Modeling Allelic Diversity of Multi-parent Mapping Populations Affects Detection of Quantitative Trait Loci

Odell, S. G.[*,1,2], **Hudson, A. I.**[2,4], **Praud, S.**[3], **Ross-Ibarra, J.**[2,4,5] **and Runcie, D.**[1]

[1]Dept. of Plant Sciences, University of California, Davis, CA, USA, [2]Dept. of Evolution and Ecology, University of California, Davis, CA, USA, [3]Limagrain, Chappes, France, [4]Center for Population Biology, University of California, Davis, CA, USA, [5]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) populations 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 We compare QTL identified using a 600K SNP array to those found using estimates of the probability of ancestry from individual founders and probability of assignment to specific haplotypes. The three methods have differing power to detect QTL for a variety of agronomic traits. Although the founder probability approach finds the most QTL, all methods are able to find unique QTL, suggesting that each method has advantages for traits with different genetic architectures. A closer look at a well-characterized flowering time QTL, qDTA8, which contains *vgt1* and *vgt2* highlights the strengths and weaknesses of each method and suggests a potential epistatic interaction. Overall, our results highlight the importance of considering different approaches to analyzing genotypic datasets, and shows the limitations of binary SNP data for identifying multi-allelic QTL.

KEYWORDS MAGIC, QTL, linkage mapping, association mapping

Introduction

The study of evolutionary quantitative genetics requires the ability to link differences in phenotype to genotypic variation. Natural and artificial selection act on phenotypes, but only heritable phenotypic variation will result in changes in population means. Maize presents an excellent model organism to study quantitative genetics due to the combination of extensive genetic and phenotypic resources and the ability to create mapping populations. In addition, maize is one of the most widely produced crops in the world and is a major source of calories for millions of people. Decades of research into maize genetics have resulted in the identification of many quantitative trait loci (QTL) that explain variation in phenotypes such as yield, flowering time, and plant height (Buckler *et al.* 2009; Wang *et al.* 2006; Wallace *et al.* 2014; Beavis *et al.* 1991; Steinhoff *et al.* 2012). Such traits are extremely agronomically important, but are also crucial in terms of fitness and local adaptation.

Researchers have discovered large-effect QTL for a number of agronomic traits in maize through the use of different types of mapping populations (Huang *et al.* 2015). The choice of mapping population comes with associated advantages and limitations. In particular, they tend to vary in three main characteristics: (1) their ability to capture genetic diversity, (2) their power to detect QTL of small effect, and (3) their population size. Multi-parent Advanced Generation Intercross (MAGIC) populations have been used in breeding to increase the genetic diversity included in a mapping population compared to biparental populations (Huang *et al.* 2012; Dell'Acqua *et al.* 2015; Highfill *et al.* 2016; Aylor *et al.* 2011; Kover *et al.* 2009; Pascual *et al.* 2015). Compared to genome-wide association panels, MAGIC populations have more power to detect low frequency alleles and can better compare allelic effects between founders because the crossing scheme increases the frequency of all parental alleles to be approximately equal. Simulations of an 8-parent MAGIC population showed that sample sizes of 300 could detect QTL accounting for 12% of variance with a power of 82% averaged across minor allele frequencies (Dell'Acqua *et al.* 2015). Lastly, a MAGIC population avoids confounding due to population structure that is encountered with GWAS because the pedigree of the lines is known.

In this study, we used a MAGIC population of 344 doubled-haploid lines derived from 16 inbred maize parents developed by Limagrain to understand how genetic models can impact the identification of QTL. Compared to previous populations in maize, this MAGIC population has a greater number of founders, reduces complexity using DH instead of recombinant inbred lines, and should have comparable power to larger nested mapping populations (Dell'Acqua *et al.* 2015; Yu *et al.* 2008). For these reasons, the Limagrain MAGIC population has great potential to reveal new insights into the genetic control of quantitative traits in maize.

In addition to the choice of mapping population, the choice of how to represent genetic information through association

* Dept. of Plant Sciences and Dept. of Evolution and Ecology, University of California, Davis, CA, USA E-mail: sgodell@ucdavis.edu

and QTL mapping can impact the power of a study to detect and analyze QTL. A bi-allelic model for QTL, often used in genome-wide association analyses (GWAS), assumes that the underlying causal variants for QTL are explained by two alleles that are segregating in the population. Most commonly, genetic variation is represented as bi-allelic SNPs which segregate in the population, with each individual possessing either a reference or alternate allele. With this bi-allelic model, hereafter referred to as $GWAS_{SNP}$, markers represent a small region of the chromosome that is in tight linkage disequilibrium with the genotyped SNP.

An alternative model for the allelic state of QTL can be used in multi-parent populations, where we assume that each founder contributes its own allele. In this model, rather than looking at individual SNPs, an allele becomes the founder identity of that region, or which parent of the population that segment of chromosome was derived from. As a result, QTL are multi-allelic, with the number of alleles equal to how many founders were used in the making of the population. We will refer to this founder model hereafter as QTL_F.

The two allelic models described above make the assumption that for each genotyped SNP, there are either two distinct alleles in the population ($GWAS_{SNP}$) or as many alleles as there are founders (QTL_F). The latter assumption, although very possible for biparental mapping populations, becomes increasingly unlikely as the number of founders increases. This is because the founders used in the making of a population are related to one another with varying degrees of distance, and therefore, most likely share ancestral haplotypes through identity-by-descent (IBD). A third allelic model takes into account shared ancestral haplotypes between founders. This model, hereafter referred to as QTL_H, allows the number of alleles at each site to vary anywhere from two to the total number of founders (here 16), based off of the number of ancestral haplotypes at that site. This has the potential to increase statistical power compared to the QTL_F model by reducing the number of parameters.

Here we present a maize MAGIC population derived from 16 parents and discuss the performance of three different models for representing allelic states: bi-allelic, founder, and ancestral haplotype allelic models for detecting QTL. Using *vgt1*, a well-characterized flowering time QTL with a strong candidate causal variant that is variable in the population, we demonstrate differences between the three methods and explore potential epistatic interactions between *vgt1* and other genetic variation in the population.

Materials and Methods

Mapping Population

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 include more details on the crosses, how many seeds per cross, and then the resulting synthetic population was intercrossed for 3 generations with around 2000 individuals per cycle (Figure 1A). 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 MAGIC DH lines were crossed to a tester MBS84 to produce 325 hybrids (Figure 1A). asher i wonder if worth looking at GERP scores and regions that are under-represented from founders. $>$ 50% of individuals dropped out of DH process Due to variation in flowering time, a subset of the lines could not be crossed to the tester.

Genotyping

The 16 founder lines and the MAGIC DH lines were all genotyped with the Affymetrix 600K Axiom SNP array (Unterseer *et al.* 2014).

A total of 503,892 SNPs from the 600K after filtering out invariant sites. and [num] sites with missing data... fill in/finish

Phenotype Data

The MAGIC F1 plants were phenotyped in five different field locations in four different years, resulting in six distinct environment-years. The environment-years included Blois, France in 2014 and 2017, Nerac, France in 2016, St. Paul, France in 2017, Szeged, Hungary in 2017, and Graneros, Chile in 2015. For each genotype, two plots of around 80 plants were grown under well-watered or water-deficit conditions. this is confusing. implies all environments had both well watered and drought The well-watered environment-years were Blois, 2014, Blois, 2017, Szeged, 2017, and Graneros, 2015. The water-deficit environment-years were Nerac, 2016 and St. Paul, 2017, where water-deficit was variable between the different environment years based off of local climate. what does based off local climate mean Measured phenotypes included days to anthesis (DTA), days to silking (DTS), plant height (PH), percent harvest grain moisture (HGM), grain yield (GY), and thousand kernel weight (TKW) (adjusted to 15% humidity), where values were averaged over plots. 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 plot were shedding pollen on approximately one quarter of the central tassel spike. On average, around 300 of the MAGIC F1 lines were grown in each environment, with a minimum of 292 and a maximum of 309. There were a total of 325 lines that had both genotype data and phenotype data from at least one environment. For each of these lines we caluclated best linear unbiased predictor (BLUP) scores for all six phenotypes, combining measurements from all environments to get estimates of the genetic contribution to the phenotype for each MAGIC line Aulchenko *et al.* (2007).

Calculation and Validation of Founder Probabilities

We used the package R/qtl2 (Broman *et al.* 2019) to determine founder probabilities of the MAGIC DH lines using the 600K genotype data and the cross type "riself16". did Karl make this for us? if so we need to include in acknowledgements. SNPs used as markers for the QTL_F approach were filtered based on linkage disequilibrium using an iterative approach where a SNP was dropped if the R^2 value of probabilities between it and the previous SNP was greater than 0.95. rewrite correct? After filtering, a total of 4,578 sites were kept.

Due to the fact that the actual crossing scheme and the cross type input into R/qtl2 differed (DH lines rather than RILs), we wanted to assess the accuracy of the founder probabilities. This was done by simulating lines using the actual crossing scheme and assessing the performance of the calc_genoprobs function of R/qtl2 in correctly identifying the founder genotype (Figure 1). We developed an R package (R Core Team (2017)), *magicsim* (https://github.com/sarahodell/magicsim) to simulate the lines using the maize consensus genetic map from (Ogut *et al.* 2015) to generate approximate recombination rates across the chromosome. We simulated 100 MAGIC populations constituting 325 lines and assessed founder assignment accuracy as the average percentage of SNPs where the predicted founder was the same

as the actual founder.

Calculation of Identity-by-Descent and Haplotype Probabilities

The identification of regions of shared genetic sequence between founder pairs allows collapsing of founders into haplotypes. IBD was measured from the 600K SNP data of the founders using the software RefinedIBD with a sliding window of 10 cM and a minimum IBD segment length of 0.2 cM (Browning and Browning 2013). The resulting segments of pairwise IBD between each of the 16 founders were used to identify distinct haplotype blocks. We did this by moving along the chromosome, starting a new haplotype block when a segment of pairwise IBD between founders started or ended. Then, within blocks, we grouped all founders that were in IBD with one another into a haplotype. Within blocks, the founder probabilities for founders within a haplotype were summed to obtain haplotype probabilities. In certain instances, the pairs of founders that were in IBD with one another in a particular haplotype block formed an incomplete graph, where not all founders were in IBD with all other founders. This sentence is confusing to me; it seems like it's saying some pairs of founders that were in IBD were then not in IBD? As someone who does not know what an incomplete graph is I'm a little lostagree! For the sake of simplicity, we continued with the assumption that all founders in a haplotype were in IBD with one another.not sure what this is referring to? why wouldn't they be? However, it is important to note that haplotypes called here do still possess genetic differences between founders, with some founders more different than others.wait you mean two founders with the same haplotype have different genotypes? if so why is refinedIBD calling them IBD? Markers for the QTL_H mapping approcah were filtered for LD using an iterative approach similar to QTL_F: for all haplotype blocks with the same number of distinct haplotypes, a SNP was dropped if the correlation of probabilities between it and the previous SNP was greater than 0.95. correct as rewritten? After filtering, a total of 11,105 sites were kept to represent haplotype blocks in the MAGIC DH lines.

