An Algorithm for Analyzing Linkages Affected by Heterozygous Translocations: QuadMap

J. D. Durrant, B. W. Gardunia, K. D. Livingstone, M. R. Stevens, and E. N. Jellen

From the Department of Plant and Animal Sciences, 275 WIDB, Brigham Young University, Provo, UT 84602 (Durrant, Stevens, and Jellen); the Department of Crop and Soil Sciences, New Beasley Lab, Texas A&M University, College Station, TX 77802 (Gardunia); and the Department of Biology, Trinity University, San Antonio, TX 78212 (Livingstone).

Address correspondence to E. N. Jellen at the address above, or e-mail: rick_jellen@byu.edu.

Introduction

Molecular linkage mapping depends on genetic differences between chosen mapping parents. To increase polymorphism, researchers often choose divergent parents. In some cases, sufficient polymorphism is found within the species of interest, but in other cases, the use of mapping populations derived from interspecific hybrids is necessary. As the genetic distance between parents increases, so does the probability of gross cytological differences including translocations, inversions, duplications, and deletions (Bourque and Pevzner 2002). These differences complicate mapping efforts due to reduction in recombination rates, false statistical association of markers from different chromosomes, and generation of segmental aneuploids (Livingstone et al. 2000). In this study we describe a software application called QuadMap that facilitates creation and comparison of multiple maximum likelihood maps of linkage groups to aid in the analysis of linkage groups affected by segregating translocations, even when the exact physical location of the translocations is unknown.

When analyzing simulated data, Livingstone et al. (2000) observed that markers spanning reciprocal translocations or inversions clustered together with markers from both chromosomes of the rearrangement. This is defined as a pseudolinkage group. Furthermore, when they created and compared multiple mappings, they found that the variance in the centimorgan distance between any two markers on different segments was, on average, 10-fold higher than that of markers on the same segment. This difference in variance could be used to unravel the false statistical associations caused by the genomic differentiation. Consequently, they determined that it was possible to map accurately around genomic rearrangements although at the time there was no software application available to implement the procedure in other experiments. QuadMap meets this need, allowing systematic implementation of the algorithm. In this study, we demonstrate the utility and function of this program using two simulations and suggest experimental situations where it may be used.

Materials and Methods

We analyzed the results for a simulated F₆ recombinant inbred line (RIL) population with two chromosomes (20 markers each) that was segregating for a translocation between markers 5 and 6. We also created a simulation of the same population and markers without a translocation. Many other simulations were used to test the program for errors during its development but are not reported in this article.

We created simulations using Tq.exe, a program provided by N. Tinker (ECORC, Agriculture and Agri-Food Canada) for simulation of populations derived from translocation heterozygotes with multiple markers. Tq.exe creates graphical genotypes and MapMaker data files for simulated individuals from F₂, RIL, doubled haploid, F₁, and parental populations. It also allows the user to control the length...
of the translocated segment, population size and type, the number of markers, and the recombination rate. Tq.exe was used to test and refine the algorithm of Livingstone et al. (2000) for analyzing the linkage relationships in populations segregating for chromosomal interchanges.

To statistically resolve the arms of the quadrivalent linkage group, we adapted the algorithm proposed by Livingstone et al. (2000), which uses the differences among multiple maps to unravel the pseudolinkage group into markers that are on the same segments. QuadMap was written in VISUAL BASIC 6.0 (Microsoft, Redmond, WA) in order to implement the steps described by Livingstone et al. (2000), with some modification. Rather than constructing linkage maps, QuadMap automates and summarizes multiple map construction using MapMaker/EXP (Lander et al. 1987). It essentially allows the user to run a bootstrap-like analysis to obtain a measure of confidence for linkage maps. To order the markers of each segment, MapMaker/EXP or a similar package should be used.

QuadMap automates the following steps. First, it asks the user for the location of the MapMaker/EXP-format input file (for format guidelines, see Green et al. 1987) and the required number of iterations (1,000 iterations were used in the simulations described in this article, for example). Analyzing data from multiple linkage groups simultaneously results in low frequencies and variances for the group in question because different linkage groups will be counted in each iteration. As the purpose of the program is to make many maps of a single linkage group, it is best to include only data from a single linkage group in the analysis at a time. Second, QuadMap permutes the order of the markers in the input file and reads those permutations into MapMaker, executing the following commands for each iteration: `prepare data, "your permuted file"; sequence all`, indicating to MapMaker that it should consider all markers in the file; and `order 3 30 3 3`, creating a maximum likelihood map from the sequence of markers starting with the first three markers that are within 30 cM of each other with a logarithm of the odds (LOD) score of three and adding new markers with an LOD score of three. MapMaker employs a greedy algorithm that assumes that existing marker orders are correct when adding new markers to the map. Consequentially, input files that list the markers in different orders can lead to maps that vary minimally. With pseudolinkage groups, the “true” map is not linear but can be represented as two linear maps corresponding to each of the parents or as one quadrivalent map. QuadMap permutes the marker order for each iteration because there are multiple linear orders that are “correct,” depending on which starting markers are chosen. When MapMaker has finished its calculations, QuadMap stores the final map from the `order output file before repeating the process with a different permuted input file. The process is repeated until it reaches the maximum iterations cutoff specified by the user.

