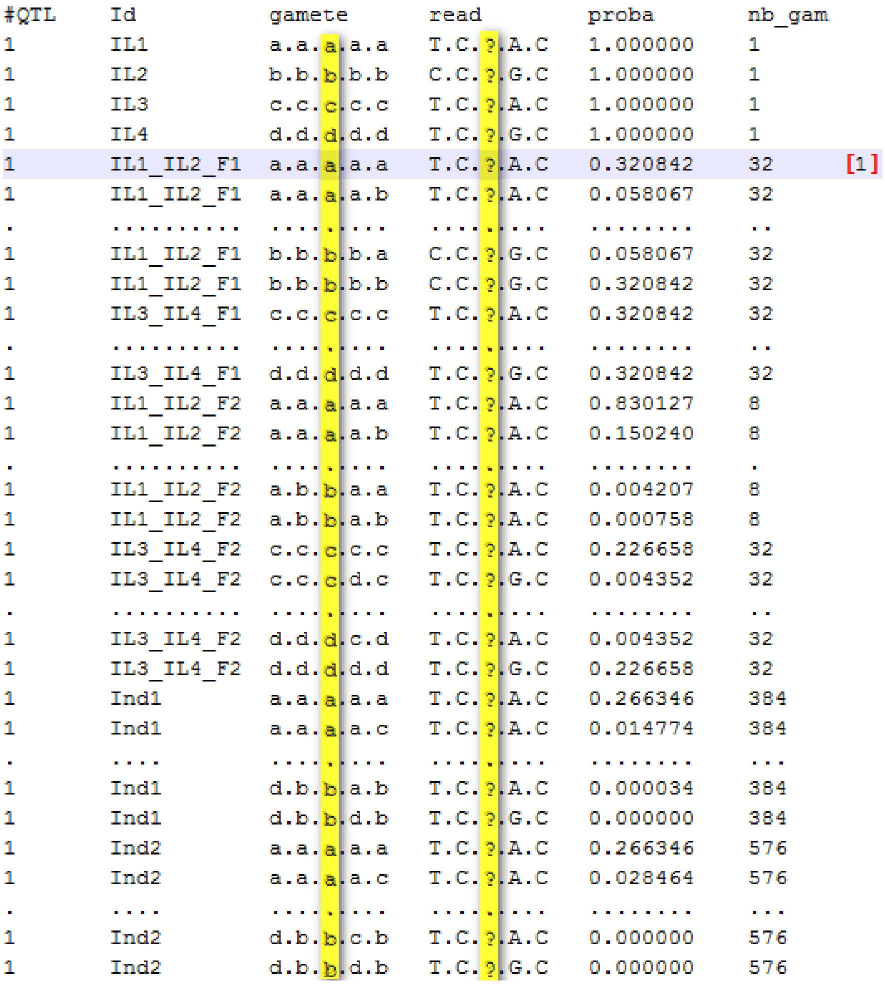
nb_haplo: number of possible diplotypes corresponding to theindividual.

<u>Note</u>: At the QTL 1, the individual IL1_IL2_F2 has 24 possible phased genotypes. One of them [1], which is the most likely given observed marker data, has a probability of 0.69. This genotype is (a/a) at the QTL position (3rd locus, see map file) and its full genotype (all loci: QTL and associated markers) is aaaaa/aaaaa. The individual Ind2 has more possible phased genotypes than Ind1 (Ind1 - 118, Ind2 - 218) because it was not genotyped (all possible diplotypes are considered).

4.2.2 _gametes_set: probabilities of gametes

The previous section underlined all the possible phased genotypes, along with their probabilities, taking into account the pedigree of individuals and their genotypes. With this information, OptiMAS determines, at each QTL, the set of possible gametes produced by each individual and estimates their probabilities.

Results for each QTL (designated as *x*) are stored in a specific folder named "qtl*x*". Files [output_folder/date/each_qtl/qtl*x*/qtl*x*_gametes_set.txt] contain probabilities for gametes according to the following structure:



```
#QTL  Id          gamete     read       proba      nb_gam
1     IL1         a.a.a.a.a  T.C.?.A.C  1.000000   1
1     IL2         b.b.b.b.b  C.C.?.G.C  1.000000   1
1     IL3         c.c.c.c.c  T.C.?.A.C  1.000000   1
1     IL4         d.d.d.d.d  T.C.?.G.C  1.000000   1
1     IL1_IL2_F1  a.a.a.a.a  T.C.?.A.C  0.320842   32      [1]
1     IL1_IL2_F1  a.a.a.a.b  T.C.?.A.C  0.058067   32
.     ..........  .........  .........  ........   ..
1     IL1_IL2_F1  b.b.b.b.a  C.C.?.G.C  0.058067   32
1     IL1_IL2_F1  b.b.b.b.b  C.C.?.G.C  0.320842   32
1     IL3_IL4_F1  c.c.c.c.c  T.C.?.A.C  0.320842   32
.     ..........  .........  .........  ........   ..
1     IL3_IL4_F1  d.d.d.d.d  T.C.?.G.C  0.320842   32
1     IL1_IL2_F2  a.a.a.a.a  T.C.?.A.C  0.830127   8
1     IL1_IL2_F2  a.a.a.a.b  T.C.?.A.C  0.150240   8
.     ..........  .........  .........  ........   .
1     IL1_IL2_F2  a.b.b.a.a  T.C.?.A.C  0.004207   8
1     IL1_IL2_F2  a.b.b.a.b  T.C.?.A.C  0.000758   8
1     IL3_IL4_F2  c.c.c.c.c  T.C.?.A.C  0.226658   32
1     IL3_IL4_F2  c.c.c.d.c  T.C.?.G.C  0.004352   32
.     ..........  .........  .........  ........   ..
1     IL3_IL4_F2  d.d.d.c.d  T.C.?.A.C  0.004352   32
1     IL3_IL4_F2  d.d.d.d.d  T.C.?.G.C  0.226658   32
1     Ind1        a.a.a.a.a  T.C.?.A.C  0.266346   384
1     Ind1        a.a.a.a.c  T.C.?.A.C  0.014774   384
.     ....        .........  .........  ........   ...
1     Ind1        d.b.b.a.b  T.C.?.A.C  0.000034   384
1     Ind1        d.b.b.d.b  T.C.?.G.C  0.000000   384
1     Ind2        a.a.a.a.a  T.C.?.A.C  0.266346   576
1     Ind2        a.a.a.a.c  T.C.?.A.C  0.028464   576
.     ....        .........  .........  ........   ...
1     Ind2        d.b.b.c.b  T.C.?.A.C  0.000000   576
1     Ind2        d.b.b.d.b  T.C.?.G.C  0.000000   576
```

Figure 4: probabilities of gametes for each individual at QTL1

Columns of this file correspond to:

#QTL: index of the QTL.

Id: corresponds to the name of each individual.

gamete: possible gamete defined according to parental origin.

read: translation of gamete in terms of observed marker alleles. Note that a given read may correspond to several gametes.

proba: probability of this specific possible gamete.

nb_gam: number of possible gametes corresponding to the individual.

Note: At the QTL number 1, the individual IL1_IL2_F1 has a probability of 100% to be heterozygous aaaaa/bbbbb (see section 4.2.1). In this case, the five loci are heterozygous and the number of possible gametes is 2^5 = 32. These 32 possible gametes have different probabilities depending on the recombination rates calculated from genetic distances (Haldane's map function used). The highest probability is that of non recombinant gametes (0.32 for both "aaaaa" and "bbbbb").

4.2.3 tab_homo_hetero: probabilities to be homozygous or heterozygous at the QTL positions

4.2.3.1 based on favorable / unfavorable allele grouping

Based on the phased genotype information (qtl*x*_haplotypes_set files), the probabilities to be homozygous / heterozygous, at the QTL positions, are computed according to favorable / unfavorable grouping of founder alleles (i.e. IL1- "a", IL2- "b", IL3 - "c", IL4- "d").

The structure for this file [output/date/tab_homo_hetero.txt] can be represented as follows:

Id	**MS**	**All**	**All**	**All**	**QTL1**	**QTL1**	**QTL1**	**QTL2**	**...**	
		(+/+)	**(-/-)**	**(+/-)**	**(+/+)**	**(-/-)**	**(+/-)**	**(+/+)**		
IL1	**0.33333**	0.33333	0.66666	0.00000	**1.00000**	0.00000	0.00000	**0.00000**	...	**[2]**
IL2	0.33333	0.33333	0.66666	0.00000	0.00000	1.00000	0.00000	1.00000	...	
IL3	0.33333	0.33333	0.66666	0.00000	0.00000	1.00000	0.00000	1.00000	...	
IL4	0.33333	0.33333	0.66666	0.00000	0.00000	1.00000	0.00000	0.00000	...	
IL1_IL2_F1	0.33333	0.00000	0.33333	0.66666	0.00000	0.00000	1.00000	0.00000	...	
IL3_IL4_F1	0.33333	0.00000	0.33333	0.66666	0.00000	1.00000	0.00000	0.00000	...	
IL1_IL2_F2	0.66137	0.65614	0.33333	0.01047	0.98658	0.00003	0.01335	0.98184	...	
IL3_IL4_F2	0.58105	0.49645	0.33422	0.16919	0.00000	**0.99996**	0.00000	0.89854	...	**[1]**
Ind1	0.58804	0.31309	0.13700	0.54989	0.00000	0.00671	0.99328	0.93928	...	**[3]**
Ind2	0.62126	0.31309	0.07055	0.61634	0.00000	0.00670	0.99328	0.93928	...	

Columns of this file correspond to:

QTL*x* (+/+): probability to be homozygous for a favorable allele at the QTL position or to be heterozygous with two different favorable alleles (when several parental alleles are considered as favorable).

QTL*x* (-/-): probability to be homozygous for an unfavorable allele at the QTL position or to be heterozygous with two different unfavorable alleles (when several parental alleles are considered as unfavorable).

QTL*x* (+/-): probability to be heterozygous at the QTL position (one favorable allele with one unfavorable).

All(+/+), All(+/-), All(-/-): mean of previous probabilities for all QTL together.

MS (Molecular Score): expected proportion of favorable alleles over all QTL (MS=All(+/+) + 0.5All(+/-), see section 4.2.4).

Note:

[1] Individual IL3_IL4_F2 has a probability of 0.99996 to be homozygous unfavorable at QTL1. The sum of the probabilities (QTL1(+/+) + QTL1(-/-) + QTL1(+/-)) is not 1.0 because some rare phased genotypes were removed via the cut-off by default (see section 4.2.6).

[2] The individual "IL1" has a molecular score of 0.333 because it is 100% homozygous favorable for the QTL1 and 0% for the two other QTL.

[3] Individual "Ind1" has a MS of 0.588 considering information at all three QTL:

Id	MS	All (+/+)	All (-/-)	All (+/-)	QTL1 (+/+)	QTL1 (-/-)	QTL1 (+/-)	QTL2 (+/+)	QTL2 (-/-)	QTL2 (+/-)	QTL3 (+/+)	QTL3 (-/-)	QTL3 (+/-)
Ind1	0.588	0.313	0.137	0.549	0.000	0.006	0.993	0.939	0.000	0.060	0.000	0.403	0.596

4.2.3.2 in terms of parental alleles

In a multi-allelic context, several parental alleles can be regrouped and considered as favorable at the QTL position (see the summarized table above). Nevertheless, it can be interesting in some cases to know the detailed probabilities of the possible genotypes in terms of the parental origin of alleles.

Results for each QTL (designated as *x*) are stored in a specific folder named "qtl*x*". Files [output_folder/date/each_qtl/qtl*x*/qtl*x*_homo_hetero.txt] contain details about parental allele origins at QTL positions according to the following structure (QTL 2 example):

```
IL1
Homo(+/+)   = 0.000000
Hetero(+/-)= 0.000000
Homo(-/-)   = 1.000000   a:a = 1.000000
IL2
Homo(+/+)   = 1.000000   b:b = 1.000000
Hetero(+/-)= 0.000000
Homo(-/-)   = 0.000000
IL3
Homo(+/+)   = 1.000000   c:c = 1.000000
Hetero(+/-)= 0.000000
Homo(-/-)   = 0.000000
IL4
Homo(+/+)   = 0.000000
Hetero(+/-)= 0.000000
Homo(-/-)   = 1.000000   d:d = 1.000000

IL1_IL2_F1
Homo(+/+)   = 0.000000
Hetero(+/-)= 1.000000   a:b = 1.000000
Homo(-/-)   = 0.000000
IL3_IL4_F1
Homo(+/+)   = 0.000000
Hetero(+/-)= 1.000000   c:d = 1.000000
Homo(-/-)   = 0.000000

IL1_IL2_F2
Homo(+/+)   = 0.981847   b:b = 0.981847
```

```
Hetero(+/-)= 0.018057   a:b = 0.018057
Homo(-/-)  = 0.000082   a:a = 0.000082

IL3_IL4_F2
Homo(+/+)  = 0.898542   c:c = 0.898542
Hetero(+/-)= 0.098733   c:d = 0.098733
Homo(-/-)  = 0.002710   d:d = 0.002710

Ind1
Homo(+/+)  = 0.939286   b:c = 0.939286
Hetero(+/-)= 0.060239   a:c = 0.008636   b:d = 0.051602
Homo(-/-)  = 0.000474   a:d = 0.000474

Ind2
Homo(+/+)  = 0.939286   b:c = 0.939286
Hetero(+/-)= 0.060239   a:c = 0.008636   b:d = 0.051602
Homo(-/-)  = 0.000474   a:d = 0.000474
```

Note: at QTL2 position, favorable alleles are "b" and "c" (see .map file). Individual "Ind1" has a very high probability to be homozygous favorable Homo(+/+) = 0.939286. We can see that this is due to a high probability of being heterozygous for favorable alleles "b" and "c" (b:c = 0.939286).

4.2.4 tab_scores: prediction of genetic value

This table summarizes and presents the molecular scores for each/all QTL and additional indexes of interest. The default structure of this file [output_folder/date/tab_scores.txt] can be represented as follows:

Id	MS	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3
IL1	0.3333	0.3333	1.0000	1	2	0	0	1.0000	0.0000	0.0000
IL2	0.3333	0.3333	1.0000	1	2	0	0	0.0000	1.0000	0.0000
IL3	0.3333	0.3333	1.0000	1	2	0	0	0.0000	1.0000	0.0000
IL4	0.3333	0.3333	1.0000	1	2	0	0	0.0000	0.0000	1.0000
IL1_IL2_F1	**0.3333**	0.3333	1.7071	0	1	2	0	**0.5000**	0.5000	0.0000
IL3_IL4_F1	0.3333	0.3333	1.7071	0	1	2	0	0.0000	0.5000	0.5000
IL1_IL2_F2	0.6613	0.6613	1.9841	2	1	0	0	0.9932	0.9908	0.0000
IL3_IL4_F2	0.5810	0.5810	1.7431	1	1	0	1	0.0000	0.9479	0.7952
Ind1	0.5880	0.5880	2.4712	1	0	1	1	0.4966	0.9694	0.2980
Ind2	0.6212	0.6212	2.5709	1	0	2	0	0.4966	0.9694	0.3977

Columns of this file correspond to:

QTL*x*: expected proportion of favorable allele (as defined after grouping) at QTL*x*, i.e. 1, 0.5, 0.0 for individuals with genotypes +/+, +/- and -/-, respectively (see table homo_hetero section 4.2.3).

MS - Molecular Score: expected proportion of favorable alleles over all QTL, i.e. the average of QTL*x* values. MS varies between 0 for an individual which does not carry any of the favorable alleles to 1 for an individual which is homozygote for the favorable alleles (i.e. it corresponds to the target genotype).

Weight (weighted MS): weighted average of QTL*x* values, to give more or less importance to the different QTL (only used via the GUI).

UC - Utility criterion: combines the molecular score with the expected variance of the MS of the gametes that can be produced by the individual. UC is based on the estimation of the expected number of favorable alleles carried by the superior 5% gametes produced by the individual. For a same MS, this criterion favors individuals with no unfavorable alleles fixed. This score ranges from 0 to the number of QTL. Note that present version of UC estimation assumes independence between QTL and should be considered as only indicative in case of linked QTL. It also assumes that the distribution of scores can be approximated by a normal distribution (which is not valid in case of small number of heterozygous QTL).

No.(+/+): number of QTL homozygous for favorable allele(s). A given QTL is considered as homozygous for favorable allele(s) when prob (+/+) exceeds a default threshold value of 0.75. This threshold can be modified via the GUI (see Fig. 9) resulting in an update of this column.

No.(-/-): number of QTL homozygous for unfavorable allele(s). A given QTL is considered as homozygous for unfavorable allele(s) when prob (-/-) exceeds a default threshold value of 0.75. This threshold can be modified via the GUI (see Fig. 9) resulting in an update of this column.

No.(+/-): number of QTL heterozygous with both favorable and unfavorable allele(s). A given QTL is considered to belong to this category when prob (+/-) exceeds a default threshold value of 0.75. This threshold can be modified via the GUI (see Fig. 9) resulting in an update of this column.

