Joint QTL Linkage Mapping for Multiple-Cross Mating Design Sharing One Common Parent

Huihui Li[1,2], Peter Bradbury[3], Elhan Ersoz[2¤], Edward S. Buckler[2,3], Jiankang Wang[1]*

1 Institute of Crop Science, The National Key Facility for Crop Gene Resources and Genetic Improvement, and CIMMYT China, Chinese Academy of Agricultural Sciences, Beijing, China, 2 Institute for Genomic Diversity, Cornell University, Ithaca, New York, United States of America, 3 United States Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, New York, United States of America

Abstract

Background: Nested association mapping (NAM) is a novel genetic mating design that combines the advantages of linkage analysis and association mapping. This design provides opportunities to study the inheritance of complex traits, but also requires more advanced statistical methods. In this paper, we present the detailed algorithm of a QTL linkage mapping method suitable for genetic populations derived from NAM designs. This method is called joint inclusive composite interval mapping (JICIM). Simulations were designed on the detected QTL in a maize NAM population and an *Arabidopsis* NAM population so as to evaluate the efficiency of the NAM design and the JICIM method.

Principal Findings: Fifty-two QTL were identified in the maize population, explaining 89% of the phenotypic variance of days to silking, and nine QTL were identified in the *Arabidopsis* population, explaining 83% of the phenotypic variance of flowering time. Simulations indicated that the detection power of these identified QTL was consistently high, especially for large-effect QTL. For rare QTL having significant effects in only one family, the power of correct detection within the 5 cM support interval was around 80% for 1-day effect QTL in the maize population, and for 3-day effect QTL in the *Arabidopsis* population. For smaller-effect QTL, the power diminished, e.g., it was around 50% for maize QTL with an effect of 0.5 day. When QTL were linked at a distance of 5 cM, the likelihood of mapping them as two distinct QTL was about 70% in the maize population. When the linkage distance was 1 cM, they were more likely mapped as one single QTL at an intermediary position.

Conclusions: Because it takes advantage of the large genetic variation among parental lines and the large population size, NAM is a powerful multiple-cross design for complex trait dissection. JICIM is an efficient and specialty method for the joint QTL linkage mapping of genetic populations derived from the NAM design.

Introduction

QTL mapping based on biparental populations has become a routine approach for genetic studies of complex traits in plants and animals. In most biparental populations, recombination has not had enough time to shuffle the genome into small fragments, and QTL are generally located in large chromosomal regions [1,2]. In addition, linkage mapping based on biparental populations can only identify QTL from the phenotypic diversity generated from the controlled cross. In contrast, association mapping is based on natural populations, searching for genotype to phenotype correlations in unrelated individuals. Though more rapid and cost-effective compared with linkage mapping, association mapping is heavily dependent on population structure, which is normally unknown [3]. Nested association mapping (NAM) population derived from a multiple-cross mating design sharing one common parent was therefore proposed to enable high power

and high resolution through joint linkage and association analysis, and to provide a broader genetic resource for quantitative trait analysis [4]. As an example, one NAM population in maize was recently reported [5,6].

Several methods have been proposed for QTL mapping on populations derived from multiple strain crosses. Rebaï and Goffinet adapted the linear regression method to the case of a diallel cross between four inbred lines [7], and afterwards presented a general linear model (GLM) for QTL mapping by combining different populations derived from diallel designs [8]. Xu [9] proposed a fixed model and a random model for multiple independent families based on the weighted least square method of linear model. By using mixture model, Liu and Zeng [10] extended composite interval mapping (CIM; [11]) to the algorithm of multiple inbred lines, and Jourjon et al. [12] implemented the CIM and iterative QTL mapping in MCQTL software to perform QTL mapping either for independent families or for families from

the diallel design. Jannink and Jansen [13] developed a multiple-QTL model for multiple related populations derived from the diallel of pairwise crosses among three inbred parents. Based on the mixed model, Crepieux et al. [14] developed a two-step IBD (identical by descent) variance component approach for breeding populations obtained from inbred parents. On the Bayesian perspective, Yi and Xu [15,16] developed a Bayesian method implemented via the Markov chain Monte Carlo algorithm for mapping QTL using complicated multiple line crosses. Fang et al. [17] extended Bayesian model selection and the Bayesian shrinkage estimation approach to multiple independent crosses.

For the NAM design, Yu et al. [4] used the principle of association mapping to investigate its genetic and statistical properties by simulation experiments. As recombination information is not used, association mapping cannot give a position estimation of the identified QTL, which is essential for gene fine-mapping and map-based cloning, especially under moderate marker density. More recently, Hayashi and Iwata [18] simulated a total of 2000 individuals in a NAM population and 84 evenly distributed markers in the whole genome to evaluate a Bayesian method for QTL mapping. Nevertheless, as Xu and Jia [19] pointed out, Bayesian models still have problems detecting QTL in real populations.

Spearheaded by the maize NAM population of Buckler et al. [5], such augmented study designs are gaining more acknowledgement and application in breeding and genetics studies for various species. Despite the fact that more data and results in this framework are quickly becoming available, there is no algorithm specifically designed to analyze data from such studies. This is to emphasize the fact that although, several methods have been proposed for simultaneous analysis of such networked multi-population QTL data [7–10,12–17], none of these methods are optimized to take full advantage of the reference design cases. In this paper, we present the algorithm of joint inclusive composite interval mapping (JICIM), which is specialty for NAM design and has been adopted in the paper published in Science by Buckler et al. [5]. We then used the QTL identified by JICIM in the real maize and *Arabidopsis* NAM populations to define genetic models in simulation experiments, so as to evaluate the statistical power and reliability of the identified QTL and the efficiency of JICIM to detect rare QTL in the NAM design. By rare QTL, in this study we mean a QTL that segregates in just one family, i.e., there are two alternative alleles having identifiable genetic effects on a trait of interest in the segregating family. In other families, either the same allele is located at the rare QTL chromosomal position, or the two alleles show no significant difference in their genetic effects on the trait of interest. In contrast, QTL that segregate in a number of families are called common QTL.

Results

Mapping results in two NAM populations

For days to silking in the maize population, 52 QTL were identified by JICIM in their one-LOD support intervals across the whole genome (Table S1), with a LOD threshold of 12.26 to guard against more than two false positives, i.e., GWER(2) = 12.26 [20]. Among the 52 QTL, the top 30 QTL (denoted as $qZ1-qZ30$, where q stands for QTL and Z for *Zea mays* L.) explain 84% of the total phenotypic variance, and were used to demonstrate the genetic architecture of maize flowering time by Buckler et al. [5]. The chromosomal distribution and effect size of the identified QTL across the 25 families (Fig. S1) supported the conclusion that large differences in flowering time among inbred maize lines are caused by the cumulative effects of numerous QTL [5]. Among all

the detected QTL, there is no rare QTL that has significant additive effect in only one family.

For flowering time in the *Arabidopsis* population [21] and when the LOD threshold of GWER(2) = 3.30, nine QTL (denoted as $qA1-qA9$, where A stands for *Arabidopsis*) are identified in one-LOD support interval across the whole genome, explaining 83% of the phenotypic variance (Table 1 and Fig. 1). Flowering time is a well studied trait in *Arabidopsis*, and numerous flowering time genes have been identified (cloned) in *Arabidopsis*, based on mutation analysis [22–24], but that the relationship between these genes and the QTL underlying variation in mapping populations derived from accession crosses is unclear. None-the-less, there are previous reports in the literature identifying the same 9 QTL that were detected by JICIM, albeit there seems to be clusters of candidate genes underlying these QTL. For instance, when the positions of the QTL detected were compared to the results from the candidate gene association mapping study of Ehrenreich et al. [25], there seems to be a cluster of three important flowering time candidate genes, i.e., FRI, LD and GA1 corresponding to the position underlying the largest QTL around marker *FRI* from the FRIGIDA gene in this study. This locus is significant in all the three families, and explains 53.84% of the phenotypic variance. A table of candidate genes identified to be significant in the association study by Ehrenreich et al. [25] and Brachi et al. [26] that correspond to the QTL intervals detected in this study is included for the purposes of linking this study to the rest of the *Arabidopsis* literature (Table 1).

