Workshop

Comparative Mapping in Salicaceae

Methods

Véronique Jorge
INRA Orléans, P1d

8-10 March 2010

Comparative Mapping - *intro* -

Definiton

Alignment of chromosomes of related species based on genetic mapping of common DNA markers

Goal

Identification of conserved/specific genomic regions

Question

Accumulation of QTL data based on heterogeneous mapping experiments (population types, sample sizes, evaluations methods, environments ...)

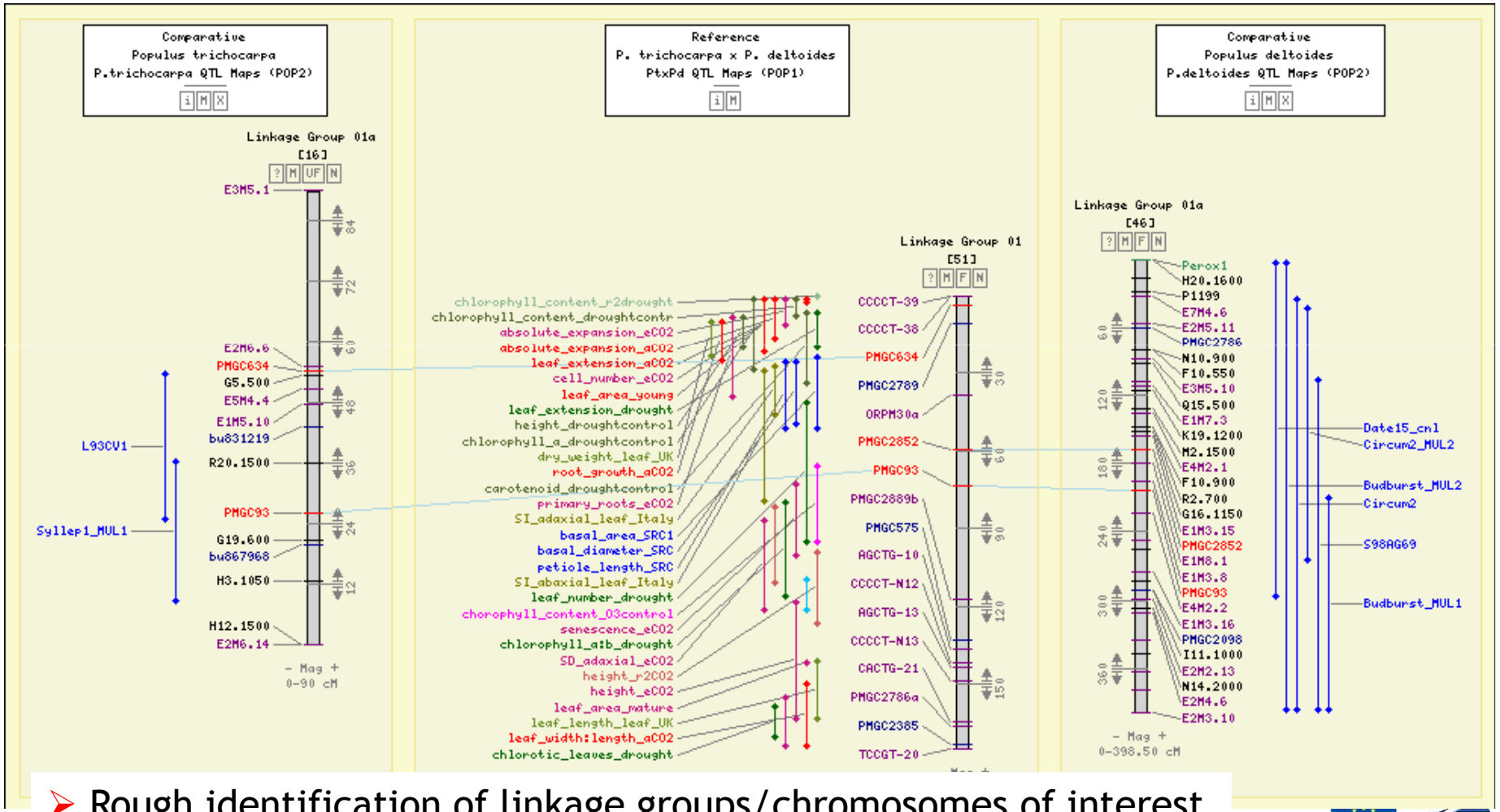
=>How many "real" QTLs are detected ?

Strategies

- ✓ Simple graphical comparison
- ✓ Map construction and QTL analysis based on mutliple mapping experiment (on raw data)
- ✓ Consensus and meta-analysis (on position data)

Comparative Mapping - *Strategies* -

Simple graphical comparison: matching marker names (Cmap)



➢ Rough identification of linkage groups/chromosomes of interest

Populus data from Salicaceae maps data base: http://www.evoltree.soton.ac.uk/evoltree/cmap/

Comparative Mapping - *Strategies* -

Map construction and QTL analysis based on mutliple mapping experiment (on raw data)

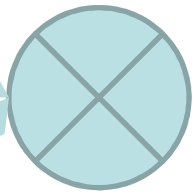
Marker and phenotypic data on pedigree 1

```
010001101110101
110111101100110
110110011010001
110111101100110
110111101100110
25 35 21 18 20
25 26 30 21 20
```

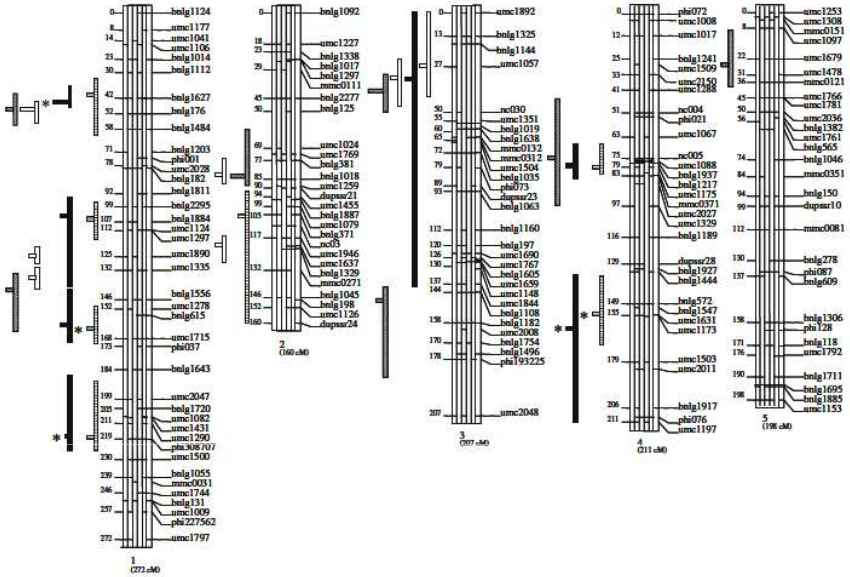
Marker and phenotypic data on pedigree 2

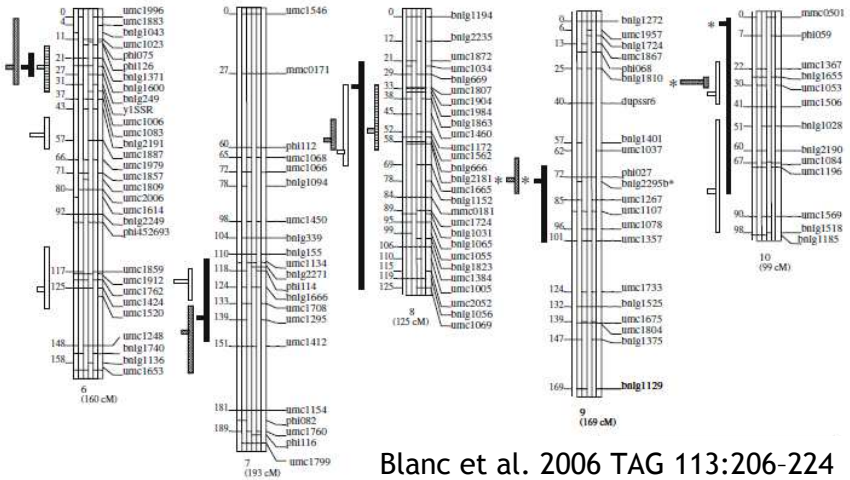
```
010001101110101
110111101100110
110110011010001
110111101100110
110111101100110
25 35 21 18 20
25 26 30 21 20
```

