The development of a MAGIC population for QTL detection in Maize

P. Dubreuil - BIOGEMMA

The development of a MAGIC population for QTL detection in Maize

Outline of the presentation

- Introduction
- The development of the MAGIC population
- Genetic characteristics of the MAGIC panel
- Phenotypic evaluation
- Association mapping (some examples)
- Conclusion and the future

Introduction

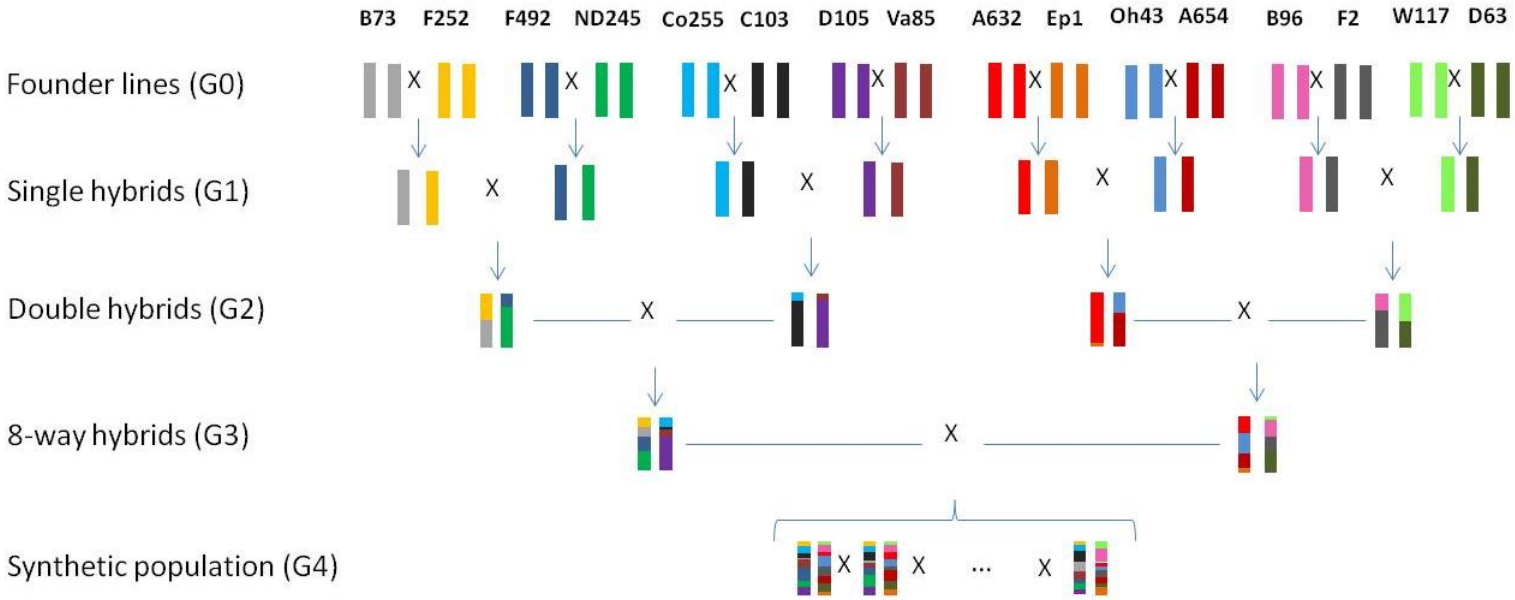
- In the context of the business model of Biogemma, **association mapping** is a method of choice to map QTL at high resolution (as compared to QTLs from linkage mapping approaches with single populations)

- Genetic diversity in Maize highly structured
  - same strong structure in the diversity panels
  - reduced power to detect genomic regions involved in the variation of traits correlated to the structure (*eg.* flowering date,…)

- MAGIC population as an alternative
  - High level of functional diversity (numerous alleles with balanced frequencies)
  - loose structure (broken by accumulating generations of intermating)
  - large panel size
  - combine resolution and power (to some extent)

Development of the MAGIC population

- Started in 2004 !
- Parental lines from very diverse origins



"Funnel" crossing scheme → synthetic population

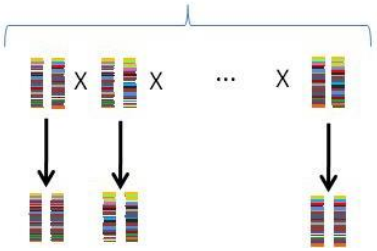
Development of the MAGIC population

3 cycles of intermating (pair-cross scheme) involving ~ 2000 individuals per cycle

Development of the MAGIC population

Intercross #3 (G7)

MAGIC lines (G8)



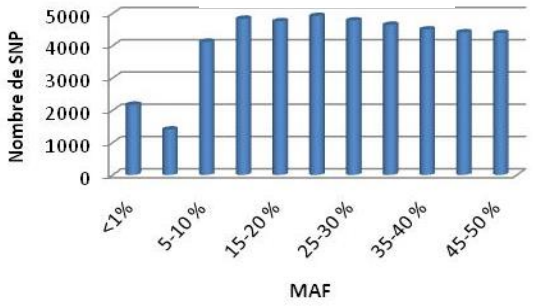
At the 3nd generation, extraction of lines using a double haploid technology

→ we expected 800, we got only 550 ! ☹

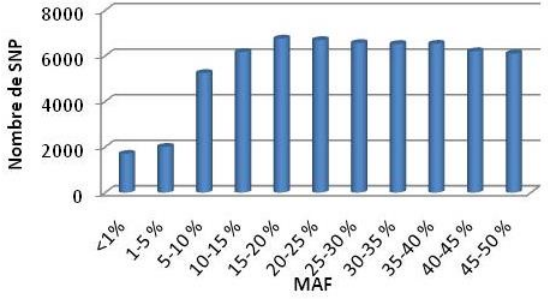
Genetic characteristics of the panel

- Genotyping with the 50K SNP Illumina bead array
  - ✓543 DH lines + 16 parental lines
- Genetic diversity (44.990 SNP effectively used)

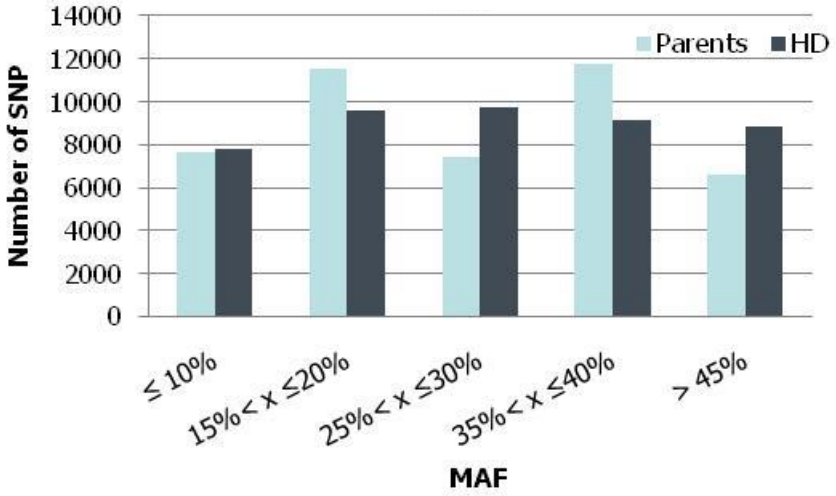
MAGIC panel



Broad-based Diversity panel



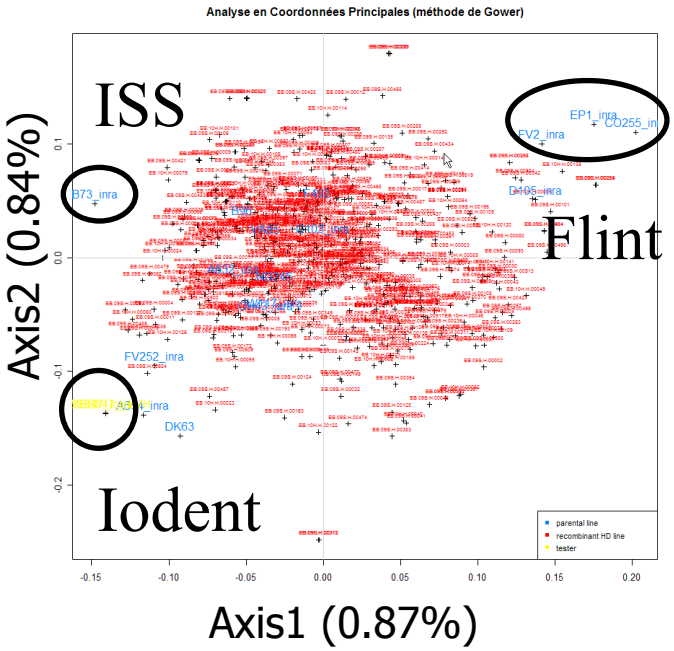
Distribution of the SNP MAF in parental lines and DH lines

