A kind of wheat MAGIC population

Stéphanie Thépot[1] Ian Mackay[2] Gwendal Restoux[3]
Christine Dillman[1] David Gouache[4] Isabelle Goldringer[1]
Jerome Enjalbert[1]

[1] UMR Génétique végétale, chemin du Moulon, 91190 Gif sur yvette-France
[2] NIAB, Huntingdon Road, Cambridge CB3 0LE - England
[3] Unite d'Ecologie, Systématique et Evolution, Orsay, France
[4] Arvalis, Institut du Végétal. Station Expérimentale, 91720 Boigneville-France

MAGIC Workshop
13/06/2013, Cambridge

Dynamic Management Context

- Complementary method to "static" conservation of genetic diversity

Dynamic Management Context

- Complementary method to "static" conservation of genetic diversity
- Aim to keep adapting diversity to changing environment

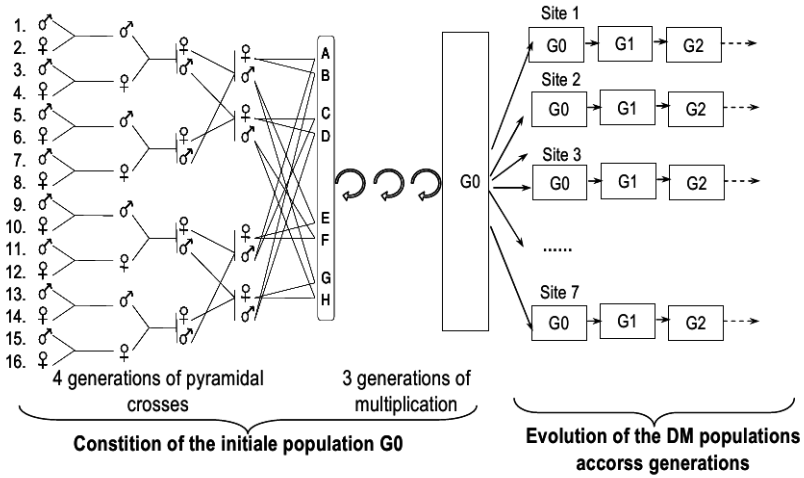
Dynamic Management Context

- Complementary method to "static" conservation of genetic diversity
- Aim to keep adapting diversity to changing environment
- Maintain global diversity in a network of populations

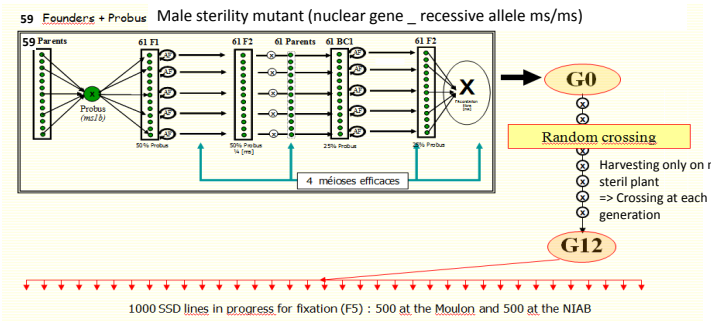
Dynamic Management project

3 gene pools studied since 1974

- 2 selfing pools (PA & PB)
  - based on a pyramidal cross of 16 parents



- One out crossing pool (PS)
  - 59 parents crossed to Probus (ms1b)
  - Harvested only on male-sterile



Dynamic Management project

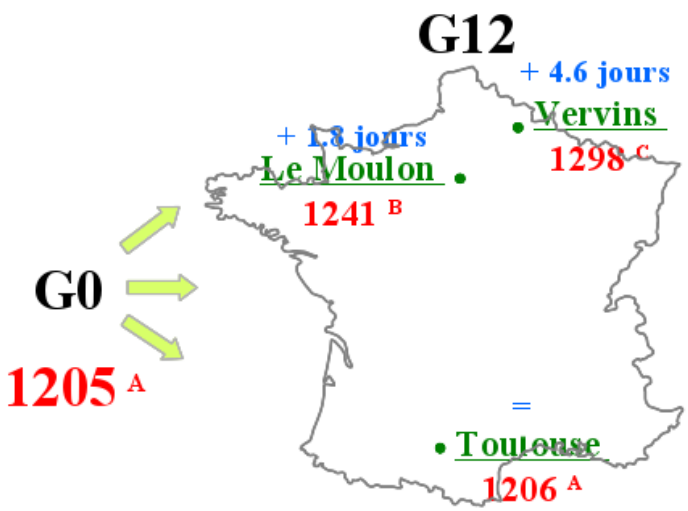
- 3 gene-pools dispatched in 12 locations over France
- Size of population : 5 000 - 10 000 plants
- No migration, isolated plots
- No human selection (random sampling in bulk harvested seeds)