Analysis



Map: JoinMap, Carthagene
QTL: MCQTL, …





- Raw data not easily available
- Connected pedigrees often needed

Blanc et al. 2006 TAG 113:206-224

Comparative Mapping - *Strategies* -

Consensus map and QTL meta-analysis based on position data

- From elaborated data (maps, position and QTL confidence intervals)
- Meta-analysis: combine data from different sources in a single study
- Widely used in medical and social science

Advantages: elaborated data retrieved from published data

Steps for QTL meta-analysis

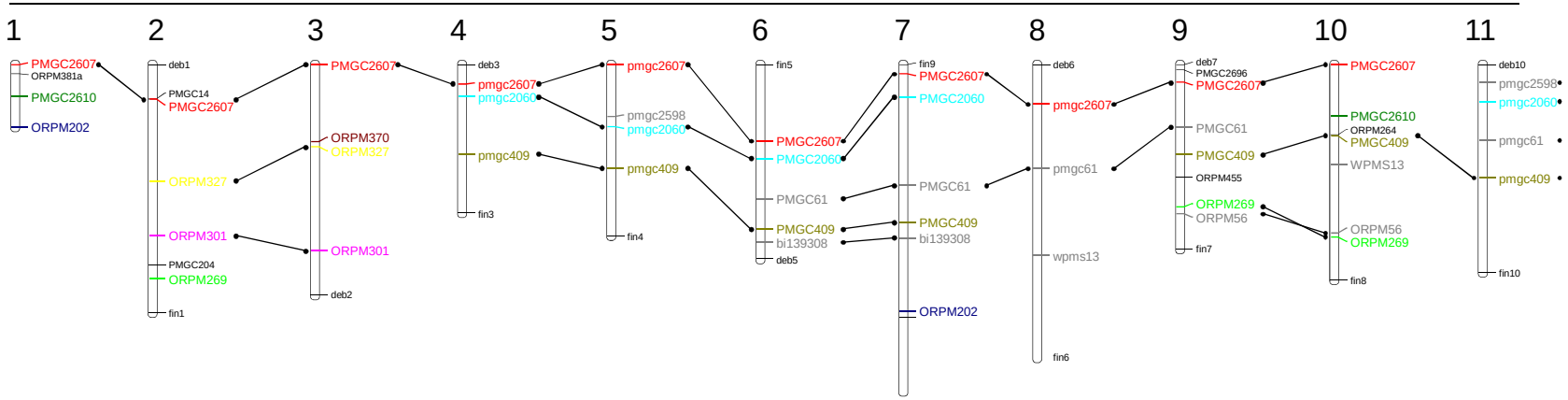
- Create a consensus map
- Project QTL on the consensus map
- QTL meta-analysis

Meta-analysis - *Methods* -

Create a consensus map

You need common markers

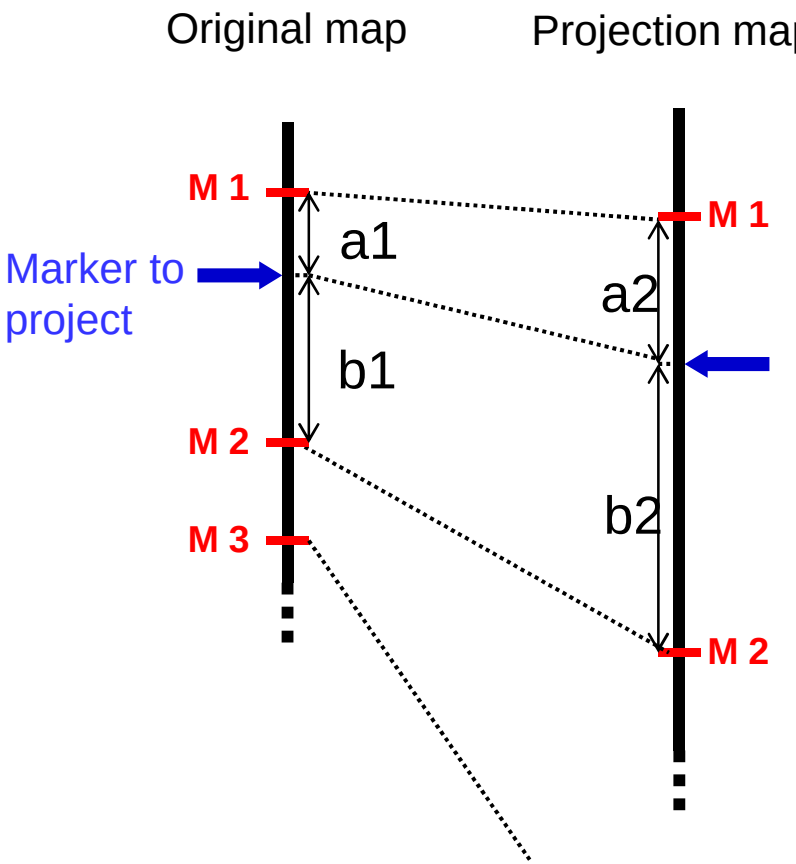
Example: case of a LG well covered by SSR (LG_VIII)



Meta-analysis - *Methods* -

Biomercator

Projection approach: homothetic projection



$$R_{M1\text{-}M2} = \frac{a2 + b2}{a1 + b1}$$

$$a2 = R_{M1\text{-}M2} \times a1$$

For markers outside interval of common markers, a global ratio is computed

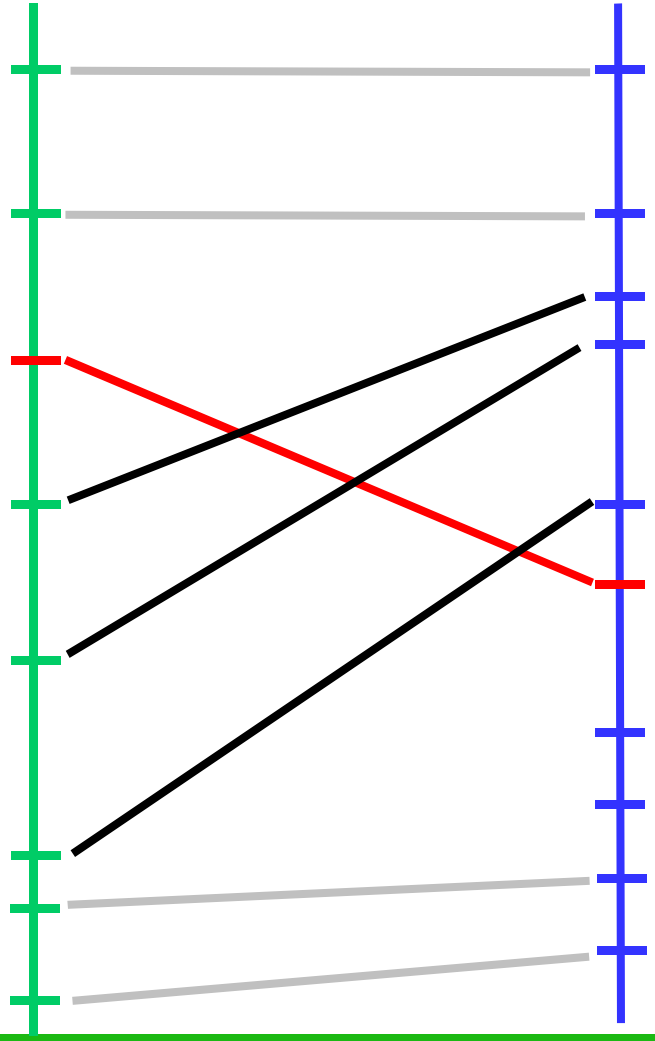
Meta-analysis - *Methods* -

Biomercator **Projection approach: homothetic projection**

Resolution of inversions between markers shared between maps:

Majority criteria

-> red marker is discarded



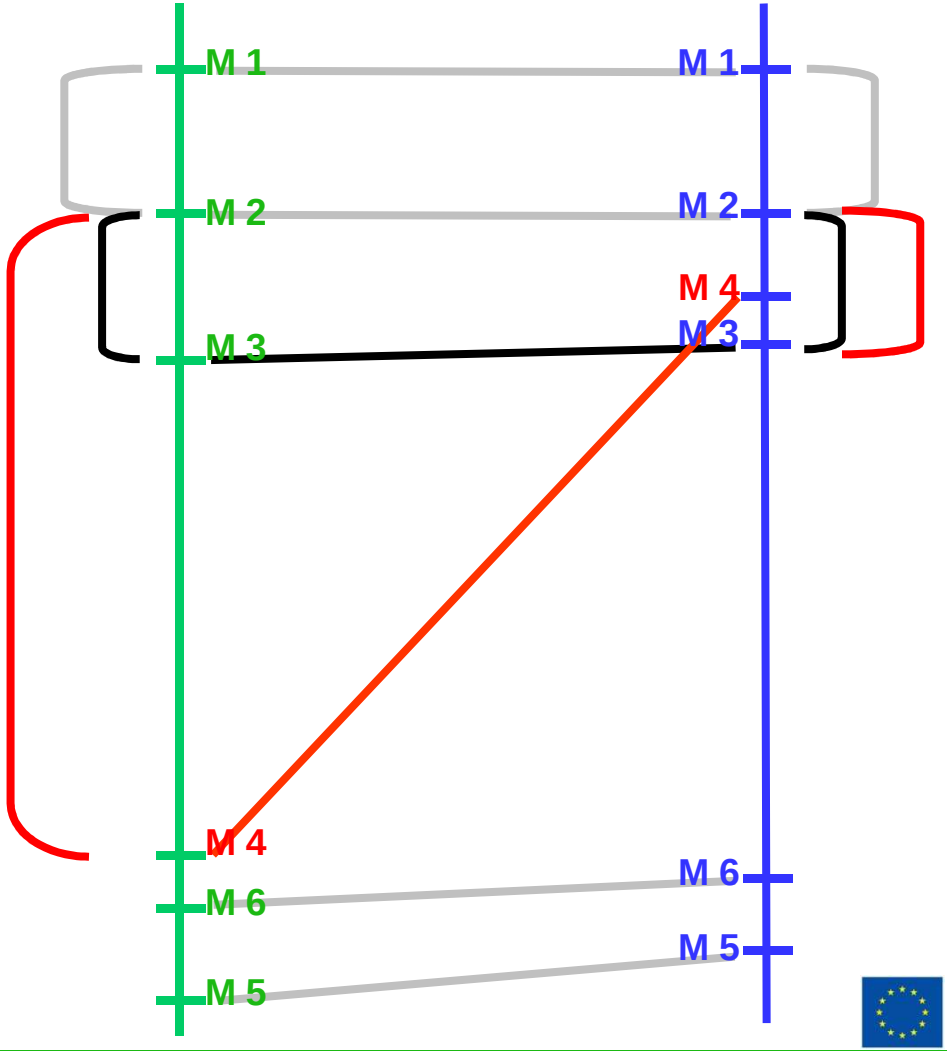
Meta-analysis - *Methods* -

Biomercator **Projection approach: homothetic projection**

Resolution of inversions between two common markers:

The homogeneity criteria for the proportion coefficient R.

-> marker M4 is discarded



Meta-analysis - *Methods* -

Biomercator **Projection approach: homothetic projection**

- Well adapted when one of the genetic map is a reference
-> all maps can be projected on the reference map

- In the other cases, iterative projections could be possible, but not optimal because the all projection orders are not equivalent.

Map1 -> Map2 = Projected_Map1
Map3 -> Projected_Map1= consensus1

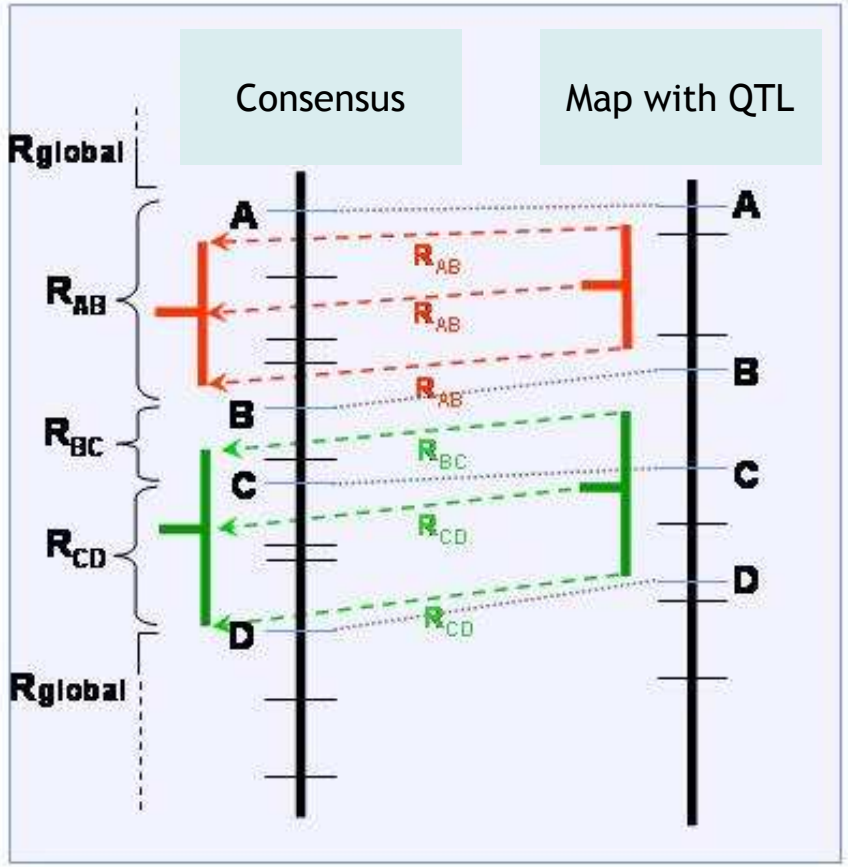
≠

Map1 -> Map3 = Projected_Map2
Map2 -> Projected_Map2= consensus2

Meta-analysis - *Methods* -

Biomercator QTL projection on the consensus map

The **3 positions** describing the QTL (start/stop confidence intervals, position of the LOD peak) are projected on the consensus map using the **R** factor of the nearest marker interval



Meta-analysis - *Methods* -

Biomercator

QTL meta-analysis

Goffinet & Gerber (2004)

Assumptions:

- Independence between QTL experiments

*"These different experiments may represent several crosses, **or** several sires, **or** different traits, **or** different locations for the same trait, **or** different environmental conditions and experimental design"*

"The quality of the results in these cases [non-independent observations] *shows that the method is robust ..."*

- QTL data set = 10 to 40
- Size of the region analysed < 200 cM

Meta-analysis - *Methods* -

Biomercator

QTL meta-analysis

Goffinet & Gerber (2004)

Analysis principle

➢ Models tested:

number of "real" positions ***k*** = 1, 2, 3, 4 or n.

