fullsibQTL: an R package for QTL mapping in biparental populations of outcrossing species — Source link ⧉

Rodrigo Gazaffi, Rodrigo R. Amadeu, Marcelo Mollinari, João Ricardo Bachega Feijó Rosa ...+3 more authors

Institutions: Federal University of São Carlos, University of Florida, North Carolina State University, Centro de Tecnologia Canavieira ...+1 more institutions

Related papers:

- R Package wgaim: QTL Analysis in Bi-Parental Populations Using Linear Mixed Models

- vqtl: An R Package for Mean-Variance QTL Mapping.

- Ensemble Learning of QTL Models Improves Prediction of Complex Traits

- IBD-based QTL detection in inbred pedigrees: A case study of cereal breeding programs

- A model for quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny

1

fullsibQTL: an R package for QTL mapping in biparental populations of outcrossing species

Rodrigo Gazaffi [1,†]**, Rodrigo R. Amadeu** [2,†]**, Marcelo Mollinari** [3]**, João R. B. F. Rosa** [4]**, Cristiane H. Taniguti** [5]**, Gabriel R. A. Margarido** [5]**, and Antonio A. F. Garcia** [5,*]

[1]*Departamento de Biotecnologia e Produção Vegetal e Animal, Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras, São Paulo, Brazil;* [2]*Blueberry Breeding & Genomics Lab, Horticultural Sciences Department, University of Florida, Gainesville, FL, USA;* [3]*Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA;* [4]*Centro de Tecnologia Canavieira (CTC), Piracicaba, SP; and* [5]*Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brazil.*

Correspondence*:
Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, 13418-900, Brazil.
augusto.garcia@usp.br

†: These authors have contributed equally to this work

2 ABSTRACT

3 Accurate QTL mapping in outcrossing species requires software programs which consider
4 genetic features of these populations, such as markers with different segregation patterns and
5 different level of information. Although the available mapping procedures to date allow inferring
6 QTL position and effects, they are mostly not based on multilocus genetic maps. Having a QTL
7 analysis based in such maps is crucial since they allow informative markers to propagate their
8 information to less informative intervals of the map. We developed `fullsibQTL`, a novel and
9 freely available R package to perform composite interval QTL mapping considering outcrossing
10 populations and markers with different segregation patterns. It allows to estimate QTL position,
11 effects, segregation patterns, and linkage phase with flanking markers. Additionally, several
12 statistical and graphical tools are implemented, for straightforward analysis and interpretations.
13 `fullsibQTL` is an R open source package with C and R source code (GPLv3). It is multiplatform
14 and can be installed from `https://github.com/augusto-garcia/fullsibQTL`.

15 **Keywords: F1 population, fullsib, composite interval mapping, interval mapping, outcross**

1 INTRODUCTION

16 Mapping quantitative trait loci (QTL) unveils the genetic architecture of quantitative traits (*e.g.* yield, fitness)
17 and has several applications, such as breeding and evolutionary studies. In plant breeding, this information
18 can be integrated into breeding programs with marker-assisted selection tools aiming to understand and
19 boost such traits. Interval mapping (Lander and Botstein, 1989), composite interval mapping (Zeng, 1993,
20 1994), and multiple interval mapping (Kao et al., 1999) are some of the available approaches to perform
21 QTL mapping. In populations developed from inbred (homozygous) lines, such as F2s, backcrosses and
22 recombinant inbred lines, these mapping methods are well established and implemented in a plethora of
23 good software, for example: Broman et al. (2003); Van Ooijen (2004); Wang et al. (2012); Meng et al.

24 (2015). However, for outcrossing species without inbred lines (*e.g.* citrus, eucalyptus, sugarcane, loblolly
25 pine, and rubber tree), there is a lack of freely available software programs for QTL analysis. For biparental
26 populations derived from the cross between non-homozygous parents, it is necessary to estimate QTL
27 number, effects and position, as well their linkage phase with markers, and segregation patterns. This
28 effort relies on a linkage map that provides the phase of markers in both parents and their complete transmission
29 pattern to the offspring, which is usually thoroughly achieved by using multipoint approaches.

