Project started in december 2009

OptiMAS: a decision support tool to conduct Marker Assisted Selection (MAS) programs

Generation
Cultivating Plant Diversity for the Resource-Poor

F. Valente, F. Gauthier, J. Joets
A. Charcosset & L. Moreau

PAG Meeting
San Diego, CA, USA
January 14-18, 2012

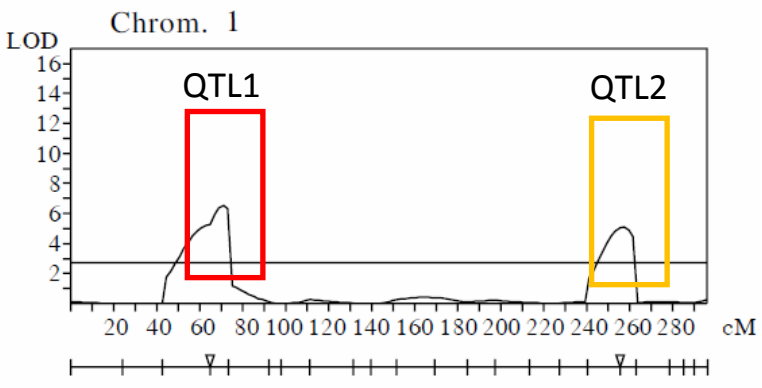
Le Moulon Génétique végétale UMR
INRA

Initial user cases: 8 crops across 32 countries

Cereals	
Maize	**Africa:** Angola, Ethiopia, Kenya, Malawi, Mozambique, Tanzania, Uganda, Zambia, Zimbabwe **Asia:** China, India, Indonesia, Thailand, The Philippines, Vietnam
Rice	**Africa:** Benin, Burkina Faso, Ethiopia, Gambia, Ghana, Guinea, Liberia, Madagascar, Mali, Mozambique, Nigeria, Rwanda, Senegal, Tanzania, Uganda **Asia:** Bangladesh, Cambodia, China, India, Indonesia, Laos, Nepal, Pakistan, Sri Lanka, Vietnam
Sorghum	Mali
Wheat	**Africa:** Ethiopia, Kenya **Asia:** China, India
Legumes	
Beans	Ethiopia, Kenya, Tanzania, Malawi
Chickpeas	**Africa:** Ethiopia, Kenya **Asia:** India
Cowpeas	Burkina Faso, Mozambique, Senegal
Roots and tubers	
Cassava	Ghana, Nigeria, Tanzania, Uganda

Breeding decisions adapted to different crops / projects

P1 (allele « a ») × *P2 (allele « b »)*



- QTL will be detected for different traits of interest and favorable alleles will be found

- Aim: to create a target genotype with all the favorable alleles at the QTL positions

Target genotype

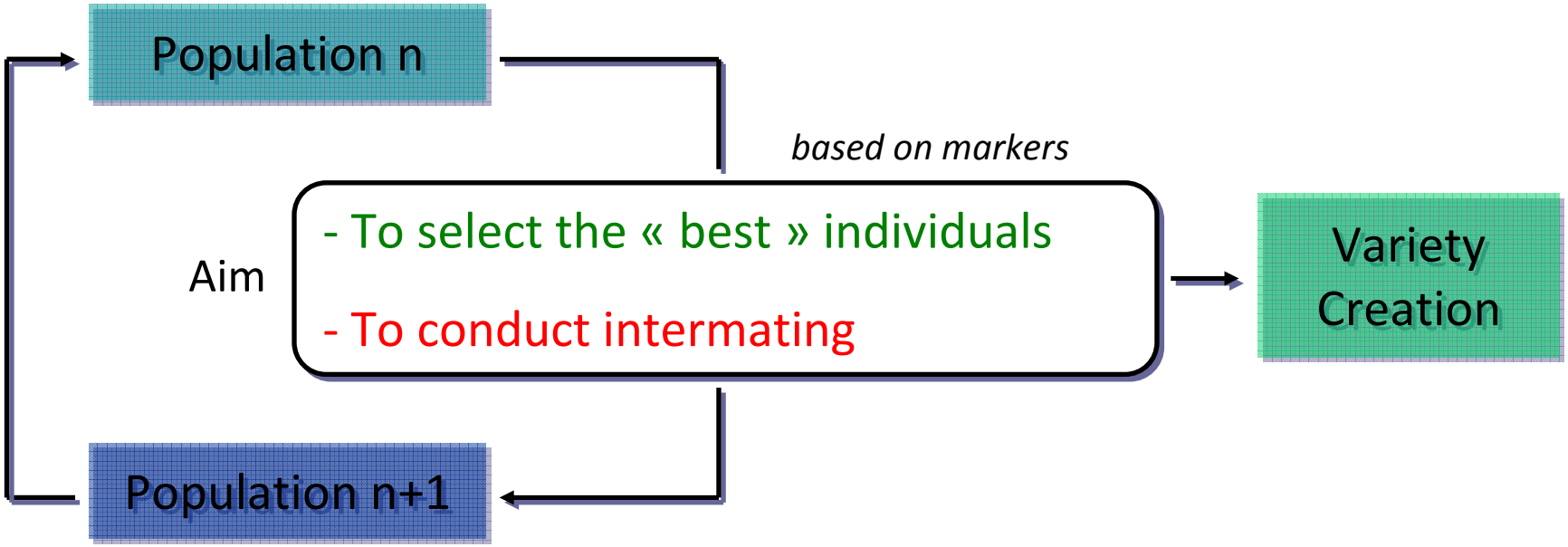
(ideotype)

Mostly biparental population(s), but several cases where populations are connected through common parent(s)

Common objective: to create new genetic materials assembling favorable QTL alleles from 2 parents or several parents (get challenging as the No. of QTLs increases!)

→ Need user oriented tool to assemble favorable alleles through Marker Assisted Recurrent Selection (MARS) strategies

Implementation of Marker Assisted Recurrent Selection in OptiMAS

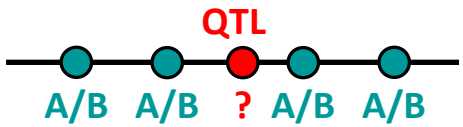


- ✓ Objective 1: to develop algorithms to compute probabilities of allele transmission through generations
- ✓ Objectif 2: to identify the best individuals
- ✓ Objective 3: to develop methodology to identify the best intermating scheme to accumulate favourable alleles and to extract varieties

Background elements on Objective 1: probabilities of allele transmission at QTLs

Aim: to evaluate individuals, we need to know which parental alleles are transmitted at the QTL position, based on the information from neighboring markers

QTL position rarely located at a marker → QTL alleles are unknown and must be inferred from flanking markers

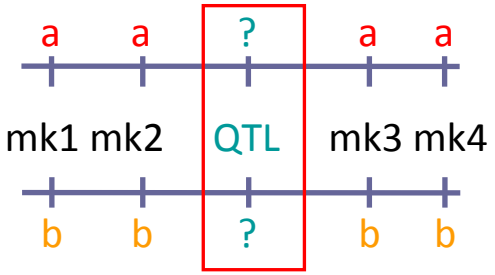


e.g 4 markers / QTL

(bi/multi allelic context)

Selection of a subset of informative markers to follow the favorable parental alleles (based on haplotypes)

Identification of all the possible phased genotypes for each QTL and computation of their probabilities



probabilities of allele transmission

- Information available (input files: pedigree/genotype & genetic map)
  - ✓ Pedigree
  - ✓ Molecular markers (observed genotypes)

Id	*P1*	*P2*	*Step*	*Cycle*	*Group*	*mrk1*	*mrk2*	*mrk3*
Fr2	a	a	IL	IL		A	A	A
Fr252	b	b	IL	IL		B	B	B
f1	Fr2	Fr252	CR	F1		A/B	A/B	A/B
ind1	f1	f1	S2	F3		A	A	A
ind2	f1	f1	S2	F3		A	A	A
ind3	f1	f1	S2	F3		A/B	-	A
ind4	f1	f1	S2	F3		B	B	A/B
ind5	f1	f1	S2	F3		B	B	B
ind6	f1	f1	S2	F3				
ind7	f1	f1	S2	F3		A	A	A

probabilities of allele transmission

- Information available (input files: pedigree/genotype & genetic map)
  - ✓ Pedigree
  - ✓ Molecular markers (observed genotypes)
  - ✓ QTL position
  - ✓ Distance between loci (Haldane → recombination rate)

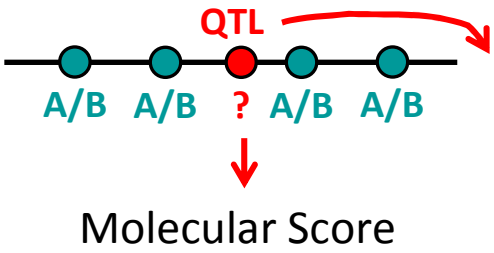
Locus	*Chr*	*QTL*	*Pos*	*All+*
marker1	1	1	42.2	
marker2	1	1	64.0	
qtl1	1	1	**70.0**	**a**
marker3	1	1	72.5	
marker4	1	1	90.8	
marker5	1	2	237.1	
marker6	1	2	252.2	
qtl2	1	2	**254.0**	**b**
marker7	1	2	259.5	
marker8	1	2	274.8	

d

Objective 1: probabilities of allele transmission

- Information available (input files: pedigree/genotype & genetic map)
  - ✓ Pedigree
  - ✓ Molecular markers (observed genotypes)
  - ✓ QTL position
  - ✓ Distance between loci (Haldane → recombination rate)

- Algorithm: compute the probabilities of allele transmission in different MAS schemes and mating designs (intercrossing, selfing, backcrossing, DH, RIL)
  - ✓ 2 versions: command line → integrated in a new GUI

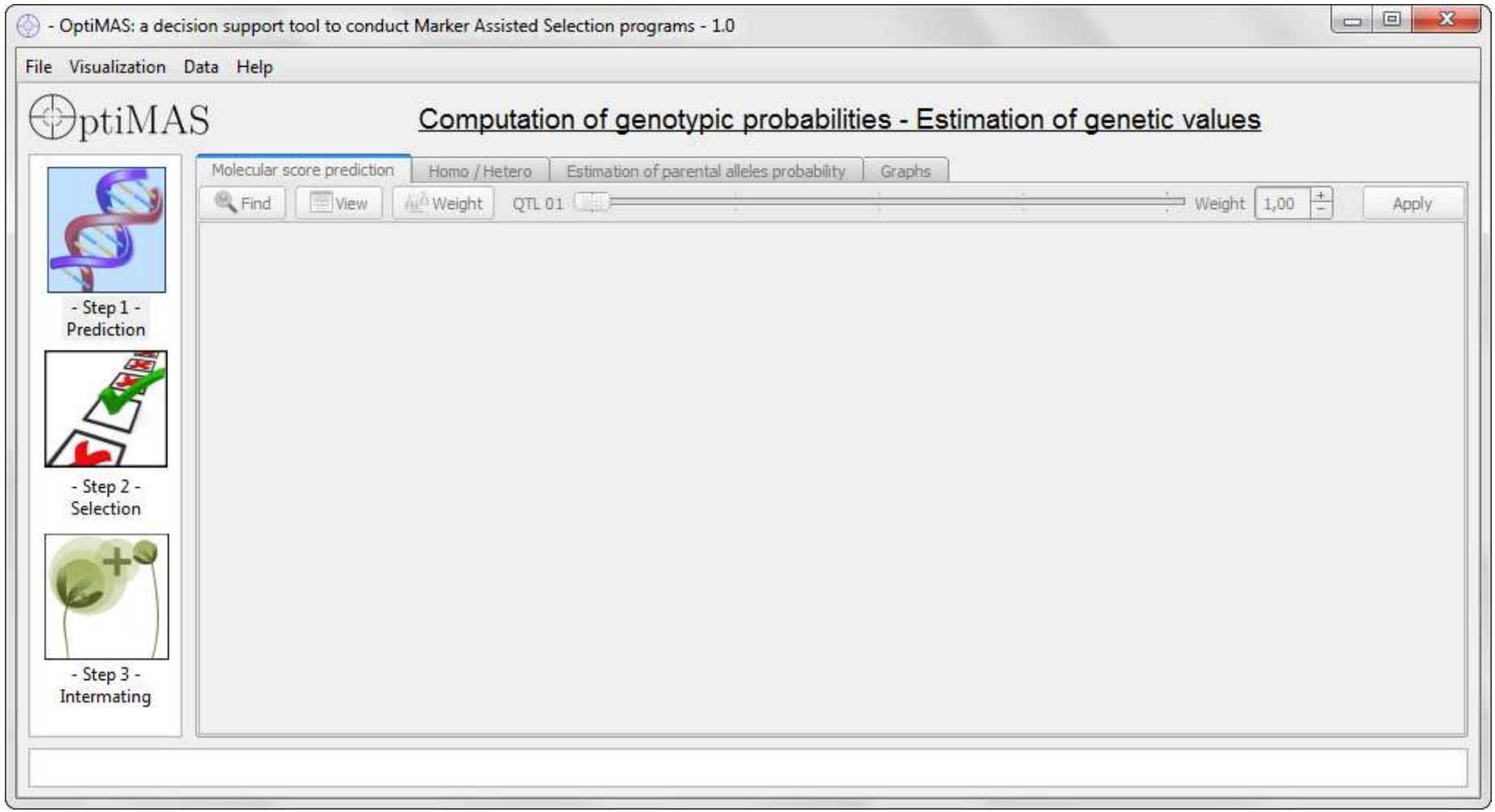


p(a/a) = 0.9, p(a/b) = 0.1, p(b/b) = 0 → probability of each genotype at QTL given markers

