

# BioMercatorV4 USER GUIDE

March 2012

Olivier Sosnoswki, Johann Joets

UMR de Génétique Végétale du Moulon  
F-91190 Gif-sur-Yvette

Introduction .....	4
Install.....	5
Memory management .....	6
Input files .....	7
BioMercator V3/V4 format.....	7
Genetic maps.....	7
Overview .....	7
Detailed format .....	7
QTLs .....	8
BioMercator V2 format.....	9
Map file.....	9
Qtl file .....	9
MetaQTL package format .....	10
MCQTL format .....	10
Files loader wizard.....	10
Workspace management.....	10
Data and analyses storage .....	11
Graphical User Interface.....	12
The menu bar .....	12
The option panel.....	13
Maps display .....	14
Zoom .....	14
Whole drawing panel zoom .....	14
Chromosome cascading zoom .....	15
Individual chromosome zoom.....	16
Dynamic map comparison .....	17
Chromosome/Linkage group Flipping.....	17
Genetic maps compilation.....	19
Overview .....	19
BioMercator V2 tools.....	19
Iterative maps compilation.....	19
QTL Meta-analyses (S. Gerber, B. Goffinet method).....	20
MetaQTL package tools .....	20
Maps connectivity analysis: Infomap .....	20
Common markers visualization: MMapView .....	22
Consensus map building: ConsMap.....	23
Qtls projection: QTLProj .....	24
Meta-analysis step 1/2 (Veyrieras): QTLClust.....	25
Meta-analysis step 2/2 (Veyrieras): MQTLView .....	27
Genome structural and functional datasets .....	29
Loading a genome version .....	29
Genetic map and genome annotation display .....	30

Functional annotation display .....	31
The Gene Ontology .....	31
GO terms representation analysis .....	32

## Introduction

BioMercator V4 is a genetic map compilation and QTL meta-analysis software providing tools to integrate these data with genome structural and functional annotation. It offers wizards to run analyses and display input and output maps through a user friendly graphical user interface (GUI). In addition to map compilation and meta-analysis methods available in previous BioMercator version, this new release includes algorithms from the MetaQTL package (<http://www.bioinformatics.org/mqtl/wiki/>).

The GUI has evolved, and now offers new ways of zooming. A cascading zoom allows to progressively enlarging a region of a map while conserving alongside the whole chromosome representation. It is now possible to use the zoom from the whole map representation to focus on zones of interest of several chromosomes. Image can be exported in various formats including SVG for subsequent image editing in third-party software.

BioMercator V4 also implements a new file loading wizard capable of reading an unlimited number of files and automatically recognizes their format.

The possibility of using MetaQTL's package algorithms directly inside BioMercator V4 came with a file format evolution; to use fully those algorithms, more data will be required than in the previous BioMercator version, while.

The user will be able to export maps and analyses in a format compliant for submission to the URGI databases (GnpIS).

## **Install**

BioMercator is a Java program; all you need is Java (v1.5 or above) installed on your machine. To install java see <http://www.java.com/>

unzip the BioMercator archive file in a directory.

**On Windows:** Double-click on the BioMercatorV4.jar to launch the program

**On other OS:** Open a terminal and execute the command line:

```
“java -jar BioMercatorV4.jar”
```

## **Memory management**

On any OS, if the software doesn't launch or exits with a memory error (when working with large genome annotation dataset), you should open a console (terminal) and type the following line:

```
“java -jar -Xms512m BioMercatorV4.jar”
```

## Input files

### ***BioMercator V3/V4 format***

## Genetic maps

Genetics maps has to be described in a text file following the format in fig . It contains all required data and can be easily edited with a text editor. (A sample file is provided with the tutorial)

## Overview

```
Organism Genus=Zea
Organism Species=mays
Parent=b73
Parent=mo17
crossType=F2 intercross
popSize=242
mappingCrossType=SF2
mappingFunction=Haldane
mapName=IBM
mapUnit=cM
mapExpansion =0
mapQuality=3
locusLocation=0
chr=chr1
lg=linkage_group 1
1      phi097  34.9
2      cdo1081a  16.7
3      umc1977  12.8  EST
4      psb201b  23.7
5      umc157  46.6
6      psb115c  11.9  EST
7      bnlg1007  23.4  EST
8      umc76  23.5
```

*Fig. 1: BioMercatorV3  
example map file*

## Detailed format

### Headers:

The header part contains 13 fields and the separator is the char 'equal' ('='); these fields are almost all mandatory, and are listed below:

- 'Organism Genus' [*mandatory*]: the organism genus
- 'Organism Species' [*mandatory*]: the organism species
- 'Parent' [*optional*]: The parent(s)
- ' CrossType ' [*mandatory*]: the cross type
- 'popSize' [*mandatory*]: The population type
- ' mapping CrossType ' [*mandatory*]: The mapping cross type
- ' mappingFunction' [*mandatory*]: The mapping function
- 'mapName' [*mandatory*]: The map name
- 'mapUnit' [*mandatory*]: The map unit (for the moment, can only be equal to 'cM')

- 'mapExpansion' [*mandatory*]: The map expansion
- 'mapQuality' [*mandatory*]: The map quality; a digit between 0 and 5 included
- 'locusLocation' [*mandatory*]: The locus location method; 0 for absolute, 1 for relative

### Chromosomes and linkage groups:

The markers of genetic maps should belong to a chromosome and a linkage group (even if chromosome has only one linkage group).

A chromosome map must start with “chr=*your\_chromosome\_name*”, and a linkage group with “lkg=*your\_linkage\_group\_name*”; a linkage group must be included in a chromosome.

For the next chromosome, just repeat the step above.

## QTLs

```
mapName=IBM
QTL1 Trait 2 ID:0 P1 Y1 chr1 linkage_group 1 3 10 360.47 351.47 369.47
QTL2 Trait3 ID:0 P1 Y1 chr1 linkage_group 1 3 22 127.84 118.84 136.84
QTL3 Trait4 ID:0 P1 Y1 chr1 linkage_group 1 3 4 169.7 160.7 178.7
QTL4 Trait4 ID:0 P1 Y1 chr1 linkage_group 1 3 14 527.75 518.75 536.75
QTL5 Trait5 ID:0 P1 Y1 chr1 linkage_group 1 3 4 402.8 393.8 411.8
QTL6 Trait6 ID:0 P1 Y1 chr1 linkage_group 1 3 7 169.7 160.7 178.7
QTL7 Trait6 ID:0 P1 Y1 chr1 linkage_group 1 3 6 616.5 607.5 625.5
QTL8 Trait 7 ID:0 P1 Y1 chr1 linkage_group 1 3 11 323.59 314.59 332.59
QTL9 Trait 7 ID:0 P1 Y1 chr1 linkage_group 1 3 9 155.95 146.95 164.95
QTL10 Trait8 ID:0 P1 Y1 chr1 linkage_group 1 3 14 508.5 499.5 517.5
QTL11 Trait 9 ID:0 P1 Y1 chr1 linkage_group 1 3 8 638.5 629.5 647.5
QTL12 Trait10 ID:0 P1 Y1 chr1 linkage_group 1 15.39 7 345 335.3 356.8
QTL13 Trait13 ID:0 P1 Y1 chr1 linkage_group 1 4.6 5.6 241.53 233.45 248.85
QTL14 Trait14 ID:0 P1 Y1 chr1 linkage_group 1 3 15 625.5 565.6 685.4
QTL15 Trait14 ID:0 P1 Y1 chr1 linkage_group 1 3 15 59.08 51.6 67.2
```

*Fig. 2: BioMercatorV3 example QTL file*

The first line corresponds to the map name; it should be as shown above, ie 'mapName=*your\_map*' and must be the same as the one in the genetic map file.

The next section of the file is tabulation-delimited and contains 12 columns, where each line corresponds to a single QTL in a given environment (for multi environments experiments). If a QTL is detected for *n* traits, then *n* lines will be completed. Effects can be specified in another file. (Spaces can be used in the columns as seen in the linkage group's)

- 1<sup>st</sup>: An ID for the QTL, this ID will be used to tag the QTL on the graphical interfaces or text outputs, and to help selecting the QTL when needed. It should be unique in order to avoid confusion
- 2<sup>nd</sup>: The trait
- 3<sup>rd</sup>: The trait ontology ID: International standard format is 'ontologyAcronym:ID'. If no ID exists, you should use 'ID:0'. Ontology ID never changes but may become obsolete.

(Visit <http://bioportal.bioontology.org/ontologies/40508>

or [http://www.gramene.org/plant/\\_ontology/index.html](http://www.gramene.org/plant/_ontology/index.html) for more details about ontologies.)

- 4<sup>th</sup>: The experiment place
- 5<sup>th</sup>: The experiment year
- 6<sup>th</sup>: The chromosome name
- 7<sup>th</sup>: The linkage group name



- 8<sup>th</sup>: The LOD Score
- 9<sup>th</sup>: The r<sup>2</sup> (numeric): the variance percentage
- 10<sup>th</sup>: The QTL most likely position (in cM)
- 11<sup>th</sup>: Confidence interval start position (in cM)
- 12<sup>th</sup>: Confidence interval stop position (in cM)

### **BioMercator V2 format**

This text format is the one used in the previous BioMercator's version but is still recognized by BioMercator V3 and V4, and is inspired from output file of cartography software MapMaker. (The following sample file is provided with the tutorial).

For more information, please refer to the previous version's documentation.

