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

Huihui Li¹,², Peter Bradbury³, Elhan Ersoz², Edward S. Buckler²,³, Jiankang Wang¹∗

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, recombinant 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. Rebai 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 IBID (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 QTLs, which is essential for gene fine-mapping and map-based cloning, especially under moderate marker density. More recently, Hayashi and Ishiwata [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 QT 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 multipopulation QT data [7–10,12–17], none of these methods are proposed for simultaneous analysis of such networked multipopulation QT 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, 32 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 32 QTL, the top 30 QTL (denoted as q1-q30) are identified in one-LOD support interval across the whole genome, explaining 93% 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) that explained 93% of the total phenotypic variance (Table 1 and Figs. 1 and 2). 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 qA6 in Table 1, had significant effects in only one family. Therefore, qA1-qA3 and qA6 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 JICIM to detect. Therefore, qA1-qA3 and qA6 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.
Table 1. Nine flowering time QTL identified in the Arabidopsis NAM population by JICIM.

<table>
<thead>
<tr>
<th>QTL</th>
<th>Chr.</th>
<th>Pos. (cM)</th>
<th>LOD score</th>
<th>PVE (%)</th>
<th>Left marker (cM)b</th>
<th>Right marker (cM)b</th>
<th>Additive genetic effect (day) in each familyc</th>
<th>PVE (%) in each family</th>
<th>Flowering genes around QTLd</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ler &gt; An-1</td>
<td>Ler &gt; Kas-2</td>
<td>Ler &gt; Kond</td>
</tr>
<tr>
<td>qA1</td>
<td>1</td>
<td>66.0</td>
<td>3.75</td>
<td>3.92</td>
<td>CIW1* (65.2)</td>
<td>F608.94 (69.4)</td>
<td>-0.28</td>
<td>0.60</td>
<td>-1.32 (-3.00)</td>
</tr>
<tr>
<td>qA2</td>
<td>1</td>
<td>107.0</td>
<td>4.69</td>
<td>1.03</td>
<td>SNP157 (104.4)</td>
<td>SNP110* (107.8)</td>
<td>-0.06</td>
<td>0.66 (3.00)</td>
<td>0.50</td>
</tr>
<tr>
<td>qA3</td>
<td>2</td>
<td>0.0</td>
<td>3.05</td>
<td>2.99</td>
<td>mrsat2.5* (0.0)</td>
<td>mrsat2.5* (0.0)</td>
<td>0.01</td>
<td>-0.19</td>
<td>-1.44 (-3.00)</td>
</tr>
<tr>
<td>qA4</td>
<td>3</td>
<td>0.0</td>
<td>13.39</td>
<td>3.60</td>
<td>SNP105* (0.0)</td>
<td>mrsat172* (2.9)</td>
<td>1.16</td>
<td>0.53 (2.00)</td>
<td>1.37 (3.00)</td>
</tr>
<tr>
<td>qA5</td>
<td>4</td>
<td>3.0</td>
<td>43.23</td>
<td>53.84</td>
<td>mrsat441 (0.0)</td>
<td>FRP* (3.6)</td>
<td>0.41</td>
<td>-2.80 (-6.00)</td>
<td>-5.62 (-11.40)</td>
</tr>
<tr>
<td>qA6</td>
<td>4</td>
<td>49.0</td>
<td>8.51</td>
<td>7.38</td>
<td>SNP295* (46.6)</td>
<td>SNP199 (49.9)</td>
<td>-0.03</td>
<td>-1.95 (-2.60)</td>
<td>-1.01</td>
</tr>
<tr>
<td>qA7</td>
<td>5</td>
<td>19.0</td>
<td>13.67</td>
<td>5.97</td>
<td>SNP136* (17.3)</td>
<td>SNP358* (20.6)</td>
<td>0.75</td>
<td>-1.39 (-3.40)</td>
<td>-1.74 (-2.8)</td>
</tr>
<tr>
<td>qA8</td>
<td>5</td>
<td>33.0</td>
<td>5.87</td>
<td>6.38</td>
<td>SNP230* (32.1)</td>
<td>nga139 (33.5)</td>
<td>0.03</td>
<td>-0.50</td>
<td>-2.09 (-5.00)</td>
</tr>
<tr>
<td>qA9</td>
<td>5</td>
<td>93.0</td>
<td>8.00</td>
<td>5.43</td>
<td>SNP101 (92.5)</td>
<td>SNP304 (93.8)</td>
<td>-0.56</td>
<td>1.38</td>
<td>-0.51</td>
</tr>
</tbody>
</table>

*aPhenotypic variance explained by each QTL;  
bMarkers with asterisks were significant in individual family mapping, reported by El-Lithy et al. (2006);  
cValues in brackets are two allele’s effects (i.e., 2a) estimated from individual family mapping.  
dAnnotations according to reports by Brachi et al. [26], and Ehrenreich et 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
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 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).

