MAPfastR: Quantitative Trait Loci Mapping in Outbred Line Crosses

Ronald M. Nelson,[*,†,1] Carl Nettelblad,[‡] Mats E. Pettersson,[*,†] Xia Shen,[*,§] Lucy Crooks,[§,] Francois Besnier,[†] José M. Álvarez-Castro,[†,††,‡‡] Lars Rönnegård,[†,§§] Weronica Ek,[†] Zheya Sheng,[*,†,***] Marcin Kierczak,[*,†] Sverker Holmgren,[‡] and Örjan Carlborg[*,†]**

*Department of Clinical Sciences, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden, †Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-75651 Uppsala, Sweden, ‡Department of Information Technology, Uppsala University, SE-75105 Uppsala, Sweden, §Department of Cell and Molecular Biology, Uppsala University, BMC SE-75124 Uppsala, Sweden, **Welcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1HH, United Kingdom, ††Department of Genetics, University of Santiago de Compostela, ES-27002 Lugo, Galiza, Spain, ‡‡Instituto Gulbenkian de Ciência, Evolutionary Genetics Group, Oeiras, Portugal, §§Statistics Unit, Dalarna University, SE-791 88 Falun, Sweden, and ***State Key Laboratory for Agro-Biotechnology, China Agricultural University,100193, Beijing, People's Republic of China

ABSTRACT MAPfastR is a software package developed to analyze quantitative trait loci data from inbred and outbred line-crosses. The package includes a number of modules for fast and accurate quantitative trait loci analyses. It has been developed in the R language for fast and comprehensive analyses of large datasets. MAPfastR is freely available at: http://www.computationalgenetics.se/?page_id=7

KEYWORDS
QTL mapping
inbred line cross
outbred line cross
quantitative genetics

Quantitative trait loci (QTL) mapping is a valuable tool for unraveling the complex genetic architecture of phenotypic traits. A number of software packages currently are available for detecting QTL from marker data (for reviews, see Manly and Olson 1999; Durrant *et al.* 2011; Zhou *et al.* 2012). Most of the software were developed for analyses of various types of crosses between inbred lines, including backcrosses and F2 crosses (R/QTL, Broman 2003; Joehanes and Nelson 2008), multicross designs (heterogeneous stocks and collaborative crosses; Mott *et al.* 2000; Jourjon *et al.* 2005; Huang and George 2011), and advanced intercross lines (Peirce *et al.* 2008). Although we have written an extension to R/QTL that enables data from outbred lines to be analyzed (Nelson *et al.* 2011), the functionality is limited. There is some software designed for outbred populations and line crosses (*e.g.*, QxPak, Pérez-Enciso and Misztal 2004 and GridQTL, Seaton *et al.* 2006), but as they are several years old, and the algorithms they use are not able to handle the large amount of data produced by current SNP chip technology (Crooks *et al.* 2011).

MAPfastR is a fast and comprehensive software package for analyzing QTL data from outbred line-crosses that has been developed for flexible analyses of large datasets. MAPfastR is distinct from other packages in several ways. Notably, MAPfastR is based on a computationally efficient algorithm that uses all available data from dense SNP-chips (*i.e.*, tens to hundreds of thousands of markers, similar to association studies) and pedigree information (Crooks *et al.* 2011). MAPfastR provides functionality for F_2 crosses and backcrosses under the assumption that different QTL alleles are fixed in the founder lines (Crooks *et al.* 2011), line-cross analyses allowing for within-line segregation (flexible interclass analysis [FIA]; Rönnegård *et al.* 2008), and tests for epistatic interactions (Carlborg and Andersson 2002). In addition to the standard functionality, the software comes with add-on packages that allow more experienced users to take advantage of modules for analyses of deep (Advanced Intercross Line) pedigrees. MAPfastR includes an online developer and community-based support

Copyright © 2013 Nelson *et al.*
doi: 10.1534/g3.113.008623
Manuscript received May 31, 2013; accepted for publication September 26, 2013; published Early Online October 11, 2013.
This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Supporting information is available online at http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.113.008623/-/DC1.
[1]Corresponding author: Department of Clinical Sciences, Swedish University of Agricultural Sciences, Box 7078, SE-75007 Uppsala, Sweden.
E-mail: ronnnie.nelson@slu.se

system. MAPfastR is implemented in the R language (with optimization of the more computationally intensive algorithms in C++), accepts several standard input formats and is available for Windows, Unix, and Mac OS.