Results found with selfing populations

- High level of genetic diversity maintained in the global network
- Detection of loci associated with local adaptation with association genetics and Fst outliers

Example of the earliness :

- Fast evolution : North-South gradient in earliness

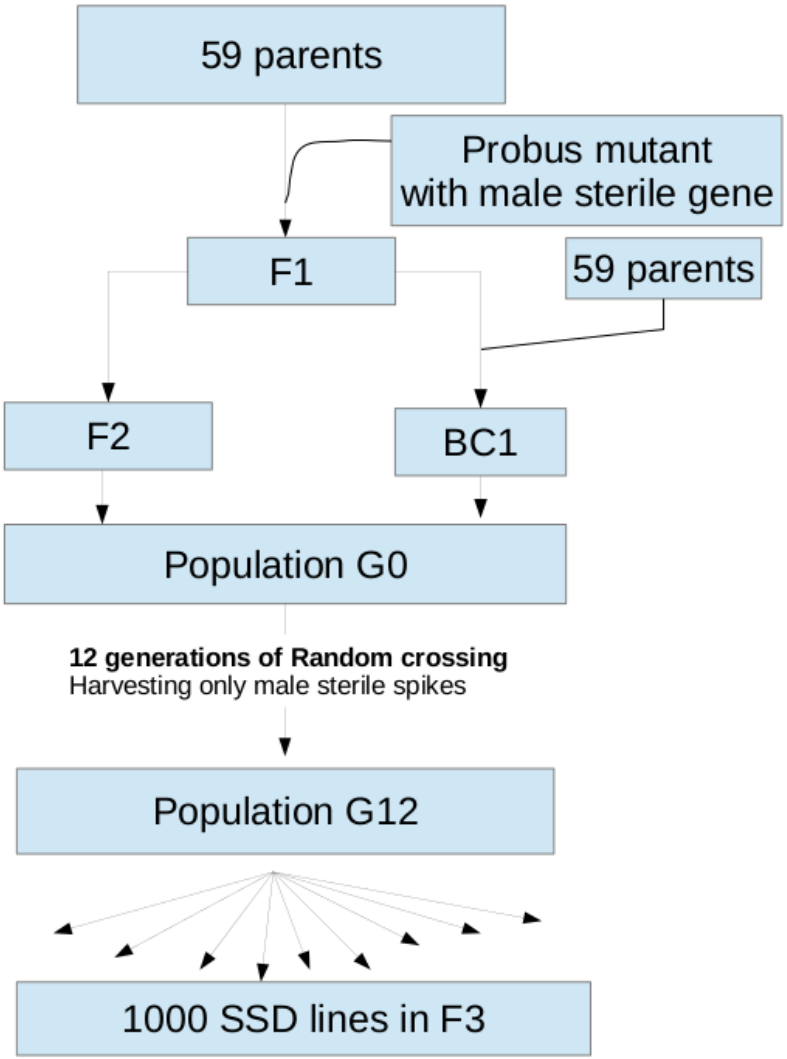


(Rhoné et al. 2010)

- Detection of major genes associated to earliness

Candidate Genes	Temporal variations			Spatial variations		
	Vervins	Le Moulon	Toulouse	Vervins - Le Moulon	Le Moulon - Toulouse	Vervins - Toulouse
VRN-1Apr	*			*		*
VRN-1Aex7				*		
VRN-1Bint1	**		**	**	**	
VRN-1Dint1	NP	NP		NP		
VRN-1ABD	*		*	*	*	
Ppd-1Dpr			**			
Ppd-1Dex8						
Ppd-1D	*	**	**			
FT-Aint1	***	NP	NP	**	NP	**
FT-Aint2a						
FT-Aint2b	*		*			
FT-Dex3						
FT-AD						
LD-Bint11a	**	*	*			
LD-Bint11b						
LD-B	*		*			
CO-B	*			**		*
GI-B		**		*		