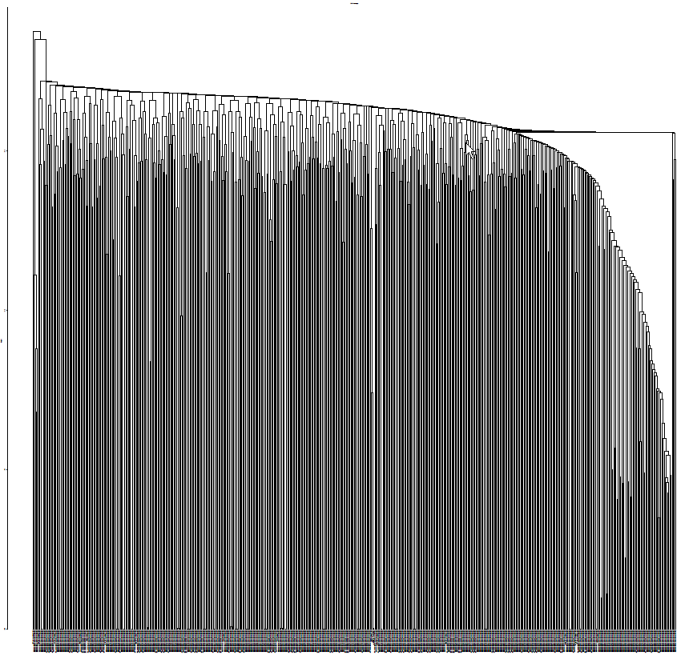


➢ High level of genetic diversity in the panel / limited genetic drift during the development of the panel (high effective population size)

Genetic characteristics of the panel

- Genetic structure
  - ✓ Use of 18218 SNP randomly sampled among those with missing data < 5%
  - ✓ PCoA & UPGMA clustering from squared MRD



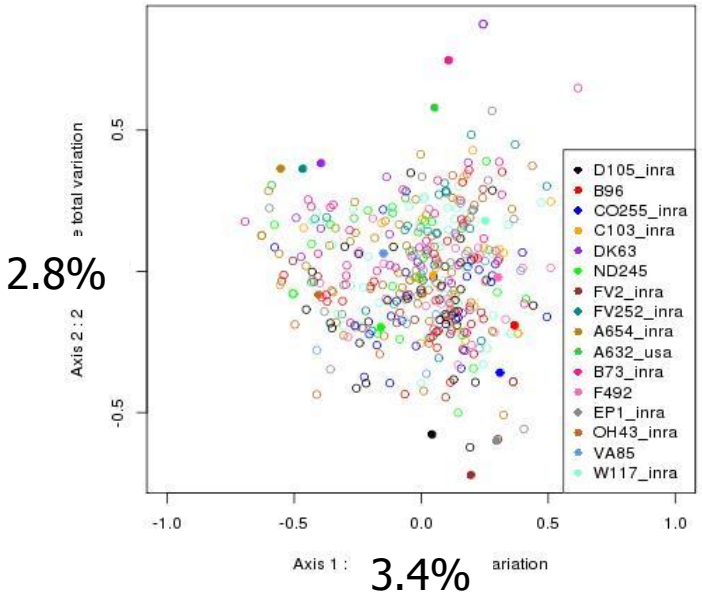


➢ No strong structure in the panel (no visible group, small part of variance explained by PCA axes, chaining effect in the dendrogram)

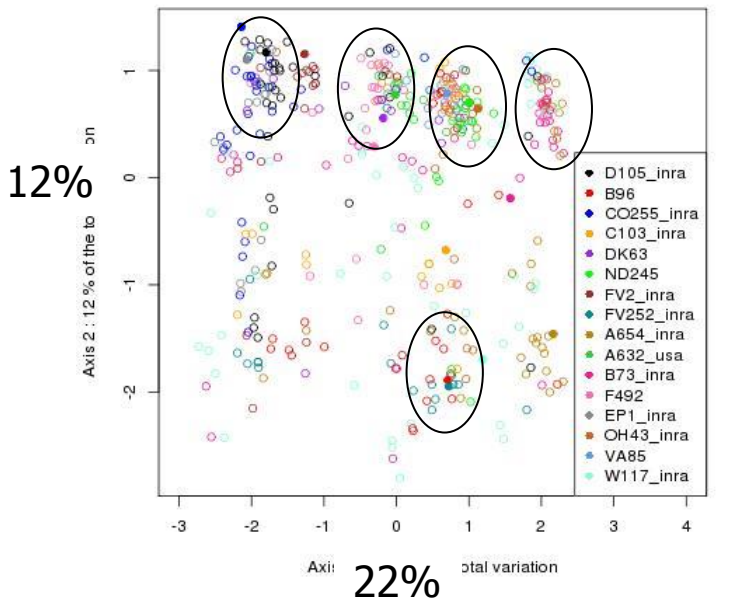
Genetic characteristics of the panel

- Genetic structure (results from R. Rincent, PhD, INRA)
  - ✓ PCoA with markers …

