BIOINFORMATICS APPLICATIONS NOTE
Vol. 28 no. 7 2012, pages 1042–1044
doi:10.1093/bioinformatics/bts049

Genetics and population analysis
Advance Access publication January 3, 2012

# xQTL workbench: a scalable web environment for multi-level QTL analysis

Danny Arends[1,†], K. Joeri van der Velde[1,†], Pjotr Prins[1,2], Karl W. Broman[3], Steffen Möller[4], Ritsert C. Jansen[1] and Morris A. Swertz[1,5,6,*]

[1]Groningen Bioinformatics Centre, University of Groningen, Groningen, [2]Laboratory of Nematology, Wageningen University, Wageningen, The Netherlands, [3]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA, [4]Institut für Neuro- und Bioinformatik, Universität zu Lübeck, [5]Genomics Coordination Centre, University Medical Centre Groningen, University of Groningen, The Netherlands and [6]EMBL-EBI, the European Bioinformatics Institute, Hinxton, UK

Associate Editor: Jeffrey Barrett

## ABSTRACT

**Summary:** *x*QTL workbench is a scalable web platform for the mapping of quantitative trait loci (QTLs) at multiple levels: for example gene expression (eQTL), protein abundance (pQTL), metabolite abundance (mQTL) and phenotype (phQTL) data. Popular QTL mapping methods for model organism and human populations are accessible via the web user interface. Large calculations scale easily on to multi-core computers, clusters and Cloud. All data involved can be uploaded and queried online: markers, genotypes, microarrays, NGS, LC-MS, GC-MS, NMR, etc. When new data types come available, *x*QTL workbench is quickly customized using the Molgenis software generator.

**Availability:** *x*QTL workbench runs on all common platforms, including Linux, Mac OS X and Windows. An online demo system, installation guide, tutorials, software and source code are available under the LGPL3 license from http://www.xqtl.org.

**Contact:** m.a.swertz@rug.nl

Received on September 30, 2011; revised on December 19, 2011; accepted on January 20, 2012

## 1 INTRODUCTION

Modern high-throughput technologies generate large amounts of genomic, transcriptomic, proteomic and metabolomic data. However, existing open source web-based tools for QTL analysis, such as webQTL (Wang *et al.*, 2003) and QTLNetwork (Yang *et al.*, 2008), are not easily extendable to different settings and computationally scalable for whole genome analyses. *x*QTL workbench makes it easy to analyse large and complex datasets using state-of-the-art QTL mapping tools and to apply these methods to millions of phenotypes using parallelized 'Big Data' solutions (Trelles *et al.*, 2011). *x*QTL workbench also supports storing of raw, intermediate and final result data, and analysis protocols and history for reproducibility and data provenance. Use of Molgenis (Swertz *et al.*, 2010a) helps to customize the software. All is conveniently accessible via standard Internet browsers on Windows, Linux or Mac (and using Java, R for the server).

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their option, the first two authors should be regarded as joint First Authors.

## 2 FEATURES

*x*QTL workbench provides visualization of QTL profiles, single and multiple QTL mapping methods, easy addition of new QTL analyses, scalable data management and analysis tracking.

### 2.1 Explore QTL profiles

Through the web interface, users can explore mapped QTLs, and underlying information by viewing QTL plots in an interactive scrollable and zoomable window. *x*QTL workbench has support for other common image formats, such as PNG, JPG, SVG and embedded postscript; useful for publishing scientific results online, and on paper. From the output, main database identifiers, such as gene, protein and/or metabolite identifiers are automatically linked-out to matching external web pages of public database such as NCBI, KEGG and Wormbase.

### 2.2 Single and multiple QTL mapping

*x*QTL workbench wraps R/qtl (Arends *et al.*, 2010; Broman *et al.*, 2003) in a web-based analysis framework offering all important QTL mapping routines, such as the EM algorithm, imputation, Haley-Knott regression, the extended Haley-Knott method, marker regression and Multiple QTL mapping. In addition, *x*QTL workbench includes R/qtlbim, a library that provides a Bayesian model selection approach for mapping multiple interacting QTL (Yandell *et al.*, 2007) and Plink, a library for association QTL mapping on single nucleotide polymorphisms (SNP) in natural populations (Purcell *et al.*, 2007).