Most loci with significant peaks in the LOD profile have been detected by QTL mapping in individual families [21]. But the LOD scores (Fig. 1) are much higher than those from individual family mapping [21], which empirically demonstrated the higher mapping power of joint linkage analysis. Four of the nine QTL identified in the *Arabidopsis* population, i.e., $qA1$ (GI/GAI) -$qA3$ (RGA) and $qA8$ in Table 1, had significant effects in only one family. Therefore, $qA1-qA3$ and $qA8$ were rare QTL in the *Arabidopsis* population. Three of the nine QTL were significant in two families, i.e., $qA4$ (RGL2) detected in families Ler×An-1 and Ler×Kond, $qA6$ (PHYD) in families Ler×Kas-2 and Ler×Kond, and $qA9$ (LFY/ELF5) in families Ler×An-1 and Ler×Kas-2. The other two QTL, i.e., $qA5$ (FRI) and $qA7$ (FLC/CO), had significant genetic effects in all three families.

Simulation results based on identified QTL

When a QTL overlapped with a marker, the detection power of JICIM was consistently higher in comparison with QTL located in the middle of a marker interval (Table 2 and Fig. 2), since a QTL located midway between markers is the most difficult scenario for QTL detection [27]. As expected, a wider support interval (SI) resulted in higher power.

For the maize QTL, detection power was over 95% for all 30 QTL when the length of the SI was 5 cM (Table 2). When we narrowed the SI down to 2 cM, the average power was 96.2% when every QTL overlapped with one marker locus, and 88.8% when every QTL was located in the middle of one marker interval. Even when the SI was 1 cM, we still achieved an average power of 90.9% when the QTL overlapped with markers, and an average power of 72.7% when the QTL was located in the middle of a marker interval. In other words, JICIM places the detected QTL within 1 cM of the true position of the QTL at about a frequency of 90% of the time in the maize NAM population, if the actual QTL is tagged and genotyped in the study. However, the expectation of genotyping of the actual functional polymorphism in this framework is not very realistic, considering the low density of markers used for QTL mapping. Similarly, Guo et al. [28]

Table 1. Nine flowering time QTL identified in the *Arabidopsis* NAM population by JICIM.

QTL	Chr.	Pos. (cM)	LOD score	PVE (%)[a]	Left marker (cM)[b]	Right marker (cM)[b]	Additive genetic effect (day) in each family[c]			PVE (%) in each family			Flowering genes around QTL[d]	
							Ler×An-I	Ler×Kas-2	Ler×Kond	Ler×An-I	Ler×Kas-2	Ler×Kond	Brachi et al. [26]	Ehrenreich et al. [25]
qA1	1	66.0	3.75	3.92	CIW1* (65.2)	F6D8.94 (69.4)	−0.28	0.60	**−1.32** (−3.00)	2.92	1.03	2.73		GI, GAI
qA2	1	107.0	4.69	1.03	SNP157 (104.4)	SNP110* (107.8)	−0.06	**0.66** (3.00)	0.50	0.14	1.27	0.39	Cluster 1	FKF1
qA3	2	0.0	3.05	2.99	msat2.5* (0.0)	msat2.5* (0.0)	−0.01	−0.19	**−1.44** (−3.00)	0.00	0.11	3.22		RGA
qA4	3	0.0	13.39	3.60	SNP105* (0.0)	nga172* (2.9)	**1.16** (1.80)	0.53 (2.00)	**1.37** (3.00)	52.03	0.80	2.94		RGL2, GAr2
qA5	4	3.0	43.23	53.84	msat4.41 (0.0)	FRI* (3.6)	0.41	**−2.80** (−6.00)	**−5.62** (−11.40)	6.50	22.84	49.46	Cluster 2	FRI, LD, GA1
qA6	4	49.0	8.51	7.38	SNP295* (46.6)	SNP199 (49.9)	−0.03	**−1.95** (−2.60)	**−1.01**	0.03	11.08	1.60	Cluster 3	PHYD
qA7	5	19.0	13.67	5.97	SNP136* (17.3)	SNP358* (20.6)	**−0.75** (−1.6)	**−1.39** (−3.40)	**−1.74** (−2.8)	21.75	5.63	4.74	Cluster 4	TFL1,ATMYB33, FLC, CO, TFL2
qA8	5	33.0	5.87	6.38	SNP236* (32.1)	nga139 (33.5)	0.03	−0.50	**−2.09** (−5.00)	0.02	0.71	6.84		
qA9	5	93.0	8.00	5.43	SNP101 (92.5)	SNP304 (93.8)	**−0.56**	**1.38**	−0.51	12.13	5.55	0.40	Cluster 5	LFY,ELF5, VIN3

a phenotypic variance explained by each QTL;

b Markers with asterisks were significant in individual family mapping, reported by El-Lithy et al. (2006);

c Values in brackets are two allele's effects (i.e., 2a) estimated from individual family mapping.

d Annotations according to reports by Brachi et al. [26], and Ehrenreichet al. [25]. In individual family mapping of Ler×An-1, qA7 was mapped at SNP 130 [21], about 6.1 cM from SNP358 (Fig. S3). The significant effects from JICIM are in bold.

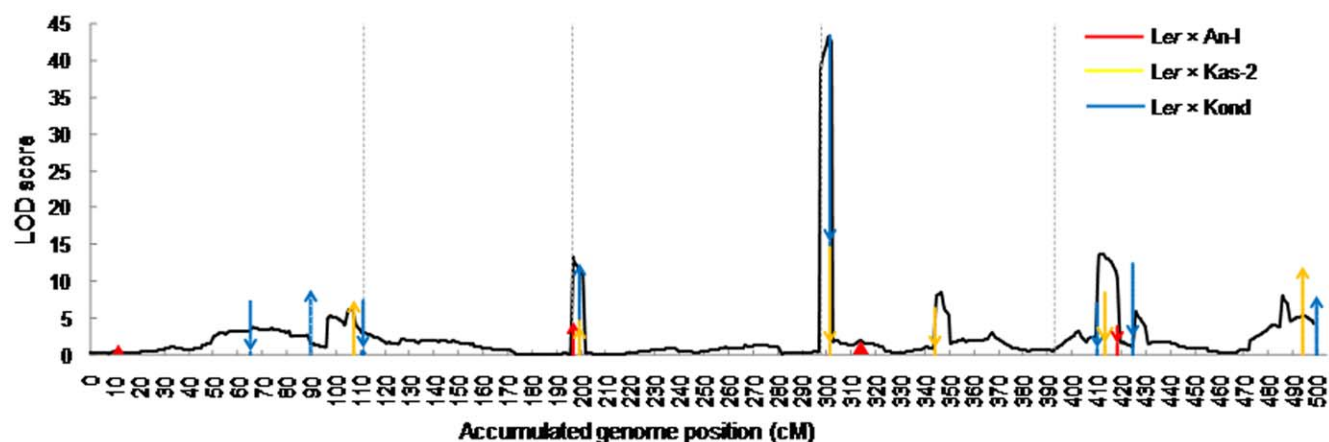doi:10.1371/journal.pone.0017573.t001

Figure 1. The one-dimensional scanning LOD profile of JICIM in the *Arabidopsis* population. The scanning step was 1 cM. Arrow size and direction represent the approximate effect size and direction of the identified QTL based on individual family mapping. Different colors of arrows indicate QTL identified in different families by individual family mapping.
doi:10.1371/journal.pone.0017573.g001

found functional markers segregating in at least five families and explaining >5% of the phenotypic variance can be detected with adequate power.

In the *Arabidopsis* population, detection power was lower (Fig. 2) than that in the maize population (Table 2), especially when SI = 1 cM, 2 cM, and 5 cM, indicating the strength of using more founders and a larger population size in the maize NAM population. The power of $qA1$-$qA3$ was much lower than that of $qA4$-$qA9$ regardless of the simulated QTL position and the length of the SI. For $qA4$-$qA9$ and when SI > = 5 cM, power was over 90.0% when the QTL overlapped with markers (Fig. 2A). When the QTL was located in the middle of a marker interval, power was over 80.0%, except for $qA8$ (Fig. 2B), which had significant effect only in the third family of the *Arabidopsis* population. These results indicated that for joint linkage mapping, rare QTL are more difficult to detect than common QTL. In the worst case, when $qA2$ was located in the middle of the marker interval and SI = 1 cM, detection power was only 10.7% (Fig. 2B), due to the small additive effect detected in just one family, and only 1.9% of the phenotypic variance was explained (Table 1).

Detection power of rare QTL in NAM populations

When rare QTL overlapped with markers and their additive effects were greater than half a day (Figs. 3A and 3B), we had a power of more than 50.0% to map QTL within 1 cM SI, 60.0% within 2 cM SI, 75.0% within 5 cM SI, and 85.0% within 10 cM SI. Much lower power can be seen in Figs. 3C and 3D for smaller-effect QTL. For example, when a rare QTL had an additive effect of 0.25 day (PVE = 0.10%) in the B73×Il14H family in the maize population and SI = 1 cM, power was 21.2% when QTL overlapped with markers. When QTL were located in the middle of marker intervals, power was lower than when QTL overlapped with markers (Fig. 3). The reduction in power depends not only on the genetic effect of rare QTL, but also on the length of the SI.