QuadMap then analyzes the maps from each run and summarizes its results in a tab-delimited text file that can be easily imported into a spreadsheet or word processing program. Frequencies, average intermarker centimorgan distances, and the centimorgan distance variance of each marker pair are stored in this file. QuadMap uses this information to determine which markers are binned together (i.e., on the same segment) and which are on different segments. The variance between markers on the same segment is expected to be low, whereas the variance between markers on different segments is up to an order of magnitude higher (Livingstone et al. 2000). The user specifies the cutoff variance, distance, and frequency used to separate the markers into their appropriate bins.

QuadMap does not determine maximum likelihood marker order and distances. For this study, we knew the marker order of our simulated data sets a priori because the Tq.exe program output is predictable. In real data sets, the final marker order could be determined using the QuadMap separated groups in conjunction with a program like MapMaker. Specifically, the MapMaker ripple command, which permutes the marker order and tests the likelihood of each map, can be used for this purpose. The map with the most support can then be drawn using a program like MapChart (R), the program used to generate the figures in this paper (http://www.plant.dlo.nl/default.asp?section=products& page=/products/mapping/right.htm) (Figures 1 and 2).

**Results**

When we constructed a single map of the translocation-affected data set without employing the QuadMap program, markers from both translocation and interstitial segments were present in the same linkage group, as expected (see Figure 1). Ideally, markers from two different chromosomes would not be associated with each other, as is the case when a simulation without a translocation is analyzed. However, due to the presence of the translocation, the linkage group was 414.4 cM in length and included all markers from all chromosomal segments.

We used QuadMap to analyze a simulated data set with no translocation. From the 1,000 permuted maps QuadMap generated, the variance in marker-pair distance for all marker pairs was 0.0 cM, and the average distance between marker pairs ranged from 3.9 to 190.7 cM. In the translocation-affected map, the variance ranged from 0.0 to 4,283.9 cM. The variance in marker-pair distance for markers on the same chromosome ranged from 0.0 to 2,987.9 cM (average = 814.4 cM), whereas the variance for markers on different chromosomes ranged from 173.6 to 4,283.9 cM (average = 1,062.0 cM). The variance in marker-pair distance for markers from the same segments, which would include four groups of markers (considering the two translocated segments and the two chromosomal segments), ranged from 0.0 to 2,649.0 cM (average = 138.4 cM), whereas the variance for markers from different segments ranged from 173.6 to 4,283.9 cM (average = 1,277.2 cM), as predicted by Livingstone et al. (2000). Once separated into the correct segments, the intermarker variance approaches that of the normal data set, and the markers from each group can be mapped using a program like MapMaker. These maps are shown in Figure 2.

To separate markers by chromosomal segment, QuadMap allows the user to set limits for average marker-pair
distance, variance in marker-pair distance, and marker-pair frequencies. Using our a priori knowledge of the marker order for this data set, we were able to separate the markers accurately by using a maximum centimorgan cutoff of 1,000 cM, a maximum variance cutoff of 172.0 cM, and a minimum frequency cutoff of 0.0. Ideally, it should be possible to segregate the markers by variance alone, as in this example. The centimorgan and frequency cutoffs should be used only to “refine” the marker segregation when variance alone is insufficient. Using these limits, which seem reasonable for most situations, the markers could be binned into the four groups shown in Figure 2. Three markers were not grouped with any segment due to high variance. These markers included T2L4 and T2L5, both right at the boundary between chromosome 1 and chromosome 2, with positions L1–L5.

Figure 1. A linkage map of the translocation-affected data set constructed without employing the QuadMap program. Markers from both translocation and interstitial segments are present in the same map. Interstitial markers are labeled “C1” for chromosome 1 and “C2” for chromosome 2, with locus positions L6 through L20. Translocation segment markers are labeled “T1” for the translocated end of chromosome 1 and “T2” for chromosome 2, with positions L1-L5.

