Method of Analysis of Multi-parent Mapping Populations Affects Detection of QTL

Odell, S. G.[*,1], **Praud, S.**[†], **Ross-Ibarra, J.**[*,‡] **and Runcie, D.**[*,‡]
[*]Dept. of Plant Sciences and Center for Population Biology, University of California, Davis, CA, USA, [†]Limagrain, Chappes, France, [‡]Genome Center, University of California, Davis, CA, USA

ABSTRACT The search for quantitative trait loci (QTL) that explain complex traits such as yield and drought tolerance has been ongoing in all crops. Methods such as bi-parental QTL mapping and genome-wide association studies (GWAS) each have their own advantages and limitations. Multi-parent advanced generation inter-crossing (MAGIC) contain more recombination events and genetic diversity than bi-parental mapping populations and reduce the confounding effect of population structure that is an issue in association mapping populations. Here we discuss the results of using a MAGIC population of doubled haploid (DH) maize lines created from 16 diverse founders to perform QTL mapping, comparing QTL identified using a 600K SNP array to those found using founder probabilities and haplotype probabilities generated by determining the regions of the MAGIC DH lines that were derived from the 16 founders and by identifying regions of identity-by-descent (IBD) between the 16 founders, respectively. The three methods have differing power and resolution for detecting QTL for a variety of agronomic traits. This highlights the importance of considering different approaches to analyzing genotypic datasets, and shows the limitations of binary SNP data for identifying multi-allelic QTL.

KEYWORDS QTL, MAGIC

Introduction

First part

A good introduction is very important. This is how you can Hufford *et al.* (2012) in line or as reference in the end (Bourne *et al.* 2017).

Second part

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Materials and Methods

Genotype Data

The MAGIC population was derived from 16 inbred maize parents representing the diversity of European flint and U.S. dent heterotic groups. The 16 founder lines were crossed in a funnel crossing scheme, and then the resulting synthetic population was intercrossed for 3 generations with around 2000 individuals per cycle. Finally, 800 lines were selected from the synthetic population to create doubled haploids (DH), resulting in 550 MAGIC DH lines at the end of the process. The 16 founder lines and the MAGIC DH lines were all genotyped with the Affymetrix 600K Axiom SNP array. In addition, the 16 founder lines were sequenced with Illumina short-read sequencing to a depth of [?]x, resulting in [45.8?] millions SNPs and [??] indels after filtering using GATK best practices.

Phenotype Data

The MAGIC DH lines were crossed to a tester MBS84 to produce 344 hybrids. Due to variation in flowering time, a subset of the lines could not be crossed to the tester. The hybrids were grown in the field in Blois, France in 2014. For each genotype, two blocks of around 80 plants were grown under well-watered conditions. Measured phenotypes included days to anthesis (DTA), days to silking (DTS), plant height, percent harvest grain moisture, grain yield, and thousand kernel weight (adjusted to 15% humidity), where values were averaged over blocks. Both flowering time phenotypes were measured as the sum of degree days since sowing with a base temperature of 6°C (48°F). Days to anthesis was considered as the growing degree days until 50% of plants in a block were flowering at 25% of the central tassel spike.

Calculation and Validation of Founder Probabilities

The package R/qtl2 [] was used to determine founder probabilities of the DH lines using the 600K genotype data and the cross type "riself16". Due to the fact that the actual crossing scheme and the cross type input into R/qtl2 differed, we wanted to assess the accuracy of the founder probabilities. This was done by simulating lines using the actual crossing scheme and

[1] Dept. of Plant Sciences, University of California, Davis, CA, USA E-mail: sgodell@ucdavis.edu

assessing the performance of the calc_genoprobs function of R/qtl2 in correctly identifying the founder genotype. We developed an R package, magicsim, (github) to simulate the lines using the maize genetic map from Ogut et al. 2015 to generate approximate recombination rates across the chromosome. For 400 simulated lines, 99.6% of SNPs were correctly assigned to the founder genotype.