MS = 0.95 → Probability to carry favorable allele at the QTL position

Presentation of mean features of OptiMAS

❑ New Graphical User Interface (GUI) organized in 3 modules corresponding to the ≠ steps of the selection program



Step 1: Computation of genotypic probabilities – Estimation of genetic values

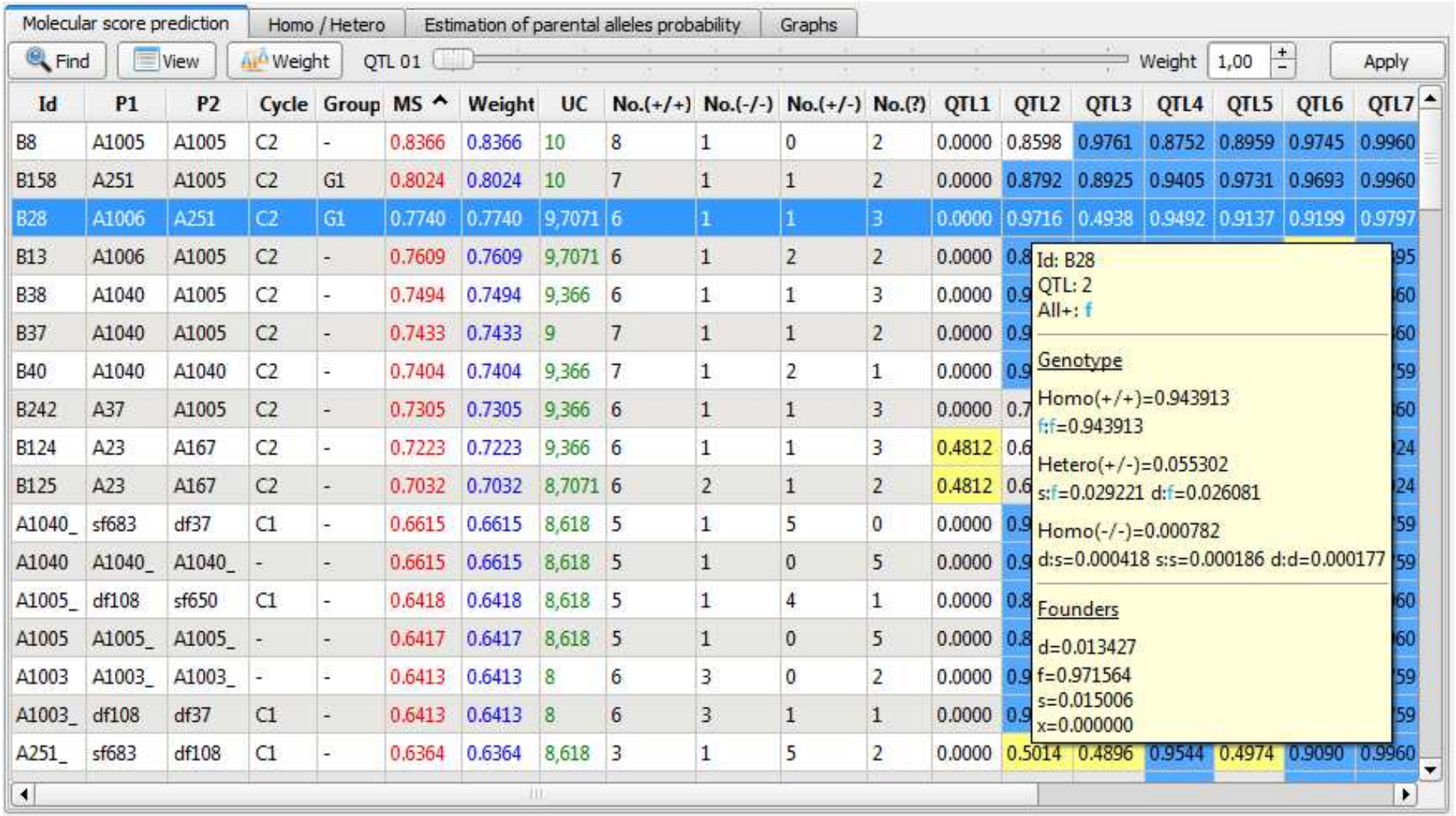
Molecular score prediction | Homo / Hetero | Estimation of parental alleles probability | Graphs

Find | View | Weight | QTL 01 | Weight 1,00 | Apply

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	0.8366	10	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	0.8024	10	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	0.7740	9,7071	6	1	1	3	0.0000	0.9716	0.4938	0.9492	0.9137	0.9199	0.9797
B13	A1006	A1005	C2	-	0.7609	0.7609	9,7071	6	1	2	2	0.0000	0.8844	0.9559	0.9179	0.9731	0.4963	0.9895
B38	A1040	A1005	C2	-	0.7494	0.7494	9,366	6	1	1	3	0.0000	0.9349	0.4907	0.9429	0.5822	0.9260	0.9860
B37	A1040	A1005	C2	-	0.7433	0.7433	9	7	1	1	2	0.0000	0.9528	0.2334	0.9172	0.8735	0.9718	0.9860
B40	A1040	A1040	C2	-	0.7404	0.7404	9,366	7	1	2	1	0.0000	0.9775	0.4376	0.9592	0.4974	0.9231	0.9759
B242	A37	A1005	C2	-	0.7305	0.7305	9,366	6	1	1	3	0.0000	0.7259	0.8959	0.8879	0.2896	0.9722	0.9860
B124	A23	A167	C2	-	0.7223	0.7223	9,366	6	1	1	3	0.4812	0.6698	0.9559	0.9429	0.2787	0.9614	0.9924
B125	A23	A167	C2	-	0.7032	0.7032	8,7071	6	2	1	2	0.4812	0.6536	0.9559	0.9429	0.0058	0.9616	0.9924
A1040_	sf683	df37	C1	-	0.6615	0.6615	8,618	5	1	5	0	0.0000	0.9771	0.4842	0.9592	0.4974	0.9231	0.9759
A1040	A1040_	A1040_	-	-	0.6615	0.6615	8,618	5	1	0	5	0.0000	0.9771	0.4842	0.9592	0.4974	0.9231	0.9759
A1005_	df108	sf650	C1	-	0.6418	0.6418	8,618	5	1	4	1	0.0000	0.8928	0.4896	0.8751	0.4878	0.9747	0.9960
A1005	A1005_	A1005_	-	-	0.6417	0.6417	8,618	5	1	0	5	0.0000	0.8927	0.4895	0.8751	0.4878	0.9747	0.9960
A1003	A1003_	A1003_	-	-	0.6413	0.6413	8	6	3	0	2	0.0000	0.9757	0.9738	0.9135	0.0000	0.4986	0.9759
A1003_	df108	df37	C1	-	0.6413	0.6413	8	6	3	1	1	0.0000	0.9757	0.9738	0.9135	0.0000	0.4987	0.9759
A251_	sf683	df108	C1	-	0.6364	0.6364	8,618	3	1	5	2	0.0000	0.5014	0.4896	0.9544	0.4974	0.9090	0.9960

- Individuals in lines, MS value [0-1] → 1 = ideotype

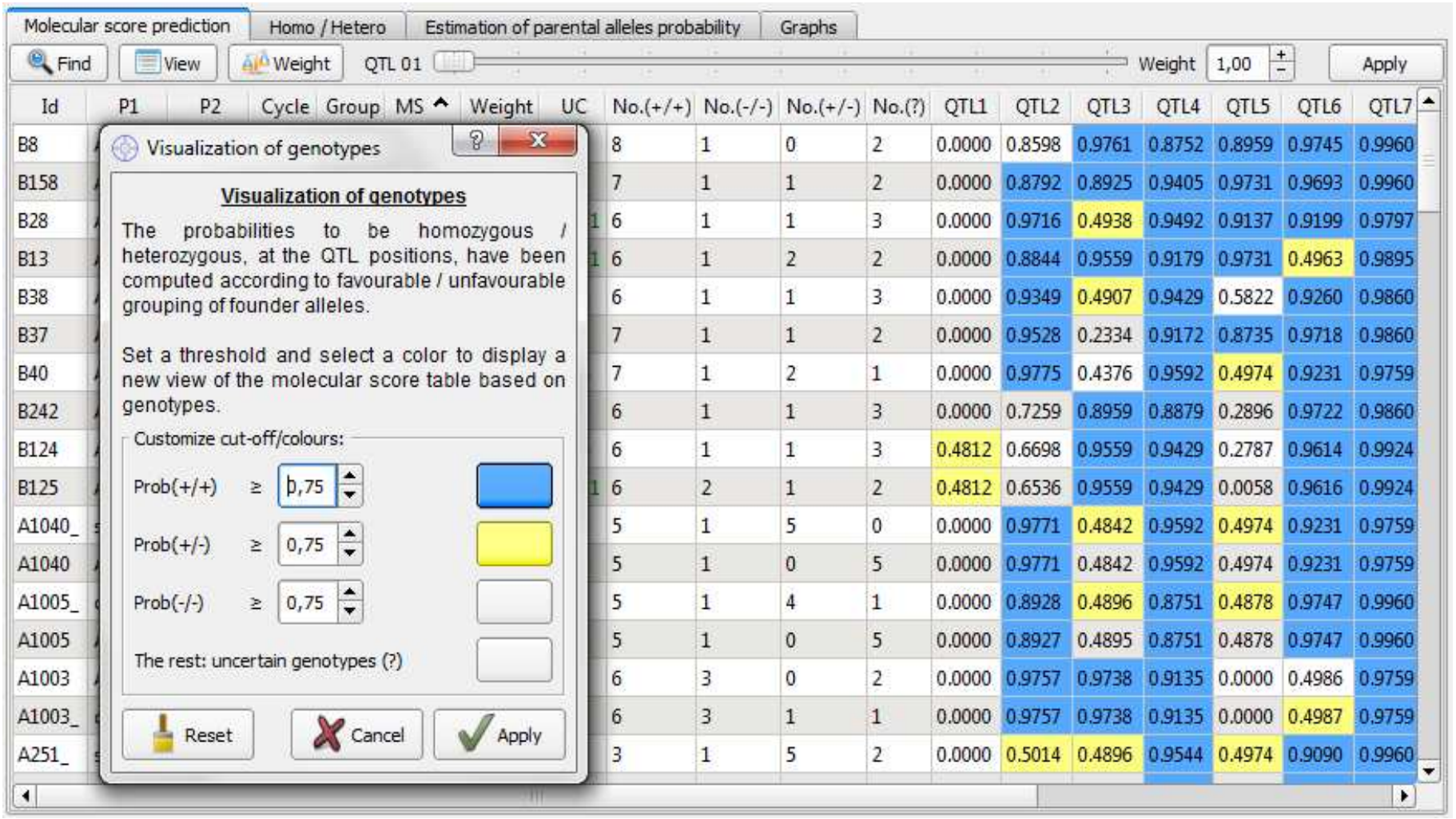
Step 1: Computation of genotypic probabilities – Estimation of genetic values



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	0.8366	10	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	0.8024	10	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	0.7740	9,7071	6	1	1	3	0.0000	0.9716	0.4938	0.9492	0.9137	0.9199	0.9797
B13	A1006	A1005	C2	-	0.7609	0.7609	9,7071	6	1	2	2	0.0000	0.8					95
B38	A1040	A1005	C2	-	0.7494	0.7494	9,366	6	1	1	3	0.0000	0.9					60
B37	A1040	A1005	C2	-	0.7433	0.7433	9	7	1	1	2	0.0000	0.9					60
B40	A1040	A1040	C2	-	0.7404	0.7404	9,366	7	1	2	1	0.0000	0.9					59
B242	A37	A1005	C2	-	0.7305	0.7305	9,366	6	1	1	3	0.0000	0.7					60
B124	A23	A167	C2	-	0.7223	0.7223	9,366	6	1	1	3	0.4812	0.6					24
B125	A23	A167	C2	-	0.7032	0.7032	8,7071	6	2	1	2	0.4812	0.6					24
A1040_	sf683	df37	C1	-	0.6615	0.6615	8,618	5	1	5	0	0.0000	0.9					59
A1040	A1040_	A1040_	-	-	0.6615	0.6615	8,618	5	1	0	5	0.0000	0.9					59
A1005_	df108	sf650	C1	-	0.6418	0.6418	8,618	5	1	4	1	0.0000	0.8					60
A1005	A1005_	A1005_	-	-	0.6417	0.6417	8,618	5	1	0	5	0.0000	0.8					60
A1003	A1003_	A1003_	-	-	0.6413	0.6413	8	6	3	0	2	0.0000	0.9					59
A1003_	df108	df37	C1	-	0.6413	0.6413	8	6	3	1	1	0.0000	0.9					59
A251_	sf683	df108	C1	-	0.6364	0.6364	8,618	3	1	5	2	0.0000	0.5014	0.4896	0.9544	0.4974	0.9090	0.9960