Regardless of whether the QTL overlapped with markers or were located in the middle of marker intervals, the variance of power across families was low ($3.9 \times 10^{-5} \sim 0.018$; Fig. 3), indicating that similar detection power can be achieved regardless of the family in which the rare QTL was segregating. For QTL having a half-day or a one-day genetic effect, power in family B73×MS71 was slightly higher than that in other families. This was due to the relatively small phenotypic variance observed in family

B73×MS71 [5], which led to decreased background noise and increased power.

We also compared the power of rare QTL from JICIM and from individual family mapping. In individual family mapping, we used the family in which a rare QTL was segregating. As an example, a rare QTL was assumed to be segregating in family B73×Il14H with various genetic effects (i.e., 0.25, 0.5, 1.0, and 1.5 days; Table 3) and PVE = 0.10%, 0.39%, 1.55%, and 3.49% for the whole maize NAM population, respectively. For a QTL with more than a 1-day effect, power from both joint linkage mapping and individual family mapping was consistently high, while individual family mapping had higher power when the length of the SI was less than 5 cM. Moreover, individual family mapping was able to narrow the SI down to 1 cM with a power of 74.5%, but 59.9% for JICIM when the QTL overlapped with markers. However, for a QTL with less than a 1-day effect, the reverse happens. Power from joint linkage mapping, despite being low, was slightly higher than that from individual family mapping. Power decreased sharply for QTL with a 0.25-day effect from individual family mapping, to below 15.0% in all cases. Similar results were observed for rare QTL segregating in other families.

Dissection of linked QTL

To investigate potential confounding when QTL are in proximity of each other, we used $qZ6$ and $qZ8$ identified in the maize NAM population to design the simulation experiment. In the actual population, they were located on chromosome 2, 49.4 cM apart, and explained 5.20% and 4.11% of the total phenotypic variance, respectively (Table 2). The genetic model used in simulation contained only the effects of $qZ6$ and $qZ8$, and no background genetic effects were included. Correlation of their effects across populations is 0.21.

When $qZ6$ and $qZ8$ were putatively located at 77 cM and 97 cM, from the mean LOD profile we can see that there were two clear peaks around the putative positions of $qZ6$ and $qZ8$ (Fig. 4A). The number of identified QTL was 108 within 75–80 cM and 79 within 95–100 cM (Fig. 5A). The additive effects for $qZ6$ and $qZ8$ were slightly overestimated (Table S2). All simulated populations showed that $qZ6$ and $qZ8$ can be mapped as two distinct QTL. When they were closer, say 5 cM apart, the two LOD peaks were closer as well (Fig. 4B). The numbers of identified

Table 2. Power (%) of the top 30 days-to-silking QTL identified in the maize NAM population under four levels of the support interval, i.e., 1, 2, 5, and 10 cM.

QTL	Chr.	Pos. (cM)	PVE[a] (%)	QTL overlapped with marker				QTL located at the center of marker interval			
				1 cM	2 cM	5 cM	10 cM	1 cM	2 cM	5 cM	10 cM
qZ1	10	41.9	9.55	96.0	97.7	98.4	99.1	76.7	89.6	98.0	98.9
qZ2	8	69.7	6.62	95.8	97.4	98.3	98.8	78.8	91.3	97.8	98.7
qZ3	9	62.9	6.34	95.2	97.5	98.0	98.8	76.9	91.4	98.1	98.7
qZ4	3	56.0	6.25	94.8	97.1	98.0	98.6	77.5	90.5	97.6	98.8
qZ5	1	84.6	5.61	96.1	97.4	97.9	98.6	76.3	90.9	97.6	98.8
qZ6	2	125.9	5.20	94.2	97.2	98.1	98.9	75.8	90.0	98.5	98.8
qZ7	9	57.3	4.33	93.6	97.3	97.9	98.7	74.4	90.0	97.4	98.6
qZ8	2	76.5	4.11	93.9	96.7	98.1	98.8	74.2	88.3	97.6	98.8
qZ9	3	105	3.88	93.6	97.0	98.1	98.9	74.3	89.0	97.4	98.8
qZ10	3	70.6	3.38	93.8	97.9	98.6	99.1	75.7	90.1	97.6	98.6
qZ11	7	76.5	3.28	89.8	96.1	98.1	98.7	72.0	89.4	96.7	98.7
qZ12	1	137.6	2.79	91.0	96.7	98.2	98.7	73.7	89.2	97.4	98.9
qZ13	4	77.2	2.76	89.1	96.4	98.5	99.0	71.7	88.6	97.5	98.8
qZ14	2	107.8	2.74	89.4	95.6	98.2	99.2	71.2	87.7	97.2	98.9
qZ15	1	181.9	2.73	90.2	96.4	98.2	98.9	71.6	89.2	97.3	98.9
qZ16	2	38.6	2.54	91.9	97.4	98.6	99.1	73.4	88.7	97.2	98.7
qZ17	5	78.4	2.52	88.8	96.2	98.6	99.2	70.3	87.8	97.8	98.8
qZ18	6	86.2	2.51	90.4	96.2	98.1	98.8	71.1	87.3	97.3	98.6
qZ19	7	49.2	2.50	91.1	96.7	98.3	99.0	69.8	87.3	97.0	99.1
qZ20	3	128.4	2.41	92.3	96.6	98.0	98.5	70.6	87.4	97.1	98.7
qZ21	5	101.9	2.35	88.1	95.2	98.4	98.6	71.0	87.3	97.2	98.7
qZ22	1	31.7	2.28	89.1	95.8	98.5	99.4	71.2	86.6	97.0	98.7
qZ23	4	47.7	2.23	88.4	96.1	98.5	98.9	70.1	87.3	97.7	98.7
qZ24	6	27.8	2.06	87.7	95.7	98.4	99.0	68.6	85.6	97.3	98.9
qZ25	1	60.8	2.02	88.0	95.1	98.9	99.0	70.4	86.4	97.0	98.6
qZ26	8	119.0	2.00	86.5	94.2	97.9	98.8	69.4	87.3	95.7	98.7
qZ27	10	82.2	1.95	89.1	96.1	97.9	98.6	69.0	87.4	97.1	98.9
qZ28	5	0.0	1.80	88.0	95.0	98.0	98.6	68.3	86.7	96.8	98.6
qZ29	8	18.3	1.60	89.1	95.8	98.5	99.3	69.4	86.4	96.3	98.4
qZ30	4	111.5	1.37	88.0	95.4	98.6	98.9	70.4	86.4	97.0	98.6

[a]Phenotypic variance explained by each QTL; QTL were ordered by PVE.
doi:10.1371/journal.pone.0017573.t002

QTL around $qZ6$ and $qZ8$ were still very high, i.e., 121 and 91, respectively (Fig. 5B). From the individual LOD profiles (Table S3), nearly 70% of the simulated populations can separate $qZ6$ and $qZ8$ properly. When $qZ6$ and $qZ8$ were 1cM apart, only one peak appeared in the mean LOD profile (Fig. 4C), and 112 QTL were detected within 75–80 cM (Fig. 5C). The estimated effects at the peak were almost equal to the sum of the two QTL effects (Table S2), indicating the two linked QTL were detected as one single QTL at an intermediary position.