Calculation of Haplotype Probabilities

The use of founder probabilities makes the assumption that all 16 founders have distinct haplotypes. This assumption is not realistic, especially considering the known varying degrees of relatedness between the founders. The identification of regions of shared genetic sequence between founder pairs would allow for the collapsing of founder probabilities into haplotype probabilities. These haplotype probabilities have the potential to increase statistical power by reducing the number of tests performed in QTL mapping. Areas of uncertainty in founder probabilities of the DH lines were associated with regions of identity-by-descent (IBD) between two or more founder lines in that region of the chromosome. Identity-by-state (IBS) was measured from the 600K genotype data of the founders using the software GERMLINE with the paramaters "'-wextend"',"'–min_m 1", and "'–err_hom 4"' as a proxy for IBD. In order to assess the quality of predicted IBD regions, we also identified IBD regions using the founder WGS data. This was done using two separate softwares. GERMLINE was run using the same parameters as above on the WGS data with all sites with missing data removed, resulting in a total of 17.[?] million SNPs for comparison of sequence similarity. The software IBDSeq was run in addition using the full 45.[?] million sites in the WGS data. Comparision of overlap between pairwise IBD regions among the three methods showed [... significant overlap between the 600K and WGS data... actual numbers]. Haplotype blocks were created by grouping regions of pairwise IBD across the 16 founders. Within blocks, the founder probabilities for founders that were in IBD were summed to obtain haplotype probabilities. This resulted in haplotype blocks with the number of unique haplotypes within blocks ranging from [2?] to 16.

Association and QTL Mapping

The R package GridLMM [] was used to run association mapping using the three different methods of representing the genotype data. The function GridLMM_ML was used with the "ML" option. The model for each was

$$Y = X\beta + Zu + \epsilon \quad (1)$$

where Y is the response variable, X is the genotype matrix, β, is the effect size, Z is the design matrix, u is the random effects, and ϵ is the error. Due to only using one environment, no fixed effects were included in the model. Significance cutoffs for p-values were obtained using permutation testing, taking the 5% cutoff from 1000 randomized permutations for each method. Code for the analyses can be found [github link].

Method Comparison

The results of the three methods of identifying QTL were compared using two main criteria: (i) presence or absence of identified QTL peaks and (ii) the resolution of those QTL peaks. The GWAS methods should be most powerful at identifying QTL for which the causal variant is biallelic and the tagged SNP is in tight LD with the causal varaint. However, for multi-allelic QTL or QTL for which LD is low between tagged SNPs, this method begins to lose power. Founder probabilities increase the odds of detecting both QTL that are multi-allelic and QTL whose causal variant is not in tight LD with any one tagged SNP [Supplemental Figure?]. Lastly, haplotype probabilities potentially improve on the power of founder probabilities to detect QTL that meet the above criteria by reducing the number of test. However, haplotype probabilities also may obscure the signal of some QTL if founders are called as being in IBD with one another when they actually differ for the causal variant. Due to the fact that the founder and haplotype probabilities take into account recent recombination events, whereas the GWAS method only uses historical recombination, we predicted that founder and haplotype mapping would result in higher resolution around QTL peaks. Higher resolution QTL are ideal in that it makes it easier to narrow down candidate genes and potential causal variants when the significant window is smaller.

Results

Here is a single column figure (Figure 1)

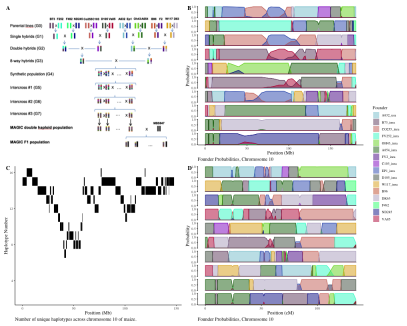

Figure 1 Important figure Describe the figure

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should

contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Discussion

Some more test

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

another figure

This is a full size figure (Figure S1) just add stars to the figure

Acknowledgments

We acknowledge the support of our coffee maker that made this work possible

References

Bourne, P. E., J. K. Polka, R. D. Vale, and R. Kiley, 2017 Ten simple rules to consider regarding preprint submission. PLoS computational biology **13**: e1005473.