### 2.3 Add new analysis tools

*x*QTL workbench supports flexible adding of more QTL analysis software: any R-based, or command-line tool, can be plugged in. All analysis results are uploaded, stored and tracked in the *x*QTL workbench database through an R-API. When new tools are added, they can build on the high-level multi-core computer, cluster and Cloud management functions, based on TORQUE/OpenPBS and BioNode (Prins *et al.*, 2011). *x*QTL workbench can be made part of a larger analysis pipeline using interfaces to R, Excel, REST and SOAP web services and Galaxy (Goecks *et al.*, 2010).

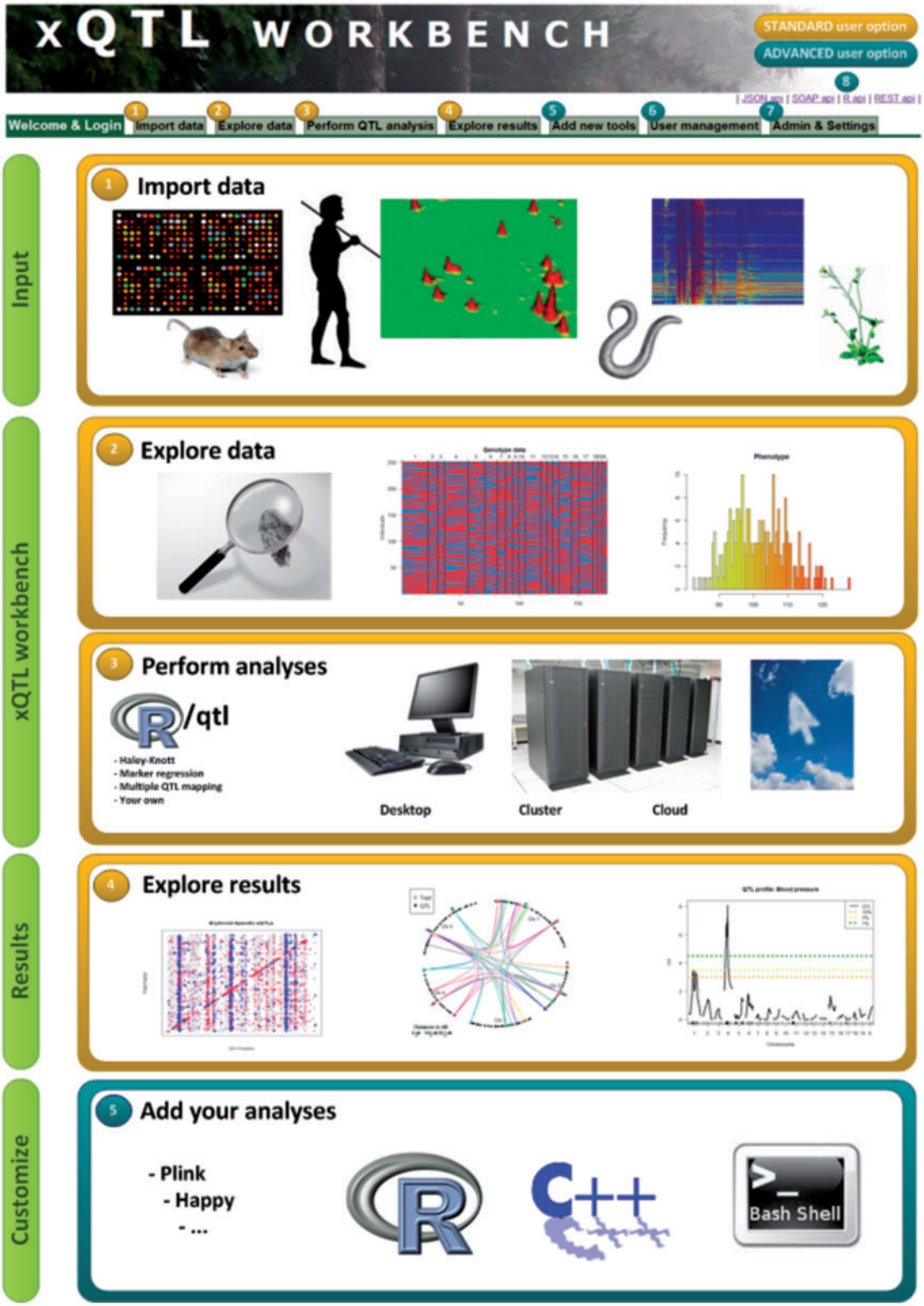


**Fig. 1.** Screenshot of xQTL workbench with all features enabled; (1) import phenotype, genotype and genetic map data, examples are given per import type; (2) search through the whole database, explore and browse your data using molgenis generated web-interfaces; (3) run R/qtl QTL mapping, the general plugin allows users to perform not only QTL mapping but also other analyze; (4) use default (or custom) plugins to explore results (e.g. Heatmaps, QTL profiles); (5) add new tools to the workbench (for Bio informaticians); (6) user management and access control of the system (Only for admins); (7) expert settings can be altered in the admin tab (Only for admins); (8) connect/share data using generated API's to R statistics, REST/JSON, SOAP.

### 2.4 Track analysis and monitor performance

When a new analysis protocol or R script is defined, this protocol can easily be applied to new data. Also, *x*QTL workbench keeps track of history. Re-use of analysis protocols can be done in an automated fashion. Previous analyses can be rerun without resetting parameters. *x*QTL workbench provides an online overview of past analyses e.g. which analyses were performed, by who, when and display settings applied.

