Mammalian Genome 12, 930–932 (2001).
DOI: 10.1007/s00335-001-1016-3

Mammalian Genome
Incorporating Mouse Genome

Map Manager QTX, cross-platform software for genetic mapping

Kenneth F. Manly, Robert H. Cudmore, Jr., Jane M. Meer

Cellular and Molecular Biology, Roswell Park Cancer Institute, Buffalo, New York 14263, USA

Received: 2 July 2001 / Accepted: 20 August 2001

Abstract. Map Manager QTX (QTX) is software for analysis of genetic mapping experiments in experimental plants and animals. It includes functions for mapping both Mendelian and quantitative trait loci. QTX is an enhanced version of Map Manager QT, rewritten with the aid of cross-platform libraries (XVT, Boulder Software Foundry, Inc.), which allow it to be compiled for multiple computer platforms. It currently is distributed for Microsoft Windows and Mac OS and is available at http://mapmgr.roswellpark.org/mmQTX.html.

Introduction

Much of the design for QTX, both analysis and user interface, derives from Map Manager QT (Manly and Olson 1999). Both programs offer similar data entry and display for Mendelian markers and quantitative traits. Both calculate two-point linkage and map distances for Mendelian markers, and offer single-locus association, simple interval mapping, and composite interval mapping for quantitative trait loci. This report will focus on differences between the two programs. These differences, described below, include additional methods for QTL analysis, improved methods for ordering marker loci and for defining marker maps, support for additional cross designs and mapping functions, and methods for defining and using subsets of data.

QTX is still under development. Current plans for further development and opportunity to comment on those plans are available at http://mapmgr.roswellpark.org/mmQTX.html.

Results and Discussion

QTL analysis methods. The development of QTX has removed some of the limitations of Map Manager QT. QTX supports crosses with dominant markers and offers a choice of three mapping functions, Haldane (Haldane 1919), Kosambi (Kosambi 1944), and Morgan (Crow 1990).

Estimation of missing marker data. The expected QTL effects of missing marker genotypes are estimated from the genotypes of adjacent markers by a Markov-chain method as described by Jiang and Zeng (1997). For efficiency, a set of conditional probabilities is calculated for each unknown genotype and stored with the genotype. These conditional probabilities can be combined, according to Bayes theorem, with expected Mendelian frequencies for each genotype to estimate a QTL effect either at the unknown marker or in the intervals flanking the unknown marker. For interval mapping, these conditional probabilities are calculated once and used repeatedly to estimate QTL effects at various positions in the intervals flanking an unknown.

This method extends to composite interval mapping. To implement composite interval mapping, Map Manager provides for each trait a list of loci and other covariates called the "background" list. The user chooses loci to add to this list according to their significance in single-locus association tests. Loci added to this list are used in multiple regression to (partially) represent the effect of a nearby quantitative trait locus while mapping another locus elsewhere in the genome. Marker loci in this background list are references to the original loci, which remain at the proper location in a chromosome where adjacent loci are available to support estimation of missing marker genotypes.

Traits as covariates. It is often useful to be able to correct traits for the effect of other continuous-valued traits, such as age or weight. To accomplish this, QTX allows traits to be included in a background list as well as marker loci. Traits in the background list are included as covariates in a multiple regression model. This allows correcting the trait of interest for any continuous independent variable.

Weighted least-squares regression. When recombinant inbred strains are used for QTL mapping, multiple individuals are available for each genotype, and these can be used to determine a trait variance for each line. Trait variances for strains may differ significantly, violating the prerequisite for least-squares regression. Weighted regression allows these variances to be used, as described in the Appendix. To use this feature, the user calculates means and variances for a phenotype and enters each set of values as a trait in QTX. In the dialog that initiates QTL mapping for the trait that represents phenotype means, QTX allows a second trait to be specified as having phenotype variances.

Marker analysis methods

Automatic construction of marker maps. A common problem at the beginning of a mapping project is to create ordered linkage groups from a group of unassigned, unordered loci. A new menu item called Make Linkage Groups performs this task. It starts a new chromosome with the most tightly linked pair of loci available and then uses Distribute (described below) to add as many other loci as possible. It repeats this procedure until no further pair of linked loci is available among the remaining loci. To allow for the possibility of handling large groups of new loci, Make Linkage Groups repeats the above procedure up to three times, at decreasing stringency levels. The first repetition uses a very stringent criterion for linkage, which will transfer a significant number of markers only in large data sets. This method is derived from the rapid chain delineation method described by Doerge (1996).

The method used by Make Linkage Groups is also available as a separate menu item called Distribute. This method is designed to move new loci from a chromosome (the source chromosome) into their best position in a group of chromosomes (the destination chromosomes) in which the loci are already ordered. The best

Correspondence to: K.F. Manly; E-mail: kenneth.manly@roswellpark.org

position for a locus is defined as the position at which its insertion causes the greatest increase in the sum of $LOD_{linkage}$ scores for adjacent loci.