Hufford, M. B., X. Xu, J. Van Heerwaarden, T. Pyhäjärvi, J.-M. Chia, *et al.*, 2012 Comparative population genomics of maize domestication and improvement. Nature genetics **44**: 808.

Supplement

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

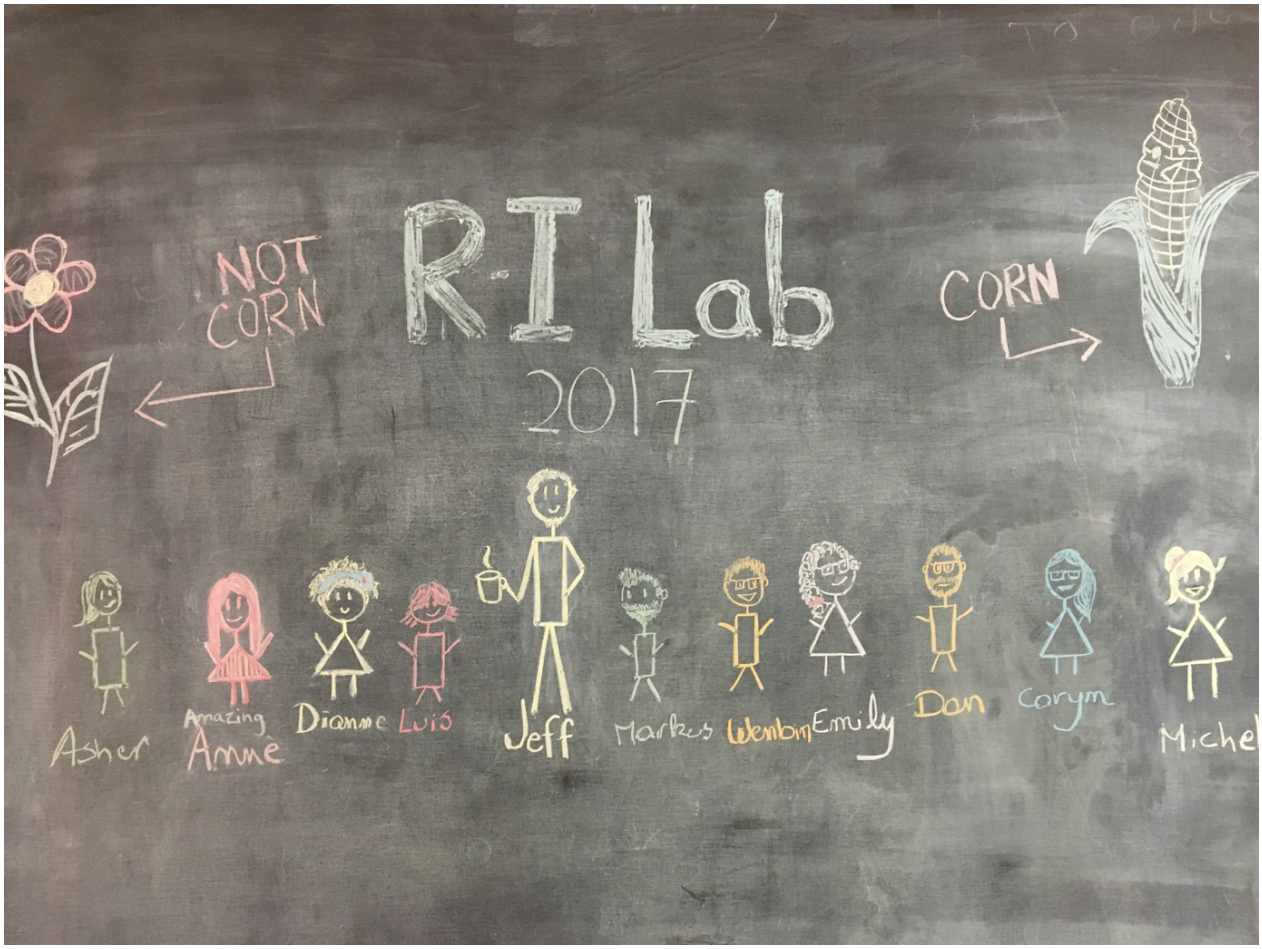

Figure S1 Supplemental figure Test test test

Table S1 Shrink a large table to fit the page

Parameter	Description
Adaptation	**Trait related parameters**
Time to optimum	Generations until new optimum is reached
Adaptation rate (haldane)	Adaptation rate until new optimum is reached. Calculated as $rate(h) = \frac{\frac{ln(x_2)}{sd_{x_{12}}} - \frac{ln(x_1)}{sd_{x_{12}}}}{t_2 - t_1}$
Final genetic variance	Genetic variance in the final generation
Fixations	**Mutations that fix after the optimum shift**
From new mutations (#)	Sum of fixed mutations in the final population that were already segregating before the optimum shift
From standing variation (#)	Sum of fixed mutations in the final population that arose after the optimum shift
Max. effect size	Maximal effect size of all fixations
Mean effect size	Mean effect size of all fixations
Mean effect size of negative fixations	Mean effect size of negative mutations
Mean effect size of positive fixations	Mean effect size of positive mutations
Mean emergence time	Mean generation when a mutation arose that fixed in the last 0.1 N generations
Mean fixation time	Mean generation in which a mutation fixed
Min. effect size	Minimal effect size of all fixations
Negative (#)	Sum of fixed mutations with negative effects in the final population
New/standing fixations	Ratio of mutations from new mutations vs. standing mutations
Proportion negative	Proportion of negative fixations from all mutations
Positive (#)	Sum of fixed mutations with positive effects in the final population
SD of effect sizes	Standard deviation of effect sizes of all fixations