Studied population : MAGIC population



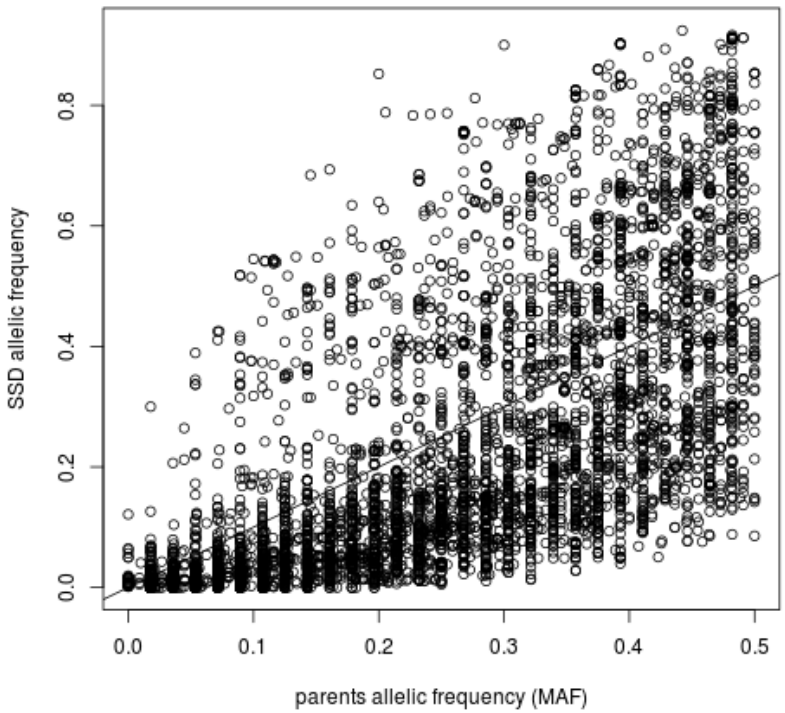
Genetic dataset

- Dataset on 380 SSD lines and 56 parents
- 9k i-select chip (Cavanagh et al. 2013)
  - 8632 SNPs on the array
  - 7270 SNPs of good quality (84%)
  - 6449 SNPs polymorphic (75%)
  - 5621 sets of unique SNP (not in full LD) (65%)
- In addition 36 Kaspar markers localised in candidate genes (height and flowering)

$\longrightarrow$ The dataset is composed of 5657 markers.

Molecular diversity

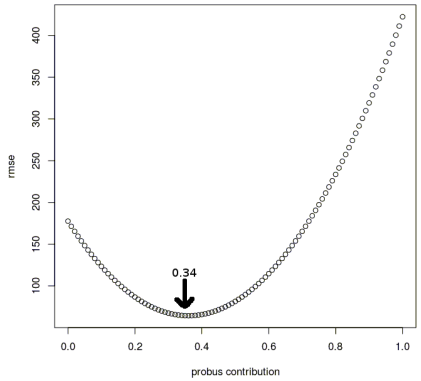
- Slight decrease of polymorphic markers in parental lines (**99.4%**) vs SSD lines (**97.7%**)
- Strong decrease of markers in full LD in parental lines (**21.2%**) vs SSD lines (**5.5%**)
- Allelic frequencies evolution :



- Loss of low frequency alleles (<0.2) in SSD lines (129 markers)
- Low allelic frequencies correlation between parents and progenies
  - due to uncorrected real parental contributions
  - or to drift

Parental contribution estimation

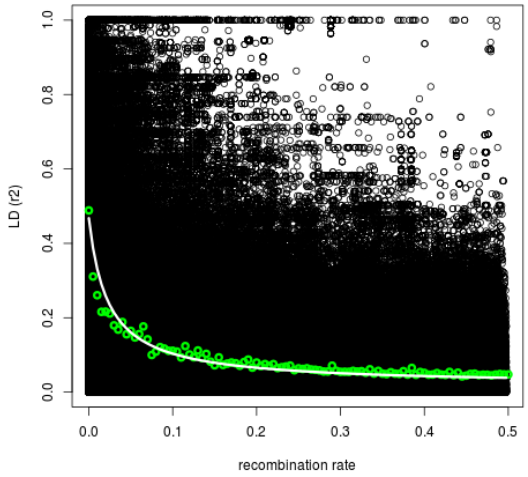
- Unknown real parental contribution due the first step of open pollination
- Hypothetic Probus contribution ranging between 25% and 50%
- 2 estimation methods for parental contributions
  - rough method : Probus contribution estimation by minimizing RMSE, assuming balanced contributions for every other parent