located NOT NEAR the centromere



NEAR the centromere

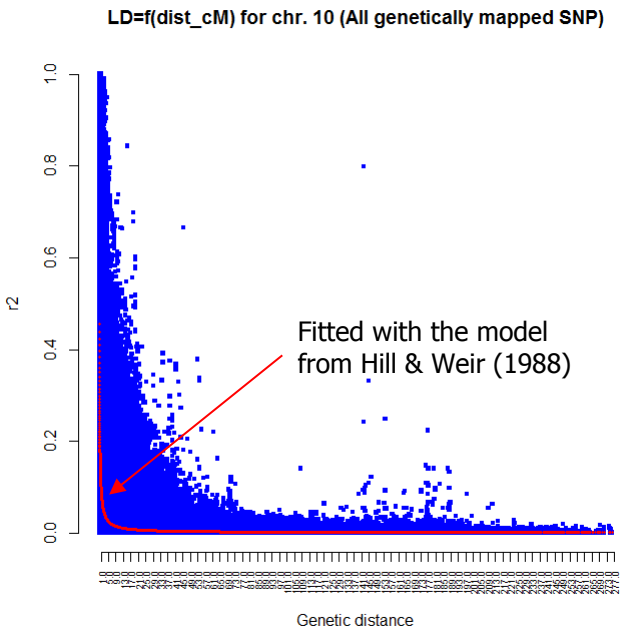


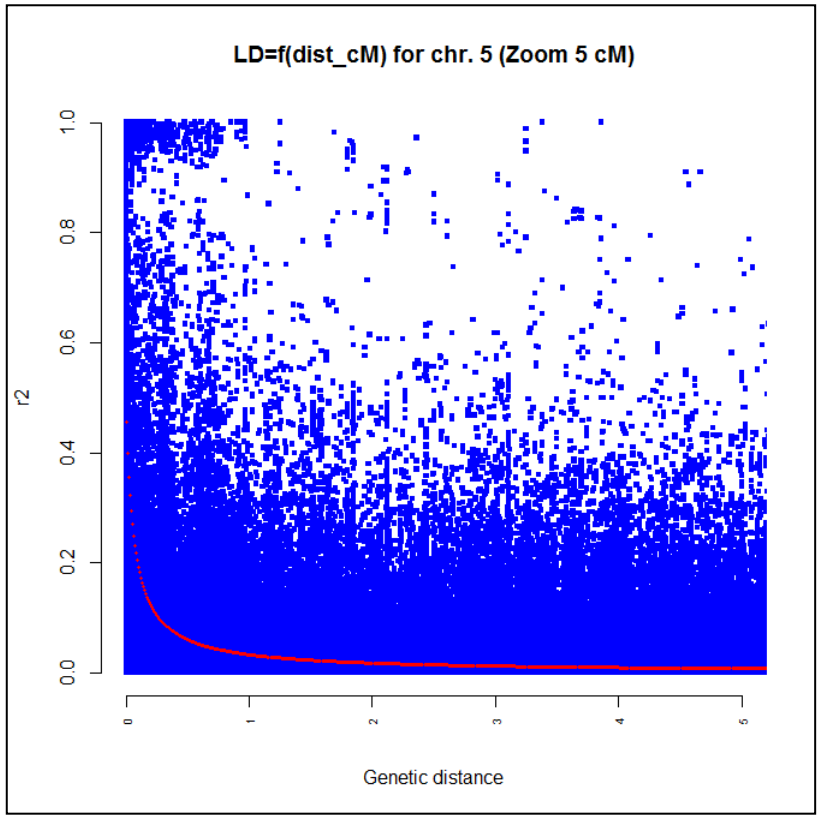
- ➢ Markers located near the centromeric regions reveals a significant structure (high portion of variance explained)
- ➢ Can be explained by low recombination frequencies near the centomeres

Genetic characteristics of the panel

- LD analysis (from SNP loci)

Relationship bw LD and genetic distance



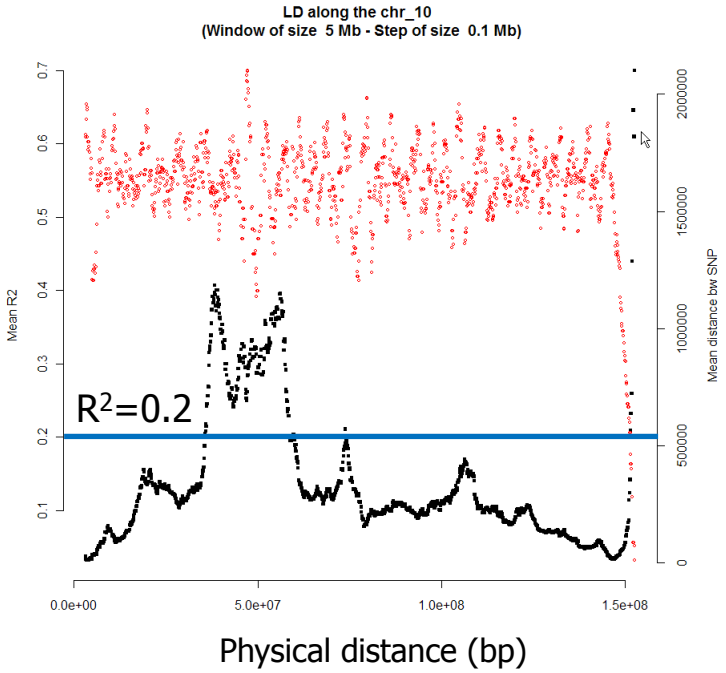


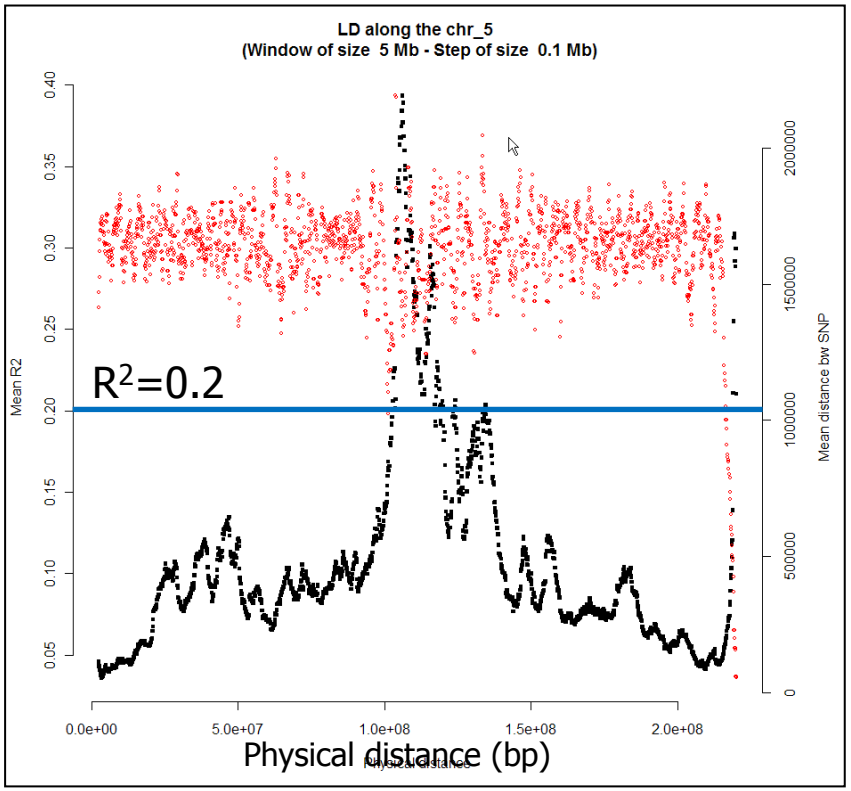
- LD decay (arbitrary threshold of 0.2) ranges between 50 to 300 Kb (conversion used 400-600 kb/cM depending on the chr.) ON AVERAGE
- but highly variable among the regions (high near the centromere)

Genetic characteristics of the panel

- LD analysis (from SNP loci)

Evolution of LD along the chrosome
(from a R script kindly provided to use by S. Nicolas)