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 x 10^-5 to 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 qA6 and qA8 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 qA6 and qA8, and no background genetic effects were included. Correlation of their effects across populations is 0.21.

When qA6 and qA8 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 qA6 and qA8 (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 qA6 and qA8 were slightly overestimated (Table S2). All simulated populations showed that qA6 and qA8 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

---

**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
QTL around \( qZ_6 \) and \( qZ_8 \) 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 \( qZ_6 \) and \( qZ_8 \) properly. When \( qZ_6 \) and \( qZ_8 \) were 1 cM 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).

<table>
<thead>
<tr>
<th>QTL</th>
<th>Chr.</th>
<th>Pos. (cM)</th>
<th>PVEa (%)</th>
<th>QTL overlapped with marker</th>
<th>QTL located at the center of marker interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 cM</td>
<td>2 cM</td>
</tr>
<tr>
<td>qZ1</td>
<td>10</td>
<td>41.9</td>
<td>9.55</td>
<td>90.0</td>
<td>97.7</td>
</tr>
<tr>
<td>qZ2</td>
<td>8</td>
<td>69.7</td>
<td>6.62</td>
<td>95.8</td>
<td>97.4</td>
</tr>
<tr>
<td>qZ3</td>
<td>9</td>
<td>62.9</td>
<td>6.34</td>
<td>95.2</td>
<td>97.5</td>
</tr>
<tr>
<td>qZ4</td>
<td>3</td>
<td>56.0</td>
<td>6.25</td>
<td>94.8</td>
<td>97.1</td>
</tr>
<tr>
<td>qZ5</td>
<td>1</td>
<td>84.6</td>
<td>5.61</td>
<td>96.1</td>
<td>97.4</td>
</tr>
<tr>
<td>qZ6</td>
<td>2</td>
<td>125.9</td>
<td>5.20</td>
<td>94.2</td>
<td>97.2</td>
</tr>
<tr>
<td>qZ7</td>
<td>9</td>
<td>57.3</td>
<td>4.33</td>
<td>93.6</td>
<td>97.3</td>
</tr>
<tr>
<td>qZ8</td>
<td>2</td>
<td>76.5</td>
<td>4.11</td>
<td>93.9</td>
<td>96.7</td>
</tr>
<tr>
<td>qZ9</td>
<td>3</td>
<td>105</td>
<td>3.88</td>
<td>93.6</td>
<td>97.0</td>
</tr>
<tr>
<td>qZ10</td>
<td>3</td>
<td>70.6</td>
<td>3.38</td>
<td>93.8</td>
<td>97.9</td>
</tr>
<tr>
<td>qZ11</td>
<td>7</td>
<td>76.5</td>
<td>3.28</td>
<td>89.8</td>
<td>96.1</td>
</tr>
<tr>
<td>qZ12</td>
<td>1</td>
<td>157.6</td>
<td>2.79</td>
<td>91.0</td>
<td>96.7</td>
</tr>
<tr>
<td>qZ13</td>
<td>4</td>
<td>77.2</td>
<td>2.76</td>
<td>89.1</td>
<td>96.4</td>
</tr>
<tr>
<td>qZ14</td>
<td>2</td>
<td>107.8</td>
<td>2.74</td>
<td>89.4</td>
<td>95.6</td>
</tr>
<tr>
<td>qZ15</td>
<td>1</td>
<td>181.9</td>
<td>2.73</td>
<td>90.2</td>
<td>96.4</td>
</tr>
<tr>
<td>qZ16</td>
<td>2</td>
<td>38.6</td>
<td>2.54</td>
<td>91.9</td>
<td>97.4</td>
</tr>
<tr>
<td>qZ17</td>
<td>5</td>
<td>78.4</td>
<td>2.52</td>
<td>88.8</td>
<td>96.2</td>
</tr>
<tr>
<td>qZ18</td>
<td>6</td>
<td>86.2</td>
<td>2.51</td>
<td>90.4</td>
<td>96.2</td>
</tr>
<tr>
<td>qZ19</td>
<td>7</td>
<td>49.2</td>
<td>2.50</td>
<td>91.1</td>
<td>96.7</td>
</tr>
<tr>
<td>qZ20</td>
<td>3</td>
<td>128.4</td>
<td>2.41</td>
<td>92.3</td>
<td>96.6</td>
</tr>
<tr>
<td>qZ21</td>
<td>5</td>
<td>101.9</td>
<td>2.35</td>
<td>88.1</td>
<td>95.2</td>
</tr>
<tr>
<td>qZ22</td>
<td>1</td>
<td>31.7</td>
<td>2.28</td>
<td>89.1</td>
<td>95.8</td>
</tr>
<tr>
<td>qZ23</td>
<td>4</td>
<td>47.7</td>
<td>2.23</td>
<td>88.4</td>
<td>96.1</td>
</tr>
<tr>
<td>qZ24</td>
<td>6</td>
<td>27.8</td>
<td>2.06</td>
<td>87.7</td>
<td>95.7</td>
</tr>
<tr>
<td>qZ25</td>
<td>1</td>
<td>60.8</td>
<td>2.02</td>
<td>88.0</td>
<td>95.1</td>
</tr>
<tr>
<td>qZ26</td>
<td>8</td>
<td>119.0</td>
<td>2.00</td>
<td>86.5</td>
<td>94.2</td>
</tr>
<tr>
<td>qZ27</td>
<td>10</td>
<td>82.2</td>
<td>1.95</td>
<td>89.1</td>
<td>96.1</td>
</tr>
<tr>
<td>qZ28</td>
<td>5</td>
<td>0.0</td>
<td>1.80</td>
<td>88.9</td>
<td>95.0</td>
</tr>
<tr>
<td>qZ29</td>
<td>8</td>
<td>18.3</td>
<td>1.60</td>
<td>89.1</td>
<td>95.8</td>
</tr>
<tr>
<td>qZ30</td>
<td>4</td>
<td>111.5</td>
<td>1.37</td>
<td>88.0</td>
<td>95.4</td>
</tr>
</tbody>
</table>