30 fullsibQTL is a free R package that performs composite interval mapping in full-sib progeny (F1
31 population) derived from a bi-parental cross between two non-homozygous parents. It requires a previously
32 estimated genetic map, considering markers with different segregation patterns, as described by (Wu
33 et al., 2002b). It implements the model proposed by Gazaffi et al. (2014), allowing to estimate the number
34 of QTLs, their position, effects, segregation patterns, and linkage phases with flanking markers. This
35 methodology has been successfully applied in different studies (Souza et al., 2013; Margarido et al., 2015;
36 Costa et al., 2016; Balsalobre et al., 2017; da Silva Pereira et al., 2017; Barreto et al., 2018; Conson et al.,
37 2018; Curtolo et al., 2020; Soratto et al., 2020), but currently there is no competing software with such
38 features.

2 FEATURES

39 fullsibQTL is a toolbox to perform QTL mapping based on composite interval mapping using the
40 model proposed by Gazaffi et al. (2014) for outcrossing species. It considers that a multipoint genetic
41 map was previously estimated with package OneMap, also freely available and widely used (Margarido
42 et al., 2007, 2020). Briefly, the model has three genetics effects, one additive effect for each parent and one
43 for the interaction of these effects (dominance). These parameters are estimated based on the maximum
44 likelihood for mixture models, with the expectation maximization algorithm. QTL genotypes probabilities
45 are calculated using a multipoint approach based on hidden Markov models (Wu et al., 2002a; Mollinari
46 et al., 2009). This allows the estimation using the information of all the markers and individuals, even if
47 there is missing data (Wu et al., 2002a).

48 The package also has a function to select cofactors using multiple linear regression through stepwise and
49 information criteria (R Core Team, 2017). Permutation tests (Churchill and Doerge, 1994) for threshold
50 determination can be implemented, combined with modifications suggested by Chen and Storey (2006)
51 for a more relaxed threshold. The model also can have other covariates. Moreover, several functions
52 were developed to provide graphical (Fig. 1) and text output. Additionally, interval mapping and inclusive
53 composite interval mapping options (Li et al., 2006) are also included. To illustrate the usage of the package,
54 a simulated dataset from Gazaffi et al. (2014) and its analysis are included in the package, with a full
55 tutorial and vignettes, explaining how to deal with partially informative markers.

56 With the results of QTL analysis, based on logarithm of the odds scores of hypothesis tests (Gazaffi et al.,
57 2014), users can visualize their genetic effects, statistical significance, segregation patterns and linkage
58 phase with flanking markers. Based on a least square approximation, fullsibQTL can also compute the
59 amount of phenotypic variation explained by each QTL in the model (the coefficient of determination, R^2).

3 CONCLUSION

60 Based on a sound QTL mapping methodology for outcrossing species, and on a reliable multipoint genetic
61 map, provides a QTL mapping toolbox to study the genetic architecture of quantitative traits in outcrossing
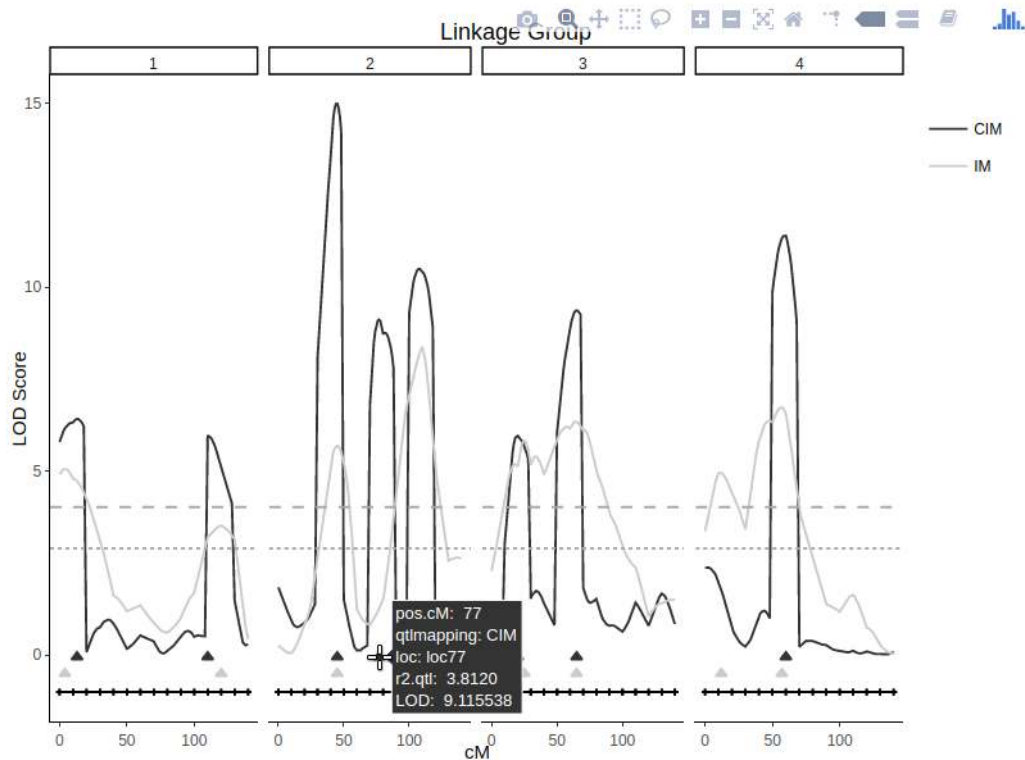
Figure 1. Screenshot of a `fullsibQTL` interactive `HTML+plotly` graphic. It provides a comparison of composite interval mapping and interval mapping approaches for a simulated population. The x-axis shows genetic map for each linkage group, and the y-axis the corresponding logarithm of the odds score for testing the presence of putative QTL. At the top right, there are buttons to zoom, box selection, etc. The mouse hover shows R^2, position, and logarithm of the odds scores.