### **Map file**

```

mapname=IBM
poptype=F2 intercross
chromosome ch1
>map
=====
Map:
Markers      Distance
  1  isu110b    3.2 cM
  2  phi097     0.0 cM
  3  tub1       0.0 cM
  4  bnlg1124   1.2 cM
  5  tub4b     38.2 cM
  6  rz444d     9.0 cM
  7  psb201b   11.8 cM
  8  umc1977    0.9 cM
  9  umc1071    2.9 cM
 10  bnlg1014   5.9 cM
 11  umc1041    9.7 cM
 12  cdc1001a  28.7 cM

```

*Fig. 3: BioMercator V2 example map file*

### **Qtl file**

The QTLs second file is tabulation-delimited text file.

mapName	name	chromosome	trait	lodscore	r2	SIM	position	from	to
IBM	QTL	ch1	Trait 2 3	10	N	360.47	351.47	369.47	
IBM	QTL	ch1	Trait3 3	22	N	127.84	118.84	136.84	
IBM	QTL	ch1	Trait4 3	4	N	169.7	160.7	178.7	
IBM	QTL	ch1	Trait4 3	14	N	527.75	518.75	536.75	
IBM	QTL	ch1	Trait5 3	4	N	402.8	393.8	411.8	
IBM	QTL	ch1	Trait6 3	7	N	169.7	160.7	178.7	
IBM	QTL	ch1	Trait6 3	6	N	616.5	607.5	625.5	
IBM	QTL	ch1	Trait 7 3	11	N	323.59	314.59	332.59	
IBM	QTL	ch1	Trait 7 3	9	N	155.95	146.95	164.95	

*Fig. 4: BioMercatorV2 example qtl file*

## MetaQTL package format

This format gathers in a single file the genetic map and the corresponding QTLs and is recognized by BioMercator V3/V4. This file can be created with the MetaDB tool from the MetaQTL package (please, refer to MetaQTL package documentation:

<http://www.bioinformatics.org/mqtl/wiki/Main/Documentation> )

Several sample files are provided with the tutorial.

```
<?xml version="1.0" encoding="UTF-8"?>
<genome-map name="Barriere_2005">
  <property name="cross.size" value="242"/>
  <property name="cross.type" value="RI1"/>
  <property name="mapping.unit" value="cM"/>
  <property name="mapping.function" value="haldane"/>
  <linkage-group name="1">
    <locus name="Barriere_2005_SD_1" type="Q" position="188.0">
      <property name="qtl.cross.size" value="242"/>
      <property name="qtl.cross.name" value="F838xF286"/>
      <property name="qtl.effect.additive" value="2.11"/>
      <property name="qtl.ci.from" value="174.0"/>
      <property name="qtl.rsquare" value="0.124"/>
      <property name="qtl.ci.lod.decrease" value="1.0"/>
      <property name="qtl.trait.name" value="SD"/>
      <property name="qtl.ci.to" value="200.0"/>
      <property name="qtl.cross.type" value="RI1"/>
    </locus>
    <locus name="bnlg1832" type="M" position="160.8"></locus>
    <locus name="bnlg1803" type="M" position="105.6"></locus>
    <locus name="bnlg615" type="M" position="214.29999999999999"></locus>
    <locus name="bnlg1866" type="M" position="143.9"></locus>
    <locus name="bnlg1014" type="M" position="0.0"></locus>
    <locus name="bnlg1178" type="M" position="74.4"></locus>
    <locus name="bnlg1112" type="M" position="60.8"></locus>
    <locus name="bnlg149" type="M" position="32.6"></locus>
    <locus name="umc1366a" type="M" position="104.0"></locus>
  </linkage-group>
  <linkage-group name="10">
    <locus name="bnlg1028" type="M" position="91.9"></locus>
    <locus name="umc1318" type="M" position="0.0"></locus>
    <locus name="umc1648" type="M" position="68.3"></locus>
    <locus name="umc1337" type="M" position="24.199999999999996"></locus>
  </linkage-group>
</genome-map>
```

Fig. 5: MetaQTL example file

In addition to this file, MetaQTL may require an ontology file in XML format in order to group several traits into a meta-trait for QTLs meta-analyses. A sample file is provided with tutorial.

## MCQTL format

XML output files from the MCQTL software can be loaded within BioMercator (see <http://carlit.toulouse.inra.fr/MCQTL/>).

## Files loader wizard

Genetic maps and QTLs loading in BioMercator V4 are done through a wizard in the menu File → Open.

The first step is the choice of the target project (the first time, you'll have to create a new project).

The next step is browsing through files to load. Clicking on *Browse* will allow you to explore your hard drive for map files. It is possible to load several files at the same time whatever the format format.

## Workspace management

A project can be seen as a directory containing maps, text files, pictures and analysis outputs. A

workspace contains projects.

It is possible to use all data contained in a workspace; meaning two maps from two different projects can be used in the same analysis; during an analysis, the first step is to choose the project where outputs will be deposited; the second step is to choose the maps to use: here, all maps from every existing projects in the workspace can be used. (See GUI part for more informations)

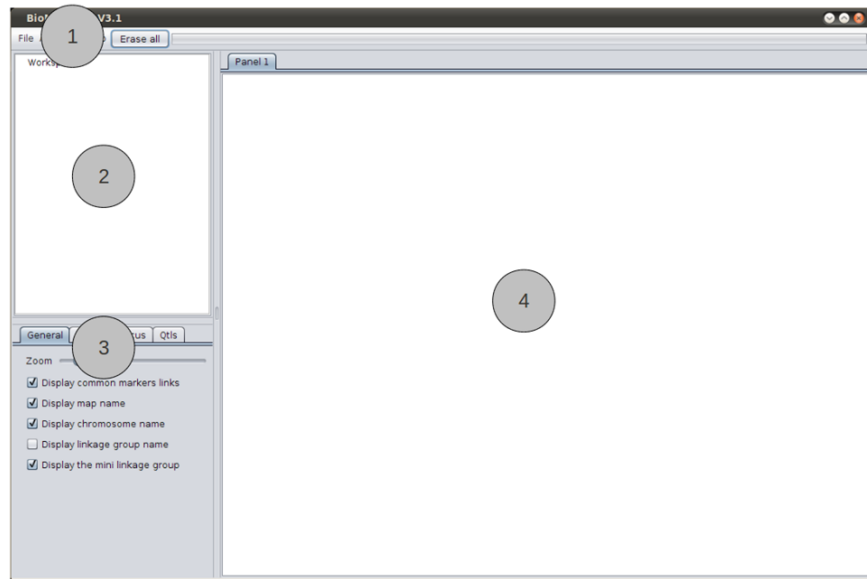
## **Data and analyses storage**

All loaded map files, compiled maps and output files (tabulated files and images) will be automatically saved and loaded at the next software launching. Data is stored on the hard drive in the BioMercatorV3 XML format in response to 3 events:

- After file loading: the data is stored in XML files in the directory "*GeneratedFiles/Your\_project*" where "*Your\_project*" is the target project you specified in the wizard.
- After performing Analysis : Analysis output is stored under the directory "*GeneratedFiles/Your\_project*" where "*Your\_project*" is the name of the target project you chose in the analysis wizard; the tabulation-delimited text files and map images can also be found in this place.
- Manual save: Data is stored under the chosen directory.

## Graphical User Interface.

1. Menu bar
2. Projects tree
3. Option tabs
4. Drawing panel



*Fig. 6: Software's overall*

### ***The menu bar***

- File
  - Genetic data loading: Launch the file loader wizard
  - Save as...: Save the map(s)
  - Export: Exports the whole display panel to an image (jpeg, or svg)
- Analyses
  - Statistics
    - Maps connectivity: gives information about maps connection
    - Common markers visualization: generate an image of aligned maps according to shared markers of the selected maps
      - Map compilations
        - Iterative compilation: map to map projection (including QTLs)
          - 1. Regression loci compilation: maps compilation in a single step (does not project QTLs)
          - 2. QTL projection: QTL projection from several maps to a target one
        - QTL meta-analyses
          - Meta-analysis (Gerber & Goffinet, 2000): meta-analysis from Gerber and Goffinet's algorithm; provides 5 models (with 1, 2, 3, 4 and  $n$  metaQtls) and Akaike's criterion to help selecting the best model

- Meta-analysis step 1/2 (Veyrieras et al, 2007): meta-analysis from Veyrieras's algorithm; creates several files for the second step; one of the file designate the best model according to several criterions.
- Meta-analysis step 2/2 (Veyrieras et al, 2007): display the meta-QTL map according to the chosen model
  - Help
  - About : Gives some information about the software
  - Language
    - French : Sets the software language to French
    - English : Sets the software language to English
  - Documentation : Launch this user guide

### ***The option panel***

- General
  - Zoom: a slider to adjust the drawing panel's zoom
  - Display common makers links: Enables/Disables dynamic map comparison
  - Display map name: Shows/hides the map name
  - Display chromosome name: Shows/hides the chromosome name
  - Display linkage group name: Shows/hides the linkage group name
  - Display the mini linkage group: Shows/hides the mini linkage group below the linkage group
- Locus
  - Display loci: Shows/hides loci
  - Display locus name: Shows/hides loci name
  - Display locus position: Shows/hides loci location
- QTL
  - Display QTLs: Hides/displays QTLs
  - QTL compact representation modes
    - Standard: standard QTLs display
    - Condensed: standard QTLs display without r2 bar
    - Lined: QTL are shown as a density curve (one by trait) along chromosome.
    - Overview: F. Chardon's overview qtls display (Chardon et al., 2004)
    - Lined (R2 weighted): QTL are shown as curves (one by trait) weighted with r2
- Expert
  - Display all loci names: Randomly shows/hides locus names when loci are too numerous
  - Automatic font size: Check it if you want the software to automatically choose a font size

## Maps display

To display a map or/and a chromosome from the project tree on the main panel, the drag and drop mouse function is the one to be used.

Dragging and dropping a map on the panel will erase any previous drawing and launch the map display: Click on the map (on the project tree) you want to show, move the mouse to the panel without releasing the mouse button; once over the panel, release the button.

Dragging and dropping a chromosome on the panel will append the chromosome's display to any previous drawing. This is obtained the same way as above.

To reorganize the chromosomes on the panel, you can also use the drag and drop: click on the selected chromosome, move the mouse where you want to insert the selected chromosome without releasing the mouse button; when done, release the button. *Note that you must drop your chromosome on another one; if not, nothing will happen.*

To suppress a chromosome from the display panel, just move the mouse on it, and right click on it to open the contextual menu, and selected 'Remove from view'; you can also do a middle click to directly suppress the chromosome.