(n = number of meta-QTL = number of initial QTLs).

k = 1 , all QTLs (n) located at a single position

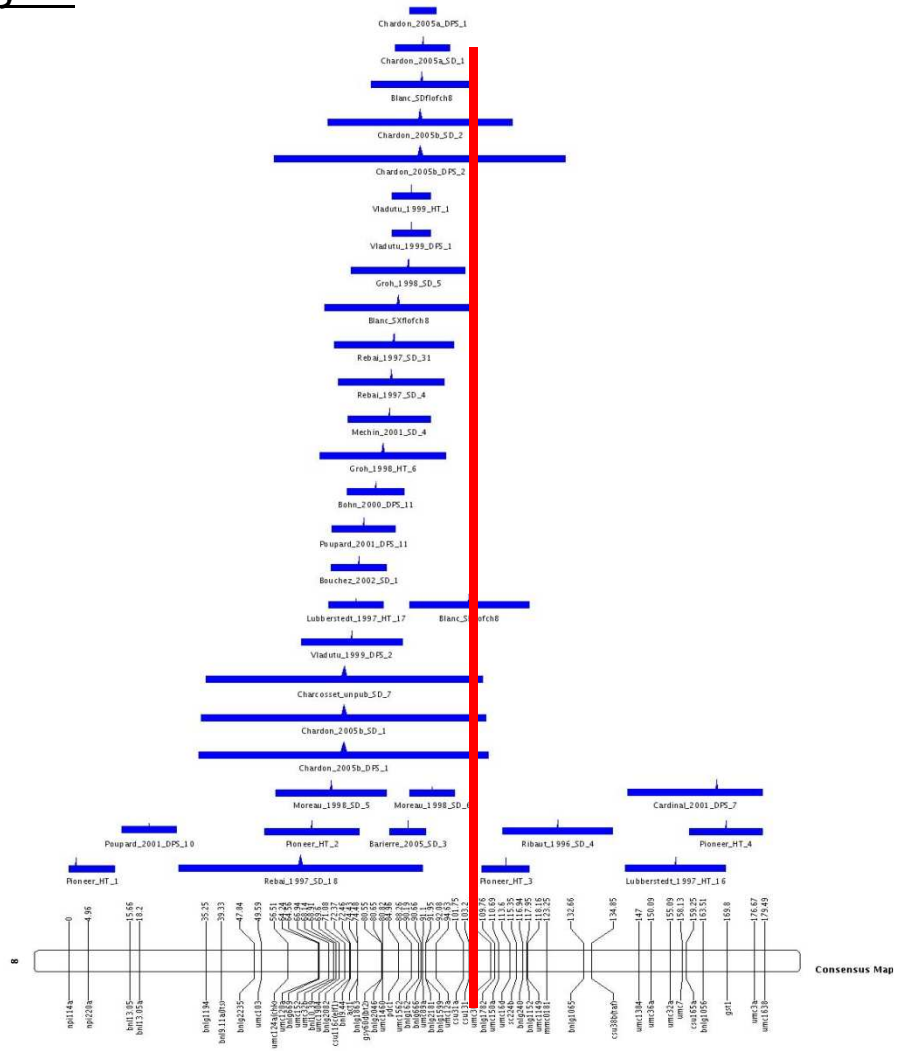
k = 2, QTLs fall into 2 groups

...

➢ For *k* ≥ 2, all possible distributions of the n QTL in k groups are considered. **The best distribution** is chosen based on a maximum likelihood method (using estimated variances of initial QTL position).

➢ To choose **the best model** (how many groups), a modified Akaike Information Criterion (AIC*) is used.

The model giving the minimal value of the AIC* is the best model.



Meta-analysis - *Methods* -

Biomercator

Data input format

Map

```
mapname=POP2tricho
poptype=BC1
chromosome LGXIX

  872  GCPM107     16.1 cM
  434  E6M51.1.     11.9 cM
  308  PMGC667      6.7 cM
   25  E4M2.7.T      9.4 cM
  822  rClasTai      4.1 cM
  433  E1M4.x.T      4.7 cM
   30  E5M5.7.T      4.2 cM
  871  GCPM104      5.0 cM
  879  rE21M18.     10.3 cM
  873  ORPM263     11.7 cM
  870  GCPM79   ----------
```

- Format analogous to Mapmaker output
- Distance between markers

QTL

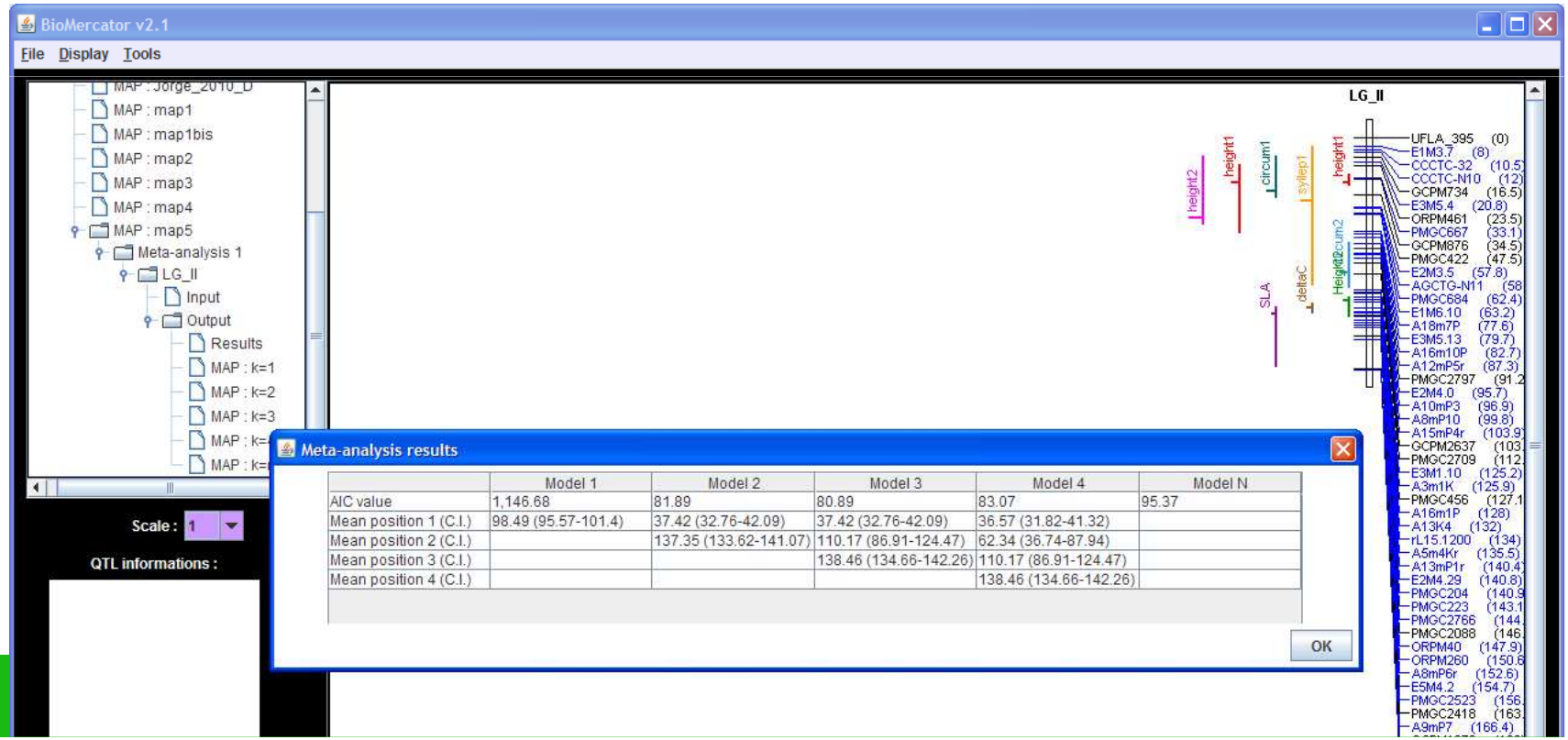
mapName	name	chromosome	trait	lodscore	r2	SIM	position	from	to
LGXIX	Rust3	LGXIX	MAX2_F	30.7	34.7	N	46.2	42.76	46.17
LGXIX	Rust4	LGXIX	MAX2_I	13.0	17.4	N	50.4	39.9	52.42
LGXIX	Rust5	LGXIX	MAX2_UK	15.0	24.2	N	41.9	39.9	45.88