62 species. It allows access to relevant information, and includes several graphical and analytical features,
63 providing an ideal environment for interaction with users. Its initial versions were extensively tested and
64 successfully applied several real world problems.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

71 Balsalobre, T. W. A., da Silva Pereira, G., Margarido, G. R. A., Gazaffi, R., Barreto, F. Z., Anoni, C. O.,
72   et al. (2017). Gbs-based single dosage markers for linkage and qtl mapping allow gene mining for
73   yield-related traits in sugarcane. *BMC Genomics* 18, 72. doi:10.1186/s12864-016-3383-x

74 Barreto, M. A., Rosa, J. R. B. F., Holanda, I. S. A., Cardoso-Silva, C. B., Vildoso, C. I. A., Ahnert, D.,
75   et al. (2018). Qtl mapping and identification of corresponding genomic regions for black pod disease
76   resistance to three phytophthora species in theobroma cacao l. *Euphytica* 214, 188. doi:10.1007/
77   s10681-018-2273-5

78 Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: Qtl mapping in experimental crosses.
79   *Bioinformatics* 19, 889–890. doi:10.1093/bioinformatics/btg112

80 Chen, L. and Storey, J. D. (2006). Relaxed significance criteria for linkage analysis. *Genetics* 173,
81   2371–2381. doi:10.1534/genetics.105.052506

82 Churchill, G. A. and Doerge, R. W. (1994). Empirical Threshold Values for Quantitative Trait Mapping.
83   *Genetics* 138, 963–971

84 Conson, A. R., Taniguti, C. H., Amadeu, R. R., Andreotti, I. A., de Souza, L. M., dos Santos, L. H.,
85   et al. (2018). High-resolution genetic map and qtl analysis of growth-related traits of hevea brasiliensis
86   cultivated under suboptimal temperature and humidity conditions. *Frontiers in Plant Science* 9, 1255.
87   doi:10.3389/fpls.2018.01255

88 Costa, E. A., Anoni, C. O., Mancini, M. C., Santos, F. R. C., Marconi, T. G., Gazaffi, R., et al. (2016).
89   Qtl mapping including codominant snp markers with ploidy level information in a sugarcane progeny.
90   *Euphytica* 211, 1–16. doi:10.1007/s10681-016-1746-7

91 Curtolo, M., Granato, L. A.-s. M., Soratto, T. A. T., Curtolo, M., Gazaffi, R., Takita, M. A. A., et al. (2020).
92   Expression Quantitative Trait Loci (eQTL) mapping for callose synthases in intergeneric hybrids of
93   Citrus challenged with the bacteria Candidatus Liberibacter asiaticusenglish. *Genetics and Molecular*
94   *Biology* 43

95 da Silva Pereira, G., Di Cassia Laperuta, L., Nunes, E. S., Chavarría, L., Pastina, M. M., Gazaffi, R., et al.
96   (2017). The sweet passion fruit (passiflora alata) crop: Genetic and phenotypic parameter estimates and
97   qtl mapping for fruit traits. *Tropical Plant Biology* 10, 18–29. doi:10.1007/s12042-016-9181-4