- Prob(++/--/+-) at the QTL position + detailed genotype in terms of parental alleles
- Estimation (probability) of parental allele transmission

Step 1: Computation of genotypic probabilities – Estimation of genetic values

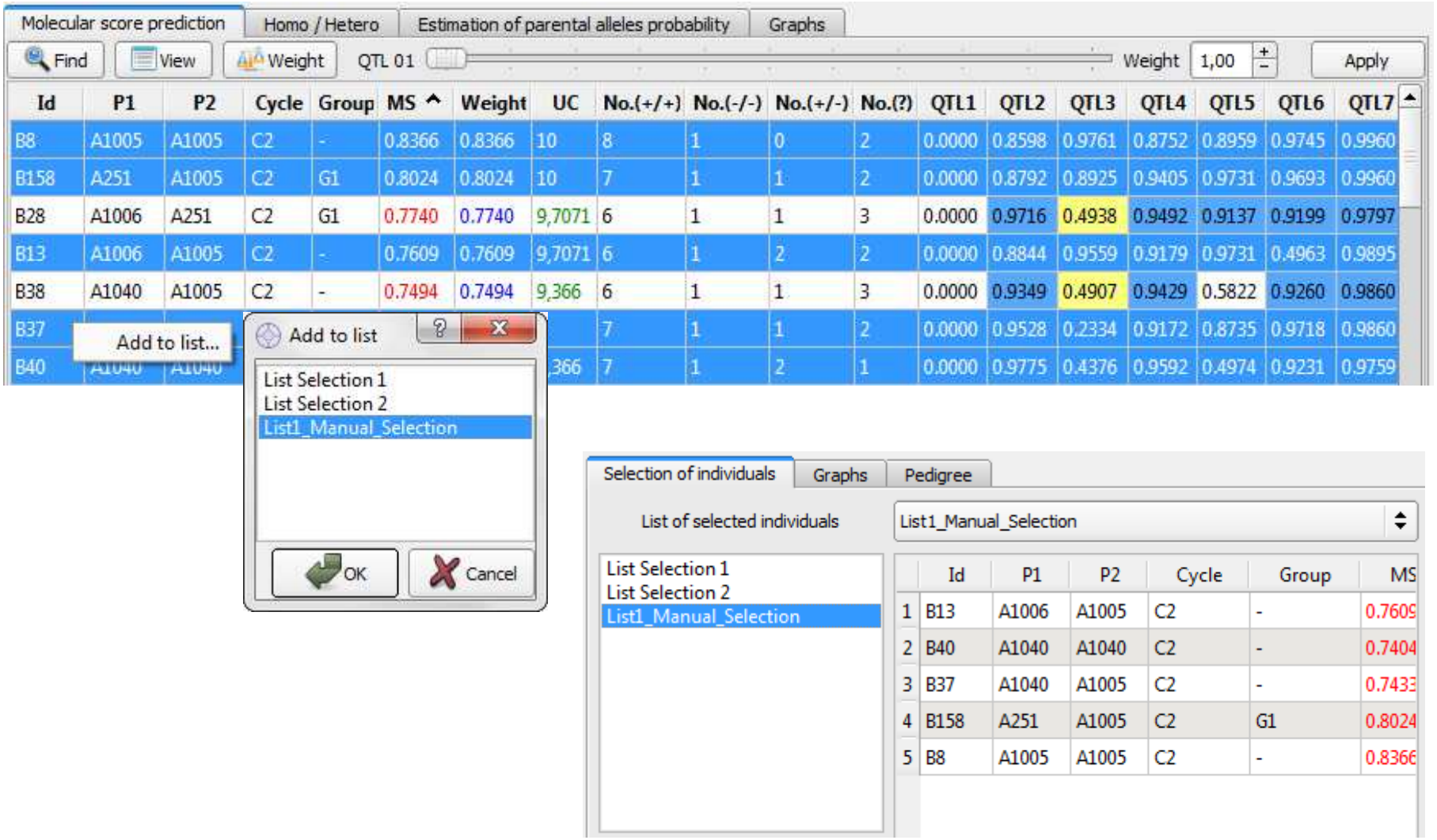


Id	P1	P2	Cycle	Group	MS	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7
B8								8	1	0	2	0.0000	0.8598	0.9761	0.8752	0.8959	0.9745	0.9960
B158								7	1	1	2	0.0000	0.8792	0.8925	0.9405	0.9731	0.9693	0.9960
B28							1	6	1	1	3	0.0000	0.9716	0.4938	0.9492	0.9137	0.9199	0.9797
B13							1	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
B124								6	1	1	3	0.4812	0.6698	0.9559	0.9429	0.2787	0.9614	0.9924
B125							1	6	2	1	2	0.4812	0.6536	0.9559	0.9429	0.0058	0.9616	0.9924
A1040_								5	1	5	0	0.0000	0.9771	0.4842	0.9592	0.4974	0.9231	0.9759
A1040								5	1	0	5	0.0000	0.9771	0.4842	0.9592	0.4974	0.9231	0.9759
A1005_								5	1	4	1	0.0000	0.8928	0.4896	0.8751	0.4878	0.9747	0.9960
A1005								5	1	0	5	0.0000	0.8927	0.4895	0.8751	0.4878	0.9747	0.9960
A1003								6	3	0	2	0.0000	0.9757	0.9738	0.9135	0.0000	0.4986	0.9759
A1003_								6	3	1	1	0.0000	0.9757	0.9738	0.9135	0.0000	0.4987	0.9759
A251_								3	1	5	2	0.0000	0.5014	0.4896	0.9544	0.4974	0.9090	0.9960

- Visualization of genotypes via colors for each QTL (how many QTL are fixed ?)

Step2: Selection of individuals

- Manual selection

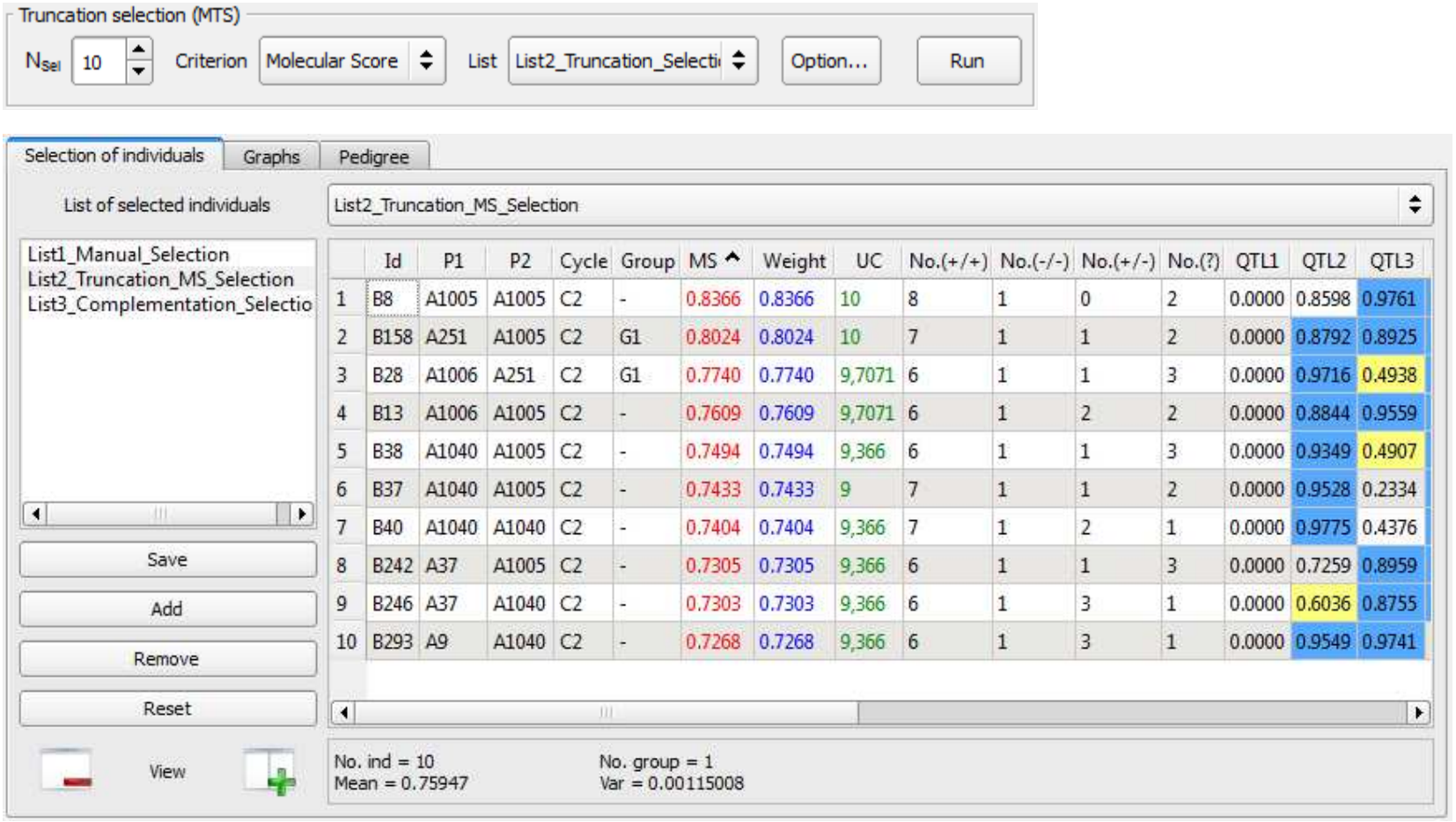


Id	P1	P2	Cycle	Group	MS	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7
B8	A1005	A1005	C2	-	0.8366	0.8366	10	8	1	0	2	0.0000	0.8598	0.9761	0.8752	0.8959	0.9745	0.9960
B158	A251	A1005	C2	G1	0.8024	0.8024	10	7	1	1	2	0.0000	0.8792	0.8925	0.9405	0.9731	0.9693	0.9960
B28	A1006	A251	C2	G1	0.7740	0.7740	9,7071	6	1	1	3	0.0000	0.9716	0.4938	0.9492	0.9137	0.9199	0.9797
B13	A1006	A1005	C2	-	0.7609	0.7609	9,7071	6	1	2	2	0.0000	0.8844	0.9559	0.9179	0.9731	0.4963	0.9895
B38	A1040	A1005	C2	-	0.7494	0.7494	9,366	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							366	7	1	2	1	0.0000	0.9775	0.4376	0.9592	0.4974	0.9231	0.9759

	Id	P1	P2	Cycle	Group	MS
1	B13	A1006	A1005	C2	-	0.7609
2	B40	A1040	A1040	C2	-	0.7404
3	B37	A1040	A1005	C2	-	0.7433
4	B158	A251	A1005	C2	G1	0.8024
5	B8	A1005	A1005	C2	-	0.8366

Step2: Selection of individuals

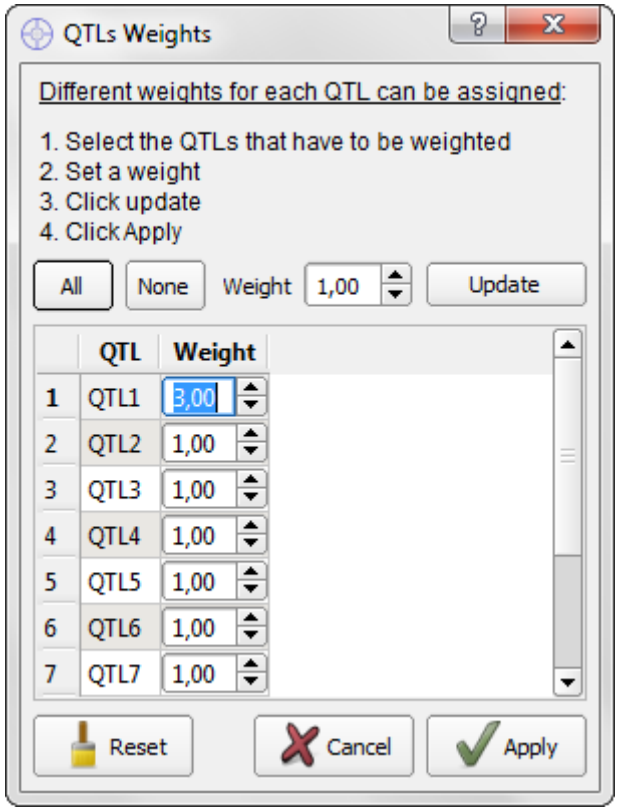
- Manual selection
- Truncation selection based on:
  - ✓ Molecular score (MS)