No.(?): number of QTL defined as uncertain. Concerns QTL which are not attributed to any of the three previous categories.

Note: At QTL1, individual IL1_IL2_F1 has a 100% probability to be heterozygous aaaaa/bbbbb (see section 4.2.1). The genotype at this QTL position is "a/b" (3rd locus) and "a" is the favorable allele. The molecular score for this individual will be: MS_{QTL1} = [p("a/b") x dose] / 2 = (1 x 1) / 2 = 0.5. The genetic value of the individual ("MS" column) is obtained by averaging the molecular score for all QTL: MS = (0.5 + 0.5 + 0.0) / 3 = 0.33.

4.2.5 tab_parents: estimated probabilities of parental alleles

Beyond global scores presented above, it can be interesting to display the probability of having received a given parental allele at individual QTL positions and globally across QTL.

The default structure for this file [output_folder/date/tab_parents.txt] can be represented as follows:

Id	**MS**	**All (a)**	**All (b)**	**All (c)**	**All (d)**	**QTL1 (a)**	**QTL1 (b)**	**QTL1 (c)**	**QTL1 (d)**	**...**
IL1	0.3333	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	...
IL2	0.3333	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	...
IL3	0.3333	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000	...
IL4	0.3333	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	...
IL1_IL2_F1	**0.3333**	0.5000	0.5000	0.0000	0.0000	**0.5000**	**0.5000**	**0.0000**	**0.0000**	...
IL3_IL4_F1	0.3333	0.0000	0.0000	0.5000	0.5000	0.0000	0.0000	0.5000	0.5000	...
IL1_IL2_F2	**0.6613**	0.5992	0.4007	0.0000	0.0000	**0.9932**	**0.0067**	**0.0000**	**0.0000**	...
IL3_IL4_F2	0.5810	0.0000	0.0000	0.5507	0.4491	0.0000	0.0000	0.4999	0.4999	...
Ind1	0.5880	**0.3328**	**0.1671**	**0.2332**	**0.2667**	0.4966	0.0033	0.0237	0.4762	...
Ind2	0.6212	0.2996	0.2003	0.2754	0.2245	0.4966	0.0033	0.2499	0.2499	...

Columns of this file correspond to:

QTL*x*(*parental allele*): expected proportion of parental allele at QTL*x*.

All(*parental allele*): average of QTLx values over all QTL.

MS (Molecular Score): expected proportion of favorable alleles over all QTL.

4.2.6 tab_check_diplo: sum of phased genotypes probabilities (with the cut-off)

Within OptiMAS, some algorithms such as selfing may become memory and/or time consuming depending on the complexity of your MAS design. To overcome this "problem" it is possible to use a cut-off on the probability of keeping a phased genotype (i.e. cut-off diplotypes = 0.000001 by default). Thus, rare phased genotypes are discarded in probability computation, which considerably reduces the computation time. At the end of a run, a file [output_folder/date/tab_check_diplo.txt] is created evaluate the impact of cut-off on such eliminations, at individual QTL locations and globally. The default structure for this file can be represented as follows:

Id	**MS**	**All**	**QTL1**	**QTL2**	**QTL3**
IL1	0.333333	1.000000	1.000000	1.000000	1.000000
IL2	0.333333	1.000000	1.000000	1.000000	1.000000
IL3	0.333333	1.000000	1.000000	1.000000	1.000000
IL4	0.333333	1.000000	1.000000	1.000000	1.000000
IL1_IL2_F1	0.333333	1.000000	1.000000	1.000000	1.000000
IL3_IL4_F1	0.333333	1.000000	1.000000	1.000000	1.000000
IL1_IL2_F2	**0.661379**	**0.999948**	0.999975	0.999987	0.999883
IL3_IL4_F2	0.581056	0.999879	**0.999962**	0.999985	0.999688
Ind1	0.588043	0.999998	0.999996	0.999999	**1.000000**
Ind2	0.621266	0.999992	0.999983	0.999999	0.999994

Columns of this file correspond to:

QTL*x*: sum of the probabilities to be homozygous (un)favorable and heterozygous at each QTL position, as displayed in the file "tab_homo_hetero.txt" (see section 4.2.3). When the "QTL" column is close to one, the sum of probabilities of removed diplotypes (via the cut-off) is negligible.

All: refers to the mean for all QTL.

MS (Molecular Score): expected proportion of favorable alleles over all QTL.

4.2.7 events_summary.log: questionable results

Questionable results are displayed in the file [output_folder/date/events_summary.log] (see below). **It is recommended to always check it before any interpretation (if a warning message appeared).** An empty file means that no errors have been found during the execution. Otherwise, this file contains the name of individuals with questionable genotypes at considered QTL positions. These are defined as individual x QTL combinations for which no diplotype passing the cut-off threshold could be found. Three main causes can be distinguished:

- Individuals have genotype data inconsistent in light of their ancestry (which may reveal either a genotyping error or an error in the declared pedigree).

- Individuals display very unlikely genotypes relative to their parents (i.e. genotypes that can only be obtained assuming unlikely recombination events).

- The cut off (threshold value) set by user for keeping diplotypes is too high.

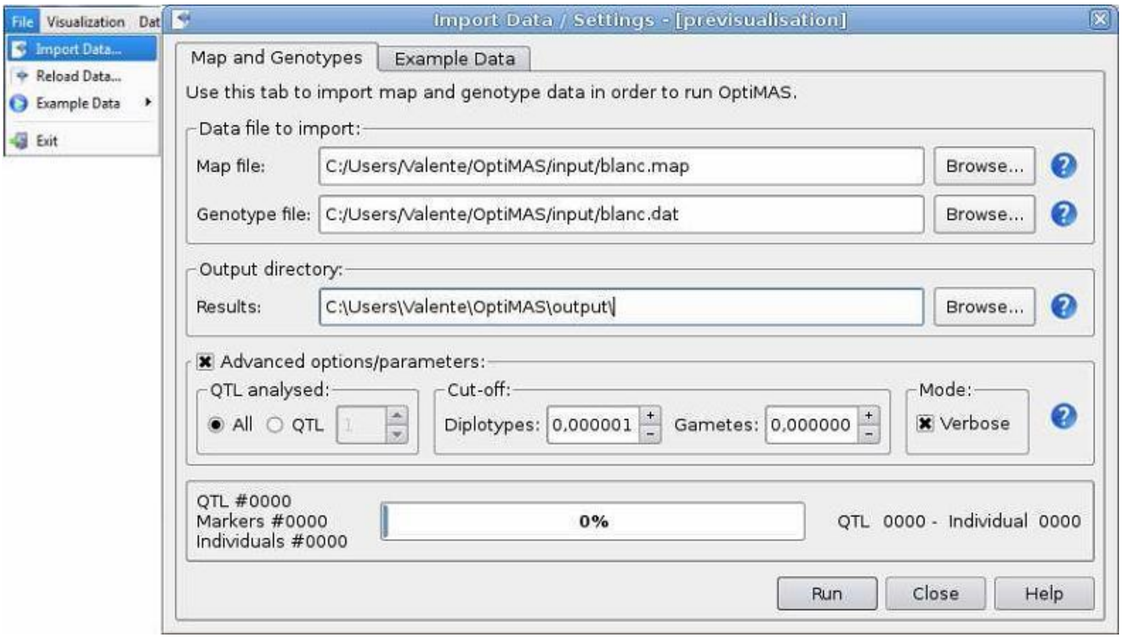
```
SUMMARY [Questionable results at markers/QTL sorted by Id]
--
Id: ind1      P1: 115_1      P2: 115_2     Cycle:F1      QTL: 1 2
Id: ind4      P1: 1814       P2: 115       Cycle:F1      QTL: 1
Id: ind33     P1: 1145       P2: 115       Cycle:F1      QTL: 1 3
Id: ind777    P1: 3734       P2: 3902      Cycle:F2      QTL: 1 4 7
```

Note that a genotyping error at generation *n* affects the results for its progeny even if at generation *n+1*, individuals at this generation have correct genotypes. See section 5.6 for more details on the procedure to adopt in case of a warning message at the end of a run.

5 OptiMAS Graphical User Interface (GUI)

5.1 Running OptiMAS GUI

To run the program, you must specify the paths to the genetic map file and the genotypes/pedigree file, see section 3) containing all information about your MAS design. ***Select File > Import Data... from the menu bar (see below).***



Use this tab to import map and genotype data in order to run OptiMAS.

Data file to import:

Map file: C:/Users/Valente/OptiMAS/input/blanc.map

Genotype file: C:/Users/Valente/OptiMAS/input/blanc.dat

Output directory:

Results: C:\Users\Valente\OptiMAS\output\

☒ Advanced options/parameters:

QTL analysed: ● All ○ QTL 1

Cut-off: Diplotypes: 0,000001 Gametes: 0,000000

Mode: ☒ Verbose

QTL #0000
Markers #0000
Individuals #0000

0%

QTL 0000 - Individual 0000

Figure 5: data set importation to run the program

Loading input/output files using the browser:

Map file: path to the genetic map input file (.map).
Remark: QTLs information files and map file has to be in the same folder, and share the same base name. *blanc.map blanc.qtlpos* and *blanc.qtll*

Genotype file: path to the genotype/pedigree file (.dat).

Output directory: path to the folder where the results will be stored. Results from each run will be stored within a new dated directory created automatically within this folder. Note that your output directory should not be in the “Program Files” folder or other specific directories with administrator privileges.

Advanced options/parameters:

QTL analyzed: by default all the QTL present in the input files will be analyzed. You can also choose to select a specific QTL to run the analysis.

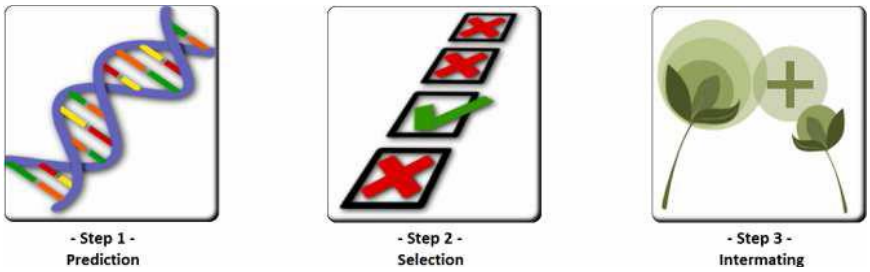
Cut-off - Diplotypes: genotypic probability below which a rare phased genotype (diplotype see definition in section 4.2.1) is removed and no more considered in subsequent computations (default value = 0.000001).

Cut-off - Gametes: gametic probability (default value = 0.000000). It corresponds to theprobability that the number of crossovers expected in the region between flanking markers exceeds a given value. Thus, unlikely gametes with number of crossovers over this value are removed and no more considered in subsequent computations. Use of this option with values up to 0.01 is recommended in case many flanking markers per QTL lead to high computation time with default option.

Verbose: Verbose mode creates two files per QTL position, reporting respectively gamete and diplotype probabilities for all individuals (default value = ON).
Warning: With large/complex data, both files may take a lot of disk space, it is then recommended to disable this option.

Click on the "Run" button to analyze the data set. The program will create output results in the folder that you specified. At the end of the run (the progress bar displays 100%), ***close the "Import data..." window by pressing the "Close" button***. If a warning message appears report to section 4.2.7 and 5.6 to analyze these questionable results before any interpretation.
Note: it is also possible to directly display the results of previous analyses by selecting ***File > Reload data.*** You can also display results from the two examples data sets provided with the program, that are located in ***File > Example Data > Biparental or Multiparental*** from the menu bar.

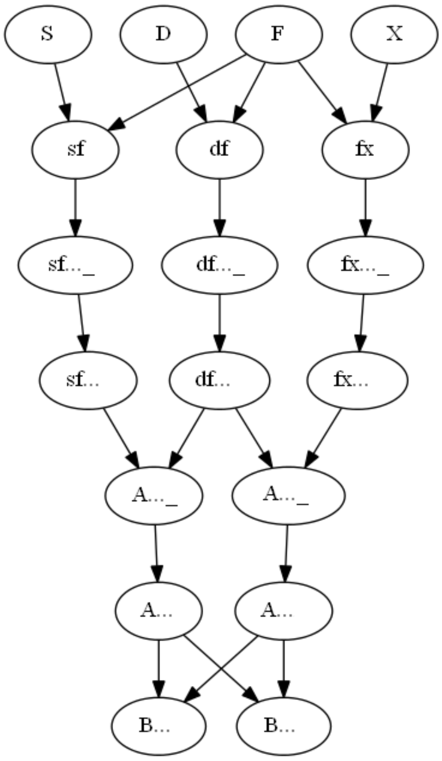
To visualize and analyze the results, the OptiMAS GUI includes three modules (on the left menu) corresponding to the different steps of the selection program (see below).



To show and use the full functionalities of OptiMAS, this analysis will focus on real data coming from a multiparental marker-assisted recurrent selection (MARS) study (Blanc *et al.,* 2006, 2008).

Six connected F2 populations, with 150 individuals each, were obtained from a half-diallel design between four unrelated maize inbred lines (*DE, F283, F810 and F9005*). Eleven QTL were detected for silking date. A set of 34 markers was selected with at least three markers (microsatellites) to follow each QTL (see blanc.map & blanc.dat files). Two cycles of MARS were performed with each time a step of selfing before intermating. In this example, OptiMAS is used at the last cycle to select the best individuals (among 297 genotyped plants) that will be used for the next cycle of MARS (see Fig.6).

To run the multiparental example data set, you must supply four input files (i.e. blanc.dat, blanc.map, blanc.qtlpos, and blanc.qtll, see section 3-4) or it is also possible to directly display the results of this previous analysis from the menu bar (i.e. ***File > Example Data > Multiparental*)**.



The four inbred lines, D (*DE*), F (*F283*), S (*F810*) and X (*F9005*) corresponding to the parental alleles (*d*, *f*, *s* and *x*, respectively) to follow through generations of selection.

Three out of six F1 plants (F2 progenies of sx, dx, sd F1 were not selected at the F2 cycle).

F1 hybrids have been selfed to obtain three F2 populations of 150 genotyped individuals. 25 F2 candidate plants were selected (first cycle of selection).

Two other selfing operations have been done. 25 families of F4 individuals were produced (without genotyping information) and crossed.

Among the progenies, 21 candidates were selected based on genotyping information (second cycle named "C1").

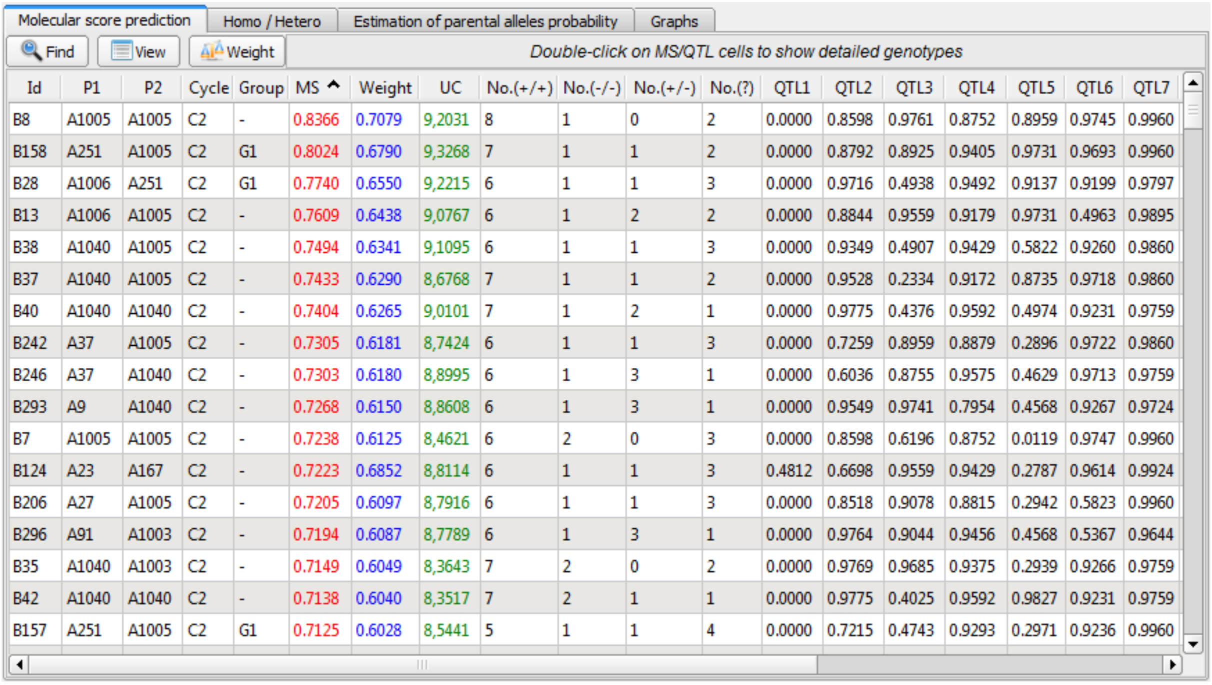
21 selfed families were produced (without genotyping information) and crossed together.

297 individuals issued from these crosses were genotyped (third cycle of selection named "C2". OptiMAS is used at this step to select the best individuals that will be used for the next cycle of MARS.