Association and QTL Mapping

The R package GridLMM (Runcie and Crawford 2019) 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 following three models were approximated by fitting each locus independently. The three methods differed in the X matrix used in the mixed linear model. The bi-allelic model ($GWAS_{SNP}$) was

$$Y = X_S\beta_S + Zu + \epsilon \quad (1)$$

where Y is the response variable, X_S is an n x p genotype matrix with reference and alternate alleles represented as 0 and 1, respectively, β_S is the effect size of the alternate allele, Z is the design matrix, u is the random effects of markers across the rest of the genome using the genomic relationship matrix, K, and ϵ is the error. GridLMM fit an n x 1 matrix for each site p. am I being dumb? why doesn't K show up here anywhere? don't we need an equation along the lines of

The founder model ($GWAS_F$):

$$Y = X_F\beta_F + Zu + \epsilon \quad (2)$$

where Y is the response variable, X_F is a (n x p) x f matrix and x_{fnp} is the probability that at site p, individuals n was derived from founder f, β_F, is the effect size of each founder allele, Z is the design matrix, u is the random effects of markers across the rest of the genome using the genomic relationship matrix, K, and ϵ is the error. GridLMM fit an n x f matrix for each site p.

The ancestral haplotype model ($GWAS_H$):

$$Y = X_H\beta_H + Zu + \epsilon \quad (3)$$

where Y is the response variable, X_H is a (n x p) x h matrix and x_{hnp} is the probability that at site p, individuals n has haplotype h, β_H, is the effect size of each haplotype allele, Z is the design matrix, u is the random effects of markers across the rest of the genome using the genomic relationship matrix, K, and ϵ is the error. GridLMM fit an n x h matrix for each site p.

Significance cutoffs for p-values were obtained using permutation testing, taking the 5% cutoff from 1000 permutations where genotypes were randomized relative to phenotypes for each method.

Model Comparison

The results of the three models were compared using two main criteria: (i) presence or absence of identified QTL peaks and (ii) the size of QTL support intervals. QTL support intervals were determined by identifying the most significant SNP for a QTL peak and demarcating the left and right bounds of the QTL as the left-most and right-most SNPs within a 100Mb window centered on the highest SNP that have a log10 p-value that is 2 log10 p-values below that of the highest SNP. The detection of QTL was compared across the three methods for each phenotype. A QTL was said to be identified across models if the QTL support interval for that QTL overlapped. The effect of the model used on the size of QTL support intervals was investigated using the QTL which were identified by all three methods (n=40). The model constructed was:

$$SupportIntervalSize = Environment - QTL + Model + \epsilon \quad (4)$$

The support interval size response variable was represented both in terms of physical distance (Mb) and genetic distance (cM).

Estimation of Effect Sizes

We used the R package *lme4qtl* to calculate standard errors of effect sizes relative to the population mean Ziyatdinov *et al.* (2018). For the QTL_F model, probabilities were dropped (converted to missing) for individual founders at some sites if the sum of probabilities for a particular founder across all MAGIC lines was less than 1 or if there where fewer than 5 MAGIC lines that had a probability greater than 0.8 for one of the founders. This same filtering was done with QTL_H effect sizes for sites with low representation of particular haplotypes. This was to ensure that effect sizes for individual founders and haplotypes could be effectively estimated. The comparison of effect sizes between methods was done using the effect size estimates of the most significant site within the QTL support interval. For methods that did not identify a QTL detected in other methods, comparison of effect sizes used the site with the highest log10 p-value that overlapped with the QTL support interval from the method that did detect the QTL. We confirmed that effect sizes calculated by GridLMM and lme4qtl matched one another, with the correlations in effect sizes between the two methods greater than 0.99.

We also compared the individual QTL identified by the three models in regards to the proportion of phenotype variance explained (PVE) by the QTL. For each model, the highest non-significant peak within the support intervals of the QTL found

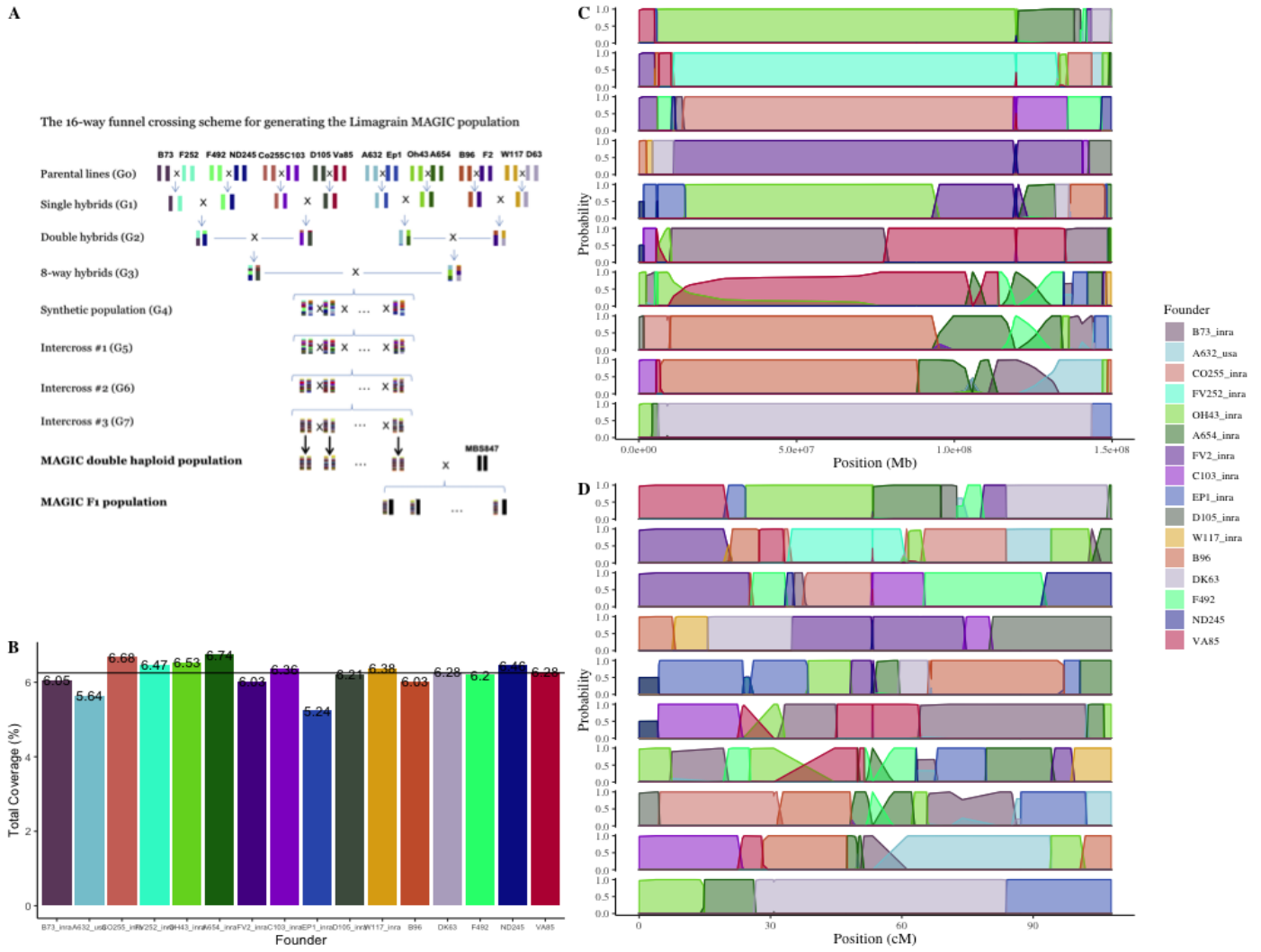

Figure 1 Structure, diversity, and founder representation of the MAGIC population(**A**) The crossing scheme of the MAGIC population. (**B**) Total coverage of each founder across the population as a percentage. Black line shows expectation of equal distribution (6.25%) (**C**) Founder probabilities for 10 MAGIC DH lines on chromosome 10 in physical distance. (**D**) Founder probabilities for 10 MAGIC DH lines on chromosome 10 in genetic distance.

in other models was used to represent the unidentified QTL. We tested for differences in PVE of QTL that were detected or undetected between models. We did this by fitting a mixed linear model with all combinations of method and whether or not a QTL was identified (i.e. QTL_F-True) as fixed effects and individual QTL as random effects. We tested for significant differences between the least square means of the estimated fixed effects using a Tukey pairwise test.

Tests for Epistasis

We ran a genome scan for epsistatic interactions with *vgt1*. The probability of the MAGIC lines having the MITE insertion at *vgt1* was calculated by summing the founder probabilities for all founders that have the $MITE^+$ allele at the site closest to the location of the MITE in the B73 APGv4 assembly (cite MaizeGDB). Lines that had uncertain allelic states at the MITE ($0.05 > \Pr(\text{MITE}) < 0.95$) were dropped for the test. Applying a Bonferroni significance threshold adjusted for the number of tests, we tested for epistasis using the 600K genotype data.

We also performed QTL mapping with QTL_F using only the $MITE^+$ MAGIC lines. This was to see if there were any other loci whose affect was only observed in the presence of the MITE. Additionally, this provided an alternative way of testing for epistasis with *vgt1* using the founder probabilities. An normal epistatic model could not be fit with founder alleles because there were not enough degrees of freedom to compare each founder. We used the model from Equation 2 using DTA BLUP scores and each independent environment and the 5% significance threshold for DTA.

Test for Over- and Under-Representation of Alleles

The rounded sum of the founder probabilities for each founder across all lines was used as an approximation of the number of lines that had that founder at a site. This observed count was compared to a null expectation of 1/16 for equal distribution across lines (approximately 21 lines per founder). We performed a χ^2 test for each site to determine if founder counts significantly deviated from null expectation, adjusting for multiple testing (p-value<xx). In order to ensure that significant χ^2 sites were not

due to sequence similarities between founders, we performed the same test with haplotype probabilities (p-value<xx). With haplotype probabilities, the null expectation was proportional to the number of founders grouped within each haplotype and the total number of unique haplotypes at the site.

Flowering Time Enrichment Test

We used a list of flowering time (FT) genes assembled by (Wang *et al.* 2017) to test for enrichment of FT genes in haplotype χ^2 peaks. Of the 907 genes, we used 88X which were aligned to chromosomes 1 through 10 in the B73 AGPv4 assembly Jiao *et al.* (2017). To determine a null distribution, we randomly sampled XX number of non-FT genes and counted the number that overlapped with regions within χ^2 peaks. We compared this number to the actual number of FT genes that overlap with χ^2 peaks.

Calculation of Flowering Time Polygenic Scores

We calculated the polygenic scores (PGS) using the software rrBLUP (cite). To determine a null distribution of FT PGS, we also calculated PGS for 100 populations of simulated MAGIC lines. We performed an X?? test to determine if the PGS of the actual DH lines were earlier than would be expected by chance.

Results

MAGIC Population

We developed a 16-parent MAGIC population using temperate inbred maize lines representative of the diversity Flint and Dent heterotic groups of North America and Europe (Figure 1A). We genotyped 334 MAGIC DH lines from the population with 600K chips and measured 6 pheotypes across 6 environments from the MAGIC F1 lines, resulting from crossing the DH lines to an inbred tester, MBS847. The distributions of the measured phenotypes were approximately normal (Figure S1). Some deviations from the assumptions of normality were observed for days to silking (DTS) and, as a result, anthesis-silking interval (ASI). Using the phenotype data from the six environments, we calculated BLUP scores for each of the MAGIC lines.

Simulation and Validation of Founder Probabilities

We partitioned the genomes of individual MAGIC lines into segments of ancestry from the 16 founders. This allowed us to determine the predicted contribution of each founder to the population (Figure 1B & D). The founder probabilities determined using R/qtl2 were able to assign founders to the actual MAGIC DH lines with high confidence (>0.90) for [percentage] of the 10 chromosomes of maize. Our simulations suggest a very high ($\mu = 99.8\%, \sigma = 0.011$) assignment accuracy (see Methods). This reinforced our confidence in the founder probabilities obtained from the actual data.