	Id	P1	P2	Cycle	Group	MS	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3
1	B8	A1005	A1005	C2	-	0.8366	0.8366	10	8	1	0	2	0.0000	0.8598	0.9761
2	B158	A251	A1005	C2	G1	0.8024	0.8024	10	7	1	1	2	0.0000	0.8792	0.8925
3	B28	A1006	A251	C2	G1	0.7740	0.7740	9,7071	6	1	1	3	0.0000	0.9716	0.4938
4	B13	A1006	A1005	C2	-	0.7609	0.7609	9,7071	6	1	2	2	0.0000	0.8844	0.9559
5	B38	A1040	A1005	C2	-	0.7494	0.7494	9,366	6	1	1	3	0.0000	0.9349	0.4907
6	B37	A1040	A1005	C2	-	0.7433	0.7433	9	7	1	1	2	0.0000	0.9528	0.2334
7	B40	A1040	A1040	C2	-	0.7404	0.7404	9,366	7	1	2	1	0.0000	0.9775	0.4376
8	B242	A37	A1005	C2	-	0.7305	0.7305	9,366	6	1	1	3	0.0000	0.7259	0.8959
9	B246	A37	A1040	C2	-	0.7303	0.7303	9,366	6	1	3	1	0.0000	0.6036	0.8755
10	B293	A9	A1040	C2	-	0.7268	0.7268	9,366	6	1	3	1	0.0000	0.9549	0.9741

Step2: Selection of individuals

- Manual selection
- Truncation selection based on:
  - ✓ Molecular score (MS)
  - ✓ Weighted MS: give more or less importance to the different QTL



List3_Truncation_Weight_Selection

	Id	P1	P2	Cycle	Group	MS	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3
1	B8	A1005	A1005	C2	-	0.8366	0.7079	10	8	1	0	2	0.0000	0.8598	0.9761
2	B158	A251	A1005	C2	G1	0.8024	0.6790	10	7	1	1	2	0.0000	0.8792	0.8925
3	B28	A1006	A251	C2	G1	0.7740	0.6550	9,7071	6	1	1	3	0.0000	0.9716	0.4938
4	B13	A1006	A1005	C2	-	0.7609	0.6438	9,7071	6	1	2	2	0.0000	0.8844	0.9559
5	B124	A23	A167	C2	-	0.7223	0.6852	9,366	6	1	1	3	0.4812	0.6698	0.9559
6	B125	A23	A167	C2	-	0.7032	0.6691	8,7071	6	2	1	2	0.4812	0.6536	0.9559
7	B110	A212	A1005	C2	-	0.6863	0.6481	9,2247	4	0	4	3	0.4376	0.7319	0.4876
8	B57	A167	A1040	C2	-	0.6860	0.6545	8,8229	4	0	6	1	0.4812	0.6066	0.4230
9	B123	A212	A91	C2	-	0.6717	0.6390	8,618	4	1	3	3	0.4590	0.6119	0.4879
10	B47	A166	A167	C2	-	0.6309	0.6570	8,366	4	2	2	3	0.8010	0.4966	0.8733

No. ind = 10
Mean = 0.727447
No. group = 1
Var = 0.00372879

Step2: Selection of individuals

- Manual selection
- Truncation selection based on:
  - ✓ Molecular score (MS)
  - ✓ Weighted MS: give more or less importance to the different QTL
  - ✓ Utility criterion: select candidates based on the possibility of obtaining superior genotype in their progeny (favor heterozygous)
- QTL complementation selection

(Hospital et al., 2000)

Step2: Selection of individuals – QTL complementation selection (QCS) *(Hospital et al., 2000)*

List4_Complementation_Selection

	Id	P1	P2	Cycle	Group	MS	Weight	UC	No.(+/+)	No.(-/-)	No.(+/-)	No.(?)	QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7
1	B8	A1005	A1005	C2	-	0.8366	0.7079	10	8	1	0	2	0.0000	0.8598	0.9761	0.8752	0.8959	0.9745	0.9960
2	B158	A251	A1005	C2	G1	0.8024	0.6790	10	7	1	1	2	0.0000	0.8792	0.8925	0.9405	0.9731	0.9693	0.9960
3	B28	A1006	A251	C2	G1	0.7740	0.6550	9,7071	6	1	1	3	0.0000	0.9716	0.4938	0.9492	0.9137	0.9199	0.9797
4	B13	A1006	A1005	C2	-	0.7609	0.6438	9,7071	6	1	2	2	0.0000	0.8844	0.9559	0.9179	0.9731	0.4963	0.9895
5	B38	A1040	A1005	C2	-	0.7494	0.6341	9,366	6	1	1	3	0.0000	0.9349	0.4907	0.9429	0.5822	0.9260	0.9860
6	B37	A1040	A1005	C2	-	0.7433	0.6290	9	7	1	1	2	0.0000	0.9528	0.2334	0.9172	0.8735	0.9718	0.9860
7	B40	A1040	A1040	C2	-	0.7404	0.6265	9,366	7	1	2	1	0.0000	0.9775	0.4376	0.9592	0.4974	0.9231	0.9759
8	B242	A37	A1005	C2	-	0.7305	0.6181	9,366	6	1	1	3	0.0000	0.7259	0.8959	0.8879	0.2896	0.9722	0.9860
9	B124	A23	A167	C2	-	0.7223	0.6852	9,366	6	1	1	3	0.4812	0.6698	0.9559	0.9429	0.2787	0.9614	0.9924
10	B125	A23	A167	C2	-	0.7032	0.6691	8,7071	6	2	1	2	0.4812	0.6536	0.9559	0.9429	0.0058	0.9616	0.9924

Take into account complementarities between candidates regarding the favorable alleles they carry

Prevent the loss of rare favourable alleles and fixation of unfavourable alleles at QTL with small effects.

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

θ_{MS}	≥	þ,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.

Step2: Selection of individuals – Comparison between lists of selected individuals

List2_Truncation_MS_Selection

	Id	P1	P2	Cycle	Group	MS	Weight	UC
1	B8	A1005	A1005	C2	-	0.8366	0.7079	10
2	B158	A251	A1005	C2	G1	0.8024	0.6790	10
3	B28	A1006	A251	C2	G1	0.7740	0.6550	9,7071
4	B13	A1006	A1005	C2	-	0.7609	0.6438	9,7071
5	B38	A1040	A1005	C2	-	0.7494	0.6341	9,366
6	B37	A1040	A1005	C2	-	0.7433	0.6290	9
7	B40	A1040	A1040	C2	-	0.7404	0.6265	9,366
8	B242	A37	A1005	C2	-	0.7305	0.6181	9,366
9	B246	A37	A1040	C2	-	0.7303	0.6180	9,366
10	B293	A9	A1040	C2	-	0.7268	0.6150	9,366

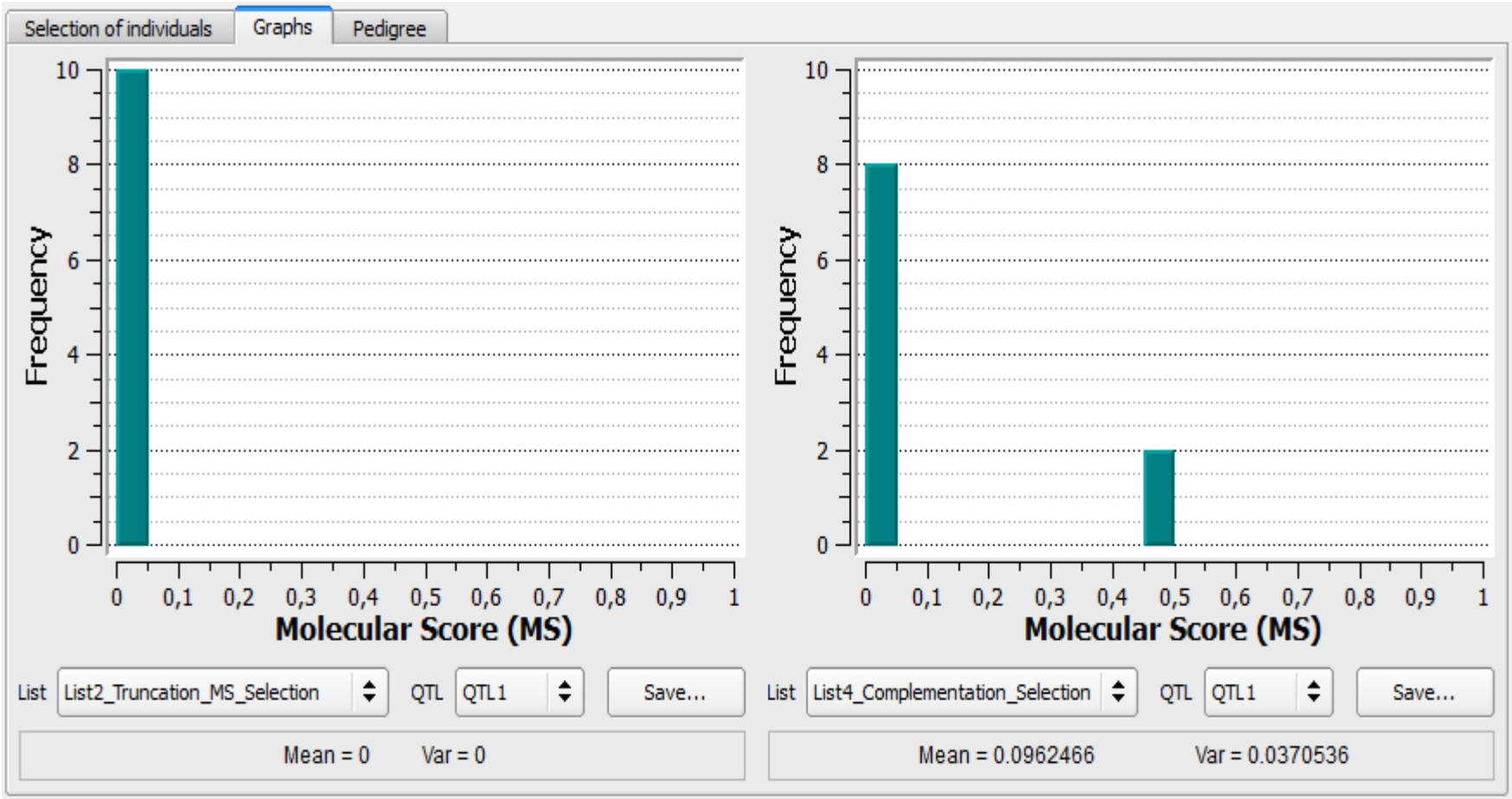
No. ind = 10
Mean = 0.75947
No. group = 1
Var = 0.00115008

List4_Complementation_Selection

	Id	P1	P2	Cycle	Group	MS	Weight	UC
1	B8	A1005	A1005	C2	-	0.8366	0.7079	10
2	B158	A251	A1005	C2	G1	0.8024	0.6790	10
3	B28	A1006	A251	C2	G1	0.7740	0.6550	9,7071
4	B13	A1006	A1005	C2	-	0.7609	0.6438	9,7071
5	B38	A1040	A1005	C2	-	0.7494	0.6341	9,366
6	B37	A1040	A1005	C2	-	0.7433	0.6290	9
7	B40	A1040	A1040	C2	-	0.7404	0.6265	9,366
8	B242	A37	A1005	C2	-	0.7305	0.6181	9,366
9	B124	A23	A167	C2	-	0.7223	0.6852	9,366
10	B125	A23	A167	C2	-	0.7032	0.6691	8,7071

