Package Iqtl - Interactive tools for R/qtl

Danny Arends [1,*],*K. Joeri van der Velde [1] Morris A. Swertz [1,4,5] Karl W. Broman [3] and Ritsert C. Jansen [1]

[1]Groningen Bioinformatics Centre, University of Groningen, NL. [2]Department of Nematology, Wageningen University, NL. [3]Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, US. [4]Genomics Coordination Centre, University Medical Center Groningen & University of Groningen, NL. [5]European Bioinformatics Institute, Hinxton, UK.

Associate Editor: XXXXXXX

ABSTRACT

Motivation

QTL analysis has become more and more challanging with the advent of new -omics techniques. not only the analysis but also the handling and (pre-)processing and storage of data. R/qtl is becomming a de-facto standard in the field of mouse and plants genetic analysis. It contains many historical algorithms for qtl analysis and map (re)contruction. Building on this community supported package, we present Iqtl for R/qtl it is build around the existsing cross object structure, and aims to be compatible and complementairy to R/qtl. Adding some experimental features for further exploration, and a HT binairy link to an XGAP database system to reduce dataset transfer times and auto-formatting to the common data structures used in R/qtl.

Availability:

Iqtl is free and open source multi-platform software for the statistical language R, and is made available under the GPLv3 license. Iqtl can be installed from http://www.Iqtl.nl/download. Also it is possible to use a CRAN install in Rgui.

1 CONTACT:

Danny.Arends@Gmail.com; Iqtl queries should be directed at the mailing list, see http://www.Iqtl.nl/list.

BRIEF OVERVIEW

The Iqtl package consisit of three mayor components:

- Helperfunctions for R/qtl mapping
- HT Contrast QTL mapping
- QTL viewing/exploring

---

*to whom correspondence should be addressed

Helperfunctions for R/qtl mapping

Iqtl is a new package to add functionality to R/qtl (Broman *et al.*, 2003) (a free and open-source qtl mapping package for R(Team, 2010).
Additional Data Over new 500 Arabidopsis phenotypes, and pre-processed QTL data formatted in R/qtl format
Quality control Scripts for e.g. batch and outlier detection, and automatic trait normalization via the VGAM package
Causal inference a basic causal inference scheme is implemented in Iqtl, based on the conditional correlation structure in the resulting QTL data. This approach is very powerful, when a large number of individuals is present in a GWA/ GWL association study.
More in and output formats to R/qtl (Happy, Sif) Iqtl adds more input formats for R/qtl, now the loading of happy format datasets is possible via the *happytocross* function.
HT Binairy connection to Molgenis Also we allow data to be stored in a Molgenis database running an XGAP datamodel. Data inside the database can be retrieved via the XGAP binary format reducing datatransfer times by 33-75% compared to plain text files.
**Differential Correlation QTL analysis** Differential correlation analysis is provided in the Iqtl package to enable researchers to obtain additional information about QTLs for differential correlation changes. This allows more detailed investigation into trait regulation.

HT Contrast QTL mapping

HT Contrast QTL mapping is designed with big data in mind. It uses a C-implementation of a basic but powerfull multiple lineair regression approach. In the first phase of the algorithm investigates the genotypes and genetic map for High Speed Model selection and QTL mapping in the second phase. It's features/aims:

- Suited for any type of experimental cross using contrasts. In the first step the Genotype matrix is analysed to create 'contrasts' as seen in (Fig x), in the second step significance of each location (aka set of contrasts) is assesed using multiple lineair regression models using:
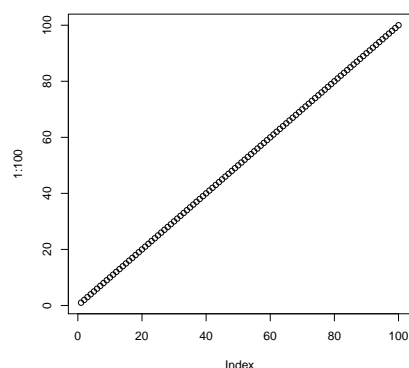
1. *Environmental cofactors* as described in Li *et al.* (2006).

2. *Genetic markers* as cofactors for more information see Jansen (1994) or Balding *et al.* (2007).

3. *Other traits* as cofactors in more complex multiple lineair model selection.

- Genomewide or marker specific contrasts. Thus giving the ability for sex chromosome handling, by allowing for more degrees of freedom at those markers.

- Model selection at multiple loci using multiple environmental cofactors. (See the list of cofactors above)

- Based on the MQM routine of Jansen (1994) as implemented in R/qtl (Arends *et al.*, 2010)

- Using SNOW to maximize the use of common multiple cores

QTL viewing/exploring

- JAVA HT qtl viewer
- JAVA Applet for qtl viewing
- SVG viewer to show small subsets online

CONCLUSIONS

We present here a package with additions to the QTL mappers toolbox in R. This includes new algorithms designed with big data and scalability in mind. It adds aims at providing a toolbox for pre-processing and post-processing as well as provide basic (but quick) analysis of QTL data from multiple origins and levels of quality.

ADDITIONAL FIGURES/SCREENSHOTS
REFERENCES

Arends, D. A., Prins, P., Jansen, R. C., and Broman, K. W. (2010). R/QTL: high throughput Multiple QTL Mapping. *Bioinformatics*, **??**(?), ???–???

Balding, D. J., Bishop, M., and Cannings, C., editors (2007). *Handbook of Statistical Genetics*. John Wiley & Sons, Ltd. ISBN: 978-0-470-05830-5.

Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19**(7), 889–890.

Jansen, R. C. (1994). Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics*, **138**(3), 871–881.

Li, Y., Alvarez, O., Gutteling, E. W., Tijsterman, M., J.Riksen, J. A., Hazendonk, E., Prins, P., Plasterk, R. H., Jansen, R. C., Breitling, R., and Kammenga, J. E. (2006). Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. *PLoS Genet*, **2**(12), e222.

Team, R. D. C. (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0.