Journal of Genetics and Genomics 46 (2019) 343–352

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Original research

Mapping quantitative trait loci using binned genotypes

Wen Yao [a, c, 1], Guangwei Li [a, 1], Yanru Cui [d], Yiming Yu [a], Qifa Zhang [a, *], Shizhong Xu [b, *]

[a] National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan, 430070, China
[b] Department of Botany and Plant Sciences, University of California Riverside, Riverside, CA, 92521, USA
[c] National Key Laboratory of Wheat and Maize Crop Science, College of Life Sciences, Henan Agricultural University, Zhengzhou, 450002, China
[d] College of Agronomy, Hebei Agricultural University, Baoding, 071001, China

ARTICLE INFO

Article history:
Received 31 December 2018
Received in revised form
16 June 2019
Accepted 21 June 2019
Available online 23 July 2019

Keywords:
Genome-wide association studies
Linear mixed model
Polygene
Proximal contamination
QTL mapping

ABSTRACT

Precise mapping of quantitative trait loci (QTLs) is critical for assessing genetic effects and identifying candidate genes for quantitative traits. Interval and composite interval mappings have been the methods of choice for several decades, which have provided tools for identifying genomic regions harboring causal genes for quantitative traits. Historically, the concept was developed on the basis of sparse marker maps where genotypes of loci within intervals could not be observed. Currently, genomes of many organisms have been saturated with markers due to the new sequencing technologies. Genotyping by sequencing usually generates hundreds of thousands of single nucleotide polymorphisms (SNPs), which often include the causal polymorphisms. The concept of interval no longer exists, prompting the necessity of a norm change in QTL mapping technology to make use of the high-volume genomic data. Here we developed a statistical method and a software package to map QTLs by binning markers into haplotype blocks, called bins. The new method detects associations of bins with quantitative traits. It borrows the mixed model methodology with a polygenic control from genome-wide association studies (GWAS) and can handle all kinds of experimental populations under the linear mixed model (LMM) framework. We tested the method using both simulated data and data from populations of rice. The results showed that this method has higher power than the current methods. An R package named binQTL is available from GitHub.

1. Introduction

Quantitative trait locus (QTL) mapping refers to a technology aiming to detect associations between molecular markers and quantitative traits. Associated markers most likely harbor QTL in their neighborhoods of the genome. If locations of the associated markers are given in the genome, the map positions of QTLs are then roughly known. Such a marker-trait association study is the prototype of interval mapping (IM) (Mackay et al., 2009). Prior to the genome era, genetic markers were often sparsely distributed across the genome and marker-trait association studies were incapable of pinpointing QTL residing between two consecutive markers. Using the two markers as anchors, Lander and Botstein (1989) developed a mixture model maximum likelihood method to investigate associations of all candidate positions between the two flanking markers with a quantitative trait. The method is called IM because the two flanking markers define an interval within which association of a locus with a trait is tested.

The IM procedure is implemented with a single locus model, i.e., only one locus is tested at a time and the entire genome is tested m times for m candidate positions of the entire genome. Quantitative traits, by definition, are controlled by multiple loci. The single locus model of IM has serious limitation because other loci not included in the model will interfere with the test and inflate the residual variance, leading to biased estimates of QTL effects and incorrect test statistics (Zeng, 1993). To address this issue, Zeng (1994) proposed a composite interval mapping (CIM) procedure that includes selected markers (cofactors) outside the tested interval to reduce the interference and correct the residual variance. The CIM method and various modified versions are now the dominant methods for

* Corresponding authors.
E-mail addresses: qifazh@mail.hzau.edu.cn (Q. Zhang), shizhong.xu@ucr.edu (S. Xu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jgg.2019.06.005

QTL mapping (Li et al., 2007). The multiple interval mapping (MIM) procedure proposed later (Kao et al., 1999) is a further improvement over CIM but has not become the main tool for QTL mapping because of its high computational cost and instability when the number of intervals becomes large (Mayer, 2005). Although CIM and its variants are widely adopted in current QTL mapping studies, they often do not offer the resolution for small and linked QTLs. To address this issue, Xu (2013a, b) and Bernardo (2013) adopted the concept of "kinship matrix" as used in genome-wide association studies (GWAS) to control polygenic background in QTL mapping. Recently, Wang et al. (2016) and Wen et al. (2018) developed a genome-wide composite interval mapping (GCIM) method for detection of small and linked QTLs by integrating polygenic background control with a multi-locus genetic model.

With the high-throughput genotyping technology, genomes of most agriculturally important organisms have been saturated with various types of markers. Genotyping by sequencing, or population sequencing, has now become the method of choice, which usually produces hundreds of thousands of markers for a mapping population (Huang et al., 2009; Xie et al., 2010). For a population derived from a cross between two inbred lines, population sequencing generates many haplotype blocks consisting of thousands of SNPs and InDels, called bins (Huang et al., 2009). Within each bin, all markers segregate with exactly the same pattern and one marker from a bin captures all information about the bin. With the high-quality reference genome sequence, it is now feasible to identify causal polymorphisms of QTL by population sequencing if the population is sufficiently large, especially with the help of other genomic information such as genome annotation and transcriptomes. To achieve such a goal, it requires a method of QTL mapping that is able to pinpoint the candidate bin harboring the causal polymorphism (SNP or InDel) rather than suggests an interval bracketed by two markers to infer the genotypes of the candidate region. Thus, the concept of IM no longer exists and the method of QTL mapping should evolve accordingly to address the new goal.

A prototype of the bin model has been developed (Xu, 2013a), but it only deals with mapping populations with two alternative genotypes per marker, e.g., backcross (BC) or recombinant inbred lines (RILs). Many new types of mapping populations have been developed in recent years including MAGIC (multi-parent advanced generation inter-cross) populations, NAM (nested association mapping) populations and many other types of experimental populations (Yu et al., 2008; Huang et al., 2015). The traditional method dealing with multiple genotypes per locus is the analysis of variance (ANOVA), first adopted by Fisher (Fisher, 1918; Pillen et al., 2003). However, ANOVA lacks a mechanism to control polygenic background.

In this study, we developed a new method for QTL mapping based on bins generated from population sequencing (binQTL). We adopted a random model methodology to map QTLs using bin genotype data from populations with arbitrary number of genotypes per locus. We also adopted the concept of "kinship matrix" as used in GWAS to control polygenic background in QTL mapping. We demonstrated the superiority of the new method using simulated and real data in comparison with ANOVA, CIM and GCIM.

