A Model Selection Based Interval Mapping Method for Autopolyploids

Dachuang Cao*, Bruce A. Craig*, & R.W. Doerge*,†
* Department of Statistics, Purdue University, West Lafayette, IN 47907
† Department of Agronomy, Purdue University, West Lafayette, IN 47907
ABSTRACT

While extensive progress has been made in quantitative trait locus (QTL) mapping for diploid species, similar progress in QTL mapping for polyploids has been limited due to the complex genetic architecture of polyploids. To date, QTL mapping in polyploids has focused mainly on tetraploids with dominant and/or codominant markers. Here, we extend this view to include any even ploidy level under a dominant marker system. Our approach first selects the most likely chromosomal marker configurations using a Bayesian selection criterion, and then fits an interval mapping model to each candidate. Profiles of the likelihood ratio test statistic and the maximum likelihood estimates (MLEs) of parameters including QTL effects are obtained via the EM algorithm. Putative QTL are then detected using resampling based significance threshold, and the corresponding parental configuration is identified to be the underlying parental configuration from which the data are observed. Although presented via pseudo-doubled backcross experiments, this approach can be readily extended to other breeding systems. Our method is applied to single-dose restriction fragment autotetraploid alfalfa data, and the performance is investigated through simulation studies.

Key words: autopolyploid, QTL mapping, interval mapping
INTRODUCTION

QTL (quantitative trait loci) mapping detects and identifies the regions of a genome associated with the variation of a quantitative trait of interest. Molecular markers have been used extensively to construct genetic maps for diploid species (Koornneef et al. 1983; Dietrich et al. 1996), and act as the foundation for QTL analyses such as interval mapping (Lander and Botstein 1989; Knapp et al. 1990; Haley and Knott 1992), composite interval mapping (Zeng 1993; Zeng 1994), and multiple QTL mapping (Jansen 1993). The statistical issues involved in diploid QTL mapping are reviewed in Doerge et al. (1997), Broman (2001), and Doerge (2002).

Polyploids are organisms having more than two complete sets of chromosomes in a cell. Polyploidy is most common in plants, but also found in some insects, amphibians, and reptiles. Some 30% to 70% of today’s angiosperms and at least half of the flowering plants are thought to be polyploid (Soltis and Soltis 1999). Not only is polyploidy important in agriculture, it is also an important evolutionary force (Soltis and Soltis 2000; Wolfe 2001; Osborn et al. 2003; Soltis et al. 2003).

To describe the current state of QTL mapping in polyploids, they must first be classified according to their homology. A polyploid with genomes all derived from the same species is called an autopolyploid. If the multiple sets of chromosomes are derived from different species, the polyploid is called an allopolyploid. A segmental allopolyploid represents species with homology in between allopolyploids and autopolyploids (Suzuki et al. 1989). For allopolyploids (or, disomic polyploids), such as bread wheat and potato, meiosis is usually restricted to the pairing of ancestral parental homologues. Thus, the genetics of allopolyploids are similar to diploids, and diploid QTL mapping methods can be car-
ried out by treating each of the two sets of homologous chromosomes from an ancestral genome as a diploid.

For autopolyploids (or, polysomic polyploids), however, the high ploidy level and homology create complications in interpreting the meiotic process. Unlike diploids whose two homologues always pair during meiosis, autopolyploids may undergo either bivalent pairing (two chromosomes pair) or multivalent pairing (more than two chromosomes pair) (Muller 1916; Newton and Darlington 1929; Darlington 1931; Sybenga 1995). Furthermore, the manner in which the paired chromosomes segregate during meiosis, especially for multivalent pairing, varies in different species (Mather 1938; Rieseberg and Doyle 1989; Jackson and Jackson 1996). Finally, the marker genotypes in autopolyploids are not always identifiable (i.e., the banding patterns are not unique since not all the alleles can be differentiated by any one marker system). Even if we are able to identify the different alleles of one locus, the number of copies (dosage) of each allele and the linkage phases (i.e., the arrangement of alleles on the informative chromosomes) between loci remain unknown.

There are two different approaches for dealing with the multiple copies of one locus for autopolyploids. With a dominant marker system, each locus is treated as biallelic. For this case, Wu et al. (1992) estimated a genetic map for autopolyploids with simplex markers, or single dose restriction fragment (SDRF) markers, which represent only one homologue and segregate 1:1 in the progeny. Ripol et al. (1999) extended the method of Wu et al. (1992) to any dominant marker with an unobservable dosage level. With co-dominant markers each locus is treated as multiallelic (Luo et al. 2000; Luo et al. 2001; Hackett et al. 2001). Each band observed on a gel is viewed as a different allele for that locus. This approach is valid for some marker systems, like single sequence
repeats (SSRs) that do not require the use of DNA digestion by restriction enzymes and whose motifs have high variation in the number of repeat units. However, since the use of SSRs may be responsible for inaccurate allele identification, Rodzen and May (2002) suggested scoring multiallelic SSR markers as individual dominant markers, unless the markers’ underlying modes of inheritance are known, simply because different loci may have different inheritance patterns. Because the availability of co-dominant markers for autopolyploids is relatively rare, and dominant markers (e.g., SDRF markers) have the advantages of being rich in plants and easily scored, the remainder of this paper is based on the dominant marker system.

An additional advantage of using single dose restriction fragment (SDRF) markers is that the only unknown factor is the linkage phase (coupling or repulsion) between the SDRFs and this can be estimated by a goodness-of-fit test. In turn, the maximum likelihood estimate (MLE) of the recombination between SDRF markers can be calculated based on the estimated linkage phase. Using SDRFs, Brouwer and Osborn (1999) estimated the first composite genetic map for tetraploid alfalfa. When considering multiple dose markers, Ripol et al. (1999) first estimated marker dosages and linkage phase, and then constructed the genetic map by computing the maximum likelihood estimate of recombination fraction based on estimated parental marker configurations. Doerge and Craig (2000) extended polyploid QTL methodology by developing an algorithmic model selection process for a single marker QTL analysis with dominant markers for autopolyploids with any even ploidy level. Recently, Luo et al. (2004) addressed both bivalent and quadrivalent chromosomal pairing during meiosis by employing a statistical framework for the analysis of tetrasomic linkage.

As a continuation of the work by Doerge and Craig (2000), we propose an interval
mapping method by employing an available genetic map to increase the power of detecting and estimating QTL locations within an autopolyplloid bivalent pairing framework. For simplicity, and in keeping with this previous work, we base our work on a pseudo-doubled backcross experiment (Grattapaglia and Sederoff 1994) and employ a model selection based interval mapping to simultaneously estimate model parameters, which include QTL location, parental marker dosages, parental QTL configuration, and QTL effect. Initially, we estimate the parental marker dosages using progeny marker data to reduce the number of potential parental configurations. Based on each putative parental configuration, interval mapping is performed to estimate QTL location and to calculate maximum likelihood estimates of QTL effect and population variation that are obtained using the EM algorithm (Dempster et al. 1977). We limit our approach to any even ploidy level with multiple dose dominant markers since odd ploidy levels are often associated with high infertility. After describing the methodology, our method is applied to single-dose restriction fragment alfalfa data with an estimated linkage map from Brouwer and Osborn (1999). Several putative QTL have been identified as associated with winter hardiness traits in bivalent tetraploid alfalfa, and act as the basis of simulation studies that investigate the performance of our approach.