### 2.5 Scalable data management

*x*QTL workbench has a consistency checking database based on XGAP specification (Swertz *et al.*, 2010b), user interfaces to manage and query genotype and phenotype datasets and support for various database back-ends including HSQL (standalone) and MySQL. Phenotype, genotype and genetic map data can be imported as text (TXT), comma separated (CSV) and Excel files. *x*QTL workbench handles and stores large data in a new and efficient binary edition of the XGAP format, named XGAPbin (extension .xbin), documented online. Such binary formats are essential when handling, storing and transporting multi-Gigabyte datasets.

### 2.6 Customizable to research needs

Additional modules for new data modalities can be added using Molgenis software generator (Swertz *et al.*, 2010b). The 'look and feel' of *x*QTL workbench is adaptable to institute or consortium style by changing a simple template, which is described in the *x*QTL workbench documentation enabling seamless integration into an existing website or intranet site, such as recently for EU-PANACEA model organism project and LifeLines biobank.

## 3 IMPLEMENTATION

We built *x*QTL workbench on top of Molgenis (Swertz *et al.*, 2004), a Java-based software to generate tailored research infrastructure on-demand (Swertz and Jansen, 2007). From a single 'blueprint' describing the whole system, Molgenis auto-generates a full application including user interface, database infrastructure, application programming interfaces in R, REST and SOAP (APIs). Molgenis' flexibility and robustness is proven by the wide range of research projects, e.g. the Nordic GWAS Control database (Leu *et al.*, 2010), EB mutation database (van den Akker *et al.*, 2011) and the Animal observation database (Swertz *et al.*, 2010a).

For data storage, the eXtensible Genotype and Phenotype (XGAP) data model was adopted (Swertz *et al.*, 2010b) and extended for big data. To support the increased demand for computational resources for included mapping routines, we added high-level cluster and cloud management functions for computation. The scalable QTL mapping routines of *x*QTL workbench are written in R and C. The choice of R ties in with the general practice of using R for QTL mapping. The user interface includes direct access to the R interpreter. Both *x*QTL workbench and Molgenis are open-source software, and source code is transparently stored and tracked in online source control repositories.

