

## Introduction

A large number of diverse applications in multidisciplinary research environments require efficient, seamless, and easy-to-use environments to address computationally intensive problems. In this work we present an approach to build Grid-enabled problem solving environments (PSE), that allow end-users to operate within familiar settings and provide transparent access to computational Grid resources. The prototype integrates a system built on the open source statistical software system R with the Grid Job Management Framework (GJMF) which provides middleware-transparent access to Grid functionality, allowing our PSE to not be limited to using a specific middleware, but rather offload Grid issues to existing arbitration layers. The presented platform was created from already existing software components.

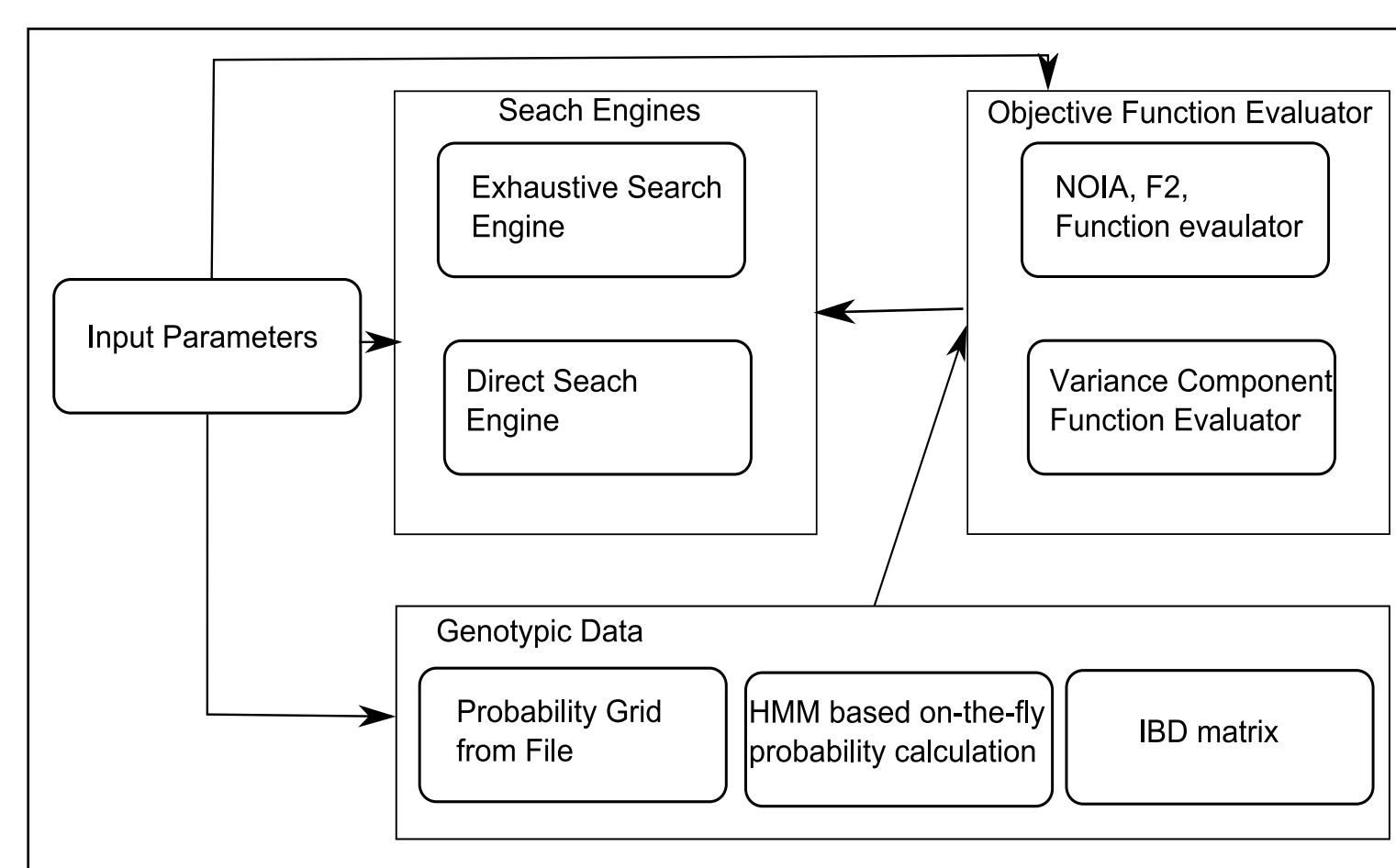
## QTL Analysis

A quantitative trait Locus (QTL) is a genomic location affecting traits like body weight, body length, total lifespan, etc. A trait can also have a molecular basis, like the concentration of some metabolite or protein in the tissue.

