GridQTL

High Performance QTL Analysis via the Grid

www.gridqtl.org.uk

'Innovating through e-Science'

An Integrative Biology Project

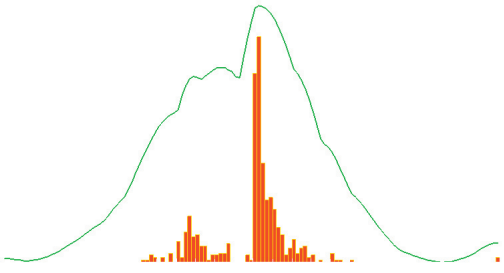
Aim: *Towards a more Predictive Biology* by the development, testing ,application and implementation of a Grid-based platform for **robust** and **fast multiple trait** mapping of **multiple Q**uantitative **T**rait **L**oci in **simple** and **complex** pedigrees.

Quantitative Trait Loci

The mapping of Quantitative Trait Loci (QTL) is the first step towards the identification of genes and causal polymorphisms that affect traits of importance in human medicine and agriculture, such as growth or blood pressure. Geneticists and Biologists use statistical models in order to estimate the position (or locus) of a QTL on a chromosome given phenotype, gene marker and pedigree data.

An Example - Obesity in Pigs:

Various populations of pigs are studied; for each pig the weight is measured and the type of each genetic marker of interest along a chromosome is noted. Statistical calculations are then carried out on this data to obtain probabilities that positions along the chromosome have an influence on the size of the weight.



Here, the plot of probability against position along a chromosome indicates the existence of a QTL that affects obesity in pigs; the most likely position of the QTL coincides with the highest peak on the graph.

QTLExpress and GridQTL

Fast, efficient and robust methods to map QTLs for single traits have already been developed by researchers at the Institute of Evolutionary Biology (IEB) at the University of Edinburgh and at the Roslin Institute near Edinburgh, two of the three collaboration groups in this project.

These methods are available via the publicly accessible web application *QTLExpress*:

http://qtl.cap.ed.ac.uk/.

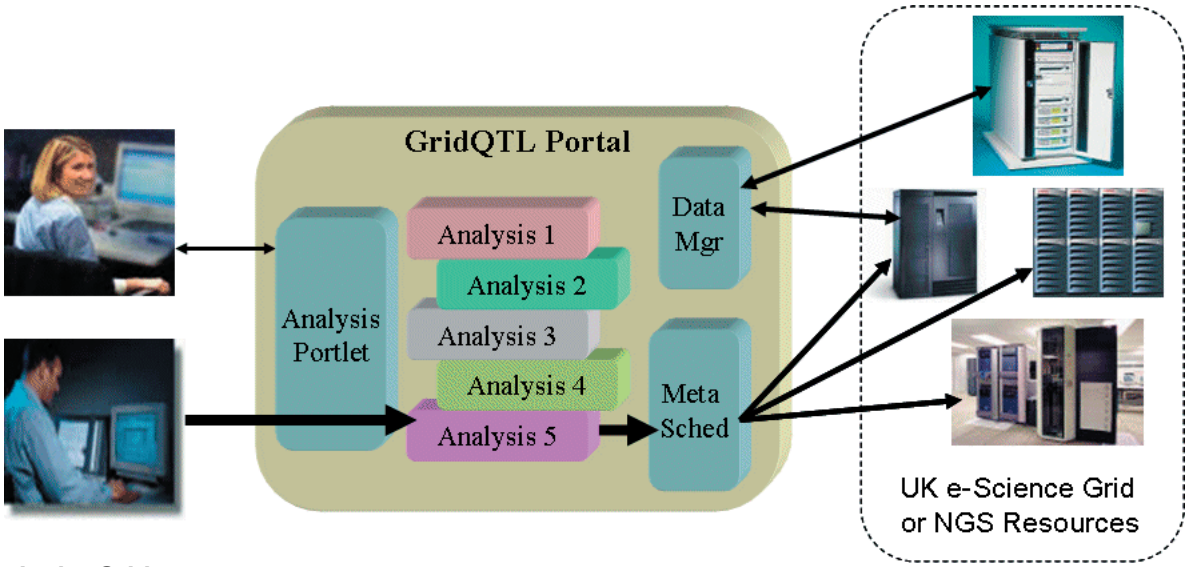
GridQTL will build on from this foundation to offer:

- Extended QTL mapping methods to enable simultaneous analyses of multiple traits, with pleiotropic models, in both simple and complex pedigrees.
- New methods to analyse the vast numbers of traits and genetic data made available from expression-QTL (eQTL) data resulting from microarray experiments.
- Methods and algorithms to detect two-QTL and higher order epistatic interactions of QTL in both simple and complex pedigrees.
- Fine-mapping of QTL positions by the use of Linkage Disequilibrium (LD) analysis.

GRIDQTL

bbsrc biotechnology and biological sciences research council

THE UNIVERSITY OF EDINBURGH



Access via the Grid

The new methods of QTL analysis are more complex, will use more samples, gene markers and data from more disparate sources, and will require greater computational power to run. This is only acheivable by the use of a Grid system that will provide the required on-demand computing and storage elements to successfully access and analyse the genetic data.

GridQTL Portal

Access to these services is via the GridQTL Web Portal shown pictorially above. It will offer the following services:

- Scalable computing and storage power to cope with the advanced QTL analysis methods.
- The ability to use any Grid system (NGS, ScotGrid, etc.).
- Execution of QTL analyses so that:
    - Parallel computation requirements are described.
    - Automatic task-level decomposition of analysis requests is performed.
    - Decomposed tasks are automatically and intelligently scheduled with job monitoring and re-start capabilities if required.
    - Notification of the outcome of all simulations is sent to the user.
- A secure and private data space for each researcher via a single logon so that:
    - Synchronisation of application input and output takes place.
    - Analysis re-start from intermediate results is available.

The GridQTL Portal will act as a gateway to a robust and growing public service for both research and industrial communities.

Project Collaborators

NeSC, National e-Science Centre, *Edinburgh*

Institute of Evolutionary Biology, *Edinburgh University*

Roslin Institute, *Edinburgh*

GridQTL will provide an essential core component of a future integrated biological system incorporating genetic, phenotypic, transcription and comparative information to allow prediction from gene sequence to consequence.