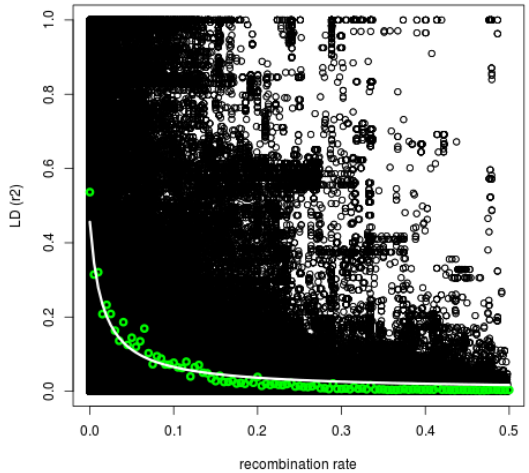
    - Estimated Probus contribution : 34%
  - accurate method : each parent contribution estimation by Bayesian method (still in progress)
- Next results were obtained with corrected Probus contribution

Linkage disequilibrium decay

parental lines



SSD lines



in green : mean local LD ; in white : Hill model fit

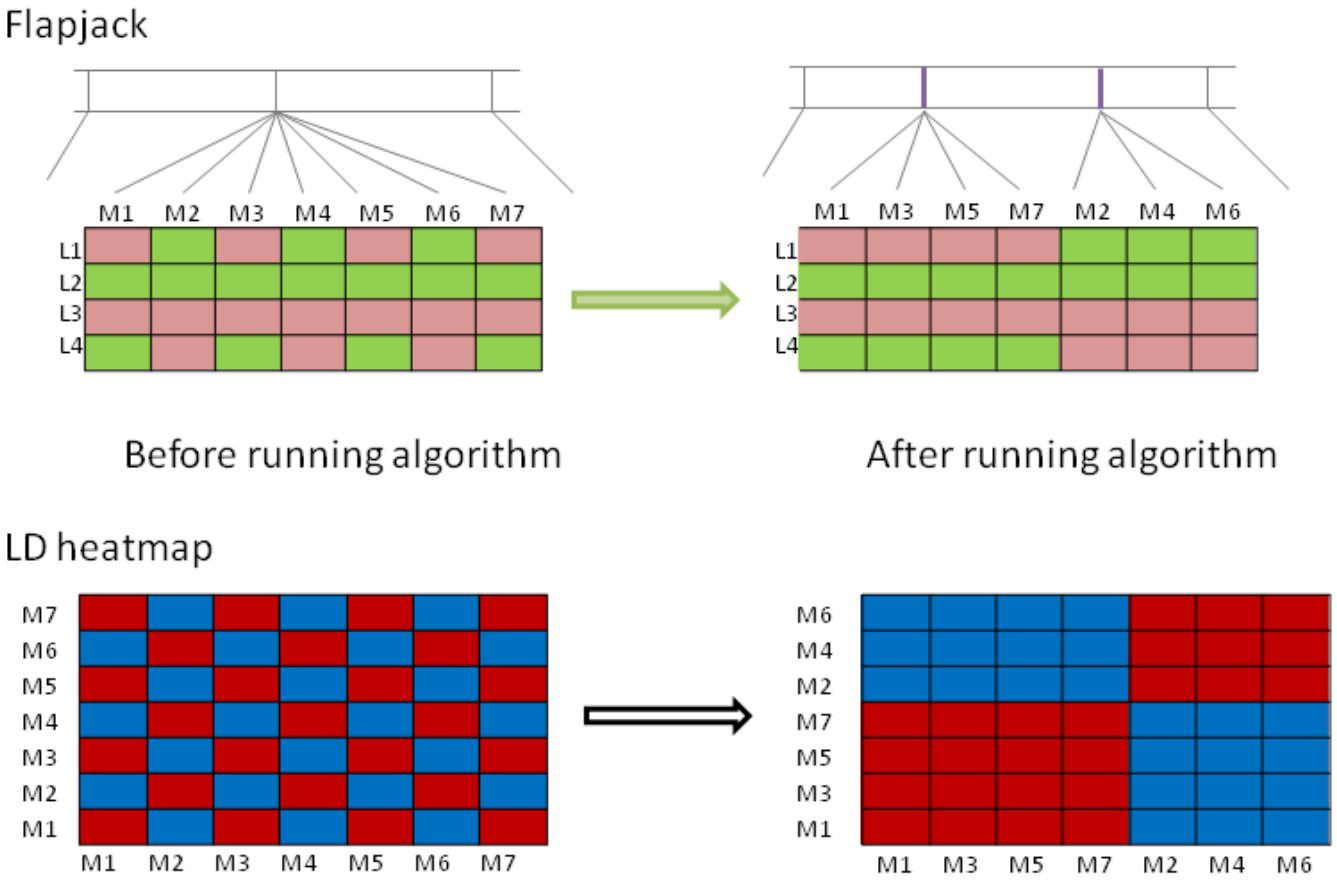
- Fast decay linkage disequilibrium with recombinant rate
- r2<0.2 within 0.04 of recombination rate
- Long distance LD is lower for SSD lines in comparison with parents
- Set of unique markers : 4434 in parental lines vs 5313 in SSD lines
- $\longrightarrow$ The 15 meioses broke linked markers