The method is designed to scale well with the number of loci in the source chromosome and the number of loci in the destination chromosomes. To achieve this goal, Distribute builds two intermediate tables. The first includes all pairs of linked loci in the source and destination chromosomes and their $LOD_{linkage}$ score, where linkage is defined by the setting of the menu item Search & Linkage Criterion. With this table, Distribute then finds the best destination location for each locus in the source chromosome, as judged by ΔLOD, the increase in total $LOD_{linkage}$ score upon insertion at that location. Distribute stores these destination locations for each locus in a second table, the transfer table, sorted by ΔLOD. When the transfer table is complete, Distribute moves the first locus, the locus with the highest ΔLOD, to its destination. Before moving the next locus, Distribute updates the transfer table to account for the change of possible destination locations. This process continues until the transfer table is empty.

The complexity of Distribute and Make Linkage Groups was estimated by using Make Linkage Groups to order randomized loci from simulated datasets. The time to identify linkage groups and order loci varies approximately as $C^{1.6}L^{2.0}$, where C is the number of linkage groups and L is the total number of loci. Sorting 1280 loci into 20 linkage groups takes less than 2 min on a 400-mHz Apple G3 Powerbook.

User-defined marker maps. Data from advanced backcrosses, especially BC3, may have too little information to define distances between markers accurately. In other cases, too, marker distances may be available that are more accurate than those estimated from marker data in the mapping cross itself. QTX supports the use of fixed marker distances specified by the user. The map distances can be entered manually or imported from a text file, and they can be entered as raw recombination fractions or as map distances derived from the Haldane (Haldane 1919) or Kosambi (Kosambi 1944) mapping functions.

Segregation distortion. Segregation distortion caused by selection against an allele at a single locus does not affect classical tests of linkage or estimates of recombination fraction. Segregation distortion caused by selection at multiple loci (Wang et al. 1994), however, invalidates the classical methods. QTX has optional methods for linkage and distance calculations that are not affected by this type of segregation distortion (Garcia-Dorado and Gallego 1992; Lorieux et al. 1995a, 1995b). These methods apply to backcrosses, recombinant inbred strains, and intercrosses with dominant alleles.

Advanced crosses. QTX supports advanced backcrosses and advanced intercrosses. Advanced backcrosses are created by successively crossing backcross progeny with the recurrent parent to yield BC2 progeny in the second generation and BC3 in the third (Tanksley and Nelson 1996). Advanced intercrosses are those in which F_2 intercross progeny are further intercrossed randomly in successive generations (Darvasi and Soller 1995). In QTX, the recombination fractions and map distances displayed for these crosses are standardized to correspond to those that would be observed in a single generation cross. Since recombination in successive generations is expected to be independent in these advanced crosses, QTX uses the Haldane mapping function to calculate map distances.

For advanced backcrosses, recombination fractions are standardized with

$$r = 1 - \sqrt{1 - 2f}$$

for BC2 progeny and with

$$r = 1 - \sqrt[3]{1 - 4f}$$

for BC3 progeny, where f is the observed recombination fraction. For advanced intercrosses, the relationship between standardized recombination fraction and observed recombination fraction is

$$f = \frac{(1 - (1 - r)^{(t-2)})(1 - 2r)}{2},$$

where t is the number of intercross generations ($t = 2$ for a standard intercross). Since this relationship does not provide an explicit solution for r, QTX calculates r by interpolation with a table of precalculated values.

For interval mapping of QTLs in advanced backcrosses, QTX uses expressions for the expected QTL genotypes given in the Appendix. Three mapping functions and two advanced generations (BC2 and BC3) are supported. For interval mapping of QTLs in advanced intercrosses, QTX treats the cross like a conventional intercross with an inflated rate of recombination. Although this is an approximation (Darvasi and Soller 1995; Jiang and Zeng 1997), it greatly simplifies calculations.

User interface

Progeny display orders and subsets. Progeny display orders are user-defined names that define the order in which progeny individuals are displayed in any window that displays progeny. A progeny order may also define a subset of the progeny, in which case only that subset is visible and available for analysis. A system-defined order that displays all progeny is always available. Map Manager QT (Manly and Olson 1999) also allowed the creation of progeny display orders. In that program, however, orders necessarily contained all progeny, and progeny orders could not be used to analyze subsets of the data.

Labels. Labels are named attributes that can be applied to chromosomes, loci, or progeny individuals. There are eight system-defined labels: Framework, Private, Tentative, Duplicate, Ours, Data Source A, Data Source B, and Data Source C. Users can define up to six labels with their own names. These are independent binary attributes in the sense that any combination of them can be either checked or unchecked for any chromosome, locus, or progeny individual. The primary function of labels is to control visibility of chromosomes, loci, or progeny individuals, through commands that allow hiding objects with a particular label. Since hidden loci or progeny are not included in analyses, labels allow easy analysis of subsets of the data.

The Framework label is unique in that it affects the operation of Ripple, a method for refinement of marker order. Ripple does not alter the order of two loci marked Framework, although it may alter the order of a Framework locus with respect to any locus not marked Framework.