Generally, the immediate influence from one or several QTL is limited. As a consequence, the QTL searches all use statistical models of varying complexity to distinguish between environmental variation, and the variation due to the QTL. Just plotting or in some other way dissecting the phenotypic variance without reference to a specific genomic location, will frequently render what seems to be a single normal distribution, with no distinguishable components.

## R Software

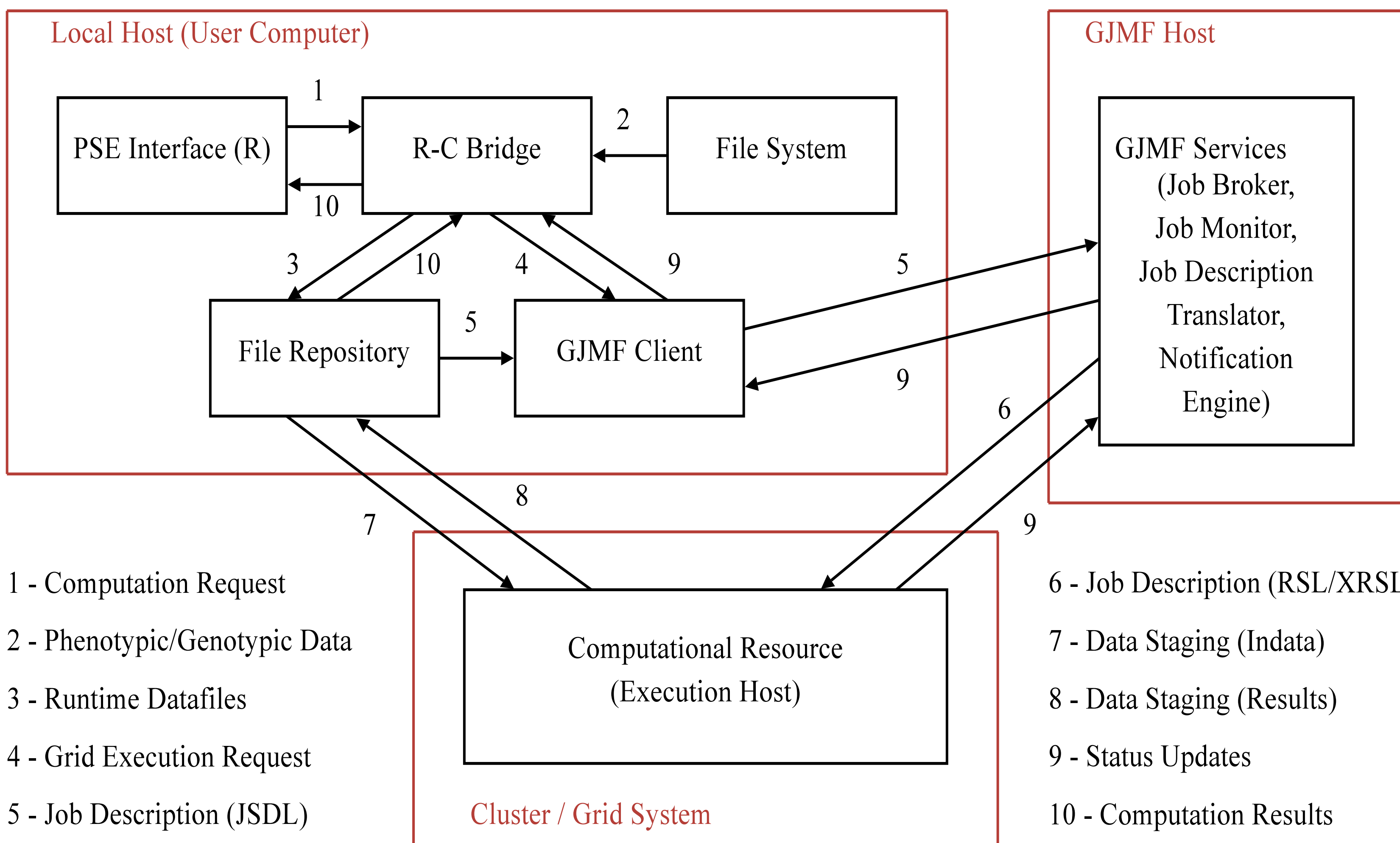
- ✓ A software package for statistical computing and graphics.
- ✓ Widely used by biologists for data analysis.
- ✓ Open source and supports different input data formats.