LD exploitation for mapping

- A genetic map of the 9k i-select array is already available (Cavanagh et al. 2013)
- 6496 markers are localised on 2494 genetic points (in average 2.6 markers on each genetic point)
- in our population :
- 6449 markers could be localised on potentially 5313 genetic points i.e. sets of unique markers (in average 1.4 markers on each genetic point)

$\longrightarrow$ Can I refine the available map for co-segregating SNPs using our population with low LD ?

LD exploitation for mapping

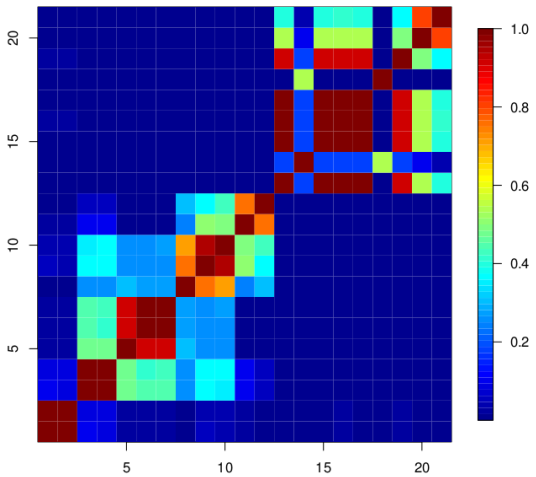


- Algorithm developed reorders markers based on LD

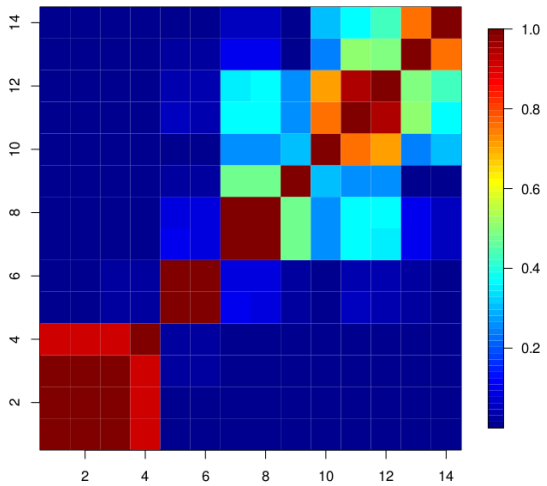
LD exploitation for mapping

Example on a sample of markers on 3B chromosome

LD heatmap with reference Australian genetic map order

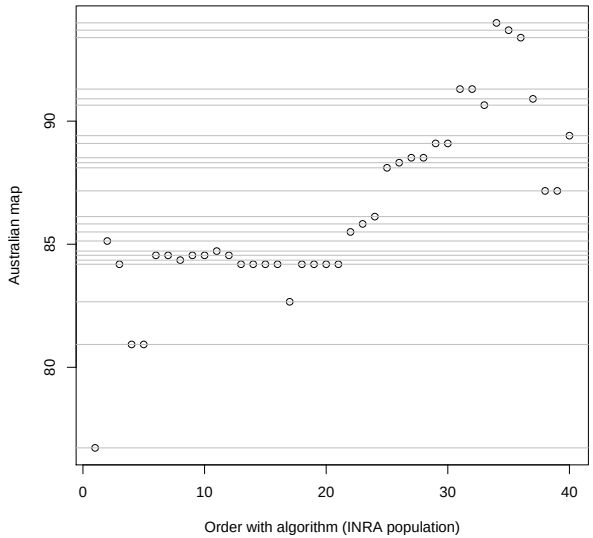


LD heatmap after running algorithm



Algorithm test on markers on 3B chromosome

- Comparison between order found with algorithm and order of the available map

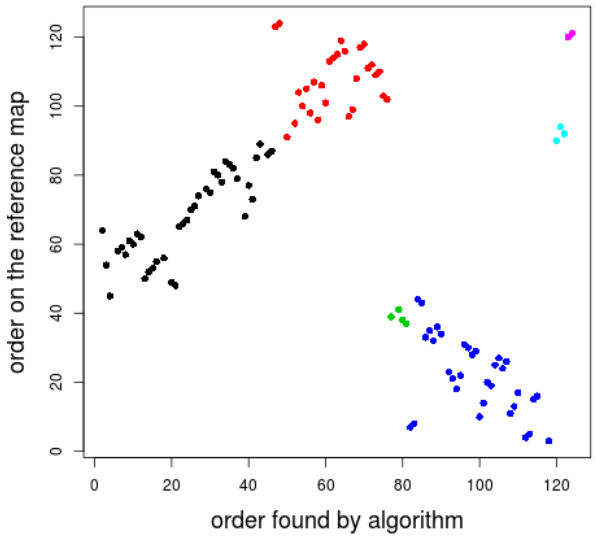


$\longrightarrow$ At fine scale, algorithm ordered markers roughly in consistancy with the avalaible map