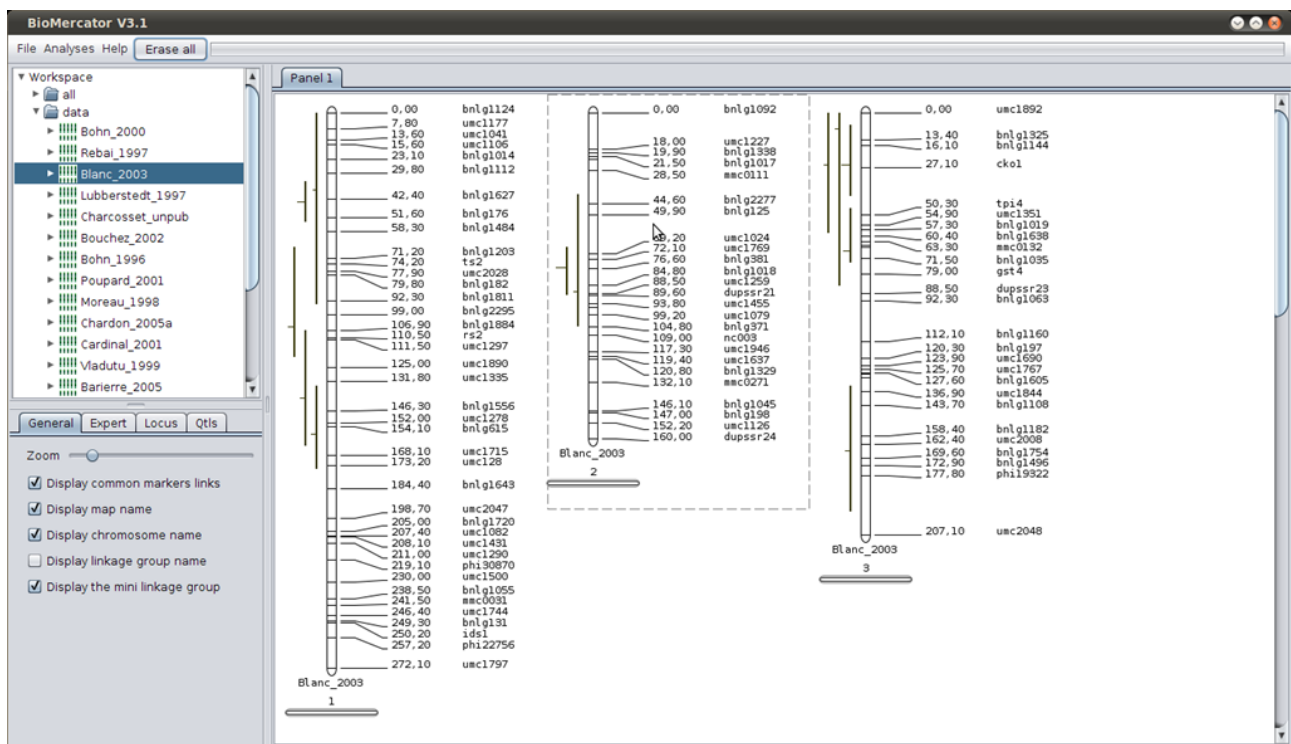


Fig. 7: Map display

## Zoom

### Whole drawing panel zoom

Inside the 'General' panel tab, a slider (called 'Map zoom') control the magnification level of the drawing panel.

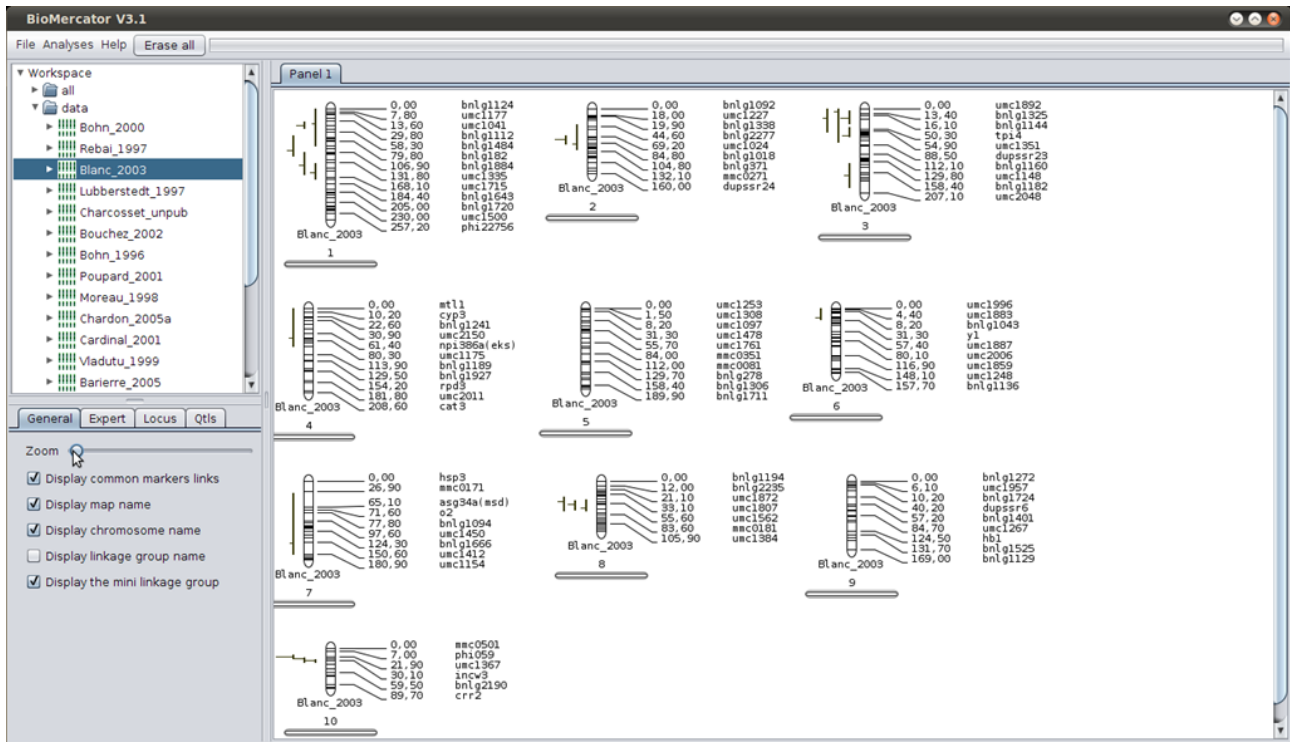


Fig. 8: Zoom – drawing panel magnification control slider

## Chromosome cascading zoom

This zoom is available only when a single linkage group or chromosome is displayed.

The zoom is activated when clicking somewhere over the displayed chromosome. A magnified chromosome appears alongside the first one. The enlarged area is depicted by a box lying on the unmagnified chromosome. It is possible to control the area which is magnified by clicking on the box edge and dragging the edge up or down. The box can also be used as a scroller. Clicking over the magnified chromosome activates an additional zoom level working like the previous one. It is therefore possible to visualize at the same time the whole chromosome and two level of magnification. It is possible to continue to activating additional zoom level.

For maps containing large number of loci, only a random subset of loci is displayed for clarity (more or all loci are displayed in magnified area). Forcing the display of all loci can be enabled from the Expert tab of the options management panel.

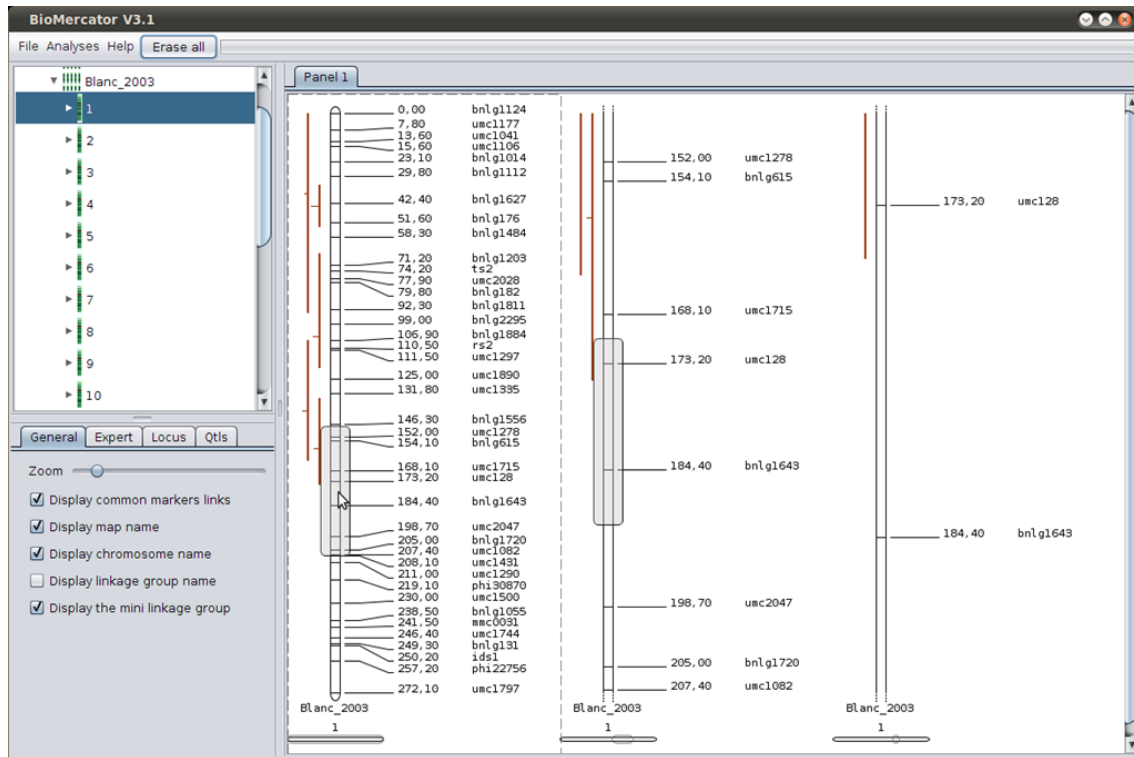


Fig. 9: Zoom - Cascading

## Individual chromosome zoom

This mode enlarges a chromosome itself. The area of the chromosome displayed is figured by a box on the mini chromosome thumbnail drawn at the bottom of each chromosome.

Two ways are available to zoom in on a chromosome:

- Right click over a chromosome to open a contextual menu, and then click on 'Zoom in' or 'Zoom out'
- Scrolling inside the chromosome with the mouse's wheel

When a chromosome is zoomed, you can scroll up and down by dragging your mouse inside. (Left click inside the chromosome, move up or down while keeping the button pressed)