METHODS

Our polyploid QTL mapping methodology is developed for organisms undergoing bivalent pairing during meiosis. Under the bivalent pairing assumption, the pairing of the chromosomes can range from preferential pairing (always pairing with the same homologous chromosome) to random pairing (equally likely to pair with any other homologous chromosome). The data analysis and simulation will be based on random pairing.
Let $n$ denote the number of progeny and $k$ denote the ploidy level. In what follows, for each locus, upper case is used to denote both the locus name and its dominant allele, and lower case represents the recessive allele (*e.g.* $M_1$ and $m_1$). The dosage of a locus refers to the number of copies of the dominant allele at that locus. For example, a locus dosage of two refers to two copies of the dominant allele at that locus. When a marker is observed, *at least* one copy of the dominant allele for that marker is present.

Marker data (presence/absence) and the quantitative trait data are measured on the $F_1$ progeny. These progeny are the result of a pseudo-doubled backcross experiment (Figure 1) where an informative parent $P_1$ is crossed with a non-informative parent $P_2$. The dosage at each locus in the parents is at least one for the informative parent, and zero for the non-informative parent. Furthermore, at most half of the chromosomes of the informative parent contain dominant alleles. Chromosomes which contain dominant alleles in the informative parent will be called informative chromosomes. The pseudo-doubled backcross design is equivalent to doubling the half of the non-informative chromosomes of the informative parent to produce the non-informative parent, and then crossing them to get $F_1$. An example of a pseudo-doubled backcross experiment for a tetraploid ($k = 4$) with two marker loci $M_1$ and $M_2$ is shown in Figure 1. In this case, the dosage for both markers is two in the informative parent. A pseudo-doubled backcross population can also be developed from diploids by choosing two homozygous parents carrying different alleles, doubling each of them to produce tetraploids, crossing them to produce an $F_1$, then crossing the $F_1$ to the parent containing recessive alleles.

**Interval Mapping:** Consider two flanking markers, $M_1$ and $M_2$, and one QTL, $Q$, with marker genetic distance or recombination fraction provided by an available genetic map. Once the parental configuration $c$ is known, interval mapping can be applied to
locate QTL. A parental configuration includes loci dosages and linkage phase. If the dominant alleles of two loci are on the same chromosome, they are in coupling phase, otherwise, they are in repulsion phase. Based on this, the five respective configurations for a tetraploid with two markers under a pseudo-doubled backcross design are called double coupling (Figure 2 – A), asymmetric coupling 12 (Figure 2 – B) asymmetric coupling 21 (Figure 2 – C) coupling (Figure 2 – D), and repulsion (Figure 2 – E). Similarly, there are three possible QTL configurations: (1) one dose of the QTL located on the first informative chromosome, (2) one dose of the QTL on the second informative chromosome, and (3) two doses of the QTL on both informative chromosomes. In what follows, configurations A to E will be used to denote a marker configuration for a tetraploid as defined above. A marker configuration letter followed by a QTL configuration number, will represent a complete parental configuration, such as “A1”, which represents a parental configuration with markers in double coupling and one QTL on the first informative chromosome.

For the $i^{th}$ individual, the observed data are $x_i = (y_i, I_i)$, where $y_i$ is the trait value, and $I_i$ is the indicator vector for marker presence (i.e., $I_{ih} = 1$ if marker $M_h$ is present and 0 otherwise). Assume for each QTL dosage level $j$ ($j = 0, 1, \cdots, k/2$), the trait $y$ has a normal distribution with mean $\mu_j$, and common variance $\sigma^2$ for all the dosages of QTL. The vector $\theta = (\mu = (\mu_0, \mu_1, \cdots, \mu_{k/2}), \sigma)$ will be used to denote model parameters. At any fixed putative position of the QTL the likelihood function of the trait is a mixture of normal distributions (Equation 1) under the assumption of one QTL,

$$L(\theta|x, c) = \prod_{i=1}^{n} \sum_{j=0}^{k/2} P(d_Q = j | I_i, c) \phi(y_i; \mu_j, \sigma^2) \overset{\triangle}{=} \prod_{i=1}^{n} \sum_{j=0}^{k/2} \omega_{j|i} \phi(y_i; \mu_j, \sigma^2), \quad (1)$$

where $P(d_Q = j | I_i, c) \overset{\triangle}{=} \omega_{j|i}$ is the probability of the $i^{th}$ progeny having dosage $j$ of
the QTL given marker presence status $I_i$, and $\phi(y_i; \mu_j, \sigma^2)$ denoting the normal density function valued at $y_i$ with mean $\mu_j$ and variance $\sigma^2$. The $\omega_{j|i}$’s are a function of parental configuration $c$, and the recombination fractions between the QTL and markers.

In the interval mapping framework (LANDER and BOTSTEIN 1989), the putative QTL is considered at incremental positions in an interval defined by adjacent markers, and a test statistic (commonly, the log-likelihood ratio) is evaluated. Let $R_h(h = 1, 2)$ denote the true recombination fractions between marker $M_h$ and QTL. At each evaluation position, let $r_h(h = 1, 2)$ denote the recombination fractions between markers defining the interval and the current QTL position. The log-likelihood ratio test statistic is calculated to test the hypothesis

$$H_0 : R_h = 0.5, \ h = 1, 2$$
$$H_a : R_h = r_h, \ h = 1, 2.$$ 

The null hypothesis assumes that the QTL is present, but unlinked to both markers, while the alternative hypothesis assumes that the QTL is present and linked to $M_1$ and $M_2$ with recombination fractions $r_h$. The log-likelihood ratio test statistic LRT is,

$$LRT = -2 \ln \frac{L(\hat{\theta}_0, R_1 = R_2 = 0.5)}{L(\hat{\theta}_a, R_1 = r_1, R_2 = r_2)},$$

where $\hat{\theta}_0$ and $\hat{\theta}_a$ are MLEs under the null and alternative hypotheses, respectively. If the test statistic is significant then the corresponding interval position with the largest LRT statistic is considered as the estimated QTL position. A permutation test can be performed to estimate the significance threshold for the test statistic (CHURCHILL and DOERGE 1994).
Because the summation format of the likelihood function (1) causes difficulty in finding the MLE of $\theta$, the complete likelihood function, containing both observable and unobservable data, is considered

$$L(\theta | x) = \prod_{i=1}^{n} \prod_{j=0}^{k/2} \left[ \omega_{ji} \phi(y_i; \mu_j, \sigma^2) \right]^{z_{ij}},$$

(2)

and the EM (expectation-maximization) algorithm (Dempster et al. 1977) employed to estimate $\theta$. Here, the unobservable data $\{z_i = (z_{i0}, \cdots, z_{i,k/2})\}_{i=1}^{n}$, where $z_{ij} = 1$ if the QTL dosage for the $i^{th}$ progeny is $j$ and 0 otherwise, are indicator vectors representing progeny QTL dosages.