$\longrightarrow$ Algorithm order markers localised on the same genetic point

Algorithm test on markers on 3B chromosome

- Comparison at chromosome scale between order found with algorithm and order of the reference australian map

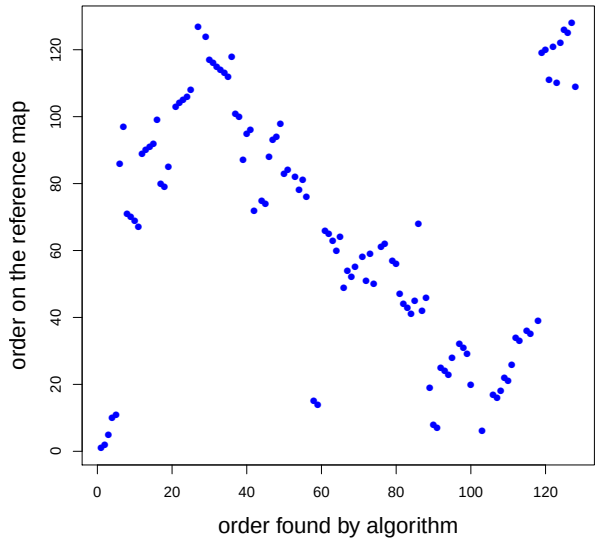


- Several independant linkage groups represented by different colors
- Pretty good global correlation if we order linkage groups
- Problems of correlation at fine scale

$\longrightarrow$ Higher marker density necessary to take all benefits of this LD

Algorithm test on markers on 3B chromosome

- Tests with NIAB MAGIC population



- Stronger LD $\longrightarrow$ one linkage group
- Pretty good global correlation

Phenotyping

Phenotyping realised on 1000 SSD lines and 60 parents

- 2009-2010
  - Moulon

Rep 0 : oct

- 2010-2011
  - Moulon

3 sowing dates, in lines

Rep 1: oct | Rep 2: oct

Rep 3 : mars | Rep 4 : mars

Rep 5 : avril

  - England

1 sowing date, in plot

Rep 14 : oct

- 2011-2012
  - Moulon

3 sowing dates, in lines

Rep 6 : oct | Rep 7: oct

Rep 8 : mars | Rep 9 : mars

Rep 10 : avril

1 sowing date, in lines with controlled light (16h)