Identity-By-Descent and MAGIC Haplotypes

In some cases, the model which inferred founder identity in the MAGIC lines had high uncertainty, with probabilities split approximately equally between two founders. We hypothesized that this uncertainty was due to the two founders having very similar genetic sequence at those regions, such that the model struggled to differentiate the two. To assess the genetic similarity of the founders, we calculated pairwise Identity-By-Descent (IBD) between all founders using the software Refined-IBD Browning and Browning (2013). Areas of uncertainty in founder probabilities of the DH lines were associated with regions of IBD between two or more founder lines in that region of the chromosome (supplemental example figure). The results showed that a total of [bp][cM] of genome were in IBD between at least two different founders. The average size of an IBD segment between two founders was 140kb (0.51 cM) with a median of 122kb (0.46).

Pairwise IBD segments sizes ranged from 8 kb (0.3 cM) to 673 kb (1.61 cM). For founder pairs that were found to be in IBD with one another, the total percentage of IBD between founders ranged from 0.0018% (F492 and VA85) to 4.3858% (B73 and A632), with an average of 0.06130%. There were no IBD segments found for 18 of 120 possible pairwise founder combinations. Amount of IBD between the founders and the tester, MBS847... The Neighbor-Joining Tree of the relatedness of the 16 founders recapitulated the IBD results (Supplemental Figure). For a particular founder pair, B73 and A632, there were large segments where the lines shared haplotypes, and the tree placed them very close together. This is consistent with pedigree of the lines, where A632 was derived from B14, a line from the same population as B73 (cite). Due to the widespread Pairwise IBD between the founders, it appeared that many founders shared ancestral haplotypes. Within individual blocks of ancestry, we collapsed founder alleles that were identical by descent into a single haplotype. The number of unique haplotypes within haplotype blocks varied across chromosomes, ranging from 6 at the lowest to 16 at the highest (Figure 2A B). The average number of unique haplotypes per haplotype block was 13 ($\mu = 12.85, \sigma = 1.71$) (Figure 2B). There was a wide range of haplotype block sizes, with the average physical size of haplotype blocks being 303.7 kb (σ = 1.71Mb) (Figure 2C)). The largest haplotype block was 39.3 Mb long on chromosome 7, which had 16 unique haplotypes. In genetic distance, haplotype block sizes range from 0 to 3.4 cM, with an average of X (Figure 2D).

QTL Mapping and Association Mapping

We performed association mapping using three models of the allelic state of QTL. The first method, $GWAS_{SNP}$, used SNP genotypes obtained from the 600K array, assuming that QTL are bi-allelic. The second method, QTL_F, used probabilities of founder identity in chromosome segments, assuming that a QTL had as many alleles as founders. The third method, QTL_H, used probabilities of haplotype identity in chromosome segments, assuming that a QTL had as many alleles as ancestral haplotypes. The three methods varied in their ability to identify QTL. Figure 3 shows the Manhattan plots from the three methods for BLUP DTA overlayed with the support intervals of significant QTL for all BLUP phenotypes. For DTA, all three methods easily identify two large QTLs, *qDTA3-2* on chromosome 3 and *qDTA8* on chromosome 8. These QTL correspond to three previously identified QTL, *vgt3* for *qDTA3-2* and *vgt1* and *vgt2* for *qDTA8*. In addition, there are multiple QTL that are found by only a subset of the models. Using BLUPs, the $GWAS_{SNP}$ method was able to identify one QTL on chromosome 5 for thousand-kernel weight, *qTKW-5* that was not found in either QTL_F or QTL_H. A second QTL on chromosome 3 for DTA, *qDTA3-1*, was only found with the QTL_H method. A DTA QTL on chromosome 9, *qDTA9* and a harvest grain moisture QTL on chromosome 3, *qHGM3-1* were only found using QTL_F. However, a QTL for DTS with overlapping support intervals to *qDTA9* was found in both QTL_F and QTL_H.

We performed QTL mapping separately in each of the 36 envi-

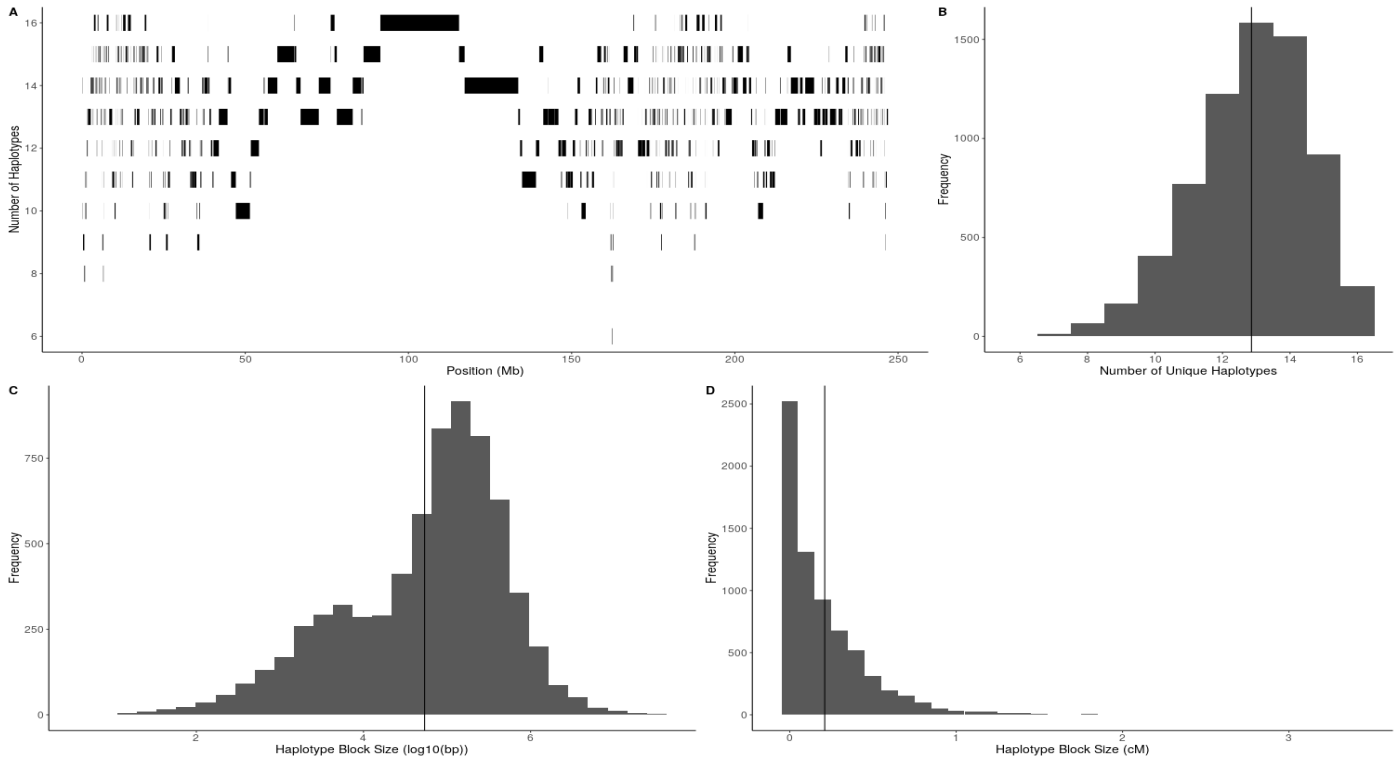

Figure 2 Diversity and Size of Haplotype Blocks The number of unique haplotypes among the 16 MAGIC founders across the 10 chromosomes of maize (**A**) The number of unique haplotypes per haplotype block and size of haplotype blocks along chromosome 4 in physical distance. (**B**) Distribution of unique haplotypes per haplotype block across the genome. (**C**) Distribution of haplotype block size in physical distance, represented as log10(bp). The black bar represents the average size. (**D**) Distribution of haplotype block size in genetic distance, represented as cM. The black bar represents the average size.

ronment:phenotype combinations, plus the across-environment averages of the 6 traits, for a total of 42 separate analyses. When we called each environment-phenotype as its own trait (i.e days to anthesis in Blois, 2014), a total of 46 unique QTL were identified from all three mapping methods. 26 (57%) QTL were identified by all three models. 6 QTL were found in both QTL_F and QTL_H and 1 QTL was found in both $GWAS_{SNP}$ and QTL_F. 7, 3, and 3 QTL were found in only $GWAS_{SNP}$, only QTL_F, and only QTL_H, respectively (S2).

Next we merged QTL of the same phenotype from different environment based on overlapping support intervals. There were 20 unique across-environment QTL identified from all methods. 10 (50%) of these across-environment QTL were found in more than one environment. 12 across-environment QTL were found using BLUPs and 2 of these QTL were not found in any individual environment. 6 (0 unique) QTL were found in Blois, 2014, 7 (1 unique) in Blois, 2017, 7 (3 unique) in Graneros, 2015, 5 (1 unique) in Nerac, 2016, 6 (3) in St.Paul, 2017, and 3 (0 unique) in Szeged, 2017.

Despite differences in the models, the power to identify and refine the location of QTL was similar across the three methods. QTL_F was able to identify the most QTL, regardless of changes in the significance threshold (Supplemental Figure). Individual QTL that were found in one method at the 5% significance threshold usually became significant in other methods when at the 10% threshold, indicating that the differences in the ability to detect these QTL between methods is mostly due to difference in power (S3). Nonetheless, there were multiple QTL that were identified in only one method. There were 10 QTL that appeared in one method and not in any others, mostly related to grain yield and thousand kernel weight traits.

There was no significant different in the size of QTL support intervals between methods, both in physical (F-value = 0.5157, p-value = 0.5989) and genetic distance (F-value = 1.1658, p-value = 0.3165). Although on average, the physical and genetic size of $GWAS_{SNP}$ support intervals were larger than those of QTL_F and QTL_H support intervals, the difference was not significant and was most likely due to a few outliers. The average physical sizes of QTL bounds were 14.4 Mb (SE = 1.18 Mb) for $GWAS_{SNP}$, 12.8 Mb (SE = 1.07 Mb) for QTL_F, and 13.4 Mb (SE = 1.09 Mb) for QTL_H. The average genetic sizes of support intervals were 6.24 cM (SE = 0.353 cM) for $GWAS_{SNP}$, 5.85 cM (SE = 0.320 cM) for QTL_F, and 5.51 cM (SE = 0.325 cM) for QTL_H. Sorry if I missed it, do you talk about whether any of the QTL are pleiotropic?

Variation in Effect Sizes

In select cases, effect sizes and allelic state of QTL seemed to explain the models' differing success in detection. Even for QTL that were identified by all methods, the estimated effect sizes of the QTL differed both across environments and between methods (include figure). For QTL that were found with some methods and not others, we attempted to examine the effect sizes to see if they provided insight into the genetic architecture of the QTL.