To implement the EM algorithm, the E (estimation) step and M (maximization) step are iterated until a convergence criterion of the log-likelihood (e.g., $\log L_{t+1} - \log L_t < 10^{-6}$) is satisfied. In the E step, the missing data $z_{ij}$ are estimated based on the observable data, and in the M step, the MLE of $\theta$ are calculated using the estimated $z_{ij}$. After $t$ iterations, estimated values of $z_{ij}^{t+1}, \mu^t$, and $\sigma^t$ are used at step $t + 1$ such that the $z_{ij}^{t+1}$ are estimated in the E step by

$$z_{ij}^{t+1} = P_{\theta}(d_Q = j | x_i, c, \mu^t, \sigma^t) = \frac{\omega_{ji} \exp\left\{ -\frac{1}{2} \left( \frac{y_i - \mu_j^t}{\sigma^t} \right)^2 \right\}}{\sum_{l=0}^{k/2} \omega_{li} \exp\left\{ -\frac{1}{2} \left( \frac{y_i - \mu_l^t}{\sigma^t} \right)^2 \right\}}.$$

Given $z_{ij}^{t+1}$, the MLE of $\theta = (\mu, \sigma)$ is obtained (M step) by maximizing (2),

$$\mu_j^{t+1} = \frac{\sum_{i=1}^{n} z_{ij}^{t+1} y_i}{\sum_{i=1}^{n} z_{ij}^{t+1}}, \quad j = 0, \cdots, k/2;$$

$$\sigma^{t+1} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{k/2} z_{ij}^{t+1} (y_i - \mu_j^{t+1})^2}.$$
Since the estimation procedure is based on a specific parental configuration (model), the model space expands quickly as the marker number and/or ploidy level increases. Enumerating each possible model for the purpose of exploring which is most likely to produce the observed data, and then estimating parameters based upon the most likely model(s) is computationally challenging. For example, under a pseudo-doubled backcross design with two markers and one QTL, there are 14 possible parental configurations for a tetraploid, 91 for a hexaploid ($k = 6$), and 390 for an octaploid ($k = 8$). To reduce these computations, we employ a model reduction step prior to parameter estimation.

**Model Reduction:** To reduce the model space, one can consider one of two approaches: either a marginal or joint statistical method. Under the marginal method, we calculate the posterior probability of dosages for each marker individually. If one particular dosage level has a posterior probability higher than a specified cutoff, only that marker dosage is considered in candidate parental configurations. Otherwise, we select the most likely dosage levels until the sum of posterior probabilities exceeds the cutoff and these dosages are used to form the candidate parental configuration set. The joint method, on the other hand, takes a similar approach, but information on all markers is considered jointly. In this case, parental marker configuration posterior probabilities are calculated directly, and candidate parental configurations are chosen using the same posterior probability rule as described for the marginal method.

Given the marginal marker presence/absence distribution (Ripol et al. 1999), the posterior probabilities for parental marker dosages of each marker are calculated using Bayes’ rule (Bayes 1783). For one marker $M$, the number of progeny with $M$ absent, $n_\emptyset$, has a binomial distribution $Bin(n, p_{d_M})$, where $d_M$ denotes the marker dosage of $M$ in the informative parent. Under the pseudo-double backcross design, $p_{d_M} = \binom{k-d_M}{k/2} / \binom{k}{k/2}$
for random pairing. Thus, given the informative parent marker dosage $d_M$,

$$P(N_\emptyset = n_\emptyset|n, d_M) \triangleq \text{Bin}(n_\emptyset; n, d_M) = \binom{n}{n_\emptyset} p_{d_M}^{n_\emptyset} (1 - p_{d_M})^{n-n_\emptyset}. $$

Let $\pi$ denote the prior distribution of marker dosage $d_M$. The posterior probability of each dosage level is

$$P(d_M|n, n_\emptyset) = \frac{\text{Bin}(n_\emptyset; n, p_{d_M})\pi(d_M)}{\sum_{d=1}^{k/2} \text{Bin}(n_\emptyset; n, p_d)\pi(d)} = \frac{p_{d_M}^{n_\emptyset} (1 - p_{d_M})^{n-n_\emptyset} \pi(d_M)}{\sum_{d=1}^{k/2} p_d^{n_\emptyset} (1 - p_d)^{n-n_\emptyset} \pi(d)}. $$

As an alternative method to the described marginal approach, we use the joint marker information and genetic map to calculate the posterior probability of a parental marker configuration. Given markers $M_1$ and $M_2$, all progeny can be classified into four sets according to the presence status of the two markers. Let $n_{obs} = (n_{00}, n_{01}, n_{10}, n_{11})$ denote the observed frequency vector in the four sets, where the $h^{th}$ ($h = 1, 2$) subscript is 1 if $M_h$ is present, and 0 otherwise. Assuming no segregation distortion and a parental configuration $c$, the distribution of $n_{obs}$ follows a multinomial distribution with probability parameter vector $p^c = (p_{00}^c, p_{01}^c, p_{10}^c, p_{11}^c)$. Table 1 lists $p^c$ for a tetraploid. The probability of observing $n_{obs} = (n_{00}, n_{01}, n_{10}, n_{11})$ is

$$P(n_{obs}|n, c) \triangleq \text{Multinomial}(n_{obs}; n, c) = \frac{n!}{\prod_{i,j=0}^{1} n_{ij}!} \prod_{i,j=0}^{1} \{p_{ij}^c\}^{n_{ij}}. $$

Let $\pi$ denote the prior distribution of marker configuration $c \in C$. The posterior probability for a specific parental marker configuration $c_0 \in C$ is

$$P(c_0|n, n_{obs}) = \frac{\text{Multinomial}(n_{obs}; n, c_0) \pi(c_0)}{\sum_{c\in C} \text{Multinomial}(n_{obs}; n, c) \pi(c)} = \frac{\prod_{i,j=0}^{1} \{p_{ij}^{c_0}\}^{n_{ij}} \pi(c_0)}{\sum_{c\in C} \prod_{i,j=0}^{1} \{p_{ij}^c\}^{n_{ij}} \pi(c)}. $$
Among the many criteria for assessing the worth of a model reduction method are ease of implementation, a high probability of selecting the correct model, and efficiency in reducing the size of the candidate model space. Even though the marginal method is easy to implement and less computationally expensive, the joint method directly estimates the posterior probability of parental marker configurations. One might expect that because there may be multiple marker configurations having the same marker dosages, especially if the marker dosages are low and ploidy high, the joint method will be more efficient in reducing the model space.

A simulation study was performed to compare the performance of the marginal and joint methods for a tetraploid under random pairing. We chose two markers and marker genetic distance varied from 0.1M to 0.5M with 0.1M increments. Sample sizes ranged from 50 to 500 with 50 increments. For each simulation setting, 10,000 data sets were generated. The cutoff probability was set to be 0.90 for both methods. Without any prior information concerning the parental marker configuration, we assumed all the possible parental marker configurations to be equally likely (i.e., the prior for the joint method is discrete uniform). Under this assumption, the prior distribution of a marker dosage is not uniform since a lower dosage is more frequent than a higher dosage. Among the five possible parental marker configurations (Figure 2; A-E) for a tetraploid, dosage 1 has a frequency of 3/5 and dosage 2 has a frequency of 2/5 for each marker locus. In turn, this frequency distribution was used as the prior of marker dosage for the marginal method.

