

Full Length Research Paper

Effect of small mapping population sizes on reliability of quantitative trait locus (QTL) mapping

C. Raghavan and B. C. Y. Collard*

Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute (IRRI), DAPO Box 7777, Metro Manila, Philippines.

Accepted 11 May, 2012

A limitation of quantitative trait loci (QTL) mapping is that accuracy of determining QTL position and effects are largely determined by population size. Despite the importance of this concept, known as the "Beavis effect" there has generally been a lack of understanding by molecular geneticists and breeders. One possible explanation for this may be that this concept has been explored by using computer simulations and that these findings may not be clearly understood. In this study, we demonstrated the effect of population size on the accuracy of determining QTL positions and effects in a simple and concise manner. Simulated data was generated for extremely large mapping populations ($n = 1000$) and smaller mapping populations ($n = 94$ or $n = 190$) were obtained by random sampling. Populations were defined to segregate for either three or five QTLs with heritabilities of $h^2 = 0.75$ or $h^2 = 0.50$. When small populations were used, errors were detected in determining QTL positions, and in some cases, QTLs were not detected (that is, false negatives) especially when $h^2 = 0.50$. More importantly, R^2 values were overestimated or underestimated. Composite interval mapping was more reliable for detecting QTLs compared to simple interval mapping. These findings have important implications for QTLs which are selected in breeding programs via marker-assisted selection.

Key words: Quantitative trait loci (QTL) mapping, Beavis effect, population size, confidence intervals, marker-assisted selection.

INTRODUCTION

More than 20 years after the landmark paper by Paterson et al. (1988), quantitative trait loci (QTL) mapping has become a widely-used method in crops for detecting genes and QTLs associated with important traits (Semagn et al., 2010). There are literally thousands of QTL maps for a wide range of traits in diverse crop species. The process of QTL mapping involves developing a mapping population that segregates for at least one trait, constructing a linkage map and performing QTL analysis (usually interval mapping methods) to identify genomic regions associated with the trait of interest (Semagn et al., 2006). Results from QTL analysis indicate the number, location, QTL effects and possible

QTL \times QTL interactions (epistasis) between loci. This information is of great interest to crop molecular geneticists and plant breeders. Furthermore, QTL data can lead to the use of markers to assist in the selection of QTLs in breeding material via marker assisted selection (MAS), although validation steps are generally recommended (Collard and Mackill, 2008; Francia et al., 2005).

One of the main limitations of QTL mapping is the accuracy of determining QTL locations and estimating QTL effects. This is largely determined by the size of the mapping population. Typical mapping population sizes consist of 100 to 200 individuals. William D. Beavis was the first researcher to specifically report on the limitations of QTL analysis caused by small sample sizes, and hence this phenomenon has been referred to as the "Beavis effect" (Beavis, 1994, 1998; Xu, 2003). This topic has received relatively little attention in the scientific literature with exceptions in maize (Melchinger et al., 1998; Utz

*Corresponding author. E-mail: bcycollard@hotmail.com.

et al., 2000; Schon et al., 2004), barley (Vales et al., 2005), and for teaching purposes (Helms and Doetkott, 2007). Despite the important implications of using small mapping populations, many plant breeders and even molecular geneticists may not be aware of them when interpreting QTL mapping results. A likely explanation for this is that, research in this area has been based on computer simulations, or explained with extensive mathematical formulas and calculations which may not be easily understood by researchers who perform wet-lab work or are involved in phenotyping for QTL mapping experiments (Xu, 2003). Furthermore, these findings have been reported primarily in conference proceedings or book chapters which may not be widely-available (Beavis, 1994, 1998).

In this study, we demonstrated the effect of sampling small population sizes on the accuracy of QTL mapping with an emphasis on QTL location and estimation of effects. We have specifically endeavoured to present our finding in simple terms so that findings and implications can be easily understood by researchers and students who are directly involved in QTL mapping.

MATERIALS AND METHODS

Computer simulation of mapping populations

Simulated data was generated using the program GREGOR (Tinker and Mather, 1993). This freely-available program can be used for simulating quantitative traits, breeding populations, linkage maps and QTLs. Four large recombinant inbred (RI) populations consisting of 1000 individuals each (called A1000, B1000, C1000 and D1000) were generated by selfing with complete homozygosity using GREGOR. These four populations are referred to as the "true" populations. RI populations were chosen because they are the most common types of mapping populations in self-pollinated species and are advantageous because they are immortal populations and trait data can be replicated. The four populations were defined to have different numbers of unlinked QTLs and broad-sense heritability (h^2) values (Figure 1). A maximum number of 1000 individuals were chosen because this was the maximum number that could be stored in computer memory while randomly sub-sampling smaller populations (described later). Furthermore, research suggests that QTL detection is close to the actual magnitude when $n = 1000$ is used (Beavis, 1998; Xu, 2003). Thus, QTLs detected in these populations were considered to the "true" values in this study.

Simulated DNA marker data and QTL data

Seven chromosomes were defined. A total of 115 simulated markers were evenly distributed on all chromosomes (15 per linkage group), and the average marker spacing was approximately 10 cM based on recommendations by Darvasi et al. (1993). Markers were designated with an "L" prefix denoting "locus". GREGOR was used to generate quantitative trait data for each of the 1000 individuals with defined heritabilities and trait means that were normally distributed. QTLs were randomly defined; however only one QTL per chromosome was allowed so as not to confound QTL analysis. All QTLs had equal additive effects and there were no dominance effects.

Population sub-sampling

Population sizes of 94 and 190 individuals were randomly and independently sampled from each of the four true populations (A1000, B1000, C1000 and D1000) consisting of 1000 individuals. While sampling from the true populations for $n = 94$ and $n = 190$, individual mapping population lines were not replaced. However, for the next sampling populations, lines were replaced and the 2nd set of 94 or 190 lines were selected; this procedure was repeated 10 times. These population sizes were defined because they conveniently fit within one or two 96-well plates routinely used in marker genotyping laboratories (including parents). A total of 80 mapping populations (data sets) were generated, representing 40 mapping populations of $n = 94$ and 40 mapping populations of $n = 190$, as 10 datasets were sampled for each population size. Data was saved as a MapMaker file (Lander et al., 1987) within GREGOR and subsequently imported into MapManager QTX (Manly et al., 2001). Populations were named with: (1) a single letter prefix according to the true population from which they were sampled; (2) three digit number indicating population size (094 or 190); (3) two digit number indicating sample number (01 to 10).

QTL analysis

Map construction and QTL analysis was performed using MapManager QTX (Manly et al., 2001). The Kosambi mapping function was used for linkage analysis. Initial genome scans for QTLs were performed using single marker analysis (marker regression command). Interval mapping (IM) and composite interval mapping (CIM) were also performed. For CIM, background markers were selected by incorporating significant markers in the analysis. The results for LOD values, position, R^2 for each QTL was recorded. Quick tests were used to determine significance levels of QTLs (Piepho, 2001). QTLs were classified as: (1) suggestive; (2) significant; and (3) highly significant according to Lander and Kruglyak (1995). Figures were produced using MapChart (Voorrips, 2002).

RESULTS AND DISCUSSION

There have been relatively few QTL mapping studies investigating the effect of population size on the accuracy of QTL position and effects. In this study, QTLs were defined in the four simulated mapping populations consisting of 1000 individuals, and QTL analysis was performed to identify the QTL positions and effects of these "true" QTLs. Random samples were then taken from these large populations to reflect the sampling which occurs when "real" mapping populations are developed. In practice, mapping populations are usually randomly selected from a larger population size or the exact number of lines is chosen. Moreover, although QTLs in the randomly sampled small mapping populations were generally located to the correct chromosome, there was considerable variation for QTL position, LOD value and R^2 between mapping populations as determined by comparison with the "true" QTL results. There were many differences between IM and CIM results. The full QTL mapping results for all populations are presented in Tables 1 to 4. False positives were also detected; these are presented in Table 5.