Internal architecture of QTL search

## Search Techniques

- ✓ Multi-dimensional model creates a large search space.
- ✓ Exhaustive search is not feasible.
- ✓ DIRECT algorithm based on Lipschitz optimization used instead.



- 1 - Computation Request
- 2 - Phenotypic/Genotypic Data
- 3 - Runtime Datafiles
- 4 - Grid Execution Request
- 5 - Job Description (JSDL)

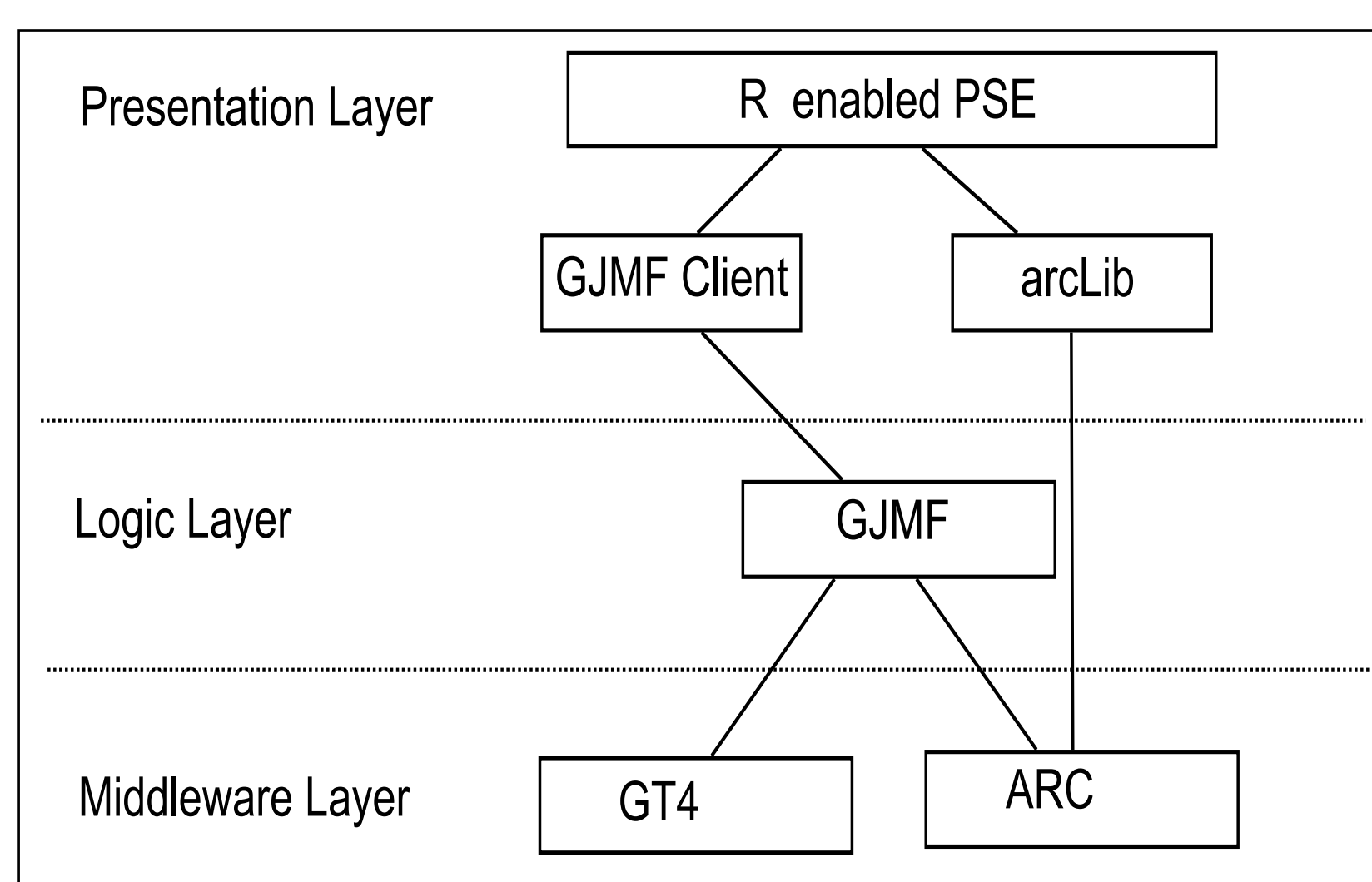
- 6 - Job Description (RSL/XRSL)
- 7 - Data Staging (Indata)
- 8 - Data Staging (Results)
- 9 - Status Updates
- 10 - Computation Results