The proportion of data sets for which the correct configuration was included in the candidate configuration space, $p_{inc}$, was recorded (Table 2) for both methods. For the marginal method, this refers to selecting the correct marker dosages for both marker loci, while for the joint method, it refers to selecting the correct marker configuration. The pro-
portion of data sets for which a unique configuration was selected, $p_{uni}$, was also recorded as a measurement of the efficiency for reducing the model space (Table 3). In general, both methods performed better with a larger sample size and shorter marker distances. With sample size 100 and above, both $p_{inc}$ and $p_{uni}$ were approximately 1.0 for all the simulation settings. With a small sample size, the extra information from the genetic map added strength to the joint method both in selecting the correct configuration and in reducing the model space. Based on the simulation results, the joint model reduction method is recommended for data analysis.

Once the model reduction step is completed, the selected sub-models are used to form the pool of candidate parental marker configurations by combining these sub-models with possible parental QTL configurations. After this, the likelihood function is constructed for each candidate parental configuration, and the EM algorithm is applied to maximize the likelihood function and estimate parameters. The parental configuration which has the maximal likelihood is then identified as the parental configuration from which the observed data have arisen.

**DATA ANALYSIS**

The proposed approach was employed to analyze a published alfalfa single-dose restriction fragment (SDRF) data set with traits related to winter hardiness (Brouwer and Osborn 1999). Alfalfa is an autotetraploid that undergoes bivalent random pairing (Bingham and McCoy 1988; Cao et al. 2004). Winter hardiness is a complex trait and one of the most important adaptations for alfalfa. Because there is no direct measurement of winter hardiness, related traits including winter injury (WI), fall growth (FG), freezing injury (FI), and unifoliate internode length (UIL), were measured in addition to
survival percentage of plants (SURV). Survival percentage (SURV) was measured by the percentage of each plant that survived in the winter (0% to 100%). Fall growth (FG) was measured by height of vertical regrowth in centimeters in early October in two successive years 1995 and 1996; freezing injury (FI) was measured by absorbance at 265 nm (ABS) and electrical conductivity per gram fresh weight (COND) in 1995 and 1996; winter injury (WI) scaled 1-5 (1 – no injury, 5 – dead) was measured in 1996 and 1997; and unifoliate internode length (UIL) was measured for all seedlings. The histogram plots of these traits can be referred to Fig. 1 in Brouwer et al. (2000).

Two genotypes, B17 and P13, representing the extremes for each trait were crossed, and a single $F_1$ plant was crossed to each parent to create two populations of 101 individuals each. Both populations were scored for 82 single-dose restriction fragment (SDRF) loci and measured for each trait in 2 years of replicated field trials. The composite map and the original analysis of the trait data and marker data can be found in Brouwer et al. (1999, 2000). Based on the genetic map, seven linkage groups were identified and corresponded to seven out of the eight chromosomes. As such in what follows, the terms for chromosome and linkage group will be used interchangeably.

Because only B17 markers were considered for the P13 x $F_1$ backcross population, and similarly only P13 markers for the B17 x $F_1$ backcross, each backcross population is equivalent to a pseudo-doubled backcross population. Our approach was applied to identify regions in the genome related to the previously described quantitative traits for both backcross populations based on the composite map. The average over the replicates in each year for each measurement was treated as a quantitative trait, and the two averages were analyzed separately. A permutation test (Churchill and Doerge 1994) with 1000 permutation for a 0.05 significance level was used to detect a significant QTL.
Seven QTL were detected for fall growth (FG), two for fall injury (FI), one for winter hardiness (SURV), and one for winter injury (WI) in the backcross to B17; and one QTL was detected for UIL in the backcross to P13 (Table 4). Among the seven QTL detected for fall growth (FG), two were mapped on chromosome 1, three on chromosome 5, and two on chromosome 8. Two pairs of QTL were detected at close map positions across two years: on chromosome 1, one QTL was detected at 0.31M for the 1995 average, and one at 0.34M for the 1996 average; on chromosome 5, one QTL was detected at 0.498M for the 1995 average, and one at 0.456M for the 1996 average. Two QTL were detected on chromosome 6 and chromosome 8 for freezing injury (FI). One QTL was detected on chromosome 8 for winter hardiness (SURV) and winter injury (WI). On chromosome 8, in the region from 0.187M to 0.247M, four QTL were detected for multiple traits: fall growth (FG 1996), freezing injury (COND 1996), winter hardiness (SURV 1995), and winter injury (1996), and two of them at the same location 0.187M. The one QTL for UIL in the P13 backcross was at the top of chromosome 8. All of the detected QTL were identified to have dosage two, except for the QTL for SURV. In general, an additional copy of a QTL was found to be associated with a higher quantitative trait mean, except for the QTL for SURV and UIL.

Brouwer et al. (2000) used a regression-based single marker analysis to identify markers associated with the previously described quantitative traits. Multiple regression analyses were performed using the same data for all 82 SDRFs to identify the best polygenic models for each trait in each backcross population. With the backcross to B17, the SDRF multiple regression models include six SDRFs for FG (1995), six for FG (1996), five for FI (COND 1996), and four for WI (1996) (Table 5 summarized from Table 2 in Brouwer et al. (2000)). Among all the detected SDRFs, all but five are located on chromosome 1, 5,
and 8, which also contain most of QTL we detected with the proposed method. Two loci for FG (1995) and three for FG (1996) are also shared by our interval mapping method. Among the other significant marker loci, four had likelihood ratio test statistics larger than the 0.05 experiment-wise threshold used in our interval mapping method, but were not local peaks. Five loci had likelihood ratio test statistics being local peaks, with two of them close to the 0.05 threshold.

Using an experiment-wise threshold based on a significance level 0.05, rather than an asymptotic-based strict cutoff based on a significance level 0.05 or 0.005 for each marker (Brouwer et al. 2000), we detected fewer QTL by controlling the type I error. The region on chromosome 8 (from 0.187M to 0.247M) that was identified by our method as being shared by multiple traits was not detected by the regression method. This is most likely the direct result of more experimental information being incorporated into the analysis (e.g., model selection, dosages, intermarker position). Since single marker analysis only explores the relationship between the quantitative trait and marker loci, if the QTL is very close to a marker locus, the estimated effect for that marker locus could be interpreted as the effect of the QTL, under the assumption that the QTL is also of single dose and in coupling with the marker. However, most of the QTL that were detected using our method were of dosage two, therefore, it is not appropriate to compare the QTL effects estimated by Brouwer et al. (2000) with our method.