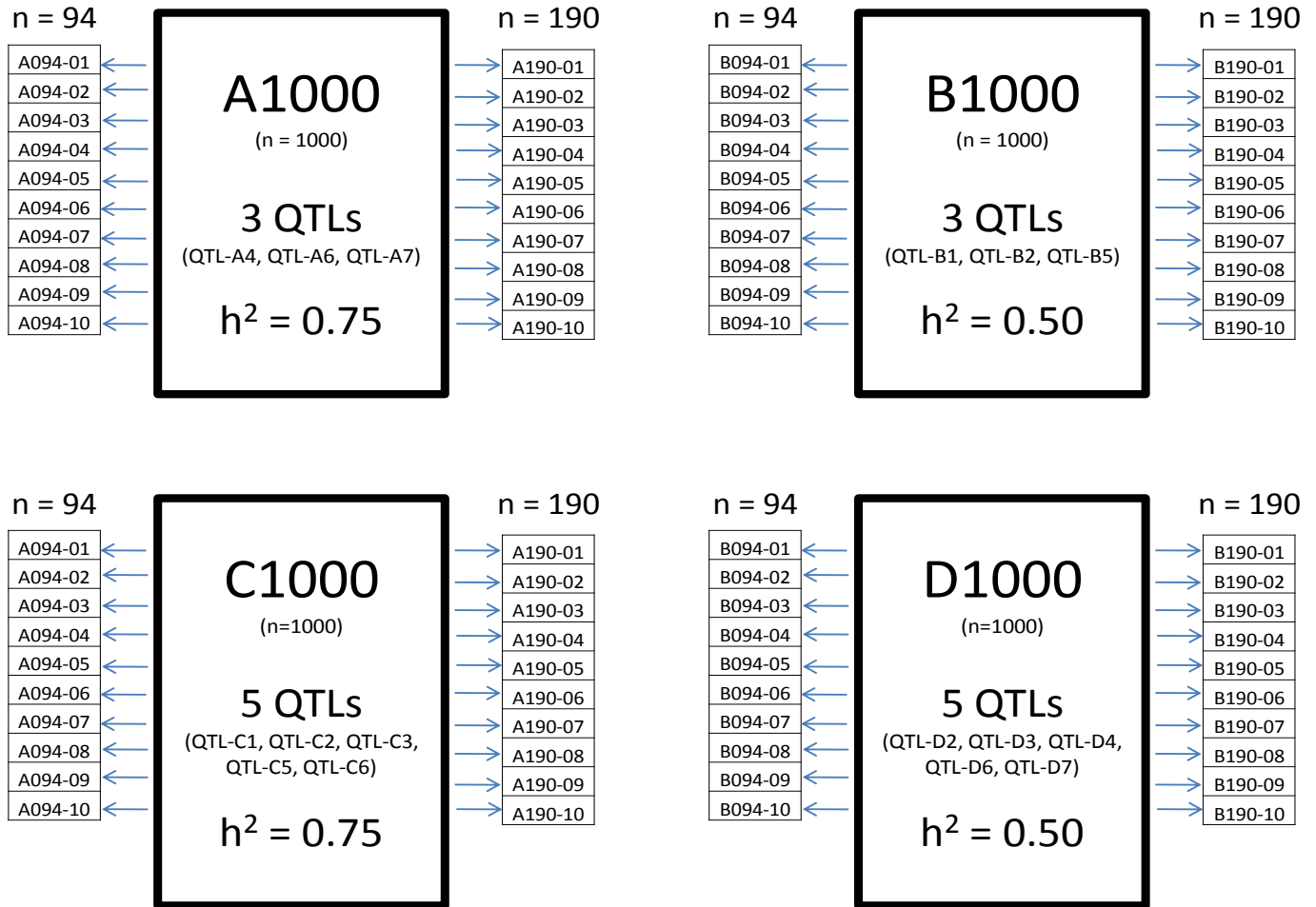


Figure 1. Overview of QTL mapping experimental design. Four true populations were simulated consisting of 1000 individuals (A1000, B1000, C1000 and D1000) with different numbers of segregating QTLs and trait heritabilities. Sampling was performed from these four true populations to derive smaller mapping populations ($n = 94$ or $n = 190$) for subsequent QTL analysis, in order to determine effect of small populations sizes on QTL mapping accuracy. Each dataset was analysed using interval mapping and composite interval mapping.

Table 1. QTL analysis results for true population “A” and sub-sampled A populations ($n = 94$ and $n = 190$) with three QTLs.

| Population | Method ^a | QTL-A4 | | | QTL-A6 | | | QTL-A7 | | | Total R ^{2e} |
|------------|---------------------|---------------------------|------------------|------------------------|--------------|-------|-----------------------|--------------|------|-----------------------|------------------------|
| | | Pos. ^b (cM) | LOD ^c | R ^{2d} (%) | Pos. (cM) | LOD | R ² (%) | Pos. (cM) | LOD | R ² (%) | |
| A1000 | IM | 22 | 60.8 | 24 | 36 | 58.6 | 24 | 53 | 70.3 | 28 | 76 |
| A1000 | CIM | 22 | 129.3 | 24 | 37 | 137.3 | 24 | 52 | 131 | 25 | 73 |
| A094-01 | IM | 22 | 7.8 | 32 | 44 | 6.1 | 26 | 45 | 11.3 | 43 | 101 |
| A094-02 | IM | 27 | 7 | 29 | 47 | 3.8 | 17 | 58 | 8.9 | 36 | 82 |
| A094-03 | IM | 30 | 3.2 | 15 | 35 | 9.3 | 37 | 54 | 5 | 22 | 74 |
| A094-04 | IM | 47 | 5.1 | 22 | 35 | 7.1 | 29 | 53 | 8.3 | 33 | 84 |
| A094-05 | IM | 23 | 3.7 | 16 | 26 | 5.7 | 25 | 53 | 3.8 | 17 | 58 |
| A094-06 | IM | 25 | 8.5 | 34 | 32 | 7.4 | 30 | 55 | 9.5 | 37 | 101 (+13) ^f |
| A094-07 | IM | 17 | 4.7 | 21 | 32 | 9.4 | 37 | 52 | 7.1 | 29 | 87 |
| A094-08 | IM | 24 | 7.3 | 30 | 31 | 5.2 | 23 | 71 | 7 | 29 | 82 |
| A094-09 | IM | 29 | 5.6 | 24 | 35 | 4.3 | 19 | 56 | 8.6 | 34 | 77 |
| A094-10 | IM | 29 | 6.4 | 27 | 33 | 5.9 | 25 | 58 | 6.2 | 26 | 78 |
| A094-01 | CIM | 24 | 8.7 | 16 | 37 | 10.1 | 18 | 55 | 18.1 | 29 | 63 |

Table 1. Contd.

| | | | | | | | | | | | |
|---------|-----|----|------|----|----|------|----|----|------|----|---------|
| A094-02 | CIM | 25 | 13.7 | 25 | 45 | 15.5 | 23 | 56 | 15.8 | 30 | 78 |
| A094-03 | CIM | 21 | 12.5 | 22 | 36 | 14.2 | 31 | 54 | 9.8 | 19 | 72 |
| A094-04 | CIM | 23 | 17.2 | 28 | 37 | 15.4 | 23 | 51 | 17.6 | 30 | 81 |
| A094-05 | CIM | 23 | 11.7 | 25 | 28 | 15.7 | 34 | 56 | 11.1 | 24 | 83 |
| A094-06 | CIM | 23 | 13.1 | 21 | 32 | 4.4 | 7 | 54 | 11.9 | 21 | 49 (+7) |
| A094-07 | CIM | 19 | 18.5 | 31 | 32 | 14.1 | 23 | 53 | 11.8 | 21 | 75 |
| A094-08 | CIM | 24 | 13.6 | 26 | 32 | 12.7 | 24 | 71 | 13.5 | 24 | 74 |
| A094-09 | CIM | 28 | 9.9 | 18 | 36 | 14.5 | 25 | 55 | 15.3 | 31 | 74 |
| A094-10 | CIM | 25 | 13.8 | 29 | 38 | 12.8 | 25 | 57 | 7 | 12 | 66 |
| A190-01 | IM | 24 | 14.9 | 30 | 39 | 12.1 | 25 | 51 | 17.4 | 34 | 89 |
| A190-02 | IM | 16 | 9.6 | 21 | 35 | 11.5 | 24 | 46 | 8.4 | 18 | 63 |
| A190-03 | IM | 21 | 14.7 | 30 | 41 | 11.1 | 24 | 51 | 9 | 20 | 74 |
| A190-04 | IM | 24 | 10.9 | 23 | 38 | 10.4 | 22 | 50 | 13 | 27 | 72 |
| A190-05 | IM | 24 | 11 | 23 | 39 | 12.7 | 27 | 53 | 13.8 | 28 | 78 |
| A190-06 | IM | 22 | 8.5 | 19 | 34 | 16.1 | 32 | 44 | 14.4 | 29 | 80 (+8) |
| A190-07 | IM | 18 | 11.1 | 24 | 38 | 16.7 | 33 | 61 | 10 | 22 | 79 |
| A190-08 | IM | 15 | 11.1 | 24 | 44 | 12.4 | 26 | 53 | 14.3 | 29 | 79 |
| A190-09 | IM | 21 | 21 | 40 | 35 | 10.5 | 23 | 60 | 8.4 | 18 | 81 |
| A190-10 | IM | 27 | 8.8 | 19 | 39 | 15.5 | 31 | 53 | 12.1 | 26 | 76 |
| A190-01 | CIM | 22 | 25.4 | 24 | 41 | 23.1 | 19 | 52 | 20.1 | 18 | 61 |
| A190-02 | CIM | 17 | 22.5 | 25 | 35 | 20.8 | 23 | 49 | 20.4 | 24 | 72 |
| A190-03 | CIM | 17 | 25.1 | 26 | 42 | 24.8 | 23 | 52 | 24 | 25 | 74 |
| A190-04 | CIM | 24 | 23.8 | 24 | 38 | 28.2 | 28 | 49 | 21.2 | 21 | 73 |
| A190-05 | CIM | 21 | 25 | 24 | 38 | 33.1 | 28 | 56 | 25.9 | 25 | 77 |
| A190-06 | CIM | 21 | 25.1 | 23 | 34 | 33.7 | 29 | 48 | 28.1 | 27 | 79 |
| A190-07 | CIM | 19 | 22.2 | 25 | 40 | 17.5 | 20 | 62 | 12.6 | 16 | 61 |
| A190-08 | CIM | 16 | 28.4 | 26 | 44 | 29.8 | 25 | 53 | 27.4 | 27 | 78 |
| A190-09 | CIM | 20 | 33.9 | 30 | 36 | 29.9 | 21 | 60 | 29.7 | 25 | 76 |
| A190-10 | CIM | 26 | 20.5 | 20 | 41 | 24.3 | 25 | 54 | 20.5 | 21 | 66 |

Total R^2 values are indicated for the total sum of effects from all QTLs and from false positives (shown in parentheses). Values in italics indicate that the defined trait heritability has been exceeded based on QTL analysis results. ^aInterval mapping method; IM, interval mapping; CIM, composite interval mapping. ^bPos., Position in centiMorgans (cM). ^cLOD, Logarithm of odds score. ^d R^2 , QTL effect explaining a proportion of the total phenotypic variance. ^eTotal R^2 is the sum of all three QTLs (A4 + A6 + A7). Total R^2 values should not theoretically exceed the total value as determined in A1000 using CIM (that is, 73%). ^fValue in parenthesis is attributable to false positive QTL effects and increase in the total phenotypic variance.