Figure 6: multiparental MAS study (Blanc *et al.*, 2006, 2008) used as example within OptiMAS

5.2 Step 1: Computation of genotypic probabilities - Estimation of genetic values

Algorithms to compute the probabilities of IBD alleles transmission throughout generations of selection have been deployed and results are displayed via three tables (corresponding to sections 4.2.4, 4.2.3 and 4.2.5) and graphs (see below).


Molecular score prediction
Homo / Hetero
Estimation of parental alleles probability
Graphs
Find
View
Weight

Double-click on MS/QTL cells to show detailed genotypes

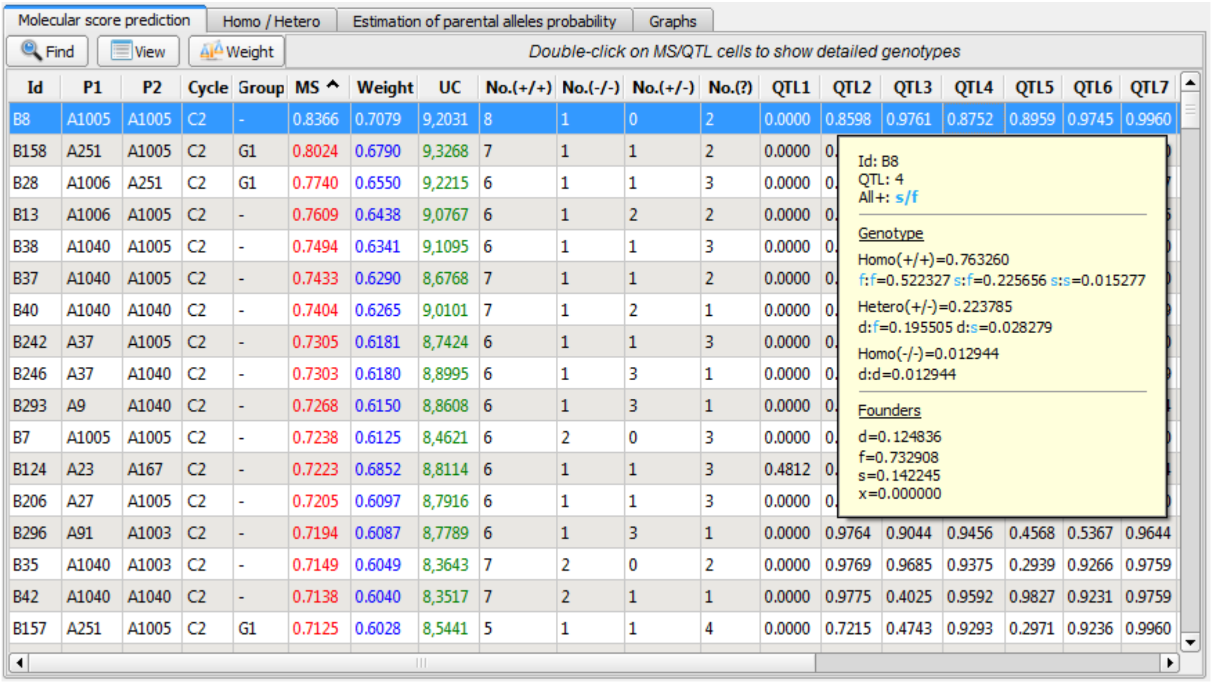
Id	P1	P2	Cycle	Group	MS ^	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7
B8	A1005	A1005	C2	-	0.8366	0.7079	9,2031	8	1	0	2	0.0000	0.8598	0.9761	0.8752	0.8959	0.9745	0.9960
B158	A251	A1005	C2	G1	0.8024	0.6790	9,3268	7	1	1	2	0.0000	0.8792	0.8925	0.9405	0.9731	0.9693	0.9960
B28	A1006	A251	C2	G1	0.7740	0.6550	9,2215	6	1	1	3	0.0000	0.9716	0.4938	0.9492	0.9137	0.9199	0.9797
B13	A1006	A1005	C2	-	0.7609	0.6438	9,0767	6	1	2	2	0.0000	0.8844	0.9559	0.9179	0.9731	0.4963	0.9895
B38	A1040	A1005	C2	-	0.7494	0.6341	9,1095	6	1	1	3	0.0000	0.9349	0.4907	0.9429	0.5822	0.9260	0.9860
B37	A1040	A1005	C2	-	0.7433	0.6290	8,6768	7	1	1	2	0.0000	0.9528	0.2334	0.9172	0.8735	0.9718	0.9860
B40	A1040	A1040	C2	-	0.7404	0.6265	9,0101	7	1	2	1	0.0000	0.9775	0.4376	0.9592	0.4974	0.9231	0.9759
B242	A37	A1005	C2	-	0.7305	0.6181	8,7424	6	1	1	3	0.0000	0.7259	0.8959	0.8879	0.2896	0.9722	0.9860
B246	A37	A1040	C2	-	0.7303	0.6180	8,8995	6	1	3	1	0.0000	0.6036	0.8755	0.9575	0.4629	0.9713	0.9759
B293	A9	A1040	C2	-	0.7268	0.6150	8,8608	6	1	3	1	0.0000	0.9549	0.9741	0.7954	0.4568	0.9267	0.9724
B7	A1005	A1005	C2	-	0.7238	0.6125	8,4621	6	2	0	3	0.0000	0.8598	0.6196	0.8752	0.0119	0.9747	0.9960
B124	A23	A167	C2	-	0.7223	0.6852	8,8114	6	1	1	3	0.4812	0.6698	0.9559	0.9429	0.2787	0.9614	0.9924
B206	A27	A1005	C2	-	0.7205	0.6097	8,7916	6	1	1	3	0.0000	0.8518	0.9078	0.8815	0.2942	0.5823	0.9960
B296	A91	A1003	C2	-	0.7194	0.6087	8,7789	6	1	3	1	0.0000	0.9764	0.9044	0.9456	0.4568	0.5367	0.9644
B35	A1040	A1003	C2	-	0.7149	0.6049	8,3643	7	2	0	2	0.0000	0.9769	0.9685	0.9375	0.2939	0.9266	0.9759
B42	A1040	A1040	C2	-	0.7138	0.6040	8,3517	7	2	1	1	0.0000	0.9775	0.4025	0.9592	0.9827	0.9231	0.9759
B157	A251	A1005	C2	G1	0.7125	0.6028	8,5441	5	1	1	4	0.0000	0.7215	0.4743	0.9293	0.2971	0.9236	0.9960

Figure 7: genetic value (molecular score) for each individual (all/each QTL)

Individuals are represented in lines followed by their pedigree, the cycle of selection and the group they belong to. Plants can be sorted regarding their molecular score for all (column MS)/each QTL (***click on the MS/QTL columns***) which is the expected proportion of favorable allele at the QTL position(s). The genetic values (MS red column) vary between 0 and 1. A value of "1" indicates that the individual corresponds to the targeted ideotype (it is homozygous for the favorable allele(s) for this QTL or all the QTL).

In this maize example, MS varies between 0.27 (X: parental line at the bottom of the table, not presented in Fig. 7) and 0.83 (B8: individual presents at the top of the table, coming from the last cycle of selection). Note that the four inbred lines D (*DE*), F (*F283*), S (*F810*) and X (*F9005*) have a MS of 0.36, 0.36, 0.45 and 0.27 respectively.

More detailed genotypes can be displayed by ***double clicking on MS or QTL cells*** (see below). This view summarizes and aggregates the information presented in the two other tables (*Homo/Hetero* and *Estimation of parental allele probabilities*, see sections 4.2.3 and 4.2.5).


Molecular score prediction
Homo / Hetero
Estimation of parental alleles probability
Graphs
Find
View
Weight

Double-click on MS/QTL cells to show detailed genotypes

Id	P1	P2	Cycle	Group	MS ^	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7
B8	A1005	A1005	C2	-	0.8366	0.7079	9,2031	8	1	0	2	0.0000	0.8598	0.9761	0.8752	0.8959	0.9745	0.9960
B158	A251	A1005	C2	G1	0.8024	0.6790	9,3268	7	1	1	2	0.0000	0.					
B28	A1006	A251	C2	G1	0.7740	0.6550	9,2215	6	1	1	3	0.0000	0.					
B13	A1006	A1005	C2	-	0.7609	0.6438	9,0767	6	1	2	2	0.0000	0.					
B38	A1040	A1005	C2	-	0.7494	0.6341	9,1095	6	1	1	3	0.0000	0.					
B37	A1040	A1005	C2	-	0.7433	0.6290	8,6768	7	1	1	2	0.0000	0.					
B40	A1040	A1040	C2	-	0.7404	0.6265	9,0101	7	1	2	1	0.0000	0.					
B242	A37	A1005	C2	-	0.7305	0.6181	8,7424	6	1	1	3	0.0000	0.					
B246	A37	A1040	C2	-	0.7303	0.6180	8,8995	6	1	3	1	0.0000	0.					
B293	A9	A1040	C2	-	0.7268	0.6150	8,8608	6	1	3	1	0.0000	0.					
B7	A1005	A1005	C2	-	0.7238	0.6125	8,4621	6	2	0	3	0.0000	0.					
B124	A23	A167	C2	-	0.7223	0.6852	8,8114	6	1	1	3	0.4812	0.					
B206	A27	A1005	C2	-	0.7205	0.6097	8,7916	6	1	1	3	0.0000	0.					
B296	A91	A1003	C2	-	0.7194	0.6087	8,7789	6	1	3	1	0.0000	0.9764	0.9044	0.9456	0.4568	0.5367	0.9644
B35	A1040	A1003	C2	-	0.7149	0.6049	8,3643	7	2	0	2	0.0000	0.9769	0.9685	0.9375	0.2939	0.9266	0.9759
B42	A1040	A1040	C2	-	0.7138	0.6040	8,3517	7	2	1	1	0.0000	0.9775	0.4025	0.9592	0.9827	0.9231	0.9759
B157	A251	A1005	C2	G1	0.7125	0.6028	8,5441	5	1	1	4	0.0000	0.7215	0.4743	0.9293	0.2971	0.9236	0.9960

Id: B8
QTL: 4
All+: s/f

Genotype

Homo(+/+)=0.763260
f:f=0.522327 s:f=0.225656 s:s=0.015277

Hetero(+/-)=0.223785
d:f=0.195505 d:s=0.028279

Homo(-/-)=0.012944
d:d=0.012944

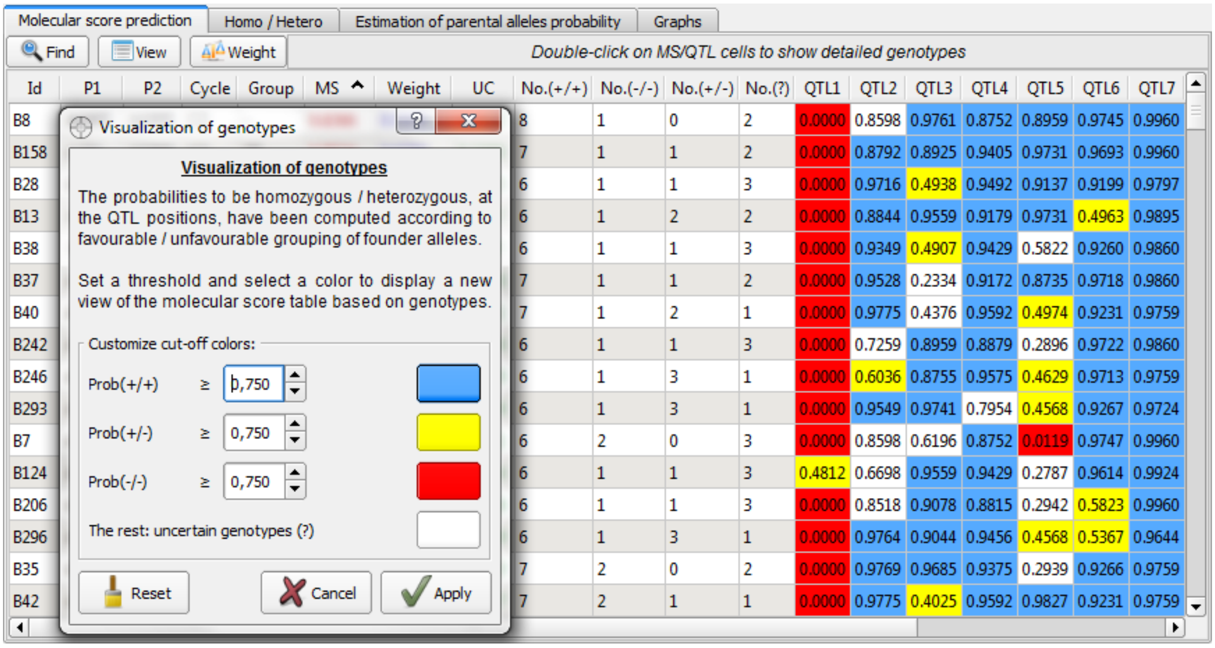
Founders

d=0.124836
f=0.732908
s=0.142245
x=0.000000

Figure 8: detailed genotype in terms of parental alleles at QTL position

Fig. 8 shows that individual B8 has a molecular score of 0.8752 at QTL 4. It has a probability of 0.763260 to be homozygous for the favorable alleles *s/f* (i.e. Homo(+/+)). This score corresponds to the sum of the probabilities of the genotypes f:f=0.522327, s:f=0.225656 and s:s=0.015277). Its MS of 0.8752 corresponds to the expected proportion of favorable allele(s) (i.e. Homo(+/+) + 1/2 Hetero(+/-)). Founders (represented by d, f, s and x alleles) indicate the expected proportion of parental alleles. We notice that this individual is issued from three parental lines D, F, S and not X (see also pedigree in Fig. 20).

A colored view of the molecular score table can be displayed to identify more easily QTL for which a given individual is considered as fixed or not (see Fig. 9). ***Press Visualization** > **Color scheme... on the menu bar.***


Molecular score prediction
Homo / Hetero
Estimation of parental alleles probability
Graphs
Find
View
Weight
Visualization of genotypes
?
Reset
Cancel
Apply

Double-click on MS/QTL cells to show detailed genotypes

Id	P1	P2	Cycle	Group	MS ^	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7
B8								8	1	0	2	0.0000	0.8598	0.9761	0.8752	0.8959	0.9745	0.9960
B158								7	1	1	2	0.0000	0.8792	0.8925	0.9405	0.9731	0.9693	0.9960
B28								6	1	1	3	0.0000	0.9716	0.4938	0.9492	0.9137	0.9199	0.9797
B13								6	1	2	2	0.0000	0.8844	0.9559	0.9179	0.9731	0.4963	0.9895
B38								6	1	1	3	0.0000	0.9349	0.4907	0.9429	0.5822	0.9260	0.9860
B37								7	1	1	2	0.0000	0.9528	0.2334	0.9172	0.8735	0.9718	0.9860
B40								7	1	2	1	0.0000	0.9775	0.4376	0.9592	0.4974	0.9231	0.9759
B242								6	1	1	3	0.0000	0.7259	0.8959	0.8879	0.2896	0.9722	0.9860
B246								6	1	3	1	0.0000	0.6036	0.8755	0.9575	0.4629	0.9713	0.9759
B293								6	1	3	1	0.0000	0.9549	0.9741	0.7954	0.4568	0.9267	0.9724
B7								6	2	0	3	0.0000	0.8598	0.6196	0.8752	0.0119	0.9747	0.9960
B124								6	1	1	3	0.4812	0.6698	0.9559	0.9429	0.2787	0.9614	0.9924
B206								6	1	1	3	0.0000	0.8518	0.9078	0.8815	0.2942	0.5823	0.9960
B296								6	1	3	1	0.0000	0.9764	0.9044	0.9456	0.4568	0.5367	0.9644
B35								7	2	0	2	0.0000	0.9769	0.9685	0.9375	0.2939	0.9266	0.9759
B42								7	2	1	1	0.0000	0.9775	0.4025	0.9592	0.9827	0.9231	0.9759

Visualization of genotypes

The probabilities to be homozygous / heterozygous, at the QTL positions, have been computed according to favourable / unfavourable grouping of founder alleles.

Set a threshold and select a color to display a new view of the molecular score table based on genotypes.

