R/qtl: high-throughput multiple QTL mapping

Danny Arends1,*, Pjotr Prins1,2,†, Ritsert C. Jansen1 and Karl W. Broman3,a

1Groningen Bioinformatics Centre, University of Groningen, Groningen, 2Department of Nematology, Wageningen University, Wageningen, The Netherlands and 3Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, USA

ABSTRACT

Motivation: R/qtl is free and powerful software for mapping and exploring quantitative trait loci (QTL). R/qtl provides a fully comprehensive range of methods for a wide range of experimental cross types. We recently added multiple QTL mapping (MQM) to R/qtl. MQM adds higher statistical power to detect and disentangle the effects of multiple linked and unlinked QTL compared with many other methods. MQM for R/qtl adds many new features including improved handling of missing data, analysis of 10,000 s of molecular traits, permutation for determining significance thresholds for QTL and QTL hot spots, and visualizations for cis–trans and QTL interaction effects. MQM for R/qtl is the first free and open source implementation of MQM that is multi-platform, scalable and suitable for automated procedures and large genetical genomics datasets.

Availability: R/qtl is free and open source multi-platform software for the statistical language R, and is made available under the GPLv3 license. R/qtl can be installed from http://www.rqtl.org/. R/qtl queries should be directed at the mailing list, see http://www.rqtl.org/list/.

Contact: kbroman@biostat.wisc.edu

Received on June 16, 2010; revised on September 15, 2010; accepted on September 30, 2010

1 INTRODUCTION

R/qtl is an extensible, interactive environment for the mapping of quantitative trait loci (QTL) in experimental crosses. It is implemented as an add-on package for the freely available and widely used statistical language/software R (R Development Core Team, 2010). Since its introduction, R/qtl (Broman and Sen, 2009) has become a reference implementation with an extensive guide to QTL mapping (Broman and Knott, 1992) and explained in the ‘Handbook of Statistical Genetics’ (Jansen, 2007). MQM has one known commercial implementation (Van Osijen et al., 2002), which has been used effectively in practical research, resulting in hundreds of papers (e.g. in mouse, plant, and fish, respectively (de Mooij-van Malsen et al., 2008; Yandell et al., 2007)). MQM provides a practical, relevant and sensitive approach for mapping QTL in experimental populations. The theoretical framework of MQM was introduced and explored by one of us (Jansen, 1994) and explained in the ‘Handbook of Statistical Genetics’ (Jansen, 2007). MQM has one known commercial implementation (Van Osijen et al., 2002), which has been used effectively in practical research, resulting in hundreds of papers (e.g. in mouse, plant, and fish, respectively (de Mooij-van Malsen et al., 2008; Yandell et al., 2007)). Now, with MQM for R/qtl, we present the first free and open source implementation of MQM, that is multi-platform, scalable and suitable for automated procedures and large datasets.

2 FEATURES

MQM for R/qtl is an automated three-stage procedure in which, in the first stage, missing genotype data are ‘augmented’. (In other words, rather than guessing one likely genotype, multiple genotypes are modelled with their estimated probabilities.) In the second stage, important marker cofactors are selected by multiple regression and backward elimination. In the third stage, a QTL is moved along the chromosomes using these preselected markers as cofactors. QTL are interval mapped using the most informative procedure for cases with large numbers of marker cofactors is included. The method lets users test different QTL models by elimination of non-significant cofactors. MQM for R/qtl brings the following advantages to QTL mapping: (1) higher power, as long as the QTL explain a reasonable amount of variation; (2) protection against over-fitting, because MQM fixes the residual variance from the full model, which allows the use of more cofactors than may be used in, for example, CIM (Zeng, 1994); (3) prevention of ‘phantom’ QTL detection (between two QTL in coupling phase); and (4) detection of negative QTL (QTL in repulsion phase).
MQM for R/qtl brings additional advantages to genetical genomics datasets with hundreds to millions of traits: (5) a pragmatic permutation strategy for controlling the FDR and prevention of locating false QTL hot spots, as discussed in Breitling et al. (2008). Marker data are permuted, while keeping the correlation structure in the trait data; (6) high-performance computing by scaling on multi-CPU computers, as well as clustered computers, by calculating phenotypes in parallel, through the message passing interface (MPI) of the SNOW package for R (Tierney et al., 2009); (7) visualizations for exploring interactions in a genomic circle plot (Fig. 1a) and cis- and trans-regulation (Fig. 1b).

A 40-page tutorial for MQM explores, both the automated procedure, and the manual procedure of adding and removing cofactors, in an Arabidopsis thaliana recombinant inbred line (RIL) metabolite (mQTL) dataset with 24 metabolites as phenotypes (Fu et al., 2007). In addition, the tutorial visually explains the effects of data augmentation, cofactor selection, model selection and tweaking of input parameters, such as cofactor significance. Genetic interactions (epistasis) are explored through effect plots, and an example is given of parallel computation. The tutorial is part of the software distribution of R/qtl and is available online.

3 CONCLUSION
MQM for R/qtl is a significant addition to the QTL mapper’s toolbox. R/qtl provides the user with the most frequently used statistical analysis methods: single-marker analysis, interval mapping, Haley–Knott regression (Haley and Knott, 1992), CIM (Zeng, 1994) and MQM (Jansen, 1994). MQM has improved handling of missing data and allows more powerful and precise detection of QTL, compared with many other methods. Not only is this new implementation of MQM available in the statistical R environment, which allows scripting for pipe-lined setups, but it is also highly scalable through parallelization and paves the way for high-throughput QTL analysis. With MQM, R/qtl is a free and high-performance comprehensive QTL mapping toolbox for the analysis of experimental populations. R/qtl now includes permutation strategies for determining thresholds.
of significance relevant for QTL and QTL hot spots, the first step towards causal inference and network analysis.

Funding: National Institutes of Health (GM074244 to K.W.B.); the Netherlands Organisation for Scientific Research/TTI Green Genetics (1CC029RP to P.P.); the Netherlands Consortium of Systems Biology (NCSB), both of which are part of the Netherlands Genomics Initiative / Netherlands Organisation for Scientific Research [to DA]; the EU 7th Framework Programme under the Research Project PANACEA (222936 to R.J.).

Conflict of Interest: none declared.

REFERENCES