- 1 map file and 1 QTL file for each map experiment
- Data are converted in *.xml* files

Meta-analysis - *Methods* -

Biomercator How the program works

- Data included in the example: *see « WS_Comp_Mapping_Salicaceae_DataSynthesis.ppt »*
- **Outputs**
  - Map projection => XXXReport.out: a report of all LG comparisons
    => XXXCommonMarkers.txt: list of common marker
    => .txt and .xml files for all projection maps created
    NB: note the order of map projection (not reported in the output files!)
  - Meta-analysis =>



	Model 1	Model 2	Model 3	Model 4	Model N
AIC value	1,146.68	81.89	80.89	83.07	95.37
Mean position 1 (C.I.)	98.49 (95.57-101.4)	37.42 (32.76-42.09)	37.42 (32.76-42.09)	36.57 (31.82-41.32)	
Mean position 2 (C.I.)		137.35 (133.62-141.07)	110.17 (86.91-124.47)	62.34 (36.74-87.94)	
Mean position 3 (C.I.)			138.46 (134.66-142.26)	110.17 (86.91-124.47)	
Mean position 4 (C.I.)				138.46 (134.66-142.26)	

Meta-analysis - *Methods* -

Steps for QTL meta-analysis

- Create a consensus map
- Project QTL on the consensus map
- QTL meta-analysis

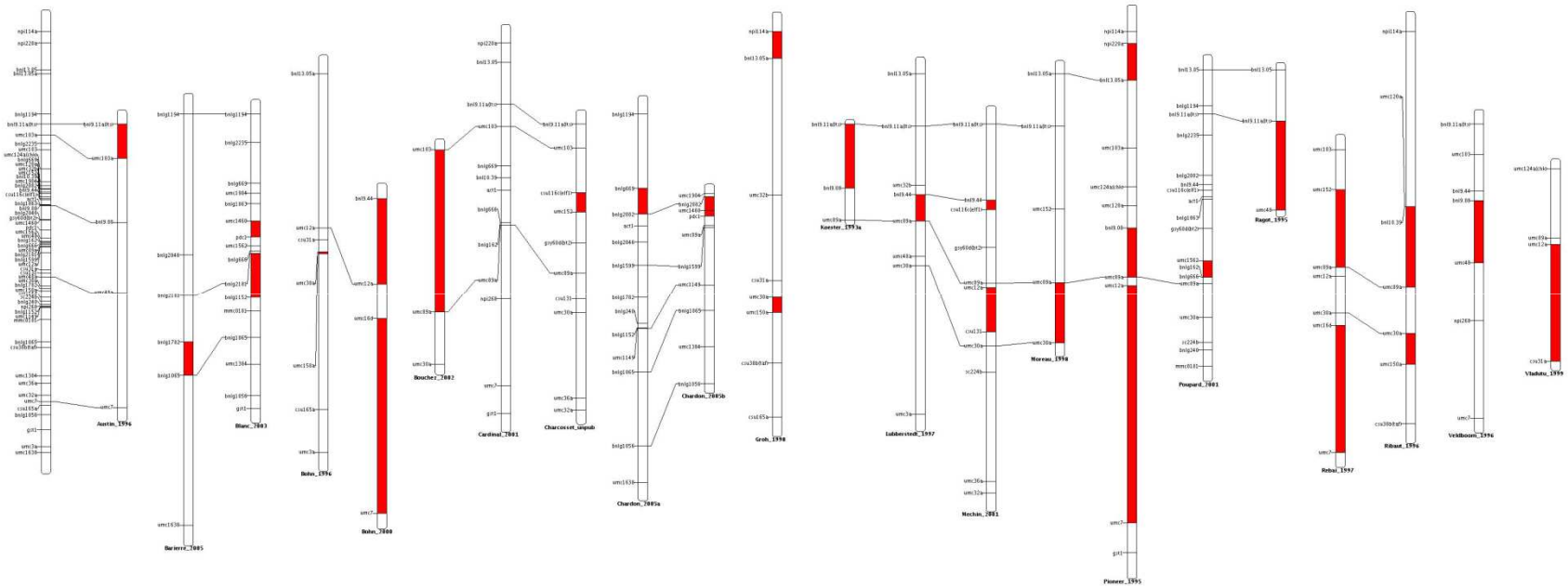
Meta-analysis - *Methods* -

MetaQTL

Consensus map

Veyrieras et al. 2007

Weighted Least Square method



Advantages:

- One step construction (vs iterative projection)
- Adjustment statistics for the consensus map (test for homogeneity between maps)
- Possibility to set a reference map.

Meta-analysis - *Methods* -

MetaQTL

Consensus map

Veyrieras et al. 2007

Evaluation of the method by simulations:
n maps with a proportion ***p*** common markers

	n	2				5				10			
	p	0.15	0.25	0.50	0.75	0.15	0.25	0.50	0.75	0.15	0.25	0.50	0.75
BC	$\overline{\text{IMSE}}$	0.28	0.44	0.82	1.24	0.33	0.52	1.01	1.62	0.43	0.64	1.02	1.65
	$\overline{\text{IMSEP}}$	0.74	0.78	0.95	1.06	0.97	**1.17**	1.62	2.08	**1.32**	**1.67**	2.45	3.30
F2	$\overline{\text{IMSE}}$	0.24	0.54	0.87	1.39	0.33	0.58	1.22	1.97	0.46	0.76	1.40	3.01
	$\overline{\text{IMSEP}}$	0.71	0.76	0.90	1.02	0.94	**1.16**	1.60	1.97	**1.30**	**1.67**	2.35	2.80

IMSE: Interval Mean Square Error,
IMSEP: Interval Mean Square Error of prediction

A measure of the quality of the consensus map:

⇒ If >1, quality of the consensus map > quality of the individual map.

⇒ Increasing the number of maps (and the proportion of common markers) increase quality of the consensus map.