Ma et al. (2002) analyzed the same alfalfa data for two traits, winter injury (WI) and fall injury (FI: COND). They identified five QTL for fall injury (FI: COND) on chromosome 4, 5, 6, and 8, and four QTL for winter injury on chromosome 1, 5, and 8. The specifics of the Ma et al. (2002) method and analysis make a direct comparison between our results and their results difficult for the following reasons. First, it is assumed in Ma
et al. (2002) that the markers and QTL have the same dosage level and are in coupling, while most of our detected QTL were of dosage two. Second, there is some evidence that the preferential pairing factor estimation process provided by MA et al. (2002) is flawed since in reality the genotypic information collected for across years should not change, yet their estimates of preferential factor do (Cao et al. 2004). Finally, in MA et al. (2002) the critical value for the significance level 5% is calculated based on 200 permutation tests. For a 5% significance level (Churchill and Doerge 1994; Doerge and Churchill 1996), 200 permutations limit full exploration of the null space, and may lead to the declaration of some QTL which are actually not significant at the 5% level.

**SIMULATION**

We used the framework provided by both the alfalfa experiment and analysis as a means to investigate the reliability of the detected QTL via a simulation study. Data were simulated by treating the estimated parental configuration and related parameters as the true values. For this purpose, five out of the twelve QTL that we detected were chosen to provide a range for the flanking marker genetic distance and configuration, QTL dosage and configuration, QTL effect, and QTL position (Table 6). The flanking marker distances were 0.358M, 0.358M, 0.098M, 0.222M, and 0.087M for the five intervals (Table 6). The flanking markers in the first two intervals were in repulsion, and the others were in coupling. The estimated QTL dosage was one for the fourth interval, and two for the other intervals. The simulation setting of the second interval was similar to that of the first interval. We included the second interval in our simulation because we were interested in investigating whether our method is able to identify the small increment from $\mu_1$ to $\mu_2$, where $\mu = (\mu_0, \mu_1, \mu_2) = (12.51, 19.18, 19.29)$ is the estimated mean vector.
for the corresponding QTL in that interval.

The five intervals settings were studied separately. For each interval, 1000 data sets with a sample size of 100 (i.e., similar to the alfalfa data) were simulated and analyzed, with results summarized in Table 6. Using the joint model selection method, the correct parental marker configuration was included in the candidate parental marker configuration pool 100% of the time for all the five cases. The overall parental configuration pool is then formed by the combination of candidate parental marker configurations and possible QTL configurations. After fitting our interval mapping model to each possible parental configuration, the one having the maximal test statistic provides estimated QTL location, marker and QTL configuration, as well as the parameter estimates. The correct configuration was identified at least 95% of the time with markers in coupling, and at least 99% with markers in repulsion, except for the third interval where the correct configuration was selected 86% of the time. Given the correct configuration, the estimated parameters were close to the true values used for simulation.

The fact that our method performed better with repulsion than coupling is due to the difference between these two configurations. This can be demonstrated by considering an example where there are two copies of the QTL on both informative chromosomes. When the markers are in coupling, one copy of the QTL is located on a chromosome containing no dominant marker alleles. Thus, not much extra information with respect to the recombination between markers and the QTL can be provided by this extra copy of the QTL. This, in turn, makes it difficult to choose between the configurations with one copy and two copies of the QTL. On the other hand, with markers in repulsion, both copies of the QTL provide information about recombination between markers and QTL. The effect of this difference is easily illustrated by comparing the second interval
with the third interval used for this simulation. Although the increment in trait means
from QTL dose 1 to dose 2 is only \(0.035\sigma\) for case 2 and \(1.22\sigma\) for case 3, case 2 still
outperformed case 3. Based on this, repulsion is not necessarily the worst case scenario
for QTL mapping in tetraploids with simplex markers, as was once thought.

Since the above simulation was mainly focused in parental marker configurations
formed by simplex markers, another simulation study was performed to serve as the
basis of investigation for how in general, QTL or marker dosages and their respective
linkage phase affect the performance of our algorithm. Under a pseudo-doubled backcross
experiment, two markers \((M_1/m_1\) and \(M_2/m_2\)) and one QTL \((Q/q)\) were used with marker
distance 0.10 M, and a sample size 100. The location of the QTL was 0.01 M, 0.03 M,
or 0.05 M. The trait mean \(\mu\) was \((10.0, 12.0)\) if QTL dosage was 1, or \((10.0, 12.0, 14.0)\) if
QTL dosage was 2, and \(\sigma\) was 1.0. 1000 data sets were simulated for each combination
of the parameter settings and each of the following configurations: double coupling (A1,
A3), asymmetric coupling 12 (B1, B2, B3), coupling (D1, D2, D3), and repulsion (E1,
E2, E3).

Our simulation results illustrate that when the correct parental configuration was
identified, the parameter estimate averages were close to the true value. The ability to
identify the correct configuration, however, depended on the underlying parental configu-
ration. The worst case scenario was when markers were in coupling on the first informative
chromosome with one copy of the QTL on the second informative chromosome (D2). For
this case the identified correct configuration proportion ranged from 63% to 66% (Figure
3). This is not surprising since with D2 under random pairing, only one third of the pair-
ings provide information concerning the recombination between marker and QTL. On the
other hand, the best situation was when two markers were in repulsion, where the correct
parental configuration was identified almost 100% regardless of the QTL configuration or location. Most of the other configurations obtained a proportion of at least 95% (Figure 3).

It is worth noting that in our earlier work (Cao et al. 2003), repulsion did not outperform configurations with higher loci dosages, such as double coupling and asymmetric coupling. The difference in results comes from the difference in the assumed chromosome pairing mechanism. Our earlier work (Cao et al. 2003) investigated the performance of our interval mapping method under preferential pairing, which specifically deals with the bivalent pairing (i.e., an informative chromosome only pairs with a non-informative chromosome). The preferential pairing mechanism is favorable to configurations with higher loci dosages because it maximizes the amount of recombination information through maximizing the disequilibrium in pairing. However, with random pairing, one third of the time one informative chromosome pairs with another informative chromosome to provide less, or even no information, if two chromosomes are homozygous. Using similar reasoning it can be demonstrated that random pairing is more informative for markers in repulsion than preferential pairing.

DISCUSSION

One of the largest challenges when mapping QTL in polyploids is the incomplete parental configuration (model) information. This is compounded by the dramatic increase in the size of the model space as the number of markers and/or ploidy level increase. To combat these challenges in autopolyploid QTL mapping, our interval mapping algorithm takes one of two approaches. The size of the model space is addressed using Bayesian model selection by first identifying potential parental marker dosages using a marginal approach,
or marker configurations based on a joint approach. We have demonstrated that both methods are effective in reducing the model space, but that the joint method is more efficient due to the fact that it incorporates the genetic map information. From the reduced model space, the candidate parental configuration pool is formed by combining each potential parental marker configuration with a possible QTL configuration. For each potential parental configuration, profiles of maximum likelihood estimates (MLEs) of QTL effects, as well as population variation and likelihood ratio test statistics are produced at each evaluation point of a grid search across the genome. Putative QTL are detected using a resampling based significance threshold. The corresponding parental configuration (and related parameter estimates) is then taken as the underlying model from which the data are produced.

