Speeding up QTL cloning in maize: power and prospects of the MAGIC maize population
Matteo Dell’Acqua, Daniel M Gatti, Elisabetta Frascaroli, Gary A Churchill, Dirk Inzé, Michele Morgante, Mario Enrico Pè.

QTL cloning relies on the efficiency by which the casual variants affecting plant and animal phenotypes can be identified

Linkage mapping
- Low marker density
- Fully known pedigree
- High MAF
- Limited variation
- Few recombinations
- Time consuming

Genome-wide associations
- High marker density
- Unknown relationships
- Panel-dependent power
- Broad variation
- Many recombinations
- Faster, cheaper

Multiparental mapping panels try to bridge linkage mapping to association panels

- >> founder lines → more diversity
- >> intermitting generations → more recombinations

Two main designs for the production of Recombinant Inbred Lines (RIL):

CC/MAGIC-like
- 8 founder lines maximizing diversity
- A632; B73; B96; F7; H99; HP301; Mo17; W153R
- Balanced breeding scheme with permutations and pooling
- 10 years in the making
- 1,636 RIL F6 produced available to collaborators
- 1/3 characterized with 50k array and 4 phenotypes in two locations

The MAGIC maize
- 8 founder lines maximizing diversity
- A632; B73; B96; F7; H99; HP301; Mo17; W153R
- Balanced breeding scheme with permutations and pooling
- 10 years in the making
- 1,636 RIL F6 produced available to collaborators
- 1/3 characterized with 50k array and 4 phenotypes in two locations

Genetic diversity

NJ tree on +54k SNP clear ascertainment bias on Mo17

PCA 1 to 4
- low structure
- no remnants of breeding funnels
- any subset of the population is equivalent

Linkage disequilibrium decay and mapping definition
- \( r^2 \) LD halves within 1-4 Mb
- LD is higher in pericentromeric regions, in some cases outside centromeres
Genome reconstruction by HMM

Each RIL genome is reconstructed in terms of founder contributions.

92 recombinations/RIL on average

Mapping power

Minor allele frequency is between 0.5 and 0.125

Trade-off between diversity and mapping power

500 RIL have ~100% power to detect QTL explaining <10% of variance

Simulation of 20 QTL with QTL i having effect 0.9 i on each chromosome (100 simulations for each MAF 1 to 4, for each sample size)

Power at 500 MM RIL is similar to that at 1,000 NAM RIL

The MAGIC maize as a platform for QTL mapping

Genotypes

Phenotypes

Founder sequences

Founder RNAseq

The LOD curve has a flat top between ~7 Mb and ~23 Mb where founder W153R and Mo17 have the low phenotype

QTL scan for grain yield

QTL on Chr 6 explaining 13% of variance in grain yield

The MAGIC maize as a platform for QTL mapping

Projection of founder data

We produced complete sequences of the founders and counted how many reads were produced in each B73 (reference) genome bin

We produced RNAseq at the fourth leaf stage for all founders and tested the differential expression of the low founders (W153R and Mo17) against all the others

Projection of founder data

We produced complete sequences of the founders and counted how many reads were produced in each B73 (reference) genome bin

We produced RNAseq at the fourth leaf stage for all founders and tested the differential expression of the low founders (W153R and Mo17) against all the others

Projection of founder data

We produced complete sequences of the founders and counted how many reads were produced in each B73 (reference) genome bin

We produced RNAseq at the fourth leaf stage for all founders and tested the differential expression of the low founders (W153R and Mo17) against all the others
Way ahead

- MAGIC maize RIL are being used to produce RIXs (recombinant intercrosses) to study heterosis
- The genome of each RIX is known from the reconstruction of the genomes of its RIL founders
- 400 RIXs have been produced, more to come (more than 1.2e6 possible combinations)
- Genotyping of the full population (1,636 RIL) in progress
- Production of phenotype-specific additional layers of RNAseq data
- Seed amplification (currently in Chile) to distribute RIL and explore additional phenotypes

Acknowledgements

The Jackson Laboratory
Gary Churchill
Daniel Mario Gatti
IGA & University of Udine
Michele Morgante
Gabriele Magris
Federica Cattonaro
VIB, UGent (PSB)
Dirk Inzé
Frederik Coppens
Joke Baute
University of Bologna
Elisabetta Frascaroli