2. Results

2.1. Simulation studies of the immortalized F_2 (IMF2) population

We simulated a hypothetical quantitative trait using the genetic map and genotypes of 1619 bins from the IMF2 population of 278 hybrids obtained by pairwise crossing of RILs derived from the cross between Zhenshan 97 and Minghui 63, the parents of the most popular rice hybrid Shanyou 63 in China. Positions and effects of the 20 simulated QTLs are given in Table S1, and the design of the simulation experiments is described in Supplementary Note S1. From this simulated sample, we performed QTL mapping using our new method (binQTL), in comparison with ANOVA, CIM and GCIM (Fisher, 1918; Zeng, 1994; Broman et al., 2003; Wen et al., 2018). The QTL effects estimated by different methods are illustrated in Fig. 1 along with the simulated "true" effects, showing the behaviors of these methods anticipated in real data analysis. Ideally, the estimated effects for the 20 bins that contain the simulated QTLs should be close to the "true" effects while the estimated effects from the rest of the bins should be close to zero. Thus, the sum of absolute differences between the estimated effects and the "true" effects across all bins provide a good criterion for evaluating different methods, the smaller the better.

The positions and effects of QTLs estimated by binQTL matched the simulated positions and effects very well as the sum of absolute differences between the estimated values of binQTL and the "true" values across all bins was 449.0, much smaller than those of ANOVA (1855.7) and CIM (1287.4) (Fig. 1A–C). Of the 20 simulated QTLs, 11 had absolute differences between the true and estimated effects smaller than 1 for binQTL; the corresponding numbers were 10 and 14 for ANOVA and CIM, respectively. Furthermore, 1200 bins out of the 1599 non-QTL bins had absolute differences between the true and estimated QTL effects smaller than 0.5 from the binQTL method, while there were only 59 and 110 non-QTL bins, respectively, had absolute differences smaller than 0.5 from the ANOVA and CIM methods (Fig. 1A–C). Such comparisons indicate that the binQTL method provides more precise estimates of QTL effects and, at the same time, has a better control of the genetic background effects than the ANOVA and CIM methods. As GCIM only outputs the estimated effects of a few predicted QTLs represented by independent markers rather than the estimated effects of all markers, we were unable to compare the performance of GCIM with the first three methods across markers of the whole genome. Of the 20 simulated QTLs, only 4 were detected by GCIM, while 7 other bins detected by this method were in the neighborhoods of simulated QTLs but outside of the assumed bins (±5 bins). Seven of the 11 QTLs detected by GCIM had absolute differences between the simulated "true" effects and the estimated effects smaller than 1 (Fig. 1D).

The LOD score test statistics from the same simulated sample are shown in Fig. 1. The binQTL method (Fig. 1A) often had lower LOD score than the ANOVA method (Fig. 1B) because the former used an efficient mechanism to control polygenic background. Of the 20 bins with simulated QTLs, 5, 8 and 5 bins passed the empirical threshold values for binQTL, ANOVA and CIM, respectively, where the thresholds were determined from 1000 simulated samples under the null hypothesis, i.e., no QTL effects existed in any bins (Table S2; see Materials and methods). However, among the 1599 bins without simulated QTLs, the binQTL method produced 44 bins that passed the threshold value, the ANOVA method showed 169 bins with LOD scores greater than the threshold value, and the CIM method had 104 bins with LOD scores passing the threshold (Fig. 1A–C). This result implies that the binQTL method has the lowest Type I error (false positive rate) compared with the ANOVA and CIM methods. For the GCIM method, there were 9 bins with LOD scores greater than 2.5 (the default threshold set by the GCIM program). We were unable to determine the LOD score threshold for the GCIM method from the 1000 simulated samples under the null model, due to the fact that the GCIM program only outputs the LOD values of a few detected bins, not from all bins of the entire genome.

We also simulated a large sample for the IMF2 population by randomly selecting 1000 hybrids from all $210 \times (210-1)/2 = 21{,}945$

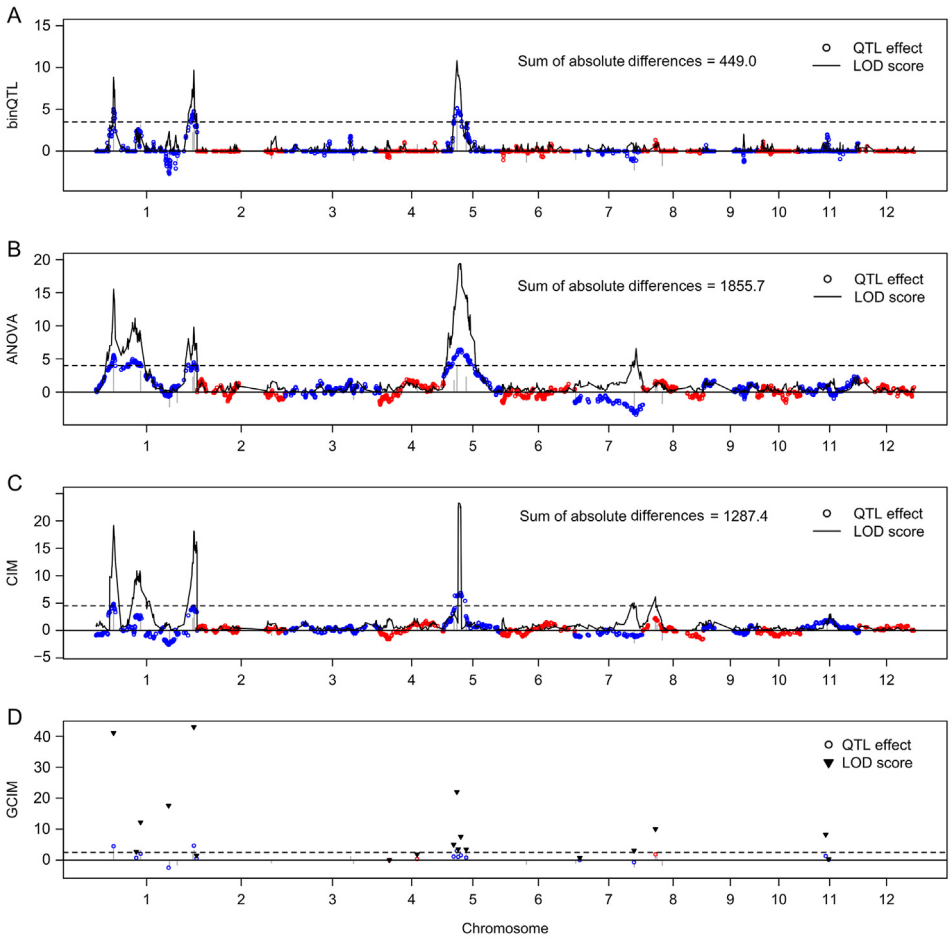

Fig. 1. QTL effects and LOD scores from a simulated IMF2 population. The grey vertical lines in all panels represent true effects of 20 simulated QTLs; the circles represent positions and effects of QTLs estimated by four different methods. The black curves in **A**–**C** represent LOD scores plotted against genome locations of bins for binQTL, ANOVA and CIM, respectively; the black triangles in **D** represent the LOD scores of detected QTLs by GCIM. The horizontal dashed lines in **A**–**C** indicate the LOD score cutoffs determined from 1000 samples simulated under the null model (no QTL effects in the bins) for binQTL, ANOVA and CIM, respectively; the horizontal dashed line in **D** indicates the default LOD score cutoff (2.5) of GCIM. The sum of absolute differences between the true QTL effects and the estimated QTL effects across all bins are shown for each method in each panel. Adjacent chromosomes are separated by alternated colors (red *vs.* blue). X-axis shows the position of each marker (bin) in the genetic map.

potential hybrids. Genotypes of the hybrids were inferred from the inbred parents. QTL mapping was conducted using binQTL, ANOVA and CIM. The estimated QTL effects from one random draw of a hybrid population (consisting of 1000 hybrids) are depicted in Fig. S1 along with the "true" effects for comparison. With such a large sample, the binQTL method had substantially high resolution and generated almost unbiased estimates of QTL effects compared with the true values. The sum of absolute differences between the estimated and true effects across all bins of the entire genome was 431.1 for the binQTL method, while the corresponding sums of absolute differences for the ANOVA and CIM methods were 2489.1 and 1297.4, respectively.