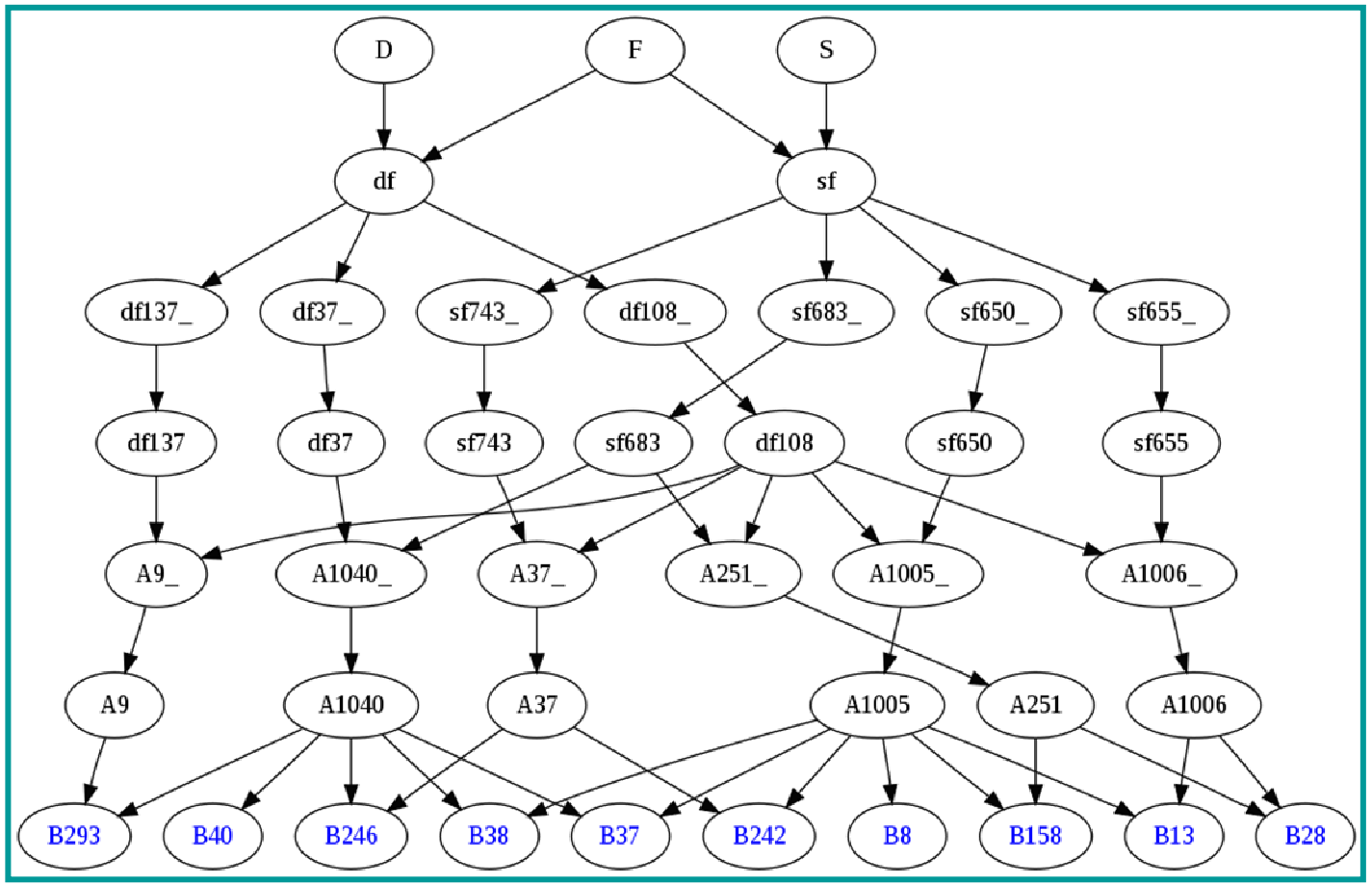
No. ind = 10
Mean = 0.756312
No. group = 1
Var = 0.00140282

Step2: Selection of individuals – Comparison between lists of selected individuals



- Distribution of individuals for each list regarding the molecular score (MS)
- Favorable allele (QTL1) lost without the QTL Complementation Selection (QCS)

Step2: Selection of individuals – Comparison between lists of selected individuals (pedigree)



IL

F1

F2

F4

C1

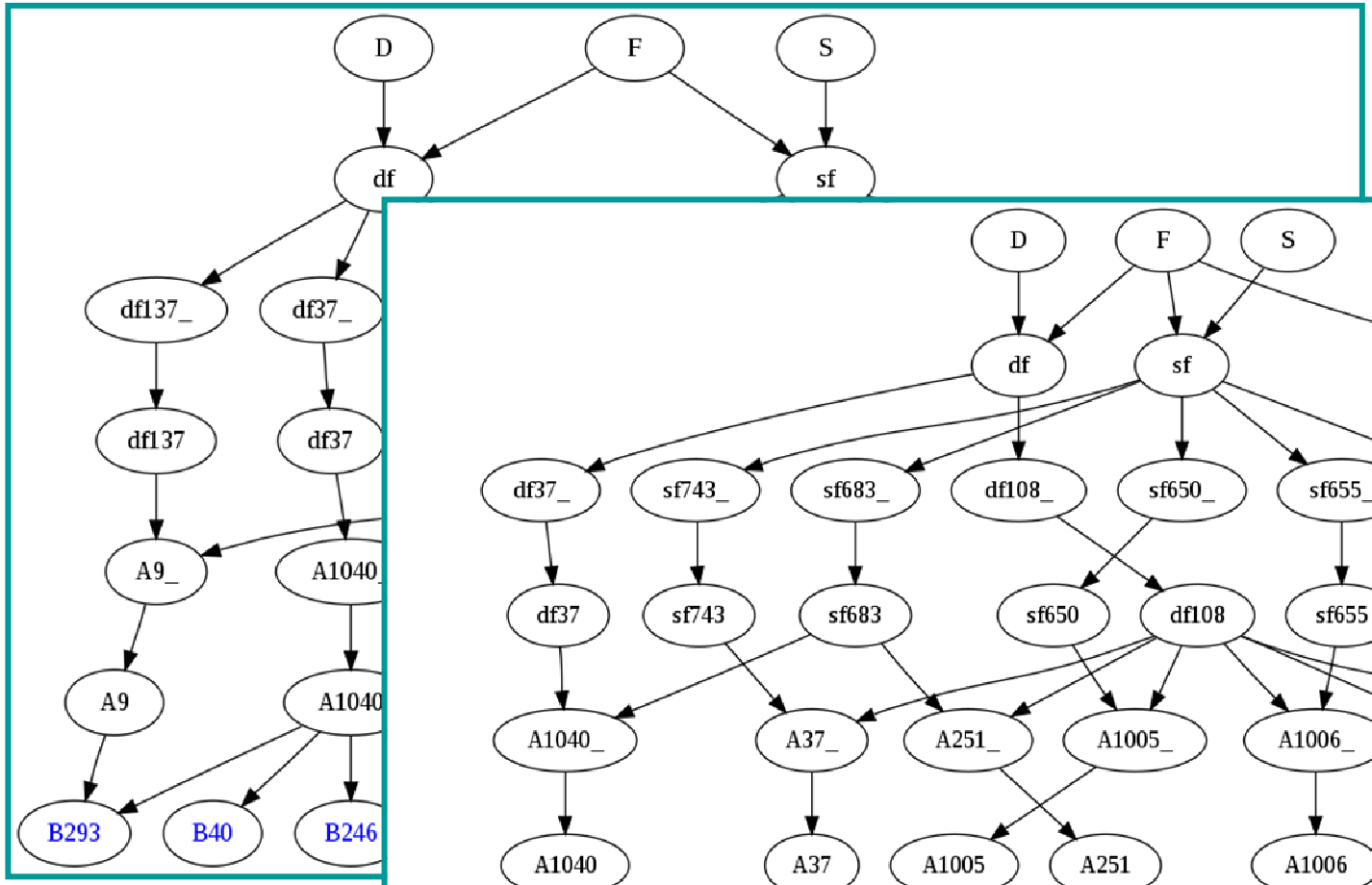
C1 selfed

C2

List2_Truncation_MS_selection

- To follow the contribution of selected individuals over generations and prevent possible bottlenecks

Step2: Selection of individuals – Comparison between lists of selected individuals (pedigree)

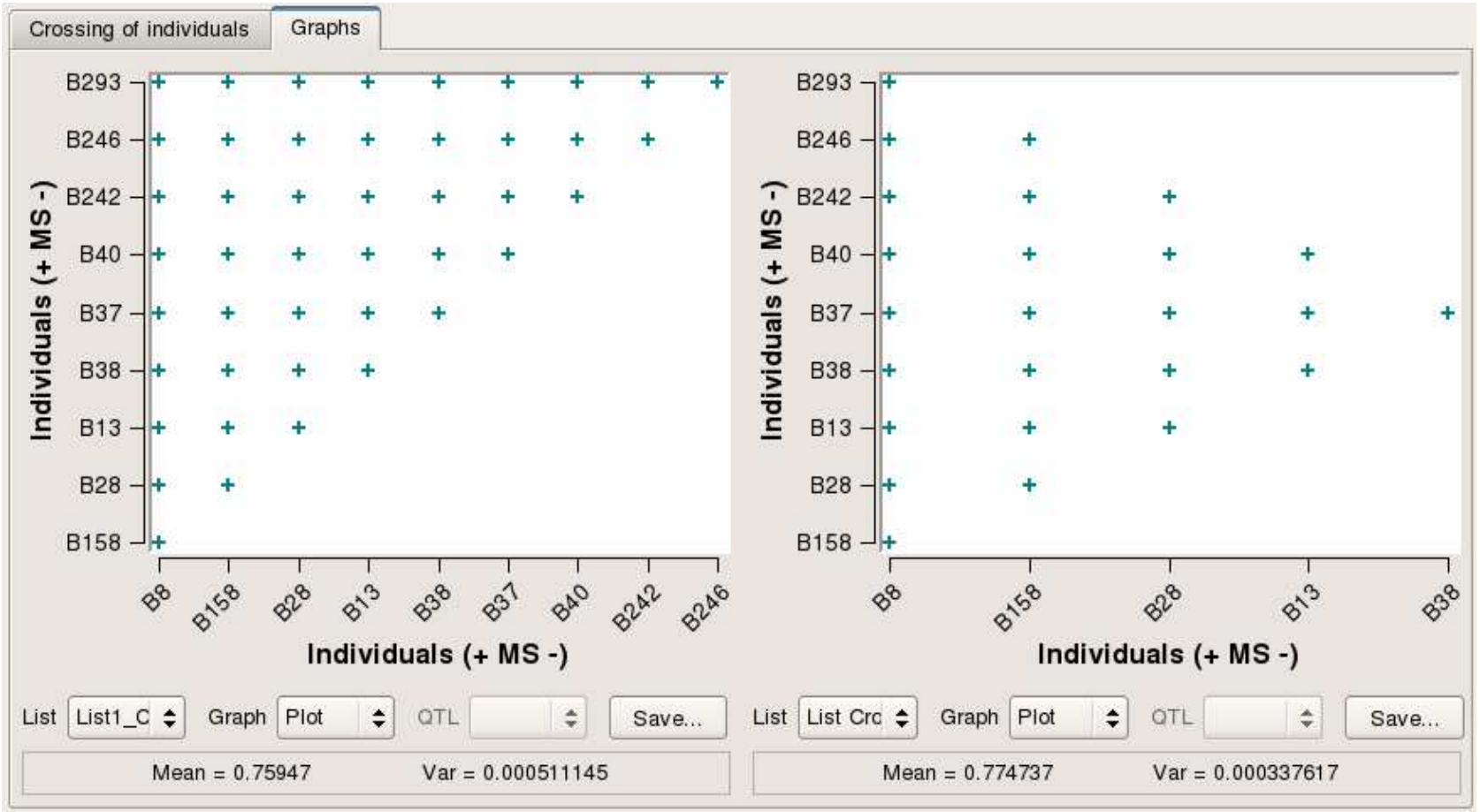


List2_Truncation_MS_

List4_Complementation_selection: 4 parental alleles present in the next generation

Step3: Identification of crosses to be made among selected individuals

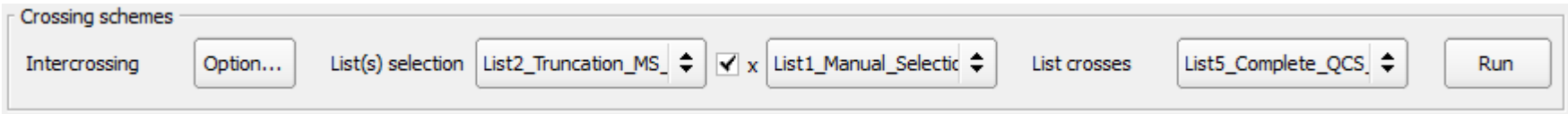
- Half-diallel between selected candidates
- « Better-half » strategy which consists of avoiding crosses between the «worst» selected individuals *(Bernardo et al., 2006)* → optimization of selection intensity



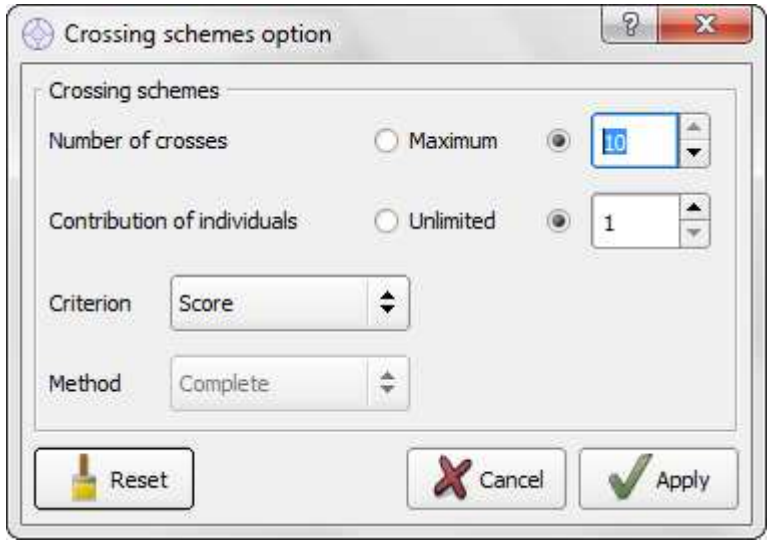
Step3: Identification of crosses to be made among selected individuals