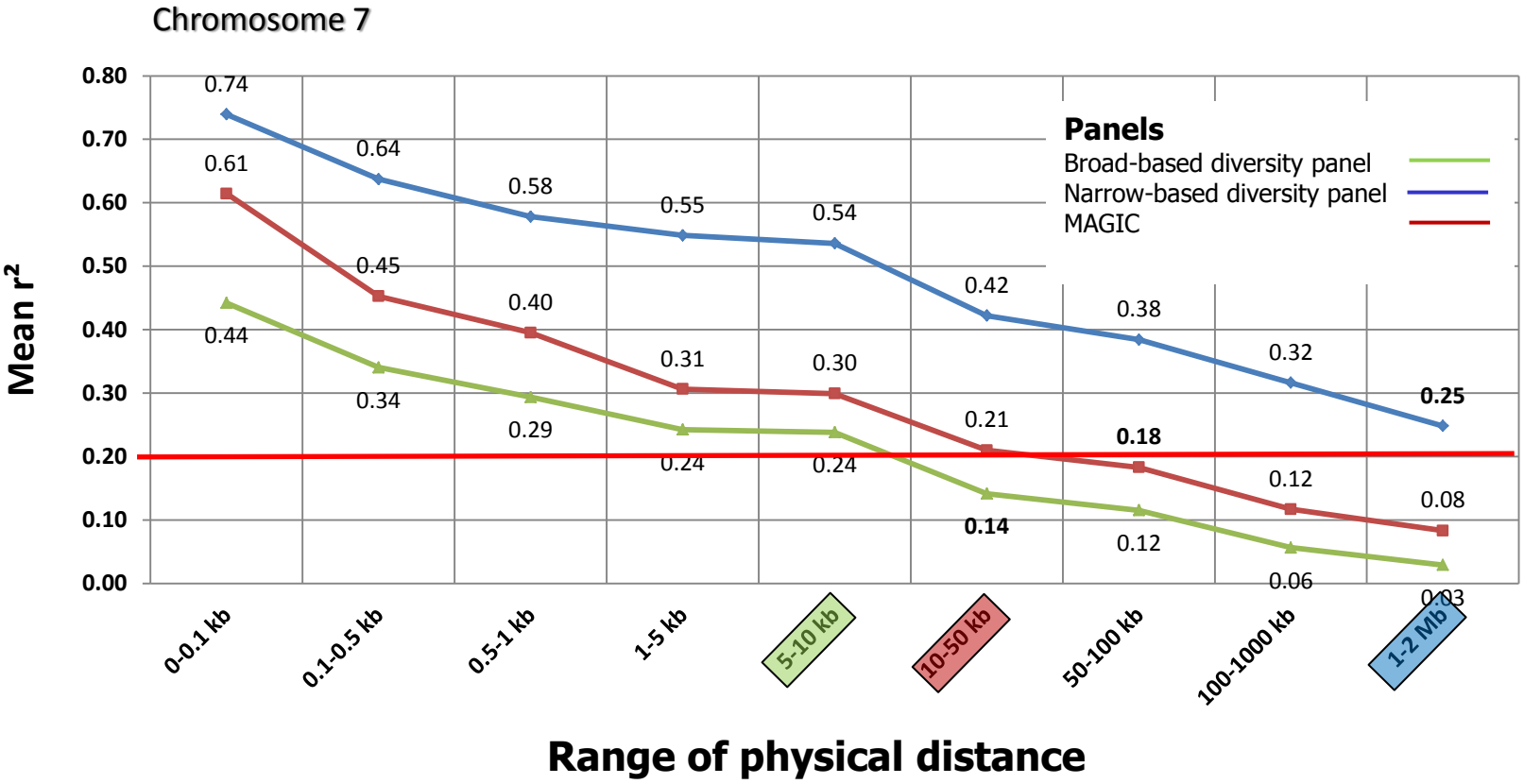




- LD decay (arbitrary threshold of 0.2) ranges between 50 to 300 Kb (conversion used 400-600 kb/cM depending on the chr.) ON AVERAGE
- but highly variable among the regions (high near the centromere)

Genetic characteristics of the panel

- Comparison of LD decay among differents panels



✓ MAGIC panel has an LD decay profil closer to that of the broad-based diversity panel than that of the narrow-base one

✓

Genetic characteristics of the panel

- Inference of parental alleles in the MAGIC lines
  - ✓ **Objectives**
    - Evaluate of the contribution of each parent
    - Draw the graphical genotype of each MAGIC line
    - Estimate of the number of recombinations
    - Run association tests / QTL detection from the parental allele genotypic information
  - ✓ **R/QTL** (K. Broman, Wisc. Univ.)
    - function **readMWril** (modified version with option *bgmagic16* for the *type* argument)
    - function **calc.genoprob** (HMM to calculate the conditional genotype probabilities)
    - 44.378 genetically mapped SNP loci with missing data < 20%

Genetic characteristics of the panel

- Results from R/QTL

Conditional probability matrix
(from calc.genoprob function)

Markers | Par. allele | MAGIC lines

marker	gen	chr	pos	ind1	ind2	ind3	ind4	ind5	ind6
snp41281_1	p1	1	1.587507957	3.95E-05	0.00015524	0.35284758	1.56E-05	1.01E-05	0.94428541
snp41281_1	p2	1	1.587507957	8.79E-08	4.95E-05	0.00020528	6.01E-07	0.00018622	9.42E-05
snp41281_1	p3	1	1.587507957	1.85E-05	0.00010812	0.01359506	6.32E-05	9.58E-06	0.01554314
snp41281_1	p4	1	1.587507957	3.96E-05	1.89E-09	0.00014138	0.00019231	1.30E-07	7.89E-08
snp41281_1	p5	1	1.587507957	8.10E-08	1.81E-05	0.0003278	9.88E-07	0.00094769	9.26E-05
snp41281_1	p6	1	1.587507957	0.00265928	0.00010541	0.07188819	0.02427094	2.55E-05	0.00452123
snp41281_1	p7	1	1.587507957	1.20E-05	0.99891375	0.00618587	1.47E-05	5.75E-06	0.00802465
snp41281_1	p8	1	1.587507957	1.29E-05	0.00010766	0.03589015	1.49E-05	6.12E-05	0.0160405
snp41281_1	p9	1	1.587507957	0.00574921	0.00019809	0.4096118	0.01471704	6.73E-06	0.00454618
snp41281_1	p10	1	1.587507957	1.31E-07	8.09E-06	0.00035257	4.02E-07	0.00163929	0.00012722
snp41281_1	p11	1	1.587507957	8.80E-08	4.27E-06	0.00027371	6.10E-07	0.99248854	0.00065972
snp41281_1	p12	1	1.587507957	8.68E-07	4.31E-06	0.00042094	2.10E-05	0.00461194	0.00018272
snp41281_1	p13	1	1.587507957	1.75E-05	0.00032667	0.04519416	9.74E-06	7.25E-06	0.00567944
snp41281_1	p14	1	1.587507957	0.83549795	2.05E-07	0.02623251	0.02041127	5.89E-09	0.0001749
snp41281_1	p15	1	1.587507957	0.15044333	1.11E-07	0.0281958	0.01418891	9.51E-09	2.50E-05
snp41281_1	p16	1	1.587507957	0.0055089	4.54E-07	0.00863718	0.92607775	6.33E-09	3.02E-06
snp41282_1	p1	1	1.670983684	7.85E-05	0.00015434	0.340983	5.00E-05	3.23E-07	0.9470684
snp41282_1	p2	1	1.670983684	2.39E-08	1.22E-06	6.02E-06	7.34E-08	0.00018498	2.22E-06
snp41282_1	p3	1	1.670983684	4.12E-05	0.00010635	0.0135909	0.00022204	4.16E-07	0.01497827
snp41282_1	p4	1	1.670983684	1.86E-06	2.52E-09	4.17E-06	8.77E-06	7.13E-06	5.73E-08
snp41282_1	p5	1	1.670983684	1.40E-08	4.45E-07	4.40E-05	1.03E-07	0.00096457	2.16E-06
snp41282_1	p6	1	1.670983684	0.00256424	0.00010379	0.07335651	0.02334652	3.43E-06	0.00440145
snp41282_1	p7	1	1.670983684	2.23E-05	0.99899603	0.00613841	5.28E-05	1.81E-07	0.00786464
snp41282_1	p8	1	1.670983684	2.13E-05	0.00010597	0.03656755	3.86E-05	8.13E-06	0.01544384
snp41282_1	p9	1	1.670983684	0.00557649	0.00019474	0.42084397	0.01419323	9.62E-07	0.00437487
snp41282_1	p10	1	1.670983684	2.17E-08	1.97E-07	4.50E-05	3.70E-08	0.00167306	2.95E-06
snp41282_1	p11	1	1.670983684	1.10E-08	1.04E-07	3.53E-05	3.90E-08	0.99245683	1.54E-05
snp41282_1	p12	1	1.670983684	3.01E-08	1.05E-07	5.41E-05	5.34E-07	0.00469896	4.38E-06
snp41282_1	p13	1	1.670983684	3.54E-05	0.00031849	0.04606302	3.68E-05	1.01E-06	0.00547215

Imputation of the most likely parental allele (probability > 0.8; missing data otherwise)

Imputed parental allele matrix

Markers | MAGIC lines

	B	C	D	E	F	G
	SNPname	chr	pos	EB-09S-H-00001	EB-09S-H-00004	EB-09S-H-00017
	PZE-1010004	1	-21.4057249	J	E	N
	SYN9660	1	-10.6848131	J	E	N
	SYN9659	1	-10.5792251	J	E	N
	SYN10558	1	-8.4033442	?	E	N
	PZE-1010010	1	-8.26039494	?	E	N
	PZE-1010010	1	-7.35644555	M	E	N
	PZE-1010011	1	-7.28654709	M	E	N
	SYN10560	1	-7.28621468	M	E	N
	PZE-1010011	1	-7.2824942	M	E	N
	PZE-1010030	1	7.84659736	J	E	N
	PZE-1010030	1	7.86897689	J	E	N
	SYN9369	1	8.6839656	J	E	N
	SYN9381	1	8.8013158	J	E	N
	SYN9380	1	8.80176787	J	E	N
	SYN9363	1	8.80634642	J	E	N
	SYN9377	1	8.83721044	J	E	N
	PZE-1010037	1	12.245822	J	E	J
	SYN11492	1	12.2460713	J	E	J
	SYN11494	1	12.7274463	J	E	J
	SYN15069	1	14.7641102	J	E	J
	SYN15073	1	14.8066174	J	E	J
	PZE-1010049	1	16.2274471	J	E	J
	SYN10467	1	16.2383036	J	E	J
	SYN10469	1	16.4382983	J	E	J
	PZE-1010045	1	16.4849832	J	E	J
	PZE-1010050	1	17.6452328	J	E	J
	PZA00393.1	1	18.1092	J	E	J
	PZE-1010050	1	18.319164	J	E	J
	PUT-163a-29	1	18.6332113	J	E	J
	SYN5934	1	19.3426051	J	E	J