We further performed power (true positive rate) and Type I error analyses using 1000 random draws each with 1000 hybrids of the simulated IMF2 population. Of the 20 simulated QTLs, 8 were detected by at least one of the four methods using the simulated LOD thresholds (binQTL, ANOVA and CIM) or default LOD threshold (GCIM). The effects of the 8 QTLs were larger than those of the other 12 QTLs except for QTL-3, which was not detected by any method. This was probably caused by the negative effect (−2.24) of the allele linked in repulsing phase with the two QTLs (QTL-1 and QTL-2) of large positive effect. Table 1 showed the powers of the 8 QTLs detected by the four methods and the genome-wide Type I errors. Due to linkage disequilibrium (LD), bins nearby a simulated QTL were often detected as being significant. Therefore, we set a window around each simulated QTL. If any bins within the window were detected, this QTL was considered to be significant. Significant bins outside these windows were counted as Type I errors. We reported the powers and Type I errors under six different window sizes (i.e., $\pm i$ bin, where $i = 0, 1, \cdots, 5$). Overall, the ANOVA method had the highest power followed by CIM and binQTL, where binQTL had higher power than CIM for some QTLs but worse for other QTLs (see Table 1). However, the higher powers of ANOVA and CIM were achieved with higher Type I errors. Neither ANOVA nor CIM controlled the Type I errors under the assumed 0.05 level, but binQTL controlled the Type I errors well below the 0.05 level. The Type I errors of GCIM were the lowest of all methods as GCIM only output the LOD values of a few detected QTLs rather than from all markers (Table 1). This feature of GCIM led to the lowest power of GCIM when the window size was small, although the power of GCIM increased quickly when the window size was increased.

A fair comparison of powers for different methods should be made under the same Type I error. Therefore, we drew receiver operating characteristic (ROC) curves for the three methods (binQTL, ANOVA and CIM). An ROC curve is a plot of power against Type I error. Fig. 2 showed the ROC curves of the binQTL, ANOVA

Table 1
Statistical powers and Type I errors for 8 simulated QTLs detected by at least one of the four methods (binQTL, ANOVA, CIM and GCIM) obtained from 1000 replicated simulation experiments of the IMF2 population.

Method	Window	QTL-1	QTL-2	QTL-5	QTL-6	QTL-11	QTL-12	QTL-13	QTL-16	Type I error
ANOVA	±0 bin	1.000	1.000	0.999	1.000	1.000	1.000	1.000	0.956	0.1061
CIM	±0 bin	1.000	0.894	1.000	1.000	0.548	0.344	0.061	0.890	0.0604
binQTL	±0 bin	1.000	0.081	0.998	1.000	0.920	1.000	0.687	0.083	0.0260
GCIM	±0 bin	0.601	0.470	0.075	0.187	0.122	0.678	0.556	0.342	0.0084
ANOVA	±1 bin	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.979	0.0986
CIM	±1 bin	1.000	0.897	1.000	1.000	0.549	0.396	0.076	0.946	0.0548
binQTL	±1 bin	1.000	0.091	0.999	1.000	0.973	1.000	0.769	0.137	0.0195
GCIM	±1 bin	0.966	0.686	0.241	0.919	0.212	0.987	0.752	0.827	0.0065
ANOVA	±2 bin	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.0925
CIM	±2 bin	1.000	0.897	1.000	1.000	0.580	0.548	0.339	0.950	0.0498
binQTL	±2 bin	1.000	0.100	0.999	1.000	0.983	1.000	0.811	0.216	0.0143
GCIM	±2 bin	0.971	0.898	0.392	0.975	0.418	0.996	0.822	0.920	0.0055
ANOVA	±3 bin	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.0876
CIM	±3 bin	1.000	0.898	1.000	1.000	0.728	0.973	0.763	0.951	0.0454
binQTL	±3 bin	1.000	0.103	1.000	1.000	1.000	1.000	0.875	0.230	0.0102
GCIM	±3 bin	1.000	0.943	0.830	0.987	0.771	1.000	0.903	0.963	0.0048
ANOVA	±4 bin	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.0823
CIM	±4 bin	1.000	0.898	1.000	1.000	0.730	0.992	0.763	0.951	0.0408
binQTL	±4 bin	1.000	0.103	1.000	1.000	1.000	1.000	0.918	0.230	0.0064
GCIM	±4 bin	1.000	0.943	0.947	0.993	1.000	1.000	0.920	0.967	0.0043
ANOVA	±5 bin	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.0768
CIM	±5 bin	1.000	0.898	1.000	1.000	0.772	1.000	0.868	0.951	0.0361
binQTL	±5 bin	1.000	0.103	1.000	1.000	1.000	1.000	0.961	0.230	0.0034
GCIM	±5 bin	1.000	0.948	0.997	0.999	1.000	1.000	0.929	0.981	0.0039

and CIM methods for each of the 20 simulated QTLs under window ±3 bins. Except for a few large QTLs where the curves of the three methods overlapped, ROC curves of the binQTL method were the highest followed by ANOVA and CIM, indicating that binQTL is more powerful than the other two methods if the Type I error is controlled at the same level. The conclusion remained the same for other window sizes (Fig. S2). We were unable to draw the ROC curve for GCIM as GCIM only outputs the LOD scores of a few detected QTLs rather than the LOD scores of all markers.

We then pooled the 20 QTLs together and drew ROC curves collectively without separating the 20 QTLs. Fig. 3 showed the overall ROC curves for the three methods under each of the six window sizes (i.e., $\pm i$ bin, where $i = 0, 1, \cdots, 5$). Under the same level of Type I error, the binQTL method had the highest powers in all cases followed by CIM and ANOVA, while the latter two were much similar to each other.

2.2. Simulation studies of the RIL population

We also simulated a trait from 1619 binned genotypes of the RIL population of 210 inbred lines of rice developed by our laboratory (Xing et al., 2002). This study demonstrated the versatility of the binQTL method to handle two genotypes per locus in RIL *vs.* three genotypes per locus in IMF2. Again, 20 QTLs were added to the genome. Details of the simulation experiment are provided in Supplementary Note S1. Positions and effects of the simulated QTLs are shown in Table S3 and illustrated in Fig. S3, which also shows the estimated effects from different methods. The LOD score test statistics are illustrated in Fig. S4. The conclusion here remained the same as the simulation study with the IMF2 population: the resolution of binQTL was comparable to CIM or GCIM and was higher than ANOVA in terms of separating neighboring peaks.