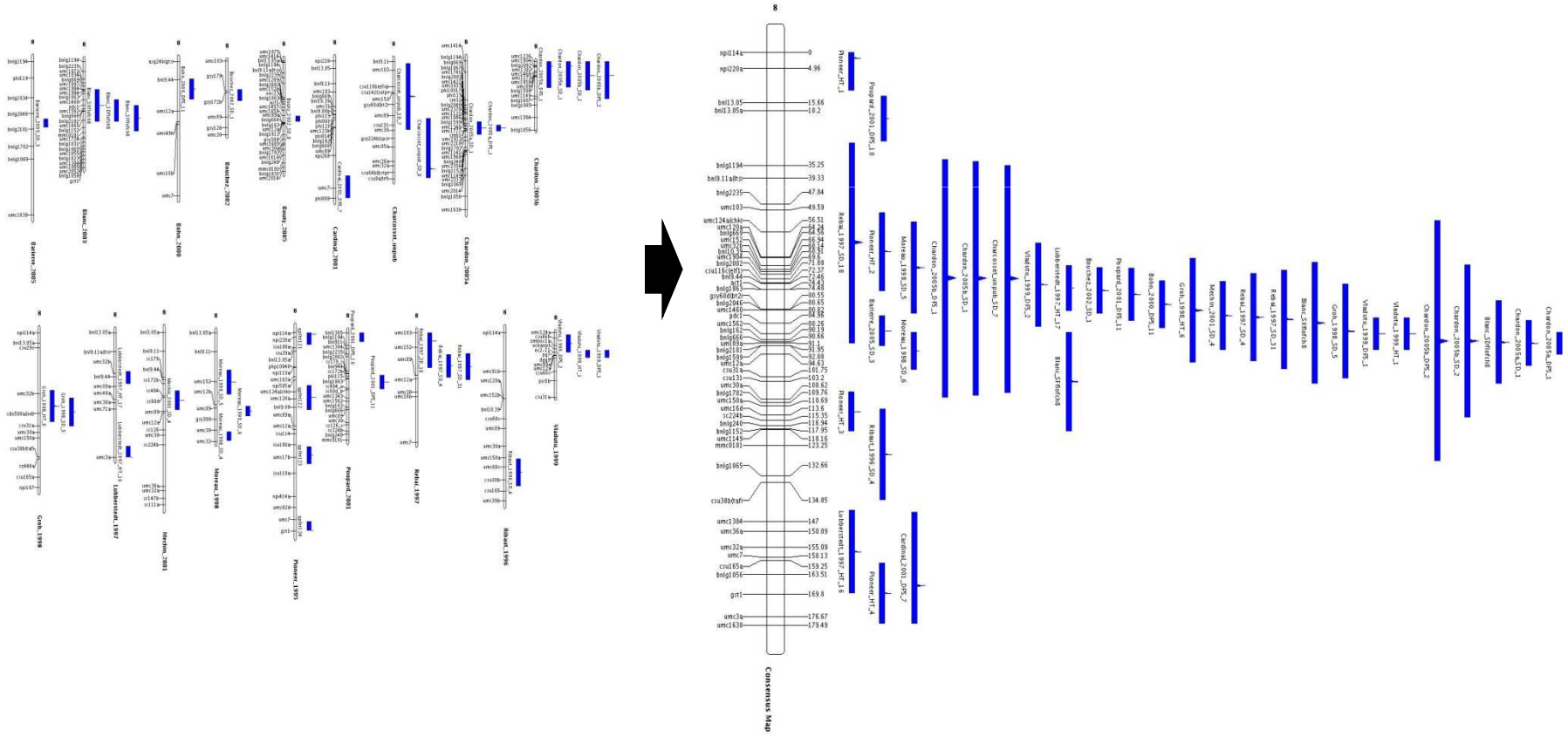
Meta-analysis - *Methods* -

MetaQTL

QTL projection on the consensus map

Veyrieras et al. 2007

The simple homothetic rule + the ratio R is weighted by the probability that the QTL lies in the intervals of markers used for R calculation



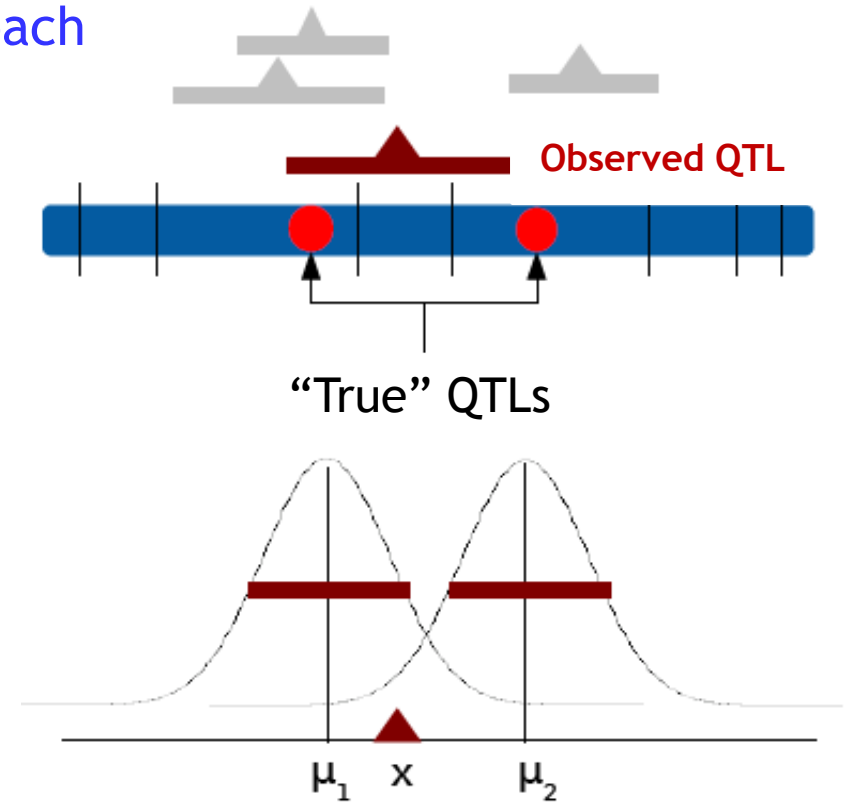
Meta-analysis - *Methods* -

MetaQTL QTL meta-analysis Vevrieras et al. 2007

Clustering based approach

k is the "true" number of QTL:

- Likelihood of observed QTL positions can be expressed as a Gaussian mixture model.
- Variances of observed QTL positions can be estimated using confidence intervals.
- Estimation of positions μ_k of the "true" QTLs by maximisation of the log-likelihood (EM algorithm)
- Criteria for choosing the "best" **k**: AIC*

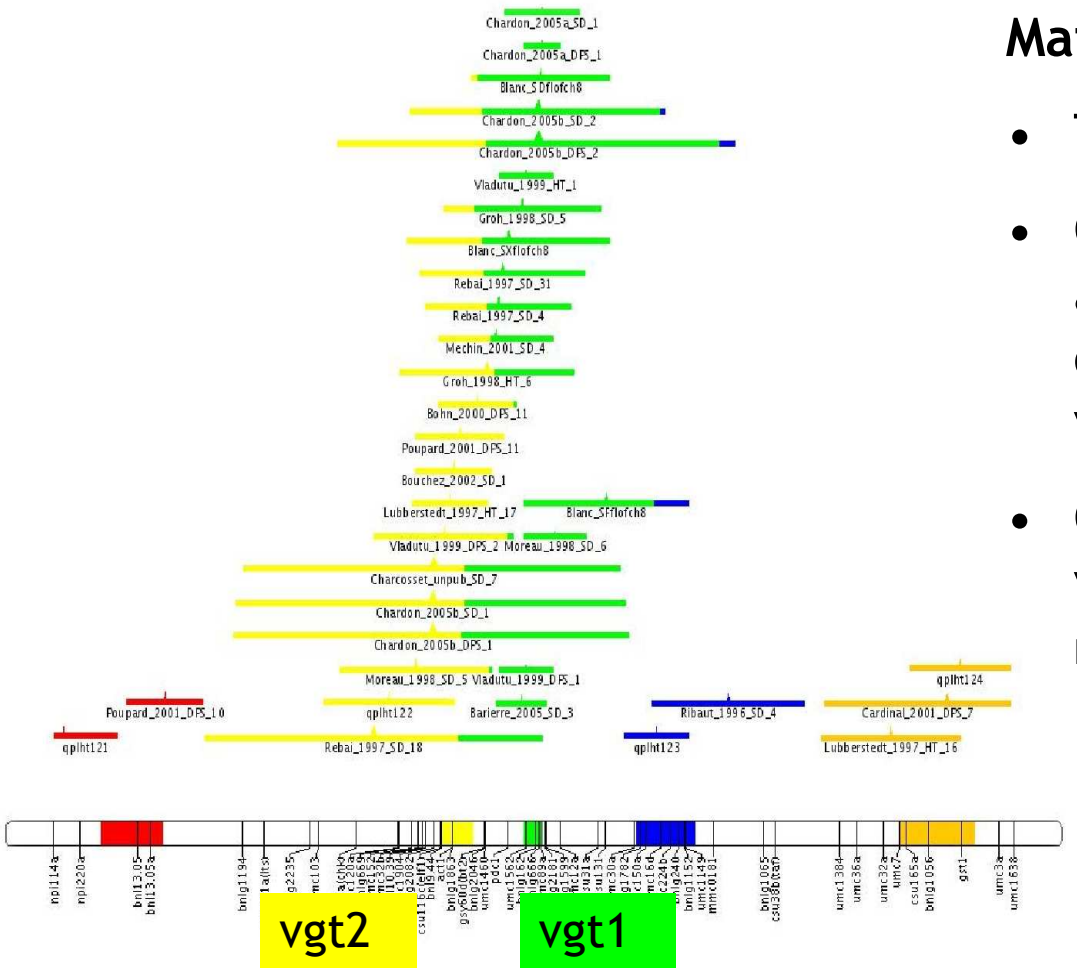


$Pr(x; \mu_1, \mu_2)$ = Gaussian mixture model

Meta-analysis - *Methods* -

MetaQTL QTL meta-analysis Veyrieras et al. 2007

Simulations and analysis of real data sets showed that the confidence intervals of the meta-QTLs are considerably reduced compared to the observed QTL