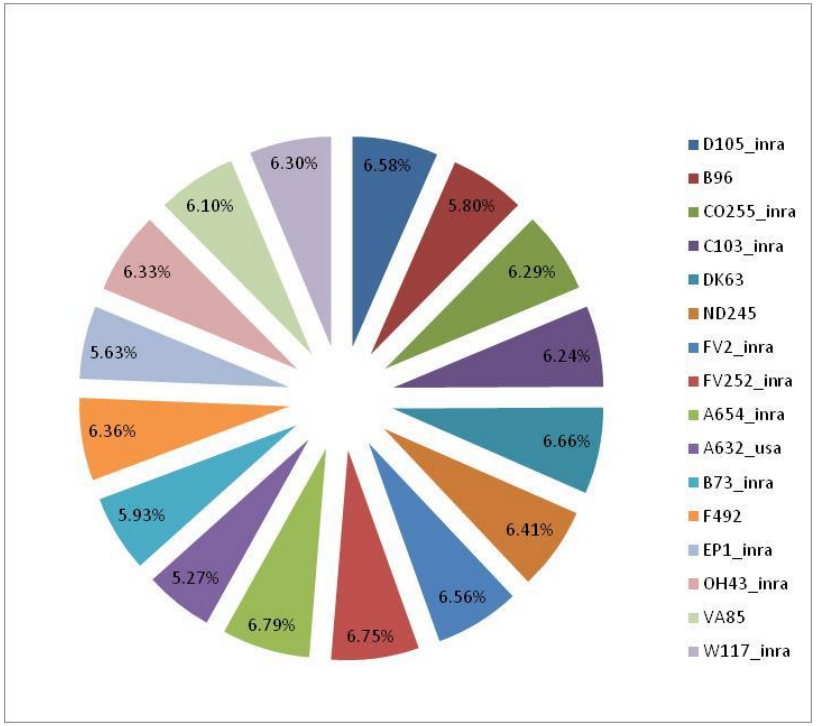
→ Very contrasted probabilities among the 16 parental alleles at a locus

→ Graphical genotypes
→ QTL detection

biogemma

Genetic characteristics of the panel

- Contribution of parental genomes to the genetic make-up of the DH lines

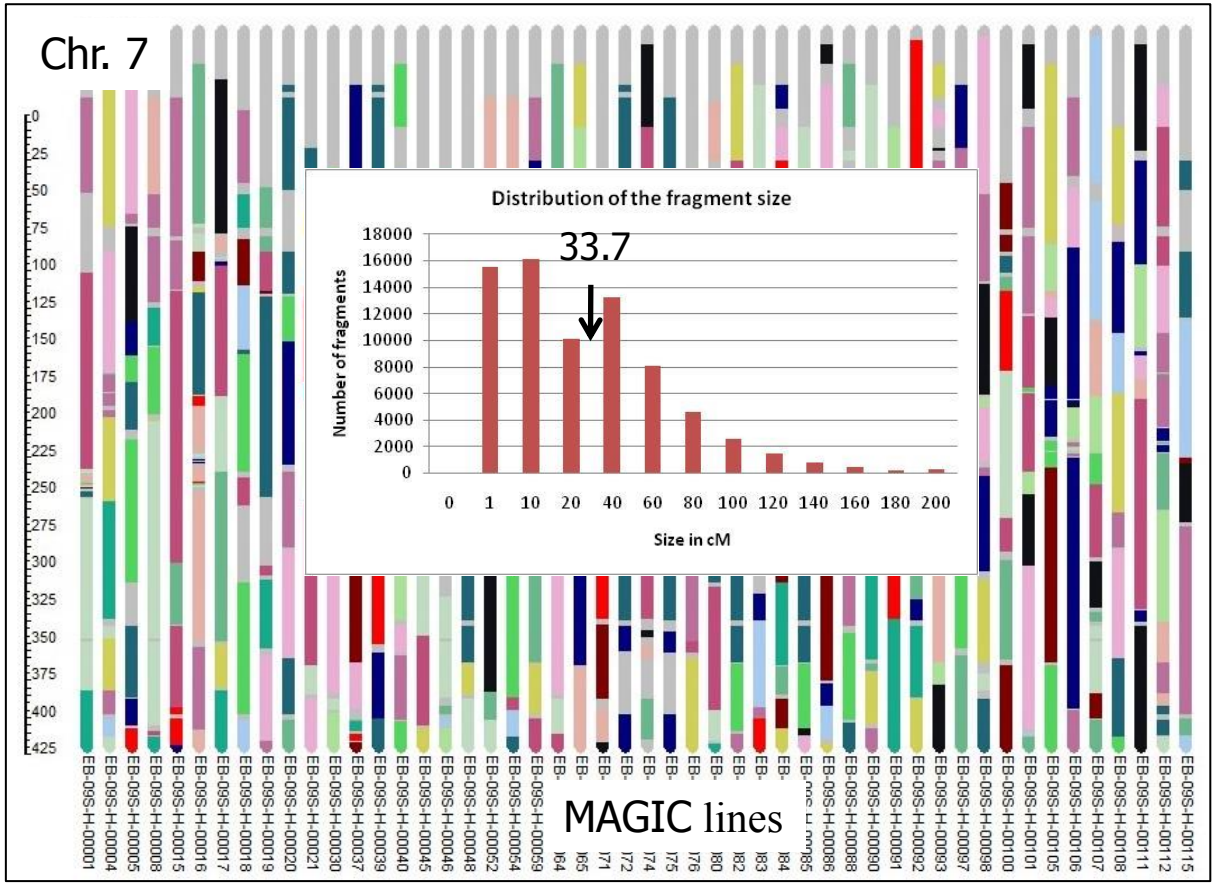


- Expected contribution from each parent 6.25%
- Estimated contributions from 5.27 to 6.75; very close to the expected one