Customize cut-off colors:

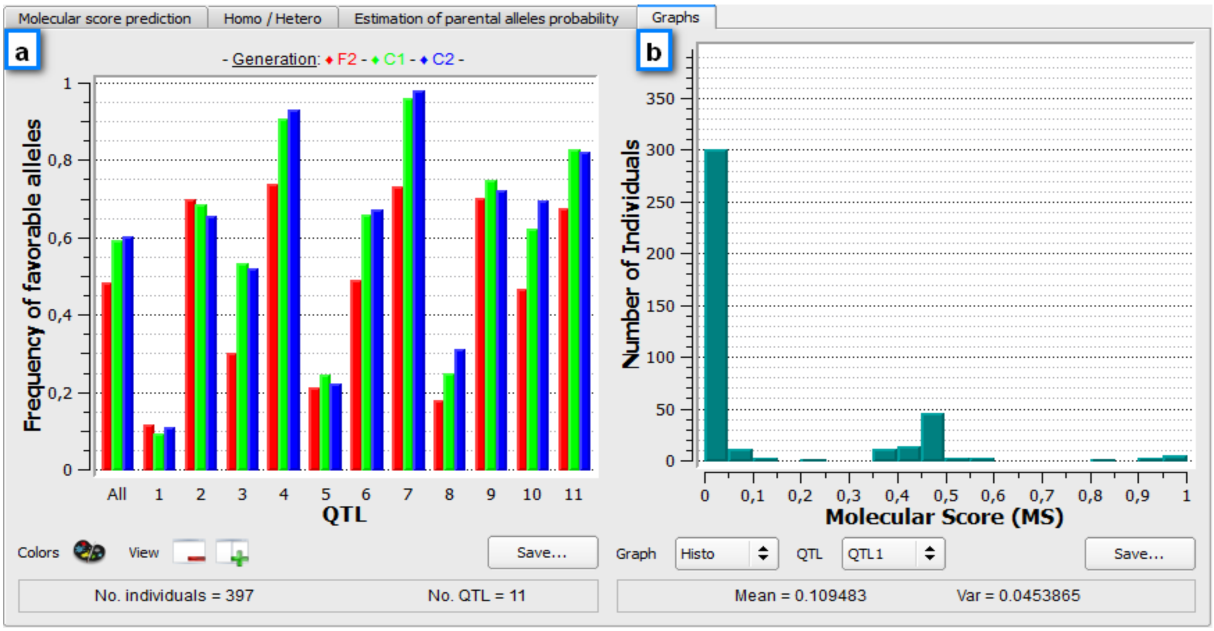
Prob(+/+)	≥	0,750	🟦
Prob(+/-)	≥	0,750	🟨
Prob(-/-)	≥	0,750	🟥
The rest: uncertain genotypes (?)			⬜

Figure 9: colored view of the molecular score table

A value of 0.75 (by default) is selected for the probability threshold to be considered as homozygous (un)favorable or heterozygous at the QTL positions. A color can be assigned to each of them. Genotypes which are not assigned to any of these categories are considered "uncertain genotypes (?)". When you apply a new set of parameters (cut-off / colors), the four corresponding columns (No.(+/+), No.(-/-), No(+/-), No.(?)) of the MS table are updated.

For example, in Fig. 9, individual B8 is considered as homozygous favorable for eight QTL (in blue, No.(+/+) = 8), homozygous unfavorable for only one QTL (in red, No.(-/-) = 1), it presents no heterozygous QTL (in grey, No.(+/-) = 0) and two QTL (in yellow, No.(?) = 2) are uncertain. Some MS at QTL positions are close to 1 (e.g. QTL3 = 0.9761) and some others are lower (e.g. QTL2 = 0.8598). This uncertainty can be due to (i) the fact that one marker flanking the QTL is heterozygous whereas the other one is homozygous favorable, which indicates that a recombination took place near the QTL position, or (ii) that there are missing data (see genotypes/pedigree file).

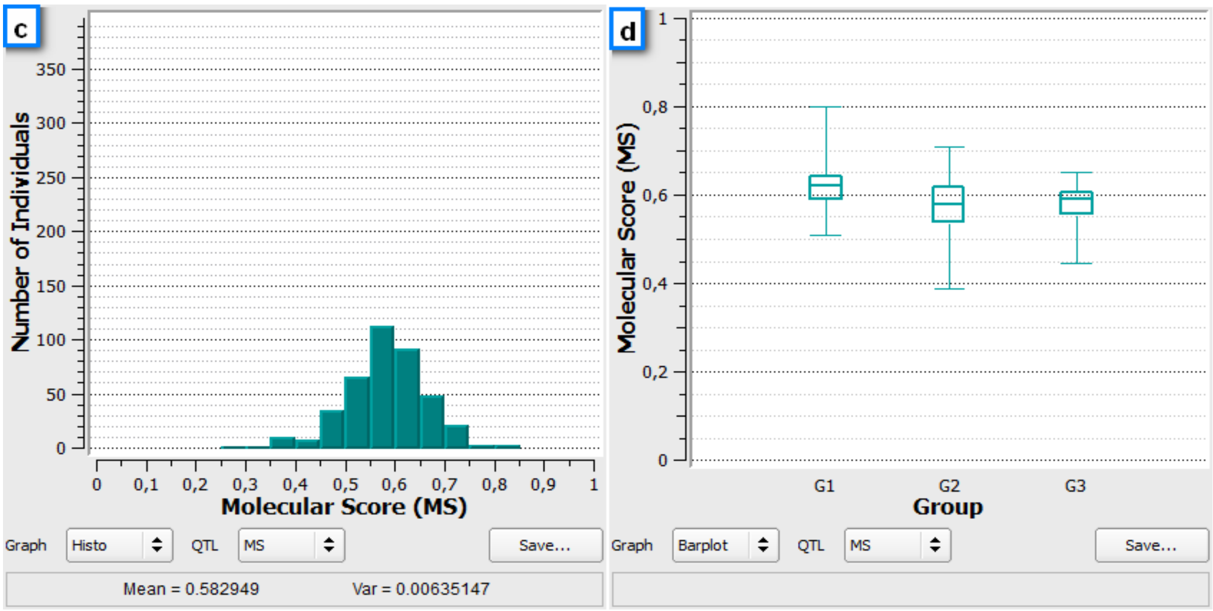
The results of the different tables can be visualized on graphs that are automatically generated by ***clicking on Graphs*** (see below).



No. individuals = 397 No. QTL = 11

Mean = 0.109483 Var = 0.0453865

The graph on Fig. 10a indicates the frequency of favorable alleles at the different generations of selection (on average and for each QTL). Note that no genetic gain is expected for the last generation (C2, in blue) because individuals are not selected yet.



Graph: Histo   QTL: MS

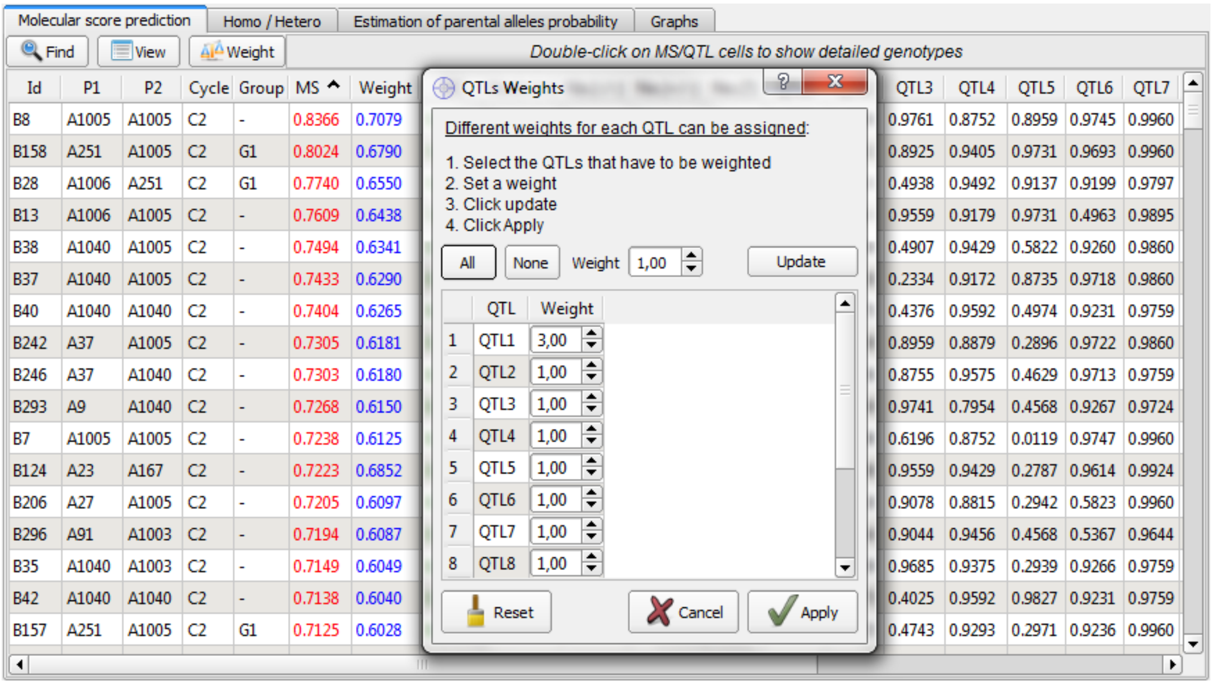
Mean = 0.582949   Var = 0.00635147

Graph: Barplot   QTL: MS

Figure 10: distribution of QTL MS, global genetic values and their evolution over the different cycles of selection

Fig. 10b and 10c show the distribution of the molecular score (for each QTL separately and on average for all QTL) whereas Fig. 10d displays the average of the MS for individuals classified according to another classification criterion (e.g. subprograms, families, etc). All the graphs can be exported (in png, svg or eps formats) by using the ***"Save…"*** button.

In the estimation of the Molecular Score (MS), OptiMAS attributes the same weight to all QTL declared in the map file. It is also possible to discard QTL and/or to attribute economical weights defined by the breeder, to compute a "Weight" index. ***Press the Weight button to open a dialog window (see below).***


Molecular score prediction
Homo / Hetero
Estimation of parental alleles probability
Graphs
Find
View
Weight
All
None
Update
Reset
Cancel
Apply

Double-click on MS/QTL cells to show detailed genotypes

Id	P1	P2	Cycle	Group	MS	Weight	QTL3	QTL4	QTL5	QTL6	QTL7
B8	A1005	A1005	C2	-	0.8366	0.7079	0.9761	0.8752	0.8959	0.9745	0.9960
B158	A251	A1005	C2	G1	0.8024	0.6790	0.8925	0.9405	0.9731	0.9693	0.9960
B28	A1006	A251	C2	G1	0.7740	0.6550	0.4938	0.9492	0.9137	0.9199	0.9797
B13	A1006	A1005	C2	-	0.7609	0.6438	0.9559	0.9179	0.9731	0.4963	0.9895
B38	A1040	A1005	C2	-	0.7494	0.6341	0.4907	0.9429	0.5822	0.9260	0.9860
B37	A1040	A1005	C2	-	0.7433	0.6290	0.2334	0.9172	0.8735	0.9718	0.9860
B40	A1040	A1040	C2	-	0.7404	0.6265	0.4376	0.9592	0.4974	0.9231	0.9759
B242	A37	A1005	C2	-	0.7305	0.6181	0.8959	0.8879	0.2896	0.9722	0.9860
B246	A37	A1040	C2	-	0.7303	0.6180	0.8755	0.9575	0.4629	0.9713	0.9759
B293	A9	A1040	C2	-	0.7268	0.6150	0.9741	0.7954	0.4568	0.9267	0.9724
B7	A1005	A1005	C2	-	0.7238	0.6125	0.6196	0.8752	0.0119	0.9747	0.9960
B124	A23	A167	C2	-	0.7223	0.6852	0.9559	0.9429	0.2787	0.9614	0.9924
B206	A27	A1005	C2	-	0.7205	0.6097	0.9078	0.8815	0.2942	0.5823	0.9960
B296	A91	A1003	C2	-	0.7194	0.6087	0.9044	0.9456	0.4568	0.5367	0.9644
B35	A1040	A1003	C2	-	0.7149	0.6049	0.9685	0.9375	0.2939	0.9266	0.9759
B42	A1040	A1040	C2	-	0.7138	0.6040	0.4025	0.9592	0.9827	0.9231	0.9759
B157	A251	A1005	C2	G1	0.7125	0.6028	0.4743	0.9293	0.2971	0.9236	0.9960

QTLs Weights

Different weights for each QTL can be assigned:

1. Select the QTLs that have to be weighted
2. Set a weight
3. Click update
4. Click Apply

Weight: 1,00

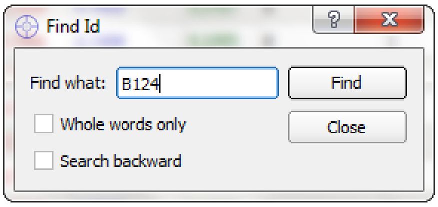
	QTL	Weight
1	QTL1	3,00
2	QTL2	1,00
3	QTL3	1,00
4	QTL4	1,00
5	QTL5	1,00
6	QTL6	1,00
7	QTL7	1,00
8	QTL8	1,00

Figure 11: weighted molecular score give more or less importance to the different QTL

We noticed in this example that the favorable allele (*x*) at QTL1 may be lost because (i) the ten best individuals of the panel have a molecular score of 0 at QTL1 (see Fig. 9, cells in red) and (ii) graphs in Fig. 10a and 10b indicate the same decay. Thus, a weight of 3.0 has been attributed to the QTL 1. ***Assign QTL weights then press Apply***. It will result in an update on the "Weight" column (in blue) and therefore produce a new classification of individuals.

In addition to the molecular score and the weight columns, OptiMAS estimates an "Utility Criterion" (UC, green column in Fig. 8) which evaluates the expected value of superior gametes of each individual by combining the molecular score with the expected variance of the MS of its gametes (see section 4.2.4 for more details).

The **"Find Id" dialog** box can be used to search and locate a specific individual in the panel. ***Press "Find" button*** (see below).



Find what: B124

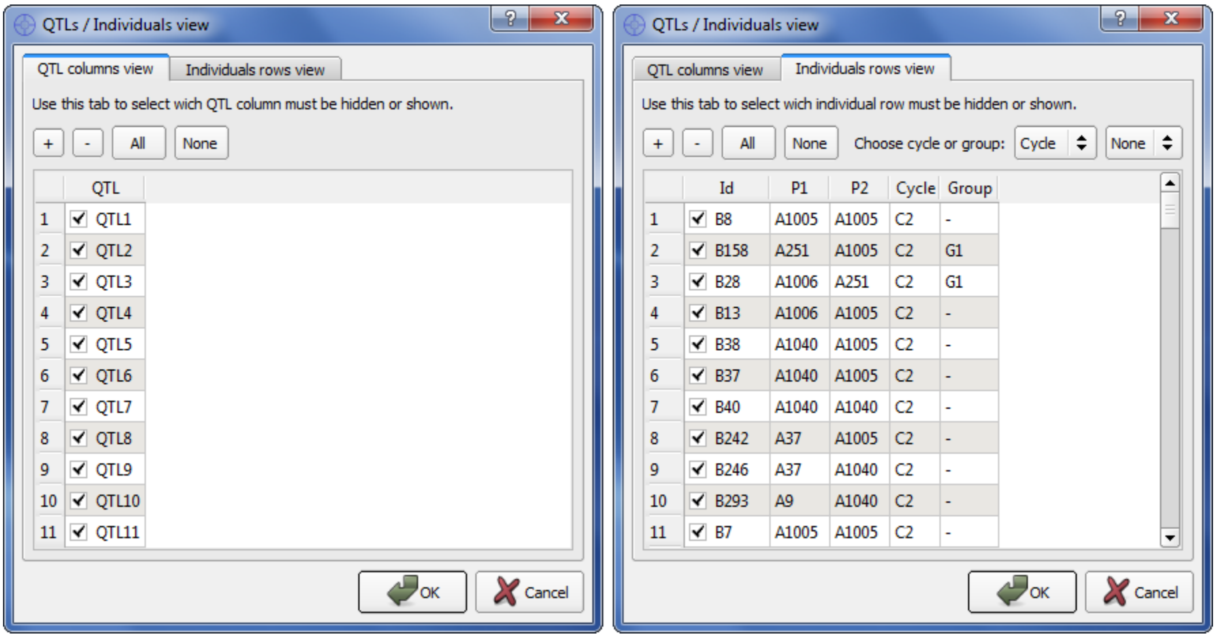
□ Whole words only

□ Search backward

Figure 12: find individual by name ("Find Id" dialog)

By default, research identifies individuals the name of which contains the declared string ("B124" in figure 12). Research is also possible via exact matching (***check "Whole words only"***). ***Enter the Id of the individual that you are looking for with the appropriate parameters into the search box and then press the "Find" button***. Any matching results will move the main display to the exact position of the individual. It will also be graphically highlighted.

The **Filter "QTL/Individuals" dialog** is used to enable or disable the display of QTL and/or individuals on the MS table. ***Press the "View" button*** to display the filter dialog (see below).



Use this tab to select wich QTL column must be hidden or shown.

	QTL
1	☑ QTL1
2	☑ QTL2
3	☑ QTL3
4	☑ QTL4
5	☑ QTL5
6	☑ QTL6
7	☑ QTL7
8	☑ QTL8
9	☑ QTL9
10	☑ QTL10
11	☑ QTL11

Use this tab to select wich individual row must be hidden or shown.