Figure 2. A linkage map of the translocation-affected data set constructed with the QuadMap program. Markers of a simulated translocation heterozygote are grouped according to chromosomal segment. We then use the MapMaker ripple command to find the maximum likelihood marker order and distances. Interstitial markers are labeled “C1” for chromosome 1 and “C2” for chromosome 2, with locus positions L6 through L20. Translocation segment markers are labeled “T1” for the translocated end of chromosome 1 and “T2” for chromosome 2, with positions L1–L5. Three markers were not grouped with any segment due to high variance. These markers included T2L4 and T2L5, both located at the translocation boundary.
the translocation and the interstitial segment. Although it appears that the current technique does not have the resolution, at least with the number of iterations analyzed here, to separate these markers into the proper chromosomal bins, this method of analysis is effective in distinguishing marker locations moving away from the breakpoint.

Discussion

The maps created for these two simulations behaved similarly to those described by Livingstone et al. (2000). The maps without translocations had low variances and comprised linkage groups of markers from a single chromosome. The translocation-affected maps had low variance between markers within a segment but high variance between markers on different segments. Livingstone et al. (2000) were able to find maps with all possible markers; therefore, pairwise comparisons of distances between all markers were possible. Because analysis of real-life data may not be as complete, QuadMap provides the user with other options to eliminate marker pairs that likely comprise markers from different segments. Specifically, such marker pairs may be present together in very few maps, providing little confidence in their location. Even if such infrequent marker pairs have low variance, QuadMap can still eliminate them by way of the frequency-cutoff parameter. Similarly, QuadMap uses the centimorgan-cutoff parameter to eliminate marker pairs that are too distant from each other to be confidently grouped to the same segment. These methods can cut down on the false-positive associations that plague pseudolinkage groups.

When using QuadMap with experimental rather than simulated data, we expect variances in marker order and distance to be higher due to marker misclassification, selection, and segregation ratio distortion. The choice of limits (i.e., cutoffs) for classification into groups is a balance between the need to associate as many markers into linkage groups as possible and the desire to prevent inaccurate, false-positive associations. We recommend making the decision as to what those limits should be based on the average variance and distance (or some other criteria specific to the application) prior to analysis to prevent the examination of data without a priori hypothesis testing in searching for significant results.

Gross cytological differences between parents can complicate mapping because of statistical association of markers from the different chromosomes involved in the rearrangement. The mapping of hexaploid oats provides a good example. In the Kanota x Ogle map linkage group 3, there is a large cluster of markers (O’Donoughue et al. 1996; Wight et al. 2003). When these markers were linked to chromosomal regions using monosomic F1 lines, Fox et al. (2001) found that some of the markers were from chromosome 17 and some were from chromosome 7C. When examined cytologically, Jellen et al. (1993, 1994, 1997) found that there was a reciprocal translocation in Ogle but not Kanota. Some may argue that such associations can be resolved by increasing marker coverage and population size. Although Wight et al. (2003) did so for the hexaploid oat map in addition to advancing the population to an F10:11, they were unable to resolve the complications in this linkage group. Other interspecific maps are likely to have similar problems, as has been seen in maps of Brassica napus (Osborn et al. 2004) and in the Solanaceae (Livingstone et al. 1999).

We created QuadMap to increase the accuracy of linkage mapping around genomic rearrangements. Accurate linkage maps are especially important for marker-assisted selection and map-based cloning because linkage relationships in the mapping population are the basis for further experimentation and investment of resources. If those associations are false, then it becomes difficult to make progress in selection or in cloning a gene of interest. The program may be of particular use when rearrangements are suspected but difficult or impossible to confirm cytologically with banding or hybrid meiotic pairing analysis. If a particular linkage group has large clusters of markers, it may be useful to map this region repeatedly, as facilitated by QuadMap, to diagnose the cluster either as a genomic rearrangement, pericentric marker clustering or as some other anomaly, although one need not have an a priori knowledge of the exact physical location of the anomaly to use the program. QuadMap, which requires a Windows (Microsoft) operating system and a functional MapMaker program, is available at http://pas.byu.edu/Faculty/enj/oatsite/QuadMap.zip.

Availability

This QuadMap software can be downloaded free of charge at http://pas.byu.edu/Faculty/enj/oatsite/QuadMap.zip.

Supplementary Data

A supplementary file is available at Journal of Heredity online (www.jhered.oxfordjournals.org).

Acknowledgments

We would like to thank Dr. Nick Tinker for providing Tq.exe, the program used to create to the simulated data sets.

References


Jellen EN, Rines HW, Fox SL, Davis DW, Phillips RL, and Gill BS, 1997. Characterization of ‘Sun II’ oat monosomics through C-banding and...


Received September 20, 2004
Accepted October 13, 2005

Corresponding Editor: Sudhir Kumar