98 Gazaffi, R., Margarido, G. R. A., Pastina, M. M., Mollinari, M., and Garcia, A. A. F. (2014). A model for
99   quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny.
100   *Tree Genetics & Genomes* 10, 791–801. doi:10.1007/s11295-013-0664-2

101 Kao, C.-H., Zeng, Z.-B., and Teasdale, R. D. (1999). Multiple interval mapping for quantitative trait loci.
102   *Genetics* 152, 1203–1216

103 Lander, E. S. and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using rflp
104   linkage maps. *Genetics* 121, 185–199

105 Li, H., Ye, G., and Wang, J. (2006). A modified algorithm for the improvement of composite interval
106   mapping. *Genetics*

107 Margarido, G. R. A., Mollinari, M., Broman, K., Ferreira, G. C., Amadeu, R. R., Taniguti, C. H., et al.
108   (2020). *onemap: Software for constructing genetic maps in experimental crosses: full-sib, RILs, F2 and*
109   *backcrosses*. R package version 2.1.3

110 Margarido, G. R. A., Pastina, M. M., Souza, A. P., and Garcia, A. A. F. (2015). Multi-trait multi-
111   environment quantitative trait loci mapping for a sugarcane commercial cross provides insights on the
112   inheritance of important traits. *Molecular Breeding* 35, 175. doi:10.1007/s11032-015-0366-6

113 Margarido, G. R. A., Souza, A. P., and Garcia, A. A. F. (2007). Onemap: software for genetic mapping in
114   outcrossing species. *Hereditas* 144, 78–79. doi:10.1111/j.2007.0018-0661.02000.x

115 Meng, L., Li, H., Zhang, L., and Wang, J. (2015). Qtl icimapping: Integrated software for genetic linkage
116   map construction and quantitative trait locus mapping in biparental populations. *The Crop Journal* 3, 269
117   – 283. doi:http://dx.doi.org/10.1016/j.cj.2015.01.001. Special Issue: Breeding to Optimize Agriculture in
118   a Changing World

119 Mollinari, M., Margarido, G. R. A., Vencovsky, R., and Garcia, A. A. F. (2009). Evaluation of algorithms
120   used to order markers on genetic maps. *Heredity* 103, 494–502. doi:10.1038/hdy.2009.96

121 R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for
122   Statistical Computing, Vienna, Austria

123 Soratto, T. A. T., Curtolo, M., Marengo, S., Dezotti, A. L., Lima, R. P. M., Gazaffi, R., et al. (2020). QTL
124   and eQTL mapping associated with host response to Candidatus Liberibacter asiaticus in citrandarins.
125   *Tropical Plant Pathology* doi:10.1007/s40858-020-00372-7

126 Souza, L. M., Gazaffi, R., Mantello, C. C., Silva, C. C., Garcia, D., Le Guen, V., et al. (2013). Qtl mapping
127   of growth-related traits in a full-sib family of rubber tree (hevea brasiliensis) evaluated in a sub-tropical
128   climate. *PLOS ONE* 8, 1–10. doi:10.1371/journal.pone.0061238

129 Van Ooijen, J. (2004). *MapQTL® 5. Software for the mapping of quantitative trait loci in experimental*
130   *populations*. Kyazma BV, Wageningen

131 Wang, S., Basten, C. J., and Zeng, Z. B. (2012). *Windows QTL Cartographer 2.5*. Department of Statistics,
132   North Carolina State University, Raleigh, NC

133 Wu, R., Ma, C.-X., Painter, I., and Zeng, Z.-B. (2002a). Simultaneous maximum likelihood estimation
134   of linkage and linkage phases in outcrossing species. *Theoretical Population Biology* 61, 349 – 363.
135   doi:http://dx.doi.org/10.1006/tpbi.2002.1577

136 Wu, R., Ma, C. X., Wu, S. S., and Zeng, Z. B. (2002b). Linkage mapping of sex-specific differences.
137   *Genetical Research* 79, 85–96. doi:10.1017/S0016672301005389

138 Zeng, Z. B. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative
139   trait loci. *Proceedings of the National Academy of Sciences* 90, 10972–10976

140 Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468