BIOINFORMATICS
Electronic edition http://www.bioinformatics.oupjournals.org

VOLUME 19
NUMBER 7
MAY 1 2003
PAGES 889–890

DOI: 10.1093/bioinformatics/btg112

R/qtl: QTL mapping in experimental crosses

Karl W. Broman[1,*], *Hao Wu*[2], *Śaunak Sen*[2,†] *and Gary A. Churchill*[2]

[1]*Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe St, Baltimore, MD 21205, USA*

[2]*The Jackson Laboratory, 600 Main St, Bar Harbor, ME 04609, USA*

Received on October 11, 2002; revised on December 13, 2002; accepted on December 20, 2002.

*To whom correspondence should be addressed.

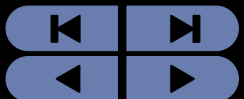
† Present address: Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94143, USA.

Abstract

Summary: *R/qtl is an extensible, interactive environment for mapping quantitative trait loci (QTLs) in experimental populations derived from inbred lines. It is implemented as an add-on package for the freely-available statistical software, R, and includes functions for estimating genetic maps, identifying genotyping errors, and performing single-QTL and two-dimensional, two-QTL genome scans by multiple methods, with the possible inclusion of covariates.*

Availability: *The package is freely available at http://www.biostat.jhsph.edu/~kbroman/qtl.*

Contact: *kbroman@jhsph.edu*



GO BACK
CLOSE FILE

Introduction

There exist numerous computer programs for QTL mapping in experimental crosses, including Mapmaker/QTL (Lander *et al.*, 1987) and Map Manager QTX (Manly *et al.*, 2001). Here we describe new QTL mapping software, R/qtl, implemented as an add-on package for the freely available statistical software, R (Ihaka and Gentleman, 1996).

R/qtl incorporates a more comprehensive set of methods than is currently available in any one package. The code is written so that new methods can be readily implemented.

Computationally intensive algorithms were coded in C, while the data manipulation and graphics functions were coded in the R language. R/qtl accepts input in a variety of formats and is available for Windows, Unix and MacOS.

Hidden Markov model technology

A key component of computational methods for QTL mapping is the hidden Markov model (HMM) technology (Baum *et al.*, 1970) for dealing with missing and partially missing genotype data. The core of R/qtl is a general implementation of the HMM technology for experimental crosses, with possible allowance for genotyping errors. Current specific implementations include backcrosses, intercrosses, and phase-known four-way crosses; the code may be extended for use with more complex crosses.

Features

R/qtl includes functions for identifying genotyping errors, visualizing genotyping data, identifying errors in marker order, and re-estimating inter-marker distances.

The user may perform single-QTL genome scans and two-dimensional, two-QTL genome scans, under a normal model, with the possible inclusion of covariates, by the EM algorithm (Dempster *et al.*, 1977; Lander and Botstein, 1989), Haley–Knott regression (Haley and Knott, 1992), and multiple imputation (Sen and Churchill, 2001). Further, R/qtl includes facilities for performing single-QTL genome scans by non-parametric interval mapping and binary trait mapping. Higher-order QTL models may be fit by multiple imputation. LOD thresholds may be estimated by permutation tests (Churchill and Doerge, 1994). Fig. 1 contains example graphs from R/qtl, for data on salt-induced hypertension in 250 backcross mice (Sugiyama *et al.*, 2001).