For QTL detected in QTL_F and not QTL_H, generally, there were higher numbers unqiue of haplotypes in these cases [stats]. Comparing effect sizes of QTL that were identified in QTL_H and not QTL_F, there tended to be fewer unique haplotypes at these

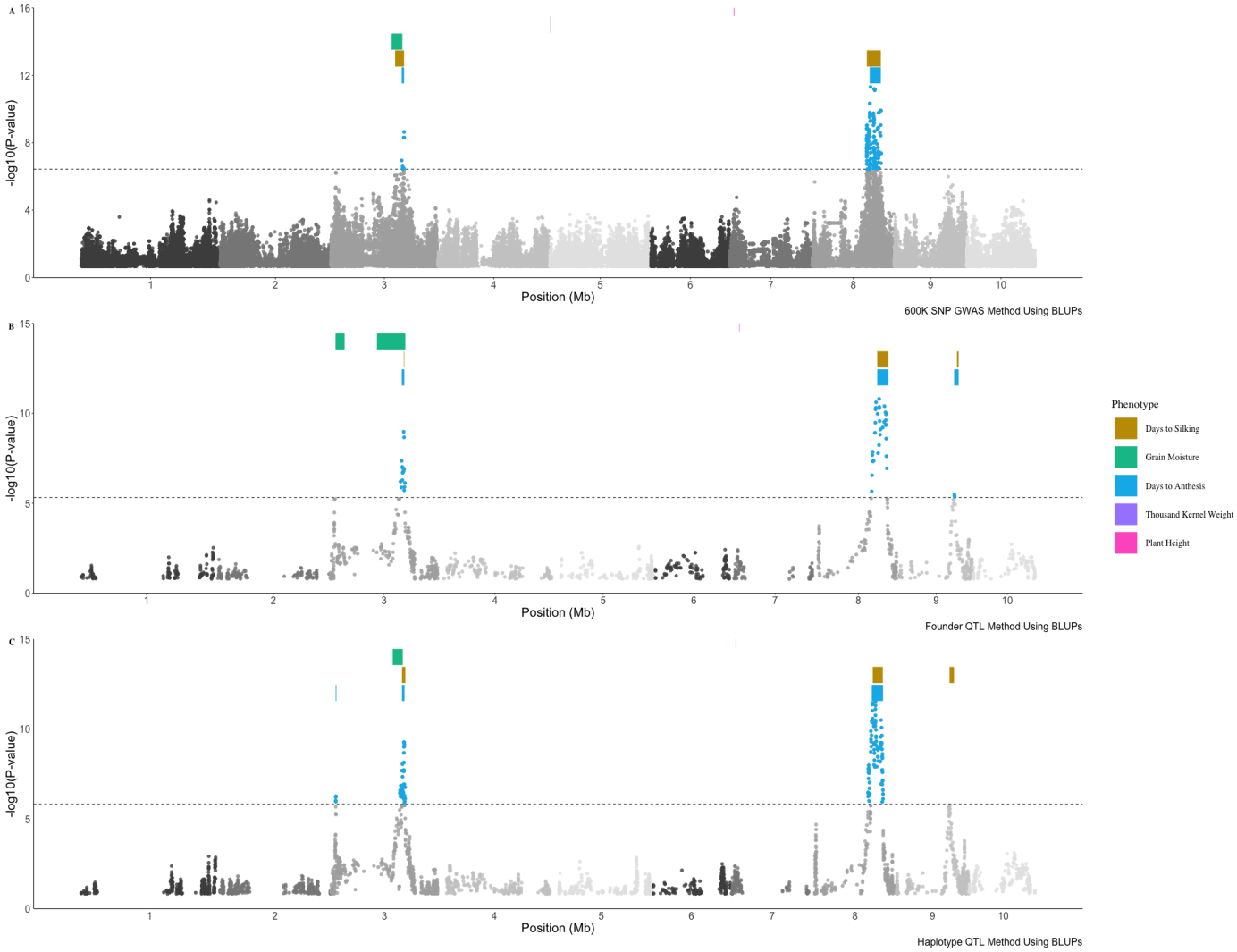

Figure 3 Results of three methods of QTL identification Colored points represent significant QTL for Days to anthesis (DTA) BLUPs above the 5% significance threshold from 1000 random permutations. Colored bars represent the location QTL support intervals for five other BLUP phenotypes (list colors and add legend, make the small intervals more visable) **top** GWAS results using the 600K SNP array. **middle** Results from QTL mapping using founder probabilities **bottom** Results from QTL mapping using haplotype probabilities

QTL regions and the effect sizes of haplotypes that grouped together more than one founder were similiar to the estimated effect sizes of those founders [stats?? figures?].

For each model, we compared QTL detected by that model to those not detected (see Methods). Comparing within models, we found that there were significant differences in the proportion of phenotypic variance explained (PVE) by QTL that were and were not detected by individual models (Figure 5). Perhaps unsurprisingly, the PVE of detected QTL were signfiicantly higher than that of undetected QTL for QTL_F (t-ratio -4.566, p-value 0.0002) and QTL_H (t-ratio -4.145, p-value 0.0009). However, QTL not detected by $GWAS_{SNP}$ had signficantly higher PVE than those detected by that method (t-ratio 3.267, p-value 0.0178). Between models, QTL detected by QTL_F and QTL_H explained significantly more phenotypic variance than those detected by $GWAS_SNP$ (t-ratio -28.883 p-value <.0001; t-ratio -31.278 p-value <.0001), although there was no significant difference between QTL_F and QTL_H (t-ratio 1.658, p-value 0.5629)

Variation around vgt1 *and* vgt2

One notable QTL that was identified by all three models using BLUPs (Figure 3) and nearly all individual environments was qDTA8, a large QTL on chromosome 8 that was strongly correlated with variation in days to anthesis, as well as days to silking. The support interval for this QTL overlapped with two previously characterized flowering time QTL, *vgt1* and *vgt2*. These known flowering time loci provided a useful benchmark for comparison of these approaches.

At *vgt1*, the founders are segregating ($MITE^+/MITE^-$) for the causal variant, a MITE insertion in a conserved non-coding sequence upstream of *ZmRap2.7* (Salvi *et al.* (2007); Castelletti *et al.* (2014)). Looking at the most significant SNP for *qDTA8* from $GWAS_{SNP}$, the alternate allele correlated strongly but imperfectly with the presence of the MITE in the founders (correlation). We expected to see QTL_F effect sizes at this locus that match the allelic state of the founders, with $MITE^+$ founders having earlier effect sizes and $MITE^-$ founders having later

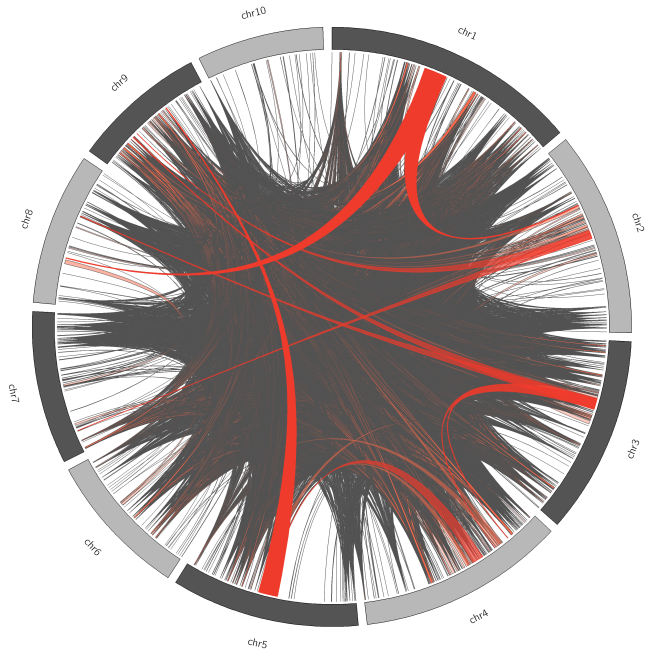

Figure 4 Interchromosomal Linkage Disequilibrium in the MAGIC Population Ribbons represent regions of $R^2 > 0.8$ between consecutive SNPs on different chromosomes. Dark, solid, red bands are regions larger than 20Mb on at least on of the chromosomes. Lighter, translucent red bands are regions greater than 1Mb and less than 20Mb. Grey bands are regions larger than 10kb.

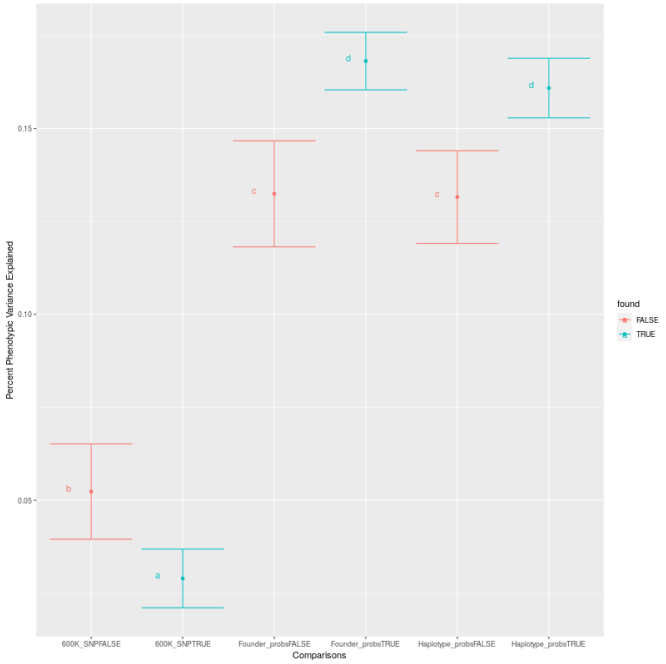

Figure 5 Percent Phenotypic Variance Explained By Detected and Undetected QTL Between Methods Color indiciates whether or not QTL were detected in that group. Letters represent significantly difference groups from a Tukey pairwise test.

effect sizes. However, for some founders the QTL_F effect sizes at *vgt1* deviated from those expections (Figure 6). Four $MITE^+$ founders, A632, F252, C103, and F492, had DTA BLUP effect size estimates later than the population average. While only one founder, F252 had a 95% confidence interval not overlapping zero, all of them had effect sizes significantly later than the other $MITE^+$ founders (pvalue). This pattern was also seen in the effect sizes estimated in individual environments (Supplemental figure). Lastly, at the most significant hit from QTL_H, founders are grouped into haplotypes consistent with their allele at the MITE, but there are still far more than 2 distinct haplotypes (12?). Analysis of the haplotype structure in the region around *vgt1* in the 16 founders showed clear differences between those that did and did not have the MITE insertion, but did not differentiate $MITE^+$ late founders from $MITE^+$ early founders (supplemental figure).

These results suggested to use that there may be an epistatic interaction between *vgt1* and the genetic background. A genome scan for epistasis between *vgt1* and other loci did not yield any significant interactions. However, QTL_F using only MAGIC lines predicted to have the MITE had two significant DTA BLUP QTL in the region of *vgt1* (figure). One of these significant sites is located in close proximity to the causal gene for *vgt2*, *ZCN8*, and is mostly explained by this linked QTL. However, the second significant site is located 15Mb downstream of *vgt1*, suggesting that some local variation around the region of *vgt1* impacts the affect of the QTL on flowering time.

Founder Representation and Linkage Disequilibrium

The MAGIC population showed an overall balance of coverage from the 16 founders (Figure 1B). Analysis of the MAGIC population showed that the overall representation of the 16 founders in the MAGIC DH lines was relatively even, with the highest percentage founder, A654, representing 6.741% and the lowest percentage founder, EP1, representing 5.237%. However, multiple chromosome regions deviated significantly from the expected equal distribution. Across individual regions of each chromosomes, 33% of founder probability sites significantly deviated from null expectations (χ^2 test p-value < 6.82e-07 (15 df)) (Supplemental Figure). The same χ^2 test performed on 100 simulated MAGIC populations of 325 individuals resulted in only 4 significant χ^2 peaks, showing that the over- and under-representation of certain founder alleles was greater than would be expected by chance. One potential reasons for this observation was that the HMM in R/qtl2 was confusing founders that were in IBD with one another. In order to ensure that these chi-squared peaks were not a result of technical error, we performed the same chi-squared test using haplotypes rather than founders. Using haplotype probabilities, XX% of sites deviated from null expectations... Together, these results suggests that the over- and under-representation of founder alleles in the MAGIC population is biological, rather than a result of model error, and perhaps evidence of selection for or against particular founder alleles.