Power and Type I error analyses from 1000 replicated simulation experiments led to similar conclusion as the IMF2 simulation study where ANOVA had higher powers than CIM, binQTL and GCIM, but comparison between CIM and binQTL varied across different QTLs (binQTL was better than CIM for some QTLs but worse for others). However, ANOVA also had the highest Type I errors while GCIM always had the lowest Type I errors (Table S4). The GCIM method had the lowest power compared to all other methods when the window size was small, but the power arose quickly as the window size increased. QTL-specific ROC curve comparisons showed that binQTL had the highest powers followed by CIM and ANOVA when the Type I error was controlled at the same level (Figs. S5 and S6). The overall ROC curves collectively for all 20 QTLs also showed that binQTL was more powerful than CIM and ANOVA (Fig. S7).

2.3. QTL mapping for 1000-grain weight (KGW) of rice

We conducted QTL mapping for KGW using the 278 IMF2 hybrids and the 210 RILs of rice with the four methods. The estimated QTL effects are shown in Fig. 4. Results of the four methods from the two populations were surprisingly consistent, indicating excellent quality of the field data. The corresponding LOD score test statistics plotted against the genome is shown in Fig. 5. All methods pinpointed to a region on chromosome 3 and another region on chromosome 5 that were strongly associated with KGW. The two major peaks were overlapped with the *GS3* gene on chromosome 3 and the *GW5* gene on chromosome 5, respectively, both being cloned for KGW (Fan et al., 2006; Weng et al., 2008). The peaks from binQTL were sharper than those from ANOVA and CIM and thus the former had higher resolution than the latter two methods.

3. Discussion

With the high-quality reference genome sequences along with the advanced population sequencing technology, genotypes of every locus of the entire genome can be determined unambiguously for all major crops. The observed sequence information allows us to define bins, which may harbor all causal polymorphisms of QTLs. Unlike traditional IM and CIM where a QTL is located to a region bracketed by two markers, we can now pinpoint a QTL to a bin. The sizes of the bins represent the resolution of QTL mapping – the smaller the bins, the higher the resolution. The resolution can be ultimately reduced to a single nucleotide by increasing the

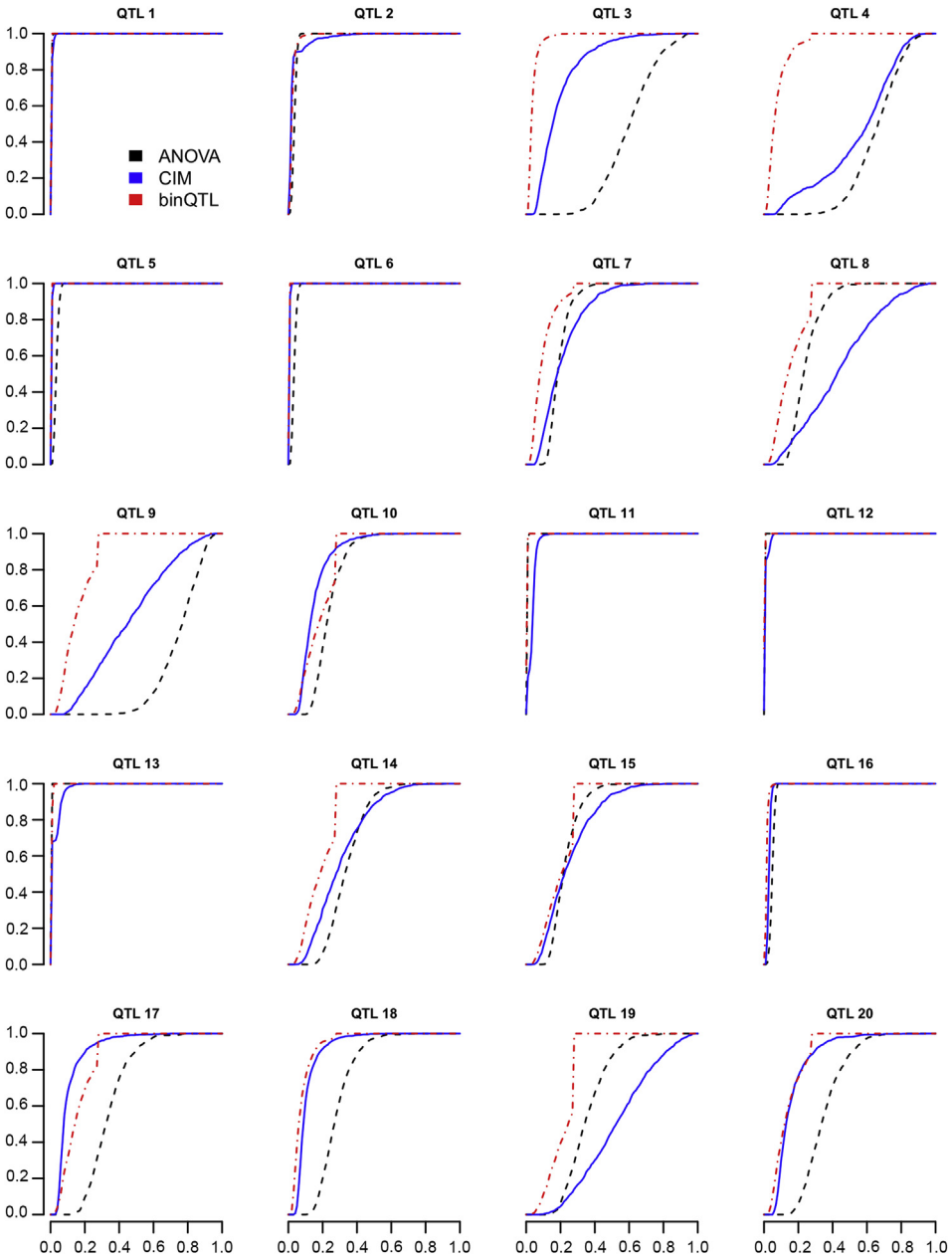

Fig. 2. ROC curves for each of the 20 simulated QTLs of the IMF2 population for three methods when the reserved window around QTL is ±3 bins. An ROC curve is defined as the plot of power (Y-axis of each plot) against Type I error (X-axis of each plot) (see Materials and methods).

sample size of a mapping population. Therefore, binQTL mapping may represent the norm of future QTL mapping technologies.