IMPLEMENTATION

MAPfastR integrates a number of published analytical tools by providing them within a comprehensive R package with a user-friendly interface and accompanying documentation. An outline of the analysis pipeline is shown in Figure 1. The main functions in the package are briefly described herein.

Data import

The first release (v 1.0) of MAPfastR supports two major input formats (CRI-MAP, Green *et al.* 1990 and triM, Crooks *et al.* 2011). All imported data are stored in a standardized R object, which consists of a list with two main components (phenotypic and pedigree data, and genotypic data) and several optional attributes (storing, for example, information on which is the heterogametic sex and which is the sex chromosome). A full description of the data object is provided in Supporting Information, File S1. Once the data import is completed, the internal data format can be used for outbred pedigree analyses by use of the main and supplementary analysis modules as well as additional custom analyses coded by the user. As output from the available functions is produced, it is appended to the data object, facilitating further analyses using the results. Support for more formats is in progress and will be provided in coming releases.

Least-squares QTL mapping

A module is provided to perform QTL mapping by least-squares regression (Haley and Knott 1992), where a user-selected phenotype is regressed onto genetic effect variables derived from genotype probabilities.

Calculation of QTL genotype probabilities using trim: The probabilities of alleles in the mapping population originating from each founder line are calculated using the triM algorithm (Nettelblad *et al.* 2009; Crooks *et al.* 2011). The algorithm uses a hidden Markov model to trace allele transmission in the pedigree. The line origins are calculated at user-defined, regular intervals along the chromosome and returned to the data object for further analysis.

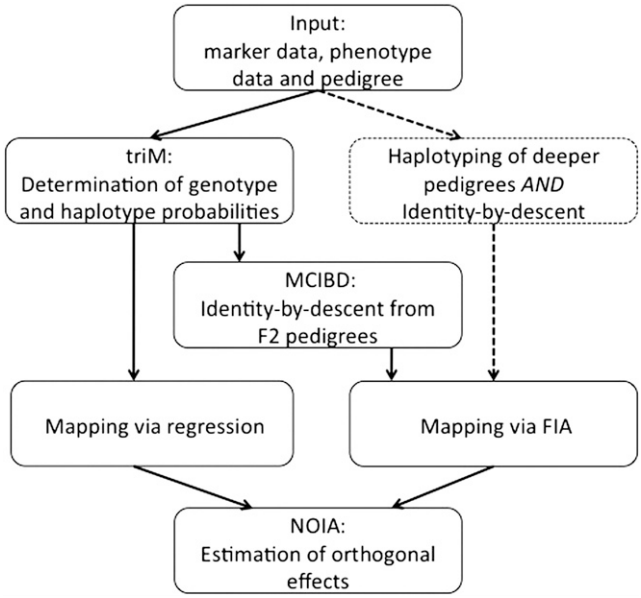

Figure 1 Outline of MAPfastR.

Regression analysis: Estimates of QTL effects are provided together with a plot of the test statistic from the fitted model across the genome, which illustrates the QTL locations. Analyses can be done on both autosomes and the homogametic sex chromosome, and for backcrosses and F2 crosses. Permutation testing can be performed by creating appropriately permuted datasets to derive an empirical significance threshold (Churchill and Doerge 1994).