Choose cycle or group: Cycle None

	Id	P1	P2	Cycle	Group
1	☑ B8	A1005	A1005	C2	-
2	☑ B158	A251	A1005	C2	G1
3	☑ B28	A1006	A251	C2	G1
4	☑ B13	A1006	A1005	C2	-
5	☑ B38	A1040	A1005	C2	-
6	☑ B37	A1040	A1005	C2	-
7	☑ B40	A1040	A1040	C2	-
8	☑ B242	A37	A1005	C2	-
9	☑ B246	A37	A1040	C2	-
10	☑ B293	A9	A1040	C2	-
11	☑ B7	A1005	A1005	C2	-

Figure 13: QTL/Individuals filter dialog

The Fig. 13 presents two tables displaying (i) the list of the QTL present in the data set (left side) and (ii) the list of all the individuals present in the dataset (right side). Individuals can be filtered by "Cycles" of selection, "Groups" or manually. ***Select and check QTL and/or individuals to follow and press the "OK" button to apply the corresponding filter.*** This refreshes immediately the MS table. This new view of the MS table can be useful if you are working with a large number of QTL and/or individuals and you want to focus on specific QTL/plants.

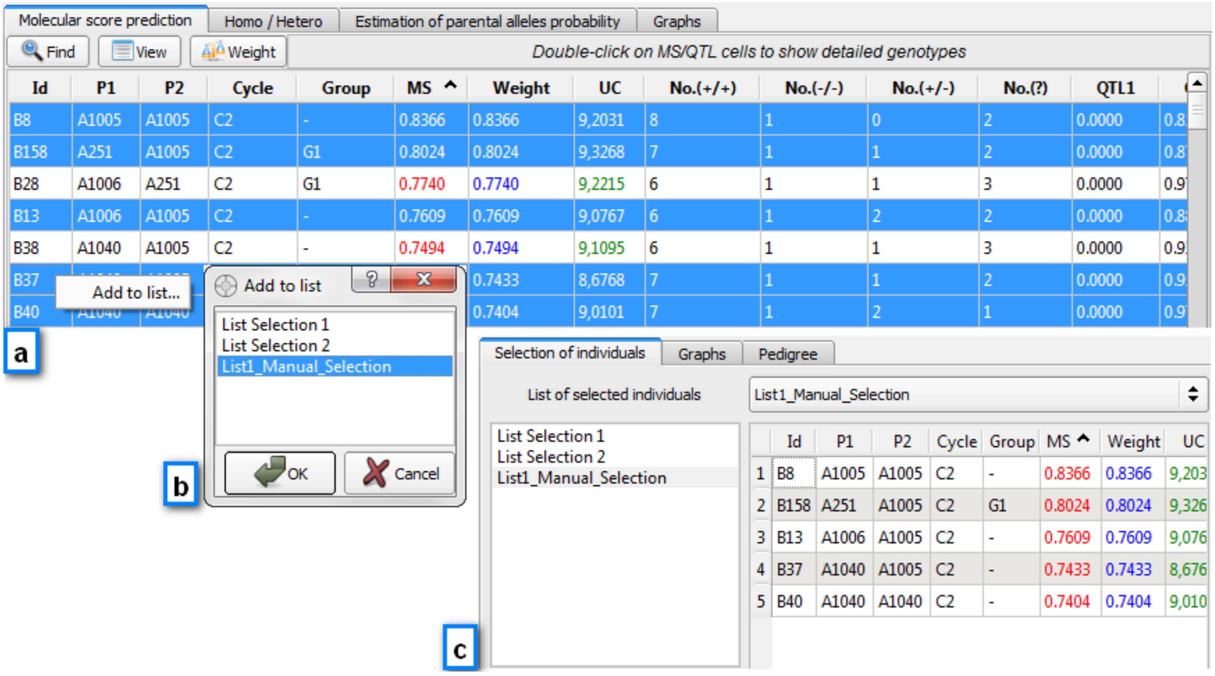
5.3 Step 2: Selection of individuals

Taking into account all the information coming from the previous tables, we can select individuals for producing the next generation.

5.3.1 Methods for selection

In OptiMAS, three different ways are possible to select individuals:

Manual selection: selection of individuals based on your own judgment (see Fig. 14). In the step 1 window (molecular score table), ***select plants via click and drag selection (or simple click + Ctrl) > Press Right click > Add to list... > Selection of your list (can be renamed by double-click) > Ok.***



Double-click on MS/QTL cells to show detailed genotypes

Id	P1	P2	Cycle	Group	MS ^	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	(
B8	A1005	A1005	C2	-	0.8366	0.8366	9,2031	8	1	0	2	0.0000	0.8
B158	A251	A1005	C2	G1	0.8024	0.8024	9,3268	7	1	1	2	0.0000	0.8
B28	A1006	A251	C2	G1	0.7740	0.7740	9,2215	6	1	1	3	0.0000	0.9
B13	A1006	A1005	C2	-	0.7609	0.7609	9,0767	6	1	2	2	0.0000	0.8
B38	A1040	A1005	C2	-	0.7494	0.7494	9,1095	6	1	1	3	0.0000	0.9
B37	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.7433	8,6768	7	1	1	2	0.0000	0.9
B40	A1040	A1040	[illegible]	[illegible]	[illegible]	0.7404	9,0101	7	1	2	1	0.0000	0.9

List Selection 1
List Selection 2
List1_Manual_Selection

List of selected individuals: List1_Manual_Selection

List Selection 1
List Selection 2
List1_Manual_Selection

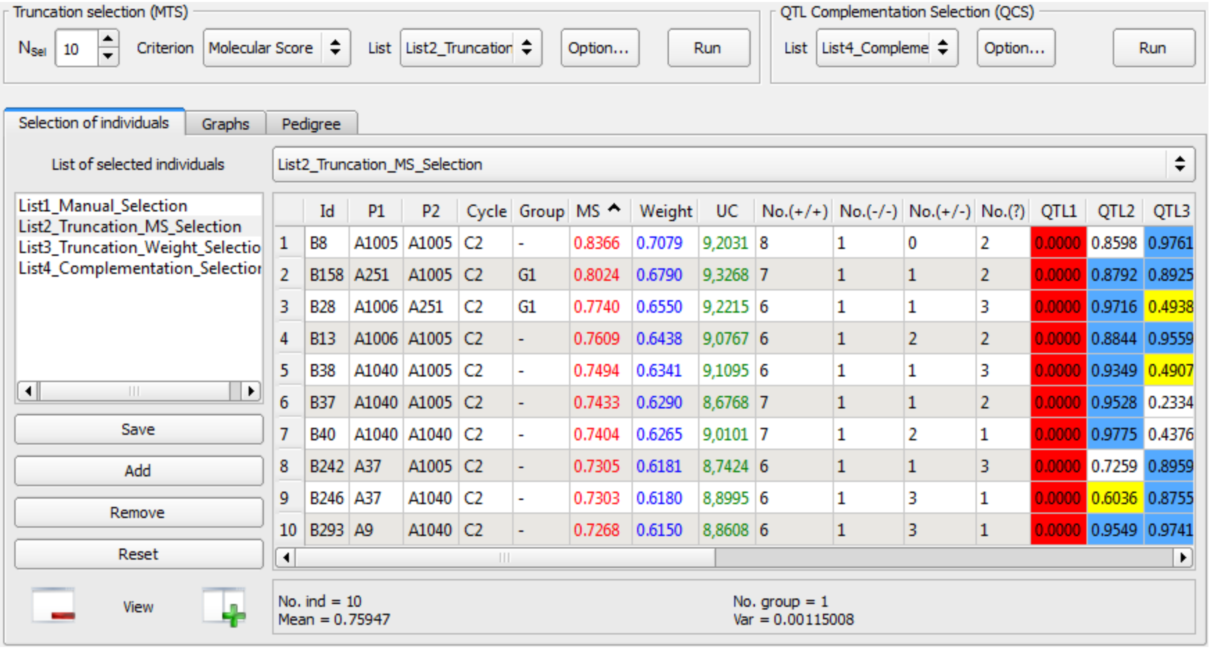
	Id	P1	P2	Cycle	Group	MS ^	Weight	UC
1	B8	A1005	A1005	C2	-	0.8366	0.8366	9,203
2	B158	A251	A1005	C2	G1	0.8024	0.8024	9,326
3	B13	A1006	A1005	C2	-	0.7609	0.7609	9,076
4	B37	A1040	A1005	C2	-	0.7433	0.7433	8,676
5	B40	A1040	A1040	C2	-	0.7404	0.7404	9,010

Figure 14: manual selection of individuals added to a list

Individuals are selected manually (a) and then added to a list of your choice (b). This new list can be accessed through the Step 2 interface (c), see above. Names of lists can be modified by the user at all steps (by double clicking on the name of the list one wants to modify).

The two next options are initiated by ***clicking on the Step 2 icon***.

Truncation selection (TS): individuals can be ranked automatically based on three possible criteria: Molecular Score (MS), Weighted MS or Utility Criterion (UC). The N_{sel} first sorted individuals are selected to generate the next population. **e.g.** ***N_{sel} = 10, Criterion = MS, List = List2_Truncation_MS_Selection, Option: Cycle = C2 (last cycle), then press Run***. A second list can be created by doing ***the same with Criterion = Weight and List = List3_Truncation_Weight_Selection.*** The list(s) of selected individuals will appear on the "Selection" page (see Fig. 15).



Truncation selection (MTS)
N_{sel} 10 Criterion Molecular Score List List2_Truncatior

QTL Complementation Selection (QCS)
List List4_Compleme

List of selected individuals: List2_Truncation_MS_Selection

List1_Manual_Selection
List2_Truncation_MS_Selection
List3_Truncation_Weight_Selectio
List4_Complementation_Selectior

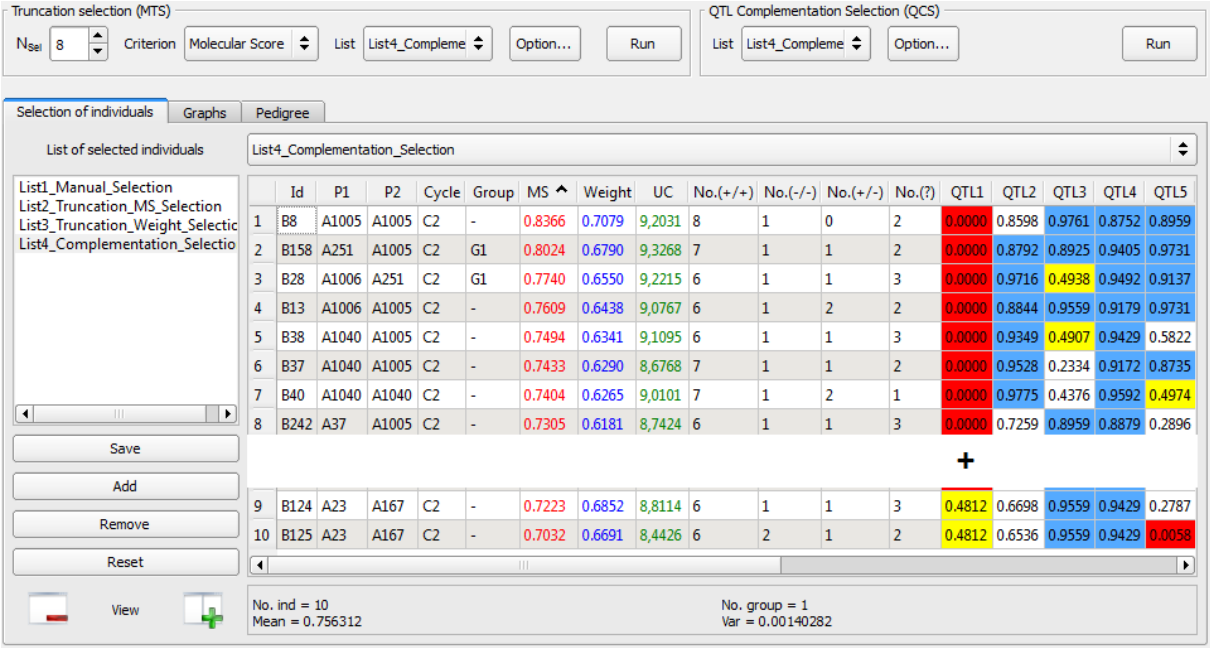
	Id	P1	P2	Cycle	Group	MS ^	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3
1	B8	A1005	A1005	C2	-	0.8366	0.7079	9,2031	8	1	0	2	0.0000	0.8598	0.9761
2	B158	A251	A1005	C2	G1	0.8024	0.6790	9,3268	7	1	1	2	0.0000	0.8792	0.8925
3	B28	A1006	A251	C2	G1	0.7740	0.6550	9,2215	6	1	1	3	0.0000	0.9716	0.4938
4	B13	A1006	A1005	C2	-	0.7609	0.6438	9,0767	6	1	2	2	0.0000	0.8844	0.9559
5	B38	A1040	A1005	C2	-	0.7494	0.6341	9,1095	6	1	1	3	0.0000	0.9349	0.4907
6	B37	A1040	A1005	C2	-	0.7433	0.6290	8,6768	7	1	1	2	0.0000	0.9528	0.2334
7	B40	A1040	A1040	C2	-	0.7404	0.6265	9,0101	7	1	2	1	0.0000	0.9775	0.4376
8	B242	A37	A1005	C2	-	0.7305	0.6181	8,7424	6	1	1	3	0.0000	0.7259	0.8959
9	B246	A37	A1040	C2	-	0.7303	0.6180	8,8995	6	1	3	1	0.0000	0.6036	0.8755
10	B293	A9	A1040	C2	-	0.7268	0.6150	8,8608	6	1	3	1	0.0000	0.9549	0.9741

No. ind = 10
Mean = 0.75947

No. group = 1
Var = 0.00115008

Figure 15: truncation selection based on three criteria (MS, Weighted MS and Utility Criterion)

QTL complementation selection (QCS): takes into account complementarities between candidate individual(s) regarding the favorable alleles they carry (see figures below). It aims at preventing the loss of rare favorable allele(s) (Hospital *et al.*, 2000). This option is important when a high number of QTL is considered.



Truncation selection (MTS)
N_{sel} 8 Criterion Molecular Score List List4_Compleme

QTL Complementation Selection (QCS)
List List4_Compleme

List of selected individuals: List4_Complementation_Selection

List1_Manual_Selection
List2_Truncation_MS_Selection
List3_Truncation_Weight_Selectic
List4_Complementation_Selectio

	Id	P1	P2	Cycle	Group	MS ^	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3	QTL4	QTL5
1	B8	A1005	A1005	C2	-	0.8366	0.7079	9,2031	8	1	0	2	0.0000	0.8598	0.9761	0.8752	0.8959
2	B158	A251	A1005	C2	G1	0.8024	0.6790	9,3268	7	1	1	2	0.0000	0.8792	0.8925	0.9405	0.9731
3	B28	A1006	A251	C2	G1	0.7740	0.6550	9,2215	6	1	1	3	0.0000	0.9716	0.4938	0.9492	0.9137
4	B13	A1006	A1005	C2	-	0.7609	0.6438	9,0767	6	1	2	2	0.0000	0.8844	0.9559	0.9179	0.9731
5	B38	A1040	A1005	C2	-	0.7494	0.6341	9,1095	6	1	1	3	0.0000	0.9349	0.4907	0.9429	0.5822
6	B37	A1040	A1005	C2	-	0.7433	0.6290	8,6768	7	1	1	2	0.0000	0.9528	0.2334	0.9172	0.8735
7	B40	A1040	A1040	C2	-	0.7404	0.6265	9,0101	7	1	2	1	0.0000	0.9775	0.4376	0.9592	0.4974
8	B242	A37	A1005	C2	-	0.7305	0.6181	8,7424	6	1	1	3	0.0000	0.7259	0.8959	0.8879	0.2896
													+				
9	B124	A23	A167	C2	-	0.7223	0.6852	8,8114	6	1	1	3	0.4812	0.6698	0.9559	0.9429	0.2787
10	B125	A23	A167	C2	-	0.7032	0.6691	8,4426	6	2	1	2	0.4812	0.6536	0.9559	0.9429	0.0058

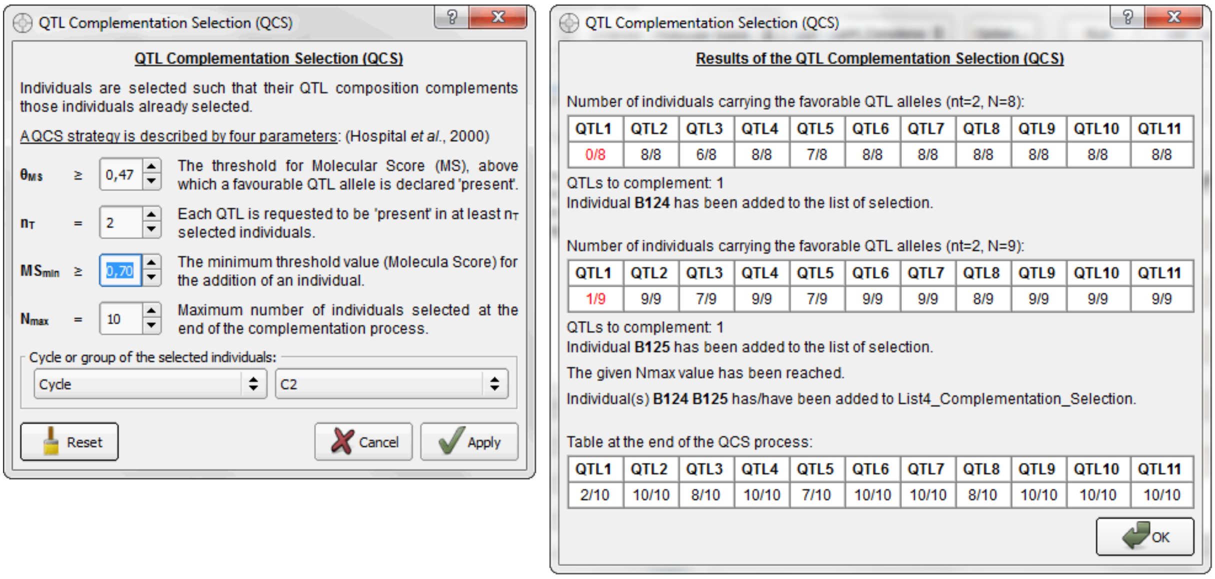
No. ind = 10
Mean = 0.756312

No. group = 1
Var = 0.00140282

Figure 16: QTL Complementation Selection

Create a new empty list by pressing the "Add" button. Rename it to "List4_Complementation_Selection" (by double-click). Apply the "Truncation Selection" on it based on the MS (N_{sel} = 8, Criterion = MS, List = List4_Complementation_Selection, Option: Cycle = C2 then press "Run").