  → No significant drift, nor selection

Genetic characteristics of the panel

- DH lines represented as mosaics of the parental genomes

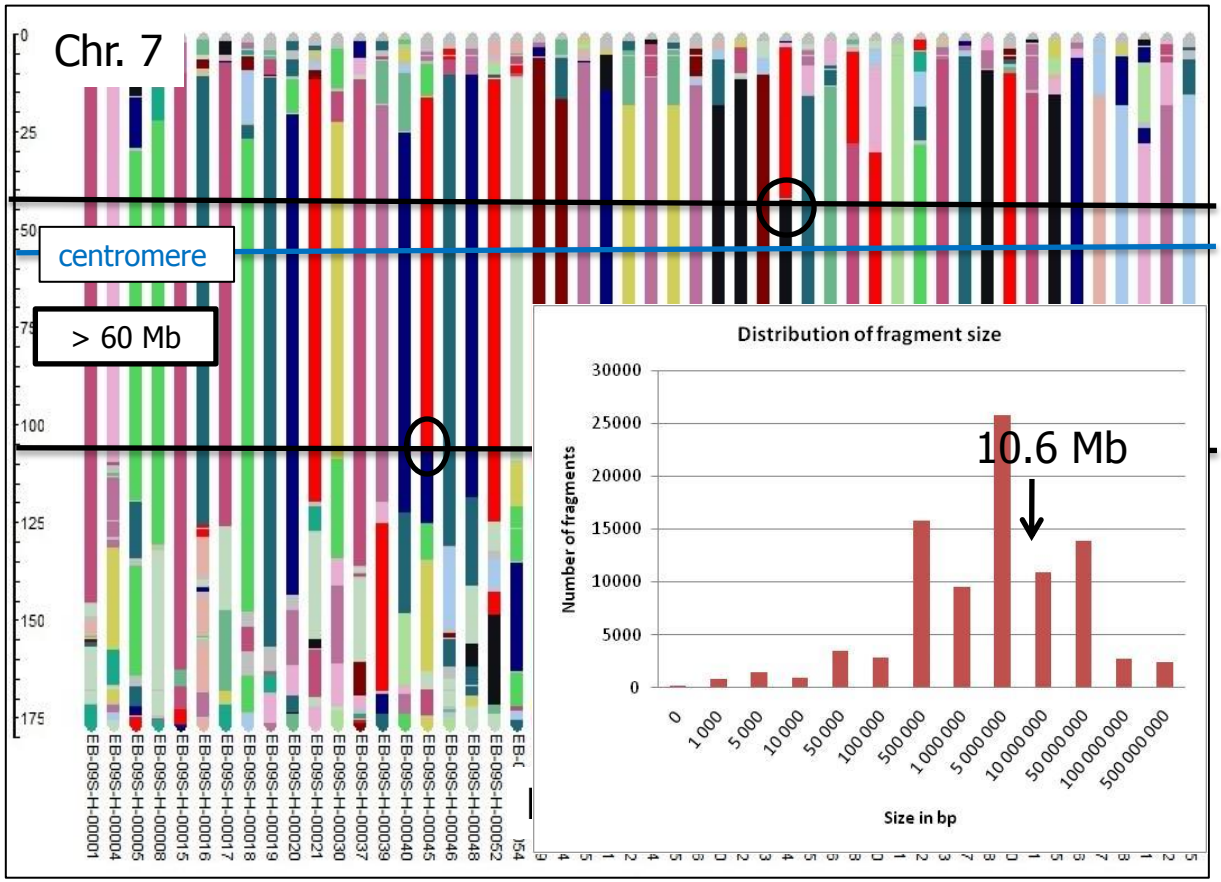


GGT v2

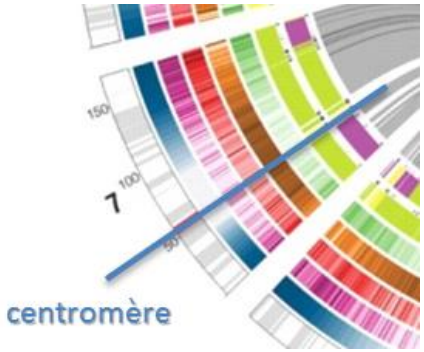
biogemma

Genetic characteristics of the panel

- DH lines represented as mosaics of the parental genomes



GGT v2



From Schnable et al., 2009

Most proximal recombinations

Phenotypic evaluation of the panel

- Summer 2011 Biomass trial
  - ✓Material evaluated
    - 317 test-cross progenies (out of 400) with suitable flowering time
  - ✓Experimental design
    - Randomized design with 2 sub-blocks grouping hybrids from a same precocity
    - 20% of the hybrids replicated twice (1.2 replicates / hybrid on average)
    - Standard sowing density (~95.000 plants/ha)
  - ✓Agronomical traits
    - Tasseling & silking
    - Plant height at harvest
    - Dry matter content at harvest (%)
  - ✓ Statistical analysis
    - spatial models with R/AsREML
- Evaluation for drought tolerance and NUE are carring out in different environments

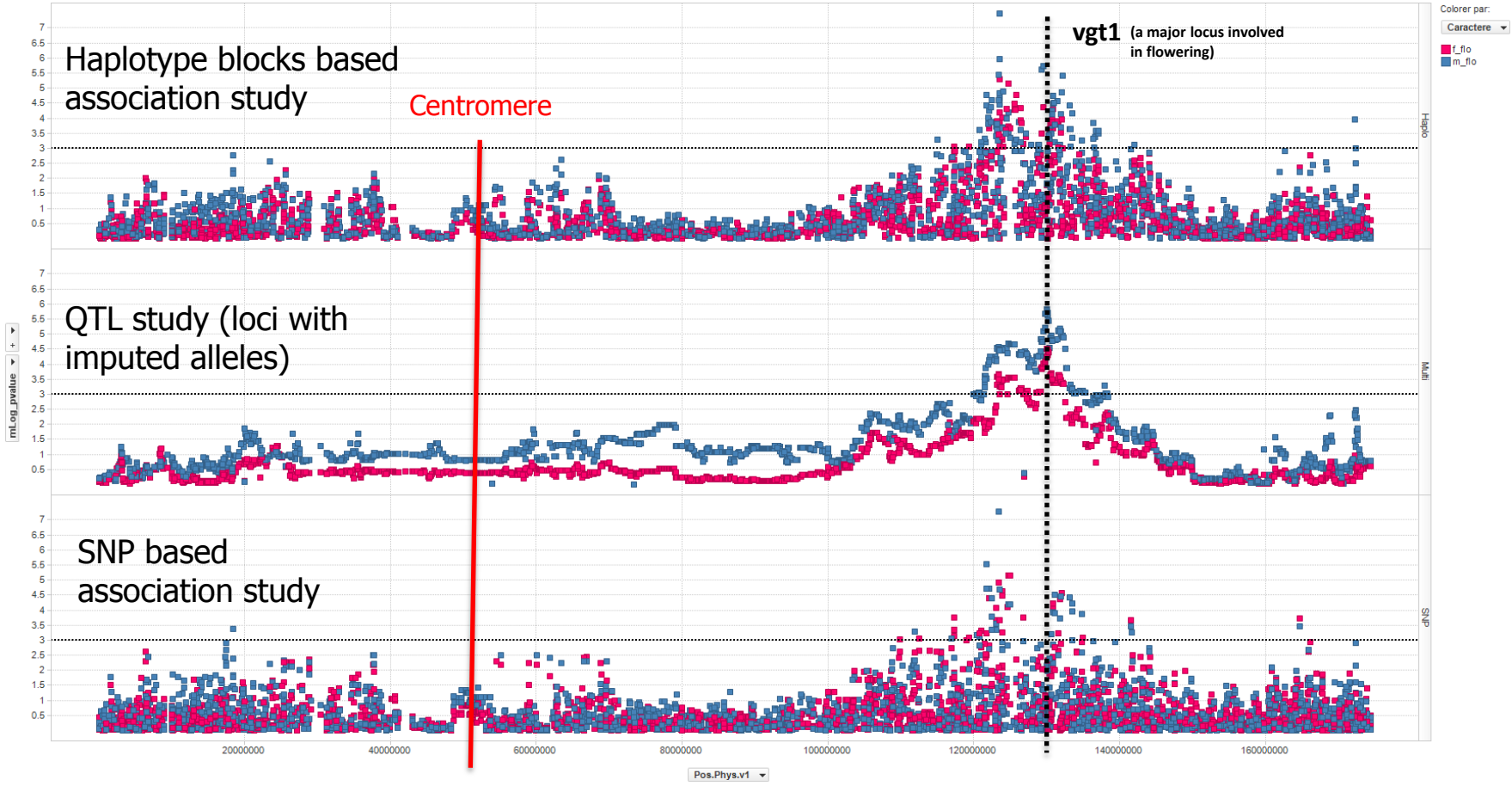
Association mapping

- Mixed model ***P = mu + L + K + R*** **,** fitted with **R/EMMA** or **R/AsREML**
  - ✓ ***P*** = adjusted means
  - ✓ ***K*** = kinship matrix (1- squared MRD)
  - ✓ ***L*** can be a **SNP,** a **haplotype** or a « **parental allele imputed** » locus

- Identification of haplotype loci
  - ✓ **based on LD extent** (Haploview)
    - Confidence Intervals *(Gabriel & al., 2002):* strong LD between each pair of markers
    - Solid Spine of LD *(Barrett et al., 2005):* strong LD between boundary markers pairs
    - Four Gamete Rule *(Wang & al., 2002):* based on two-marker halotypes frequencies
  - ✓ **based on a sliding windows**
    - Window of 100 kb / Step 50 Kb