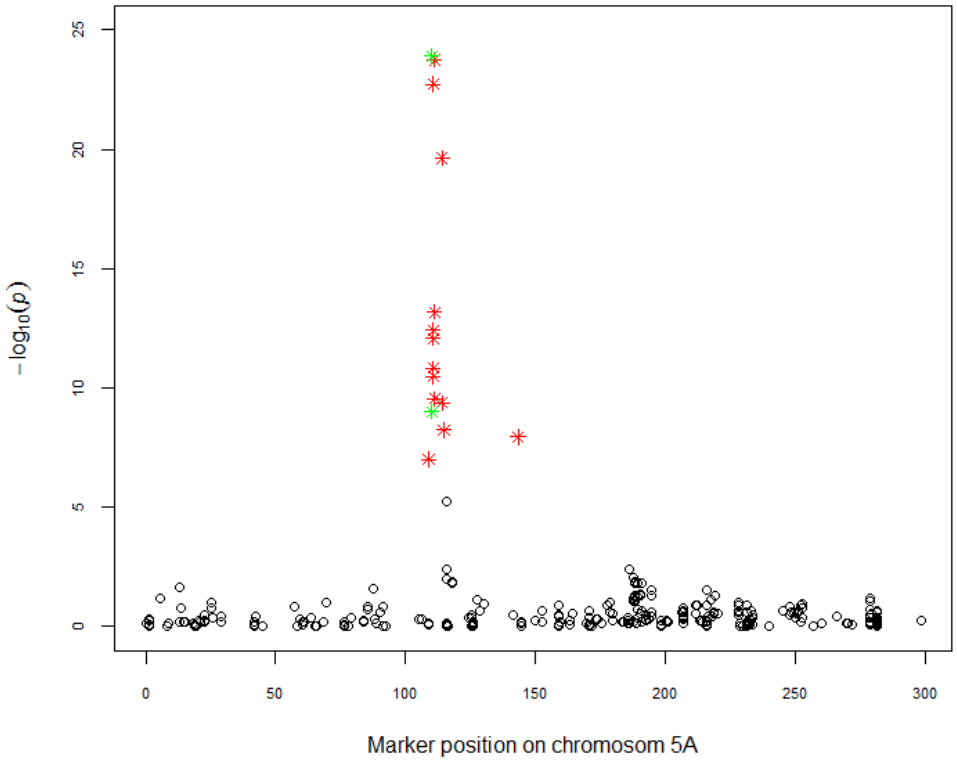
Rep 11 : oct

1 sowing date, in plot

Rep 12 : oct | Rep 13 : oct

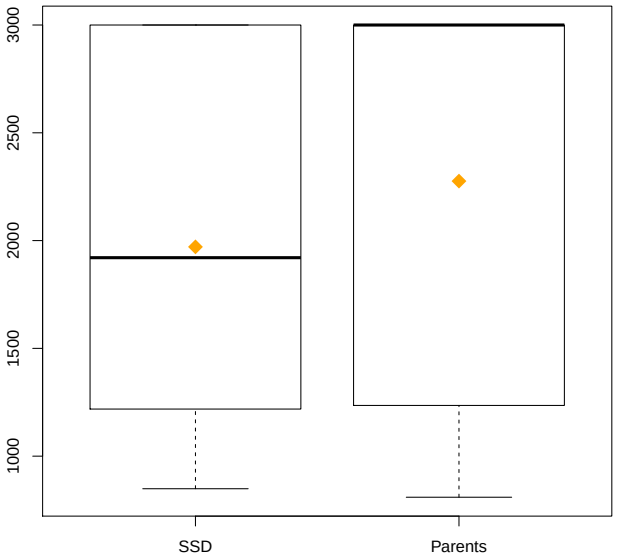
Application : Association genetics

- Illustration with basic association tests done marker by marker
- Association tests performed on heading date score of the April sowing
- Strong associations detected around the location of candidate gene Vrn1A