An important objective of this study was to determine the accuracy of QTL position in the small mapping populations by comparing the detected QTL's location with the true location. There were often large inaccuracies in QTL position as indicated in Tables 1 to 4. There were many examples of the QTL positions being detected with errors of >20 cM (in other words, QTLs were detected in the smaller mapping populations that were >20 cM away from the true position for both population sizes (e.g. A094-04, B190-10, C094-10, C190-07 and D094-02). The most extreme case of inaccuracy in QTL position was detected in mapping population C094-01 using IM (46 cM away from true position). However, for the same QTL the accuracy was higher when CIM was used (16 cM away from true position). Examples of how QTL position varied between the different mapping populations are shown in Figure 2. Most importantly, flanking markers containing the QTL

were often incorrectly identified. Consistent with QTL mapping theory, the accuracy was higher in mapping populations consisting of 190 individuals compared to 94. Confidence intervals are often reported for QTL position (Hackett, 2002; Lander and Botstein, 1989). The variation in QTL position in this study strongly supports the use of confidence intervals for reporting QTL mapping results.

False negative QTL results (failing to detect a QTL at a defined position) were obtained in datasets when the trait heritability was 0.50 (for C or D populations; Figure 2). Although less false negatives were detected when CIM was used, this finding suggests that many minor QTLs with small effects may be undetected in real mapping populations, especially in small population sizes. A total of 32 false positive across all samples (n = 94 and n = 190) QTLs were detected, although the majority were only detected as "suggestive" and using IM (the suggestive significance level was defined by Lander and

Table 2. QTL analysis results for true population “B” and sub-sampled B populations (n = 94 and n = 190) with three QTLs.

| Population | Method ^a | QTL-B1 | | | QTL-B2 | | | QTL-B5 | | | Total R ² ^e |
|------------|---------------------|---------------------------|------------------|------------------------------------|--------------|------|-----------------------|--------------|------|-----------------------|-----------------------------------|
| | | Pos. ^b (cM) | LOD ^c | R ² ^d (%) | Pos. (cM) | LOD | R ² (%) | Pos. (cM) | LOD | R ² (%) | |
| B1000 | IM | 20 | 37.9 | 16 | 50 | 34.3 | 15 | 35 | 42.4 | 18 | 49 |
| B1000 | CIM | 20 | 68 | 19 | 50 | 59.5 | 16 | 35 | 70.3 | 19 | 54 |
| B094-01 | IM | 23 | 6.1 | 26 | 53 | 6.2 | 26 | 31 | 3.1 | 14 | 66 |
| B094-02 | IM | 21 | 2.3 | 11 | 54 | 2.8 | 13 | 34 | 2.9 | 13 | 37 |
| B094-03 | IM | 11 | 4 | 18 | 39 | 4.1 | 18 | 38 | 4.2 | 19 | 55 |
| B094-04 | IM | 14 | 4 | 18 | 54 | 8.7 | 35 | 36 | 5.5 | 24 | 77 (+13) ^f |
| B094-05 | IM | 24 | 5.4 | 23 | 53 | 2.3 | 11 | 43 | 6.5 | 27 | 61 |
| B094-06 | IM | 19 | 1.8 | 8 | 58 | 3.3 | 15 | 40 | 3.9 | 17 | 40 |
| B094-07 | IM | 19 | 3 | 14 | 43 | 3.4 | 15 | 29 | 4.6 | 20 | 49 |
| B094-08 | IM | 20 | 3.1 | 14 | 46 | 3.4 | 15 | 29 | 8.7 | 35 | 64 |
| B094-09 | IM | 18 | 1.5 | 7 | 47 | 4.7 | 21 | 24 | 4.2 | 18 | 46 |
| B094-10 | IM | 21 | 7.1 | 29 | 51 | 2.6 | 12 | 59 | 3.5 | 16 | 57 |
| B094-01 | CIM | 24 | 7.3 | 23 | 54 | 6.9 | 22 | 31 | 2.7 | 7 | 52 |
| B094-02 | CIM | 20 | 5.3 | 17 | 50 | 6.3 | 20 | 36 | 5 | 16 | 53 |
| B094-03 | CIM | 19 | 6.2 | 17 | 36 | 7.1 | 20 | 37 | 8.4 | 22 | 59 |
| B094-04 | CIM | 15 | 4.8 | 11 | 51 | 13.8 | 31 | 33 | 8.1 | 19 | 61 |
| B094-05 | CIM | 21 | 5 | 15 | 44 | 4.9 | 13 | 42 | 5 | 15 | 43 |
| B094-06 | CIM | 21 | 2.8 | 9 | 58 | 5.6 | 18 | 36 | 6.6 | 22 | 49 |
| B094-07 | CIM | 18 | 8.4 | 20 | 43 | 6.5 | 16 | 34 | 8.8 | 23 | 59 (+7) |
| B094-08 | CIM | 19 | 6.9 | 17 | 37 | 3.4 | 8 | 30 | 9.3 | 26 | 51 |
| B094-09 | CIM | 16 | 6.4 | 19 | 47 | 8.6 | 26 | 31 | 5.9 | 16 | 61 |
| B094-10 | CIM | 20 | 8.7 | 23 | 52 | 6.3 | 16 | 58 | 6.4 | 17 | 56 |
| B190-01 | IM | 12 | 6.9 | 15 | 50 | 5.2 | 12 | 36 | 4.6 | 11 | 38 |
| B190-02 | IM | 22 | 10.2 | 22 | 51 | 10.4 | 22 | 39 | 9.2 | 20 | 64 |
| B190-03 | IM | 21 | 7.7 | 17 | 49 | 8.3 | 18 | 39 | 9.6 | 21 | 56 |
| B190-04 | IM | 25 | 9.4 | 20 | 48 | 5.6 | 13 | 32 | 5.9 | 13 | 46 |
| B190-05 | IM | 24 | 6.3 | 14 | 50 | 7.3 | 16 | 35 | 10 | 21 | 51 |
| B190-06 | IM | 19 | 1.8 | 8 | 58 | 3.3 | 15 | 40 | 3.9 | 17 | 40 |
| B190-07 | IM | 19 | 3 | 14 | 43 | 3.4 | 15 | 29 | 4.6 | 20 | 49 |
| B190-08 | IM | 20 | 3.1 | 14 | 41 | 3.2 | 15 | 29 | 8.7 | 35 | 64 |
| B190-09 | IM | 18 | 1.5 | 7 | 47 | 4.7 | 21 | 29 | 4.3 | 19 | 47 |
| B190-10 | IM | 21 | 7.1 | 29 | 51 | 2.6 | 12 | 59 | 3.5 | 16 | 57 |
| B190-01 | CIM | 18 | 14 | 22 | 49 | 10.2 | 16 | 36 | 8.3 | 13 | 51 |
| B190-02 | CIM | 23 | 16.9 | 20 | 52 | 17.5 | 21 | 39 | 15.7 | 19 | 60 |
| B190-03 | CIM | 21 | 13.6 | 17 | 49 | 15.6 | 20 | 39 | 15.6 | 21 | 58 |
| B190-04 | CIM | 24 | 15 | 23 | 48 | 10.7 | 16 | 32 | 9.3 | 14 | 53 |
| B190-05 | CIM | 22 | 13.4 | 18 | 49 | 12.8 | 18 | 34 | 15.8 | 22 | 58 |
| B190-06 | CIM | 21 | 2.8 | 9 | 58 | 5.6 | 18 | 39 | 6.6 | 22 | 49 |
| B190-07 | CIM | 18 | 8.4 | 20 | 43 | 6.5 | 16 | 34 | 8.8 | 23 | 59 (+7) |
| B190-08 | CIM | 19 | 6.9 | 17 | 37 | 3.4 | 8 | 30 | 9.3 | 26 | 51 |
| B190-09 | CIM | 16 | 6.4 | 19 | 47 | 8.6 | 26 | 31 | 5.9 | 16 | 61 |
| B190-10 | CIM | 13 | 9.4 | 25 | 52 | 6.3 | 16 | 58 | 6.4 | 17 | 58 |

Total R² values are indicated for the total sum of effects from all QTLs and from false positives (shown in parentheses). Values in italics indicate that the defined trait heritability has been exceeded based on QTL analysis results. Shaded cells indicate QTLs detected using the “suggestive” criterion according to Lander and Kruglyak (1995). ^aInterval mapping method; IM, interval mapping; CIM, composite interval mapping; ^bPos., position in centiMorgans (cM); ^cLOD, logarithm of odds score. ^dR², QTL effect explaining a proportion of the total phenotypic variance. ^eTotal R² is the sum of all three QTLs (B1 + B2 + B5). Total R² values should not theoretically exceed the total value as determined in B1000 using CIM (that is, 54%). ^fValue in parenthesis was attributable to false positive QTL effects and increases the total phenotypic variance.

Table 3. QTL analysis results for true population "C" and sub-sampled C populations (n = 94 and n = 190) with five QTLs.