We estimated recombination rates and linkage disequilibrium across the MAGIC lines. The founder probabilities showed levels of recombination in the MAGIC population that are consistent with expectations from the simulated MAGIC populations [I need to figure this out]. The intra-chromosomal LD structure show LD breakdown consistent with many recombination

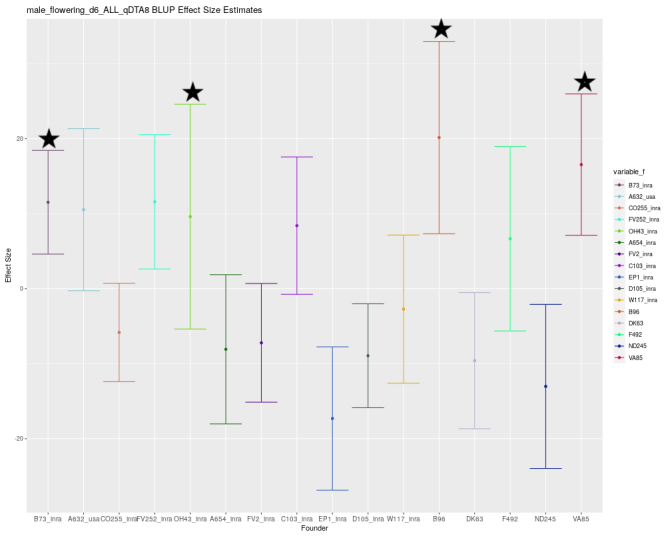

Figure 6 Estimated founder effect sizes for qDTA8 Estimates of founder effect sizes centered around the population mean for Days to Anthesis (DTA) BLUP phenotypes using QTL_F. Positive effect sizes indicate later flowering and negative effect sizes indicate earlier flowering. [Stars indicate founders that do not possess the MITE insertion ($MITE^-$)]

events. The average linkage disequilibrium R^2 50 for the population was XX. Unexpectedly, there was a large amount of high inter-chromosomal LD (4). Of a total of 9,796,630 SNP pairs with an R^2 greater than or equal to 0.9, 365,345 (3.7%) of those pairs came from different chromosomes. The number and size of inter-chromosomal high LD regions were more than would be expected by chance. In 100 simulated populations, there were no SNP pairs with R^2 greater than 0.9 detected between chromosomes. Interestingly, a large segment of interchromosomal-LD between chromosome 3 and chromosome 8 (4) overlaps with the support intervals for *qDTA8* and *qDTA3-2*, corresponding to *vgt1* *vgt2* and *vgt3*, respectively.

Evidence of Selecion on Flowering Time

We had reason to believe that selection on flowering time occurred during the making of the MAGIC population. When MAGIC DH lines were crossed to the inbred tester to produce MAGIC F1s, the DH lines needed to nick with the tester in order for crosses to be made. The tester, MBS847, is a generally later flowering line is $MITE^-$ at *vgt1* Chardon *et al.* (2004). For this reason, we would expect that, if selection on flowering time had occurred, most of the genetic variation that was selected against would have been for earlier flowering alleles. We tested to see if genes involved in flowering were enriched in χ^2 peaks, where some founderss and haplotypes were significantly over- or under-represented. However, the number of FT genes found within χ^2 peaks was not significantly different than were expected by chance (S4). Therefore, there is no evidence that the over- and under-representation of particular founders and haplotypes was due to the presence of flowering time genes within these regions.

Due to the fact that MBS847 is earlier flowering, we would expect that earlier flowering alleles would be selected against and be at a lower frequency in the resulting MAGIC F1s. The polygenic score (PGS) for DTA and DTS for the lines from the 100 simulated MAGIC populations were .. compared to the observed PGS of the actual MAGIC lines.

Discussion

We used three models of QTL allelic states to identify QTL in the MAGIC population, a bi-allelic model ($GWAS_SNP$), a founder multi-allelic model (QTL_F), and an ancestral haplotype mult-allelic model (QTL_H). The $GWAS_{SNP}$ method 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 variant. However, for multi-allelic QTL or QTL for which LD is low between tagged SNPs, this method should have lower power. QTL_F, which assumes that all founders possess distinct alleles, increases 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. The higher number of parameters that must be fit by this model may also reduce power. Lastly, QTL_H potentially improves on the power of QTL_F to detect QTL that meet the above criteria by reducing the number of parameters that must be estimated. There is, however, the potential for QTL_H to 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 QTL_H and QTL_F take into account recent recombination events, whereas $GWAS_{SNP}$ uses historical recombination, we predicted that founder and haplotype mapping would result in higher resolution of QTL support intervals. Higher resolution QTL are ideal in that they makes it easier to narrow down candidate genes and potential causal variants when the support interval is smaller.

The results of using the three models of genetic architecture to identify QTL suggest that each has its own advantages and disadvantages in terms of how many and which QTL they can identify. Overall, the QTL_F performed the best in terms of the number of QTL identified, although there were multiple QTL identified uniquely in all models. We conclude that if the goal of a study is to find as many QTL as possible, than it would be most useful to employ all three models to maximize QTL identification. The lack of significant difference in the genetic size of support intervals for the QTL found in the three methods suggests that the differences in QTL detection ability are more a result of differences in statistical power than a difference in ability to account for recombination events.

Previous studies have used variations of these methods to identify QTL, and some have directly compared them. The use of combined linkage and association analyses, sometimes referred to as linkage disequilibrium - linkage association (LDLA) was first proposed by Meuwissen and Goddard (2001), who used predicted IBD probabilities between parents using an evolutionary model and applied them to linkage mapping. LDLA has been used in multiple studies to enhance QTL detection in multi-parent populations in nmaize (Giraud *et al.* (2017); Yu *et al.* (2008); McMullen *et al.* (2009)) and other organisms Hérault *et al.* (2018). Jansen *et al.* (2003) used a haplotype-based method for QTL mapping and showed through simulation that this strategy could reduce the number of estimated parameters and, therefore, increase power. Different means of determining ancestral haplotype blocks from parental sequences have been used, with clusthaplo Leroux *et al.* (2014), an extension of the software MCQTL (Jourjon *et al.* (2004)), being a commonly used algorithm in recent studies. Bayesian frameworks have also been implemented in real (Pérez-Enciso (2003) and simulated Bink *et al.* (2012) multi-parent populations. Bardol *et al.* (2013) showed the

importance of employing different bi-allelic and multi-allelic models for detect QTL for complex traits. Giraud *et al.* (2014), for example, used both a haplotype- and founder- based approach to map in two nested association mapping populations of Northern European flint and dent maize lines created by (Bauer *et al.* 2013) genotyped with a 50K SNP array Ganal *et al.* (2011). Giraud *et al.* (2014) found used clusthaplo to determine haplotype blocks based on IBD between parents and used discrete founder and haplotype values in their models. To our knowledge, our method of converting genotype data into haplotypes, QTL_H has not been done previously. The use of the package GridLMM Runcie and Crawford (2019) allowed us to use continuous rather than discrete representations of founder and haplotype state, which allowed for the encorporation of uncertainty due to genotyping or model error into our tests for association.

Previous studies have compared similar models in different maize populations and using different methods of calculating haplotype blocks between parents (Leroux *et al.* (2014) and Giraud *et al.* (2014) Garin *et al.* (2017)). Interestingly, the performance of the three models differs across studies. Giraud *et al.* (2014) found that their haplotype model outperformed the founder and SNP models in terms of the number of QTL identified using EU-NAM Flint and Dent maize populations. In contrast, Garin *et al.* (2020) found that in the EU-NAM Flint population, the bi-allelic model detected a larger number of unique QTL, compared to parental or ancestral haplotype models.

The performance of the three models seems to depend heavily on the diversity of the parents used to generate the population. For populations with more diverse founders, it would be expected that there would be fewer shared haplotypes between founders, reducing the efficacy of a haplotype model Giraud *et al.* (2014). The fact that the QTL_F model outperformed the QTL_H and $GWAS_{SNP}$ in this population suggests that the MAGIC population contains a relatively more diverse representation of European and North American flint and dent than populations used in previous studies. It is also possible that the structure of multi-parent populations has an effect on the performance of the three models, compared to previous studies which used nested association mapping (Giraud *et al.* (2014); Garin *et al.* (2020)) and and diallel populations Bardol *et al.* (2013).

Differences in the estimated effect sizes across models offer suggestions as to the reason for their differences in QTL detection. QTL that were only found in the $GWAS_{SNP}$ method most likely have a bi-allelic causal variant. It is likely that the increased number of parameters in the QTL_F and QTL_H models reduce statistical power when the true number of functional alleles is low. Multi-allelic QTL were more likely to be identifed by the QTL_F or QTL_H models and not the $GWAS_{SNP}$ method unless the effect size was large. For QTL that were not identified by the QTL_H method, particularly for QTL that were successfully identified by QTL_F, the most likely reason is a failure of QTL_H to accurately represent the true haplotype structure of the QTL region.

The QTL_H method's assumption that founders share ancestral haplotype varied in its ability to improve QTL detection. For QTL that were identified in the QTL_H method and not the QTL_F method, the lower number of unique haplotypes suggests that QTL_H was more successful in finding these QTL due to improved power when there were fewer functional alleles than founders. The lack of significant improvement of the QTL_H method relative to the others in some cases is most likely the result of the formation of haplotypes based off of sequence information that do not accurately reflect the functional haplotypes of the founder for particular phenotypes. Grouping of two or more founders into the same haplotype that do not share the same functional alleles underlying a QTL would result in a decrease in the ability to accurately estimate haplotype effect sizes and a reduction in signal.

It should also be noted that in the process of calculating pairwise IBD and assigning founders to haplotypes requires an inherently arbitrary significance threshold for calling founders as in IBD or not. In our case, this resulted in some incomplete haplotype graphs, meaning that in some cases, most, but not all founders grouped into the same haplotype were in pairwise IBD with the other founders in the haplotype. On the whole, most QTL were found by all three methods, so there was limited ability to draw reliable conclusions about underlying mechanisms that caused the methods to perform better or worse. Generally, the comparison of QTL detection and effect size estimates suggested that the methods failed and succeeded on a QTL-by-QTL basis. This is to be expected, as each QTL is the result of a distinct causal variant with a different number of alleles within the population. Whether QTL that appeared in only one method are due to false positives or true differences in the methods' abilities to identify QTL with different genetic architectures cannot be determined.

Many of the QTL identified in the MAGIC have underlying candidate genes or have been found in previous studies, providing support to their biological reality . Multiple flowering time QTL support intervals overlap or are close by previously identifed flowering time genes and QTL. *qDTA9* is nearby the previously identified maize flowering time gene *ZmCCT9* Huang *et al.* (2018). *qDTA3-2* overlaps with *Vgt3*, whose underlying gene was identified as *ZmMADS69* Liang *et al.* (2019). *qDTA3-1* is nearby a recently identified flowering time QTL also associated with phasphatidylcholine levels Rodríguez-Zapata *et al.* (2021). The support interval for *qDTA8* overlaps with two flowering time QTL, *vgt1*, which we discuss in length, and another, *vgt2*. The causal gene for this QTL is *ZCN8*, which is the maize ortholog of FT in *Arabidopsis* Lazakis *et al.* (2011). Variation in the promoter region of *ZCN8* between temperate maize and teosinte suggests that earlier flowering alleles were under selection during the precess of maize domestication (Guo *et al.* 2018);(Bouchet *et al.* 2013). Overlap of other QTL with previously identified QTL... It is interesting to note that there is strong overlap in the support intervals of QTL found on chromosome three between flowering time and harvets grain moisture (Figure 3), perhaps due to developmental pleiotropy linking flowering time and the moisture of kernels at harvest. Three GxE QTL were detected in this population (cite Asher's paper) using the QTL_F model, but neither overlapped with the main effect QTL we detected in this study. One of these QTL appeared to be a false positive resulting from low representation of one of the founders in the region, indicating the potential for low founder sample size to confound in QTL_F.

