

Applications Note

R/qtlbim: QTL with Bayesian Interval Mapping in Experimental Crosses

Brian S. Yandell^{1,2,*}, Tapan Mehta³, Samprit Banerjee³, Daniel Shriner³, Ramprasad Venkataraman³, Jee Young Moon¹, W. Whipple Neely¹, Hao Wu⁴, Randy von Smith⁵ and Nengjun Yi³

¹Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706, USA; ²Department of Horticulture, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706, USA; ³Section on Statistical Genetics, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294, USA; ⁴Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205 USA; and ⁵The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA.

Associate Editor: Keith Crandall

ABSTRACT

Summary: R/qtlbim is an extensible, interactive environment for the Bayesian Interval Mapping of QTL, built on top of R/qtl (Broman *et al.* 2003), providing Bayesian analysis of multiple interacting quantitative trait loci (QTL) models for continuous, binary and ordinal traits in experimental crosses. It includes several efficient Markov chain Monte Carlo (MCMC) algorithms for evaluating the posterior of genetic architectures, *i.e.* the number and locations of QTL, their main and epistatic effects, and gene-environment interactions. R/qtlbim provides extensive informative graphical and numerical summaries, and model selection and convergence diagnostics of the MCMC output, illustrated through the vignette, example and demo capabilities of R (R Development Core Team 2006).

Availability: The package is freely available from cran.r-project.org.

Contact: byandell@wisc.edu or nyj@ms.soph.uab.edu

1 INTRODUCTION

The freely available QTL mapping package, R/qtlbim (www.qtlbim.org), provides a comprehensive framework for Bayesian model selection of the genetic architecture of complex traits in experimental crosses. Classical approaches to model selection in QTL mapping, such as multiple interval mapping as implemented in QTL Cartographer (Basten, Weir and Zeng 2002), largely rely on stepwise model selection with separate fits to each possible model. The Bayesian approach has the advantage of sampling across the more probable models, and providing graphical summaries that can compare many models at once. R/qtlbim can infer multiple QTL in the presence of epistasis (gene-gene interaction) and gene-environment interactions.

R/qtlbim is built on the widely used R/qtl framework (Broman *et al.* 2003), which provides many graphical tools for data checking and classical model selection. R/qtlbim shares the extensibility features of R/qtl. Computationally intensive algorithms are written in C, with data manipulation and graphics in R. R/qtlbim is available across Window, Linux and Mac OSX platforms and accepts a variety of input formats via R/qtl.

2 MCMC TECHNOLOGY

Central to R/qtlbim is the Markov chain Monte Carlo (MCMC) technology (Yi 2004; Yi *et al.* 2004; Yi *et al.* 2005). MCMC samples are drawn from the posterior distribution of genetic architecture, including number and location of genetic loci, gene action effects at all loci, epistatic interactions between pairs of loci, fixed and random covariates, and gene-environment interactions. These MCMC samples are then summarized and interpreted with graphs to infer key aspects of the genetic architecture.

MCMC sampling provides a mechanism to study the full Bayesian posterior distribution for any particular genetic architecture. Further, using a prior to model uncertainty of genetic architecture leads to MCMC samples across multiple genetic architectures. This brings model selection formally into a Bayesian framework in which the genetic architecture is just another parameter to be estimated.

3 FEATURES

R/qtlbim contains several efficient MCMC algorithms to search for genetic architectures that are most probable. It includes graphical and tabular summaries that assess the contribution of individual loci and pairs of loci while adjusting for effects of all other possible loci and covariates via model averaging.

Graphical tools allow the user to examine the MCMC samples by loci and genotypic effects in a variety of ways. They include estimation of Bayes factors for model selection on the number of QTL, the pattern of QTL across chromosomes, and the patterns of epistatic and gene-environment interactions. High posterior density (HPD) regions provide estimates of QTL analogous to LOD support intervals. Suggestions for model selection are provided in vignettes, examples and demos using the R (R Development Core Team 2006) interactive documentation system. The primary vignette, *qtlbim.pdf*, gives an overview of the package. Each command has a help page and corresponding example, which can be adapted to the user's own data. The command *demo(qtlbim)* leads the user through several demonstrations of the package capabilities.

The *scan.pdf* vignette provides more detail on R/qtlbim extensions of typical interval mapping scans. The R/qtl package

* To whom correspondence should be addressed.

(Broman *et al.* 2003) offers genome scans of the classical log odds (LOD) of Lander and Botstein (1989) and the Bayesian log posterior density (LPD) of Sen and Churchill (2001). These only consider the effect of one (*scanone*) or two (*scantwo*) QTL on a phenotype. R/qtlbim's scan routines, principally *qb.scanone* and *qb.scantwo*, use R/qtl's generic plot routines for 1-D and 2-D scans, respectively. However, our philosophy differs in important ways beyond changes in the calling sequence and our use of the *qb* prefix. Our scans consider the contribution of a given locus, or pair of loci, to the LPD after adjusting for all other possible QTL and covariates by model averaging over all genetic architectures that contain the QTL(s) being examined. These marginal scans are partitioned into contributions from main effects, epistatic interactions and gene-environment interactions. Another important distinction is that we provide functions to estimate marginal heritability, Bayes factors, means by genotypes, and other features in addition to LPD.