Discussion

Efficiency of JICIM for complex trait dissection in the NAM design

In the maize and *Arabidopsis* NAM populations, the LOD score of JICIM was much higher than that from individual family mapping, indicating the high mapping power of JICIM. In the maize population, JICIM had a more than 85% chance of locating

QTL within 1 cM SI when QTL overlapped with markers, and a more than 68% chance when QTL were located in the middle of marker intervals (Table 2). For linkage mapping in a biparental population with a size of hundreds, the mapping resolution is always in the 10 cM magnitude due to limited number of recombination that can be surveyed in such population sizes, not to mention that the effect sizes are usually grossly overestimated in such studies. Cross-validation was suggested as a viable option for diminishing the effects of such sampling and resolution issues. In such populations, only if an individual QTL is explaining more than 10% of phenotypic variance, a resolution below 5 cM can be achieved [2], and rigorous testing via bootstrapping and other re-sampling strategies are necessary for accurate estimation of the effect sizes. In addition, although not comparable to the scale of association mapping in diverse populations where hundreds of alleles per locus can be simultaneously tested, in JICIM multiple alleles at the order of tens can easily be tested simultaneously (Tables S1 and 1).
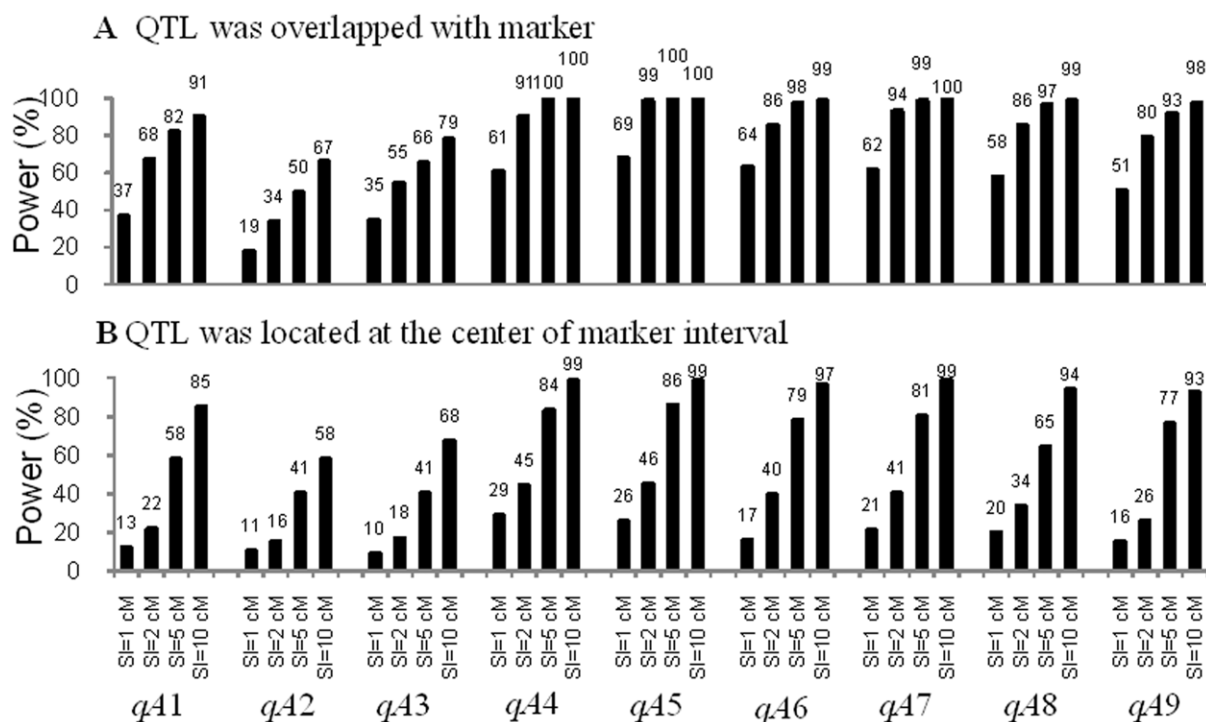
Figure 2. Power of the 9 identified QTL in the _Arabidopsis_ population. Power was calculated as the proportion of runs where each QTL was detected within four levels of the support interval (SI = 1 cM, 2 cM, 5 cM, and 10 cM). Two QTL distribution scenarios were considered, i.e., QTL overlapped with markers (A), and QTL at the center of marker intervals (B).
doi:10.1371/journal.pone.0017573.g002

Flowering time is an adaptive trait that is highly correlated with population structure, as demonstrated in maize [29] and _Arabidopsis_ [30]. Joint linkage mapping of NAM populations better avoids the confounding effect of population structure, and identified most genetic variation of flowering time (Tables S1 and 1). Thus joint linkage mapping provides us with a promising approach to study the genetics of flowering time.

Efficiency of the NAM design for rare QTL detection

The power of QTL detection in a mating design involving multiple crosses depends not only on QTL effects, but also on allele frequency (Table 2; Fig. 2). For the maize population, high detection power has been observed for QTL segregating in 2–4 families (Tables S1 and 1). The power of rare QTL was around 80% for 3-day effects in the _Arabidopsis_ population (Fig. 2), and for a 1-day effect in the maize population (Fig. 3) within 5 cM SI. We were interested in finding out whether it is more efficient to use individual family mapping for rare QTL. Simulation results showed that to map rare QTL with more than a 1-day effect, joint linkage mapping was not of merit, since adding more families with non-significant QTL would increase the sample variance; this, in turn, could cause the QTL signals to become vague, and the position of the QTL could be distorted by neighboring QTL in other families. For small-effect QTL, the power of individual family mapping was slightly lower than the power of joint linkage mapping. One possible reason may be the actual genetic architecture used in simulation as the background effect. In maize NAM, over 98% of the QTL alleles had less than a 1-day effect on days to silking ([5] and Fig. S1). In this sense, the background effects from other families may be helpful for rare QTL with less than a 1-day effect. Considering the experiment expense and the statistical work load of joint linkage mapping, individual family QTL mapping is still useful, especially for rare QTL.

Integration of linkage maps in NAM populations for joint QTL mapping

When building an integrated linkage map for joint linkage mapping or joint association-linkage mapping, variation in recombination frequency among families is commonly observed in some chromosome regions [6]. To avoid the confounding effect of this variation, JICIM recognizes the recombination frequencies across families by recalculating the crossovers in each family. To accentuate the situation further, Buckler et al. [5] found an association between flowering time QTL with larger effects and reduced local recombination rates in the maize NAM population. This is not completely unexpected since flowering time is a characteristic life history trait that may have influence over the species evolution. For regions around _vegetative to generative transition 1_ (_Vgt1_) locus on chromosome 8 [31], recombination frequencies were significantly different across QTL effects estimated in each family and across the population structure of the 26 parental lines [6]. Under such circumstances where there are more sources of confounding for the estimation of the QTL effect sizes, stepwise regression seems to provide a more accurate effect estimation (see Fig. 5 in Buckler et al. [4]) than JICIM (Table S1). However, once the marker density across the _Vgt1_ region is increased, JICIM increase the mapping resolution and give a much more accurate position and effect estimation of _Vgt1_. At the same time, another early flowering QTL located about 5 cM upstream from _Vgt1_ [32] was confirmed by JICIM in the maize population (Fig. S2).

Further considerations

Simulations based on putative genetic models are commonly used to investigate the efficiency of a QTL mapping method, where QTL detection power and false positive rate can both be properly estimated [33–36]. But there is a concern whether the
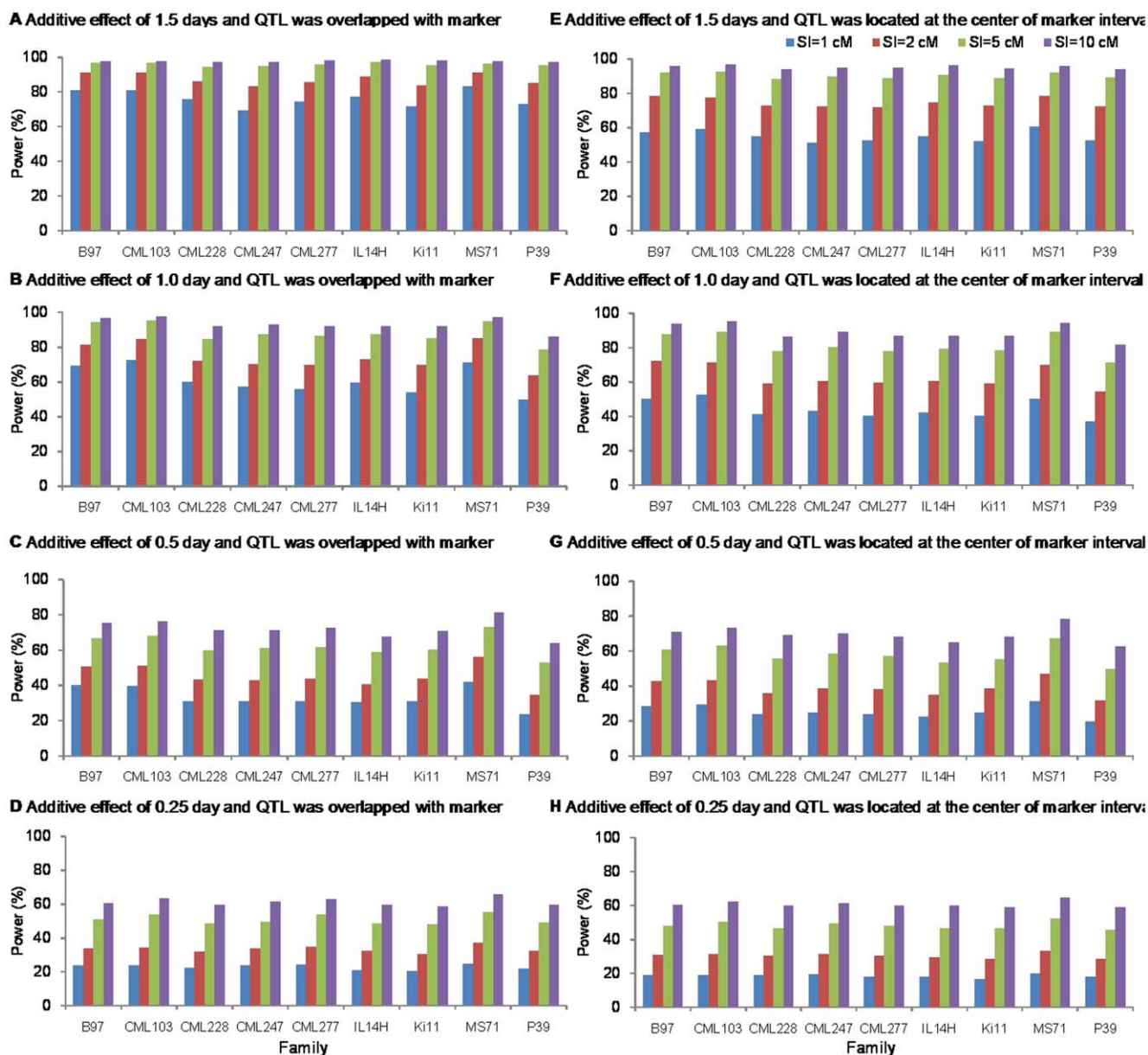
Figure 3. Power of the 9 simulated rare QTL in the maize NAM population. Two scenarios of QTL positions were simulated, i.e., QTL overlapped with markers (A–D), and QTL located at the center of marker intervals (E–H). Four levels of additive effects (1.5 days for A and E, 1.0 day for B and F, 0.5 day for C and G, and 0.25 day for D and H), corresponding to PVE = 13.76%, 6.12%, 1.53%, and 0.38%, were considered. Power was calculated as the proportion of runs where QTL were detected within four levels of the support interval (SI = 1 cM, 2 cM, 5 cM, and 10 cM).
doi:10.1371/journal.pone.0017573.g003