The MAGIC population presented here provides a useful resource for investigating quantitative trait variation in temperate maize. As a multi-parent population, it has the advantages of increased diversity compared to bi-parental mapping populations, no population structure , and higher power to detect QTL with lower allele freqeuencies. In contrast, it's disadvantages include lower power for detecting epistasis. In addition, Although theoretically the design should result in equal distribution of parents, this does not always appear to be the case. The complex

crossing scheme has the potential to introduce the influence of selection. Purifying selection may actually reduce the genetic variation that we are attempting to study.

Simulations of the MAGIC population provide an opportunity to validate assignment of founder identities, as well as generate null expectations for various aspects of the population. We made our own package due to the complicated nature of the crossing scheme.

The population displayed unexpected patterns of inter-chromosomal LD and uneven founder representation. High over- or under-representation of some founder alleles may be due to these founders being in close IBD with another founder, resulting in uncertainty in the founder probablities. Another possible explanation is that there was selection against these founders during the breeding process. The high levels of inter-chromosomal LD detected in the MAGIC population deviates significantly from null expectations obtained from simulations. If we assume that the simulated populations are accurate representations of the construction of the MAGIC population without selection, this result would suggest that the differences in founder representation observed in the actual population may be due to selection for or against certain founder alleles.

Another interesting observation of the population was the relatively high levels of inter-chromosomal LD, which deviated significantly from that obtained from simulations. Inter-chromosomal LD has been detected in multiple populations of domesticated organisms, where breeding has resulted in the preservation of certain combinations of favorable alleles between chromosomes Robbins *et al.* (2010) Malysheva-Otto *et al.* (2006). Strong selection and positive or negative epistasis in natural populations have also been shown to create a pattern of inter-chromosomal LD (Kulminski (2011) Gupta *et al.* (2021) Hench *et al.* (2019) Petkov *et al.* (2005)). Both of these results suggest that forces other than those of random segregation have operated on the MAGIC population. Another potential consequence of inter-chromosomal LD is the chance for confounding of association analyses, namely resulting in the detection of "ghost" QTL. Although for the discussed QTL detected that were in LD with one another, *qDTA8* and *qDTA3-2*, their effects on flowering time were independent, the chance for false positives and innacurate support intervals due to LD structure is still worth noting.

There was reason to suspect that there had been purifying selection on flowering time over the course of the creation of the MAGIC population. However, there was no enrichment of flowering time genes in regions where some founders were more or less represented than expected by chance. Polgenic scores of flowering time compared to simulations...

Due to the fact that it is segregating for *vgt1*, this population provides an opportunity to further study the mechanism behind the QTL's affect on flowering time.

Flowering time is a highly important trait both from a breeding and evolutionary standpoint. In addition to being a crucial agronomic trait, it contributes to local adaptation for annual plants such as maize, ensuring that individuals can reproduce within the growing season of their environments [cite?]. Flowering time in maize has been shown to be a highly polygenic trait controlled by many, mostly small-effect loci Buckler *et al.* (2009), although some larger effect QTL have been identified. Once such QTL is *vgt1*, located on chromosome 8 Salvi *et al.* (2007). The relatively large effect of *vgt1* compared to most other identified maize flowering time QTL has made it a target of a great deal of study, and it has been identified in multiple populations [cite]. Previous research has shown that variation in flowering time at this site is strongly correlated with a MITE insertion about 70kb upstream of the flowering time regulator, *ZmRAP2.7*, an *APETALA*-like transcription factor, with the presence of the MITE associated with an earlier flowering time Castelletti *et al.* (2014). Within maize heterotic groups, Flint maize lines tend to possess the early-flowering allele of *vgt1* (*MITE+*), while dents (such as B73) tend to carry the late-flowering allele(*MITE-*) [cite?]. The frequency of the MITE in maize populations follows a latitudinal gradient, suggesting that the early allele was selected for during the process of maize adaptation to temperate climates Navarro *et al.* (2017). As a negative regulator of flowering time, this reduced expression results in an earlier flowering phenotype []. It has also been shown that there a differentially-methylated regions between B73, landrace maize, and teosinte Xu *et al.* (2020). The hypothesized method, then, is that the MITE represses expression of *ZmRAP2.7*, possibly due to change in methylation around the insertion, resulting in earlier flowering Castelletti *et al.* (2014). However, the MITE has not been experimentally shown to cause a decrease in *ZmRap2.7* expression and results in earlier flowering. Further, there has been some evidence of epistatic interactions, potentially with another flowering time QTL, *vgt3* which can impact the effect of the QTL (Alain Charcosset, personal communication). However, a recent study using multiple multi-parent populations suggested that variation in the effect of *vgt1* in different genetic backgrounds was due to local genetic variation surrounding *vgt1*, rather than epistasis with distant loci (Rio *et al.* 2020). This finding suggests two possibilities: either (1) that the causal variant underlying *vgt1* is some as-yet unidentified variant that is in tight, but imperfect linkage disequilibrium with the MITE insertion, or (2) that the MITE insertion directly impact flowering time, and that another variant nearby has a modifiying effect on the MITE. *ZCN8* functions somewhere downstream of *ZmRAP2.7*, and is negatively regulated by it. Variation in the promoter region of *ZCN8* between temperate maize and teosinte suggests that earlier flowering alleles were under selection during the precess of maize domestication (Guo *et al.* 2018);(Bouchet *et al.* 2013) One benefit of using founder and haplotype approaches lies in the potential to dissect the effects of individuals founders and/or haplotypes within QTL. This allowed us to look more closely at a well-characterized, large effect flowering time QTL, *vgt1* and observe an interesting pattern of effect sizes that deviated from our expectations based on previous research. The results of $MITE^{+}$ QTL_{F} are consistent with previous research that showed that local genomic differences around *vgt1* affect allelic effects in an admixed population of Flint and Dent lines (Rio *et al.* (2020)). This opens up new areas of inquiry for future studies.

Conclusion

Acknowledgments

Limagrain gets authorship, no need to include here. Please cite the NSF GF grant for some of your funding.

References

Aulchenko, Y. S., D.-J. de Koning, and C. Haley, 2007 Genomewide rapid association using mixed model and regression: A fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics **177**: 577–585.

Aylor, D. L., W. Valdar, W. Foulds-Mathes, R. J. Buus, R. A. Verdugo, *et al.*, 2011 Genetic analysis of complex traits in the emerging collaborative cross. Genome Research **21**: 1213–1222.

Bardol, N., M. Ventelon, B. Mangin, S. Jasson, V. Loywick, *et al.*, 2013 Combined linkage and linkage disequilibrium qtl mapping in multiple families of maize (zea mays l.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism. Theoretical and applied genetics **126**: 2717–2736.

Bauer, E., M. Falque, H. Walter, C. Bauland, C. Camisan, *et al.*, 2013 Intraspecific variation of recombination rate in maize. Genome Biology **14**: R103.

Beavis, W. D., D. Grant, M. Albertsen, and R. Fincher, 1991 Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. Theoretical and Applied Genetics **83**: 141–145.

Bink, M. C., L. R. Totir, C. J. ter Braak, C. R. Winkler, M. P. Boer, *et al.*, 2012 Qtl linkage analysis of connected populations using ancestral marker and pedigree information. Theoretical and Applied Genetics **124**: 1097–1113.

Bouchet, S., B. Servin, P. Bertin, D. Madur, V. Combes, *et al.*, 2013 Adaptation of maize to temperate climates: mid-density genome-wide association genetics and diversity patterns reveal key genomic regions, with a major contribution of the vgt2 (zcn8) locus. PLoS One **8**.

Broman, K. W., D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, *et al.*, 2019 R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multiparent populations. Genetics **211**: 495–502.

Browning, B. L. and S. R. Browning, 2013 Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics **194**: 459–471.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, *et al.*, 2009 The genetic architecture of maize flowering time. Science **325**: 714–718.

Castelletti, S., R. Tuberosa, M. Pindo, and S. Salvi, 2014 A mite transposon insertion is associated with differential methylation at the maize flowering time qtl vgt1. G3: Genes, Genomes, Genetics **4**: 805–812.

Chardon, F., B. Virlon, L. Moreau, M. Falque, J. Joets, *et al.*, 2004 Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics **168**: 2169–2185.

Dell'Acqua, M., D. M. Gatti, G. Pea, F. Cattonaro, F. Coppens, *et al.*, 2015 Genetic properties of the magic maize population: a new platform for high definition qtl mapping in zea mays. Genome Biology **16**: 167.

Ganal, M. W., G. Durstewitz, A. Polley, A. Bérard, E. S. Buckler, *et al.*, 2011 A large maize (zea mays l.) snp genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the b73 reference genome. PLOS ONE **6**: e28334.

Garin, V., M. Malosetti, and F. van Eeuwijk, 2020 Multi-parent multi-environment qtl analysis: an illustration with the eu-nam flint population. Theoretical and Applied Genetics **133**: 2627–2638.

Garin, V., V. Wimmer, S. Mezmouk, M. Malosetti, and F. van Eeuwijk, 2017 How do the type of qtl effect and the form of the residual term influence qtl detection in multi-parent populations? a case study in the maize eu-nam population. Theoretical and Applied Genetics **130**: 1753–1764.

Giraud, H., C. Bauland, M. Falque, D. Madur, V. Combes, *et al.*, 2017 Linkage analysis and association mapping qtl detection models for hybrids between multiparental populations from two heterotic groups: Application to biomass production in maize (zea mays l.). G3 Genes | Genomes | Genetics **7**: 3649–3657.

Giraud, H., C. Lehermeier, E. Bauer, M. Falque, V. Segura, *et al.*, 2014 Linkage disequilibrium with linkage analysis of multiline crosses reveals different multiallelic qtl for hybrid performance in the flint and dent heterotic groups of maize. Genetics **198**: 1717–1734.

Guo, L., X. Wang, M. Zhao, C. Huang, C. Li, *et al.*, 2018 Stepwise cis-regulatory changes in zcn8 contribute to maize flowering-time adaptation. Current Biology **28**: 3005–3015.e4.

Gupta, S., A. Harkess, A. Soble, M. Van Etten, J. Leebens-Mack, *et al.*, 2021 Inter-chromosomal linkage disequilibrium and linked fitness cost loci influence the evolution of nontarget site herbicide resistance in an agricultural weed. bioRxiv p. 2021.04.04.438381.

Hench, K., M. Vargas, M. P. Hœppner, W. O. McMillan, and O. Puebla, 2019 Inter-chromosomal coupling between vision and pigmentation genes during genomic divergence. Nature Ecology Evolution **3**: 657–667.

Hérault, F., M. Damon, P. Cherel, and P. Le Roy, 2018 Combined gwas and ldla approaches to improve genome-wide quantitative trait loci detection affecting carcass and meat quality traits in pig. Meat Science **135**: 148–158.