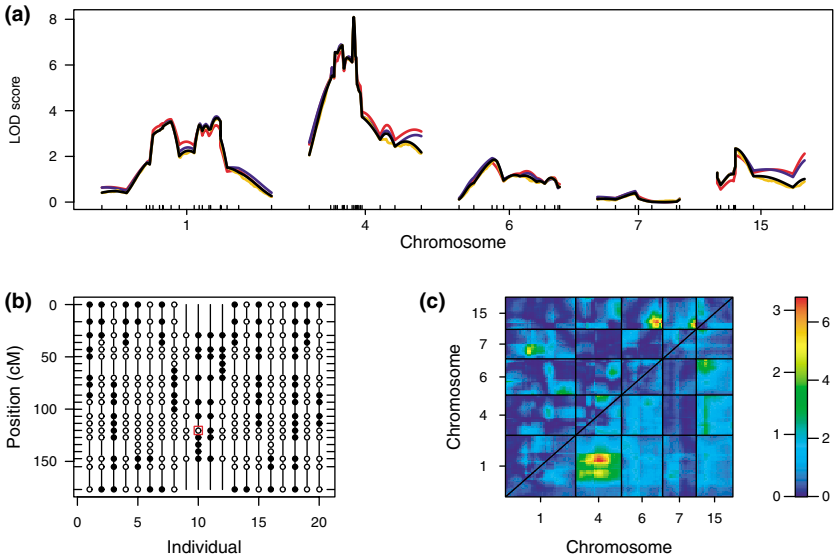


Fig. 1. Example graphs from R/qtl, based on the backcross data from Sugiyama *et al.* (2001). (**a**) LOD curves for selected chromosomes, calculated by standard interval mapping (black), Haley–Knott regression (blue), multiple imputation (orange), and non-parametric interval mapping (red). (**b**) Chromosome 1 genotype data for 20 individuals, with open and filled circles corresponding to homozygous and heterozygous genotypes, respectively; a possible genotyping error is flagged in red. (**c**) LOD scores for a two-QTL genome scan. Values below the diagonal correspond to a test of two versus one QTL; values above the diagonal correspond to a test for two-locus epistasis. In the color scale, the numbers to the right and left correspond to the values below and above the diagonal, respectively.

GO BACK
CLOSE FILE

Future development

R/qtl is under continual development. Our current efforts focus on the fit of higher-order QTL models by multiple interval mapping (Kao *et al.*, 1999), techniques for model comparison and model search for such multiple-QTL models, and the proper treatment of the X chromosome in QTL mapping.

Future plans include the coordinated analysis of multiple traits, analysis of recombinant inbred lines with random line effects, and analysis of multiple-QTL models for binary traits. Further, in collaboration with Kenneth F. Manly and colleagues at the Roswell Park Cancer Institute, we are developing a graphical user interface for R/qtl.

References

Baum,L.E., Petrie,T., Soules,G. and Weiss,N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.*, **41**, 164–171.

Churchill,G.A. and Doerge,R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics*, **138**, 963–971. MEDLINE Abstract

Dempster,A.P., Laird,N.M and Rubin,D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc.* B, **39**, 1–38.

Haley,C.S. and Knott,S.A. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69**, 315–324.

Ihaka,R. and Gentleman,R. (1996) R: a language for data analysis and graphics. *J. Comp. Graph. Stat.*, **5**, 299–314.

Kao,C.-H., Zeng,Z.-B. and Teasdale,R.D. (1999) Multiple interval mapping for quantitative trait loci. *Genetics*, **52**, 1203–1216.

Lander,E.S. and Botstein,D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199. MEDLINE Abstract

Lander,E.S., Green,P., Abrahamson,J., Barlow,A., Daly,M.J., Lincoln,S.E. and Newburg,L. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **1**, 174–181. MEDLINE Abstract

GO BACK

CLOSE FILE

Manly,K.F., Cudmore,Jr,R.H. and Meer,J.M. (2001) Map Manager QTX, cross-platform software for genetic mapping. *Mamm. Genome*, **12**, 930–932. MEDLINE Abstract

Sen,Ś. and Churchill,G.A. (2001) A statistical framework for quantitative trait mapping. *Genetics*, **159**, 371–387. MEDLINE Abstract

Sugiyama,F., Churchill,G.A., Higgins,D.C., Johns,C., Makaritsis,K.P., Gavras,H. and Paigen,B. (2001) Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics*, **71**, 70–77. MEDLINE Abstract