predefined genetic model could represent the realistic model. Simulations in this study used real phenotypic data as the background effect, which represents the realistic model. In this case, it is less likely to distinguish the false positives from the true positives [37]. Instead, the genome wide 2-error rate was used to control the false positives. The power thus determined could be overestimated when the moving QTL is in the vicinity of the previously identified QTL. The simulation study where the residual of the linear model (details in Material and Methods) is used as the background may provide more convinced power.

In this study, we only considered the QTL fusion of two identified QTL under three linkage distances. To make more inclusive conclusions, further studies will be needed by including more factors, such as linkage distance, correlation of
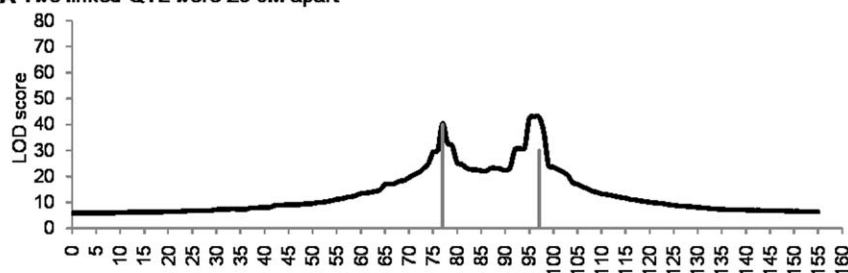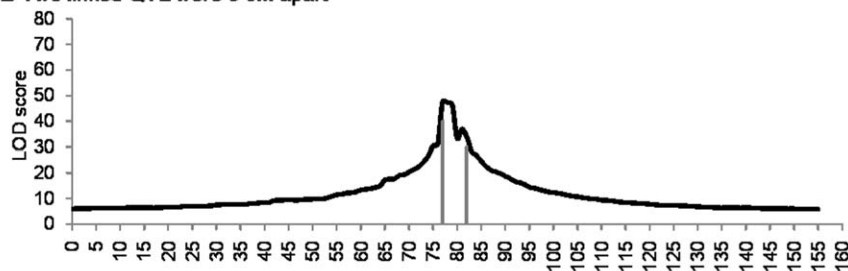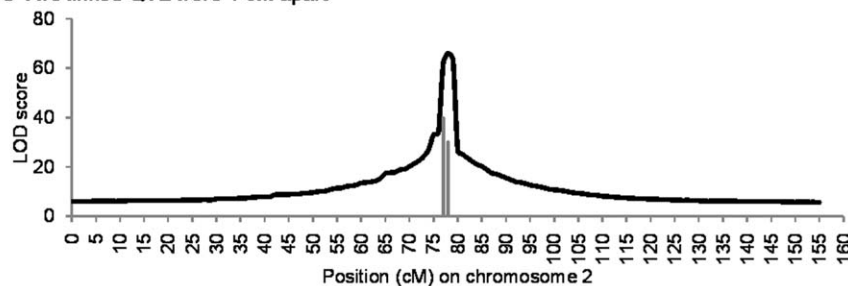
two QTL effects across populations, two QTL segregating in the same populations or different population, and the allelic effects that are correlated positively or negatively across populations, etc.

We used multiple families sharing one common parent to illustrate the algorithm of JICIM. Theoretically, there is no limit to how much JICIM may be extended to other mating designs involving multiple crosses, such as the eight-way cross [38] and the diallel cross [39]. In biparental populations, ICIM is able to detect dominance [36] and digenic epistasis [35]. Further studies are needed to extend JICIM to other mating designs and genetic models including dominance and epistasis. In order to maximize QTL mapping power, comparison of different mating designs, and the balance between the number of families and the number of individuals in each family of a NAM design, can be further investigated by JICIM.

Table 3. Simulated power (%) from JICIM and ICIM for rare QTL segregating in family B73×Il14H under four levels of support interval, i.e., 1, 2, 5, and 10 cM.
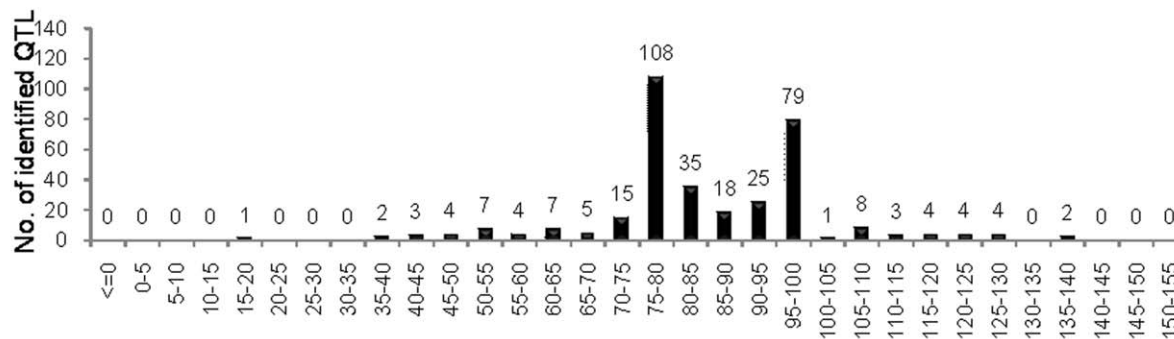
Method	Additive effect (day)	QTL overlapped with marker				QTL located at the center of marker interval			
		1 cM	2 cM	5 cM	10 cM	1 cM	2 cM	5 cM	10 cM
JICIM in the maize NAM population	1.50	77.4	89.1	97.5	98.6	55.2	75.0	91.0	96.4
	1.00	59.9	73.2	87.6	92.0	42.6	60.5	79.7	87.1
	0.50	30.4	40.8	58.8	67.7	22.6	35.1	53.3	64.9
	0.25	21.2	32.5	48.6	59.8	18.1	29.7	46.9	59.9
ICIM in family B73×Il14H	1.50	84.7	92.2	94.8	94.8	51.6	71.2	87.1	93.5
	1.00	74.5	84.7	93.2	94.4	44.4	61.7	80.6	87.1
	0.50	25.3	42.4	53.2	61.0	21.9	33.3	49.5	57.3
	0.25	4.9	6.8	8.8	11.9	3.0	4.0	7.1	10.5