The binQTL method for QTL mapping was developed under the linear mixed model (LMM) framework, much like the mixed linear model implemented in GWAS (Yu et al., 2006). It incorporates a marker inferred kinship matrix into the LMM to capture the polygenic background effect. The polygene serves the same function as the cofactors in CIM (Zeng, 1994), but the result is much more robust because cofactor selection has been avoided. The binQTL method has added several new features over the existing bin model (Xu, 2013a). One feature is that the binQTL method is sufficiently flexible to handle any number of genotypes per locus. For example, the binQTL method can analyze BC, RIL, F_2 or any other mapping populations, as long as the number of genotypes per locus is a reasonable finite number, say <10. We accomplished this by treating the genotypic effects as random effects and testing the genotypic variance (not the effects) using the likelihood ratio test. Another feature is that we reserved a three-bin window around the scanned bin to avoid competition of the tested bin with its polygenic counterpart, i.e., avoid proximal contamination (Listgarten et al., 2012). We did this by removing the triplet (three bins) from the kinship matrix. This led to a change in kinship matrix for every bin scanned. Special algorithm of Woodbury matrix identities (Woodbury, 1949) was adopted here to ease the computational burden. This feature is more important in QTL mapping than in GWAS because the number of bins is often smaller in a linkage population than in a random population in association studies. The smaller the number of bins, the stronger the competition exists between the scanned bin and its counterpart in the polygene.

Although the binQTL method was particularly designed for QTL

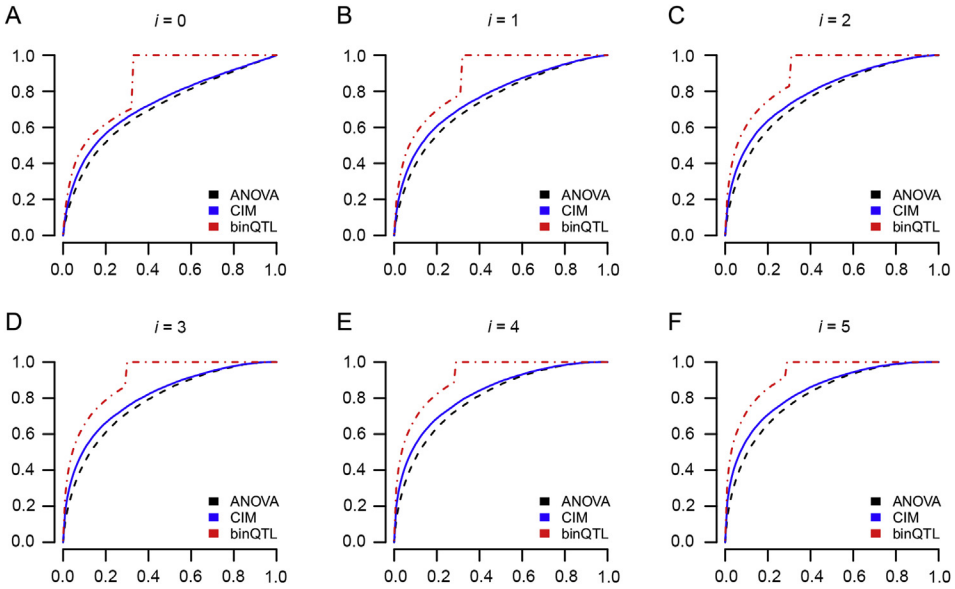

Fig. 3. ROC curves for all 20 simulated QTLs of the IMF2 population for three methods. An ROC curve is defined as the plot of power (Y-axis of each plot) against Type I error (X-axis of each plot). The six panels represent ROC curves of six window sizes ($\pm i$ bin, where $i = 0, 1, 2, 3, 4, 5$).

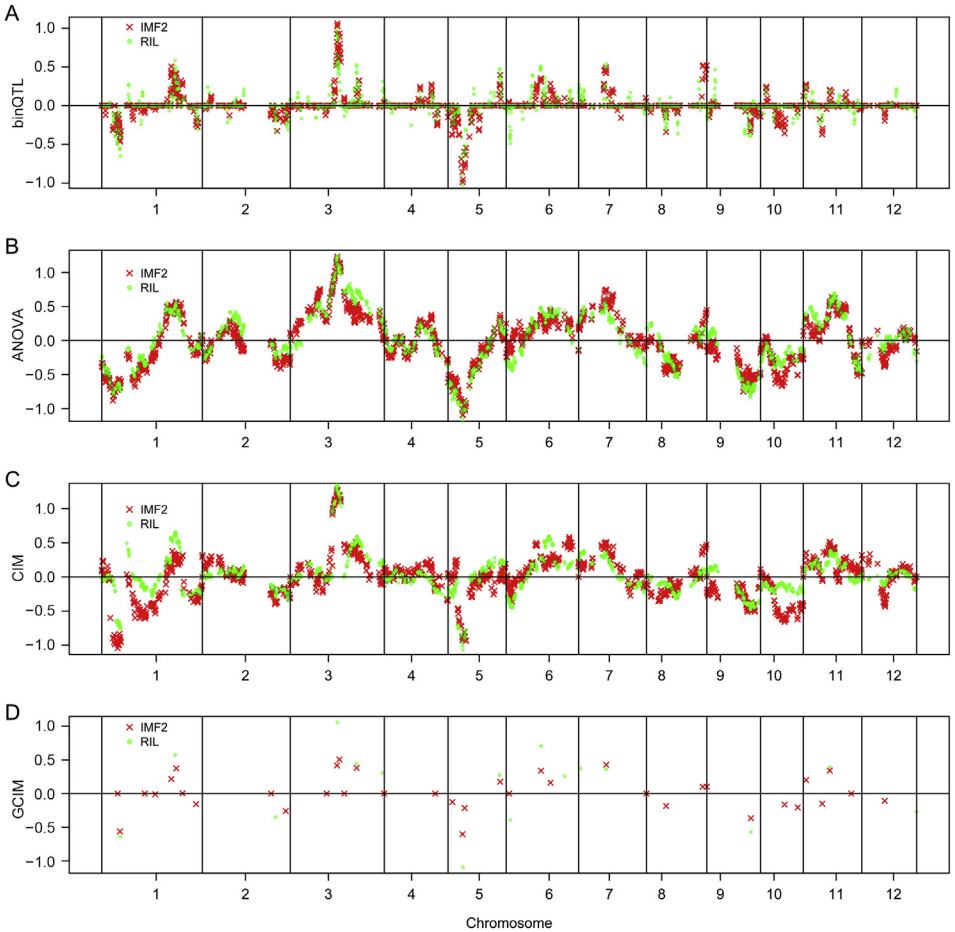

Fig. 4. Estimated QTL effects of KGW from the IMF2 and RIL populations of rice using four methods. **A:** binQTL. **B:** ANOVA. **C:** CIM. **D:** GCIM. Different chromosomes are separated by the black vertical lines. X-axis shows the position of each marker (bin) in the genetic map.