FIA

FIA (Rönnegård *et al.* 2008) is an algorithm developed for analyses of outbred line-cross data where it is not reasonable to assume fixation of different QTL alleles in the founder lines. The analysis is performed by the following two steps.

IBD estimation using MCIBD: Identity by descent (IBD) matrices are estimated from the QTL genotype probabilities calculated from triM using the Monte Carlo Identity-By-Descent Matrix Estimation (MCIBD) algorithm (Shen *et al.* 2011). These matrices are used in the second step of the FIA analysis.

FIA: The variance-component-based FIA analysis scans the genome and provides estimates of genetic effects at regular spaced, user-defined locations in the genome as well as estimates of the likelihood that the QTL is fixed or segregating within the founder lines. The significance testing is based on a score-statistic and empirical significance thresholds are derived by permutation (Rönnegård *et al.* 2008).

Estimation of genetic effects using the Natural and Orthogonal InterAction model

The Natural and Orthogonal InterAction (NOIA) model is a unified model that ensures genetic effect estimates are orthogonal and enables effects to be translated from one population to another, aiding biological interpretation (Álvarez-Castro and Carlborg 2001). This allows users to estimate, for example, interaction effects that are comparable between populations and construct high-order genotype-phenotype maps for further analyses of interactions (*e.g.*, Álvarez-Castro and Rouzic 2008; Le Rouzic *et al.* 2008; Le Rouzic and Álvarez-Castro 2008).

Variance-component-based analysis of deep intercross pedigrees

An external module for performing analyses of deep pedigrees is provided as an unsupported add-on function for advanced users (Besnier *et al.* 2011). When this module is used, individuals from Advanced Intercross Lines generated from outbred founders can be haplotyped and an IBD matrix created that can be used to screen the genome for QTL using the FIA module for variance-component-based analysis.

RESULTS

Each of the functions has been extensively tested during development (Crooks *et al.* 2011; Ek *et al.* 2012; Shen *et al.* 2012). The complete pipeline has also been thoroughly tested to ensure that the package performs as a whole. Sample code for a complete analysis of an outbred line-cross and the resulting output is available in the supplementary documentation and example files.

In conclusion, MAPfastR is a comprehensive, fast, and accurate software that is able to perform various methods for QTL mapping in outbred line-cross data. It can also be used for analyzing data from inbred line crosses, where the computational efficiency may be a benefit. Add-on functions for the analysis of deeper pedigrees are also provided for advanced users. MAPfastR is under ongoing development to extend and improve its functionality and is extensively documented, with support available through an online forum for community and developers alike (https://groups.google.com/d/forum/mapfastr).

ACKNOWLEDGMENTS

This work was supported by the Future Research Leader program of the Swedish Foundation for Strategic Research and a EURYI Award from the European Science Foundation to Ö.C.

LITERATURE CITED

Álvarez-Castro, J. M., and Ö. Carlborg, 2001 A unified model for functional and statistical epistasis and its application in quantitative trait Loci analysis. Genetics 176: 1151–1167.

Álvarez-Castro, J. M., and A. L. Rouzic, 2008 How to perform meaningful estimates of genetic effects. PLoS Genet. 4: e1000062.

Besnier, F., P. Wahlberg, L. Rönnegård, W. Ek, L. Andersson *et al.*, 2011 Fine mapping and replication of QTL in outbred chicken advanced intercross lines. Genet. Sel. Evol. 43: 3.

Broman, K. W., 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889–890.

Carlborg, Ö., and L. Andersson, 2002 Use of randomization testing to detect multiple epistatic QTLs. Genet. Res. 79: 175–184.

Churchill, G. A., and W. R. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics 138: 963–971.

Crooks, L., C. Nettelblad, and Ö. Carlborg, 2011 An improved method for estimating chromosomal line origin in QTL analysis of crosses between outbred lines. Genes. Genom. Genet. 1: 57–64.