| Population | Method ^a | QTL-C1 | | | QTL-C2 | | | QTL-C3 | | | QTL-C5 | | | QTL-C6 | | | Total R ^{2e} |
|------------|---------------------|---------------------------|------------------|------------------------|--------------|-------|-----------------------|--------------|-------|-----------------------|--------------|-------|-----------------------|---------------------------|-------|-----------------------|-----------------------|
| | | Pos. ^b (cM) | LOD ^c | R ^{2d} (%) | Pos. (cM) | LOD | R ² (%) | Pos. (cM) | LOD | R ² (%) | Pos. (cM) | LOD | R ² (%) | Pos. (cM) | LOD | R ² (%) | |
| C1000 | IM | 105 | 37.4 | 16 | 39 | 36.7 | 16 | 55 | 33.2 | 14 | 31 | 48 | 20 | 101 | 38.9 | 16 | 82 |
| C1000 | CIM | 104 | 110 | 14 | 39 | 112.1 | 15 | 56 | 116.3 | 15 | 31 | 123.1 | 17 | 101 | 111.1 | 14 | 75 |
| C094-01 | IM | 100 | 2.2 | 10 | 85 | 2.5 | 11 | 66 | 2.1 | 10 | 41 | 5 | 22 | 102 | 4.4 | 19 | 72 |
| C094-02 | IM | 103 | 3.9 | 18 | 42 | 4.4 | 19 | 59 | 4.5 | 20 | 38 | 5.6 | 24 | 121 | 4.9 | 21 | 102 |
| C094-03 | IM | 102 | 4.4 | 19 | 45 | 4.8 | 21 | 48 | 1.9 | 9 | 36 | 4.8 | 21 | 107 | 7.7 | 31 | 101 |
| C094-04 | IM | 104 | 4 | 18 | 43 | 3.6 | 16 | 56 | 3.7 | 16 | 35 | 3.5 | 16 | 113 | 5.1 | 22 | 88 |
| C094-05 | IM | 103 | 6.7 | 28 | 41 | 2.9 | 13 | 57 | 3.4 | 16 | 24 | 2.6 | 12 | 117 | 3.1 | 14 | 83 |
| C094-06 | IM | 96 | 2.8 | 13 | 27 | 2.1 | 10 | 48 | 5.1 | 22 | 33 | 6.6 | 28 | 100 | 3.7 | 16 | 89 |
| C094-07 | IM | 100 | 4.6 | 20 | 32 | 4.4 | 19 | 62 | 4.6 | 20 | 32 | 3.2 | 15 | 95 | 6.6 | 28 | 102 |
| C094-08 | IM | 98 | 5.6 | 24 | 40 | 4.3 | 19 | 59 | 3.4 | 15 | 37 | 3 | 14 | Not detected ^f | | | 72 |
| C094-09 | IM | 103 | 4 | 18 | 35 | 5 | 22 | 54 | 3 | 14 | 31 | 3.3 | 15 | Not detected | | | 69 |
| C094-10 | IM | 131 | 4.5 | 20 | 31 | 2.8 | 13 | 36 | 5.8 | 25 | 28 | 3.6 | 16 | 113 | 3.7 | 17 | 91 |
| C094-01 | CIM | 96 | 6.9 | 12 | 55 | 8.7 | 13 | 65 | 6.6 | 11 | 41 | 11.6 | 23 | 102 | 8 | 14 | 73 |
| C094-02 | CIM | 101 | 8.7 | 12 | 43 | 9.1 | 13 | 65 | 6.9 | 9 | 37 | 10.1 | 15 | 118 | 8.1 | 12 | 61 |
| C094-03 | CIM | 102 | 11.1 | 14 | 44 | 8.9 | 11 | 48 | 12.9 | 17 | 31 | 9.7 | 12 | 107 | 12.9 | 16 | 70 |
| C094-04 | CIM | 104 | 7.4 | 15 | 43 | 8.6 | 14 | 56 | 9.4 | 21 | 35 | 4.6 | 9 | 107 | 7.1 | 13 | 72 |
| C094-05 | CIM | 103 | 11.2 | 18 | 44 | 8.8 | 13 | 58 | 8.1 | 12 | 27 | 8.9 | 13 | 119 | 8.7 | 13 | 69 |
| C094-06 | CIM | 98 | 4.2 | 8 | 34 | 5.8 | 10 | 49 | 8.2 | 17 | 33 | 6.9 | 14 | 106 | 13.6 | 18 | 67 |
| C094-07 | CIM | 98 | 14.8 | 16 | 32 | 15.1 | 16 | 62 | 16.1 | 18 | 33 | 12.8 | 13 | 89 | 11.3 | 10 | 73 |
| C094-08 | CIM | 98 | 10.7 | 22 | 42 | 8.1 | 16 | 52 | 7 | 13 | 37 | 7.7 | 15 | Not detected | | | 66 |
| C094-09 | CIM | 104 | 13.2 | 23 | 35 | 13.5 | 25 | 58 | 8.9 | 10 | 31 | 7.2 | 11 | 109 | 5.5 | 8 | 77 |
| C094-10 | CIM | 128 | 5.7 | 11 | 38 | 9.3 | 15 | 57 | 10.3 | 16 | 22 | 5.7 | 10 | 114 | 6.8 | 13 | 65 |
| C190-01 | IM | 107 | 5.8 | 13 | 44 | 9.1 | 20 | 64 | 6.5 | 15 | 29 | 8.6 | 19 | 104 | 6.6 | 15 | 82 |
| C190-02 | IM | 93 | 14.4 | 29 | 40 | 4.7 | 11 | 53 | 6.6 | 15 | 29 | 6.9 | 16 | 111 | 6.8 | 15 | 86 |
| C190-03 | IM | 97 | 7 | 16 | 46 | 9.4 | 20 | 54 | 3.9 | 9 | 33 | 11 | 23 | 97 | 7.6 | 17 | 85 |
| C190-04 | IM | 96 | 6.6 | 15 | 40 | 6.4 | 14 | 57 | 9.4 | 20 | 33 | 7.3 | 16 | 101 | 8.3 | 18 | 83 |
| C190-05 | IM | 105 | 6.5 | 15 | 42 | 6.1 | 14 | 53 | 4.5 | 10 | 36 | 7.7 | 17 | 111 | 8.8 | 19 | 75 |
| C190-06 | IM | 102 | 5 | 11 | 40 | 5.4 | 12 | 54 | 4.6 | 11 | 28 | 7.9 | 17 | 103 | 8.8 | 19 | 70 (+7) |
| C190-07 | IM | 125 | 10.8 | 23 | 37 | 5.7 | 13 | 38 | 3.7 | 9 | 30 | 10 | 22 | 107 | 7.8 | 17 | 84 |
| C190-08 | IM | 102 | 6.4 | 14 | 41 | 7.3 | 16 | 63 | 6.1 | 14 | 31 | 9.9 | 21 | 102 | 9.7 | 21 | 86 |
| C190-09 | IM | 100 | 4.7 | 11 | 38 | 6.1 | 14 | 50 | 6.1 | 14 | 34 | 9.5 | 21 | 96 | 9 | 20 | 80 |
| C190-10 | IM | 104 | 14.4 | 29 | 33 | 6.1 | 14 | 58 | 6.9 | 16 | 29 | 10 | 22 | 94 | 5.2 | 12 | 93 |
| C190-01 | CIM | 107 | 22.2 | 15 | 41 | 21.2 | 14 | 64 | 18.5 | 12 | 30 | 23.7 | 16 | 103 | 26.8 | 19 | 76 |
| C190-02 | CIM | 97 | 25.1 | 20 | 39 | 19 | 14 | 57 | 18.3 | 13 | 29 | 19.8 | 15 | 114 | 21.5 | 14 | 76 |

Table 3. Contd.

| | | | | | | | | | | | | | | | | | |
|---------|-----|-----|------|----|----|------|----|----|------|----|----|------|----|-----|------|----|----|
| C190-03 | CIM | 96 | 16.3 | 11 | 47 | 21.9 | 16 | 56 | 20.9 | 15 | 32 | 24.9 | 19 | 99 | 16.3 | 11 | 72 |
| C190-04 | CIM | 97 | 18.6 | 14 | 41 | 22.1 | 16 | 59 | 18.1 | 13 | 32 | 19.9 | 15 | 101 | 17.5 | 13 | 71 |
| C190-05 | CIM | 105 | 15.1 | 12 | 41 | 18 | 18 | 54 | 11.8 | 11 | 34 | 18.8 | 19 | 112 | 16.3 | 14 | 74 |
| C190-06 | CIM | 100 | 23.2 | 20 | 38 | 16.9 | 13 | 55 | 25.6 | 19 | 28 | 17.3 | 14 | 105 | 20.4 | 17 | 83 |
| C190-07 | CIM | 126 | 15.3 | 13 | 38 | 16.7 | 14 | 55 | 17.4 | 12 | 30 | 16.8 | 14 | 106 | 18.5 | 16 | 69 |
| C190-08 | CIM | 101 | 18.5 | 13 | 39 | 26.6 | 17 | 63 | 20.1 | 14 | 29 | 24.4 | 18 | 102 | 18.4 | 13 | 75 |
| C190-09 | CIM | 100 | 20.4 | 15 | 39 | 19.5 | 13 | 50 | 21.3 | 15 | 35 | 22.7 | 17 | 95 | 23.4 | 18 | 78 |
| C190-10 | CIM | 104 | 24.2 | 17 | 33 | 17.8 | 11 | 61 | 19.8 | 12 | 30 | 21.9 | 15 | 91 | 20.6 | 13 | 68 |

Total R^2 values are indicated for the total sum of effects from all QTLs and from false positives (shown in parentheses). Values in italics indicate that the defined trait heritability has been exceeded based on QTL analysis results. Shaded cells indicate QTLs detected using the “suggestive” criterion according to Lander and Kruglyak (1995). ^aInterval mapping method. IM, interval mapping; CIM, composite interval mapping. ^bPos., Position in centiMorgans (cM). ^cLOD, logarithm of odds score. ^d R^2 , QTL effect explaining a proportion of the total phenotypic variance. ^eTotal R^2 is the sum of all five QTLs (C1 + C2 + C3 + C5 + C6). Total R^2 values should not theoretically exceed the defined total value as determined in C1000 using CIM (that is, 75%). N.D. = QTLs not detected at significant or suggestive levels, indicating false negatives because the QTL was defined in this position in the true population.