- Half-diallel between selected candidates
- « Better-half » strategy which consists of avoiding crosses between the worst selected individuals *(Bernardo et al., 2006)* → optimization of selection intensity

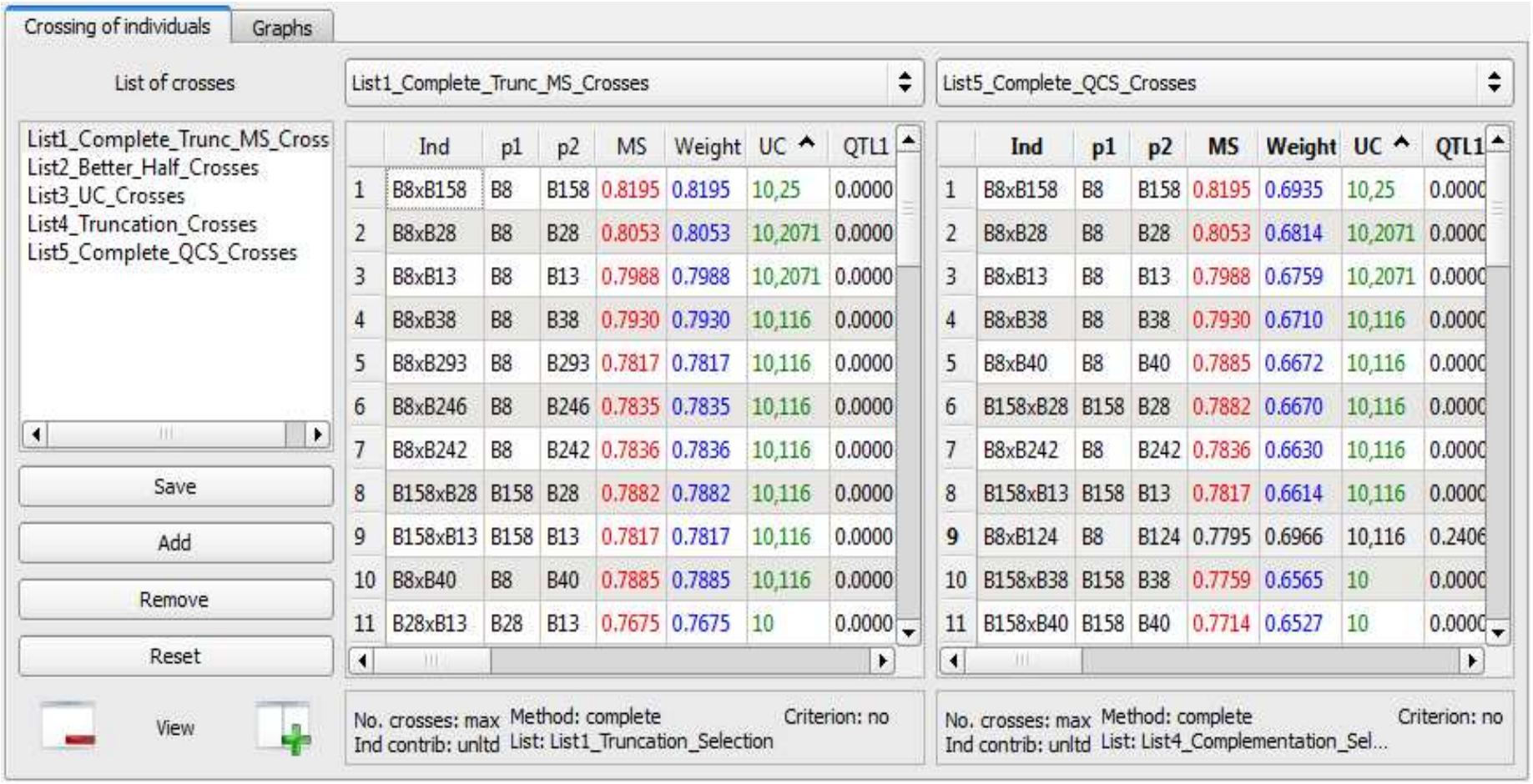
- 2 lists: factorial design



- Constraints:
  - ✓ Maximum number of crosses
  - ✓ Contribution of individuals



Step3: Identification of crosses to be made among selected individuals



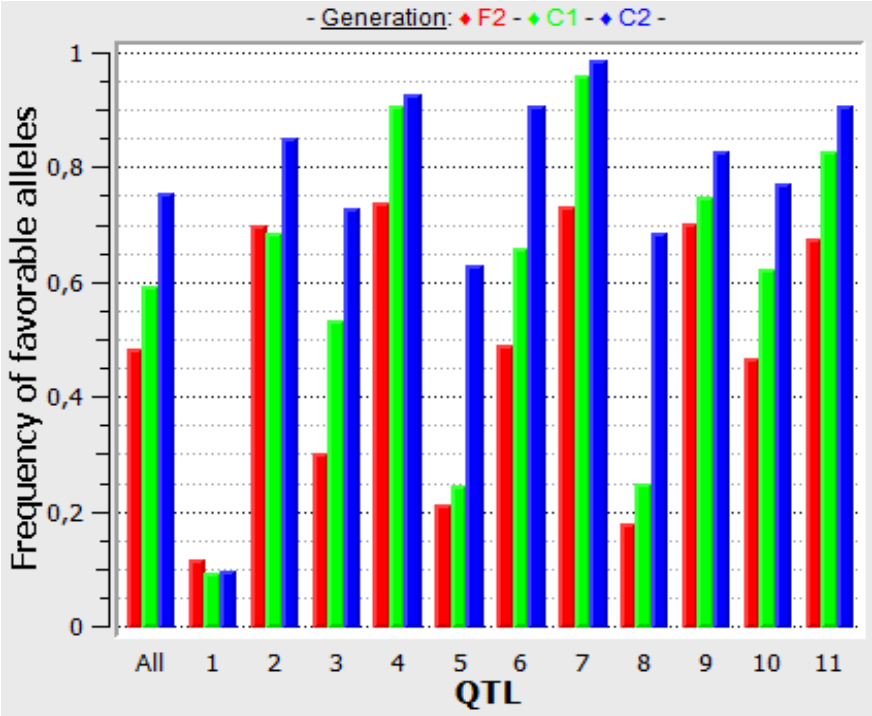
	Ind	p1	p2	MS	Weight	UC	QTL1
1	B8xB158	B8	B158	0.8195	0.8195	10,25	0.0000
2	B8xB28	B8	B28	0.8053	0.8053	10,2071	0.0000
3	B8xB13	B8	B13	0.7988	0.7988	10,2071	0.0000
4	B8xB38	B8	B38	0.7930	0.7930	10,116	0.0000
5	B8xB293	B8	B293	0.7817	0.7817	10,116	0.0000
6	B8xB246	B8	B246	0.7835	0.7835	10,116	0.0000
7	B8xB242	B8	B242	0.7836	0.7836	10,116	0.0000
8	B158xB28	B158	B28	0.7882	0.7882	10,116	0.0000
9	B158xB13	B158	B13	0.7817	0.7817	10,116	0.0000
10	B8xB40	B8	B40	0.7885	0.7885	10,116	0.0000
11	B28xB13	B28	B13	0.7675	0.7675	10	0.0000

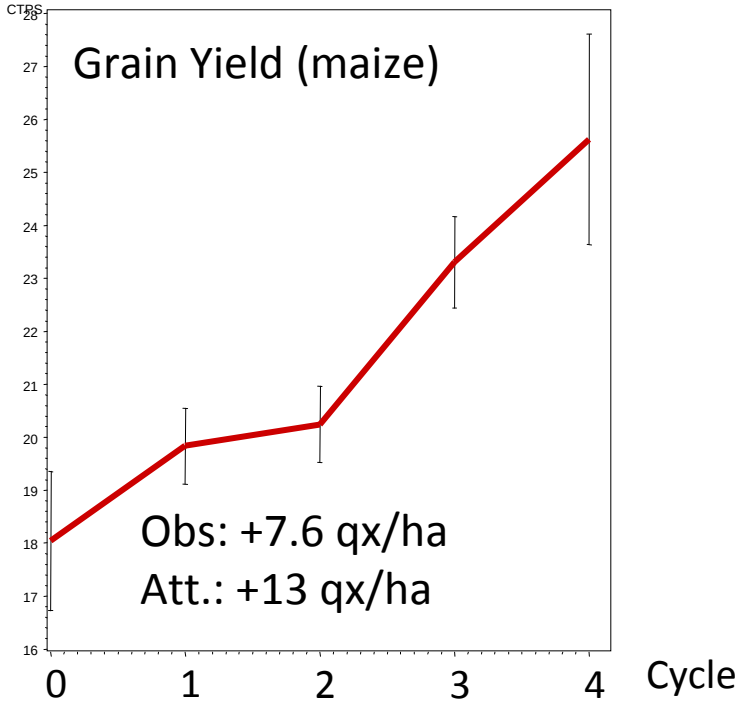
	Ind	p1	p2	MS	Weight	UC	QTL1
1	B8xB158	B8	B158	0.8195	0.6935	10,25	0.0000
2	B8xB28	B8	B28	0.8053	0.6814	10,2071	0.0000
3	B8xB13	B8	B13	0.7988	0.6759	10,2071	0.0000
4	B8xB38	B8	B38	0.7930	0.6710	10,116	0.0000
5	B8xB40	B8	B40	0.7885	0.6672	10,116	0.0000
6	B158xB28	B158	B28	0.7882	0.6670	10,116	0.0000
7	B8xB242	B8	B242	0.7836	0.6630	10,116	0.0000
8	B158xB13	B158	B13	0.7817	0.6614	10,116	0.0000
9	B8xB124	B8	B124	0.7795	0.6966	10,116	0.2406
10	B158xB38	B158	B38	0.7759	0.6565	10	0.0000
11	B158xB40	B158	B40	0.7714	0.6527	10	0.0000