Ek, W., A. S. Sahlqvist, L. Crooks, R. Sgonc, H. Dietrich *et al.*, 2012 Mapping QTL affecting a systemic sclerosis-like disorder in a cross between UCD-200 and red jungle fowl chickens. Dev. Comp. Immunol. 38: 352–359.

Durrant, C., M. A. Swertz, R. Alberts, D. Arends, S. Moller *et al.*, 2011 Bioinformatics tools and database resources for systems genetics analysis in mice–a short review and an evaluation of future needs. Brief. Bioinform. 13: 135–142.

Green, P., K. Fallsand, and S. Crooks, 1990 *CRI-MAP Documentation, version 2.4*, Washington University School of Medicine, St. Louis, MO.

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

Huang, B. E., and A. W. George, 2011 R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. Bioinformatics 27: 727–729.

Joehanes, R., and J. C. Nelson, 2008 QGene 4.0, an extensible Java QTL-analysis platform. Bioinformatics 24: 2788–2789.

Jourjon, M. F., S. Jasson, J. Marcel, B. Ngom, and B. Mangin, 2005 MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinformatics 21: 128–130.

Le Rouzic, A., and J. M. Álvarez-Castro, 2008 Estimation of genetic effects and genotype-phenotype maps. Evol. Bioinform. 4: 225–235.

Le Rouzic, A., J. M. Álvarez-Castro, and Ö. Carlborg, 2008 Dissection of the genetic architecture of body weight in chicken reveals the impact of epistasis on domestication traits. Genetics 179: 1591–1599.

Manly, K. F., and J. M. Olson, 1999 Overview of QTL mapping software and introduction to map manager QT. Mamm. Genome 10: 327–334.

Mott, R., C. J. Talbot, M. G. Turri, A. C. Collins, and J. Flint, 2000 A method for fine mapping quantitative trait loci in outbred animal stocks. Proc. Natl. Acad. Sci. USA 97: 12649–12654.

Nelson, R. M., X. Shen, and Ö. Carlborg, 2011 qtl.outbred: Interfacing outbred line cross data with the R/qtl mapping software. BMC Res. Notes 4: 154.

Nettelblad, C., S. Holmgren, and L. Crooks, 2009 cnF2freq: efficient determination of genotype and haplotype probabilities in outbred populations using markov models. Bioinform. Comp. Biol. 5462: 307–319.

Peirce, J. L., K. W. Broman, L. Lu, E. J. Chesler, G. Zhou *et al.* 2008 Genome Reshuffling for Advanced Intercross Permutation (GRAIP): simulation and permutation for advanced intercross population analysis. PloS One 3: e1977.

Pérez-Enciso, M., and I. Misztal, 2004 Qxpak: a versatile mixed model application for genetical genomics and QTL analyses. Bioinformatics 20: 2792–2798.

Rönnegård, L., F. Besnier, and Ö. Carlborg, 2008 An improved method for quantitative trait loci detection and identification of within-line segregation in F2 intercross designs. Genetics 178: 2315–2326.

Seaton, G., J. Hernandez, J. A. Grunchec, I. White, J. Allen *et al.*, 2006 GridQTL: a grid portal for QTL mapping of compute intensive datasets. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, August 13–18, 2006.

Shen, S., L. Rönnegard, and Ö. Carlborg, 2011 How to deal with genotype uncertainty in variance component quantitative trait loci analyses. Genet. Res. 93: 333–342.

Shen, X., M. Pettersson, L. Rönnegård, and Ö. Carlborg, 2012 Inheritance beyond plain heritability: variance-controlling genes in *Arabidopsis thaliana*. PLoS Genet. 8: e1002839.

Zhou, J. J., A. Ghazalpour, E. M. Sobel, J. S. Sinsheimer, and K. Lange, 2012 Quantitative trait loci association mapping by imputation of strain origins in multifounder crosses. Genetics 190: 459–473.

Communicating editor: R. B. Brem