doi:10.1371/journal.pone.0017573.t003

Figure 4. Mean LOD profile of the second chromosome from 100 simulation runs. Two linked QTL were respectively located at 77 cM and 97 cM (A), 77 cM and 82 cM (B), and 77 cM and 78 cM (C) on chromosome 2. The two vertical lines indicate the two QTL. The length of each line was proportional to the size of its corresponding QTL. The mean LOD profile of the other chromosomes was almost equal to zero, and thus was not shown.
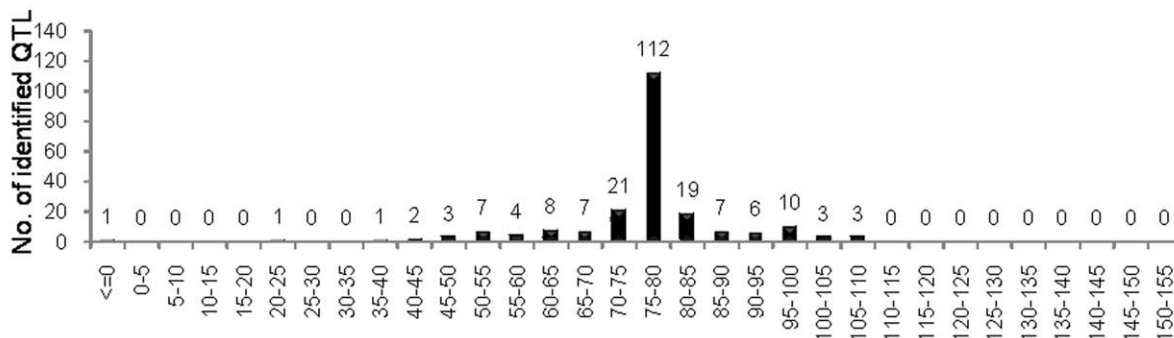doi:10.1371/journal.pone.0017573.g004

Figure 5. Number of identified QTL in each marker interval on the second chromosome. Two linked QTL were respectively located at 77 cM and 97 cM (A), 77 cM and 82 cM (B), and 77 cM and 78 cM (C) on chromosome 2. The LOD threshold was 12.26.
doi:10.1371/journal.pone.0017573.g005

Materials and Methods

Statistical models in joint QTL mapping

For multiple crosses sharing one common parent, we assume that there are F ($F \geq 1$) families with a size of n_f individuals in the fth family, and a total population size of N ($N = \sum_{f=1}^{F} n_f$). The proposed statistical method consists of two steps. In the first step, GLM was used where population and population-by-marker interactions were treated as fixed effects. Each marker has $F+1$ levels (the common parent and the other F founders) in the genetic linkage map. These parameters were included in the following model,

$$\mathbf{Y} = b_0 + \alpha\mathbf{u} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (1)$$

where $\mathbf{Y}$ is the vector of phenotypic values; b_0 is the intercept; $\mathbf{u}^T = (u_1, u_2, \ldots, u_F)$ is the effect vector indicating the cross effect of each founder with the common parent; α is the $N \times F$ incidence matrix relating each u_f ($f = 1, 2, \ldots, F$) to $\mathbf{Y}$; $\boldsymbol{\beta}$ is the $[(F+1)m] \times 1$ effect vector of the $N \times [(F+1)m]$ incidence matrix $\mathbf{X}$ where m is the number of markers; and ε is the vector of residuals. To avoid over-fitting, we used stepwise regression to estimate the parameters in model [1]. If the regression variable did not enter into the model, the corresponding coefficient was set at 0.

Table 4. Genotypic distribution of QTL in the fth family under four marker types in mapping interval $(k, k+1)$, adapted with modification from Table 1, Li et al. (2007).

Group	Sample size	Frequency	Marker genotype		Frequency of QTL genotype		Distribution of Δy_i
			k	$k+1$	Q_fQ_f	q_0q_0	
1	n_{f1}	$\frac{1}{2}(1-r_{k,k+1})$	+	+	p_{f1}	1- p_{f1}	$p_{f1}N(\mu_f,\sigma_f{}^2)+(1-p_{f1})N(\mu_0,\sigma_f{}^2)$
2	n_{f2}	$\frac{1}{2}r_{k,k+1}$	+	-	p_{f2}	1- p_{f2}	$p_{f2}N(\mu_f,\sigma_f{}^2)+(1-p_{f2})N(\mu_0,\sigma_f{}^2)$
3	n_{f3}	$\frac{1}{2}r_{k,k+1}$	-	+	p_{f3}	1-p_{f3}	$p_{f3}N(\mu_f,\sigma_f{}^2)+(1-p_{f3})N(\mu_0,\sigma_f{}^2)$
4	n_{f4}	$\frac{1}{2}(1-r_{k,k+1})$	-	-	p_{f4}	1-p_{f4}	$p_{f4}N(\mu_f,\sigma_f{}^2)+(1-p_{f4})N(\mu_0,\sigma_f{}^2)$

Note: $n_{f1}+n_{f2}+n_{f3}+n_{f4}=n_f$, where n_f is the size of the fth family. Q_f is the allele in the fth parent, and Q_0 is the allele in the common parent.
$p_{f1}=(1-r_{j,q})(1-r_{q,j+1})/(1-r_{j,j+1})$, $p_{f2}=(1-r_{j,q})r_{q,j+1}/r_{j,j+1}$, $p_{f3}=1-p_{f2}$, and $p_{f4}=1-p_{f1}$, where $r_{j,q}$, $r_{q,j+1}$ and $r_{j,j+1}$ are the recombination frequencies in the fth family between marker j and the putative QTL, between the putative QTL and marker $j+1$, and between markers j and $j+1$, respectively. "+" and "-" denote two types of homozygote for the marker genotype. $N(\mu_f,\sigma_f{}^2)$ and $N(\mu_0,\sigma_f{}^2)$ represent the distributions for the two QTL genotypes Q_fQ_f and q_0q_0 in the fth family, respectively. All missing marker types were assigned to either the common parent genotype, or the non common parent genotype through an imputation algorithm that considers the linkage relationship between markers within each family [37]; thus this table assumes no missing genotypic data.
doi:10.1371/journal.pone.0017573.t004

Based on coefficient estimation, one-dimensional scanning is conducted in the second step, which is similar to the second step of inclusive composite interval mapping (ICIM) [33,34]. At a testing position in a marker interval $(k, k+1)$, all individuals in the fth family can be classified into four groups based on the allelic states of the two flanking markers (Table 4). All missing marker types were assigned to either the common parent genotype, or the non common parent genotype through an imputation algorithm that considers linkage relationship between markers within each family [40]. Phenotypic value in model [1] can be adjusted by the estimated coefficients in the first step to exclude the influence of QTL on other chromosomal regions, i.e.,

$$\Delta y_{if} = y_{if} - \alpha_{if}u_f - \sum_{j\neq k,k+1} \hat{b}_j x_{ifj},$$

where $i = 1, 2, \ldots, n_f, f = 1, 2, \ldots, F$ and the current scanning position is within the interval determined by the kth and $k+1^{th}$ marker.

If there is a QTL (with the two alleles in the fth family denoted as Q_f and Q_0) at the testing position, Δy_{if} $(i = 1, 2, \ldots, n_f)$ follows a mixture distribution consisting of the distribution of $N(\mu_f,\sigma_f{}^2)$ for individuals with QTL genotype Q_fQ_f, and the distribution of $N(\mu_0,\sigma_f{}^2)$ for individuals with QTL genotype Q_0Q_0. Proportions of the two component distributions depend on the recombination frequencies between the QTL and its two flanking markers. The existence of a QTL at the current mapping position can be tested by the following hypotheses:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_F = \mu_0 \text{ vs.}$$

H_A: at least one of $\mu_1, \mu_2, \ldots,$ and μ_F not equal to μ_0.
The log-likelihood function under the alternative hypothesis H_A is,

$$L_A = \sum_{f=1}^{F}\sum_{j=1}^{4}\sum_{\substack{i=1 \\ i\in S_j}}^{n_f} \log\left[p_{fj}\Phi(\Delta y_{if};\mu_f,\sigma_f{}^2)+(1-p_{fj})\Phi(\Delta y_{if};\mu_0,\sigma_f{}^2)\right], (2)$$

where S_j denotes the jth marker type group $(j = 1, 2, 3,$ and 4; Table 4); p_{fj}, the proportion of QTL genotypes Q_fQ_f in the jth

marker group of the fth family, was calculated from the recombination frequencies between two flanking markers in the current scanning interval, and between either side of the flanking marker and the putative QTL located between the interval; μ_f and μ_0 are the means for QTL genotypes Q_fQ_f and Q_0Q_0, respectively, and $\Phi(\bullet; \mu_\bullet,\sigma_\bullet{}^2)$ represents the probability density of the normal distribution. The expectation and maximization (EM) algorithm was used to estimate the $F+1$ means and F variances in equation [2], and their maximum likelihood estimates are represented by $\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_F$, and $\hat{\sigma}_1{}^2, \hat{\sigma}_2{}^2, \ldots,$ and $\hat{\sigma}_F{}^2$, from which the additive effect of the putative QTL in each family (denoted by $\hat{a}_f$) can be estimated as $\hat{a}_f = \frac{1}{2}(\hat{\mu}_f - \hat{\mu}_0)$.