## 4 CONCLUSION

*x*QTL workbench provides a total solution for web-based analysis: major QTL mapping routines are integrated for use by experienced and inexperienced users. Researchers can upload raw data, run analyses, explore mapped QTL and underlying information, and link-out to important databases. New algorithms can be flexibly added, immediately available to all users. Large analyses can be easily executed on a cluster or in the Cloud. Future work include visualizations and search options to explore the results. We also had an EU-SYSGENET workshop that envisioned further integration of xQTL with analysis tools like HAPPY, databases like GeneNetwork, and the workflow manager TIQS (Durrant *et al.*, 2011).

## ACKNOWLEDGEMENTS

We thank Konrad Zych for Figure 1.

*Funding*: National Institutes of Health (GM074244 to KB); Netherlands Organisation for Scientific Research (NWO)/TTI Green Genetics (1CC029RP to P.P.); NWO (Rubicon 825.09.008 to M.A.S), Centre for BioSystems Genomics (CBSG), Netherlands Consortium of Systems Biology (NCSB) (to D.A.), Netherlands Bioinformatics Center (NBIC) (to M.A.S.), all part of Netherlands Genomics Initiative/NWO; Target/LifeLines co-funded by the European Regional Development Fund and NWO (to M.A.S.); and EU-FP7 Projects PANACEA (222936 to K.J.v.d.v.) and EURATRANS (241504 to R.C.J.).

*Conflict of Interest*: none declared.

## REFERENCES

Arends,D. *et al.* (2010) R/qtl: high throughput multiple QTL mapping. *Bioinformatics*, **26**, 2990–2992.

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

Durrant,C. *et al.* (2011) Bioinformatics tools and database resources for systems genetics analysis in miceña short review and an evaluation of future needs. *Brief. Bioinform.* http://bib.oxfordjournals.org/content/early/2011/07/08/bib.bbr026.full.

Goecks,J. *et al.* (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome. Biol.*, **11**, R86.

Leu,M. *et al.* (2010) Nordicdb: a nordic pool and portal for genome-wide control data. *Eur. J. Hum. Genet.*, **18**, 1322–1326.

Prins,P. *et al.* (2011) Scalable Computing in Evolutionary Genomics. In Anisimova,M. (ed.), Evolutionary Genomics: statistical and computational methods. Humana-Springer.

Purcell,S. *et al.* (2007) Plink: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.

Swertz,M.A. *et al.* (2004) Molecular Genetics Information System (MOLGENIS): alternatives in developing local experimental genomics databases. *Bioinformatics*, **20**, 2075–2083.

Swertz,M.A. *et al.* (2010a) The molgenis toolkit: rapid prototyping of biosoftware at the push of a button. *BMC Bioinformatics*, **11**(Suppl. 12), S12.

Swertz,M.A. *et al.* (2010b) Xgap: a uniform and extensible data model and software platform for genotype and phenotype experiments. *Genome. Biol.*, **11**, R27.

Swertz,M.A. and Jansen,R.C. (2007) Beyond standardization: dynamic software infrastructures for systems biology. *Natl Revl. Genetl.*, **8**, 235–243.

Trelles,O. *et al.* (2011) Big data, but are we ready? *Nat. Rev. Genet.*, **12**, 224–224.

van den Akker,P. *et al.* (2011) The international dystrophic epidermolysis bullosa patient registry: an online database of dystrophic epidermolysis bullosa patients and their col7a1 mutations. *Huml. Mutatl.*, **32**, 1100–1107.

Wang,J. *et al.* (2003) Webqtl. *Neuroinformatics*, **1**, 299–308.

Yandell,B.S. *et al.* (2007) R/qtlbim: QTL with Bayesian interval mapping in experimental crosses. *Bioinformatics*, **23**, 641–643.

Yang,J. *et al.* (2008) Qtlnetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics*, **24**, 721–723.