GRIDQTL: A GRID PORTAL FOR QTL MAPPING OF COMPUTE INTENSIVE DATASETS.

George Seaton[1], Jules Hernandez[1], Jean-Alain Grunchec[1], Ian White[1], John Allen[2], D.J. De Koning[3], Wenhua Wei[3], Dave Berry[2], Chris Haley[3], Sara Knott[1,4].

[1]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK.
[2]NeSC: National e-Science Centre, 15 South College St., Edinburgh EH8 9AA, UK.
[3]Roslin Institute, Roslin, Midlothian EH25 9PS, UK.
[4]Correspondence.

INTRODUCTION.

QTL mapping is an essential tool for the understanding of the genetic basis of complex traits, including health and production traits in farm animals and disorders and quantitative risk factors for disease in animal and human populations. QTL Express has provided a user-friendly and web-accessible analysis tool that has seen wide use for the analysis of experimental data, particularly from studies of outbred species. With the advent of microarrays, high-density multiple trait gene expression datasets are now prevalent. Paralleled with the availability of dense SNP marker maps for thousands of individuals this results in a computationally intensive and complex QTL mapping problem. The GridQTL project aims to provide an expanded and improved QTL analysis tool that harnesses Grid technologies to deal with greatly increased computational demands.

COMPUTING METHODS FOR QTL MAPPING

The first stage of the project has included the re-factoring of existing QTL Express Java servlets (http://qtl.cap.ed.ac.uk, Seaton et *al*. 2002), for simple single trait QTL analysis for one and two QTL allowing them to run in the new GridQTL portal environment (see below). Recent developments in QTL Express, such as the addition of a module for QTL analysis for general pedigrees (George *et al*., 2000) and a power calculator for general pedigrees will also be made available through GridQTL.

In conjunction with this, a suite of new programs is being developed that will be capable of: QTL mapping methods to simultaneously analyse multiple traits with pleiotropic models (with the ability to reduce the dimensionality of multi-trait information) for simple and complex pedigrees; new approaches to enable efficient QTL mapping of multi-trait data resultant from microarray experiments; methods and algorithms to detect 2 QTL with epistatic interactions of QTL across a range of pedigree structures; variance components methods to fine-map QTL by linkage disequilibrium analysis.

All applications written for the grid will be produced to a set of open standards thus allowing collaborators to produce third party QTL analysis modules that can be integrated into the GridQTL infrastructure.

PORTAL ENVIRONMENT AND GRID IMPLEMENTATION

An intuitive web-based user graphical interface will be implemented using the GridSphere portal project (http://www.gridsphere.org, Novotny *et al*., 2004) which will host JSR 168 compliant Java portlets designed specifically for QTL analysis job submission, job status querying, file management, data manipulation and result visualisation through a UK grid service (initially The National Grid Service, http://www.ngs.ac.uk).

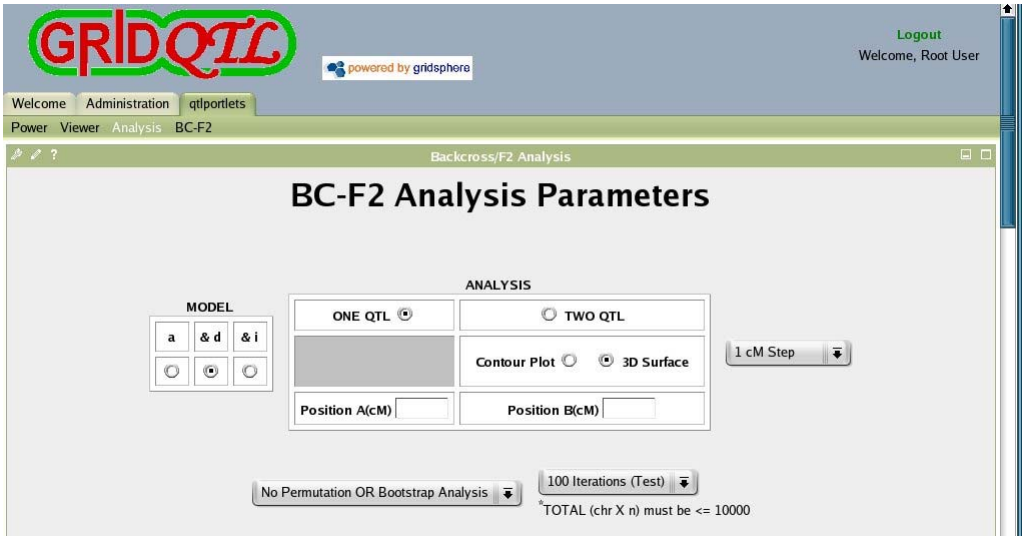

Figure 1. Example of GridQTL portal GUI using the GridSphere project.

Services made available by the Grid hardware via the GridQTL portal and grid middleware (Globus Toolkit, MyProxy *etc*) will include: the ability to use any grid system, ultimately including non-UK grids; scalable computing and storage power to cope with the increasingly complex datasets and QTL mapping methodologies; a secure persistent per-user data space via a single authenticated logon with options for data exchange; intelligent task scheduling and monitoring allowing optimum analysis performance; parallel computation and analysis re-start.

AVAILABILITY

Further information on the progress of the GridQTL portal project can be found at http://www.gridqtl.org.uk. It is intended that the GridQTL portal will be available as a public service for both academic and commercial communities.

REFERENCES

George, A. W., Visscher, P. M. and Haley, C. S. (2000) *Genetics* **156:** 2081-2092.
Novotny, J., Russell, M., Wehrens, O. (2004) *Euromicro* 2004: 412-419.
Seaton, G., Haley, C.S., Knott, S.A., Kearsey, M., Visscher, P.M. (2002) *Bioinformatics* 18: 339-340.