Our simulation studies have demonstrated that if the correct parental configuration is identified, our interval mapping method provides parameter estimates close to true values. However, the chance of identifying the correct configuration depends on the true parental configuration, chromosome pairing system, and genetic distance between markers and the QTL. This helps us address a question raised in Doerge and Craig (2000), “for which situation is the linkage mostly affected?” Comparatively, among these factors, the true parental configuration is the most influential one. Since a configuration consists of both loci dosages and linkage phase, the influence of one cannot be considered without considering the effect of the other. For example, a configuration with higher marker/QTL dosages does not always imply a greater chance of identifying the correct configuration as shown in the case of repulsion. However, a configuration with higher dosages is often more informative because it often has a higher disequilibrium, which as we all know is the basis for linkage analysis. Furthermore, such a configuration is less sensitive to the chromosome
pairing system, and thus consistently provides high power for detecting QTL.

Our simulation results also help to answer some of the remaining questions presented by Doerge and Craig (2000). First, “if a molecular marker is found to be tightly linked to a QTL, should the dosage of the marker agree with the dosage of the QTL?”. Not necessarily. There is no causal relationship between the dosages of the QTL and the flanking markers. Although as discussed previously, a configuration with the marker and QTL having the same dosage often has a higher disequilibrium, and thus a higher power of detecting QTL. Second, “should the models which are controlled by dosage levels be weighted for the purpose of representing more realistic results?”. Weighting the candidate models may be a better way than depending solely on the one “best” model, when the uncertainty about parental configuration is high. However, even configurations with same marker dosage levels could differ to a greater extent biologically, therefore further investigation is needed to find both a statistically and biologically sound weighting scheme. Finally, “would models with dosage levels more similar to each other be more likely, especially with close linkage (short genetic distance)?”. This answer depends on the linkage phase since it is possible for two models with exactly the same dosage levels and genetic association to be biologically different due to different linkage phases, and therefore they are different in likelihood.

We applied our approach to a published alfalfa data set for quantitative traits that are related to winter hardiness. Our analysis confirms several putative QTL identified by a regression-based single marker analysis (Brouwer et al. 2000), and it also identified a new region on chromosome 8 that was shared by several traits. Our analysis results demonstrate that for a majority of the detected QTL, there is a positive or negative correlation between the quantitative trait mean and QTL dosage level, i.e., $\mu_0 \leq \mu_1 \leq$
\[
\cdots \leq \mu_{dQ} \text{ or } \mu_0 \geq \mu_1 \geq \cdots \geq \mu_{dQ}.
\]
If such an ordering of means is available before analysis, an order restricted version of interval mapping can be performed to improve power of detection. With an order constraint, the problem falls into the category of isotonic regression and it can be computed using the pool-adjacent-violator algorithm (Robertson et al. 1988). We have performed simulations (not shown) to compare the power between an order restriction and without. In fact, even when the only information is that the means are ordered, and the direction remains unknown, there is a slight improvement in the performance. The improvement is most evident when the order of the means is increasing or decreasing with respect to the QTL dosage, is known \textit{a priori}.

Thus far it has been assumed that the QTL dosage is at most half the ploidy level, which is a reasonable assumption given the pseudo-doubled backcross experimental design that we considered. In practice, if this assumption does not hold, we can simply change the upper bound for the QTL dosage to the ploidy level. This change will result in more candidate QTL configurations to choose from, but the method and the procedure remain the same. Similarly, if the breeding design is altered from a pseudo-doubled backcross design, our approach can be adapted. Also, if co-dominant markers are considered, they can be easily incorporated in the model by including parameters such as allelic effect and interaction.

The effect of the assumptions we place on the genetic map that is used for QTL mapping in polyploids is one area of research that remains unaddressed. When the genetic map is estimated, the most likely parental marker configuration in turn has to be estimated since recombination requires a parental marker configuration. Because of this, the parental marker configuration used for estimating the genetic map should be consistent with the one used to locate QTL. However, if the most likely parental marker configuration is not
the correct configuration, the reliability of the map and the mapped QTL are affected. By selecting the pool of potential parental configurations based on observed progeny data, our method can serve as a further verification for the estimated genetic map. A framework is currently under development to allow variability or uncertainty in estimating the map to be considered when mapping QTL.

Acknowledgment We thank Dr. Tom Osborn for providing the data set and continued work in this area, and we thank Dr. Gary Churchill for his comments this work.

LITERATURE CITED


25


Figure 1: A pseudo-doubled backcross experiment for a tetraploid with two loci. Each locus is assumed to be biallelic: $M_1/m_1$ and $M_2/m_2$, where $M_1$ and $m_1$ denote two alleles of the first locus and $M_1$ is the dominant allele, and similarly for $M_2$ and $m_2$. For each locus, there are two copies of the dominant allele in the informative parent $P_1$. $P_2$ is the non-informative doubled-haploid produced by doubling a haploid of $P_1$, which only contains recessive alleles.
Figure 2: Possible parental marker configurations for a tetraploid with two marker loci ($M_1/m_1$ and $M_2/m_2$) under a pseudo-doubled backcross design. A. Double coupling; B. Asymmetric coupling 12; C. Asymmetric coupling 21; D. Coupling; E. Repulsion.
Figure 3: The proportion identifying the correct parental (marker and QTL) configuration for a tetraploid under a pseudo-doubled backcross experiment. The trait mean $\mu$ is (10.0, 12.0) if QTL dosage is 1, or (10.0, 12.0, 14.0) if QTL dosage is 2. Population standard deviation $\sigma$ is 1.0. With sample size 100, 1000 data sets are simulated under 11 parental configurations corresponding to numbers 1 to 11 on the horizontal axis: A1, A3, B1, B2, B3, D1, D2, D3, E1, E2, and E3. The letters denote the marker configurations: (A) double coupling, (B) asymmetric coupling 12, (D) coupling, and (E) repulsion. The numbers denote the QTL configurations: (1) one dose of QTL present on the first informative chromosome, (2) one dose of QTL present on the second informative chromosome, and (3) two doses of QTL present on both informative chromosomes. $d(M_1, Q)$ is the genetic distance between the QTL and the left end flanking marker.
Table 1: Multinomial distribution probability parameters $\mathbf{p}^c = (p_{00}^c, p_{01}^c, p_{10}^c, p_{11}^c)$ in a tetraploid with two markers under random pairing.

|   | $^{*}(M_1M_2|M_1M_2)$ | $(M_1M_2|M_1)$ | $(M_1M_2|M_2)$ | $(M_1|M_2)$ | $(M_2|M_1)$ |
|---|----------------------|-----------------|-----------------|-------------|-------------|
| $p_{00}^c$ | $\frac{r^2-2r+1}{6}$ | $\frac{1-r}{6}$ | $\frac{1-r}{6}$ | $\frac{1-r}{2}$ | $\frac{1+r}{6}$ |
| $p_{01}^c$ | $\frac{-r^2+2r}{6}$ | $\frac{2+r}{6}$ | $\frac{r}{6}$ | $\frac{r}{2}$ | $\frac{2-r}{6}$ |
| $p_{10}^c$ | $\frac{-r^2+2r}{6}$ | $\frac{r}{6}$ | $\frac{2+r}{6}$ | $\frac{r}{2}$ | $\frac{2-r}{6}$ |
| $p_{11}^c$ | $\frac{r^2-2r+5}{6}$ | $\frac{3-r}{6}$ | $\frac{3-r}{6}$ | $\frac{1-r}{2}$ | $\frac{1+r}{6}$ |

* Only dominant alleles are shown for informative chromosomes, separated by “|”.
† $r$ denotes the estimated recombination fraction between the two markers.
Table 2: Estimated $p_{inc}$, probability of selecting the correct configuration in the candidate configuration (model) space, for a tetraploid with two markers, 10,000 simulated data sets, and a cutoff of 0.90 for both the marginal and joint model reduction methods under random pairing.