Highfill, C. A., G. A. Reeves, and S. J. Macdonald, 2016 Genetic analysis of variation in lifespan using a multiparental advanced intercross drosophila mapping population. BMC Genetics **17**: 113.

Huang, B. E., A. W. George, K. L. Forrest, A. Kilian, M. J. Hayden, *et al.*, 2012 A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnology Journal **10**: 826–839.

Huang, B. E., K. L. Verbyla, A. P. Verbyla, C. Raghavan, V. K. Singh, *et al.*, 2015 Magic populations in crops: current status and future prospects. Theoretical and Applied Genetics **128**: 999–1017.

Huang, C., H. Sun, D. Xu, Q. Chen, Y. Liang, *et al.*, 2018 *ZmCCT9* enhances maize adaptation to higher latitudes. Proceedings of the National Academy of Sciences **115**: E334–E341.

Jansen, R. C., J.-L. Jannink, and W. D. Beavis, 2003 Mapping quantitative trait loci in plant breeding populations. Crop Science **43**: 829–834.

Jiao, Y., P. Peluso, J. Shi, T. Liang, M. C. Stitzer, *et al.*, 2017 Improved maize reference genome with single-molecule technologies. Nature **546**: 524.

Jourjon, M.-F., S. Jasson, J. Marcel, B. Ngom, and B. Mangin, 2004 Mcqtl: multi-allelic qtl mapping in multi-cross design. Bioinformatics **21**: 128–130.

Kover, P. X., W. Valdar, J. Trakalo, N. Scarcelli, I. M. Ehrenreich, *et al.*, 2009 A multiparent advanced generation inter-cross to fine-map quantitative traits in arabidopsis thaliana. PLoS genetics **5**.

Kulminski, A. M., 2011 Complex phenotypes and phenomenon of genome-wide inter-chromosomal linkage disequilibrium in the human genome. Experimental Gerontology **46**: 979–986.

Lazakis, C. M., V. Coneva, and J. Colasanti, 2011 Zcn8 encodes a potential orthologue of arabidopsis ft florigen that integrates both endogenous and photoperiod flowering signals in maize. Journal of Experimental Botany **62**: 4833–4842.

Leroux, D., A. Rahmani, S. Jasson, M. Ventelon, F. Louis, *et al.*, 2014 Clusthaplo: a plug-in for mcqtl to enhance qtl detection using ancestral alleles in multi-cross design. Theoretical and Applied Genetics **127**: 921–933.

Liang, Y., Q. Liu, X. Wang, C. Huang, G. Xu, *et al.*, 2019 Zmmads69 functions as a flowering activator through the zmrap2.7-zcn8 regulatory module and contributes to maize flowering time adaptation. New Phytologist **221**: 2335–2347.

Malysheva-Otto, L. V., M. W. Ganal, and M. S. Rœder, 2006 Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (hordeum vulgare l.). BMC Genetics **7**: 6.

McMullen, M. D., S. Kresovich, H. S. Villeda, P. Bradbury, H. Li, *et al.*, 2009 Genetic properties of the maize nested association mapping population. Science **325**: 737–740.

Meuwissen, T. H. and M. E. Goddard, 2001 Prediction of identity by descent probabilities from marker-haplotypes. Genetics Selection Evolution **33**: 605.

Navarro, J. A. R., M. Willcox, J. Burgueño, C. Romay, K. Swarts, *et al.*, 2017 A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nature genetics **49**: 476.

Ogut, F., Y. Bian, P. J. Bradbury, and J. B. Holland, 2015 Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. Heredity **114**: 552–563.

Pascual, L., N. Desplat, B. E. Huang, A. Desgroux, L. Bruguier, *et al.*, 2015 Potential of a tomato magic population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. Plant Biotechnology Journal **13**: 565–577.

Pérez-Enciso, M., 2003 Fine mapping of complex trait genes combining pedigree and linkage disequilibrium information: A bayesian unified framework. Genetics **163**: 1497–1510.

Petkov, P. M., J. H. Graber, G. A. Churchill, K. DiPetrillo, B. L. King, *et al.*, 2005 Evidence of a large-scale functional organization of mammalian chromosomes. PLOS Genetics **1**: e33.

R Core Team, 2017 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rio, S., T. Mary-Huard, L. Moreau, C. Bauland, C. Palaffre, *et al.*, 2020 Disentangling group specific qtl allele effects from genetic background epistasis using admixed individuals in gwas: an application to maize flowering. PLoS genetics **16**: e1008241.

Robbins, M. D., S.-C. Sim, W. Yang, A. Van Deynze, E. van der Knaap, *et al.*, 2010 Mapping and linkage disequilibrium analysis with a genome-wide collection of snps that detect polymorphism in cultivated tomato. Journal of Experimental Botany **62**: 1831–1845.

Rodríguez-Zapata, F., A. C. Barnes, K. A. Blœcher-Juárez, D. Gates, A. Kur, *et al.*, 2021 Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time. bioRxiv p. 2021.01.25.426574.

Runcie, D. E. and L. Crawford, 2019 Fast and flexible linear mixed models for genome-wide genetics. PLOS Genetics **15**: e1007978.

Salvi, S., G. Sponza, M. Morgante, D. Tomes, X. Niu, *et al.*, 2007 Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proceedings of the National Academy of Sciences **104**: 11376–11381.

Steinhoff, J., W. Liu, J. C. Reif, G. Della Porta, N. Ranc, *et al.*, 2012 Detection of qtl for flowering time in multiple families of elite maize. Theoretical and applied genetics **125**: 1539–1551.

Unterseer, S., E. Bauer, G. Haberer, M. Seidel, C. Knaak, *et al.*, 2014 A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k snp genotyping array. BMC Genomics **15**: 823.

Wallace, J. G., P. J. Bradbury, N. Zhang, Y. Gibon, M. Stitt, *et al.*, 2014 Association mapping across numerous traits reveals patterns of functional variation in maize. PLOS Genetics **10**: e1004845.

Wang, L., T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, *et al.*, 2017 The interplay of demography and selection during maize domestication and expansion. Genome Biology **18**: 215.

Wang, Y., J. Yao, Z. Zhang, and Y. Zheng, 2006 The comparative analysis based on maize integrated qtl map and meta-analysis of plant height qtls. Chinese Science Bulletin **51**: 2219–2230.

Xu, G., J. Lyu, Q. Li, H. Liu, D. Wang, *et al.*, 2020 Evolutionary and functional genomics of dna methylation in maize domestication and improvement. Nature Communications **11**: 5539.

Yu, J., J. B. Holland, M. D. McMullen, and E. S. Buckler, 2008 Genetic design and statistical power of nested association mapping in maize. Genetics **178**: 539–551.

Ziyatdinov, A., M. Vázquez-Santiago, H. Brunel, A. Martinez-Perez, H. Aschard, *et al.*, 2018 lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. BMC Bioinformatics **19**: 68.

Supplement

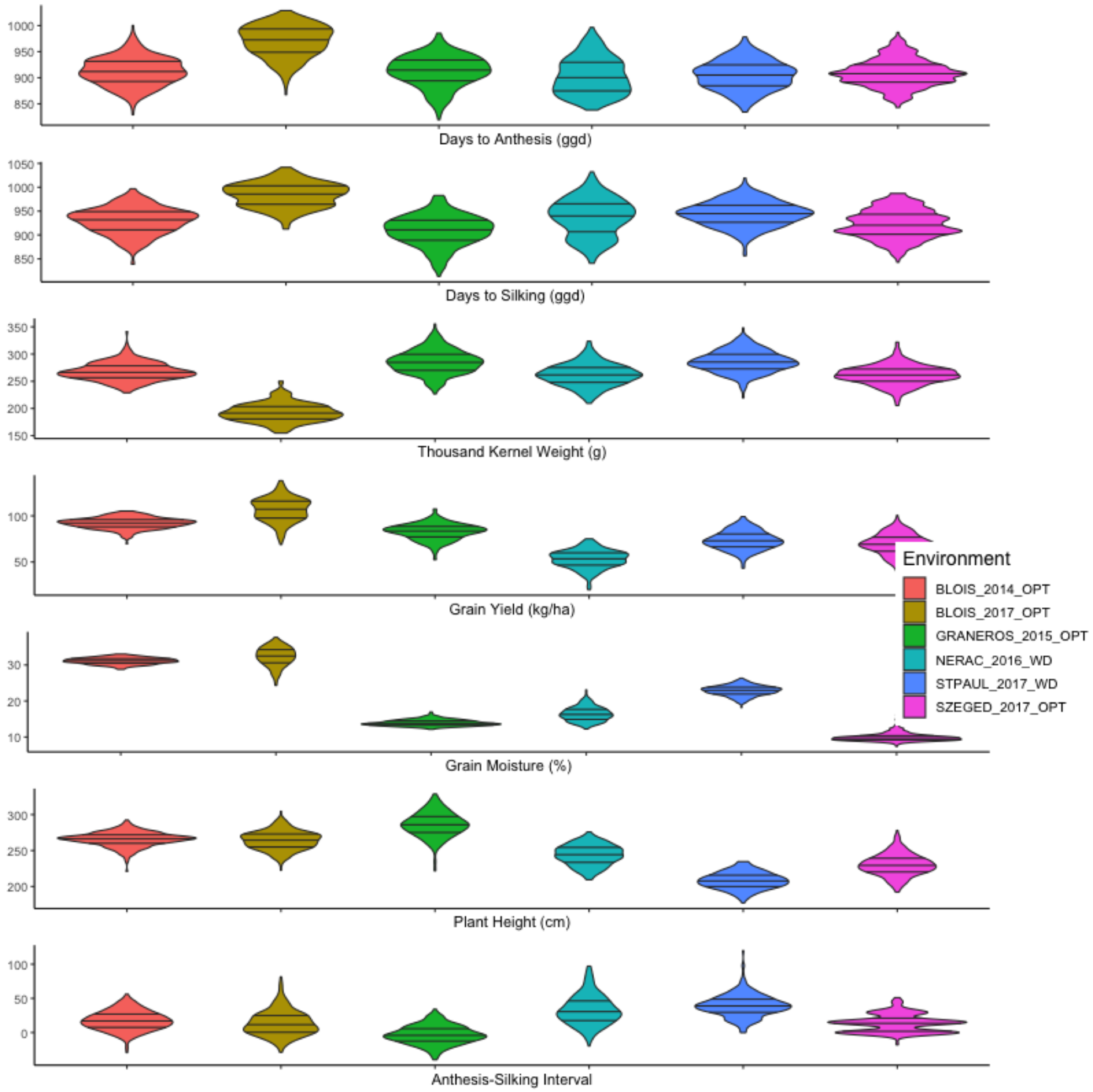

Figure S1 Distribution of phenotypes The density plots of the six measured phenotypes. The vertical bar represents the mean, and the grey shading shows two standard deviations. The heritability of each trait is shown on the right.