GWAS for flowering

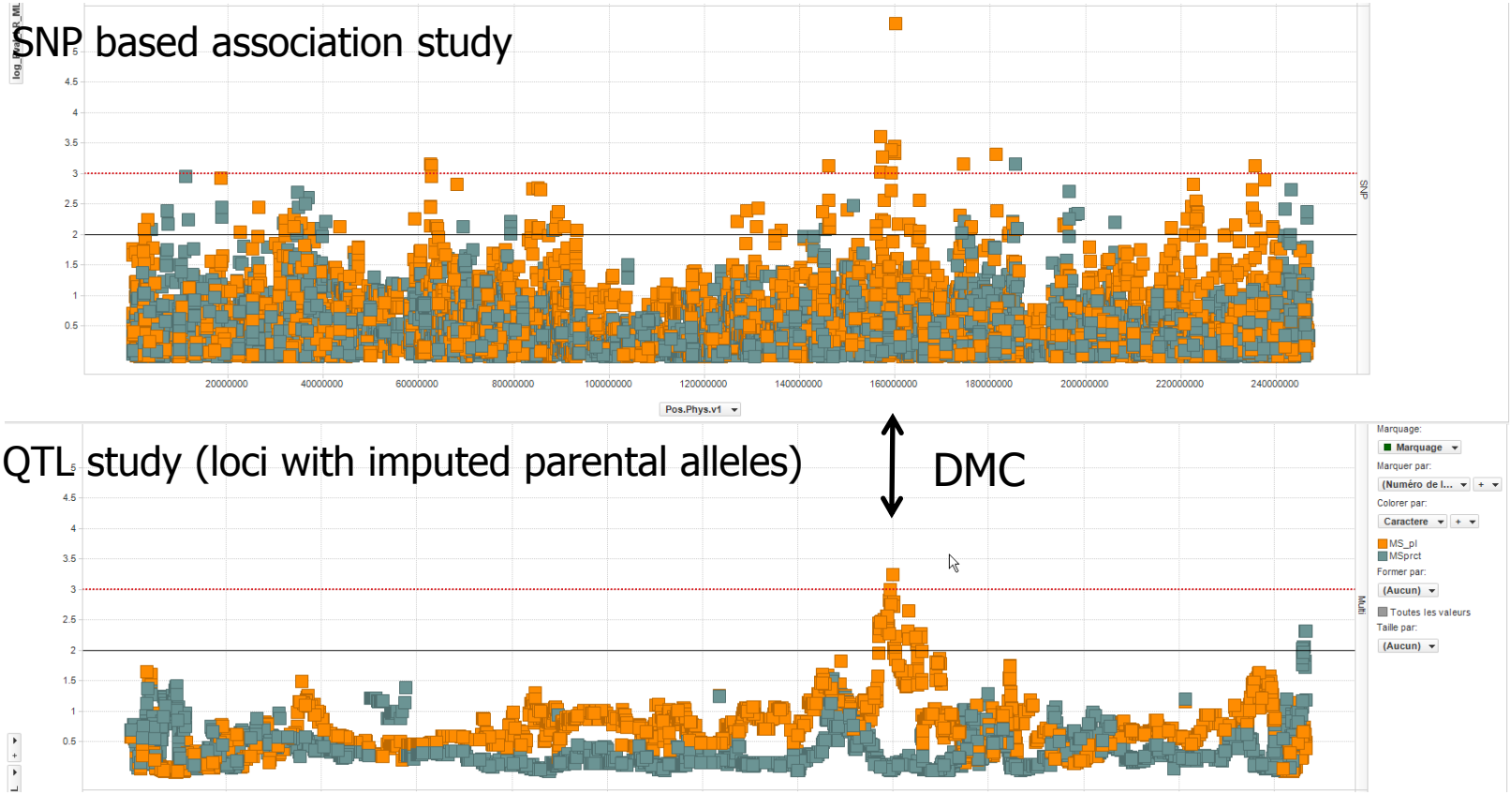
- Example of Mahattan plot for flowering times, here for chr.8



✓The profiles match well / high physical LD bw multiallelic loci clearly visible (huge redondance bw adjacent loci) / low basal noise for the QTL study

GWAS for flowering and biomass

- Example of Mahattan plot for dry matter content, here chr. 4



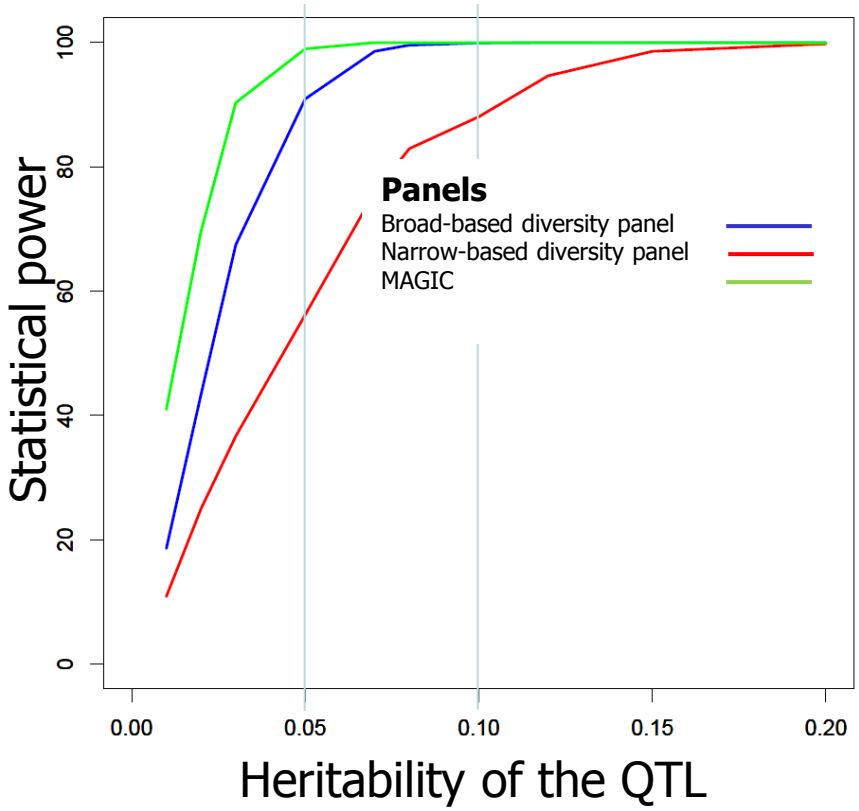
- Good agreement between both approaches → better confidence in the localisation of true genetic factor