<table>
<thead>
<tr>
<th></th>
<th>50</th>
<th>100</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>M2</td>
<td>M1</td>
</tr>
<tr>
<td>Marginal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M1M2</td>
<td>M1M2)</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>(M1M2</td>
<td>m1M2)</td>
<td>0.992</td>
<td>0.992</td>
</tr>
<tr>
<td>(M1M2</td>
<td>M1m2)</td>
<td>0.992</td>
<td>0.992</td>
</tr>
<tr>
<td>(M1M2</td>
<td>m1m2)</td>
<td>0.988</td>
<td>0.987</td>
</tr>
<tr>
<td>(m1M2</td>
<td>M1m2)</td>
<td>0.983</td>
<td>0.987</td>
</tr>
<tr>
<td>Joint</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M1M2</td>
<td>M1M2)</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>(M1M2</td>
<td>m1M2)</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>(M1M2</td>
<td>M1m2)</td>
<td>1.000</td>
<td>0.999</td>
</tr>
<tr>
<td>(M1M2</td>
<td>m1m2)</td>
<td>0.999</td>
<td>0.998</td>
</tr>
<tr>
<td>(m1M2</td>
<td>M1m2)</td>
<td>0.999</td>
<td>0.999</td>
</tr>
</tbody>
</table>

* $n$ denotes sample size.
† Marker genetic distance expressed in Morgan (M).
Table 3: Estimated $p_{uni}$, probability of selecting a unique configuration in the candidate configuration (model) space, for a tetraploid with two markers, 10,000 simulated data sets, and a cutoff of 0.90 for both the marginal and joint model reduction methods under random pairing.

<table>
<thead>
<tr>
<th></th>
<th>50</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1 0.3 0.5</td>
<td>0.1 0.3 0.5</td>
<td>0.1 0.3 0.5</td>
<td>0.1 0.3 0.5</td>
<td>0.1 0.3 0.5</td>
</tr>
<tr>
<td></td>
<td>Marginal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(M1M2</td>
<td>M1M2)</td>
<td>0.997 0.996 0.995</td>
<td>1.000 1.000 1.000</td>
<td>1.000 1.000 1.000</td>
</tr>
<tr>
<td></td>
<td>(M1M2</td>
<td>m1M2)</td>
<td>0.901 0.904 0.910</td>
<td>0.991 0.992 0.991</td>
<td>0.999 0.999 0.999</td>
</tr>
<tr>
<td></td>
<td>(M1M2</td>
<td>m1m2)</td>
<td>0.903 0.908 0.900</td>
<td>0.991 0.991 0.989</td>
<td>0.999 0.999 0.999</td>
</tr>
<tr>
<td></td>
<td>(M1M2</td>
<td>m1m2)</td>
<td>0.860 0.842 0.841</td>
<td>0.983 0.982 0.979</td>
<td>0.997 0.998 0.998</td>
</tr>
<tr>
<td></td>
<td>(m1M2</td>
<td>M1M2)</td>
<td>0.813 0.821 0.820</td>
<td>0.982 0.978 0.982</td>
<td>0.998 0.998 0.998</td>
</tr>
<tr>
<td></td>
<td>Joint</td>
<td>0.1 0.3 0.5</td>
<td>0.1 0.3 0.5</td>
<td>0.1 0.3 0.5</td>
<td>0.1 0.3 0.5</td>
</tr>
<tr>
<td></td>
<td>(M1M2</td>
<td>M1M2)</td>
<td>0.992 0.989 0.983</td>
<td>0.999 1.000 0.999</td>
<td>1.000 1.000 0.999</td>
</tr>
<tr>
<td></td>
<td>(M1M2</td>
<td>m1M2)</td>
<td>0.999 0.991 0.979</td>
<td>1.000 0.999 0.999</td>
<td>1.000 1.000 1.000</td>
</tr>
<tr>
<td></td>
<td>(M1M2</td>
<td>m1m2)</td>
<td>0.999 0.989 0.978</td>
<td>1.000 1.000 0.999</td>
<td>1.000 1.000 1.000</td>
</tr>
<tr>
<td></td>
<td>(M1M2</td>
<td>m1m2)</td>
<td>0.999 0.981 0.867</td>
<td>0.999 0.999 0.981</td>
<td>1.000 1.000 0.997</td>
</tr>
<tr>
<td></td>
<td>(m1M2</td>
<td>M1M2)</td>
<td>0.999 0.987 0.900</td>
<td>1.000 0.999 0.985</td>
<td>1.000 1.000 0.997</td>
</tr>
</tbody>
</table>

* $n$ denotes sample size.
† Marker genetic distance expressed in Morgan (M).
Table 4: Estimated locations and effects of QTL associated with winter hardiness and its related traits for autotetraploid alfalfa. Data are from Brouwer and Osborn (1999). Significant QTL were detected by an experiment-wise thresholds calculated from 1000 permutation tests at 0.05 significance level.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Year</th>
<th>( ^a \text{Ch} )</th>
<th>Marker Interval</th>
<th>( ^b \text{Pos} )</th>
<th>( ^c d_1 )</th>
<th>( ^d d_Q )</th>
<th>( ^e \mu )</th>
<th>( ^f \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ^g \text{FG} )</td>
<td>95</td>
<td>1</td>
<td>uwg284–vg2b9</td>
<td>0.31</td>
<td>0.31</td>
<td>2</td>
<td>12.24, 21.45, 24.61</td>
<td>2.61</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>1</td>
<td>uwg284–vg2b9</td>
<td>0.34</td>
<td>0.34</td>
<td>2</td>
<td>12.51, 19.18, 19.29</td>
<td>3.17</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>5</td>
<td>vg2a2–vg2e5</td>
<td>0.498</td>
<td>0.001</td>
<td>2</td>
<td>12.30, 21.65, 23.75</td>
<td>2.82</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>5</td>
<td>hg2b12–vg2a2</td>
<td>0.456</td>
<td>0.05</td>
<td>2</td>
<td>12.70, 18.48, 22.08</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>5</td>
<td>mtsc35–ugac473</td>
<td>0.29</td>
<td>0.02</td>
<td>2</td>
<td>15.23, 18.46, 20.90</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>8</td>
<td>ugac291–ugac540</td>
<td>0.187</td>
<td>0.10</td>
<td>2</td>
<td>13.02, 18.60, 21.53</td>
<td>3.08</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>8</td>
<td>ugac109–uwg169</td>
<td>0.704</td>
<td>0.06</td>
<td>2</td>
<td>15.57, 18.95, 19.95</td>
<td>3.62</td>
</tr>
<tr>
<td>( ^h \text{COND} )</td>
<td>96</td>
<td>6</td>
<td>ugac281–vg2c2</td>
<td>0.356</td>
<td>0.09</td>
<td>2</td>
<td>52.42, 59.05, 94.60</td>
<td>13.30</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>8</td>
<td>ugc291–ugac540</td>
<td>0.247</td>
<td>0.16</td>
<td>2</td>
<td>50.91, 58.36, 92.89</td>
<td>12.64</td>
</tr>
<tr>
<td>( ^i \text{SURV} )</td>
<td>95</td>
<td>8</td>
<td>ugac291–ugac540</td>
<td>0.187</td>
<td>0.10</td>
<td>1</td>
<td>0.54, 0.14</td>
<td>0.229</td>
</tr>
<tr>
<td>( ^j \text{WI} )</td>
<td>96</td>
<td>8</td>
<td>ugc291–ugac540</td>
<td>0.217</td>
<td>0.13</td>
<td>2</td>
<td>2.66, 4.40, 4.82</td>
<td>0.513</td>
</tr>
<tr>
<td>( ^k \text{UIL} )</td>
<td>-</td>
<td>8</td>
<td>ugac235–ugac291</td>
<td>0.001</td>
<td>0.001</td>
<td>2</td>
<td>9.34, 3.35, 12.74</td>
<td>1.847</td>
</tr>
</tbody>
</table>