- In each case: computation of the expected MS of the progeny

Conclusion

- OptiMAS tested on a multiparental connected design *(Blanc et al., 2008)*

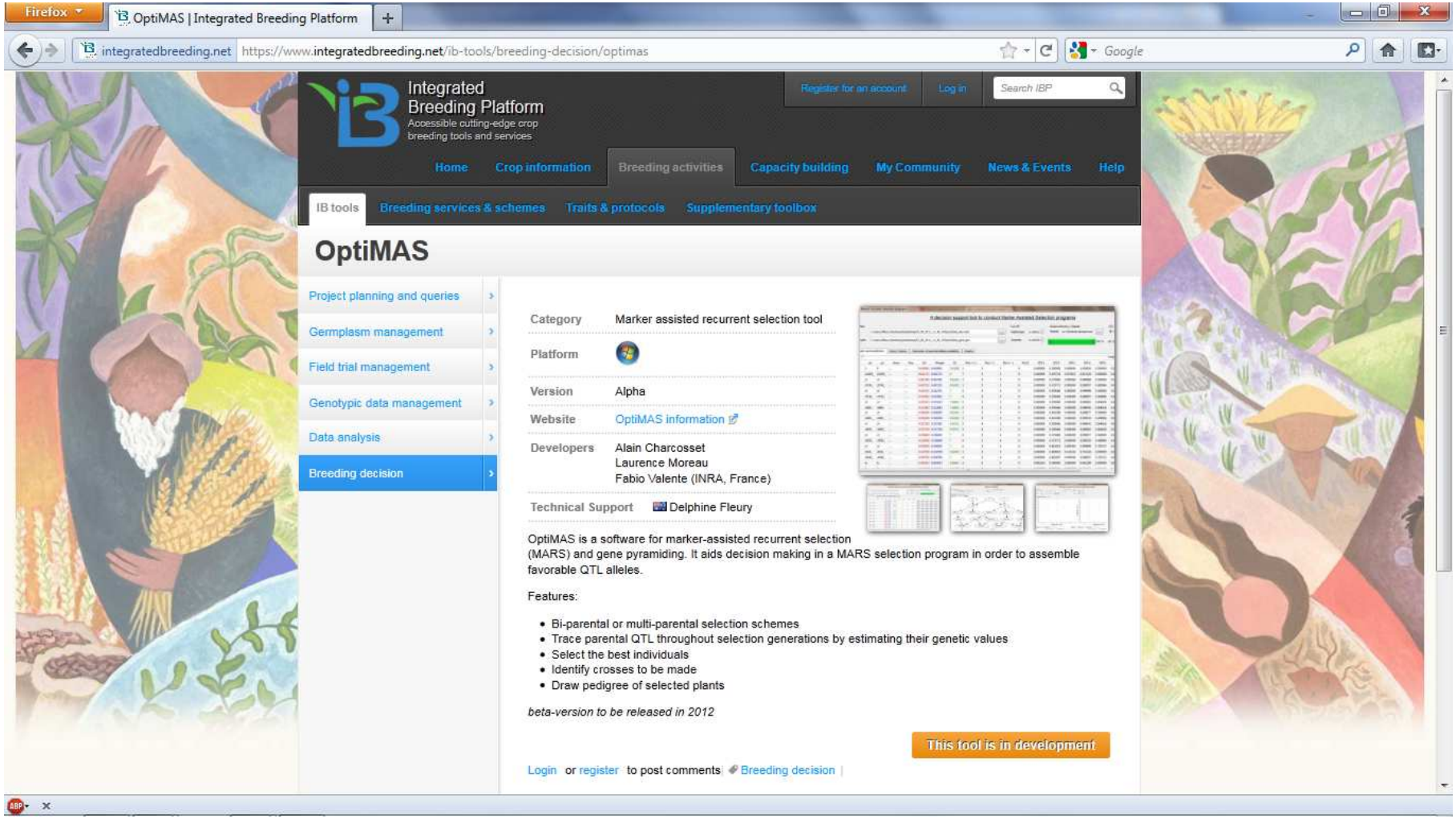




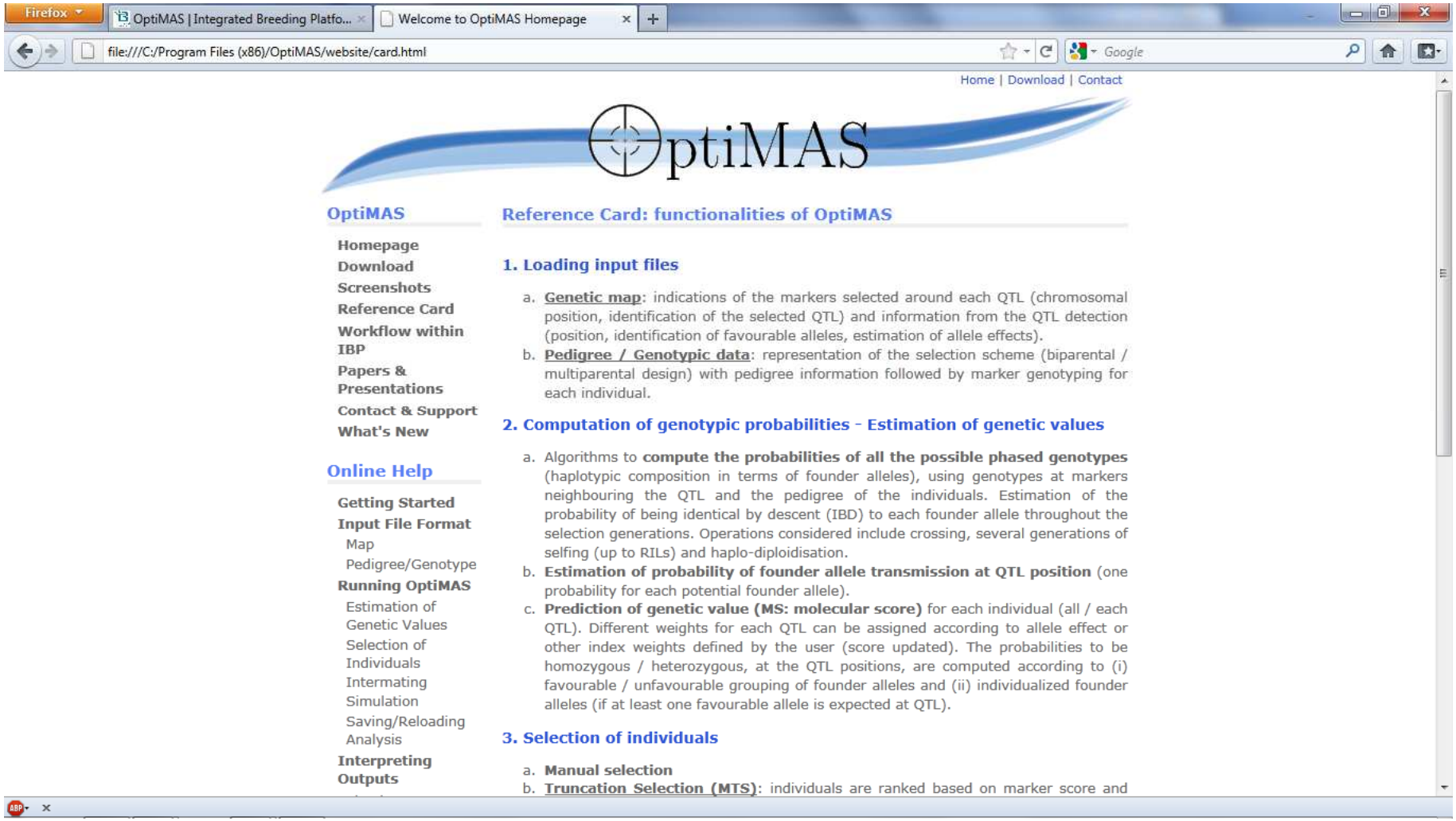
Creation of new genetic material

- Development of 2 version of the tool:
  - ✓ Command line: C-ANSI language
  - ✓ Graphical Interface (GUI): C++ (Qt & Qwt libraries)
- OptiMAS release: available online soon for Windows & Linux then Mac OSX

Integrated Breeding Platform Website



Tutorial/Documentation



Acknowledgments

- Bill & Melinda Gates Foundation
- Guylaine Blanc & J.B Veyrieras for developing the first prototype of the program
- Nicolas Bardol & Delphine Fleury (GCP) for beta testing

Thank you

Questions welcome

Generation
Cultivating Plant Diversity for the Resource-Poor

INRA

LE MOULON
Génétique végétale UMR

Identification of crosses to be made among selected individuals

- Complete : all possible crosses between the selected individuals.
- Better Half : avoid crosses between the last selected genotypes.
  - → Optimization of selection intensity
- Predefined number of crosses to be made based on :
  - ✓ Molecular score.
  - ✓ Utility criterion *[in progress]* : expected mean and variance at the next generation (favor heterozygous…).

▪ Favor couples that will produce the best individuals as possible in the next generation.

$$U_{\text{couple}} = \begin{matrix}\text{Mean of the parental} \\ \text{values of the couple}\end{matrix} + i\,\sigma_{\text{couple}}$$

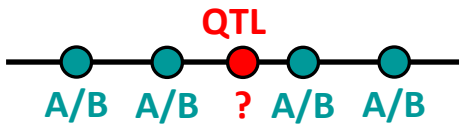
Variance = only heterozygotes genotypes at QTL contribute to variance in the next generation.

Selection of individuals

- Aim: to obtain an ideotype combining all the favorable alleles at QTL positions.
- Prediction of genetic value based on the information from neighboring markers.

⟹ Which parental alleles are transmitted at the QTL position ?

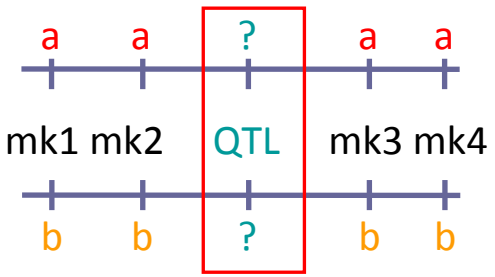
- QTL position rarely located at a marker → QTL alleles are unknown and must be inferred from flanking markers.



e.g 4 markers / QTL

Selection of a subset of informative markers to follow the favourable parental alleles (based on haplotypes).

Identification of all the possible phased genotypes for each QTL and computation of their probabilities



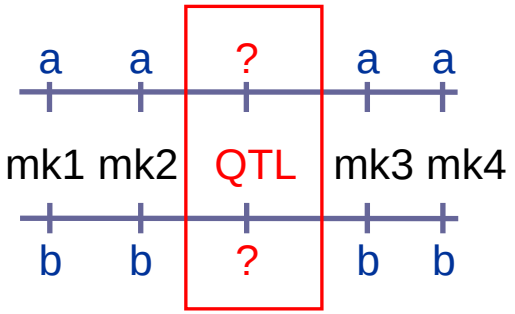
Following « phased genotypes »

- Information available
  - ✓ Pedigree
  - ✓ Distance between loci (Haldane → recombination rate)
  - ✓ Molecular markers (observed genotypes)

⟹ Identification of all the possible phased genotypes for each QTL and computation of their probabilities.

mk1	mk2	QTL	mk3	mk4
A/B	A/B	?	A/B	A/B

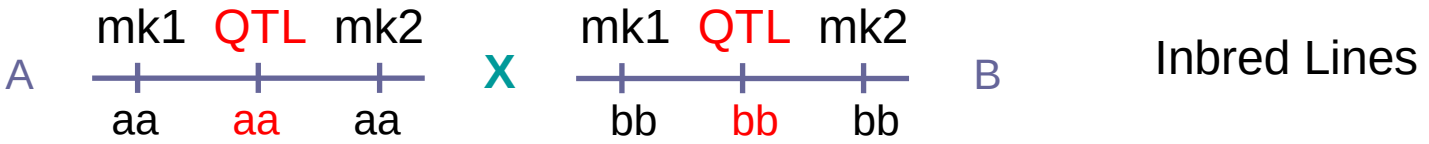
Read / Real



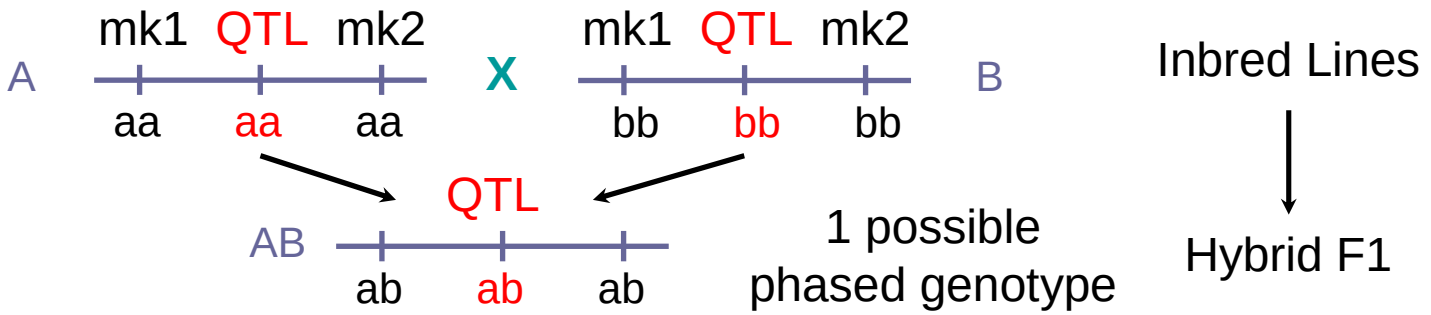
- ✓ Phase (haplotypes) unknown → all possibilities considered

Phased genotype: contains no ambiguity on alleles transmitted from parents and the phase.