SD of negative effect sizes	Standard deviation of effect sizes of negative fixations
SD of positive effect sizes	Standard deviation of effect sizes of positive fixations
Total (#)	Sum of fixed mutations in the final population
Sweeps	**Mutations that fix faster than 99% of neutral fixations**
Hard sweeps (#)	Sum of selective sweeps from new mutations
Proportion of hard sweeps	Porportion of hard selective sweeps of all selective sweeps
Proportion of sweeps from standing	Proportion of selective sweeps from stainding variation of all selection sweeps
Sweeps (#)	Sum of selective sweeps
Sweeps from standing variation (#)	Sum of selective sweeps from mutations that were already segregating before the optimum shift
Sweeps/fixations	Ratio of sweeps vs. fixations
Segregating sites	**Mutations that segregate in the final generation**
Max. effect size	Maximal effect size of segregating sites
Mean effect size	Mean effect size of segregating sites
Mean effect size of negative sites	Mean effect size of segregating sites with negative effects
Mean effect size of positive sites	Mean effect size of segregating sites with positive effects
Mean frequency of all sites	Mean allele frequency of segregating sites
Mean frequency of negative sites	Mean allele frequency of segregating sites with negative effects
Mean frequency of positive sites	Mean allele frequency of segregating sites with positive effects
Min. effect size	Minimal effect size of segregating sites
Negative (#)	Sum of segregating sites with negative effect
Positive (#)	Sum of segregating sites with positive effect
Proportion of negative sites	Proportion of segregating sites with negative effect of all segregating sites
Standard deviation of effect sizes	Standard deviation of effect sizes of all segregating sites
Total (#)	Sum segregating sites in the final generation