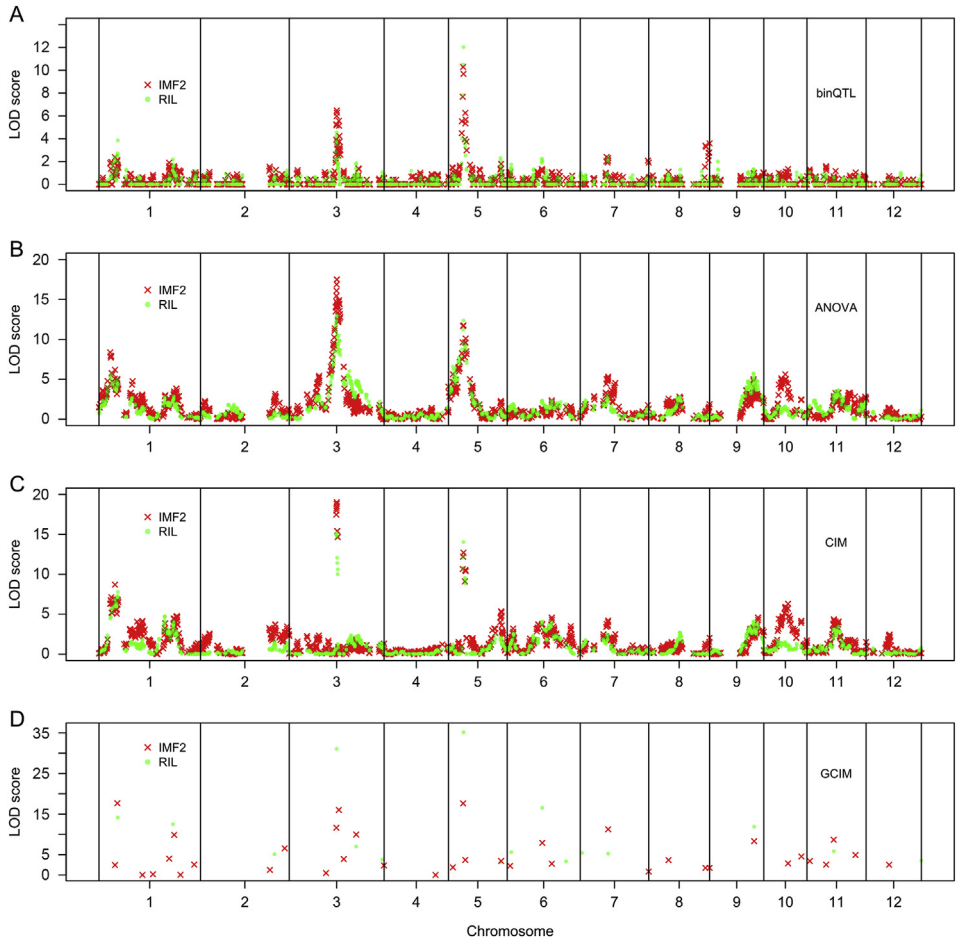

Fig. 5. LOD scores of KGW plotted against genome locations of bins for four methods. **A:** binQTL. **B:** ANOVA. **C:** CIM. **D:** GCIM. from the IMF2 and RIL populations of rice. Different chromosomes are separated by the black vertical lines. X-axis shows the position of each marker (bin) in the genetic map.

mapping, it can be used for GWAS with minor modifications. We demonstrated that the method can be used for a population with two genotypes (RIL) or three genotypes (IMF2) per locus. We further tested binQTL on a mouse MAGIC population with eight genotypes per locus. The genotype data of the MAGIC population was obtained from a previous study while the phenotype data were simulated with 20 QTLs (Table S5) (Collaborative Cross Consortium, 2012; Wei and Xu, 2016). The QTL peaks identified by binQTL matched very well with most of the 20 simulated QTLs, indicating the high efficiency of binQTL for MAGIC populations (Fig. S8). In more general situations, four or more genotypes per locus may occur in a natural population; the binQTL method can also be applied to QTL analysis of such populations because the program can detect the number of genotypes per locus automatically and estimate the genotypic variance. In addition, copy number variation and InDel markers can be incorporated into the SNP map for GWAS because the binQTL method can be modified to deal with variable number of genotypes across loci.

The key to the versatility of binQTL is the genotypic model with random effects. When the genotypic effects are treated as random effects, the variance of the genotypic effects is the parameter of interest. The test statistic for testing a variance is the likelihood ratio test. Under the null hypothesis, the test statistic follows a mixture distribution of χ_0^2 and χ_1^2 with an equal weight. This mixture distribution causes the discontinued ROC curves of binQTL at a particular point of Type I error (Figs. 2 and 3). The other two methods CIM and ANOVA do not have this behavior because their test statistics are not of mixture distribution.

The concept of bins applies to experimental populations created by crossing a few inbred lines. A bin contains a group of markers with perfect LD. The bin concept may be extended such that a generalized bin can be defined in a random association population to reduce the bin number. Adjacent bins with high LD (not necessarily perfect LD) can be combined into a generalized bin. Xu (2013a) called bins defined this way artificial bins and also provided the method for doing so. The generalized bin model is significant in several aspects. It allows us to map epistasis of bin pairs (bin-by-bin interactions) because the number of bin pairs can be made sufficiently small to a manageable number. Method was also proposed for constructing an epistatic kinship matrix to control the polygenic epistatic background effect (Xu, 2013a). The smaller number of bins may even allow the use of a multiple QTL model for association studies.

4. Materials and methods

4.1. Statistical methods

Let y be an $n \times 1$ vector of phenotypic values of a trait for n individuals and define Z_k as an $n \times q_k$ design matrix (dummy variables) for the genotypes of bin k, where q_k is the number of genotype classes of bin k. We now introduce the following mixed model for y

$$y = X\beta + Z_k\gamma_k + \xi_{-k} + \varepsilon$$

where X is a design matrix for r covariates, β is a $r \times 1$ vector of the covariates themselves (fixed effects), γ_k is a $q_k \times 1$ vector of genotypic effects (random) for bin k (one effect per genotype), ξ_{-k} is an $n \times 1$ vector of polygenic effects (excluding effects in the neighborhood of bin k) and ε is a $n \times 1$ vector of residual errors. The random effects and residual errors are assumed to be independent and normally distributed

$$\gamma_k \sim N\left(0, \varphi_k^2\right)$$
$$\xi_{-k} \sim N\left(0, K_{-k}\varphi^2\right)$$
$$\varepsilon \sim N\left(0, I\sigma^2\right)$$

where ϕ_k^2 is the variance of bin k, ϕ^2 is the polygenic variance, σ^2 is the residual error variance and K_{-k} is an $n \times n$ kinship matrix inferred from all bins except the kth bin and its two neighbors. The bin specific kinship matrix is defined as

$$K_{-k} = c\sum_{k' \neq k}^{m} Z_{k'}Z_{k'}^T = c\sum_{k'=1}^{m} Z_{k'}Z_{k'}^T - cZ_kZ_k^T = K - cZ_kZ_k^T$$

where $K = c\sum_{k'=1}^{m} Z_{k'}Z_{k'}^T$ is the kinship matrix inferred from all bins and c is a normalization constant (see Supplementary Note S1 in detail for definitions of kinship matrices).