\( ^a \) Ch denotes the chromosome number.
\( ^b \) Pos denotes the estimated composite map position of the QTL.
\( ^c \) \( d_1 \) denotes the genetic distance between the detected QTL and left end marker of the interval.
\( ^d \) \( d_Q \) denotes the estimated QTL dosage.
\( ^e \) \( \mu = (\mu_0, \mu_1, \ldots, \mu_{d_Q}) \) is the estimated mean vector.
\( ^f \) \( \sigma \) is the estimated common standard deviation of the trait.
\( ^g \) The trait fall growth (FG) was measured by height of vertical regrowth in centimeters in early October in two successive years 1995 and 1996.
\( ^h \) Electrical conductivity per gram fresh weight (COND) was measured in 1995 and 1996 to represent freezing injury (FI).
\( ^i \) Winter survival percentage of plants (SURV) was measured in 1995 and 1996 to represent winter hardiness.
\( ^j \) Winter injury (WI) was scaled 1-5 (1 – no injury, 5 – dead), and was measured in 1996 and 1997.
\( ^k \) UIL denotes the unifoliate internode length.
Table 5: Estimated locations of QTL affecting traits related to winter hardiness in Brouwer et al. (2000). (A) Fall growth (FG) 1995 average. (B) Fall growth (FG) 1996 average. (C) Freezing injury (FI – COND) 1996 average. (D) Winter injury (WI) 1996 average.

<table>
<thead>
<tr>
<th></th>
<th>A Ch Marker</th>
<th>B Ch Marker</th>
<th>C Ch Marker</th>
<th>D Ch Marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>aCh</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>vg2b9</td>
<td>*35.8</td>
<td>*35.8</td>
<td>d35.8</td>
</tr>
<tr>
<td>1</td>
<td>mtsc9</td>
<td>51.4</td>
<td>d46.1</td>
<td>e51.4</td>
</tr>
<tr>
<td>3</td>
<td>uwg96</td>
<td>d24.3</td>
<td>*27.0</td>
<td>e23.1</td>
</tr>
<tr>
<td>3</td>
<td>mtsc14</td>
<td>e46.1</td>
<td>*29.3</td>
<td>e30.9</td>
</tr>
<tr>
<td>5</td>
<td>vg2a2</td>
<td>*49.8</td>
<td>*30.9</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ugac191</td>
<td>63.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Locus close to the QTL identified by the interval mapping method.

*a* Ch denotes the chromosome.

*b* Pos is the estimated composite map position of the QTL.

*c* LRT at this locus is higher than 0.05 threshold, but not a peak.

*d* LRT at this locus is a peak close to the 0.05 threshold.

*e* LRT at this locus is a peak, not close to the 0.05 threshold.
Table 6: Estimated QTL location and effect for data simulated from five intervals containing detected QTL as indicated by alfalfa data and analysis. For each interval, 1000 data sets of sample size 100 were simulated.

<table>
<thead>
<tr>
<th>QTL</th>
<th>Trait</th>
<th>(^a)Pos</th>
<th>(^b)Mconfig</th>
<th>(d(M_1, M_2))</th>
<th>(a\bar{d}_1)</th>
<th>(\sigma)</th>
<th>(\bar{\mu})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FG</td>
<td>0.31</td>
<td>E</td>
<td>0.358</td>
<td>0.31</td>
<td>2.61</td>
<td>12.24, 21.45, 24.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>999</td>
<td>0.30(0.066)</td>
<td>2.55</td>
<td>12.32, 21.49, 24.54</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>FG</td>
<td>0.34</td>
<td>E</td>
<td>0.358</td>
<td>0.34</td>
<td>3.17</td>
<td>12.51, 19.18, 19.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>992</td>
<td>0.30(0.092)</td>
<td>3.04</td>
<td>12.53, 19.21, 19.29</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>FG</td>
<td>0.456</td>
<td>D</td>
<td>0.098</td>
<td>0.05</td>
<td>2.96</td>
<td>12.70, 18.48, 22.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>862</td>
<td>0.059(0.036)</td>
<td>2.82</td>
<td>12.68, 18.58, 21.89</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>SURV</td>
<td>0.187</td>
<td>D</td>
<td>0.222</td>
<td>0.10</td>
<td>0.23</td>
<td>0.54, 0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>955</td>
<td>0.097(0.041)</td>
<td>0.23</td>
<td>0.54, 0.14</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>UIL</td>
<td>0.001</td>
<td>D</td>
<td>0.087</td>
<td>0.001</td>
<td>1.85</td>
<td>9.34, 3.35, 12.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>954</td>
<td>0.015(0.027)</td>
<td>1.86</td>
<td>9.14, 3.41, 12.50</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Pos denotes the position of the detected QTL on the composite genetic map.

\(^b\) Mconfig denotes marker configuration (D is coupling and E is repulsion).

\(^c\) \(d_1\) denotes the distance between the QTL and the left end marker.

\(^d\) \(\sigma\) is the common standard deviation of the trait for all QTL dosage levels.

\(^e\) \(\mu = (\mu_0, \mu_1, \cdots, \mu_d_Q)\) is the mean vector of the trait corresponding to QTL dosages.

\(^f\) Freq denotes the frequency of selecting the correct parental configuration out of 1000 data sets.

\(^g\) \(\bar{d}_1\) denotes the average of the 1000 estimated QTL locations from the 1000 simulated data sets.

\(^h\) \(s(\bar{d}_1)\) denotes the sample standard deviation of the 1000 estimated QTL locations from the 1000 simulated data sets.

\(^i\) \(\bar{\sigma}\) denotes the average of population standard deviation estimates.

\(^j\) \(\bar{\mu}\) denotes the average of population mean vector estimates.