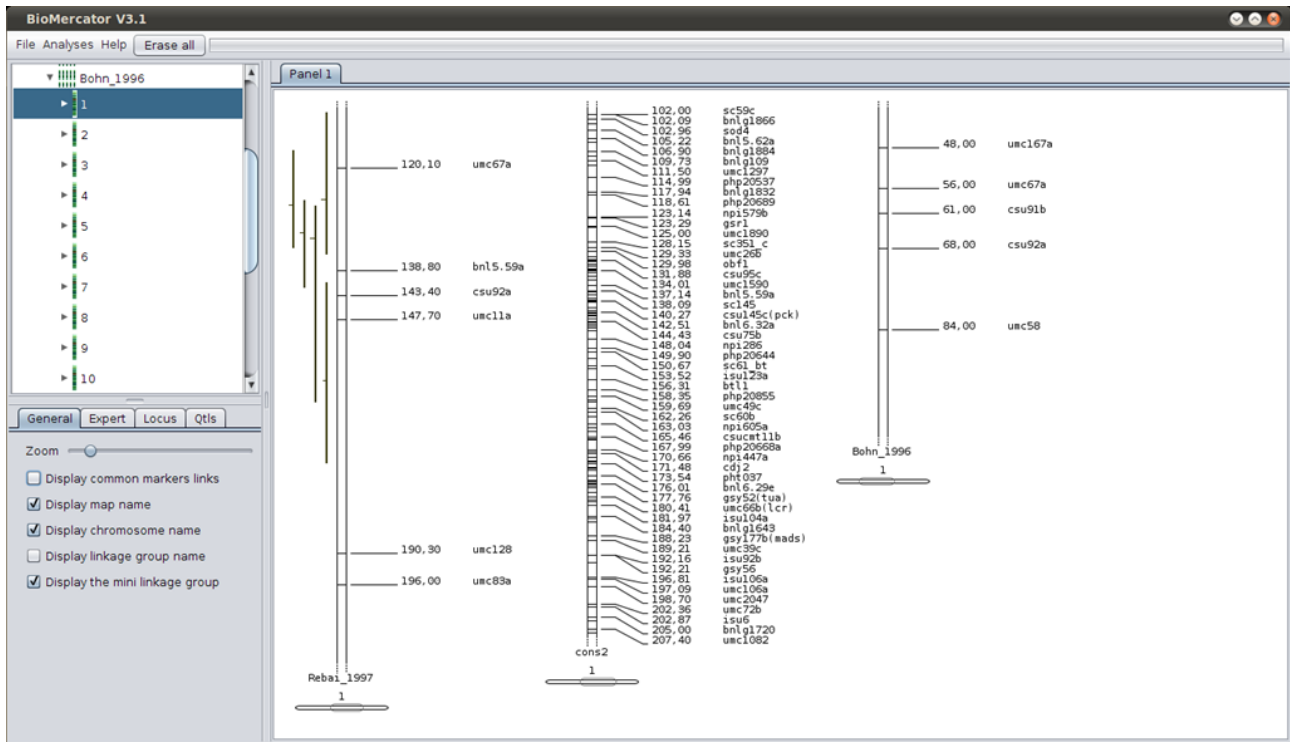


Fig. 10: Zoom - Individual

## Dynamic map comparison

BioMercator V3/V4 dynamically compares chromosomes from several maps (fig 11). This is done automatically, each time at least two chromosome are displayed in the drawing panel (make sure the option 'dynamic map comparison' is checked in the option management panel). Instantly, the links (red when inversion is detected, blue otherwise) will appear between the chromosomes; you can rearrange the chromosomes as described earlier to rearrange the comparison. Links are drawn between loci sharing the same name from adjacent chromosomes, whatever the chromosomes. Therefore it may be useful to rearrange chromosomes order to inspect all possible links.

## Chromosome/Linkage group Flipping

If a whole linkage group is inverted, right click on it and choose "Reverse" to flip the linkage group; the map will be kept reversed for all subsequent analyses.

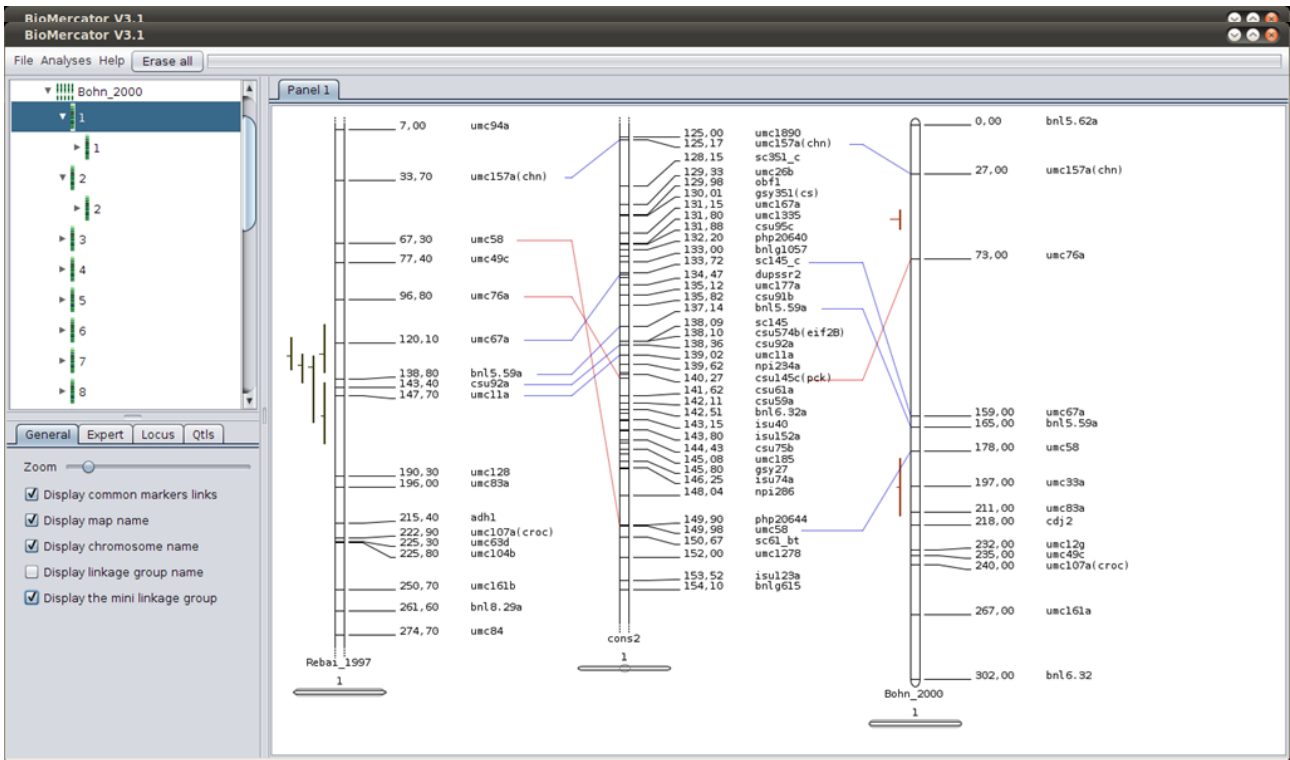


Fig. 11: Dynamic comparisons

**NB:** It is also possible to zoom when comparing maps.

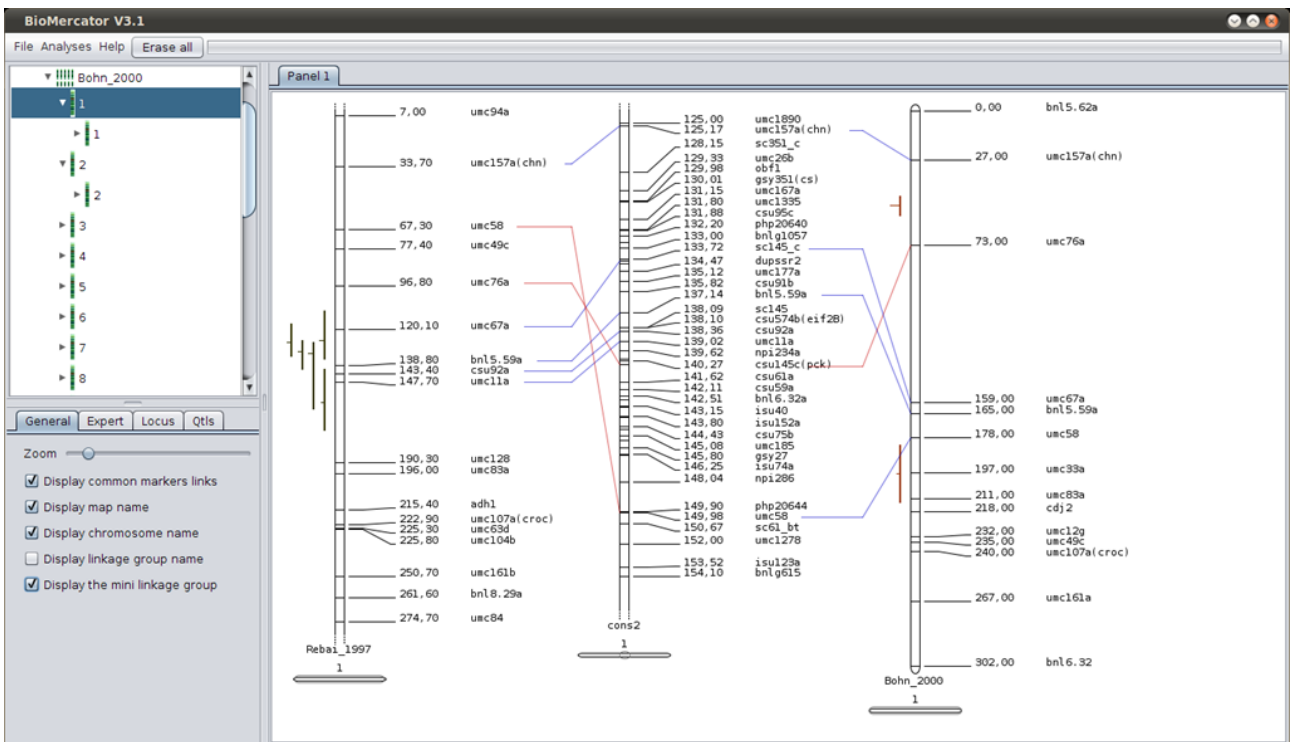


Fig. 12: Dynamic comparisons of zoomed maps

## Genetic maps compilation

### Overview

BioMercator V3/V4 includes BioMercator V2 and MetaQTL's algorithms. Each time an analysis is launched, a wizard appears and the first step is always the same: selecting the project target where the outputs will be stored; then, depending on the analysis to perform, the next steps can be maps choosing, parameters setting, output file name setting. The name (for maps or files) must be unique regarding the whole workspace; if it is not, it will be rejected a red box will surround the name input text field.

A progress window is displayed until the end of a computation. If an analysis do not terminate properly, the error message (if any) originating from the algorithm is displayed here. In certain circumstances, the metaQTL packages do not return any error message. Please refers to the metaQTL package documentation.

*NB: If one of the MetaQTL packages analyses fail, it often means the parameters where improperly set. In such a case, please refer to MetaQTL documentation.*

### BioMercator V2 tools

#### Iterative maps compilation

**Method:** The objective is to project QTLs, genes and other loci from a genetic map to another one, in order to pool iteratively all information onto one single map. This computation is only based on loci position data. For each pair of homologous chromosomes, common loci (sharing a same name) are listed. Inverted common loci can be automatically discarded from the analysis. Then, a specific distance ratio is computed for each interval between pairs of shared loci. A global ratio is computed for projecting loci located above the first interval of common markers and below the last interval of common markers. At last, QTLs and/or remaining loci position on the target map are computed by application of the appropriate distance ratio (homothetic projection). As no criteria are provided to assess the quality of projection process, users are encouraged to examine the resulting map. Individual maps can be iteratively projected on the compiled map. In this case, carefully deciding the order of maps in projection process is crucial. We suggest beginning projection iterations with maps showing highest quality in loci order, in order to limit error propagation.

#### Input:

- the source map (on the wizard's left side)
- the target map

**Output:** The consensus map where the projected loci are colored in blue.

**Information:** In order to project a map on one another, common loci without inversion must be present. In case of inversions, the analysis can still be launch only if the box 'Resolve inversions automatically' is checked (in this case, inverted loci are discarded before projection).