The linear mixed model has an expectation $E(y) = X\beta$ and a variance

$$\mathrm{var}(y) = V_k = K\phi^2 + I\sigma^2 - cZ_kZ_k^T\phi_k^2 = V - cZ_kZ_k^T\phi_k^2$$

where $V = K\phi^2 + I\sigma^2$ is the overall variance matrix. The expectation and variance allow us to construct the following restricted log likelihood function to estimate variances involved in the mixed model

$$L(\theta) = -\frac{1}{2}\ln\left|V_k\right| - \frac{1}{2}\ln\left|X^TV_k^{-1}X\right| - \frac{1}{2}(y - X\widehat{\beta})V_k^{-1}(y - X\widehat{\beta})$$

where $\theta = \{\phi_k^2, \phi^2, \sigma^2\}$ is the parameter vector and

$$\widehat{\beta} = \left(X^TV_k^{-1}X\right)^{-1}X^TV_k^{-1}y$$

The Newton-Raphson iterative algorithm was used to estimate the parameters. Eigenvalue decomposition and Sherman–Morrison–Woodbury matrix identity were adopted to ease the computational burden (Woodbury, 1949). Details of the algorithm are provided in Supplementary Notes S2 and S3.

To test the association of bin k with the trait of interest under the null hypothesis $H_0 : \phi_k^2 = 0$, we used the likelihood ratio test defined as

$$\tau_k = -2[L(\widehat{\theta}_0) - L(\widehat{\theta})]$$

where $\widehat{\theta}_0 = \{\widehat{\phi}^2, \widehat{\sigma}^2\}$ is the parameter vector with $\phi_k^2 = 0$, $L(\widehat{\theta}_0)$ and $L(\widehat{\theta})$ are the log likelihood values evaluated under the null and full models, respectively. Under the null model, τ_k follows approximately a mixture of two chi-square distributions

$$\tau_k \sim \frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$$

from which an appropriate p-value is calculated. The entire QTL mapping involves testing all bins sequentially until the entire genome is scanned. Please see Supplementary Note S4 for p-value calculation of the above mixture chi-square distribution. The new method of QTL mapping is called binQTL, which is compared with three existing methods, ANOVA, CIM and GCIM.

4.2. Simulation studies using bin genotypes of inbred and hybrid rice

The purpose of the simulation studies is to examine the power (true positive rate) and Type I error (false positive rate) of the new method in comparison to CIM, ANOVA and GCIM. To make the simulation experiments as close to reality as possible and also to simplify the simulation, we took advantage of existing mapping populations in rice and used the genotypes of 1619 bins of 210 recombinant inbred lines (RILs) and their 278 immortalized F_2 (IMF2) crosses to simulate genetic and phenotypic values of a hypothetic quantitative trait. The two populations of rice (*Oryza sativa*) were derived from the cross between Zhenshan 97 and Minghui 63 (Hua et al., 2002; Xing et al., 2002), the parents of a widely grown hybrid Shanyou 63. The first population (RIL) was derived by single-seed descent from the cross between the two parents. The second population (IMF2) was generated via crossing by randomly pairing of the 210 RILs (Hua et al., 2002, 2003; Zhou et al., 2012). The genomic data are represented by 1619 bins inferred from ~270,000 SNPs of the rice genome (Xie et al., 2010; Yu et al., 2011; Zhou et al., 2012). All SNPs within a bin have exactly the same segregation pattern and thus one SNP is sufficient to represent the entire bin. The bin genotypes of the 210 RILs were coded as "1" for the Zhenshan 97 genotype and "0" for the Minghui 63 genotype (two genotypes per locus). Genotypes of the hybrids were deduced from the genotypes of the two parents (Hua et al., 2003) and thus there are three possible genotypes per locus for the hybrid population with the homozygote of Minghui 63 coded as "1", the homozygote of Zhenshan 97 coded as "−1" and the heterozygote coded as "0". In fact, the numerical value of the genotype of a hybrid is the sum of the numerical genotypic codes of the two parents subtracted by 1.

We simulated 20 QTLs with effects and positions listed in Table S1 for the IMF2 population. The 20 QTLs are distributed on 9 of the 12 chromosomes. The numerical codes of genotypes for the 20 QTLs are correlated due to linkage disequilibrium (LD), and the variance-covariance matrix is listed in Table S6. We added a residual error sampled from $N(0, \sigma^2)$ to the total genetic value of each individual to form the phenotypic value of the trait, where $\sigma^2 = 10$ is the residual error variance. The genetic variance and proportion of phenotypic variance contributed by each QTL are listed in Table S1. The 20 QTLs collectedly contribute 0.8499 of the phenotypic variance. However, the contribution by each individual QTL varies from 0.009 to 0.148 with the majority of the QTLs having contribution less than 0.05. Details regarding the experimental design and the theoretical analysis are given in Supplementary Note S1.

Similar to the simulated IMF2 population, we simulated 20 QTLs with effects and positions listed in Table S3 for the RIL population. The variance-covariance matrix of the numerical codes for the 20 QTLs is listed in Table S7. The residual error variance is $\sigma^2 = 10$. The genetic variance and proportion of phenotypic variance contributed by each QTL are listed in Table S3. The 20 QTL collectedly contribute 0.8471 of the phenotypic variance. The contribution by each individual QTL varies from 0.001 to 0.08 with the majority of the QTLs having contribution less than 0.03. Details regarding the experimental design and the theoretical analysis are given in Supplementary Note S1.

We further generated another 1000 independent samples, all

having the same genotype array but with independent random draws for the residuals to form different phenotypes. In each sample, 20 QTLs were simulated with positions uniformly distributed along the genome and effects randomly sampled from mean 0 and variance 4, i.e., $N(0,4)$. For each sample, we performed QTL mapping using four methods (binQTL, ANOVA, CIM and GCIM) and recorded the LOD scores for each bin. To determine the critical values of the LOD scores used to declare associations of bins with the trait, we also simulated 1000 additional samples under the null model (no QTLs were added to the bins). For each null sample, we picked up the highest LOD score across all bins. From the 1000 null samples, we had 1000 such highest LOD scores. The 95th percentile of the 1000 LOD scores was used as the critical value (empirical threshold).

We are now back to the 1000 samples with true QTLs added to the bins. If the LOD score of a bin was larger than the critical value, the bin was declared as a QTL. Due to LD, bins in the neighborhood of a simulated QTL often showed associations. Therefore, we reserved a window $\pm i$ bin around each bin with true QTL, where i took one of the six values: 0, 1, 2, 3, 4 and 5. For example, if $i = 1$, the window covers the QTL (bin in the middle) and the two neighboring bins. If the LOD scores of any bins in the window passed the critical value, this simulated QTL was declared as being detected. The situation of $i = 0$ is a special case where no window was reserved. Of the 1000 samples, the proportion of samples with significant detections for a simulated QTL is the power for that QTL. In addition to these QTL specific powers, we also recorded the overall power for all QTLs collectively. For example, if 15.5 QTL (on average) are detected out of the 20 QTLs across all 1000 samples, the overall power should be $15.5/20 = 0.775$.

The Type I error was calculated using all bins outside the reserved windows of all the simulated QTLs. The total number of significant bins outside the reserved windows divided the total number of bins outside the reserved windows is the empirical Type I error. Of course, we took the average Type I errors across the 1000 simulated samples.