The eight first individuals (B8... B242, see Fig. 16) are added to the list. The colored view points that we are losing the favorable allele for QTL 1 (in red, QTL1 = 0 for all individuals). As an example, the QCS is applied in order to find two candidates such that their QTL composition complements those eight individuals already selected (see Fig. 16 & 17).



QTL Complementation Selection (QCS)

QTL Complementation Selection (QCS)

Individuals are selected such that their QTL composition complements those individuals already selected.

A QCS strategy is described by four parameters: (Hospital *et al.*, 2000)

θ_{MS} ≥ 0,47 — The threshold for Molecular Score (MS), above which a favourable QTL allele is declared 'present'.

n_T = 2 — Each QTL is requested to be 'present' in at least n_T selected individuals.

MS_{min} ≥ 0,70 — The minimum threshold value (Molecula Score) for the addition of an individual.

N_{max} = 10 — Maximum number of individuals selected at the end of the complementation process.

Cycle or group of the selected individuals: Cycle — C2

QTL Complementation Selection (QCS)

Results of the QTL Complementation Selection (QCS)

Number of individuals carrying the favorable QTL alleles (nt=2, N=8):

QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7	QTL8	QTL9	QTL10	QTL11
0/8	8/8	6/8	8/8	7/8	8/8	8/8	8/8	8/8	8/8	8/8

QTLs to complement: 1
Individual **B124** has been added to the list of selection.

Number of individuals carrying the favorable QTL alleles (nt=2, N=9):

QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7	QTL8	QTL9	QTL10	QTL11
1/9	9/9	7/9	9/9	7/9	9/9	9/9	8/9	9/9	9/9	9/9

QTLs to complement: 1
Individual **B125** has been added to the list of selection.

The given Nmax value has been reached.

Individual(s) **B124 B125** has/have been added to List4_Complementation_Selection.

Table at the end of the QCS process:

QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7	QTL8	QTL9	QTL10	QTL11
2/10	10/10	8/10	10/10	7/10	10/10	10/10	8/10	10/10	10/10	10/10

Figure 17: QTL Complementation Selection algorithm with parameters

Select the "List4_Complementation_Selection" in the QCS section. Click on the QCS "Option..." button to set up the parameters shown in Fig. 17 (on the left side) and press "Apply". The result appears in a pop-up window (on the right).

<u>The QCS is described by five parameters</u>:

θ_{MS}: corresponds to the MS QTL threshold (QTLx > 0.47 by default), above which a favorable QTL allele is declared "present". In this case and depending on the threshold value, not only individuals considered as homozygous for the favorable allele at QTL position will be taken into account (e.g. heterozygous individuals, see Fig. 16, in yellow).

n_T: means that each favorable QTL allele is requested to be "present" in at least n_T selected individuals (here n_T = 2).

MS_{min}: the minimum threshold value ($MS_{min} \geq 0$, by default) for the addition of an individual. In this example, individuals with MS (genetic value) < 0.7 are not considered.

N_{max}: the maximum number of individuals selected at the end of the QCS process. Here, up to two individuals will be added to the final subset of selected individual (N_{max} = 10).

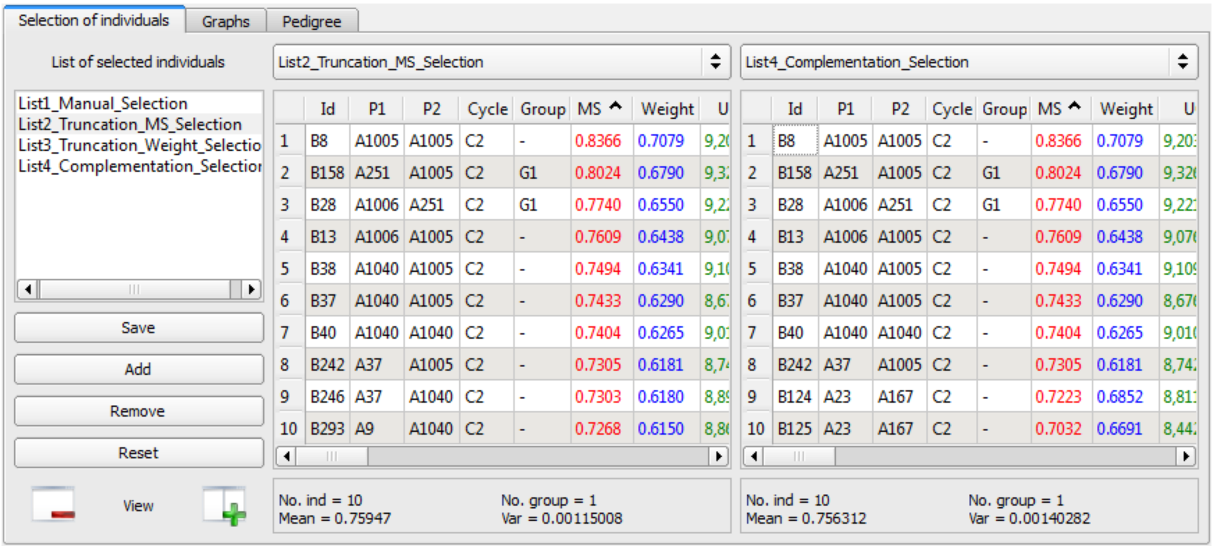
Cycle/Group: optional information regarding the generation of selection or another classification criterion in the program (e.g. first cycle, second cycle, F2, F4, subprograms, families, etc). "C2" (cycle 2) was selected instead of "None" (no selection, by default) in order to select individuals that belong to the last cycle of selection.

This approach can be applied to any predefined list (including the possibility to consider an empty list). In this example, the first eight individuals (N_0'= 8) were selected via the truncation selection (based on the MS criterion). Then: (i) the QTL for which the favorable alleles are "present" in fewer than n_t selected individuals are identified (see Fig. 17, in red); (ii) among the remaining individuals (belonging to *Cycle*), taken in order of decreasing MS (with MS ≥ MS_{min}), OptiMAS searches for the individual having favorable alleles "present" at the largest number of those QTL identified in (i). This individual is added to the subset of selected individuals (i.e. $N = N_0' + 1$). Steps (i) and (ii) are iterated until either of the following conditions is met: favorable alleles at all QTL are present in at least n_t individuals of the selected subset, or the number of individuals in the selected subset reaches the given N_{max} value, or it is not possible to find an individual in step (ii).

Note that in step (ii), MS is a secondary criterion; individuals are taken based on their ability to complement the subset of already selected individuals. Thus, with the present set of parameters, two individuals (B124, B125) have been added to the selected list (see more details on the right side of Fig. 17).

5.3.2 Display and comparison of lists of selected individuals

The different lists of selected individuals can be compared in two parallel tables (see below) displayed by default (which can be reduced to one list by clicking on the "-" button).



List1_Manual_Selection
List2_Truncation_MS_Selection
List3_Truncation_Weight_Selectio
List4_Complementation_Selectior

List2_Truncation_MS_Selection

	Id	P1	P2	Cycle	Group	MS	Weight	U
1	B8	A1005	A1005	C2	-	0.8366	0.7079	9,20
2	B158	A251	A1005	C2	G1	0.8024	0.6790	9,3
3	B28	A1006	A251	C2	G1	0.7740	0.6550	9,2
4	B13	A1006	A1005	C2	-	0.7609	0.6438	9,0
5	B38	A1040	A1005	C2	-	0.7494	0.6341	9,1
6	B37	A1040	A1005	C2	-	0.7433	0.6290	8,6
7	B40	A1040	A1040	C2	-	0.7404	0.6265	9,0
8	B242	A37	A1005	C2	-	0.7305	0.6181	8,7
9	B246	A37	A1040	C2	-	0.7303	0.6180	8,8
10	B293	A9	A1040	C2	-	0.7268	0.6150	8,8

No. ind = 10 No. group = 1
Mean = 0.75947 Var = 0.00115008

List4_Complementation_Selection

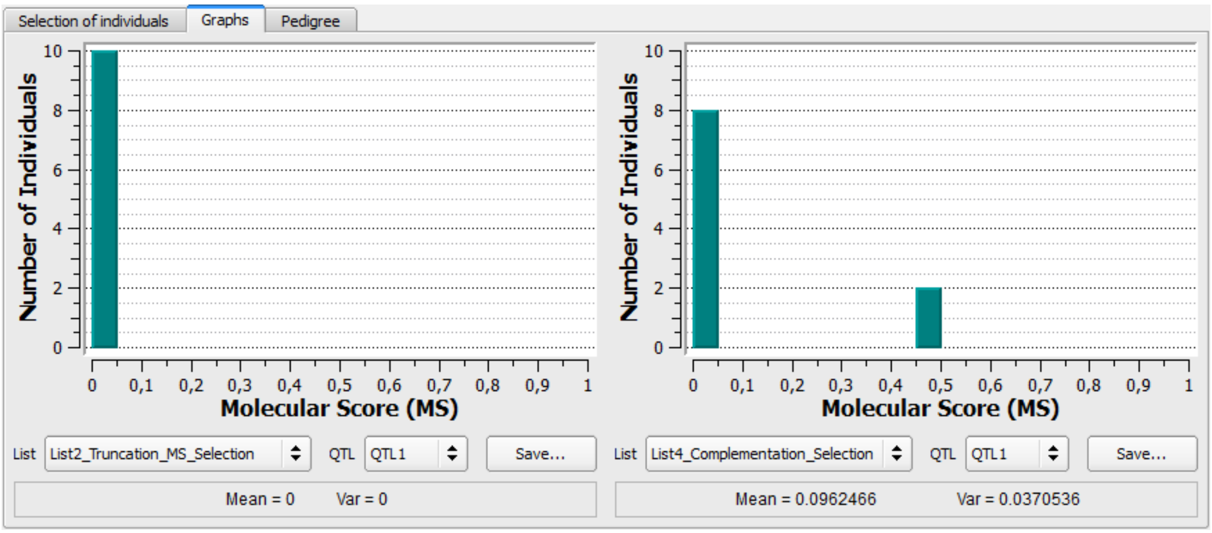
	Id	P1	P2	Cycle	Group	MS	Weight	U
1	B8	A1005	A1005	C2	-	0.8366	0.7079	9,20
2	B158	A251	A1005	C2	G1	0.8024	0.6790	9,32
3	B28	A1006	A251	C2	G1	0.7740	0.6550	9,22
4	B13	A1006	A1005	C2	-	0.7609	0.6438	9,07
5	B38	A1040	A1005	C2	-	0.7494	0.6341	9,10
6	B37	A1040	A1005	C2	-	0.7433	0.6290	8,67
7	B40	A1040	A1040	C2	-	0.7404	0.6265	9,01
8	B242	A37	A1005	C2	-	0.7305	0.6181	8,74
9	B124	A23	A167	C2	-	0.7223	0.6852	8,81
10	B125	A23	A167	C2	-	0.7032	0.6691	8,44

No. ind = 10 No. group = 1
Mean = 0.756312 Var = 0.00140282

Figure 18: comparison between lists of selected individuals (via tables)

If we compare the two lists of ten individuals selected via the MS (on the left) and QCS (on the right), we can see that the two last individuals (*B246 and B293)* are replaced by *B124* and *B125* in the QCS list. The selection of these two individuals will bring two more unfavorable alleles (QTL5 and QTL8 in red, not presented in Fig. 18) in the next generation.

The different lists of individuals can also be compared via the *Graphs* section window (***click on the Graphs table***, see Fig. 19).



List List2_Truncation_MS_Selection QTL QTL1

Mean = 0 Var = 0

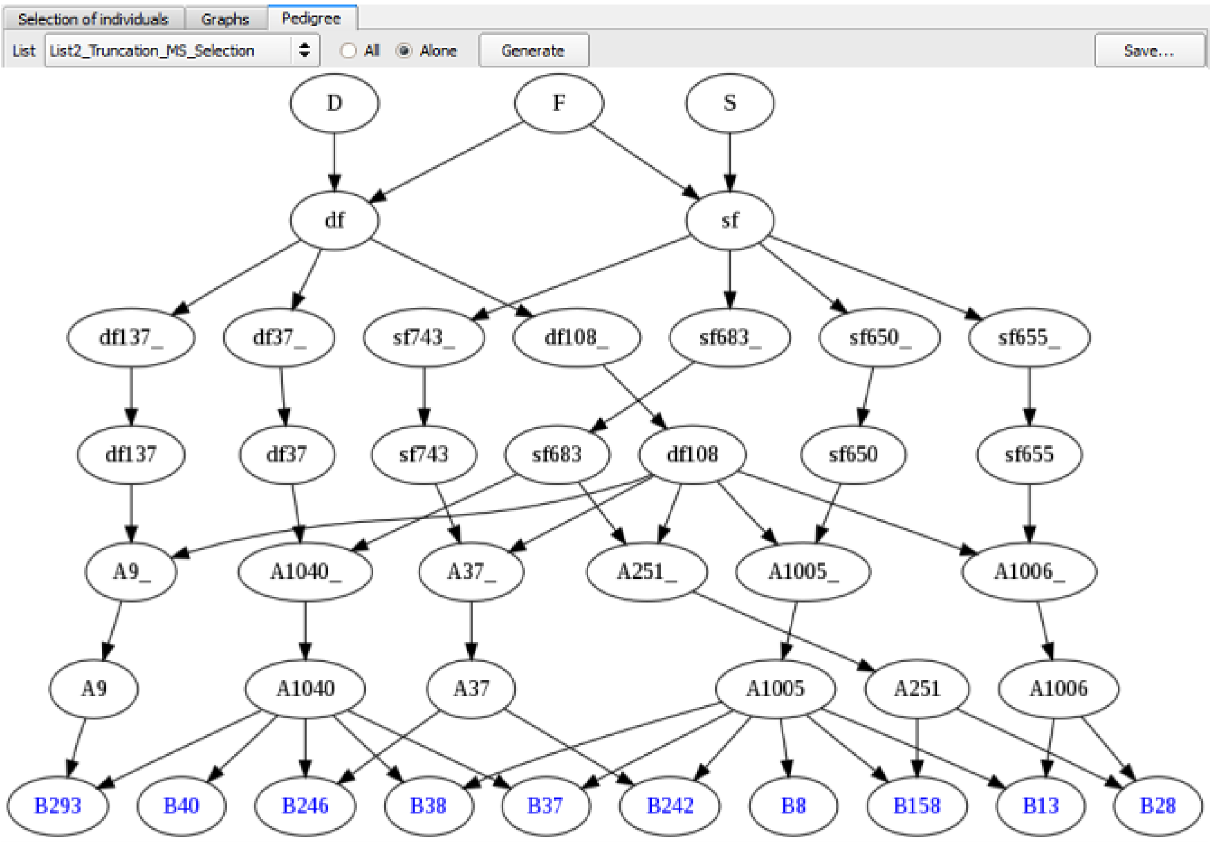
List List4_Complementation_Selection QTL QTL1

Mean = 0.0962466 Var = 0.0370536

Figure 19: comparison between lists of selected individuals (via graphs)

This example shows that use of truncation selection based on MS leads to the selection of individuals all carrying the unfavorable allele at QTL1 (see above, histogram on the left). On the right side, the two individuals (*B124*, *B125*) added via the QCS procedure are observed with a MS comprised between 0.45 and 0.5. In addition, if we ***select QTL7 in both lists***, we notice that the QTL7 will be fixed for the favorable allele at the next generation (the ten individuals have a MS > 0.95 in the graph, not represented in Fig. 19).