In QTL mapping for a bi-parental population, when we calculated p_{fj}, the recombination frequencies between the two flanking markers in the current scanning interval and between either side of the flanking marker and the putative QTL located within the interval were estimated from the marker positions in the linkage map through mapping function, such as the Morgan, Haldane, and Kosombi function, etc. [41]. To deal with the variation in recombination frequency across families in joint linkage mapping by a consensus linkage map, recombination frequencies were estimated from the observed crossovers in the fth family.

Under the null hypothesis H_0, Δy_{if} $(i = 1, \ldots, n_f)$ within each family follow a normal distribution denoted as $N(\mu_0,\sigma_f{}^2)$. The mean and variance of this distribution can be estimated as,

$$\hat{\mu}_0 = \frac{1}{N}\sum_{f=1}^{F}\sum_{i=1}^{n_f}\Delta y_{if} \text{ and } \hat{\sigma}_f{}^2 = \frac{1}{n_f}\sum_{i=1}^{n_f}(\Delta y_{if}-\hat{\mu}_0)^2.$$

Thus, the maxima log-likelihood function under the null hypothesis H_0 is,

$$L_0 = \sum_{f=1}^{F}\sum_{i=1}^{n_f}\ln f(\Delta y_{if};\hat{\mu}_0,\hat{\sigma}_f{}^2).$$

The LOD score at the testing position can be calculated from the log-likelihoods under the two hypotheses. If LOD score decays

at least one in both sides of the higher-than-LOD threshold peak, the position corresponding to that higher-than-LOD threshold peak is declared as the existence of a QTL. In this sense, each QTL was identified by its one-LOD support interval.

Phenotypic variance explained by each QTL

In a biparental population, such as a RIL population, the theoretical genetic variance for each QTL is a^2, where a is the additive effect of that QTL, defined as half of the difference between the two homozygous QTL types. Ignoring segregation distortion, the phenotypic variance explained (PVE) is,

$$PVE = \frac{a^2}{V_P},$$

where V_P is the phenotypic variance observed in the biparental population. For the NAM design, assuming there are F families, the genotypic value of the common parent QTL genotype and those of the other F parents are μ_0, $\mu_0 + 2a_1$, ..., $\mu_0 + 2a_F$, respectively, and the corresponding frequencies of these $F+1$ genotypic values are $\frac{1}{2}$, $\frac{n_1}{2N}$, ..., $\frac{n_F}{2N}$, respectively. Therefore, the genetic variance of the $F+1$ QTL genotypes in the NAM population can be calculated as,

$$V_{QTL} = 2\sum_{f=1}^{F}\frac{n_f}{N}a_f^2 - \left(\sum_{f=1}^{F}\frac{n_f}{N}a_f\right)^2,$$

Thus,

$$PVE = \frac{2\sum_{f=1}^{F}\frac{n_f}{N}a_f^2 - \left(\sum_{f=1}^{F}\frac{n_f}{N}a_f\right)^2}{V_P},$$

where $V_P = \sum_{f=1}^{F}\frac{n_f}{N}V_{Pf}$ and V_{Pf} is the phenotypic variance of the fth family. The phenotypic variance used here is the weighted average of variances of the F families. The difference between family means contributes to the phenotypic variance of the whole NAM population, which is not suitable to calculate PVE of any QTL in the NAM population. It can be proved that the weighted variance of the F families is equal to the variance of the NAM population adjusted by family means. It should be noted that the total genotypic variance of all the identified QTL cannot be viewed as the sum of genotypic variance of individual QTL, due to the possible linkage between them.

The maize and *Arabidopsis* NAM populations

The maize NAM population has 25 biparental families sharing the common parent B73. From each cross, 200 RILs were produced, resulting in a total of 5000 lines. 4699 RILs and 1106 markers were used to construct the consensus linkage map, which has a total length of 1400 cM, and one marker every 1.3 cM on average [6]. All the genetic markers show polymorphism between B73 and all the other 25 parents. The best linear unbiased prediction of days to silking for each RIL across environments [5] was used as the phenotypic data in joint QTL linkage mapping.

To give another example, we combined three *Arabidopsis* RIL families [21], i.e., L*er* (Landsberg *erecta*) ×Antwerp (An-1), L*er* ×

Kashmir (Kas-2), and L*er* × Kondara (Kond), in a multiple cross design sharing one common parent. For our purposes, we only obtained the raw marker and phenotype data from the original study, and reconstructed the genetic maps and QTL scans. A total of 109 genetic markers showed polymorphism between L*er* and the other 3 parents. A similar strategy as used in the maize population [6] was adopted to construct the consensus map in the *Arabidopsis* population by MAPMAKER 3.0 [42]. The L*er* allele was designated as the "A" parent allele, the other three parent alleles were designated as the "B" parent alleles, and heterozygous loci were converted to missing data. In addition, markers that were non-polymorphic in a particular family were converted to missing data. (Marker SNP395 has a chromosomal conflict, since it could be located on either chromosome 4 or 5; therefore this marker was not used for further joint linkage mapping.) The consensus map thus obtained consists of 108 marker loci, and has a total genetic length of 499.7 cM and an average marker density of 4.81 cM (Fig. S3).

Model selection criteria in JICIM

Permutation tests were conducted using days to flowering to determine the criteria of model selection in the first step of JICIM. To permute the data, overall and population means were fitted to the original data, and predicted values and residuals were calculated. The residual was randomized, and then added back to the predicted value [43]. Associations of all markers for the updated dataset were re-calculated and the lowest P-value for all markers was identified. This procedure was repeated 1000 times to establish a distribution of P-values to test the null hypothesis. For both populations, the P-value corresponding to the overall type I error $\alpha = 0.05$ was approximately 10^{-4}. Thus in the first step of JICIM, we used $P = 10^{-4}$ as the probability for markers entering into the model, and $P = 2 \times 10^{-4}$ as the probability for markers moving out of the model.

Power simulation of the identified QTL

To determine the reliability of identified QTL and the mapping resolution of JICIM, genetic effects and allele frequency in simulation were the same as those of QTL identified in the maize or *Arabidopsis* population, which represented the general properties of the observed genetic architecture. To avoid simplifying simulated model, and to simulate the genetic scheme as genuinely as possible, we used the real phenotype as the background, that is,

$$P_i = P_i^{real} + a_f g_i$$

where $i = 1, 2, ..., n_f$; P_i is the simulated phenotype for the ith individual in the fth family, P_i^{real} is the real phenotype for the ith individual in the fth family, a_f is the additive effect of one of the identified QTL in maize or *Arabidopsis* in the fth family, and g_i is the genotype of the identified QTL for the ith individual in the fth family.

We considered two scenarios of QTL distribution in our simulation experiments. In the first scenario, the number of simulation runs was equal to the number of markers, i.e., 1106 for the maize population, and 108 for the *Arabidopsis* population. In the first simulation run, the QTL was located at the first marker locus; in the second run, the QTL was at the second marker locus; and so on. Thus, for each individual, the QTL genotype was the same as its overlapped marker. In the second scenario, the number of simulations was equal to the number of marker intervals, and QTL were assigned to the middle of marker intervals. In the first

simulation run, the QTL was located in the middle of the first marker interval; in the second run, the QTL was in the middle of the second marker interval; and so on. Thus the QTL genotype has equal probabilities of being the genotype on either side of the two flanking markers. That is, the QTL genotype was assigned as the genotype of left flanking markers with a probability of 0.5, and as the genotype of right flanking markers with a probability of 0.5. The number of runs was 1096 for the maize population, and 103 for the *Arabidopsis* population.

Each of the top 30 and 9 identified QTL in the maize and *Arabidopsis* populations, respectively, was simulated to determine whether its effect could be detected by the LOD threshold from permutation tests within a pre-fixed length of the support interval (SI) centered at the true QTL position. The proportion of detected QTL across markers or marker intervals was used as the detection power of each simulated QTL. Because simulation in this study used real phenotypic data as the background effect, it was less likely to distinguish the false positives from the true positives in the simulated population [37]. Therefore, we used the genome-wide k-error rate (GWER(k)) [20] to guard against more than k false positives.

