TetraploidMap for Windows: Linkage Map Construction and QTL Mapping in Autotetraploid Species

CHRISTINE ANNE HACKETT, IAIN MILNE, JOHN E. BRADSHAW, AND ZEWEI LUO

From Biomathematics and Statistics Scotland, Scottish Crop Research Institute (SCRI), Invergowrie, Dundee DD2 5DA, UK (Hackett); SCRI, Invergowrie, Dundee DD2 5DA, UK (Milne and Bradshaw); and the School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK (Luo).

Address correspondence to Dr. C. A. Hackett at the address above, or e-mail: christine@bioss.ac.uk.

An earlier program, TetraploidMap, enabled linkage analysis to be performed for autotetraploid species, with a text-based input and output. The current program, TetraploidMap for Windows, is considerably enhanced, and now goes beyond linkage analysis to perform quantitative trait locus (QTL) interval mapping, with a range of models and thresholds assessed by permutation tests. A Windows-based interface facilitates data entry and exploration. TetraploidMap for Windows is freely available from the Web site of Bioinformatics and Statistics Scotland at http://www.bioss.ac.uk/ (user-friendly software).

Linkage analysis and QTL mapping in autopolyploid plant species are more complicated than in diploid species, and so methodology and software have developed more slowly. However, several important plant species are polyploid (including tetraploid potato and alfalfa, hexaploid sweetpotato, and octaploid sugarcane). Hackett and Luo (2003) developed TetraploidMap, a suite of Fortran 90 programs for linkage analysis in autotetraploid species, based on dominant or codominant marker information scored in 2 parents and their full-sib progeny. The linkage analysis is based on a simple model for inheritance in autotetraploids, with random pairing of the 4 homologous chromosomes to give 2 pairs of bivalents at meiosis, but includes a test for double reduction to assess the adequacy of this part of the model. A significant result here can occur for other reasons than double reduction, such as segregation distortion or genotyping problems. This software has proved suitable for linkage and QTL analysis in potato and alfalfa but would be unsuitable for linkage mapping in species with substantial double reduction.

A linkage map is usually constructed as a forerunner to identifying QTLs associated with important economic traits, such as yield or disease resistance. Theory for interval mapping of QTLs was extended from diploid to autotetraploid species by Hackett et al. (2001), and this has been applied to a potato population by Bradshaw, Bryan, et al. (2004), Bradshaw, Pande, et al. (2004), and Bryan et al. (2004). QTLs of large effect were identified for plant maturity, field resistance to late blight, and quantitative resistance to the white potato cyst nematode.

TetraploidMap for Windows extends TetraploidMap to include a choice of methods for analyzing phenotypic data, including simple analysis of variance (ANOVA) for each marker in turn, QTL interval mapping for a range of models, and permutation tests. In potato, interval mapping proved superior to simple ANOVA. There are also several enhancements to the linkage analysis routines, particularly the development of a user-friendly Windows-based interface for importing data, displaying and outputting results as linkage maps and QTL profile plots.

Methods

Linkage Map Estimation

The steps to estimate the linkage map are the same as for the original TetraploidMap: analysis of single marker segregation, clustering into linkage groups, estimation of recombination frequency between all pairs of markers within a linkage group, and ordering based on pairwise data. A further step, inference of marker phase, is required before any QTL interval mapping can be performed. The enhancements in TetraploidMap for Windows are described here.

The first step is to determine the most likely parental genotypes for each marker, given the segregation data. The most informative type among the configurations for a dominant marker is the simplex marker, present as a single copy in one parent and absent in the other, that is, AOOO × OOOO or vice versa. An enhancement of TetraploidMap for Windows is a cluster analysis of the simplex markers to identify which are linked in coupling. Alphabetic codes display this information at subsequent stages. Duplex markers (AAOO × OOOO) are expected to be linked to 2 simplex groups in coupling and to 2 in repulsion; a chi-square test shows significant duplex–simplex associations to identify homologous chromosomes. This aids the partitioning of markers.

The markers are grouped by cluster analysis using a similarity measure derived from the chi-square test for independent segregation. The coding of the simplex markers aids interpretation. The cluster analysis has been enhanced by a graphical interface to investigate changes in the marker...
partition by sliding a line representing the similarity threshold.