Days to pollen shed QTL

Greatest QTL on Chr 8, already described (19% variance)

In QTL interval:
- differential expression (D) on founders’ coefficients (B)
- Imputed SNPs form founders sequencing (C)

- 211 Kb haplotype
- IncRNA
- 500Kb from ZCN8
- GATA-domain

Started from 12 lines maximizing genetic diversity

- some exotic lines included
- photoperiod issue; exotic lines were not synchronized with established lines

MAGIC maize breeding scheme

35 subfamilies

Way ahead

- MAGIC maize RIL are being used to produce RIXs (recombinant intercrosses) to study heterosis
- The genome of each RIX is known from the reconstruction of the genomes of its RIL founders
- 400 RIXs have been produced, more to come (more than $1.2 \times 10^6$ possible combinations)
- Genotyping of the full population (1,636 RIL) in progress
- Production of phenotype-specific additional layers of RNAseq data
- Seed amplification (currently in Chile) to distribute RIL and explore additional phenotypes

Acknowledgements

The Jackson Laboratory
Gary Churchill
Daniel Mario Gatti
IGA & University of Udine
Michele Morgante
Gabriele Magris
Federica Cattonaro
VIB, UGent (PSB)
Dirk Inzé
Frederik Coppens
Joke Baute
University of Bologna
Elisabetta Frascaroli

Days to pollen shed QTL

Greatest QTL on Chr 8, already described (19% variance)

In QTL interval:
- differential expression (D) on founders’ coefficients (B)
- Imputed SNPs form founders sequencing (C)

- 211 Kb haplotype
- IncRNA
- 500Kb from ZCN8
- GATA-domain

Started from 12 lines maximizing genetic diversity

- some exotic lines included
- photoperiod issue; exotic lines were not synchronized with established lines

MAGIC maize breeding scheme

35 subfamilies
Inbreeding procedure

<table>
<thead>
<tr>
<th>Season</th>
<th>Year</th>
<th>F_n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer</td>
<td>2007</td>
<td>F₁ (100%)</td>
</tr>
<tr>
<td>Summer</td>
<td>2008</td>
<td>F₂ (50%)</td>
</tr>
<tr>
<td>Summer</td>
<td>2009</td>
<td>F₃ (25%)</td>
</tr>
<tr>
<td>Winter</td>
<td>2010</td>
<td>F₄ (12.5%)</td>
</tr>
<tr>
<td>Summer</td>
<td>2010</td>
<td>F₅ (6.25%)</td>
</tr>
<tr>
<td>Winter</td>
<td>2011</td>
<td>F₆ (3.125%)</td>
</tr>
<tr>
<td>Summer</td>
<td>2011</td>
<td>F₆ Sib (3.125%)</td>
</tr>
<tr>
<td>Summer</td>
<td>2012</td>
<td>F₇ Sib (3.125%)</td>
</tr>
</tbody>
</table>

35 8-ways F₁ pools (5,000 individuals)

1,636 RILs-BW F₆

529 RILs-BW F₅

IBS regions and genome reconstruction are affected by genotyping technique
ddRAD sequencing used to expand the genotyping

50k genotyping VS GBS genotyping

Hidden Markov Model

Three inputs for the HMM:
1. Prior probabilities for each genotype state (from which we start, meaning 1/8 or 1/9)
2. Emission probabilities (probability of observing each allele given the genotype state at each marker – starting from allele frequency in the founders)
3. Transition probabilities (probability of observing a recombination between markers, calculated from K. Broman’s work \( r_4 - r_1 + 2r \))

Credit: Dan Gatti

Then, Expectation-Maximization (EM) algorithm:
- E-step, calculation of probabilities as per HMM
- M-step, update the probabilities based on the MAGIC genotype
the model is run until the log-likelihood of the HMM differs by less than 1/1000th of the initial log-likelihood

Filtered

Smoothed
Transition Probability Matrix

Credit: Dan Gatti

HMM Example

Credit: Dan Gatti

HMM Example

Credit: Dan Gatti

HMM Example

Credit: Dan Gatti