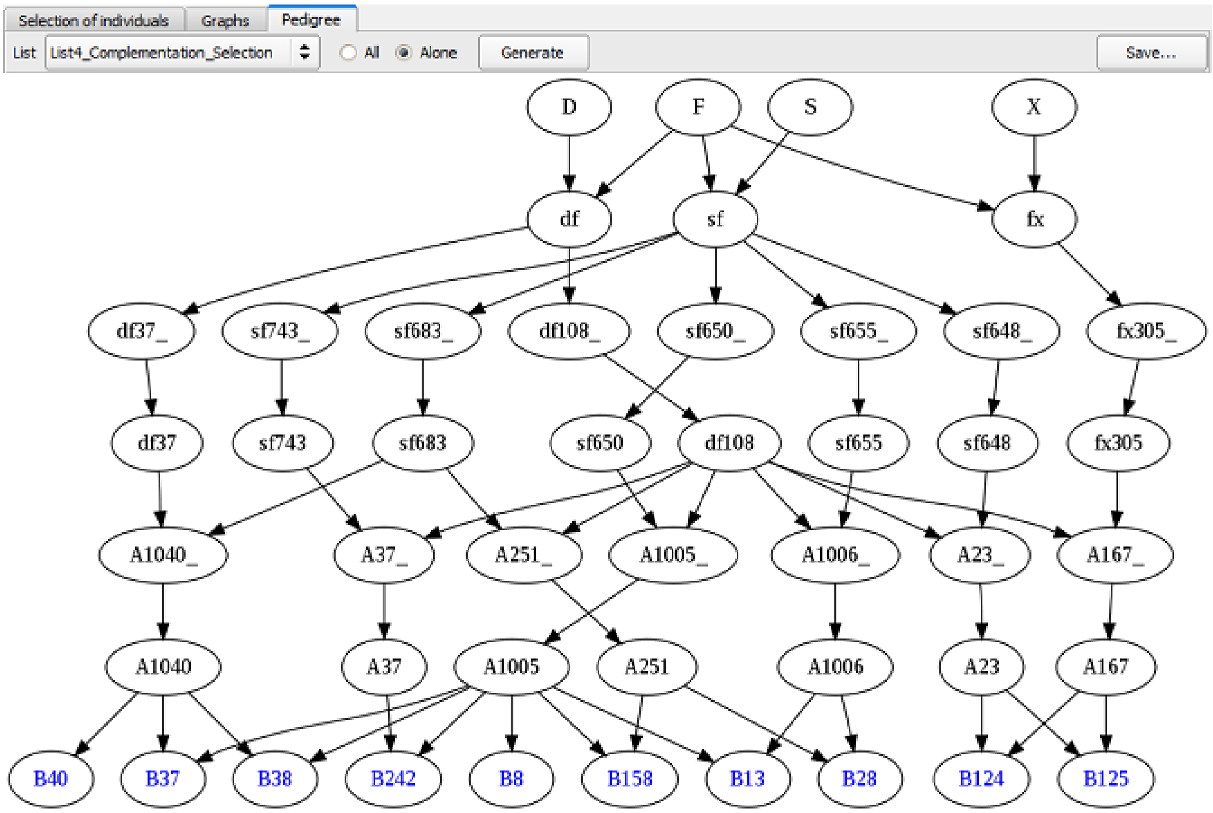
To visualize the origin of the selected individuals of each list, the user can display their pedigree (see below). ***Move to "Pedigree" table, select the list of your choice, check "Alone" (to only have the individuals of the selected list) and click on the Generate button.***



List List2_Truncation_MS_Selection ○ All ● Alone

Figure 20: pedigree of individuals selected via the Truncation Selection method

It is possible to compare pedigrees coming from different lists of selection (see Fig. 20 & 21, Truncation Selection list vs. QCS, respectively). Note that the two individuals added via the QCS bring the parental allele *x*. So, if we use the QCS list to produce the next generation, the four parental alleles will be present in the next cycle.



List: List4_Complementation_Selection ○ All ● Alone

Figure 21: pedigree of individuals selected via the QCS method

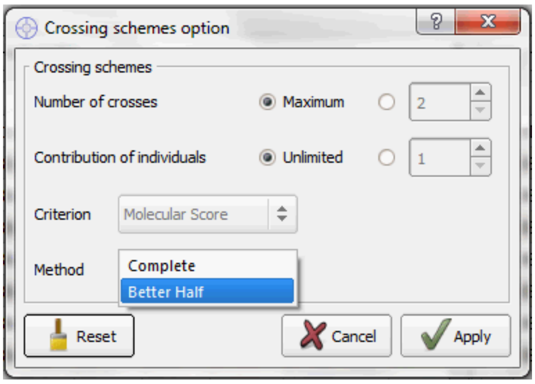
This representation is useful to follow the contribution of selected individuals over generations of selection and to prevent possible bottlenecks (individuals coming from a reduced number of parents at a given generation), in order to limit risk of drift (which may lead for instance to the fixation of an undesired phenotypic type for traits not considered in the MARS process). It also can be used to maintain diversity for selection on traits complementary to those considered for the MARS process.

5.4 Step 3: Identification of crosses to be made among selected individuals

Now that your list(s) of selected individuals is/are established, it is necessary to identify the crosses to be made to initiate the next MARS cycle. We addressed crosses between individuals of a single list (diallel design) or two complementary lists (**factorial design**). The diallel situation can be managed with three options: (i) the automatic definition of the whole list of possible crosses according to a **half-diallel** (complete method, see Fig. 22, 23, 24 and 25), (ii) the **"better-half"** strategy (Bernardo *et al.*, 2006) which consists of avoiding crosses between selected individuals with the lowest scores (see Fig. 22, 23, 24 and 25) and (iii) application of **constraints on the contribution of parents or on the maximum number of crosses** (see Fig. 26 and 27). In this last case, best crosses are determined according to either the (weighted) molecular score or the utility criterion. Then, lists of crosses created via the different methods can be analyzed and compared via graphs.

Example of instructions to create the two lists of crosses that will contain results of the half-diallel and the better-half processes among the previously selected candidates:

Rename the two empty lists of crosses (by double-click) in List1_Half_Diallel_from_QCS and List2_Better_Half_from_QCS. Then, select the appropriate method via the "Option..." button (see Fig. 22 & 23).



Crossing schemes option

Crossing schemes

Number of crosses ● Maximum ○ 2

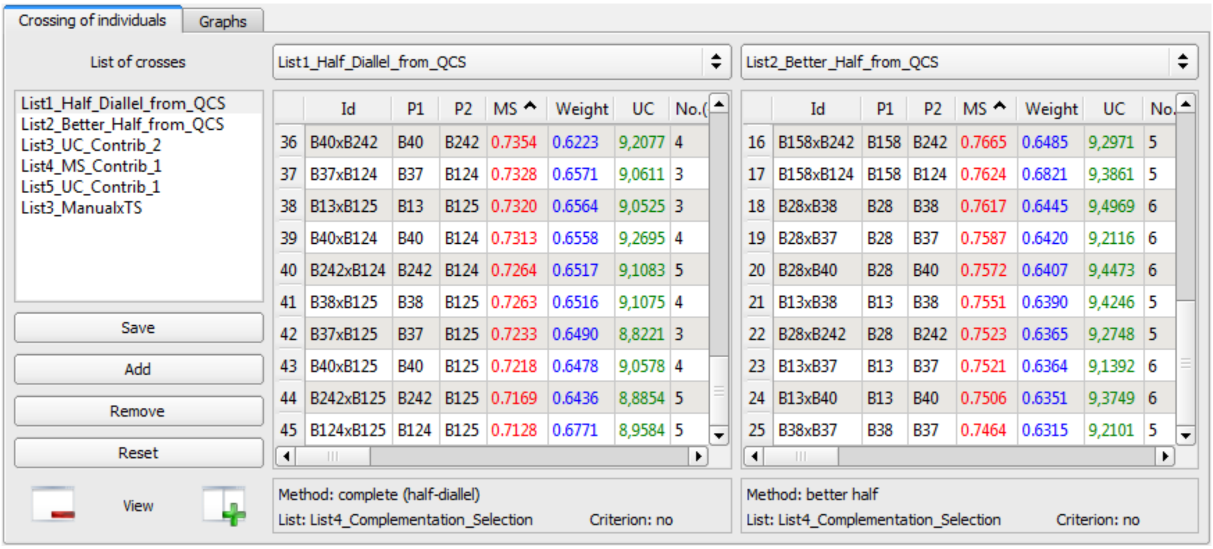
Contribution of individuals ● Unlimited ○ 1

Criterion: Molecular Score

Method: Complete / Better Half

Figure 22: crossing schemes options (method selection)

Choose the list of selected candidates coming from the step 2 (e.g. List4_Complementation_Selection) and the adequate list of crosses (e.g. List1_Half_Diallel_from_QCS & List2_Better_Half_from_QCS). Finally, click on the Run button to see the results of these crosses stored to the appropriate list (see below).



List1_Half_Diallel_from_QCS
List2_Better_Half_from_QCS
List3_UC_Contrib_2
List4_MS_Contrib_1
List5_UC_Contrib_1
List3_ManualxTS

List1_Half_Diallel_from_QCS

	Id	P1	P2	MS	Weight	UC	No.(
36	B40xB242	B40	B242	0.7354	0.6223	9,2077	4
37	B37xB124	B37	B124	0.7328	0.6571	9,0611	3
38	B13xB125	B13	B125	0.7320	0.6564	9,0525	3
39	B40xB124	B40	B124	0.7313	0.6558	9,2695	4
40	B242xB124	B242	B124	0.7264	0.6517	9,1083	5
41	B38xB125	B38	B125	0.7263	0.6516	9,1075	4
42	B37xB125	B37	B125	0.7233	0.6490	8,8221	3
43	B40xB125	B40	B125	0.7218	0.6478	9,0578	4
44	B242xB125	B242	B125	0.7169	0.6436	8,8854	5
45	B124xB125	B124	B125	0.7128	0.6771	8,9584	5

Method: complete (half-diallel)
List: List4_Complementation_Selection Criterion: no

List2_Better_Half_from_QCS

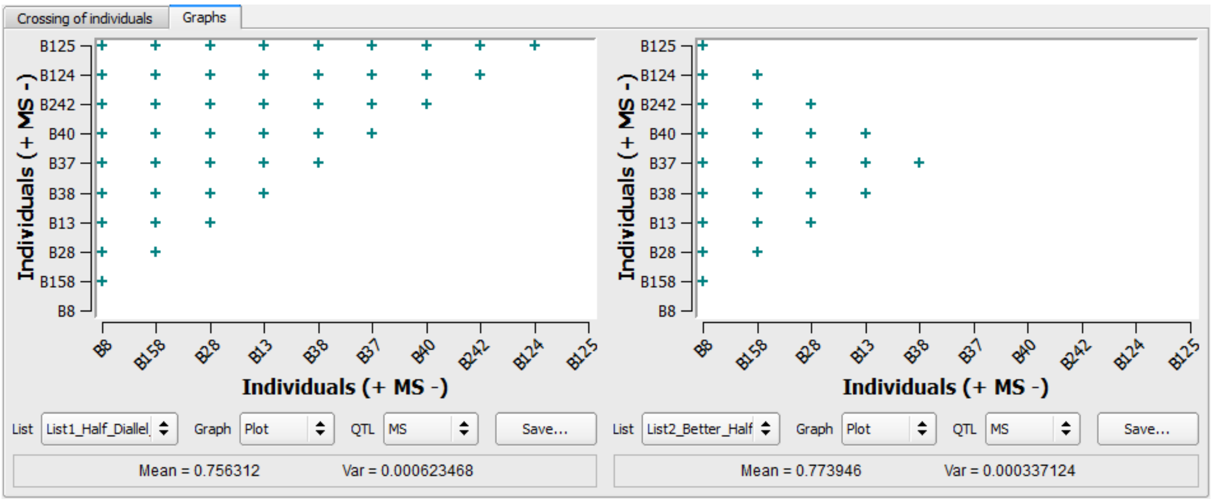
	Id	P1	P2	MS	Weight	UC	No.
16	B158xB242	B158	B242	0.7665	0.6485	9,2971	5
17	B158xB124	B158	B124	0.7624	0.6821	9,3861	5
18	B28xB38	B28	B38	0.7617	0.6445	9,4969	6
19	B28xB37	B28	B37	0.7587	0.6420	9,2116	6
20	B28xB40	B28	B40	0.7572	0.6407	9,4473	6
21	B13xB38	B13	B38	0.7551	0.6390	9,4246	5
22	B28xB242	B28	B242	0.7523	0.6365	9,2748	5
23	B13xB37	B13	B37	0.7521	0.6364	9,1392	6
24	B13xB40	B13	B40	0.7506	0.6351	9,3749	6
25	B38xB37	B38	B37	0.7464	0.6315	9,2101	5

Method: better half
List: List4_Complementation_Selection Criterion: no

Figure 23: comparison of the two lists (half-diallel vs. better-half) of couples generated

On the left table (see *List1_Half_diallel_From_QCS*), all the individuals were crossed together. For each of the 45 resulting pairs, a virtual individual is created and OptiMAS computes the expected molecular score of the progeny of the cross for all and each QTL. On the right table, the *better-half* strategy leads to 25 couples (see also Fig. 24). QTL columns can still be sorted, in order for instance to identify pairs having a score of 0. Then, these crosses can be deleted (***right click on the selected pair and press the "delete..." button***).

Meanwhile, graphs are automatically generated to display a view of the selected crosses based on the different strategies (see below).



List List1_Half_Diallel Graph Plot QTL MS

Mean = 0.756312 Var = 0.000623468

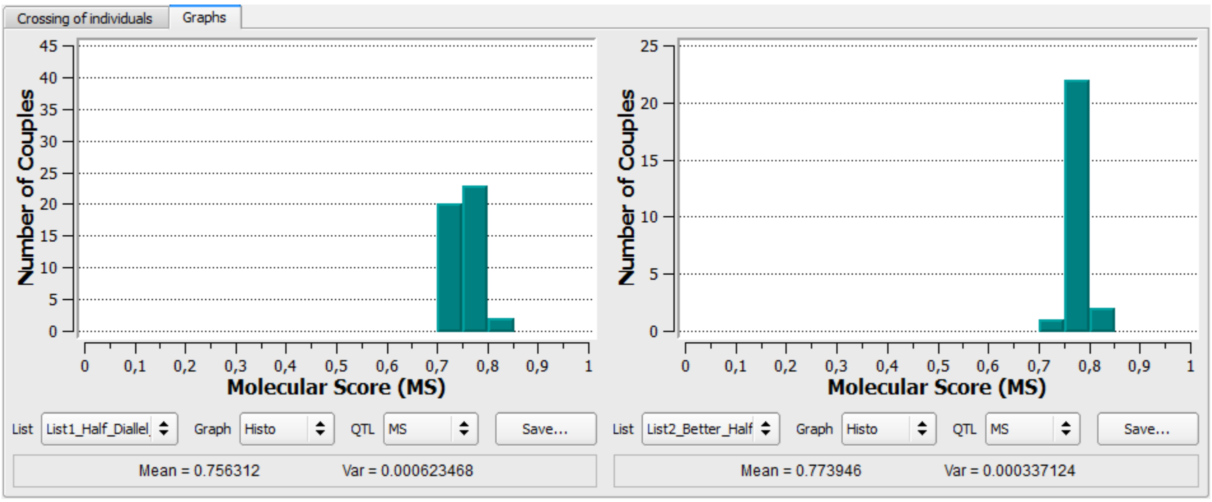
List List2_Better_Half Graph Plot QTL MS

Mean = 0.773946 Var = 0.000337124

Figure 24: half-diallel vs. better-half strategies

In the figure above, individuals are ranked on the two axes based on their genetic value (MS from highest to lowest). On the left side, the "*Complete"* procedure (half-diallel) has been applied. All the individuals were crossed together with no limitation on their contribution whereas on the right side (*better-half* method), crosses between individuals having lowest MS have been avoided (i.e. *B37* to *B125*).

Lists of crosses created via the different methods and/or constraints can also be analyzed and compared via histograms (i.e. ***select Graph = Histo***, see below).



List List1_Half_Diallel Graph Histo QTL MS

Mean = 0.756312 Var = 0.000623468

List List2_Better_Half Graph Histo QTL MS

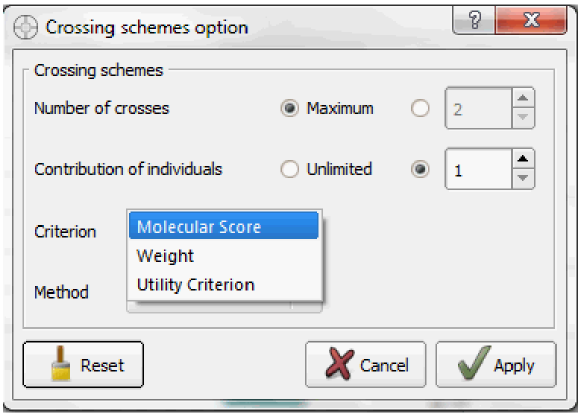
Mean = 0.773946 Var = 0.000337124

Figure 25: comparison of the two lists (half-diallel Vs better-half) of couples generated

As expected from the higher relative contribution of individual B8, the average MS is higher for *Better Half* option (0.756312 vs. 0.773946) and the variance of MS among generated couples is lower (6.23 10^{-4} vs. 3.37 10^{-4}).