## QTL Meta-analyses (S. Gerber, B. Goffinet method)

**Method:** QTL meta-analysis is a useful tool to synthesize QTL information from independent experiments and to refine the chromosomal region involved in trait variation control. The QTL meta-analysis algorithm developed by Goffinet and Gerber (2000) can help to determine if  $N$  QTLs linked to a same trait or related ones, detected in independent experiments and located in a same region, are consistent with 1-, 2-, 3-, 4- or  $N$ - QTL models ( $N$ -QTLs model being the case where there are as much "real" QTLs as input QTLs). For each of those five models, the most likely QTL arrangement, assuming a Gaussian distribution, is determined by means of the maximum likelihood method (Fig. 6). Then, an Akaike-type statistical criterion indicates the best model among the five ones. For each model, consensus QTL positions are determined as the mean of QTL distribution maximizing the likelihood.

Meta-analysis computing is based on the position of each input QTL, and on the variance of this position, assessed through confidence interval values. Goffinet and Gerber jumped directly from a model with 4 QTLs to a model with  $N$  QTLs and did not extend their algorithm to intermediary models (5, 6, 7 ...). The reason for this is that Goffinet and Gerbers's study was based on a model chromosome with a length of 200 cM, and that 4 hypothetical non-overlapping QTLs of 50 cM-long can be found at most on this chromosome. The Akaike-type criteria values were determined using simulations processes. These values are optimized for use under precise conditions: - QTLs linked to a same trait and located on a same chromosome should all come from strictly independent experiments. Including several QTLs detected in a same experiment for a given trait and located on the same chromosome may introduce biases in the analysis. - The QTL dataset should include 10 to 40 QTLs - this set of QTLs should lie within a genome region no longer than 200 cM. Users working on larger chromosomes, or on a larger set of QTLs, should segment their working set to fit the above conditions, and repeat meta-analysis experiment on each subset.

**Inputs:** The chromosome on which to compute the meta-analysis. To select QTLs or traits to be included in the analysis click the wizard "QTLs choice" button.

### Output:

- A text file listing the 5 models (1, 2, 3, 4 and  $n$  metaQtl) and their corresponding Akaike criterion value. The model to consider is the one with the lowest Akaike
- A map containing the 4 models along with the selected Qtl and metaQtl

## MetaQTL package tools

*Informations hereafter are adapted from the original metaQTL's documentation. For more details, the reader is asked to refer himself to the MetaQtl's documentation.*

## Maps connectivity analysis: Infomap

**Method:** Before performing the construction of the consensus map, it is important to evaluate how the input maps are connected together. For each linkage group, InfoMap displays some descriptive statistics about the marker maps and for each pair of mapping experiments the program looks for common marker sequences. A common marker sequence is a set of at least two common markers for which the order in the two linkage groups is consistent. The marker order is said to be consistent even if the sequence is completely inverted between the two linkage groups.

```
\includegraphics[width=5cm]{images/BM_mqtl_infoMap_0.jpeg}
```

```
\caption{Inversion de marqueurs}
```

For example, in the above figure the number of shared loci sequences is 2. The first one (in red) involved 2 markers (umc116, umc5b), which order is reversed between the two chromosomes, and the second one (in blue) involved 3 markers (umc110, bn116.06, umc168).

**Input:**

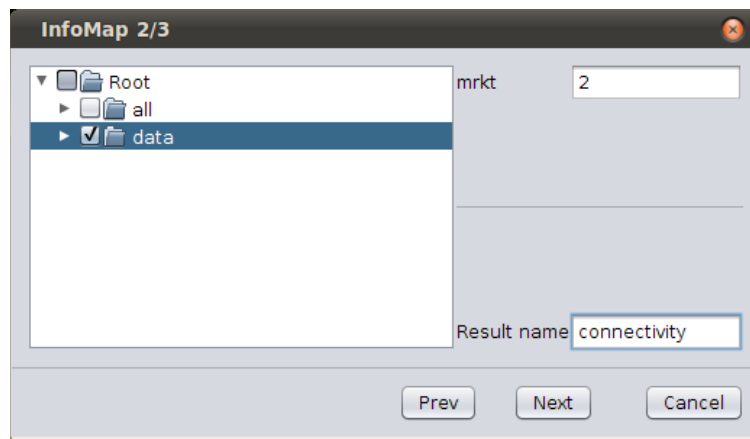
- Genetic maps : The maps to compare

**Parameters:**

- mrkt: The minimum number of common marker between maps (default is 2).

**Output:**

- \*\_cmp.txt: lists the markers connection chromosome by chromosome
- \*\_mrk.txt: lists the number of common markers chromosome by chromosome



*Fig. 15: InfoMap analysis*

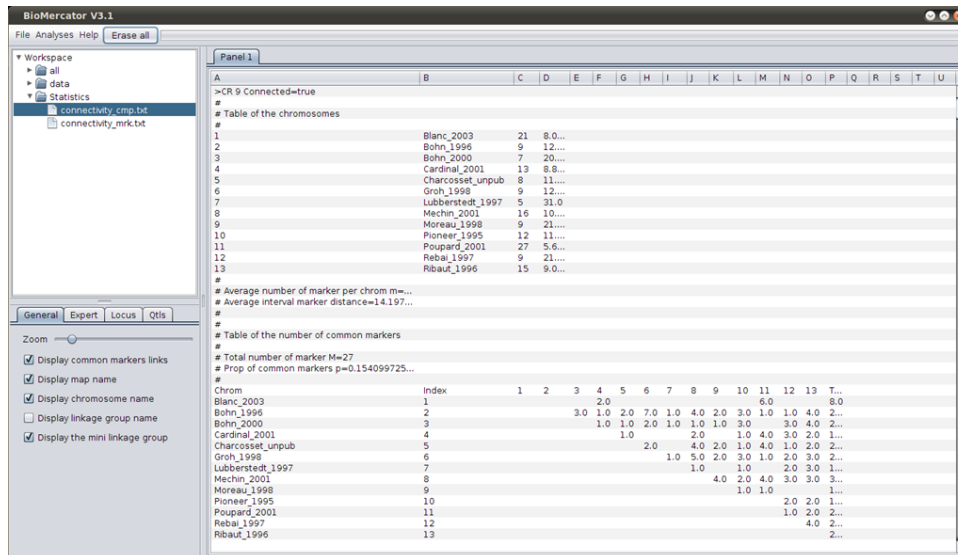


Fig. 16: InfoMap analysis - result

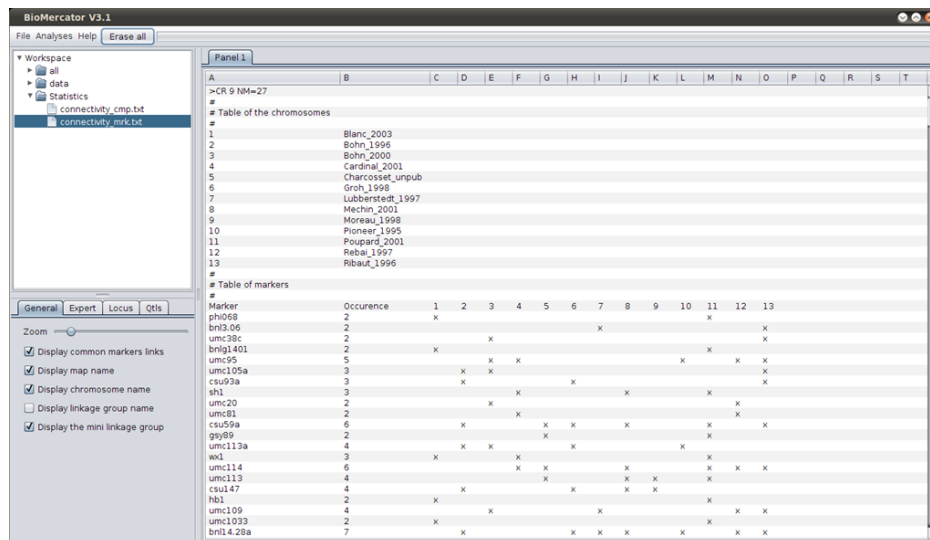


Fig. 17: InfoMap analysis - result3

## Common markers visualization: MMapView

**Method:** MMapView is designed to draw markers and QTL for one linkage group for several genetic maps. It helps to visualize common markers between maps but also to display marker interval distance heterogeneities.

### Input:

- The reference map
- The map(s) to compare to

**Output:** The image of the common markers connection for each of the chosen chromosome.

## Consensus map building: ConsMap

**Method:** The method implemented in MetaQTL is based on a Weighed Least Square (WLS) strategy. By contrast to iterative projection procedure, this method makes it possible to integrate all maps in a single step. It is also possible to fix a genetic map as reference.

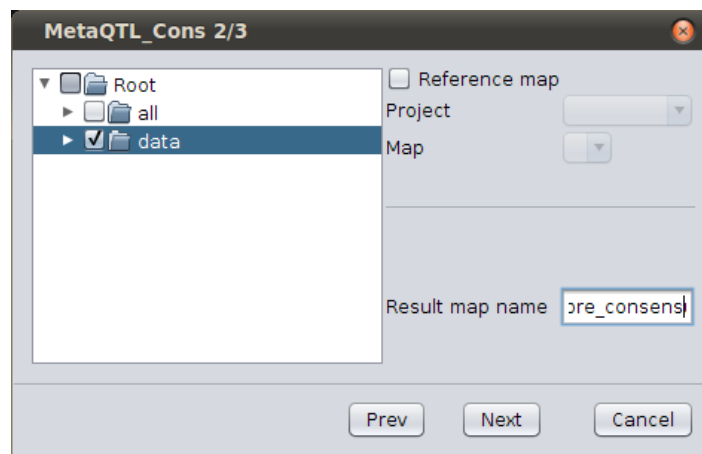
### Input:

- The maps to compile
- *A reference map fixing the order of the loci.*

### Output:

- The consensus map.

**Information:** For this analysis, the population size is necessary (to be set in the map txt file); if this value is missing, an event will occur to inform you and will stop the analysis.