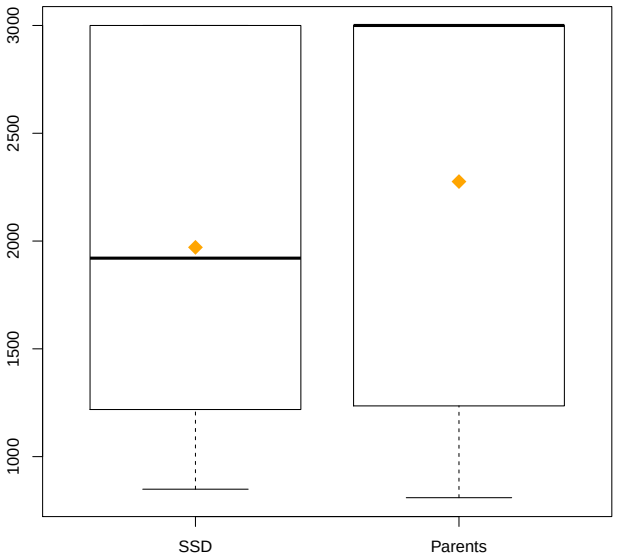
Application : Association genetics

- Significant phenotypic temporal evolution

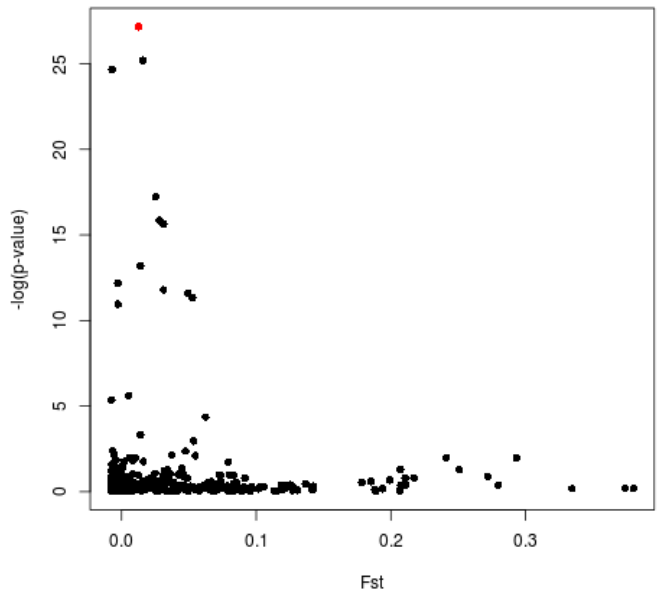


Application : Association genetics

- Significant phenotypic temporal evolution



Correlation between Fst between parents and progenies and significance of association



$\longrightarrow$ Low allelic frequency variation of strong effect allele could involve a strong phenotypic variation ?

$\longrightarrow$ More general question : what is the most adapted method to QTL detection with our population ?

Conclusion & Perspectives

Population with good potential :

- High molecular diversity : 60 parents

Conclusion & Perspectives

Population with good potential :

- High molecular diversity : 60 parents
- Low LD : 15 meioses
  $\longrightarrow$ Trade off between mapping precision and power of de novo QTL discovery

Conclusion & Perspectives

Population with good potential :

- High molecular diversity : 60 parents
- Low LD : 15 meioses
  $\longrightarrow$ Trade off between mapping precision and power of de novo QTL discovery

- Interest for mapping

Conclusion & Perspectives

Population with good potential :

- High molecular diversity : 60 parents
- Low LD : 15 meioses
  $\longrightarrow$ Trade off between mapping precision and power of de novo QTL discovery

- Interest for mapping
  - Validation at fine scale with 3B physical map
  - Refine LD
  - Genotyping with 400k SNPs planned

Conclusion & Perspectives

Population with good potential :

- High molecular diversity : 60 parents
- Low LD : 15 meioses
  $\longrightarrow$ Trade off between mapping precision and power of de novo QTL discovery

- Interest for mapping
  - Validation at fine scale with 3B physical map
  - Refine LD
  - Genotyping with 400k SNPs planned

- Interest for QTL detection

Conclusion & Perspectives

Population with good potential :

- High molecular diversity : 60 parents
- Low LD : 15 meioses
  $\longrightarrow$ Trade off between mapping precision and power of de novo QTL discovery

- Interest for mapping
  - Validation at fine scale with 3B physical map
  - Refine LD
  - Genotyping with 400k SNPs planned

- Interest for QTL detection
  - QTL detection without pedigree : Happy, DSPRqtl
  - QTL detection with multiple environments

Thank you for your attention