Maize example:

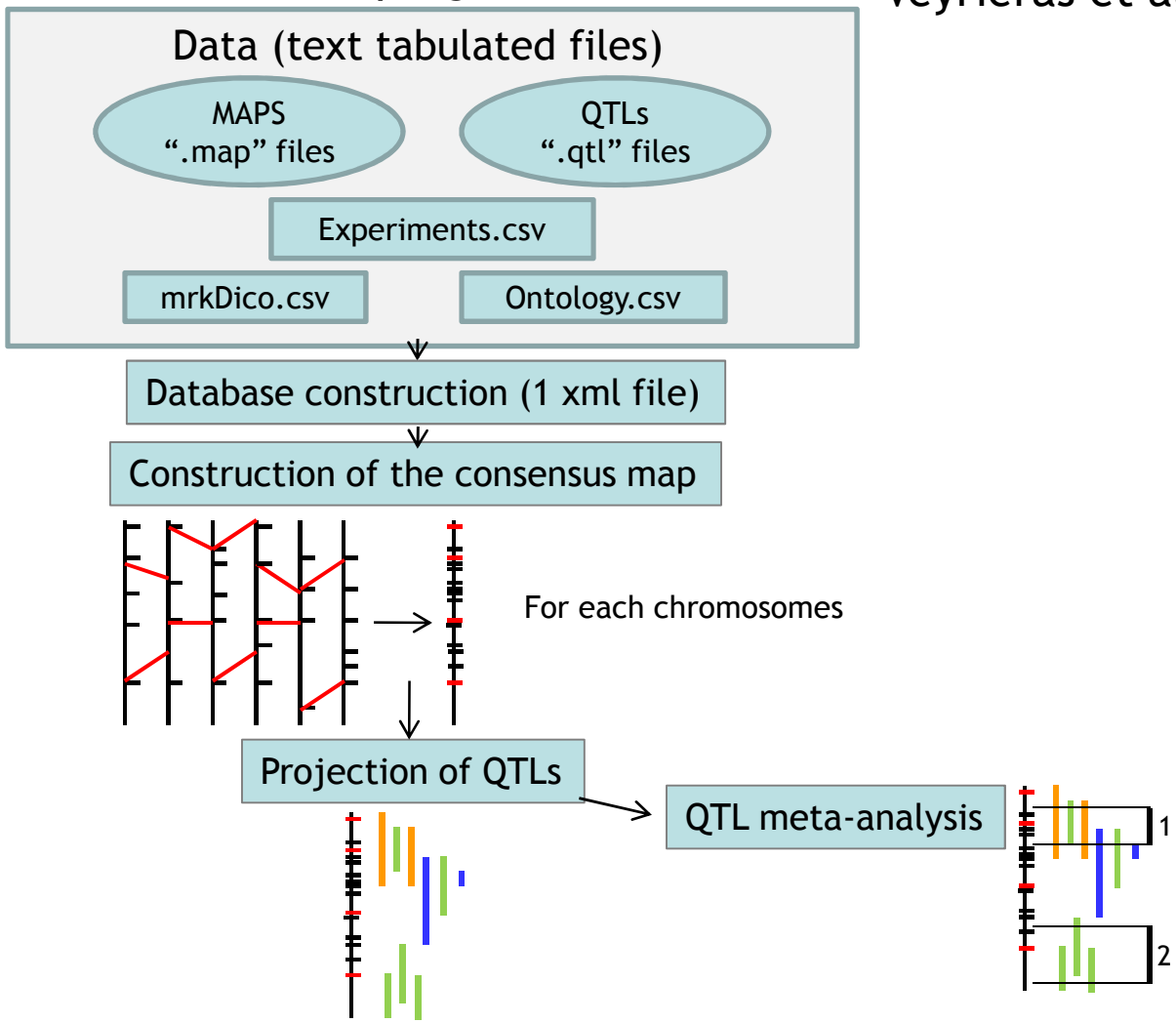
- The best model **k** = 5 QTLs.
- Colors -> quantitative assignation of QTL -> 75% of observed QTLs correspond to vgt1 and/or vgt2.
- Confidence intervals reduced: vgt1 (~ 4cM) include a fine mapped QTL (Salvi et al., 2000).

Meta-analysis - *Methods* -

MetaQTL

How the program works

Veyrieras et al. 2007



Not user friendly: command lines ...

Meta-analysis - *Methods* -

MetaQTL

Data input format

Veyrieras et al. 2007

Experiments.csv

Tabulated text

map	mapping.function	mapping.unit	mapping.expansion	cross.type	cross.name	cross.size
Rae_2009	Kosambi	cM	0	BC	POP1	210
Jorge_2010_T	Kosambi	cM	0	BC	POP2	332
Jorge_2010_D	Kosambi	cM	0	BC	POP2	332
Novaes_2009	Kosambi	cM	0	BC	52-124	396
Paolucci_2010_6K3	Kosambi	cM	0	BC	POP4	140
Paolucci_2010_14P11	Kosambi	cM	0	BC	POP4	140

mrkDico.csv

Tabulated text

ORPM127	ORPM127a
ORPM15	ORPM015
ORPM16	ORPM016
ORPM23	ORPM023
ORPM26	ORPM026
ORPM276	ORPM276b
ORPM28	ORPM028a
ORPM29	ORPM029
ORPM30	ORPM030b
ORPM312	ORPM312a
ORPM49	ORPM049x
ORPM50	ORPM050
ORPM64	ORPM064
PMGC204	PMGC204a

Two columns
Aliases or synonyms for a list of markers

Meta-analysis - *Methods* -

MetaQTL

Data input format

Veyrieras et al. 2007

Ontology.csv

Tabulated text

```
term_id term_name   parent_id   synonyms    definition
1   my_ontology
2   growth      1
3   leaf_traits 1
4   height1     2       Stem heigth at the end of the first growing season
5   circum1     2       Circumference at the end of the first growing season
6   height2     2       Stem heigth at the end of the second growing season
7   circum2     2       Circumference at the end of the second growing season
10  syllep1     2       Number of sylleptics branches at the end of the first growing season
```

1 my_ontology

2 growth

- height1 — Stem heigth at the end of the first growing season
- Circum1 — Circumference at the end of the first growing season
- height2 — Stem heigth at the end of the second growing season
- Circum2 — Circumference at the end of the second growing season
- syllep1 — Number of sylleptics branches at the end of the first growing season

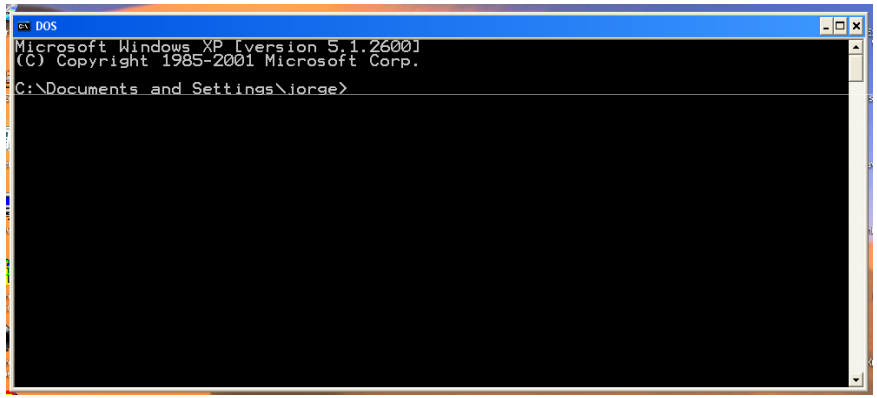
Meta-analysis - *Methods* -

MetaQTL How the program works Veyrieras et al. 2007

On Windows

- Create the environment variable *CLASSPATH*
- Launch a command line interface (MS-DOS):

  Start > All programs > Accessories > Command Line Interface



Advise: save command lines in a txt file

Meta-analysis - *Methods* -

MetaQTL

How the program works

Veyrieras et al. 2007

Command structure

metaqtl_commands.txt

On DOS, execute command lines in the directory where the 'metaqtl.jar' is.