Comparison of panels for detection power

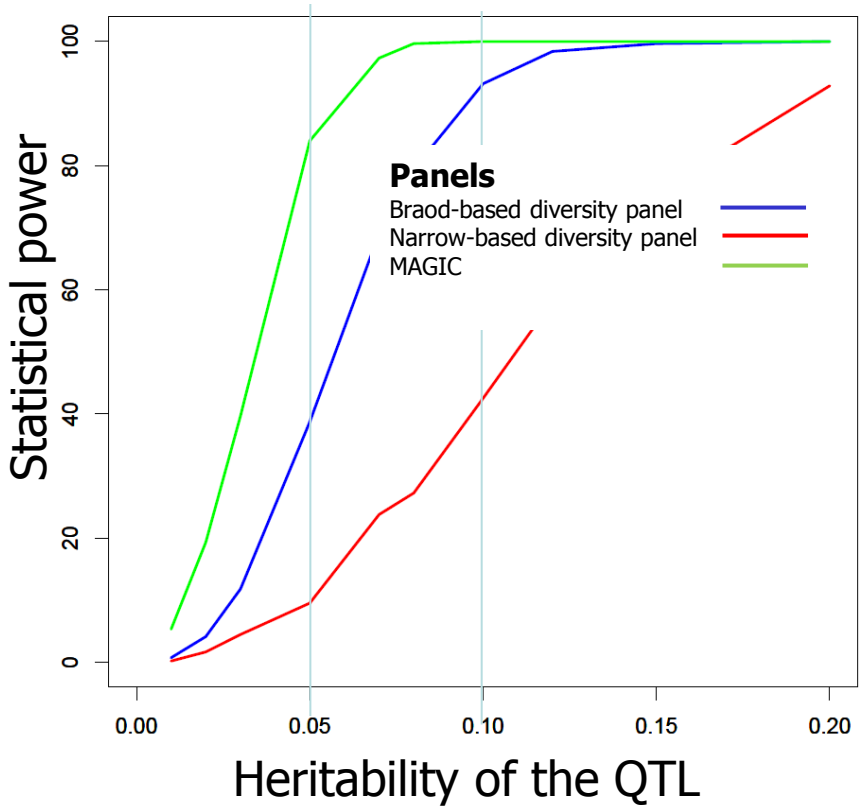
- 3 panels compared
  - ✓ Broad-based diversity panel
  - ✓ Narrow-based diversity panel (breeding lines)
  - ✓ MAGIC panel
- Based on a simulation study
  - ✓ Simulation of a polygenic trait from the SNP data
    - ➢ Genetic effect
      - ❖ One SNP (all each at a turn) as a QTL explaining a fraction of the phenotypic variance
      - ❖ 100 other SNP loci evenly sampled in the genome to simulate the genetic background
    - ➢ Environmental effect
      - ❖ drawn in a normal distribution to obtain a given heritability of the trait
  - ✓ Each QTL subjected to association study using the appropriate model (Q+K or K only depending of the panel). Statistical power as the ratio between the number of QTL detected over the total number of tests

Comparison of panels for detection power

Heritability of the trait = 0.4
α-risk = 0.001



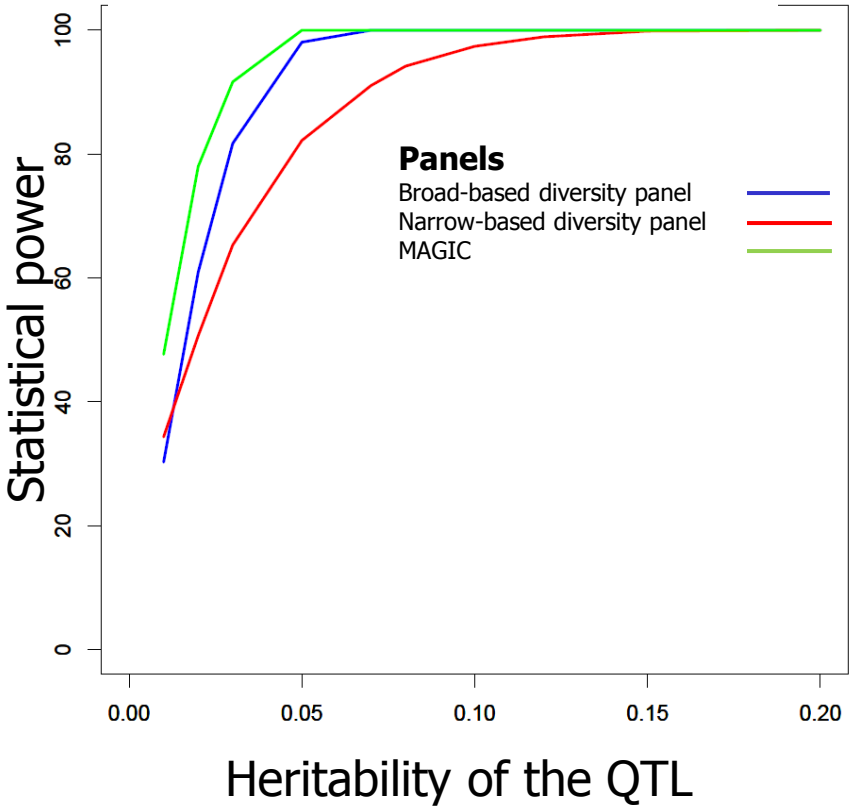
Heritability of the trait = 0.4
α-risk = 1.10-6



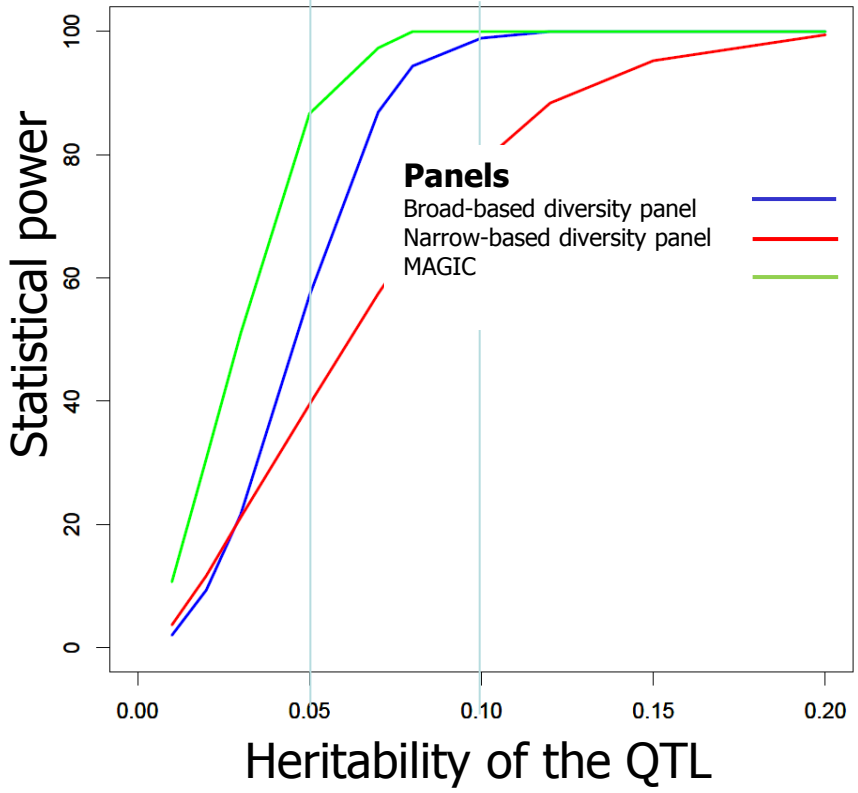
- MAGIC panel more powerful that the 2 diversity panels. The broad-based panel may suffer from a high structure whereas the narrow-based panel suffers from a lack of diversity and a smaller size

Comparison of panels for detection power

Heritability of the trait = 0.9
α-risk = 0.001



Heritability of the trait = 0.9
α-risk = 1.10-6



- For highly heritable traits, the MAGIC panel is still more powerful though the difference between the panels are lesser

Conclusions

- The MAGIC panel is appropiate for QTL and association mapping

  ✓ High allele diversity (~ broad-based diversity panel)

  ✓ Balanced frequencies of parental alleles

  ✓ No structure of the diversity (except around the centromere)

  ✓ Highly variable LD but not too strong in the telomeric parts (where most genes lie)

  ✓ High power (>0.8) to detect with a high confidence (1.10^{-6}) a loci explaining a modest part (5%) of a trait with a medium heritability (0.4)

  ✓ Detection of genomic regions that colocalize with previously mapped QTL/MetaQTL

The near future and after

- Continue the phenotypic evaluation
  - ✓ improved precision of phenotypic evaluation → more sensitivity and power to detect association
  - ✓ more characterized environments → modelling GxE
  - ✓ investigate QTL x E interaction, estimate environmental specific allele effects
- Increase marker density
  - ✓ 200K SNP GBC
  - ✓ 500K SNP Axiom array
- Extract a new panel of DH lines from an more advanced generation of intermating to increase resolution
  - ✓ the 11th generation will be achieved this year

Thank you for your attention

Project partly supported by the **KBBE-CornFed** program (2009-2013)

Special thanks to:

- **Clément Buet** who has performed most of the statistical analyses
- **Renault Rincent** for his contribution to the statistical analyses
- **Karl Broman** for having helped us to infer parental alleles
- **Morgan Renault** for having developed the MAGIC pop
- **Alain Murigneux** who has initiated the project
- **Sébastien Praud** the head of the GGC team

Contact: pierre.dubreuil@biogemma.com