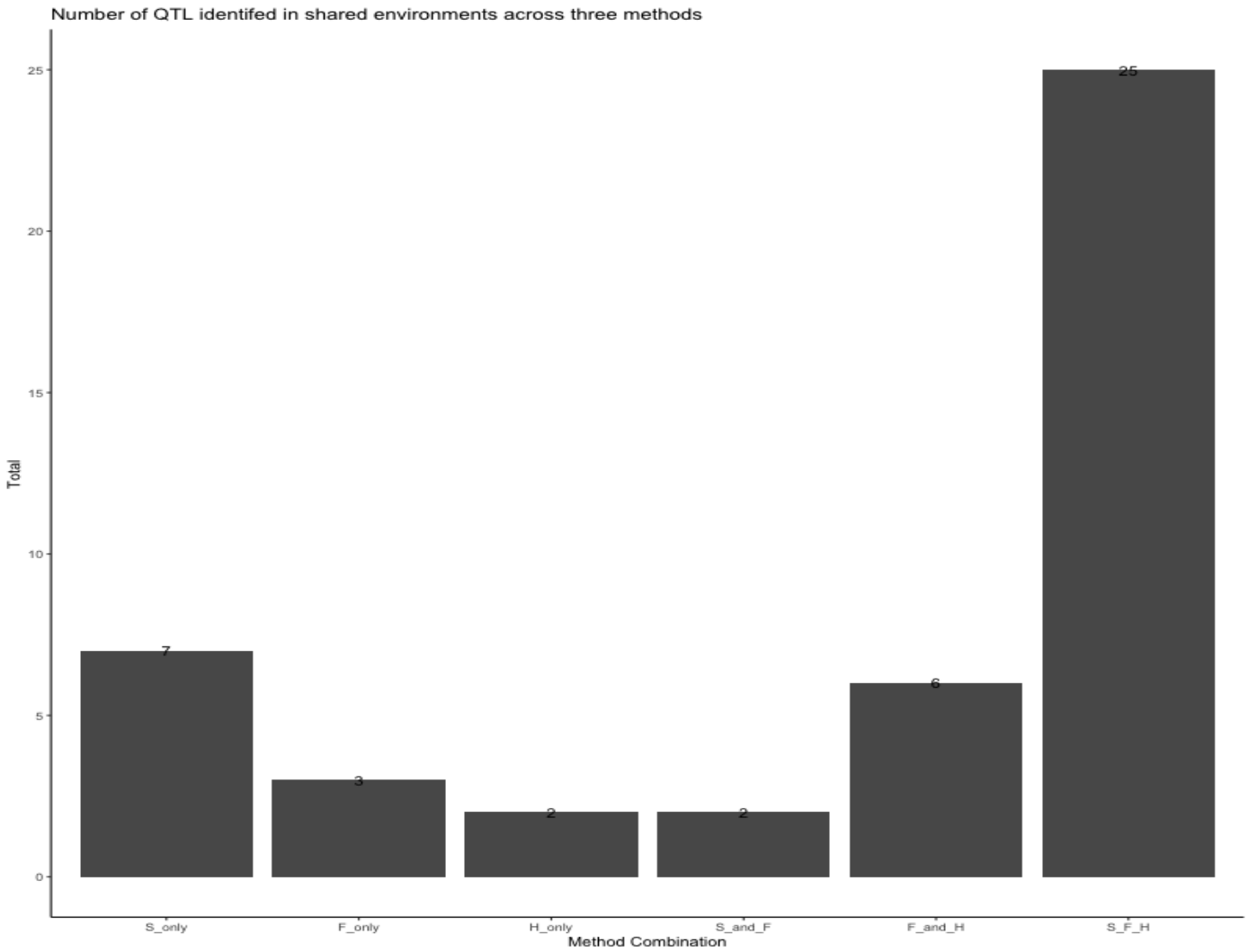

Figure S2 Number of Environment-Specific QTL Found by Method S: $GWAS_{SNP}$, F: QTL_F, H: QTL_H. Environment-specific QTL are QTL for a phenotype that were identified in one environment. QTL were called as found by multiple methods if their support intervals overlapped (see Methods).

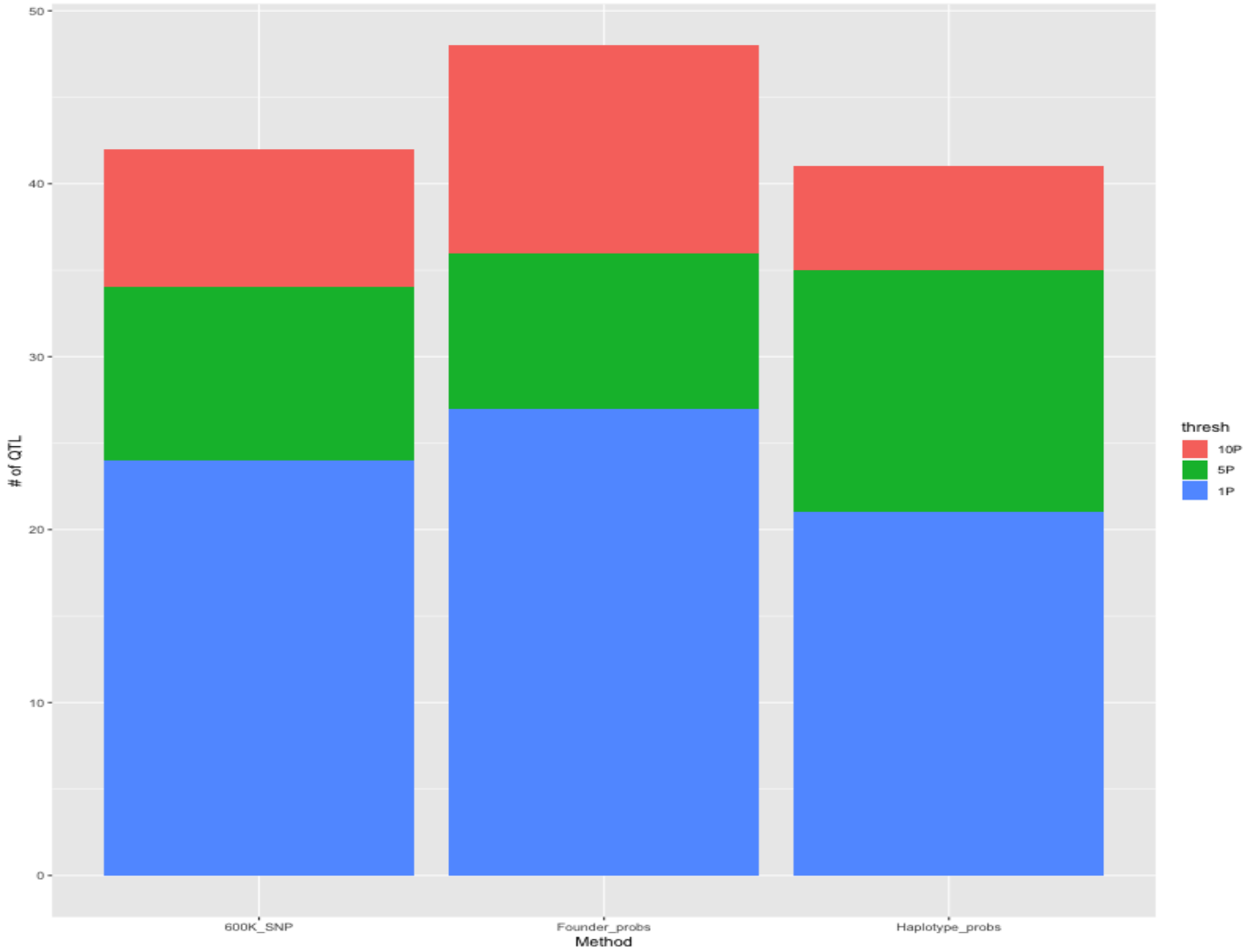

Figure S3 Total Number of Environment-QTL Found By Method By Significance Threshold S: $GWAS_{SNP}$, F: QTL_F, H: QTL_H. Environment-specific QTL are QTL for a phenotype that were identified in one environment. QTL were called as found by multiple methods if their support intervals overlapped (see Methods).

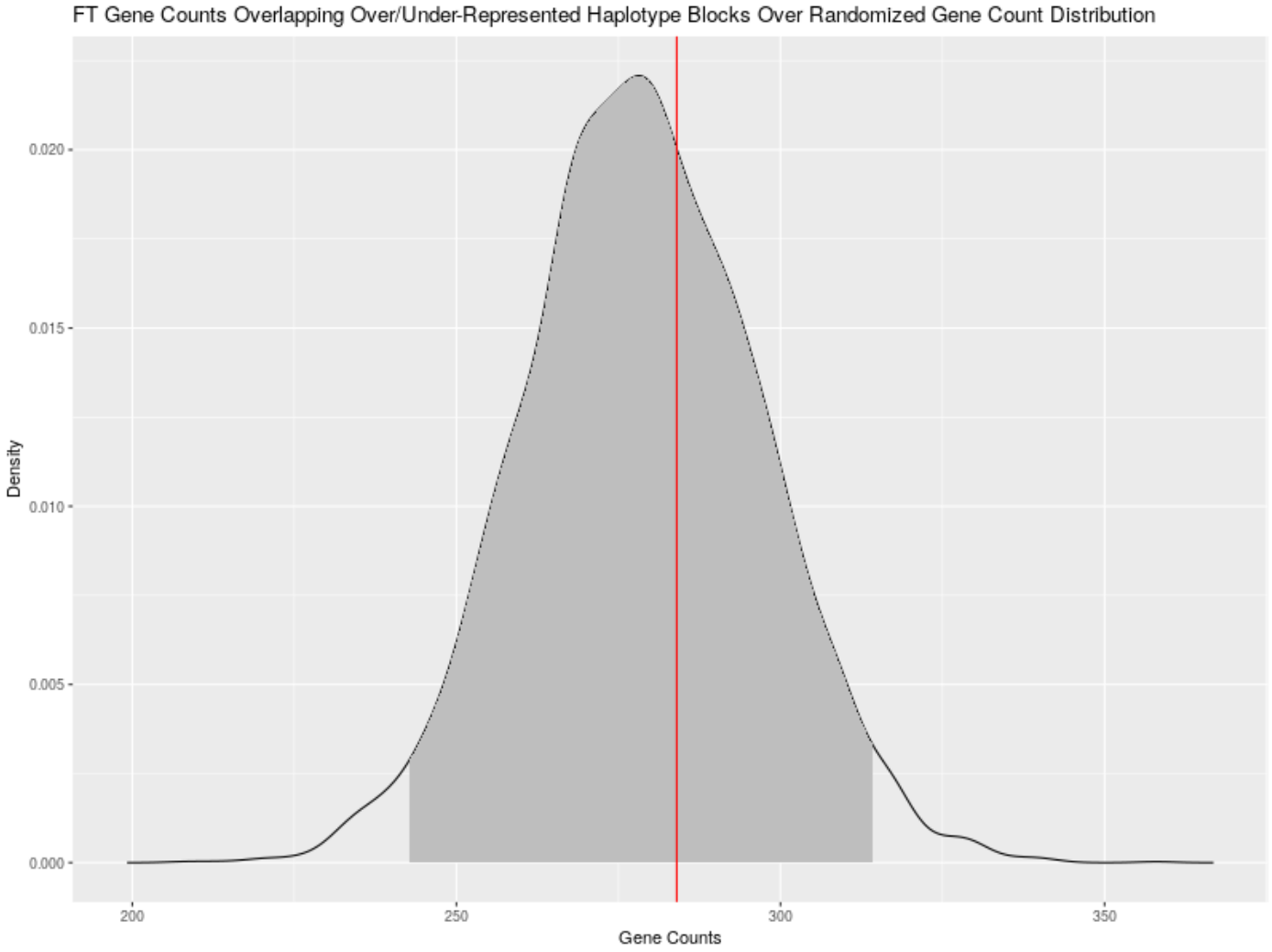

Figure S4 Enrichment of Flowering Time Genes in χ^2 Peaks Density plot of 1,000 permutations of number of 904 randomly selected genes that overlap with χ^2 peaks for over- or under-representation of haplotypes in the MAGIC population. Red line indicates the actual number (284) of FT genes from the list of 904 that overlapped with χ^2 peaks.