Table 4. QTL analysis results for true population “D” and sub-sampled D populations (n = 94 and n = 190) with five QTLs.

| Population | Method ^a | QTL-D2 | | | QTL D3 | | | QTL-D4 | | | QTL-D6 | | | QTL-D7 | | | Total R^2 ^e |
|------------|---------------------|---------------------------|------------------|---------------------------|---------------------------|------|---------------------------|---------------------------|------|---------------------------|---------------------------|------|---------------------------|---------------------------|------|---------------------------|--------------------------|
| | | Pos. ^b (cM) | LOD ^c | R^2 ^d (%) | Pos. ^b (cM) | LOD | R^2 ^d (%) | Pos. ^b (cM) | LOD | R^2 ^d (%) | Pos. ^b (cM) | LO D | R^2 ^d (%) | Pos. ^b (cM) | LOD | R^2 ^d (%) | |
| D1000 | IM | 81 | 30.1 | 13 | 104 | 26.1 | 11 | 52 | 22.6 | 10 | 90 | 20.2 | 9 | 87 | 30.3 | 13 | 56 |
| D1000 | CIM | 84 | 55.6 | 13 | 103 | 50 | 12 | 51 | 41.8 | 10 | 91 | 46.4 | 11 | 87 | 55.6 | 13 | 59 |
| D094-01 | IM | 73 | 5.4 | 23 | 104 | 2.6 | 12 | 50 | 1.8 | 8 | 81 | 4.4 | 19 | 85 | 2.1 | 10 | 72 |
| D094-02 | IM | 52 | 4.5 | 20 | 96 | 4.4 | 19 | Not detected ^f | | | 114 | 1.9 | 9 | 87 | 5.8 | 25 | 73 |
| D094-03 | IM | 83 | 3.4 | 15 | Not detected | | | 44 | 5.8 | 25 | 88 | 2.5 | 12 | 103 | 2.2 | 10 | 62 |
| D094-04 | IM | Not detected | | | Not detected | | | 58 | 3.4 | 15 | 109 | 2.8 | 13 | 78 | 3.4 | 15 | 43 |
| D094-05 | IM | 80 | 5.2 | 23 | 98 | 3.2 | 14 | 78 | 4.2 | 19 | Not detected | | | 72 | 2.6 | 12 | 68 |
| D094-06 | IM | Not detected | | | 88 | 5 | 22 | 48 | 2.9 | 13 | 85 | 3.2 | 15 | 84 | 2.6 | 12 | 62 |
| D094-07 | IM | 91 | 3 | 13 | 105 | 5.1 | 22 | 51 | 3.3 | 15 | Not detected | | | 95 | 1.8 | 8 | 58 |
| D094-08 | IM | 92 | 5.1 | 22 | 93 | 2.8 | 13 | 52 | 1.9 | 9 | 82 | 2.3 | 11 | 95 | 2 | 9 | 64 |
| D094-09 | IM | 100 | 2.8 | 13 | Not detected | | | 34 | 3.4 | 15 | 81 | 3.1 | 14 | 101 | 4 | 18 | 60 |
| D094-10 | IM | 76 | 2.5 | 12 | 98 | 3.5 | 16 | Not detected | | | Not detected | | | 101 | 2.1 | 10 | 38 |
| D094-01 | CIM | 73 | 6.2 | 16 | 104 | 4.8 | 12 | 49 | 3.2 | 8 | 82 | 5.2 | 13 | 85 | 2.2 | 5 | 54 |
| D094-02 | CIM | 74 | 6 | 13 | 95 | 5.4 | 13 | Not detected | | | 114 | 3.3 | 8 | 87 | 6.7 | 17 | 51 |
| D094-03 | CIM | 84 | 3.7 | 8 | 89 | 3.2 | 7 | 49 | 6.9 | 17 | 88 | 5.4 | 13 | 108 | 2.8 | 6 | 51 |
| D094-04 | CIM | Not detected | | | Not detected | | | 59 | 4.9 | 16 | 104 | 3 | 10 | 79 | 3.4 | 11 | 37 |
| D094-05 | CIM | 87 | 7.5 | 19 | 99 | 3.9 | 9 | 76 | 4.7 | 11 | Not detected | | | 73 | 5.7 | 14 | 53 |

Table 4. Contd.

| | | | | | | | | | | | | | | | | | |
|---------|-----|--------------|------|----|--------------|------|----|--------------|------|----|--------------|------|----|-----|------|----|----|
| D094-06 | CIM | Not detected | | | 88 | 5.1 | 17 | 47 | 4.1 | 12 | 85 | 3.9 | 12 | 82 | 2.5 | 7 | 48 |
| D094-07 | CIM | 88 | 3.3 | 10 | 105 | 4.2 | 14 | 50 | 3.2 | 10 | Not detected | | | 89 | 2.4 | 7 | 41 |
| D094-08 | CIM | 93 | 5.8 | 13 | 94 | 6.4 | 15 | 51 | 3.4 | 7 | 86 | 5.2 | 12 | 91 | 5.1 | 11 | 58 |
| D094-09 | CIM | 97 | 4.1 | 11 | Not detected | | | 95 | 2.9 | 7 | 71 | 3.4 | 9 | 112 | 4.4 | 13 | 40 |
| D094-10 | CIM | 76 | 2.9 | 11 | 98 | 4.1 | 16 | Not detected | | | Not detected | | | 83 | 2.1 | 7 | 34 |
| D190-01 | IM | 86 | 3.8 | 9 | 117 | 3.9 | 9 | 45 | 8.1 | 18 | 90 | 2.3 | 6 | 77 | 5.7 | 13 | 55 |
| D190-02 | IM | 88 | 11.7 | 25 | 100 | 6.6 | 15 | 47 | 4.6 | 11 | 82 | 2.7 | 6 | 78 | 4.8 | 11 | 68 |
| D190-03 | IM | 84 | 5.3 | 12 | 105 | 5.3 | 12 | 40 | 2.8 | 7 | 87 | 7.7 | 17 | 96 | 4.3 | 10 | 58 |
| D190-04 | IM | 90 | 6.6 | 15 | 94 | 4.2 | 10 | 52 | 6.9 | 15 | Not detected | | | 87 | 6.4 | 14 | 54 |
| D190-05 | IM | 86 | 6.2 | 14 | 101 | 5.8 | 13 | 40 | 5.9 | 13 | 92 | 3.9 | 9 | 93 | 6.3 | 14 | 63 |
| D190-06 | IM | 92 | 6.2 | 14 | 98 | 3.8 | 9 | 55 | 7.3 | 16 | 95 | 4.4 | 10 | 84 | 2.3 | 6 | 55 |
| D190-07 | IM | 93 | 7.4 | 16 | 109 | 5.9 | 13 | 48 | 3.5 | 8 | 91 | 7.6 | 17 | 87 | 3.3 | 8 | 62 |
| D190-08 | IM | 91 | 7.5 | 17 | 100 | 6.9 | 15 | 53 | 4.8 | 11 | 85 | 3.8 | 9 | 93 | 3.3 | 8 | 60 |
| D190-09 | IM | 76 | 5.2 | 12 | 113 | 4 | 9 | 56 | 2.9 | 7 | 88 | 4.9 | 11 | 84 | 4 | 9 | 48 |
| D190-10 | IM | 86 | 5.8 | 13 | 119 | 3.9 | 9 | 55 | 5.1 | 12 | 96 | 4.3 | 10 | 91 | 6.2 | 14 | 58 |
| D190-01 | CIM | 84 | 5.9 | 8 | 119 | 6.5 | 9 | 45 | 8.7 | 12 | 80 | 6.9 | 9 | 78 | 8.4 | 12 | 50 |
| D190-02 | CIM | 88 | 16.5 | 20 | 100 | 8.5 | 9 | 46 | 8.9 | 10 | 82 | 6.9 | 7 | 77 | 9.5 | 10 | 56 |
| D190-03 | CIM | 83 | 11.2 | 13 | 104 | 9.2 | 11 | 39 | 10.8 | 13 | 88 | 12.4 | 15 | 94 | 10.7 | 13 | 65 |
| D190-04 | CIM | 89 | 7.2 | 10 | 93 | 5.4 | 7 | 53 | 8.2 | 12 | Not detected | | | 88 | 10 | 15 | 44 |
| D190-05 | CIM | 84 | 11.2 | 13 | 100 | 13.8 | 16 | 45 | 11 | 12 | 94 | 12.3 | 13 | 93 | 7 | 8 | 62 |
| D190-06 | CIM | 94 | 6 | 8 | 98 | 6.6 | 9 | 55 | 11 | 15 | 95 | 10.2 | 14 | 82 | 7.9 | 10 | 56 |
| D190-07 | CIM | 92 | 10.7 | 13 | 110 | 8.5 | 11 | 40 | 7.3 | 9 | 91 | 10.7 | 14 | 86 | 5.1 | 6 | 53 |
| D190-08 | CIM | 90 | 12.3 | 16 | 101 | 9.2 | 11 | 52 | 8.5 | 10 | 86 | 8.3 | 10 | 93 | 4.7 | 5 | 52 |
| D190-09 | CIM | 82 | 8.2 | 11 | 113 | 4.9 | 7 | 56 | 5.5 | 8 | 89 | 9.1 | 13 | 82 | 6.6 | 10 | 49 |
| D190-10 | CIM | 86 | 9.8 | 12 | 114 | 9 | 11 | 54 | 7.6 | 9 | 95 | 11.8 | 15 | 95 | 9.4 | 11 | 58 |

Total R^2 values are indicated for the total sum of effects from all QTLs. Values in italics indicate that the defined trait heritability has been exceeded based on QTL analysis results. Shaded cells indicate QTLs detected using the “suggestive” criterion according to Lander and Kruglyak (1995). Interval mapping method. IM, interval mapping; CIM, composite interval mapping. ^bPos., Position in centiMorgans (cM). ^cLOD, Logarithm of odds score. ^d R^2 , QTL effect explaining a proportion of the total phenotypic variance. ^eTotal R^2 is the sum of all five QTLs (D2 + D3 + D4 + D6 + D7). Total R^2 values should not theoretically exceed the total value as determined in D1000 using CIM (that is, 75%). ^fQTL not detected at significant or suggestive levels, indicating false negatives because the QTL was defined in this position in the true population.