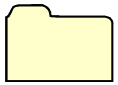
Command for the program language

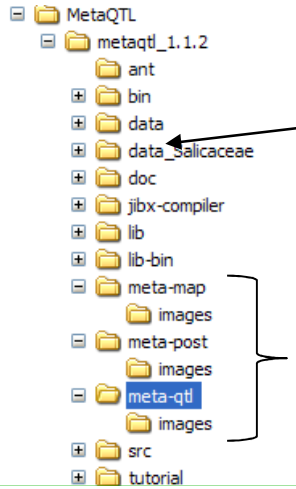
Call the metaqtl program, and the module MetaDB

```
>java -classpath metaqtl.jar org.metaqtl.main.MetaDB
        -e data/experiments.csv
        -m data/genetic-map
        -q data/qtl-map
        -t data/trait_ontology.csv
        -d data/marker_dictionary.csv -o xml
```

Call data

Directory structure





Your data

Directories to create:

meta-map, images

meta-qtl, images

meta-post, images

Meta-analysis - *Methods* -

MetaQTL How the program works Veyrieras et al. 2007

Modules

MetaDB	*Creating the database*
InfoMap	*Evaluations/basic comparisons of the maps*
MapView	*Visualisation of one map*
MMapView	*Visualisation of several maps*
ConsMap	*Construction of the consensus map*
QTLProj	*Projection of the QTLs on the consensus map*
QTLClust, QTLTree	*Clustering of the QTLs*
Xml2A, A2Xml	*Converting files*

Meta-analysis - *Methods* -

MetaQTL How the program works Veyrieras et al. 2007

Data included in the example: data from maize
~18 maps

map	mapping.function	mapping.unit	mapping.expansion	mapping.cross	cross.name
Barierre_2005	haldane cM	0	RI1	F838xF286	RI1	242
Blanc_2003	haldane cM	0	SF2	DexF283xF810xF9005	SF2	299
Bohn_1996	haldane cM	0	SF2	CML131xCML67	SF2	215
Bohn_2000	haldane cM	0	SF2	DO6xD408	SF2	226
Bouchez_2002	haldane cM	0	BC	F2xMBS847	BC	217
Cardinal_2001	haldane cM	0	RI1	B73xB52	RI1	200
Charcosset_unpub	haldane cM	0	RI1	F2xF252	RI1	129
Chardon_2005a	haldane cM	0	SF3	F7pxF2	SF3	150
Chardon_2005b	haldane cM	0	SF3	F7pxGaspe	SF3	150
Groh_1998	haldane cM	0	RI1	CML131xCML67	RI1	166
Lubberstedt_1997	haldane cM	0	SF2	KW1265xD146	SF2	380
Mechin_2001	haldane cM	0	SF5	F2xMBS847	SF5	100
Moreau_1998	haldane cM	0	SF3	F2xF252	SF3	300
Pioneer_1995	haldane cM	0	SF4	unknown	SF4	976
Poupard_2001	haldane cM	0	SF6	F2xMBS847	SF6	86
Rebai_1997	haldane cM	0	SF2	EuropeanxAmerican	SF2	1200
Ribaut_1996	haldane cM	0	SF2	Tropical	SF2	272
Vladutu_1999	haldane cM	0	SF2	E20xN28	SF2	88

Traits involved in flowering time

Term_id	Term_name	Parent_id	Synonyms	Definition
1	trait ontology			
2	FT	1	Flowering Time	
3	DPS	2	Days to pollen shed	
4	SD	2	Silking date	
5	HT	2	Plant height	
6	LN	2	Leaf number	

Meta-analysis - *Comparison of methods* -

	Biomercator (Goffinet & Gerber, 2000)	**MetaQTL** (Veyrieras et al., 2007)
Assumptions	- Independent QTL experiments - QTL data set = 10 to 40 - Size of the region analysed < 200 cM	- Independent QTL experiments
Construction of the consensus map	Homothetic projection Iterative	Weighted Least Square method One step
QTL projection	Homothetic projection	"Weighted" homothetic projection
QTL meta-analysis	Model tested: ***k*** = 1, 2, 3, 4 or n	Model tested: ***k*** = 1 to n Gaussian mixture model
Environment	Graphical interface	Command lines

Meta-analysis - *Software improvements-*

MetaQTL project

http://urgi.versailles.inra.fr/projects/MetaQTL/index.php

Objectives

- Develop a new version of Biomercator
  - integrate powerful statistical methods from MetaQTL in a friendly user graphical interface
  - implement methods for the integration of physical maps (genome sequence)
  - visualisation of meta-QTLs in a genome browser (links to GnpMap: http://urgi.versailles.inra.fr/data/gnpMap/index.php)

References

Methodology and program references

Arcade, A., A. Labourdette, M. Falque, B. Mangin, F. Chardon, A. Charcosset, and J. Joets. 2004. BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. Bioinformatics 20:2324-2326.

Goffinet, B., and S. Gerber. 2000. Quantitative Trait Loci: A Meta-analysis. Genetics 155:463-473.

Veyrieras, J.-B., B. Goffinet, and A. Charcosset. 2007. MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments. BMC Bioinformatics 8:49.

Applications in crops with MetaQTL methodology

Loffler, M., C.C. Schon, and T. Miedaner. 2009. Revealing the genetic architecture of FHB resistance in hexaploid wheat (Triticum aestivum L.) by QTL meta-analysis. Molecular Breeding 23:473-488.

Marandel, G., J. Salava, A. Abbott, T. Candresse, and V. Decroocq. 2009. Quantitative trait loci meta-analysis of Plum pox virus resistance in apricot (Prunus armeniaca L.): new insights on the organization and the identification of genomic resistance factors. Molecular Plant Pathology 10:347-360.

References

Applications in crops with Goffinet & Gerber / Biomercator methodology

Ballini, E., J.B. Morel, G. Droc, A. Price, B. Courtois, J.L. Notteghem, and D. Tharreau. 2008. A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. Molecular Plant-Microbe Interactions 21:859-868.

Chardon, F., B. Virlon, L. Moreau, M. Falque, J. Joets, L. Decousset, A. Murigneux, and A. Charcosset. 2004. Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics 168:2169-2185.

Lanaud, C., O. Fouet, D. Clement, M. Boccara, A.M. Risterucci, S. Surujdeo-Maharaj, T. Legavre, and X. Argout. 2009. A meta-QTL analysis of disease resistance traits of Theobroma cacao L. Molecular Breeding 24:361-374.

Hanocq, E., A. Laperche, O. Jaminon, A.L. Laine, and J. Le Gouis. 2007. Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. Theoretical and Applied Genetics 114:569-584.

Khowaja, F.S., G.J. Norton, B. Courtois, and A.H. Price. 2009. Improved resolution in the position of drought-related QTLs in a single mapping population of rice by meta-analysis. Bmc Genomics 10.

Shi, J., R. Li, D. Qiu, C. Jiang, Y. Long, C. Morgan, I. Bancroft, J. Zhao, and J. Meng. 2009. Unraveling the complex trait of crop yield with quantitative trait loci mapping in Brassica napus. Genetics 182:851-861.

Rong, J., E.A. Feltus, V.N. Waghmare, G.J. Pierce, P.W. Chee, X. Draye, Y. Saranga, R.J. Wright, T.A. Wilkins, O.L. May, C.W. Smith, J.R. Gannaway, J.R. Wendel, and A.H. Paterson. 2007b. Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176:2577-2588.