*Phenotypic variance explained by each QTL; QTL were ordered by PVE.*

doi:10.1371/journal.pone.0017573.t002
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 ([3] 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 2. Power of the 9 identified QTL in the Arabidopsis population.](https://www.plosone.org/doi/10.1371/journal.pone.0017573.g002)
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 [53]. 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.

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
Table 3. Simulated power (%) from JICIM and ICIM for rare QTL segregating in family B73×II14H under four levels of support interval, i.e., 1, 2, 5, and 10 cM.

<table>
<thead>
<tr>
<th>Method</th>
<th>Additive effect (day)</th>
<th>QTL overlapped with marker</th>
<th>QTL located at the center of marker interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 cM</td>
<td>2 cM</td>
</tr>
<tr>
<td>JICIM in the maize NAM population</td>
<td>1.50</td>
<td>77.4</td>
<td>89.1</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>59.9</td>
<td>73.2</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>30.4</td>
<td>40.8</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>21.2</td>
<td>32.5</td>
</tr>
<tr>
<td>ICIM in family B73×II14H</td>
<td>1.50</td>
<td>84.7</td>
<td>92.2</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>74.5</td>
<td>84.7</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>25.3</td>
<td>42.4</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>4.9</td>
<td>6.8</td>
</tr>
</tbody>
</table>

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
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 \( f \)th family, and a total population size of \( N = \sum_{f=1}^{F} n_f \). 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,

\[
Y = b_0 + au + X\beta + \epsilon
\]  

(1)

where \( Y \) is the vector of phenotypic values; \( b_0 \) is the intercept; \( u^T = (u_1, u_2, \ldots, u_F) \) is the effect vector indicating the cross effect of each founder with the common parent; \( \alpha \) is the \( N \times F \) incidence matrix relating each \( u_f (f = 1, 2, \ldots, F) \) to \( Y \); \( \beta \) is the \([F+1] \times m \) effect vector of the \( N \times ([F+1] \times m) \) incidence matrix \( X \) where \( m \) is the number of markers; and \( \epsilon \) is the vector of residuals. To avoid overfitting, 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.

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
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 \(k^{th}\) 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 [37]; thus this table assumes no missing genotypic data.

doi:10.1371/journal.pone.0017573.t004

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample size</th>
<th>Frequency</th>
<th>Marker genotype</th>
<th>Frequency of QTL genotype</th>
<th>Distribution of (Ay_f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(n_f)</td>
<td>(\frac{1}{2}(1-r_{k,k+1}))</td>
<td>+</td>
<td>(p_{\mu_1})</td>
<td>(1-p_{\mu_1})</td>
</tr>
<tr>
<td>2</td>
<td>(n_f)</td>
<td>(\frac{1}{2}r_{k,k+1})</td>
<td>+</td>
<td>(p_{\mu_2})</td>
<td>(1-p_{\mu_2})</td>
</tr>
<tr>
<td>3</td>
<td>(n_f)</td>
<td>(\frac{1}{2}r_{k,k+1})</td>
<td>-</td>
<td>(p_{\mu_3})</td>
<td>(1-p_{\mu_3})</td>
</tr>
<tr>
<td>4</td>
<td>(n_f)</td>
<td>(\frac{1}{2}(1-r_{k,k+1}))</td>
<td>-</td>
<td>(p_{\mu_4})</td>
<td>(1-p_{\mu_4})</td>
</tr>
</tbody>
</table>