Constraints on the contribution of parents or on the maximum number of crosses to be done can be applied (see Fig. 26, 27). In this case crosses determination is optimized based on the expected molecular score and UC of the crosses.

This will be exemplified with two new lists of crosses. ***Press two times the "Add" button to create two new lists of crosses and rename them List4_MS_Contrib_1 and List5_UC_Contrib_1. Then Press the "Option..." button to open the "Crossing schemes option" window (see below).***



Crossing schemes option

Crossing schemes

Number of crosses ● Maximum ○ 2

Contribution of individuals ○ Unlimited ● 1

Criterion Molecular Score

Weight

Utility Criterion

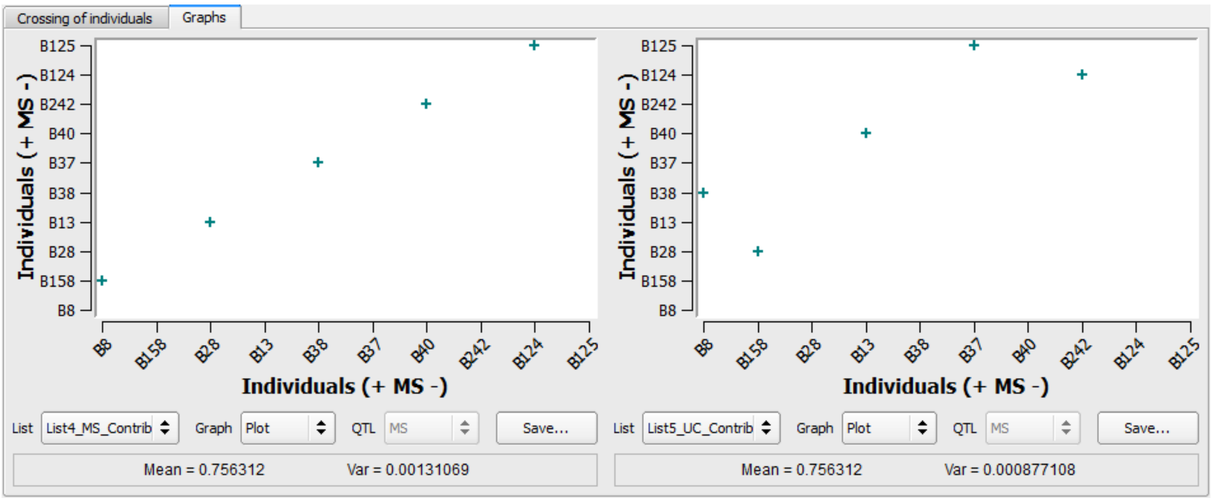
Method

Figure 26: crossing schemes options (constraints)

To specify that each of the ten candidates must be crossed only once and that there is no constraint on the maximum number of crosses to be done**,** ***select Contribution of individuals = 1*** and leave ***Number of crosses = Maximum.*** The criterion box will appear as in Fig. 26. ***Select Criterion = Molecular Score for the first list and press the "Apply" button.***

To apply this strategy to candidates previously selected via the QCS method, ***select List selection(s) = List4_Complementation_Selection. Then, select the list where crosses will be stored by selecting List crosses = List4_MS_Contrib_1. Press the "Run" button to create this first list.***

To apply the same constraints based the utility criterion of the pairs (see Fig. 27 on the right), ***do the same thing with Criterion = Utility Criterion and List(s) selection = List5_UC_Contrib_1.***



List List4_MS_Contrib Graph Plot QTL MS Save...

Mean = 0.756312 Var = 0.00131069

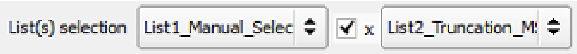
List List5_UC_Contrib Graph Plot QTL MS Save...

Mean = 0.756312 Var = 0.000877108

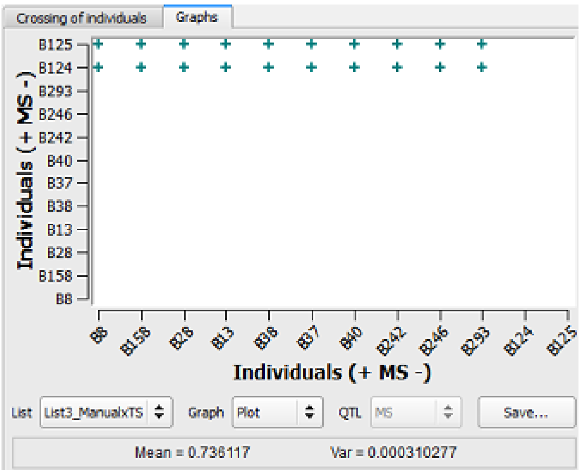
Figure 27: constraints on the contribution of parents based on the MS or the Utility Criterion

On the left side, individuals with the higher MS are crossed together (e.g. B8xB158, B28xB13, etc) whereas on the right side, the use of the utility criterion resulted in different pairs (e.g. *B8xB38*).

A factorial design between two lists of selected plants can be applied. Checking the box between the two lists enables the possibility to select a second list of selected candidates (see below).



List(s) selection: List1_Manual_Selec ☑ x List2_Truncation_M:



List: List3_ManualxTS
Graph: Plot
QTL: MS

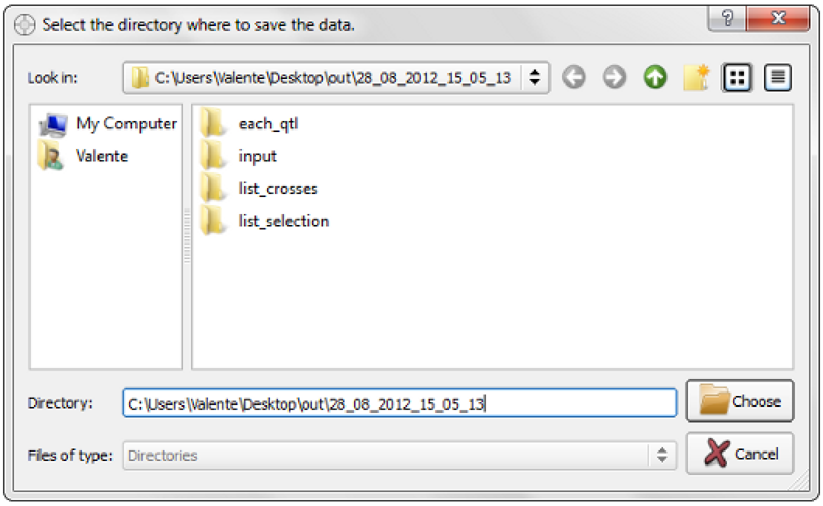
Mean = 0.736117 Var = 0.000310277

Figure 28: factorial design applied between two lists of selection

Thus, a list containing only two individuals (*B124* and *B125)* previously found via the QCS method has been crossed with the ten individuals having the highest MS (i.e. List1_TS_MS) resulting in a factorial design displayed in Fig. 28.

5.5 Saving/reloading your previous analysis

Lists of selected individuals and crosses can be saved into the results folder. On the menu bar, ***Press Data*** **>** ***Save all...*** **>** ***then press the "Choose" button.***



Select the directory where to save the data.

Look in: C:\Users\Valente\Desktop\out\28_08_2012_15_05_13

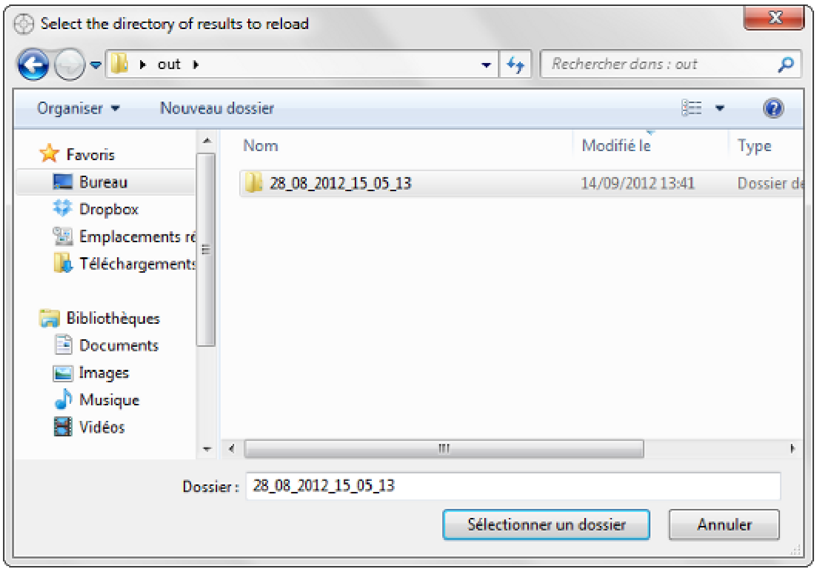
Directory: C:\Users\Valente\Desktop\out\28_08_2012_15_05_13

Files of type: Directories

Figure 29: saving your current analysis

By default all the lists are saved to the appropriate directory specified in Fig. 5.

Once the files are saved, you can close OptiMAS and reopen your analysis later by selecting the results folder previously saved and loading it (see below). On the menu bar ***Press File*** **>** ***Reload Data...*** **>** ***Select your previous folder*** **>** ***OK.***

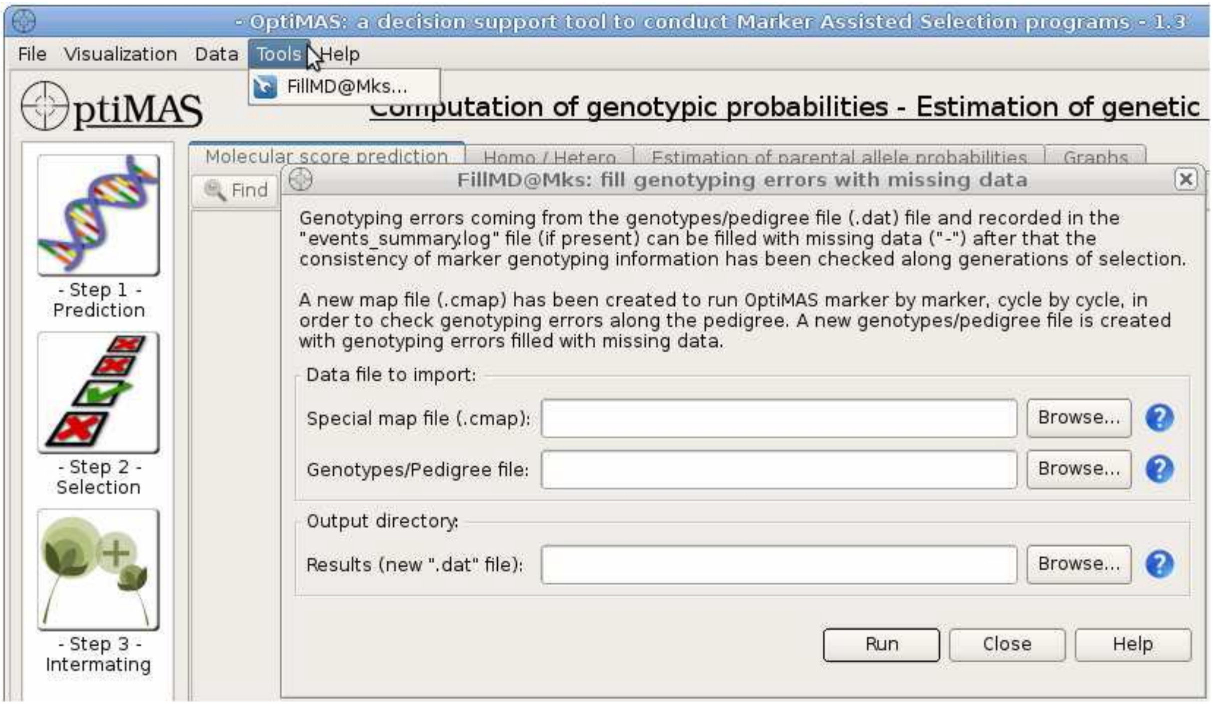


Nom	Modifié le	Type
28_08_2012_15_05_13	14/09/2012 13:41	Dossier de

Dossier : 28_08_2012_15_05_13

5.6 FillMD@Mk: a tool to replace genotyping errors with missing data

Genotyping errors coming from the genotypes/pedigree file (.dat) file and recorded in the "events_summary.log" file (if present) can be filled with missing data ("-") after that the consistency of marker genotyping information has been checked along generations of selection. ***Select Tools*** **>** ***FillMD@Mks...*** from the menu bar (see below).



- OptiMAS: a decision support tool to conduct Marker Assisted Selection programs - 1.3
File
Visualization
Data
Tools
Help
FillMD@Mks...
OptiMAS
Computation of genotypic probabilities - Estimation of genetic
Molecular score prediction
Homo / Hetero
Estimation of parental allele probabilities
Graphs
Find
- Step 1 -
Prediction
- Step 2 -
Selection
- Step 3 -
Intermating
FillMD@Mks: fill genotyping errors with missing data
Browse...
Run
Close
Help

Genotyping errors coming from the genotypes/pedigree file (.dat) file and recorded in the "events_summary.log" file (if present) can be filled with missing data ("-") after that the consistency of marker genotyping information has been checked along generations of selection.

A new map file (.cmap) has been created to run OptiMAS marker by marker, cycle by cycle, in order to check genotyping errors along the pedigree. A new genotypes/pedigree file is created with genotyping errors filled with missing data.

Data file to import:

Special map file (.cmap):

Genotypes/Pedigree file:

Output directory:

Results (new ".dat" file):

Figure 31: data set importation to run the FillMd@Mks tool

Loading input/output files using the browser:

Map file: path to a new genetic map/QTL position input file (.cmap). This file has been created at the end of a previous run of OptiMAS in order to re-run OptiMAS marker by marker and localize genotyping errors at individual marker position.

Genotype file: path to the genotype/pedigree file (.dat). This is the same file used to run OptiMAS the first time.

Output directory: path to the folder where a new genotype/pedigree file (.dat) with inconsistent genotyping data replaced by missing data (“-“) will be stored. Note that your output directory should not be in the “Program Files” folder or other specific directories with administrator privileges.

Click on the “Run” button to create this new genotype/pedigree file (new_*file*.dat). Use this new .dat file to re-run OptiMAS.

6 Future work

Next steps coming soon:

- Addition of a quantitative score column (Genomic Selection prediction or phenotypic value).
- Development of a simulation procedure (Step 4) to produce a “virtual” next generation (with information on the number of individuals needed to reach the ideotype).
- Computation of diversity score based on pedigree (effective population size).
- New algorithm to compute genotypic probabilities with no limitation on the number of flanking markers around the QTL position (option to use a sliding window to compute probabilities along the genome).
- Linkage between QTL in the estimation of the utility criterion.
- Manage the QTL position uncertainty in score computation.
- Computation of diversity score based on markers outside QTL.
- Handle allelic effects at QTL in order to compute expected gain for different traits with the possibility to weight them to compute indexes.
- A wizard to help users who want to run automatically basic options of the tool.

7 How to cite this program

In publications including results from the use of this program, please specify the version of the software you used.

Valente, F., Gauthier, F., Bardol, N., Blanc, G., Joets, J., Charcosset, A. & Moreau, L. (*in prep*) OptiMAS: decision support for marker-assisted selection.

8 Contact

Please send bug reports and/or requests for new features to Fabio Valente (fvalente@moulon.inra.fr) or Laurence Moreau (moreau@moulon.inra.fr).

9 Acknowledgments

We are grateful to colleagues within INRA and the Generation Challenge Program (GCP) for enlightening user oriented input to this project. This development has benefited from the advices and beta-testing of Delphine Fleury and Mark Sawkins.

This program is part of the IBP within GCP and is funded by the Bill and Melinda Gates Foundation.

10 References

Bernardo, R., Moreau, L. and Charcosset, A. (2006) Number and fitness of selected individuals in marker-assisted and phenotypic recurrent selection, *Crop Sci.*, **46**, 1972-1980.

Blanc, G., Charcosset, A., Mangin, B., Gallais, A. and Moreau, L. (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize, *Theor. Appl. Genet.*, **113**, 206-224.

Blanc, G., Charcosset, A., Veyrieras, J.B., Gallais, A. and Moreau, L. (2008) Marker-assisted selection efficiency in multiple connected populations: a simulation study based on the results of a QTL detection experiment in maize, *Euphytica*, **161**, 71-84.

Hospital, F., Goldringer, I. and Openshaw, S. (2000) Efficient marker-based recurrent selection for multiple quantitative trait loci, *Genet. Res.*, **75**, 357-368.

Huang, Y.F., Madur, D., Combes, V., Ky, C.L., Coubriche, D., Jamin, P., Jouanne, S., Dumas, F., Bouty, E., Bertin, P., Charcosset, A. and Moreau, L. (2010) The Genetic Architecture of Grain Yield and Related Traits in Zea maize L. Revealed by Comparing Intermated and Conventional Populations, *Genetics*, **186**, 395-U612.

Ribaut, J.M., de Vicente, M.C. and Delannay, X. (2010) Molecular breeding in developing countries: challenges and perspectives, *Curr. Opin. Plant Biol.*, **13**, 213-218.