Kruglyak (1995); Table 5). By definition, suggestive QTLs occur once by chance in a genome-wide scan; therefore, caution is always required when interpreting QTLs with this classification (Lander and Kruglyak, 1995). However, four false positive QTLs were significant

using IM (LOD scores between 2.7 and 3.6) and three were significant using CIM (LOD scores were between 3.0 and 4.5). This number is relatively high considering that only 80 datasets were generated and supports the practice of confirmation of QTL results in independent

populations (Collard et al., 2006; Tucker et al., 2007). Variation in LOD score due to sampling bias has rarely been reported. This is important because it is often used in conjunction with a R^2 value by breeders or geneticists to determine whether a QTL could be used for MAS. As has

Table 5. False positives detected in sub-sampled mapping populations.

| Population | QTL | IM ^a | | | | CIM ^a | | | |
|------------|-----|------------------------|------------------|-----------------------------|-------------------------|------------------|-----|----------------|------------|
| | | Pos. ^b (cM) | LOD ^c | R ² ^d | Sig. level ^e | Pos. (cM) | LOD | R ² | Sig. level |
| A094-06 | A5 | 58 | 2.9 | 13 | S | 58 | 4.5 | 7 | HS |
| A190-08 | A5 | 90 | 1.9 | 5 | Sg | 88 | 1.9 | 1 | Sg |
| A190-04 | A2 | 98 | 2.2 | 5 | Sg | 97 | 1.3 | 1 | NS |
| A190-05 | A2 | 55 | 1.9 | 4 | Sg | 105 | 0.7 | 0 | NS |
| A190-06 | A2 | 19 | 3.6 | 8 | S | 112 | 0.7 | 0 | NS |
| A190-07 | A1 | 112 | 1.6 | 4 | NS | 107 | 2.7 | 3 | Sg |
| B094-07 | B6 | 115 | 2.3 | 11 | Sg | 102 | 3.0 | 7 | S |
| B190-07 | B6 | 115 | 2.3 | 11 | Sg | 102 | 3.0 | 7 | S |
| B094-02 | B7 | 39 | 1.7 | 8 | Sg | 38 | 0.9 | 3 | NS |
| B094-04 | B7 | 4 | 2.7 | 13 | S | 6 | 1.6 | 3 | NS |
| B190-04 | B7 | 18 | 2.0 | 5 | Sg | 2 | 1.8 | 2 | Sg |
| B190-09 | B7 | 6 | 1.7 | 4 | Sg | 6 | 1.1 | 2 | NS |
| B190-10 | B7 | 94 | 1.9 | 5 | Sg | 93 | 1.7 | 2 | Sg |
| B094-10 | B3 | 73 | 2.2 | 10 | Sg | 17 | 0.5 | 1 | NS |
| B190-07 | B3 | 44 | 1.9 | 5 | Sg | 46 | 1.5 | 2 | NS |
| C094-02 | C7 | 25 | 2.1 | 10 | Sg | 24 | 0.9 | 1 | NS |
| C190-02 | C7 | 44 | 1.8 | 4 | Sg | 48 | 2.0 | 1 | Sg |
| C190-06 | C7 | 85 | 3.0 | 7 | S | 92 | 1.9 | 1 | Sg |
| C190-07 | C7 | 15 | 1.8 | 4 | Sg | 32 | 1.2 | 1 | NS |
| C190-09 | C7 | 17 | 2.6 | 6 | Sg | 16 | 0.9 | 1 | NS |
| C094-09 | C4 | 86 | 1.7 | 8 | Sg | 26 | 1.8 | 2 | Sg |
| C190-05 | C4 | 49 | 1.8 | 4 | Sg | 24 | 1.3 | 1 | NS |
| C190-06 | C4 | 47 | 1.9 | 5 | Sg | 61 | 0.9 | 1 | NS |
| D1000 | D5 | 100 | 2.6 | 1 | Sg | 63 | 0.7 | 0 | NS |
| D094-10 | D5 | 0 | 1.5 | 7 | NS | 0 | 1.7 | 6 | Sg |
| D190-03 | D5 | 53 | 2.6 | 6 | Sg | 45 | 1.3 | 1 | NS |
| D190-06 | D5 | 12 | 2.2 | 5 | Sg | 112 | 1.4 | 2 | NS |
| D190-09 | D5 | 30 | 2.1 | 5 | Sg | 30 | 0.6 | 1 | NS |
| D094-07 | D1 | 76 | 1.9 | 9 | Sg | 60 | 1.6 | 5 | NS |
| D094-09 | D1 | 14 | 2.5 | 12 | Sg | 14 | 1.5 | 4 | NS |
| D190-07 | D1 | 99 | 1.7 | 4 | Sg | 99 | 2.0 | 2 | Sg |
| D190-09 | D1 | 13 | 1.7 | 4 | Sg | 57 | 0.5 | 1 | NS |

Some QTLs were detected using either IM or CIM, but more false positives were detected using IM. The LOD scores indicate that some false QTLs were >2.5 which is sometimes used as the threshold for declaring a QTL as significant. ^aInterval mapping method; IM, interval mapping; CIM, composite interval mapping; ^bPos., position in centiMorgans (cM); ^cLOD, logarithm of odds score. ^dR², QTL effect explaining a proportion of the total phenotypic variance. ^eSignificance levels were defined as HS, highly significant; S, significant; Sg, suggestive; NS, not significant (shaded).

been detected experimentally (Keurentjes et al., 2005), LOD scores increased as population size increased.

One of the most important findings in this paper was the detection of variation in R² values, especially for IM results and for small population sizes (n = 94). R² values are an indicator of the utility of a QTL for introgression in breeding programs. In this study, QTLs were often overestimated or underestimated by more than 10%. The highest overestimation of R² value was for B094-04 (19% larger than the true QTL effect using IM; 15% larger using CIM). As the population size increased, the QTL effects decreased as can be observed by comparing results for population sizes of n = 94 versus n = 190 (Figure 4). The

variation detected between random samples of a large mapping population should warn researchers that in typically-used population sizes in molecular genetics research, R² values are only an estimate of the QTL effect and that random sampling can cause errors in estimating the true effect of a QTL.

Due to the development of high-throughput DNA marker genotyping methods [especially for single nucleotide polymorphism (SNP) markers], some linkage maps for QTL mapping will have considerably more markers than the ones used for the construction of "framework" maps in this study. Although marker density was not tested in this study, previous research strongly

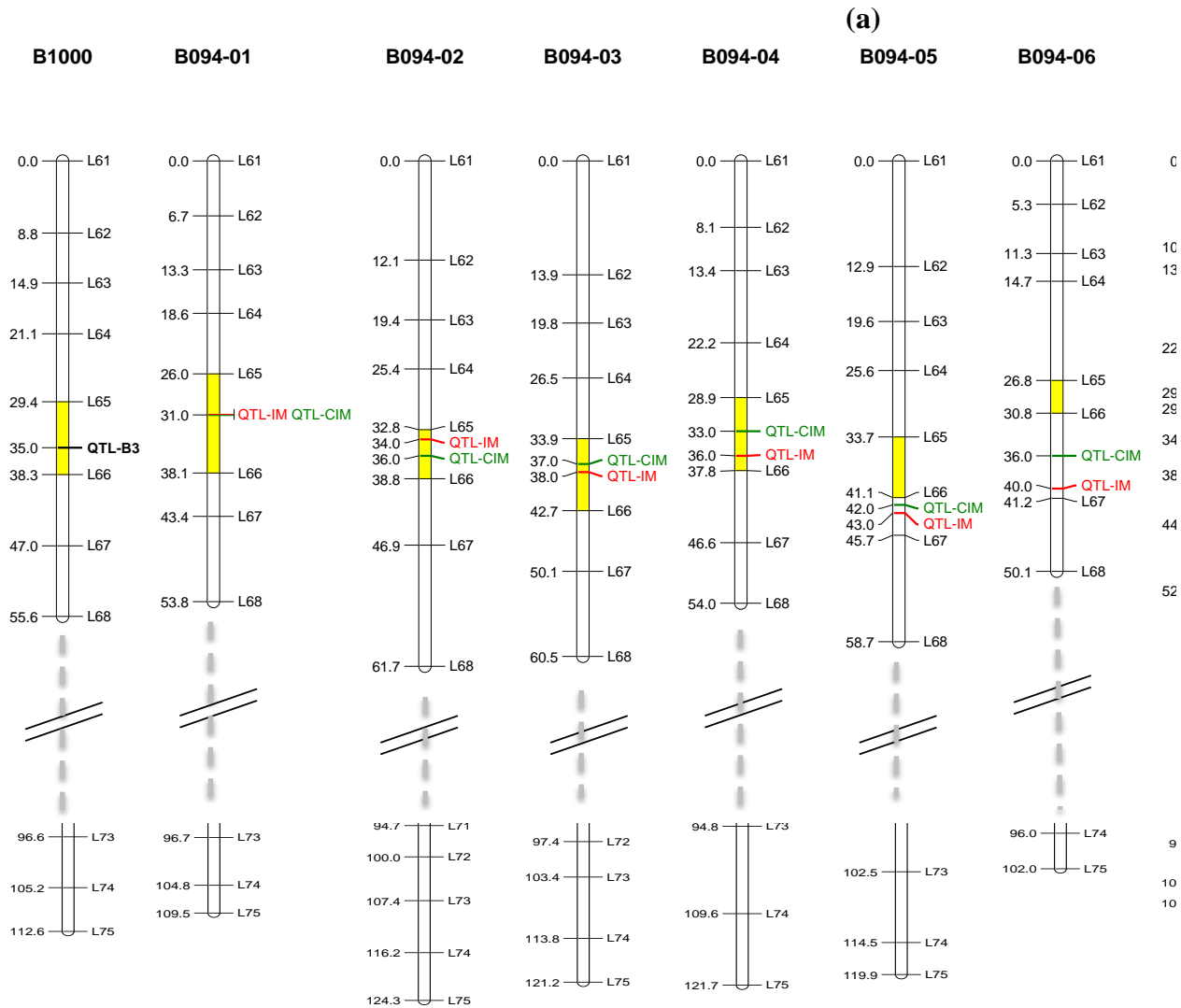


Figure 2. (a) Linkage maps indicating QTL positions (detected by IM (red) and CIM (green)) of 10 random sub-sampled populations (n=94) from population B1000 (shown left). Note the difference in linkage maps between populations, especially the total map length. The QTL containing region is shown in yellow for all populations. The true QTL (called QTL-B3) is located between markers L65 and L66. Note that the QTL was incorrectly located by both methods in populations B094-05 and B094-06. The QTL was incorrectly located by IM in populations, B094-07, B094-08 and B094-09. (b) Linkage maps indicating QTL positions (detected by IM (red) and CIM (green)) of 10 random sub-sampled populations (n=94) from population D1000 (shown left). Note the difference in total map lengths for each linkage group. The QTL containing region is shown in yellow for all populations. The true QTL (called QTL-D3) is located between markers L27 and L28, closer to marker L27. Note that the QTL was incorrectly located in populations D094-02 and D094-10 using IM and CIM, and incorrectly located in populations D094-05 using IM only. False negatives were detected in populations D094-04 and D094-06.