Rare QTL Simulation

To investigate the possibility of incorrectly mapping linked QTL that segregate in only a few populations as a single QTL, we simulated nine rare QTL in maize population, that segregate in one family with four effects, i.e., 1.5, 1.0, 0.5 and 0.25 days, corresponding to PVE = 3.49%, 1.55%, 0.39%, and 0.10%, respectively. The nine rare QTL were assumed to be segregating in families B73×B97, B73×CML103, B73×CML228, B73×CML247, B73×CML277, B73×IL14H, B73×Ki11, B73×MS71, and B73×P39, respectively. Phenotypic variances in these families are 2.61, 2.08, 7.85, 8.26, 9.82, 5.41, 9.17, 2.62 and 5.59, respectively. For each of the 36 combinations of nine rare QTL and four genetic effects, two QTL distribution scenarios were considered for power simulation, as described in the previous section.

Linked QTL Simulation

To evaluate potential confounding when QTL are located in proximity of each other, we designed additional simulation experiments based on the maize NAM linkage map and two linked QTL explaining similar phenotypic variance in the maize NAM population. To keep constant the phenotypic variance explained by the two QTL, the phenotypic variance of each simulated family was equal to that of the maize NAM population. The error variance in each simulated family was calculated by $V_\varepsilon = V_P - a_1^2 - a_2^2$, where V_P is the phenotypic variance of the family in the real maize NAM population, and a_1 and a_2 are the additive effects of the two linked QTL, respectively. Thus, a_1, a_2, and V_ε were used to simulate the phenotypic value of the fth simulated family. No background genetic effect was considered, and this was generally adopted to compare different QTL mapping methods [33–36]. The principle described above has been used to investigate the effect of population size and marker density on the detection of coupling and repulsive linkage in ICIM [2].

Supporting Information

Figure S1 A The distribution of QTL across 25 maize NAM families; B Histogram of additive allele estimates for the 52 days to silking QTLs for 25 founder lines relative to B73. Count of effects increasing flowering time above the line, and decreasing flowering time below the line.
(TIF)

Figure S2 LOD score profile for chromosome 8 from the joint inclusive composite interval mapping (JICIM) and general linear model (GLM). Both JICIM and GLM were implemented when including a miniature transposon (*MITE*) as a marker, which associated a previously identified vgt1 allele from northern germplasm. In this case, *vgt1* were accurately identified, while another gene (*vgt2*) associated flowering time were also identified.
(TIF)

Figure S3 The consensus linkage map from three RIL families in the *Arabidopsis* NAM population.
(TIF)

Table S1 Fifty-two identified QTL for maize NAM population by JICIM.
(XLS)

Table S2 The estimated QTL effects from 100 simulation runs to test QTL confounding.
(XLS)

Table S3 LOD profiles on chromosome 2 for the 100 simulation runs when QZ6 and QZ8 were located at 77cM and 82cM on chromosome 2.
(XLS)

Author Contributions

Conceived and designed the experiments: HL ESB JW. Performed the experiments: HL. Analyzed the data: HL PB EE JW. Contributed reagents/materials/analysis tools: ESB. Wrote the paper: HL JW.

References

1. Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In *Proceedings of the Forty-Ninth Annual Corn & Sorghum Industry Research Conference* Washington, DC: American Seed Trade Association. pp 250–266.
2. Li H, Hearne S, Banziger M, Li Z, Wang J (2010) Statistical properties of QTL linkage mapping in biparental genetic populations. Heredity 105: 257–267.
3. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6: 95–108.
4. Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178: 539–551.
5. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, et al. (2009) The genetic architecture of maize flowering time. Science 325: 714–718.
6. McMullen MD, Kresovich S, Villeda HS, Bradbury PJ, Li H, et al. (2009) Genetic properties of the maize nested association mapping population. Science 325: 737–740.
7. Rebaï A, Goffinet B (1993) Power of tests for QTL detection using replicated progenies derived from a diallel cross. Theor Appl Genet 86: 1014–1022.
8. Rebaï A, Goffinet B (2000) More about quantitative trait locus mapping with diallel designs. Genet Res 75: 243–247.
9. Xu S (1998) Mapping quantitative trait loci using multiple families of line crosses. Genetics 148: 517–524.
10. Liu Y, Zeng Z-B (2000) A general mixture model approach for mapping quantitative trait loci from diverse cross designs involving multiple inbred lines. Genet Res Camb 75: 345–355.
11. Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136: 1457–1468.
12. Jourjon M-F, Jasson S, Marcel J, Ngom B, Mangin B (2005) MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinformatics 21: 128–130.
13. Jannink J-L, Jansen R (2001) Mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics 157: 445–454.

14. Crepieux S, Lebreton C, Servin B, Charmet G (2004) Quantitative trait loci (QTL) detection in multicross inbred designs: recovering QTL identical-by-descent status information from marker data. Genetics 168: 1737–1749.
15. Yi N, Xu S (2001) Bayesian mapping of quantitative trait loci under complicated mating designs. Genetics 157: 1759–1771.
16. Yi N, Xu S (2002) Linkage analysis of quantitative trait loci in multiple line crosses. Genetica 114: 217–230.
17. Fang M, Jiang D, Chen X, Pu L, Liu S (2008) Bayesian analysis of genetic architecture of quantitative traits using data of crosses of multiple inbred lines. Genetica 134: 367–375.
18. Hayashi T, Iwata H (2009) Bayesian QTL mapping for multiple families derived from crossing a set of inbred lines to a reference line. Heredity 102: 497–505.
19. Xu S, Jia Z (2007) Genome-wide analysis of epistatic effects for quantitive traits in barley. Genetics 175: 1955–1963.
20. Chen L, Storey JD (2006) Relaxed significance criteria for linkage analysis. Genetics 173: 2371–2381.
21. El-Lithy ME, Bentsink L, Hanhart CJ, Ruys GJ, Rovito D, et al. (2006) New Arabidopsis recombination inbred line population genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. Genetics 172: 1867–1876.
22. Komeda Y (2004) Genetic regulation of time to flower in Arabidopsis thaliana. Annu Rev Plant Biol 55: 521–535.
23. Baurle I, Dean C (2006) The timing of developmental transitions in plants. Cell 125: 655–664.
24. Roux F, Touzet P, Cuguen J, Le Corre V (2006) How to be early flowering: an evolutionary perspective. Trends Plant Sci 11: 375–381.
25. Ehrenreich IM, Hanzawa Y, Chou L, Roe JL, Kover PX, et al. (2009) Candidate gene association mapping of Arabidopsis flowering time. Genetics 183: 325–335.
26. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, et al. (2010) Linkage and association mapping of Arabidopsis thaliana flowering time in nature. PLoS Genet 6(5): e1000940. doi:10.1371/journal.pgen.1000940.
27. Darvasi A, Weinreb A, Minke V, Wellert JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134: 943–951.
28. Guo B, Sleper DA, Beavis WD (2010) Nested association mapping for identification of functional markers. Genetics 186: 373–383.
29. Flint-Garcia SA, Thuillet A-C, Yu J, Pressoir G, Romero SM, et al. (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44: 1054–1064.
30. Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, et al. (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1: e60. doi:10.1371/journal.pgen.0010060.
31. Salvi S, Sponza G, Morgante M, Tomes D, Niu X, et al. (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci USA 104: 11376–11381.
32. Chardon F, Virlon B, Moreau L, Falque M, Joets J (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.
33. Li H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. Genetics 175: 361–374.
34. Wang J (2009) Inclusive composite interval mapping of quantitative trait genes. Acta Agronomica Sinica 35: 239–245.
35. Li H, Ribaut J-M, Li Z, Wang J (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. Theor Appl Genet 116: 243–260.
36. Zhang L, Li H, Li Z, Wang J (2008) Interactions between markers can be caused by the dominance effect of QTL. Genetics 180: 1177–1190.
37. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38: 203–208.
38. The Complex Trait Consortium (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet 36: 1133–1137.
39. Christie BR, Shattuck VI (1992) The diallel cross: design, analysis and use for plant breeders. Plant Breed Rev 9: 9–36.
40. Zhang L, Wang S, Li H, Deng Q, Zheng A, et al. (2010) Effects of missing marker and segregation distortion on QTL mapping in F_2 populations. Theor Appl Genet 121: 1071–1082.
41. Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. SunderlandMassachusetts: Sinauer Associates, Inc.
42. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, et al. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174–181.
43. Anderson MJ, Ter Braak CJF (2003) Permutation tests for multi-factorial analysis of variance. J Stat Comput Sim 73: 85–113.