Different methods may have different Type I errors and thus the power comparison may not be fair if their Type I errors are not controlled at the same level. Therefore, we drew a receiver operating characteristic (ROC) curve for each QTL with the three methods (binQTL, ANOVA and CIM). The ROC curve is a plot of the power against the Type I error. When the three ROC curves are plotted together, we can compare their powers at the same level of Type I error: the further away from the diagonal line, the higher the efficiency. The ROC curve can be QTL specific and population specific. We presented both in the simulation studies.

4.3. QTL mapping for grain weight in inbred and hybrid rice

We applied the new method of QTL mapping to 1000-grain weight (KGW) of rice in an IMF2 population of 278 hybrids and an RIL population of 210 inbred lines. The trait was measured in two consecutive years for the IMF2 population (1998 and 1999) and in three consecutive years for the RIL population (1997, 1998 and 1999) at the Huazhong Agricultural University Experimental Station (Hua et al., 2002). The heritability for KGW estimated from the replicated experiment was 0.79 using data of the IMF2 population (see Supplementary Note S5 for the ANOVA method used for estimating the broad sense heritability). In this study, we took the average KGW of multiple years as the original phenotype for QTL mapping. All four methods (ANOVA, CIM, GCIM and binQTL) were used in the real data QTL mapping. The CIM method implemented in the R/qtl package (Broman et al., 2003) was used to scan QTLs in this study. The QTL effect of the CIM method was calculated using WinQTLCart (Wang et al., 2012).

4.4. Data and software package

The phenotypes (KGW) and genotypes (1619 bins) of the 278 IMF2 hybrids and the 210 RILs are available in Tables S8–S11. We developed an R package, named binQTL, to implement the new method of QTL mapping. The R package has a graphic interface for convenience of users. The R code is stored in GitHub (https://github.com/venyao/binQTL) along with the binQTL.shiny code for graphic interface (https://github.com/venyao/binQTL.shiny). binQTL.shiny is deployed at http://150.109.59.144:3838/binQTL.shiny/ for online use.

Acknowledgments

The work was supported by the National Key Research and Development Program (2016YFD0100802) to Q.Z. and the National Science Foundation Collaborative Research grant (DBI-1458515) to S.X.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgg.2019.06.005.

References

Bernardo, R., 2013. Genomewide markers as cofactors for precision mapping of quantitative trait loci. Theor. Appl. Genet. 126, 999–1009.
Broman, K.W., Wu, H., Sen, Ś., Churchill, G.A., 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19, 889–890.
Collaborative Cross Consortium, 2012. The genome architecture of the collaborative cross mouse genetic reference population. Genetics 190, 389–401.
Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., Li, X., Zhang, Q., 2006. *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor. Appl. Genet. 112, 1164–1171.
Fisher, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. T. Roy. Soc. Edin. 52, 399–433.
Hua, J., Xing, Y., Wu, W., Xu, C., Sun, X., Yu, S., Zhang, Q., 2003. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc. Natl. Acad. Sci. U. S. A. 100, 2574–2579.
Hua, J.P., Xing, Y.Z., Xu, C.G., Sun, X.L., Yu, S.B., Zhang, Q., 2002. Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 162, 1885–1895.
Huang, B.E., Verbyla, K.L., Verbyla, A.P., Raghavan, C., Singh, V.K., Gaur, P., Leung, H., Varshney, R.K., Cavanagh, C.R., 2015. MAGIC populations in crops: current status and future prospects. Theor. Appl. Genet. 128, 999–1017.
Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., Dong, G., Sang, T., Han, B., 2009. High-throughput genotyping by whole-genome resequencing. Genome Res. 19, 1068–1076.
Kao, C.H., Zeng, Z.B., Teasdale, R.D., 1999. Multiple interval mapping for quantitative trait loci. Genetics 152, 1203–1216.
Lander, E.S., Botstein, D., 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185–199.
Li, H., Ye, G., Wang, J., 2007. A modified algorithm for the improvement of composite interval mapping. Genetics 175, 361.
Listgarten, J., Lippert, C., Kadie, C.M., Davidson, R.I., Eskin, E., Heckerman, D., 2012. Improved linear mixed models for genome-wide association studies. Nat. Methods 9, 525.
Mackay, T.F.C., Stone, E.A., Ayroles, J.F., 2009. The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. 10, 565–577.
Mayer, M., 2005. A comparison of regression interval mapping and multiple interval mapping for linked QTL. Heredity 94, 599.
Pillen, K., Zacharias, A., Léon, J., 2003. Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 107, 340–352.
Wang, S.-B., Wen, Y.-J., Ren, W.-L., Ni, Y.-L., Zhang, J., Feng, J.-Y., Zhang, Y.-M., 2016. Mapping small-effect and linked quantitative trait loci for complex traits in backcross or DH populations via a multi-locus GWAS methodology. Sci. Rep. 6, 29951.
Wang, S., Basternand, J., Zeng, Z., 2012. Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm.
Wei, J., Xu, S., 2016. A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. Genetics 202, 471–486.
Wen, Y.J., Zhang, Y.W., Zhang, J., Feng, J.Y., Dunwell, J.M., Zhang, Y.M., 2018. An efficient multi-locus mixed model framework for the detection of small and

linked QTLs in F_2. Brief. Bioinform. https://doi.org/10.1093/bib/bby058.
Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., Lei, C., Zhang, X., Cheng, Z., Guo, X., Wang, J., Jiang, L., Zhai, H., Wan, J., 2008. Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. Cell Res. 18, 1199–1209.
Woodbury, M.A., 1949. The Stability of Out-Input Matrices. University of Chicago Press, Chicago, p. 93.
Xie, W., Feng, Q., Yu, H., Huang, X., Zhao, Q., Xing, Y., Yu, S., Han, B., Zhang, Q., 2010. Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc. Natl. Acad. Sci. U. S. A. 107, 10578–10583.
Xing, Z., Tan, F., Hua, P., Sun, L., Xu, G., Zhang, Q., 2002. Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor. Appl. Genet. 105, 248–257.
Xu, S., 2013a. Genetic mapping and genomic selection using recombination breakpoint data. Genetics 195, 1103–1115.
Xu, S., 2013b. Mapping quantitative trait loci by controlling polygenic background effects. Genetics 195, 1209–1222.
Yu, H., Xie, W., Wang, J., Xing, Y., Xu, C., Li, X., Xiao, J., Zhang, Q., 2011. Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. PLoS One 6, e17595.
Yu, J., Holland, J.B., McMullen, M.D., Buckler, E.S., 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178, 539.
Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S., 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38, 203.
Zeng, Z.B., 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. U. S. A. 90, 10972–10976.
Zeng, Z.B., 1994. Precision mapping of quantitative trait loci. Genetics 136, 1457–1468.
Zhou, G., Chen, Y., Yao, W., Zhang, C., Xie, W., Hua, J., Xing, Y., Xiao, J., Zhang, Q., 2012. Genetic composition of yield heterosis in an elite rice hybrid. Proc. Natl. Acad. Sci. U. S. A. 109, 15847–15852.