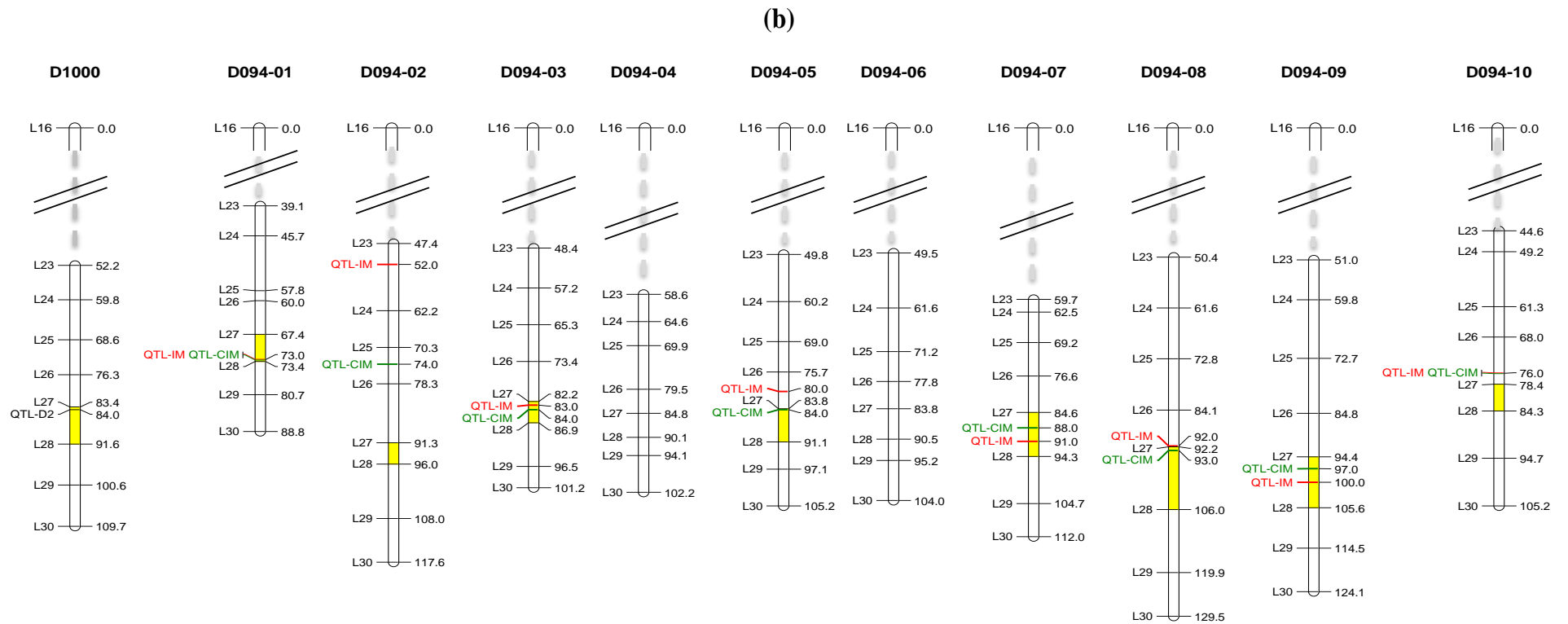


Figure 2. Contd.

suggests that QTL mapping results do not differ much when different average marker spacing is used (Darvasi et al., 1993; Darvasi and Soller, 1994). For example, previous research has indicated that marker spacing between 10 and 50 cM are sufficient for preliminary QTL mapping; this was the basis of simulating linkage maps with an average marker spacing of 10 cM in this study. In fact, the study by Collard et al. (2009) indicated that too many markers were detrimental to the accuracy of QTL mapping because marker orders were “unstable” (that is, different results were

obtained when mapping software programs were used to analyse an identical data set). Previous research in maize (Melchinger et al., 1998; Utz et al., 2000) has indicated that QTLs could be overestimated due to sampling bias, leading to an over inflated estimate of the utility of QTLs for MAS. While this was demonstrated here, it was also demonstrated that QTLs could be severely under-estimated potentially leading to useful QTLs being ignored for potential MAS. The largest underestimation of R^2 value was for A094-06 using CIM (17% lower than the true effect). This

also highlights the value of determining trait heritabilities so that the proportion of genetic variance explained can be accounted for. The variation detected for QTL effects in this study implies that QTL mapping results should be confirmed in independent populations.

Another important finding in this paper was the increase in accuracy of CIM versus IM, at least for RI populations between 94 and 190 individuals. This was based on higher LOD scores and reduced standard deviations for QTL positions; LOD scores from CIM results were higher than

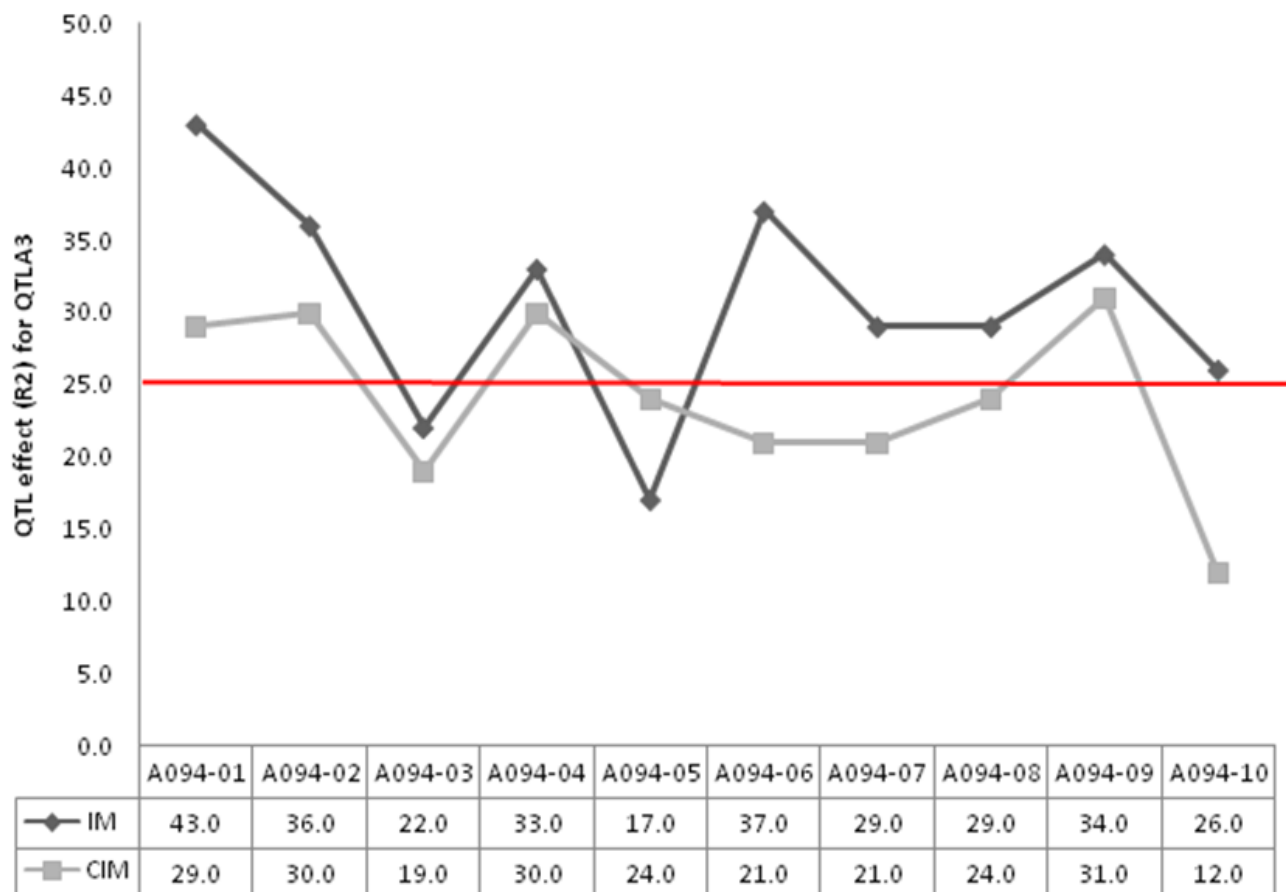


Figure 3. Graph showing variation in estimation of QTL effect (that is, R^2 value of QTL-A3) using IM (dark grey line with diamond shape co-ordinates) and CIM (light grey line with square co-ordinates) in 10 populations ($n = 94$) randomly sampled from A1000. The true effect of this QTL is 25% of the phenotypic variance; however, the observed magnitude varies markedly between different mapping populations (all derived by sampling from A1000).

those from IM results, indicating that a QTL was more likely. This implies that re-analysing QTL data generated from IM should be performed using CIM. This also supports map curation efforts such as described in Lehmsiek et al. (2005).