After partitioning, the markers in each group are ordered. A 2-point analysis calculates the recombination frequency and lod score for each pair of markers in each possible phase and identifies the phase with the highest likelihood among those with a recombination frequency less than 0.5. These 2-point data are then used to calculate the best order for the markers in the linkage group by optimizing a weighted least squares criterion (Stam 1993). As the simulated annealing algorithm used by TetraploidMap is very slow for large numbers of markers, 3 methods are now available: an initial ordering based on a seriation algorithm (Buotow and Chakravarti 1987), a ripple search, and simulated annealing. The first 2 of these enable a rapid exploration to see whether any markers have been wrongly allocated to a linkage group. Simulated annealing can then be used for a final ordering.

In order to proceed to a QTL analysis, the phases of the ordered markers need to be inferred. A panel of pairwise results shows the most likely phase for any pair of markers, together with their recombination frequency and lod score. The coding of the simplex markers also simplifies this step. Inferred phase information can be typed directly into the panel of results. With the phase information, the marker maps can be displayed as a single order or as 4 homologous chromosomes, and this display can be exported as a graphics file (.png) or as a text file.

Analysis of Phenotypic Data
The most important new feature of TetraploidMap for Windows is the inclusion of methods to analyze phenotypic traits, such as yield. The analysis of a phenotypic trait begins with ANOVA for each simplex, duplex, and double-simplex (AOOO x AOOO) marker in turn to compare trait means with and without the marker. Significances for both the usual parametric ANOVA and the nonparametric Kruskal-Wallis test are shown. This identifies those linkage groups most closely associated with a trait.

Once markers in a linkage group have been ordered and their phase has been determined, QTL interval mapping can be carried out for one or more traits. This considers the QTL alleles inherited from one parent only. The trait distribution is modeled as a normal mixture model with 6 components, corresponding to the 6 possible QTL genotypes inherited from that parent (Q_12, Q_13, Q_14, Q_23, Q_24, and Q_34), where Q_i is the QTL genotype of an individual inheriting homologous chromosomes i and j from its parent. This model is fitted for putative QTL locations at 2-cM intervals along the linkage group. The QTL genotypes cannot be observed, but their probabilities can be calculated, conditional on the marker phenotypes in that linkage group. Details of this are given in Hackett et al. (2001). The likelihood of a QTL at any location is expressed as the lod score, the log10 of the ratio of the likelihood of the data given a QTL at this point to the likelihood of no QTL. TetraploidMap for Windows displays a plot of the lod score against position on the linkage group, together with estimates of the means and standard errors associated with each QTL genotype at the position corresponding to the maximum lod. The program can perform a permutation test (Churchill and Doerge 1994) to assess the significance of the QTL.

The full model with 6 components can be compared against a range of simpler models. One possibility is that the QTL is controlled by a simplex allele, that is, the parental genotypes are QQQQ crossed with qqqq, so that QTL genotypes Q_12, Q_13, and Q_14 are all equal to Qq, and are associated with significantly different trait values from Q_23, Q_24, and Q_34, which are all equal to qq. Another possibility is that the QTL is controlled by a dominant duplex allele, that is, the parental genotypes are QQqq crossed with qqqq. In this case, Q_34 will be associated with qq and will be significantly different from the remaining 5 genotypes, which will be QQ or Qq. There are 4 such possible simplex models and 6 possible duplex models, and TetraploidMap for Windows compares the full model with each of these simpler ones, using a likelihood ratio test, to see if any of them is adequate to model the trait means. This comparison is initially based on the trait means at the maximum lod position for the full model. If a simpler model represents the data as well as the full model, the interval mapping can be repeated for the entire linkage group based on that model. QTL lod profiles for the full and simpler models can be exported as a graphics file (.png) or as a text file.

Availability
TetraploidMap for Windows is written in Java 1.5 and provides a seamless user interface to the underlying Fortran 90 analysis routines (which are available for Windows platforms only at this time). It is freely available from the Web site of Biomathematics and Statistics Scotland at http://www.bioss.ac.uk/(user-friendly software). The Web site provides a full installer for the software that will automatically ensure a suitable version of Java is downloaded and installed along with the application itself. A full user guide is supplied along with a test data set.

Funding
UK Biotechnology and Biological Science Research Council; (213/GAT09080) Scottish Executive Environment and Rural Affairs Department.

References
Bradshaw JE, Pande B, Bryan GJ, Hackett CA, McLean K, Stewart HE. 2004. Interval mapping of quantitative trait loci for resistance to late blight (Phytophthora infestans (Mont.) de Bary), height and maturity in a tetraploid...


**Corresponding Editor:** William Tracey