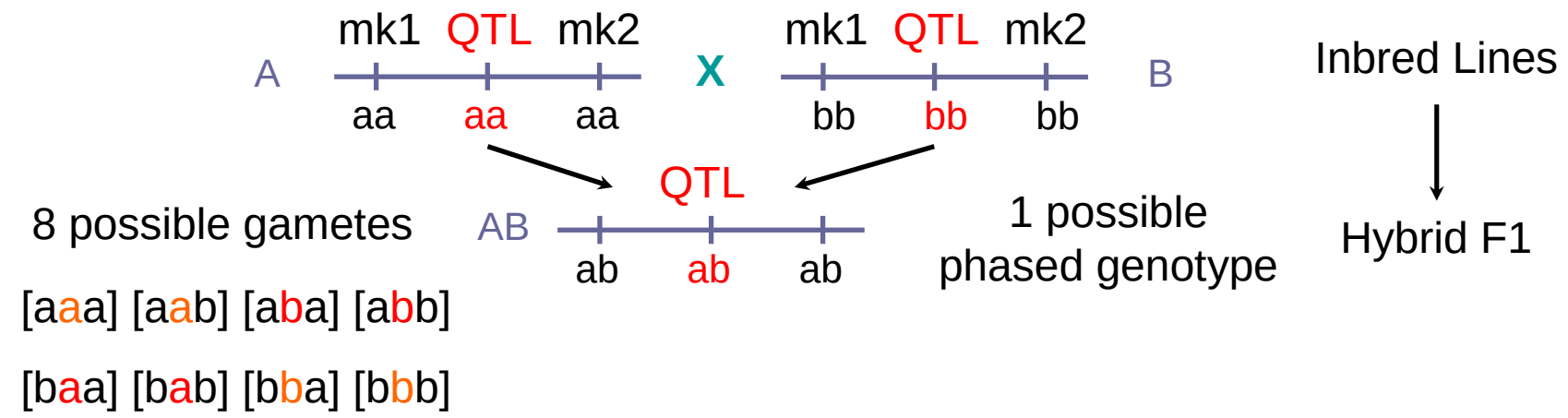
New algorithms: adaptability to different marker selection schemes and mating systems



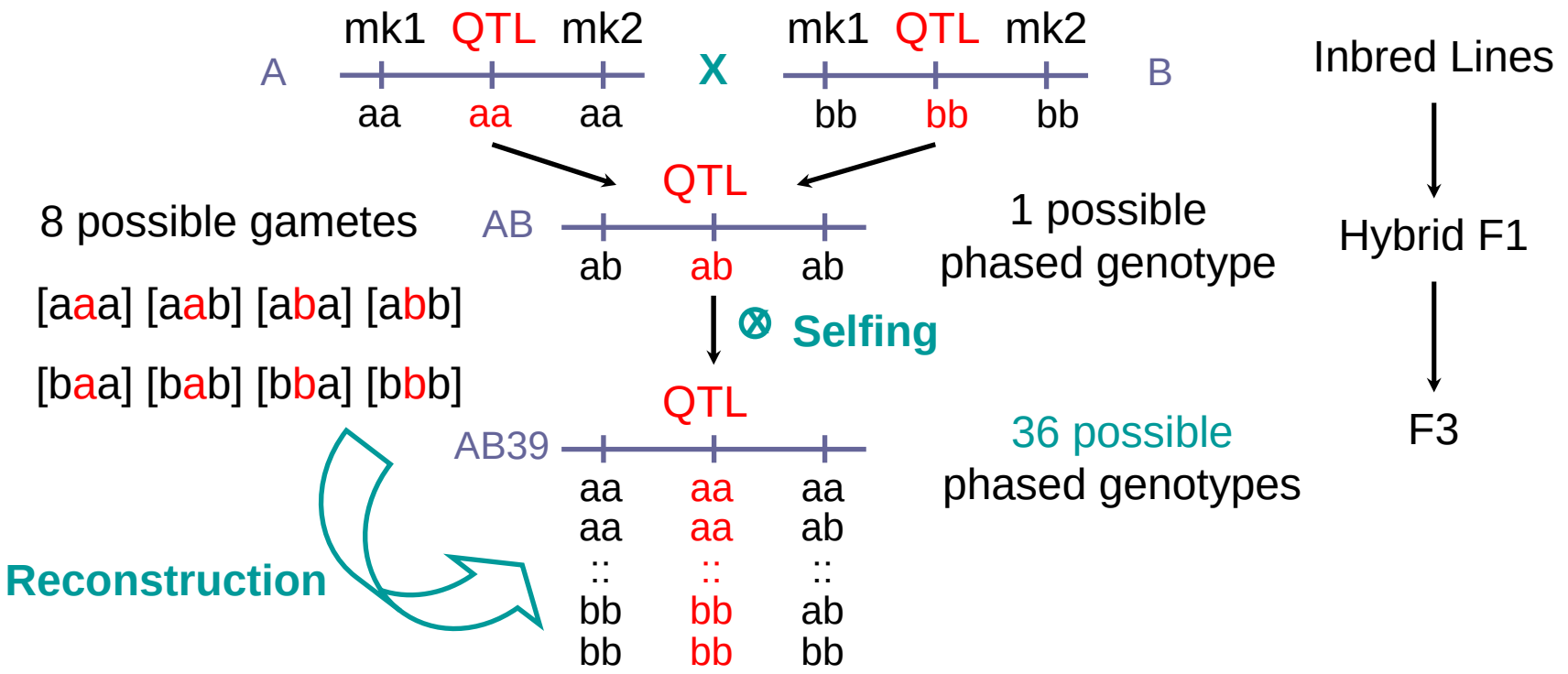
New algorithms: adaptability to different marker selection schemes and mating systems



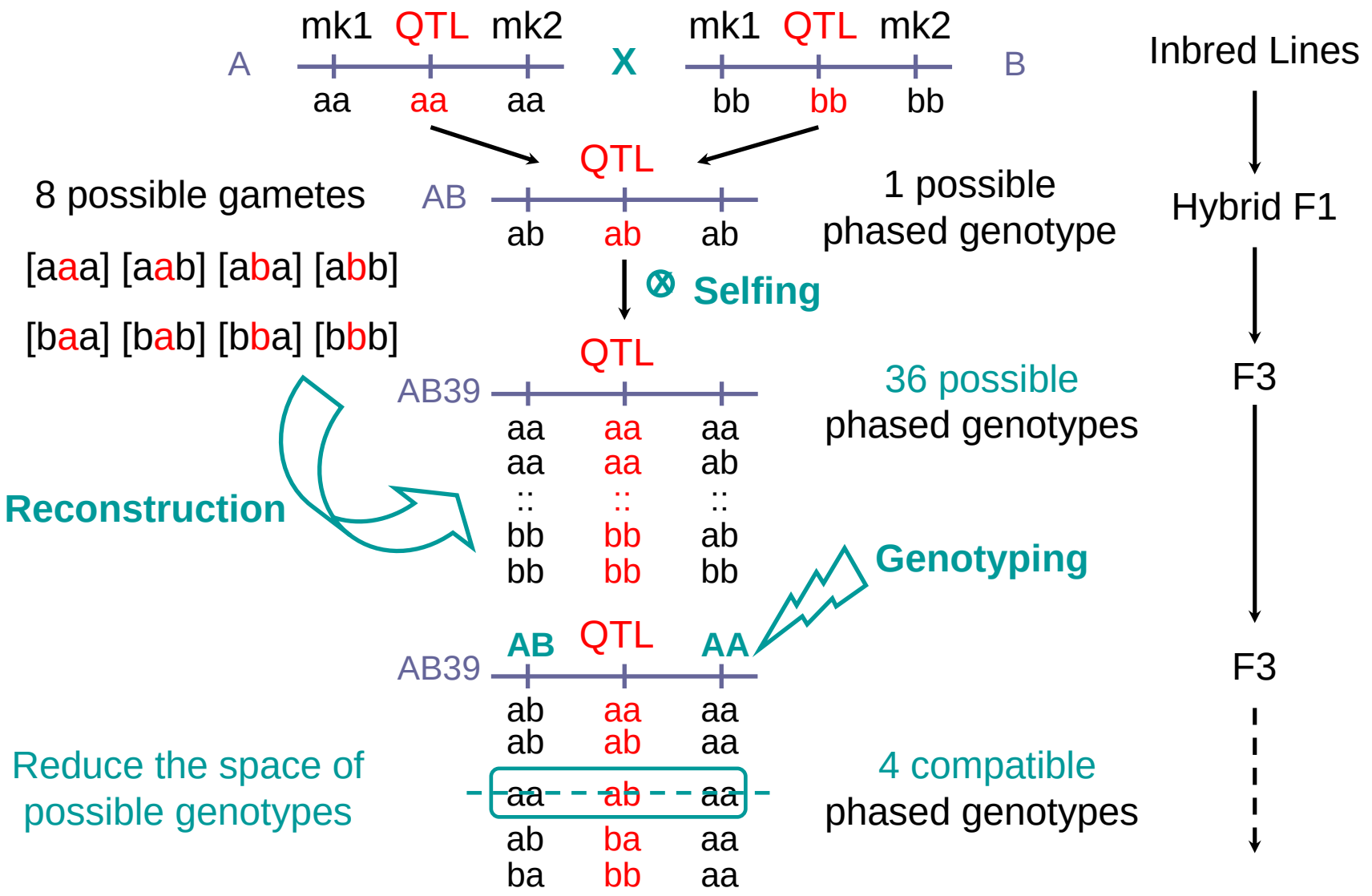
New algorithms: adaptability to different marker selection schemes and mating systems



New algorithms: adaptability to different marker selection schemes and mating systems



New algorithms: adaptability to different marker selection schemes and mating systems



Estimation of genetic values

- _scores: genetic value (molecular score) for each individual

✓ Each QTL: expected dose of favorable allele (0 or 1) for all possible phased genotypes , weighted by the probability of each phased genotype.

$$M_{each} = \frac{\sum \theta_q \cdot p_G}{2}$$

$$M_{all} = \frac{\sum M_{each}}{nb\_qtl}$$

2 classes:
- Favorable
- Not favorable

θ_q : genotype (homo $^{-/+}$, hetero) at QTL position (0, 1, 2)

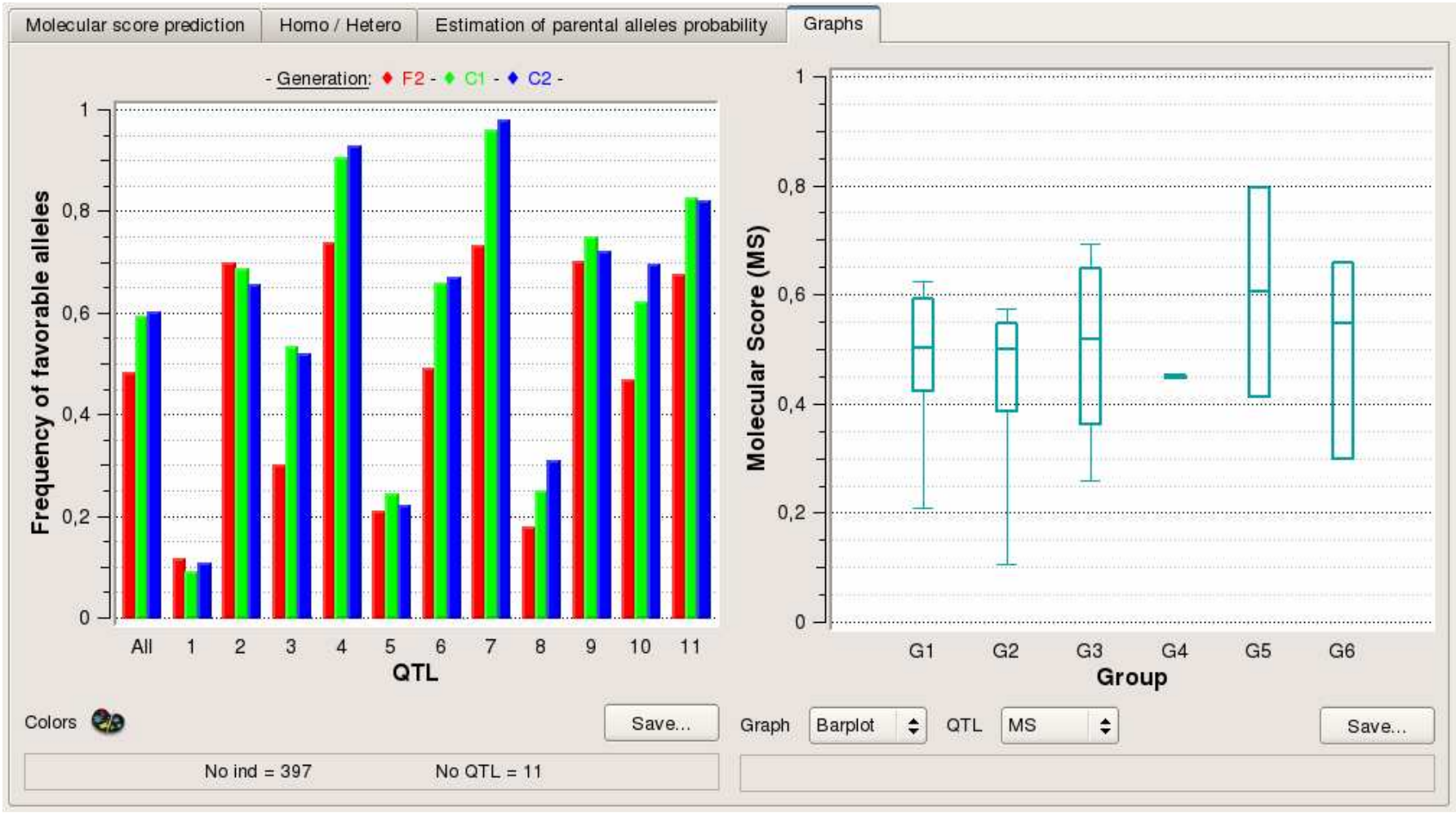
p_G: probability of the current genotype

→ Option: effect associated with parental allele(s) at QTL

→ QTL are considered as being independent (unlinked)

Geno	θ
Homo $^{-}$	0
Hetero	1
Homo $^{+}$	2

Step 1: Computation of genotypic probabilities – Estimation of genetic values (graphs)



- Mean frequency of favorable alleles at all/each QTL through generation of selection
- Intra-group variation (ears)