Note: \(n_f + n_{f+1} + n_f = n_j\) where \(n_j\) is the size of the \(f^{th}\) family. \(Q_i\) is the allele in the \(f^{th}\) parent, and \(Q_j\) is the allele in the common parent. \(p_1 = (1-r_{j,k+1})(1-r_{j,k+1})/2, p_2 = (1-r_{j,k+1})(1-r_{j,k+1}), p_3 = 1-p_{1}, p_4 = 1-p_{1}\), where \(r_{j,k+1} = \) 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; \(s_j\) and \(s_{j+1}\) are the recombinant frequencies between the two flanking markers in the current scanning interval, and between either side of the flanking marker and the putative QTL located between the interval; \(\mu_j\) and \(\mu_0\) are the means for QTL genotypes \(Q_jQ_j\) and \(Q_0Q_0\), respectively, and \(\Phi(\cdot, \mu_j, \sigma^2)\) represents the distribution 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 \(\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_F,\) \(\sigma_j^2, \sigma_{j+1}^2, \ldots, \) and \(\sigma_0^2\), from which the additive effect of the putative QTL in each family (denoted by \(\bar{a}_f\)) can be estimated as \(\bar{a}_f = \frac{1}{\sqrt{2}}(\bar{\mu}_i - \bar{\mu}_0)\).

In QTL mapping for a bi-parental population, when we calculated \(\hat{\mu}_i\), 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 Kosambi 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 \(f^{th}\) family.

Under the null hypothesis \(H_0\), \(Ay_f\) \((i = 1, \ldots, n_f)\) within each family follow a normal distribution denoted as \(N(\mu_0, \sigma^2)\).

The mean and variance of this distribution can be estimated as,

\[
\bar{\mu}_0 = \frac{1}{n_f} \sum_{i=1}^{n_f} Ay_f \quad \text{and} \quad \bar{\sigma}_0^2 = \frac{1}{n_f} \sum_{i=1}^{n_f} (Ay_f - \bar{\mu}_0)^2.
\]

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

\[
L_0 = \sum_{f=1}^{F} \sum_{i=1}^{n_f} \ln(\hat{\phi}(Ay_f; \bar{\mu}_0, \bar{\sigma}_0^2)).
\]

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 \( \sigma^2 \), where \( \sigma \) 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,

\[
PVE = \frac{\sigma^2}{\bar{V}_P},
\]

where \( \bar{V}_P \) 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 \( \mu_0, \mu_0 + 2\alpha_i, \ldots, \mu_0 + 2\alpha_F \), respectively, and the corresponding frequencies of these \( F+1 \) genotypic values are \( \frac{1}{2N}, \frac{n_1}{2N}, \ldots, \frac{n_F}{2N} \) respectively. Therefore, the genetic variance of the \( F+1 \) QTL genotypes in the NAM population can be calculated as,

\[
V_{QTL} = 2 \sum_{j=1}^{F} \left( \frac{n_j}{N} \sigma_j^2 \right) - \left( \sum_{j=1}^{F} \frac{n_j}{N} \alpha_j \right)^2.
\]

Thus,

\[
PVE = \frac{2 \sum_{j=1}^{F} \left( \frac{n_j}{N} \sigma_j^2 \right) - \left( \sum_{j=1}^{F} \frac{n_j}{N} \alpha_j \right)^2}{\bar{V}_P},
\]

where \( \bar{V}_P = \sum_{f=1}^{F} \frac{n_f}{N} V_{PF} \) and \( V_{PF} \) is the phenotypic variance of the \( f^{th} \) 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., Ler (Landsberg erecta) × Antwerp (An-1), Ler × Kashmir (Kas-2), and Ler × 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 from the original study, and reconstructed the genetic maps and QTL scans. A total of 109 genetic markers showed polymorphism from the original study, and reconstructed the genetic maps and QTL scans. A total of 109 genetic markers showed polymorphism from the original study, and reconstructed the genetic maps and QTL scans. A total of 109 genetic markers showed polymorphism from the original study, and reconstructed the genetic maps and QTL scans.
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 1006 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 \((\text{GWER}(k))\) [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 \( \times \) B97, B73 \( \times \) CML103, B73 \( \times \) CML228, B73 \( \times \) CML247, B73 \( \times \) CML277, B73 \( \times \) L14H, B73 \( \times \) Ki11, B73 \( \times \) MS71, and B73 \( \times \) 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 \( V_e = V_P - a_1^2 - a_2^2 \), 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 \( V_e \) were used to simulate the phenotypic value of the \( f_{ij} \)-th 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 JICIM [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 vtgl allele from northern germplasm. In this case, vtgl was accurately identified, while another gene (vtg2) 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 PB EE JW. Contributed reagents/materials/analysis tools: ESB. Wrote the paper: HL JW.

**References**