CONCLUSION AND RECOMMENDATION

The cost and time required for genotyping or phenotyping often restricts the size of a mapping population used, so what should researchers do if only population sizes of under <200 can be used (as have been used by the authors!)? In many cases, there are practical and financial reasons which influence the development of small population sizes (e.g. available glasshouse space for single seed descent for RI populations or technical constraints for the production of doubled-haploid lines in tissue culture laboratories). First, researchers should be careful in interpreting QTL mapping results and not automatically assume that the results could be suitable

for MAS. Secondly, QTL results should be confirmed in independent populations (Collard et al., 2006; Tucker et al., 2007). Thirdly, the population could be expanded to include more individuals after a preliminary QTL mapping study. In the last two recommendations, only the most tightly-linked markers to QTLs need to be used for marker genotyping in the “extended” population. This was proposed many years ago by Lander and Krugliak (1995). Furthermore, cross validation techniques (based on re-sampling) have been proposed for improving the accuracy of QTL mapping (Melchinger et al., 2004; Schon et al., 2004) and the results in this study support the investigation of these methods for mapping QTLs.

In conclusion, QTL results are a product of the mapping population used, and are largely influenced by population size and to a lesser extent, method for QTL analysis. It is recommended that population sizes of at least 190 lines are used when the trait heritability is 0.50 or higher. It is hoped that the results presented in this paper will assist plant breeders, molecular geneticists and other researchers in the execution of new experiments and

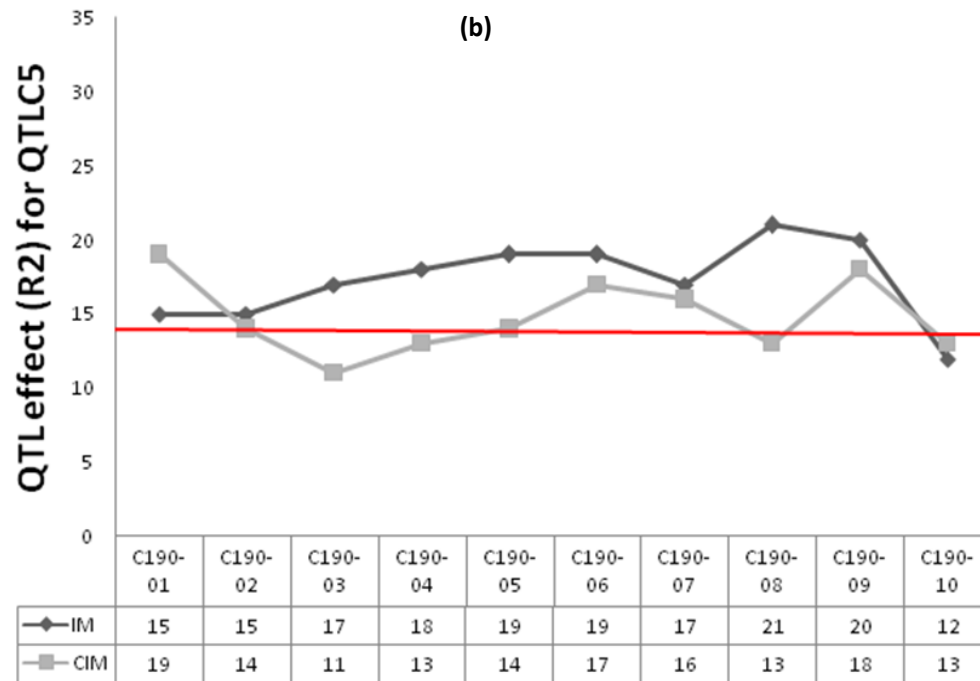
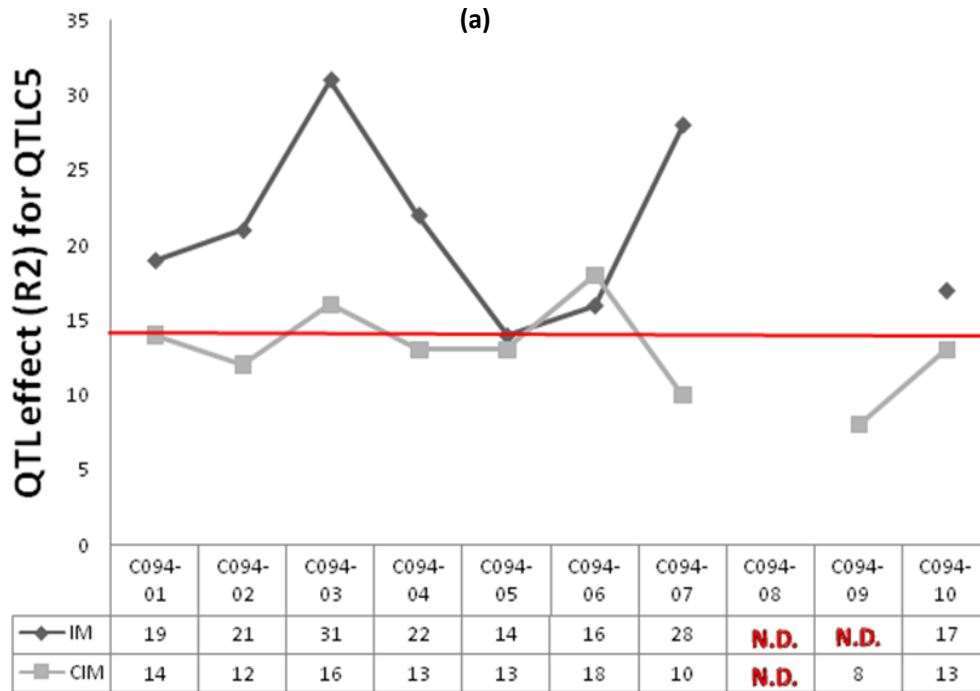


Figure 4. Comparison of variation in estimation of QTL effect (that is, R^2) using IM dark grey line with diamond shape co-ordinates) and CIM (light grey line with square co-ordinates) for populations consisting of 94 (a) and 190 (b) randomly sampled individual lines from C1000. The true effect of this QTL is 14% of the phenotypic variance. Note that when $n = 94$, this QTL was not detected (N.D.) in three populations. When $n = 190$, the variation in estimation in effect of QTL-C5 improves considerably.

interpretation of QTL mapping results in the future, especially for RI mapping populations of less than 200 individuals.

REFERENCES

Beavis WD (1994). The power and deceit of QTL experiments: lessons from comparative QTL studies. 49th Annual Corn and Sorghum

- Industry Research Conference. Am. Seed Trade Assoc., pp. 250-266.
- Beavis WD (1998). QTL Analyses: Power, precision and accuracy. In: A.H. Paterson (Ed.), *Molecular Dissection of Complex Traits*. CRC Press, Boca Raton.
- Collard BCY, Jolley R, Bovill WD, Wildermuth GB, Sutherland MW (2006). Confirmation of QTL mapping and marker validation for crown rot seedling resistance in wheat line 2-49. *Aust. J. Agric. Res.*, 57: 967-973
- Collard BCY, Mace E, McPhail M, Wenzl P, Cakir M, Fox G, Poulsen D, Jordan D (2009). How accurate are the marker orders in crop linkage maps generated from large marker datasets? *Crop Pasture Sci.*, 60: 362-372.
- Collard BCY, Mackill DJ (2008). Marker-assisted selection: an approach for precision plant breeding in the 21st century. *Phil. Trans. R. Soc. B*. 363: 557-572.
- Darvasi A, Soller M (1994). Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. *Theor. Appl. Genet.*, 89: 351-357.
- Darvasi A, Weinreb A, Minke V, Weller 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.
- Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, Vale G (2005). Marker assisted selection in crop plants. *Plant Cell Tis. Org.*, 82: 317-342.
- Hackett C (2002). Statistical methods for QTL mapping in cereals. *Plant Mol. Biol.*, 48: 585-599.
- Helms TC, Doetkott C (2007). An educational software for simulating the sample size of molecular marker experiments. *J. Nat. Resour. Life Sci. Educ.*, 36: 53-57.
- Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M (2007). Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics*, 175: 891-905
- Lander ES, Botstein D (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121: 185-199.
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987). Mapmaker an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, 1: 174-181.
- Lander ES, Kruglyak L (1995). Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat. Genet.*, 11: 241-247.
- Lehmensiek A, Eckermann PJ, Verbyla AP, Appels R, Sutherland MW, Daggard GE (2005). Curation of wheat maps to improve map accuracy and QTL detection. *Aust. J. Agric. Res.*, 56: 1347-1354.
- Manly KF, Cudmore Robert Jr H, Meer JM (2001). Map Manager QTX, cross-platform software for genetic mapping. *Mamm. Genome*, 12: 930-932.
- Melchinger AE, Utz HF, Schon CC (1998). Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics*, 149: 383-403.
- Melchinger AE, Utz HF, Schon CC (2004). QTL analyses of complex traits with cross validation, bootstrapping and other biometric methods. *Euphytica*, 137: 1-11
- Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature*, 335: 721-726.
- Piepho HP (2001). A quick method for computing approximate thresholds for quantitative trait loci detection. *Genetics*, 157: 425-432
- Schon CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics*, 167: 485-498.
- Semagn K, Bjornstad A, Ndjiondjop MN (2006). Principles, requirements and prospects of genetic mapping in plants. *Afr. J. Biotech.*, 5: 2569-2587.
- Semagn K, Bjornstad A, Xu YB (2010). The genetic dissection of quantitative traits in crops. *Electron. J. Biotechnol.*, 13(5): 14.
- Tinker NA, Mather DE (1993). GREGOR: software for genetic simulation. *J. Hered.*, 84: 237.
- Tucker DM, Griffey CA, Liu S, Brown-Guedira G, Marshall DS, Maroof MAS (2007). Confirmation of three quantitative trait loci conferring adult plant resistance to powdery mildew in two winter wheat populations. *Euphytica*, 155: 1-13.
- Utz HF, Melchinger AE, Schon CC (2000). Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics*, 154: 1839-1849.
- Vales MI, Schon CC, Capetini F, Chen XM, Corey AE, Mather DE, Mundt CC, Richardson KL, Sandoval-Islas JS, Utz HF, Hayes PM (2005). Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust. *Theor. Appl. Genet.*, 111: 1260-1270.
- Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.*, 93: 77-78.
- Xu S (2003). Theoretical basis of the Beavis effect. *Genetics*, 165: 2259-2268.