Jobs (p)	d = 3	d = 4	d = 5
1	32	331	3749
2	24	221	2245
4	13	222	1628
8	10	142	1104
16	9	78	838
32	-	79	849
64	-	-	602

Table-1: QTL analysis complexity. Increase in the search dimensionality, increases the execution time.

## GJMF

- Grid Job Management Framework provides:
- ✓ Middleware independent job management.
  - ✓ Reliability.
  - ✓ Robustness.
  - ✓ Hierarchical architecture, composed of high level and low level services.



## Architecture

- ✓ Allows end-users with limited experience to access grid/HPC resources.
- ✓ In proposed PSE, both local and remote resources can be utilized for the advanced QTL analysis.
- ✓ Using the current GJMF setup, the grid jobs can be submitted to ARC and GT4 resources.
- ✓ GJMF's notification based job monitoring is used to determine the status of the submitted jobs.
- ✓ Proposed model allows scientists to access all the capabilities of R softwares with transparent access to the computational power of the grid environment.

## Results

- ✓ PSE enables biologists to exploit existing software within a grid framework.
- ✓ Table-2 shows the gain by merging the local and grid resources, especially for search in higher dimensions.
- ✓ Thanks to using serialization of program state, application code was kept mostly unchanged when transformed into a distributed grid application.
- ✓ In proposed PSE, both local and remote resources can be utilized for QTL analysis, with search state serialized for transfer onto grid.

Dim	Local runtime(s)	Grid runtime(s)	Grid Jobs (Speedup)
1	72	13	10 (6)
2	1765	353	10 (5)
3	26613	1558	20 (17)
4	352789	2565	207 (138)

Table-2: Randomization test locally and on the grid.

## References

- Related Paper <http://user.it.uu.se/~carln/realpaperV.pdf>  
 QTL Analysis <http://www.it.uu.se/research/project/ctrait> ; <http://www.computationalgenetics.se>  
 GJMF <http://www.gird.se/gjmf>

## Project Members

- Mahen Jayawardena (Uppsala Univ) [mahen.jayawardena@it.uu.se](mailto:mahen.jayawardena@it.uu.se) Per-Olov Östberg (Umeå Univ) [p-o@cs.umu.se](mailto:p-o@cs.umu.se)  
 Carl Nettelblad (Uppsala Univ) [carl.nettelblad@it.uu.se](mailto:carl.nettelblad@it.uu.se) Erik Elmroth (Umeå Univ) [elmroth@cs.umu.se](mailto:elmroth@cs.umu.se)  
 Salman Zubair Toor (Uppsala Univ) [salman.toor@it.uu.se](mailto:salman.toor@it.uu.se) Sverker Holmgren (Uppsala Univ) [sverker.holmgren@it.uu.se](mailto:sverker.holmgren@it.uu.se)