A third vignette, *hyperslide.pdf*, offers a way to automate model selection for genetic architecture of a complex trait. By default, it analyzes the hypertension data of Sugiyama *et al.* (2001) illustrated below. However, this vignette is an interactive object that can be reconfigured with another dataset using the *qb.sweave* command, via the Sweave framework (Leisch 2002). Sweave is an advanced feature using the LaTeX (www.latex-project.org) type-setting system, which must be separately installed.

An example of graphical output from R/qtlbim is provided in Figure 1 using the salt-induced hypertension data of Sugiyama *et al.* (2001) that is available in R/qtl. Note the improvement in LPD over R/qtl scanone by (1) adjusting for affects of all other possible QTL and (2) providing marginal evidence for epistasis.

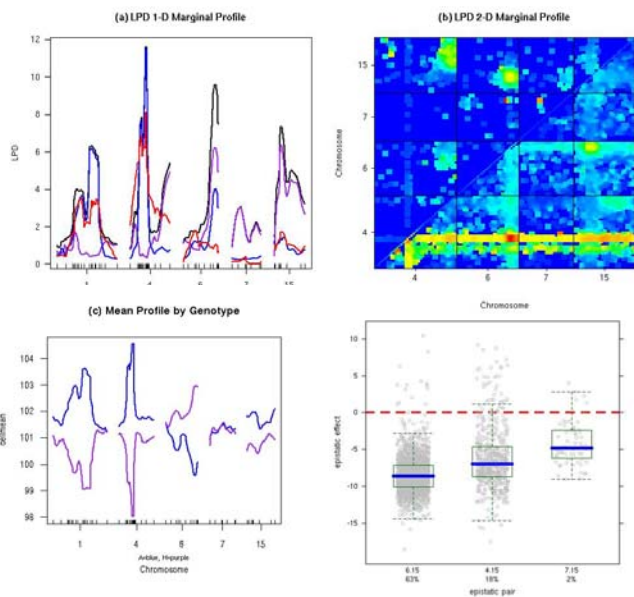


Figure 1. Example graphs from R/qtlbim based on the backcross of Sugiyama *et al.* (2001). (a) Marginal LPD one-dimensional

profile scan on chromosomes 1, 4, 6, 7, 15 (red = 1-QTL LPD scan from R/qtl, blue = LPD marginal scan for main effects, purple = LPD marginal scan for epistasis, black = LPD marginal for all effects of QTL); (b) marginal 2-D genome scan for LPD on chromosomes 4, 6, 7 and 15; (c) marginal means by genotype profiled across chromosomes; (d) epistatic effects for chromosome pairs 6.15 (chr 6 and 15, with 67% of MCMC samples included this pair), 4.15 (19%) and 7.15 (2%).

4 FUTURE DEVELOPMENT

R/qtlbim is under continual development. Future plans include proper treatment of the X chromosome (Broman *et al.* 2006) and extensions to correlated traits and experimental crosses derived from multiple inbred lines and outbred populations. We are also investigating ways to assess false discovery and check the fit of a model to data and prior assumptions. More extensive graphics for gene-environment interactions and for ordinal traits are planned. We intend to build on the graphical user interface for R/qtl that is under development, and we are in close communication with the R/qtl development team.

ACKNOWLEDGEMENTS

Special thanks to Karl Broman, who provided useful feedback and coordinated R/qtl releases and changes with us. This work is supported by National Institutes of Health (NIH) Grants R01 GM069430 (NY). In addition, RVS has partial support from NIH GM070683; BSY has partial support from NIH/PA-02-110, NIH/NIDDK 5803701 and NIH/NIDDK 66369-01; and NY has partial support from NIH HL80812, ES09912 and DK067487.

REFERENCES

Basten CJ, Weir BS, Zeng ZB (2002) QTL Cartographer, Version 1.16. Department of Statistics, North Carolina State University, Raleigh, NC.
 Broman KW, Sen S, Owens SE, Manichaikul A, Southard-Smith EM, Churchill GA (2006) The X chromosome in quantitative trait locus mapping. *Genetics* 00: 000-000. dx.doi.org/10.1534/genetics.106.061176
 Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19:889-890, www.rqtl.org.
 Leisch F (2002) Dynamic generation of statistical reports using literate data analysis. In W Härdle and B Rönz, editors, *Compstat 2002 - Proceedings in Computational Statistics*, pages 575-580. Physika Verlag, Heidelberg, Germany.
 Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185-199.
 R Development Core Team (2006) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-90051-07-0, www.R-project.org.
 Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. *Genetics* 159: 371-387.
 Sugiyama F, Churchill GA, Higgins DC, Johns C, Makaritsis KP, Gavras H, Paigen B (2001) Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics* 71: 70-77.
 Yi N (2004) A unified Markov chain Monte Carlo framework for mapping multiple quantitative trait loci. *Genetics* 167: 967-975.
 Yi N, Xu S, George V, Allison DB (2004) Mapping multiple quantitative trait loci for ordinal traits. *Behavior Genetics* 34: 3-15.
 Yi N, Yandell BS, Churchill GA, Allison DB, Eisen EJ, Pomp D (2005) Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics* 170: 1333-1344.