MCQTL: multi-allelic QTL mapping in multi-cross design

Marie-Françoise Jourjon, Sylvain Jasson, Jacques Marcel, Baba Ngom and Brigitte Mangin*

INRA, Unité de Biométrie et d’Intelligence Artificielle, B.P. 27, 31326 Castanet-Tolosan Cedex, France

Received on April 29, 2004; revised on June 28, 2004; accepted on July 27, 2004
Advance Access publication August 19, 2004

ABSTRACT

Summary: The aim of the MCQTL software package is to perform QTL mapping in multi-cross designs. It allows the analysis of the usual populations derived from inbred lines and can link the families by assuming that the QTL locations are the same in all of them. Moreover, a diallel modelling of the QTL genotypic effects is allowed in multiple related families.

The implemented model is a linear regression model. A composite interval mapping and an iterative QTL mapping are implemented to deal with multiple QTL models. Marker cofactor selections by forward or backward stepwise methods are implemented as well as computation of threshold test value by permutation.

Availability: The program is available on request after signing a licence agreement; free of charge for academic and non-profit organizations at http://www.genoplante.com (Bioinformatics products).

Contact: contactbioinf@genoplante.com

INTRODUCTION

Many QTL mapping computer programs are available to analyse data collected in experimental crosses. Most of them are designed to analyse line crosses, one family at once. They were reviewed in Manly and Olson (1999), except for the most recent one, R/qtl (Broman et al., 2003) and are listed in the web site http://linkage.rockefeller.edu/soft.

Combining different line crosses for mapping QTL has been shown by simulation studies to be powerful, whatever the choice of the QTL model: random effect, fixed intra-family effect or diallel modelling (Muranty, 1996; Xie et al., 1998; Xu, 1998; Rebai and Goffinet, 2000). Moreover, this strategy was successfully applied with mouse inbred lines (Hitzemann et al., 2002) and is current in dairy cattle (Khatkar et al., 2004). However, few mapping computer programs are available to perform QTL analysis in multiple families. QTL express (Seaton et al., 2002) permits analysis of complex pedigree as the half-sib design. It fits a general linear regression model and is used via a web server. INTERQRL (Jannink and Wu, 2003) and FlexQTL™(Bink et al., 2004) perform QTL mapping in a Bayesian framework.

MCQTL is a local UNIX application. It is designed to perform QTL mapping using linear regression model in multiple families of line crosses with fixed intra-family QTL effects or diallel modelling. An original method to automatically find a multiple QTL model is implemented. Obviously, the analysis of a single cross is also feasible.

METHODS

MCQTL package is comprised of three software applications. The first component, TranslateData reads data from MAPMAKER (Lincoln et al., 1993) like files. The second component, ProbaPop computes QTL genotype probabilities given marker information at each chromosome location for each family and stores them in XML formatted files. The last component, Multipop builds the pooled model using the genotype probabilities, computes Fisher test and estimates the model parameters. This three-step procedure is similar to the two-step procedure of Seaton et al. (2002) and therefore has the same flexibility and robustness. Indeed, users can add or drop a family and modify the model without re-computing the QTL genotype probabilities.

The intra-family model is a usual regression model (Haley and Knott, 1992) with genetic cofactors and a single QTL. Let \( c \) denote the cross between two parent lines \( i, j \), the phenotypic value \( Y_{ck} \) of the \( k \)-th individual is modelled by

\[
Y_{ck} = \mu_c + \sum_{l=1}^{L} \sum_{ij} p^l_{ck,ij} \theta_{c,ij} + \epsilon_{ck},
\]

where \( \mu_c \) is the global mean in the cross \( c \), \( L - 1 \) is the number of genetic cofactors, \( p^l_{ck,ij} \) is the probability of the \( k \)-th individual having genotype \( ij \) at the QTL or cofactor locus \( l \) given the marker information, \( \theta_{c,ij} \) is the mean of the \( ij \) genotype at locus \( l \) in cross \( c \) and \( \epsilon_{ck} \) the residual error.

*To whom correspondence should be addressed.
Two models are implemented to link multiple families. The first one assumes the same locations of cofactors and QTL for all the crosses but intra-family parameters. For the second one, $\theta_{c,ij}$ are assumed to be independent of the cross. This implies that additive allelic effects depend only on the parent lines. In both models, intra-family residual variances are assumed equal. In this current version, only an additive model is implemented.

A resampling method by permutation of the trait data is implemented to set chromosome or genome-wide significance.
thresholds of a single QTL model (Churchill and Doerge, 1994). The adaptation to multiple cross design is obtained by limiting permutations of the trait data to intra-family permutations. Two scan methods are implemented (i) a composite interval mapping (Zeng, 1993) that consists in dropping out the cofactors belonging to a chromosome while it is scanned and (ii) an iterative QTL mapping that aims to automatically find a multiple QTL model (Charcosset et al., 2001). A marker cofactor selection usually precedes the scan step. So, a forward stepwise method on the whole genome and (ii) an iterative QTL mapping that aims to automatically find a multiple QTL model (Charcosset et al., 2001). A marker cofactor selection usually precedes the scan step. A forward stepwise method chromosome by chromosome by limiting permutations of the trait data to intra-family permutations. Two scan methods are implemented (i) a composite interval mapping (Zeng, 1993) that consists in dropping out the cofactors belonging to a chromosome while it is scanned and (ii) an iterative QTL mapping that aims to automatically find a multiple QTL model (Charcosset et al., 2001). A marker cofactor selection usually precedes the scan step. So, a forward stepwise method on the whole genome and

**IMPLEMENTATION**

The TranslatedData application was developed in Java and runs using a 1.4 or later JVM. ProbaPop and MultiPop, being computationally intensive, are coded in C++. A Sun-OS and a Linux archive respectively tested on Sun OS 5.9 sparc Ultra-Enterprise and Linux 2.4.4 intel computers are available. GnuPlot 3.7 or latter version (available from http://www.gnuplot.info) is required.

**REFERENCES**