*Fig. 18: ConsMap analysis*

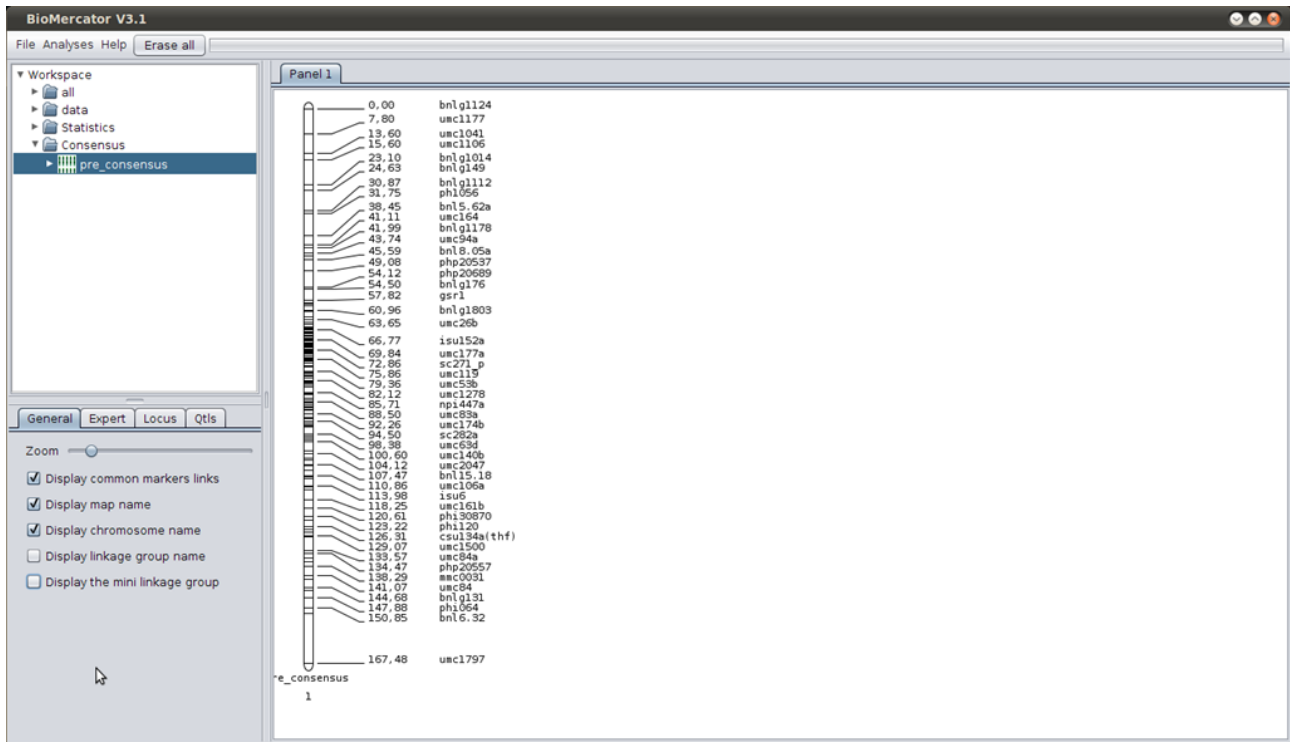


Fig. 19: ConsMap analysis - map result

## Qtls projection: QTLProj

**Method:** QTLProj implements a dynamic algorithm in order to find the optimal context to do the projection. An optimal context consists in a pair of common marker which flanks the QTL in the original map and for which the distance estimate is consistent between the maps. Two parameters control the behavior of the algorithm to find such a configuration: the minimal value of the ratio of the flanking marker interval distances and the minimal p-value obtained by testing the homogeneity of the flanking marker interval distances between the original and the reference map.

### Input:

- The source QTLs maps
- The target map

### Output:

- The target map with all projected QTLs



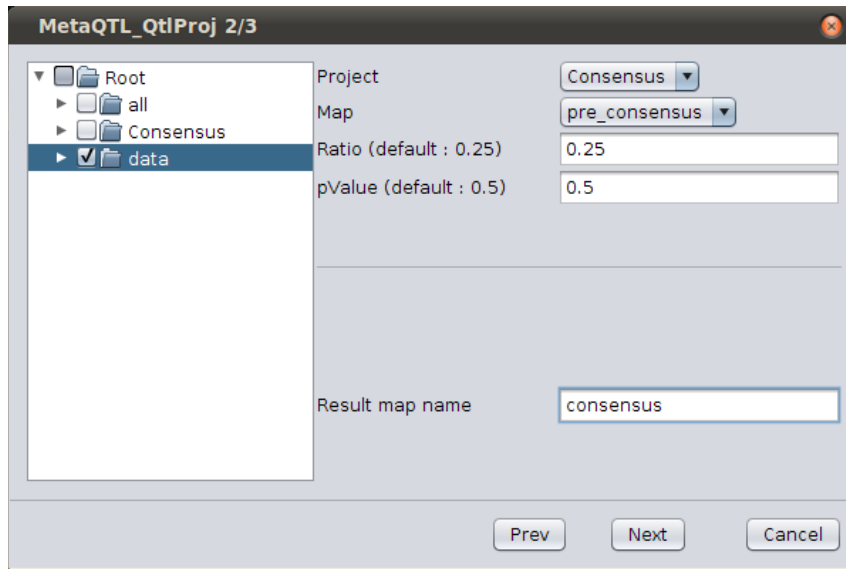


Fig. 20: QTLProj analysis

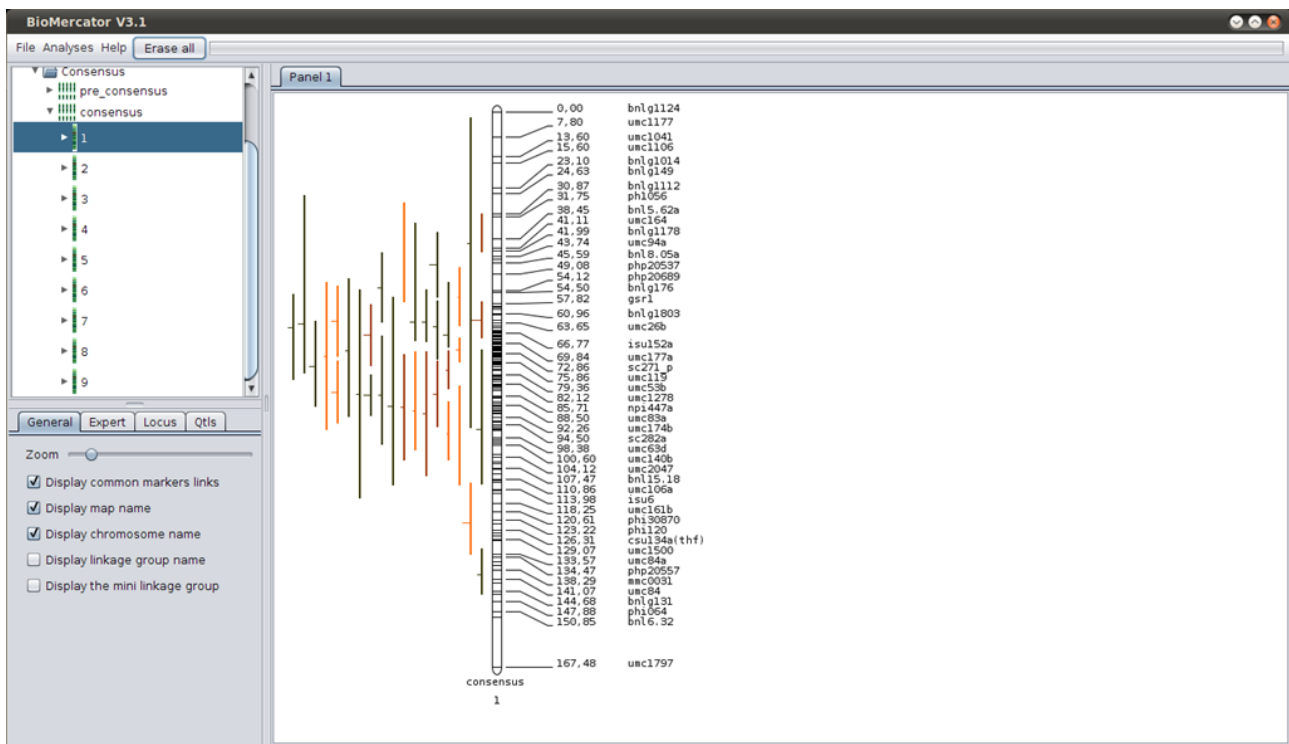


Fig. 21: QTLProj analysis - map result

## Meta-analysis step 1/2 (Veyrieras): QTLClust

**Method:** Here we want to address the following question: How many “real” QTL do the QTL detected in the different mapping experiments represent – one, two, three, four... or as many as the number detected throughout the studies? The meta-analysis of QTL can be viewed as a clustering procedure. To do so, MetaQTL implements two kinds of clustering algorithm. Whatever the procedure used to perform the clustering, the QTL locations are assumed to be normally distributed around their true locations with variances which can be derived from the reported CI or r-square values. This Gaussian and unbiased approximation comes from the classical asymptotic Gaussian

distribution of the maximum-likelihood estimation of the parameters.

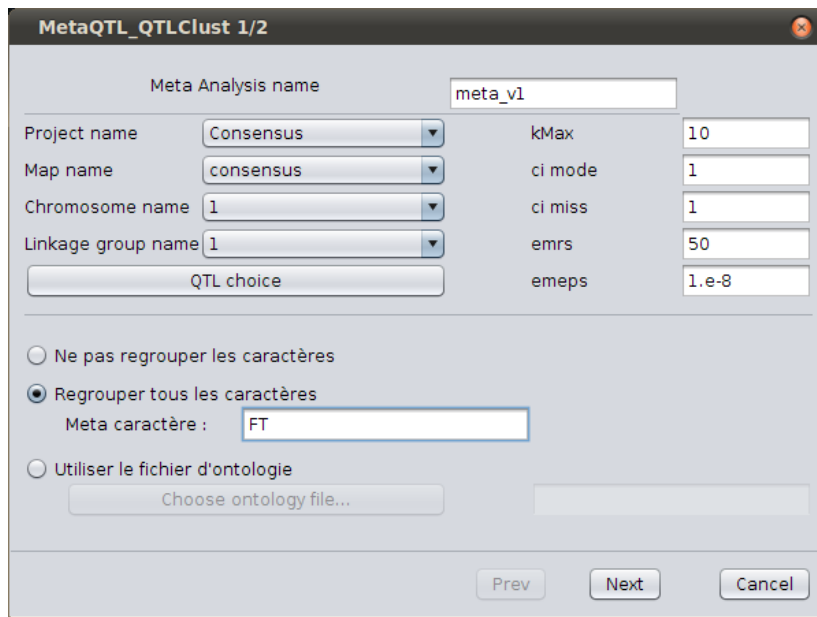
ClustQTL implements a clustering procedure based on a Gaussian mixture model which parameter estimates are obtained by applying a EM-algorithm.