Acknowledgments. The development of the software described here was first supported by a grant from the Rockefeller Foundation, which allowed the development of Map Manager XP. QTX was developed from Map Manager XP, with support from grant HG01656 from the National Human Genome Research Institute, by the addition of functions for mapping quantitative trait loci. We thank Rosemary Elliott for a critical reading of the manuscript.

References

Carbonell EA, Gerig TM, Balansard E, Asins MJ (1992) Interval mapping in the analysis of nonadditive quantitative trait loci. Biometrics 48, 305–315

Crow JF (1990) Mapping functions. Genetics 125, 669–671
Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141, 1199–1207
Doerge RW (1996) Constructing genetic maps by rapid chain delineation. J Agric Genomics 2.
Garcia-Dorado A, Gallego A (1992) On the use of the classical tests for detecting linkage. J Hered 83, 143–146
Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8, 299–309
Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69, 315–324
Jiang C, Zeng Z-B (1997) Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetica 101, 47–58
Knapp SJ, Bridges WC Jr, Birkes D (1990) Mapping quantitative trait loci using molecular marker linkage maps. Theor Appl Genet 79, 583–592
Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugenics 12, 172–175
Lorieux M, Goffinet B, Perrier X, Leon DGD, Lanaud C (1995a) Maximum-likelihood models for mapping genetic markers showing segregation distortion. I. Backcross populations. Theor Appl Genet 90, 73–80
Lorieux M, Perrier X, Goffinet B, Lanaud C, Leon DGD (1995b) Maximum-likelihood models for mapping genetic markers showing segregation distortion II. F2 populations. Theor Appl Genet 90, 81–89
Manly KF, Olson JM (1999) Overview of QTL mapping software and introduction to Map Manager QT. Mamm Genome 10, 327–334
Martinez O, Curnow RN (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85, 480–488
Moreno-Gonzalez J (1992) Genetic models to estimate additive and non-additive effects of marker-associated QTL using multiple regression techniques. Theor Appl Genet 85, 435–444
Tanksley SD, Nelson JC (1996). Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92, 191–203
Wang GL, Mackill DJ, Bonman JM, McCouch SR, Champoux MC, Nelson RJ (1994) RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics 136, 1421–1434

Appendix

Weighted regression. Regression coefficients are calculated by standard weighted regression:

$$b = [X' \cdot W \cdot X]^{-1} X' \cdot W \cdot Y$$

where b is the vector of regression coefficients, X is the matrix of marker genotypes, Y is the vector of trait values, and W is a diagonal matrix with the reciprocals of the trait variances on the diagonal. X' denotes the transpose of X, and $[X]^{-1}$ denotes the inverse of X.

The likelihood ratio statistic for weighted regression is calculated by:

$$LRS = \frac{(Y - X_0 b_0)' W (Y - X_0 b_0)}{(Y - X_1 b_1)' W (Y - X_1 b_1)}$$

where X_0 and b_0 are the marker genotypes and regression coefficient estimates for the null (or reduced) hypothesis, and X_1 and b_1 are the same for the full hypothesis.

Advanced backcrosses. The following expressions for expected QTL effect are straightforward extensions of similar expressions previously published for backcrosses and intercrosses (Knapp et al. 1990; Carbonell et al. 1992; Haley and Knott 1992; Martinez and Curnow 1992; Moreno-Gonzalez 1992; Jiang and Zeng 1997). These expressions are unique in that they provide for multiple mapping functions in a unified way. The model upon which these expressions are based is that of a QTL of unknown strength at a known (or assumed) position between two marker loci of known phenotype, where both the distance between the marker loci and the interference model are known (or assumed).

Table 1. Expected QTL effect for advanced backcross.

	A	H	B	C	D
A	−q3/q	(q3−q1)/(1−q)	–	(q3−q1)/(1−q)	–
H	(q3−q2)/(1−q)	±(q3−q1−q2−1)/(2n−q−2)	(q2−q3)/(1−q)	–	–
B	–	(q1−q3)/(1−q)	q3/q	–	(q1−q3)/(1−q)
C	(q3−q2)/(1−q)	–	–	±(q3−q1−q2−1)/(2n−q−2)	–
D	–	–	(q2−q3)/(1−q)	–	±(q3−q1−q2−1)/(2n−q−2)

Expected QTL genotype for a QTL at a hypothetical location between two flanking marker loci. The headings of the rows and columns are the phenotypes of the left and right flanking markers, respectively. Phenotype symbols are: A, maternal; H, heterozygote; B, paternal; C, dominant paternal; D, dominant maternal. In the table, $q = (1 - R)^n$, $q1 = (1 - R1)^n$, $q2 = (1 - R2)^n$; $q3 = q$ for Morgan mapping function, $q3 = (1 - R - 2\,R\,R1\,R2)^n$ for Kosambi mapping function, $q3 = (1 - R - R1\,R2)^n$ for Haldane mapping function; n = number of backcross generations, either 2 or 3. R1 and R2 are the recombination fractions between the QTL and the flanking markers, and R is the recombination fraction between the flanking markers.