**Input:**

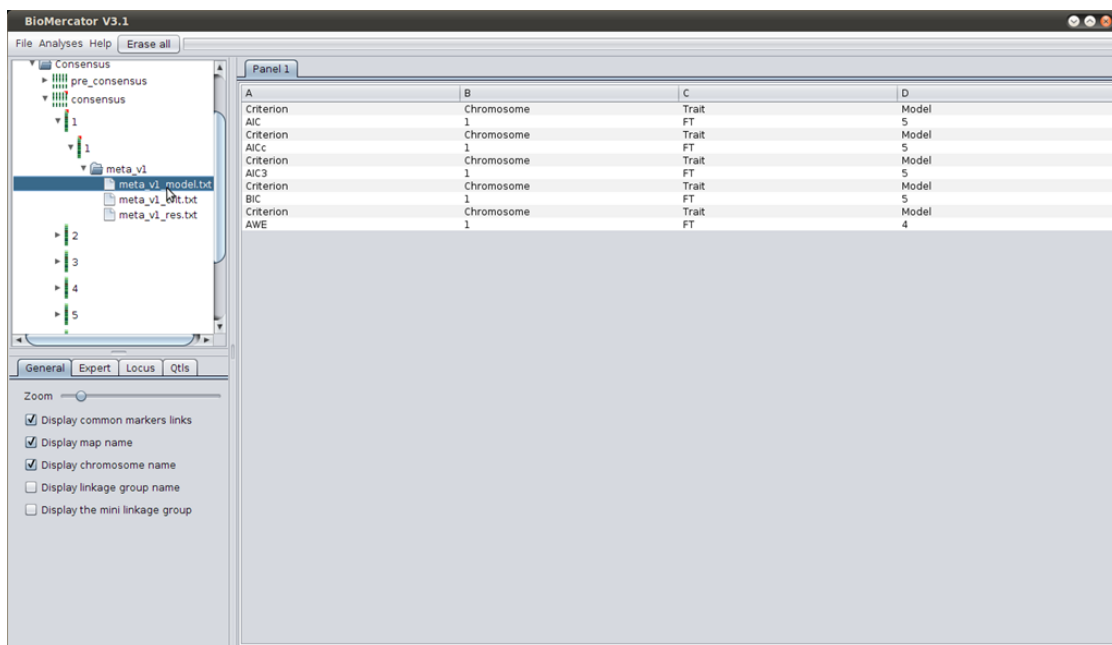
- The chromosome on which to perform the meta-analysis
- Some 'MetaQTL software' parameters such as :
  - kmax: The maximal number of clusters.
  - ci mode: The CI computation mode.
  - ci miss: The imputation mode for missing CI.
  - Emrs: The number of random starting points for the EM algorithm
  - emeps: the convergence threshold for the EM algorithm
- The choice to dismiss one or several Qtls from the analysis
- The choice of regrouping the traits into meta-traits:
  - no regrouping at all
  - Regrouping all traits
  - Regrouping according to an ontology file

**Output:** The output of ClustQTL is divided into 3 plain text files:

- \*\_res.txt: this files contains a summary of the clustering for each linkage group
- \*\_crit.txt: this file summarizes the values of the model choice criteria
- \*\_model.txt: This file gives the optimal number of QTL location according to the model choice criteria. The file is organized as a table with 4 columns. The first column indicates the name of the criterion, the second one the name of the chromosome, the third one the name of the trait and the last one the number of QTL



*Fig. 22: Meta analysis - step 1*



*Fig. 23: Meta analysis - First result*

## Meta-analysis step 2/2 (Veyrieras): MQTLView

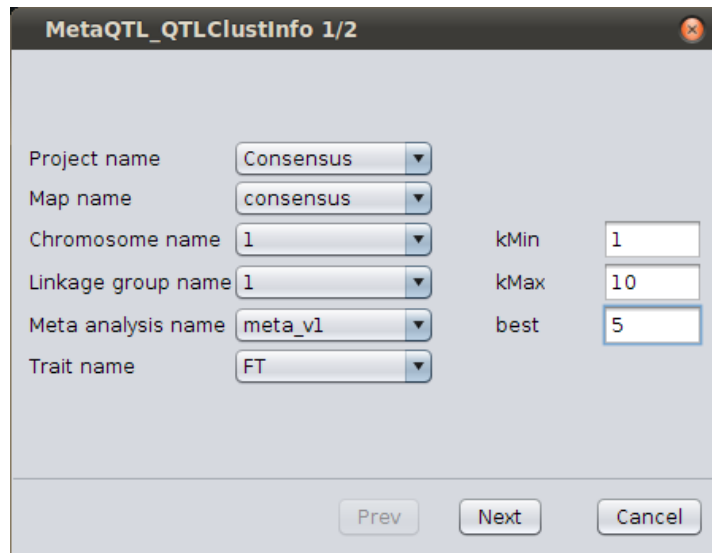
**Method:** MQTLView provides a way to represent the results of the QTL clustering depending on the algorithm used to perform it.

### Input:

- The chromosome
- The meta analysis created previously

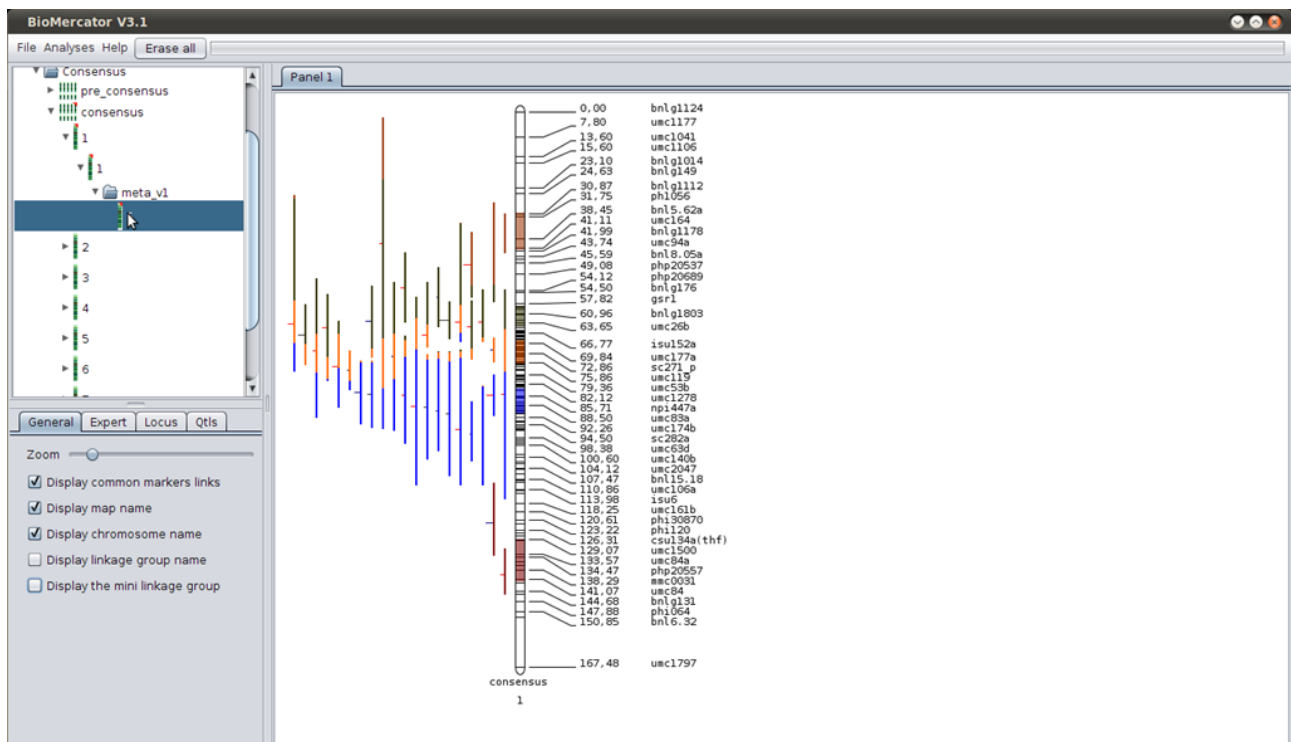
- The trait or meta-trait chosen
- The number of metaQTLs for the model (as suggested in the \*\_model.txt file)

**Output:**



*Fig. 24: Meta analysis - step 2*

- The model with the number of metaQTLs chosen as a new map in the project tree panel



*Fig. 25: Meta analysis - Visualisation*

## Genome structural and functional datasets

In BioMercator V4 only, genetic maps can be connected to genome annotation. This functionality is very useful to easily and quickly inventory genes included in QTL or metaQTL confidence intervals (CI). When functional annotation is available as GO terms it is possible to run some basic statistical analysis of GO term representation within CI. BioMercator V4 does not offer tools to align genetic and physical maps but need an anchoring file to connect these maps. Such information is usually generated by genome project consortia as a way to refine physical maps.

As genome (sequence and annotation) are splitted in various files subjected to more or less frequent update that may have consequence on features location it is crucial to ensure that the different annotation files corresponds to the proper genome assembly version. To help user dealing with genome assembly and annotation version, these data are managed as genome version in BioMercator.

A genome version includes 3 files

1. A structural annotation GFF file (location of all genes)
2. A functional annotation file (Only GO term are useful for BioMercator)
3. An anchoring file (Containing loci from genetic maps with corresponding sequence position)

Those files allow the software to link the genetic maps to the sequence, providing the user with gene display and analysis tools inside a genetic map region (a QTL confidence interval for instance)

### Loading a genome version

The genome version wizard allows loading any tabulated file (with any separator, such as comma, semi-colon, colon...) as long as necessary columns are present. By default, structural annotation files loading are set on GFF3 format.

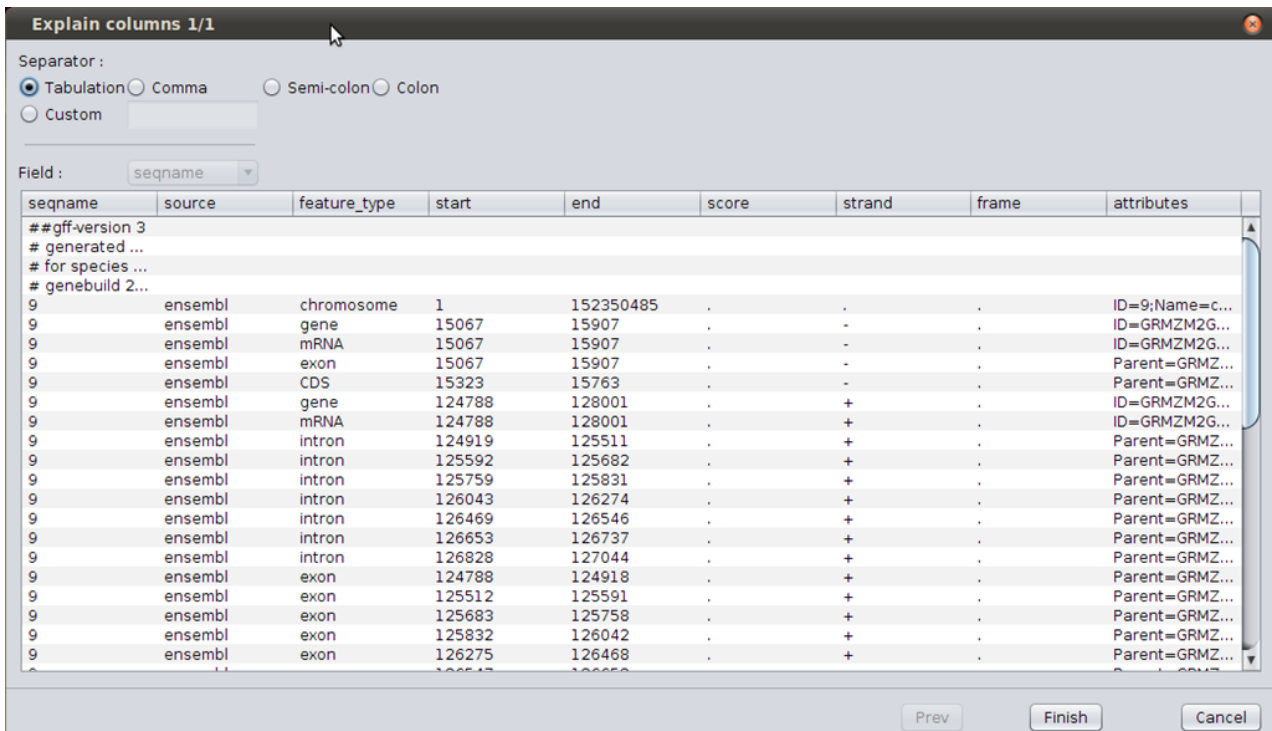
The following fields are **mandatory fields**:

- **Structural annotation**
  - The chromosome name: “seqname”
  - The genome features type: “feature\_type”
  - The starting position on the sequence (in pb): “start”
  - The ending position on the sequence (in pb): “end”
  - The attributes containing the element's and its parent's ID (if applicable): “attributes”
- **Functional annotation**
  - The transcript's name: “transcript”
  - The origin database: “db” *For the moment, BioMercator only uses terms from the Gene Ontology database (db: “GO”)*
  - *The corresponding term: “accession”*
- **Anchors**
  - The locus: “locus”
  - the chromosome name: “chromosome”
  - The locus starting position on the sequence (in pb): “start”

- The locus ending position on the sequence (in pb): “end”

If one or several new version of these file become available from genome project consortium, a new genome version has to be created. A new genome version may be based on one or more files included in a previous genome version (if only functional annotation is upgraded for instance).

**To remember:** A genome version can be used for several genetic maps; the more an anchor file contains anchor loci, the more accurate the display and analyses will be.



## Genetic map and genome annotation display

Once the genome version loaded, the next step is the display of a genetic map alongside the genome physical.

First, a chromosome (the first chromosome of a consensus map for instance) must be displayed in the graphical panel (by dragging it from the explorer tree); then, a genome version must be chosen in the “Genome Version” tab; a progression bar is displayed until the genomic data is loaded.

The physical map and the structural and functional annotation are then displayed (Fig 27):

1. A vertical line corresponding to the physical map at the right of the genetic chromosome.
2. Two 'scrollers' positioned on the genetic and the physical chromosomes.
3. A genome window (right side) containing several horizontal lines, featuring genome structural annotation (genes). The genome area displayed is the one corresponding to the scroller of the genetic and physical chromosome representations (left side). The scroller can be resized by dragging up or down their edges. When one of the scroller is resized, the other one is automatically adjusted.
4. Blue bars, representing anchor loci
5. Blue and red lines, representing links between maps (the red corresponding to the inversions)

The scrollers are used to explore the chromosome; they are both movable and re-sizable. Clicking

on a QTL will directly resize the scroller around its confidence interval and the corresponding structural annotation is displayed.

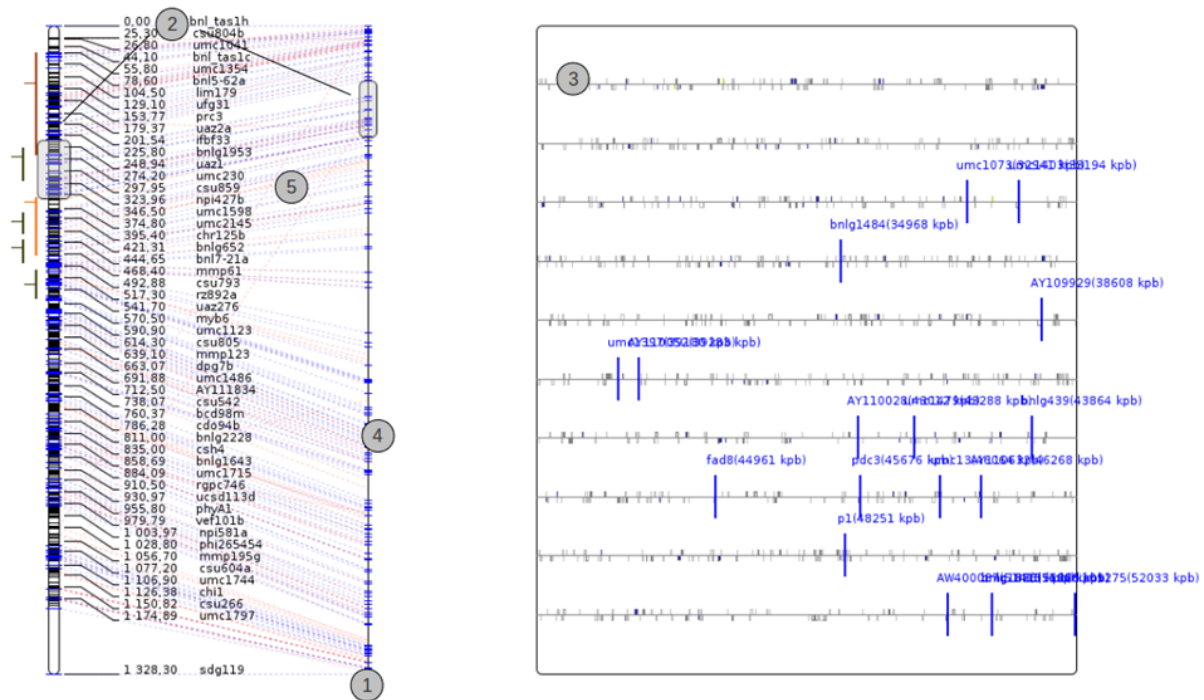


Fig. 27: Genome version overall

## Functional annotation display

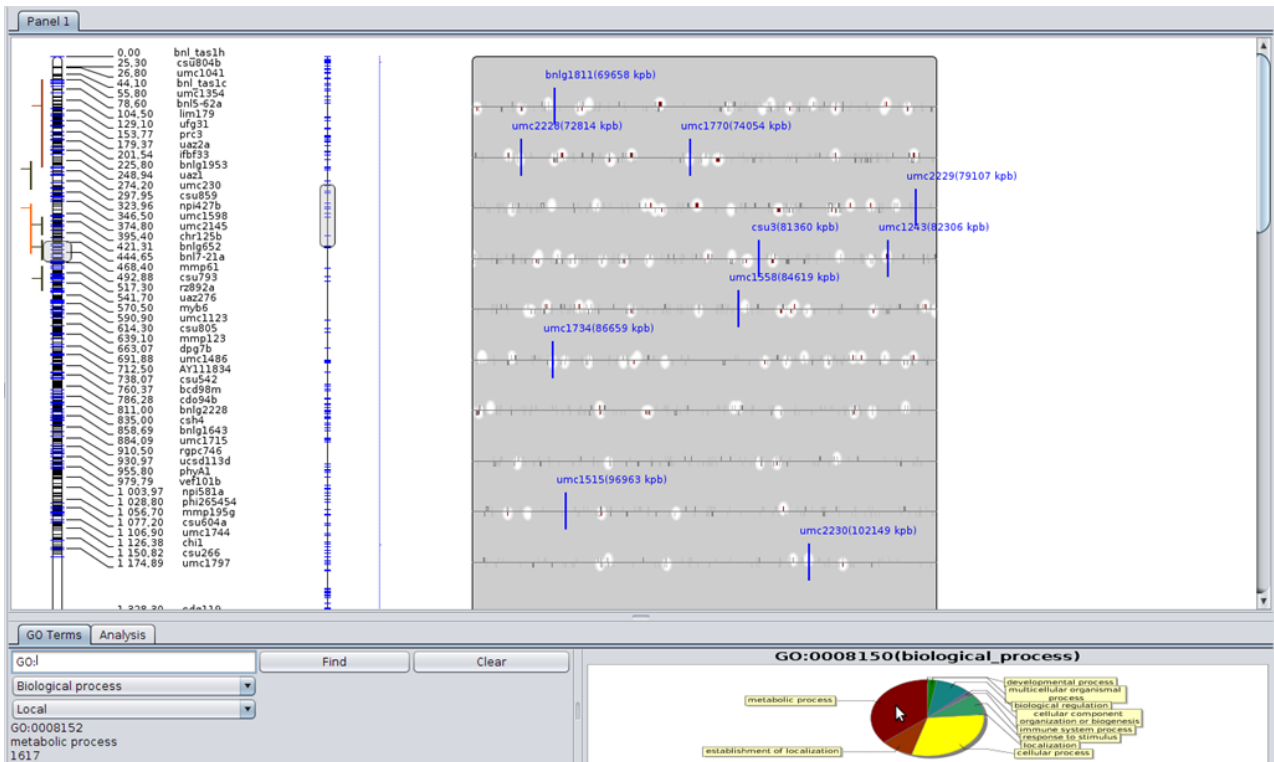
### The Gene Ontology

The Gene Ontology (<http://www.geneontology.org/>) is a graph-structured set of terms (corresponding to a biological process or function). The functional annotation files usually provided by genome project consortia associate GO term to genes.

As explained above, the right window is a blow up of the sequence framed region (by the scroller); in this window are represented genes found in the structural annotation file where position around the line (above or below) depends on their 'strand' field, and color corresponds to their GO term(s).

At the software's bottom, in the frame named "GO terms", information about the GO terms can be found: a pie chart represents the terms proportion in the blow up genome window; every time the scroller is moved or resized, the pie chart is refreshed and shows the highest levels in the GO hierarchy; the pie chart colors are the ones used to display the genes.

This pie chart is interactive: clicking on a section (corresponding to a GO term) highlights genes containing the GO terms selected; another click on the same section allows to get deeper in the Gene Ontology hierarchy; this protocol is possible as long as the hierarchy level contains at least one gene (in the window). A GO term search is also possible through the appropriate field.



## GO terms representation analysis

At the software's bottom, the second tab named “Analysis” contains several fields and a button; this button launches a GO term over-representation analysis on a given interval.

As seen earlier, clicking on a QTL positions the annotations window around its confidence interval; the fields will automatically be filled with these values, and it will be possible to launch the analysis.

The result is displayed in a multi-columns list:

- pValue: The analysis is done using a hypergeometric law. The pValue represents the probability to obtain this distribution by chance
- GO term density: The ratio of the number of genes associated to the GO term in the interval and relatively to the whole chromosome
- Go term ID: The GO term id
- Go term name: the Go term name
- Ontology: The ontology to which the term belongs



