Germinate 2: Storage, Visualization and Analysis of High Volume Genotypic, Phenotypic and Pedigree Data

Paul Shaw[1], Iain Milne[1], Linda Cardle[1], Micha Bayer[1], Andy Flavell[2], Robbie Waugh[1], Gordon Stephen[1], Martin Graham[3], Jessie Kennedy[3] and **David Marshall**[1]

[1]Scottish Crop Research Institute, Invergowrie, Dundee, Scotland DD2 5DA.
[2]Division of Plant Sciences, University of Dundee at SCRI
[3]School of Computing, Edinburgh Napier University, Edinburgh, Scotland

Introduction

Germinate 2 is a generic plant data management platform designed to hold phenotypic, genotypic, pedigree and passport data and was developed at SCRI. We have recently added components to the system specifically developed and optimised to deal with high throughput genotyping technologies including Illumina SNP and DART, large scale field trial experiments and integration with R. Germinate allows the integration of different data types into a number of export formats suitable for external analysis tools as well as seemless connectivity with analysis tools developed at SCRI.
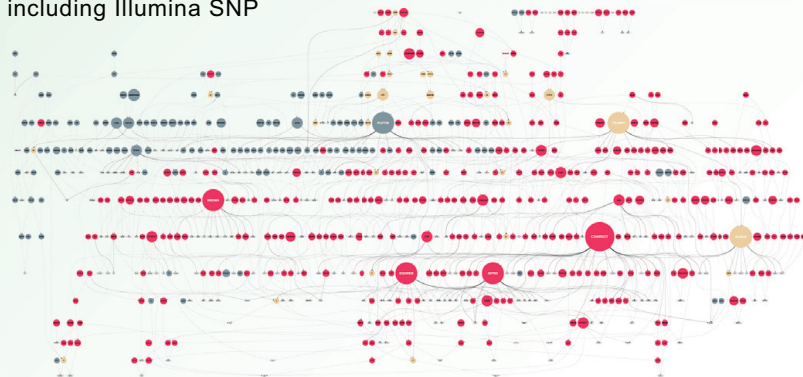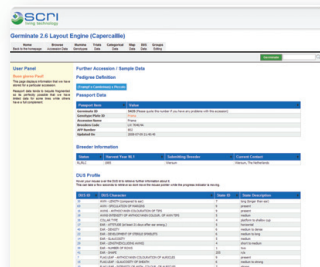

Pedigree Visualization
Representation of complex plant pedigrees are possible from the Germinate database.
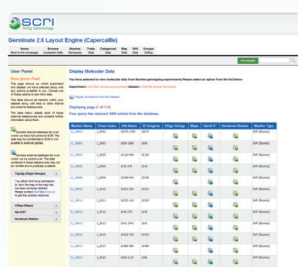
Data Pooling
Information sources for markers are brought together in one place including both internal and external data sources.


Line Passport and Additional Information
All information including annotations on a particular line are available on a single line profile page.

Data Export Options

Germinate facilitates the export of data in a number of formats. Data is grouped into experiments and datasets and data is exported using a marker and line grouping which allows the intersection of 2 groups to be exported. In this way complex groupings can be easily defined. The use of these logical groupings also means that if multiple datasets exist on the same genotyping plate they can be distinguised and treated separately for analysis.

Data can be exported to our genotype visualization application Flapjack, the comparative map viewer Strudel or any other application provided a suitable output format can be generated. The SCRI tools contain embedded information to allow additional data to be retrieved from the Germinate database from which it was exported.

Pedigree Visualization

We have implemented an integrated system using the freely available GraphViz libraries (http://www.graphviz.org/) to visualize plant pedigrees as directed acyclic graphs and allowing the overlaying of user defined data directly from the Germinate 2 database platform. In this way phenotypic of genotypic data can be overlayed on top of a pedigree layout diagram which may be helpful in the understanding of data or to gauge data coverage across defined pedigrees.

We hope to further improve this by adding error checking capabilities using the mathematical graph representation (DAG) to act as a scaffolding onto which inference or error checking algorithms can be constructed using Perl.
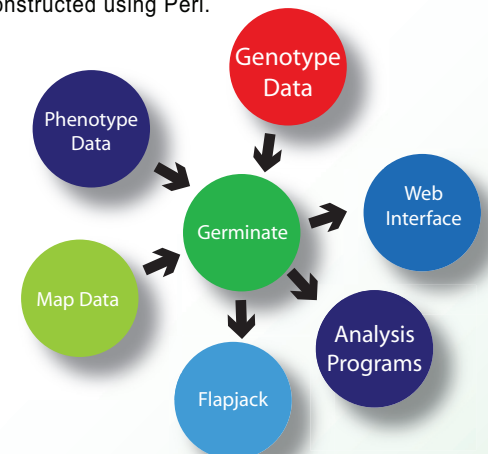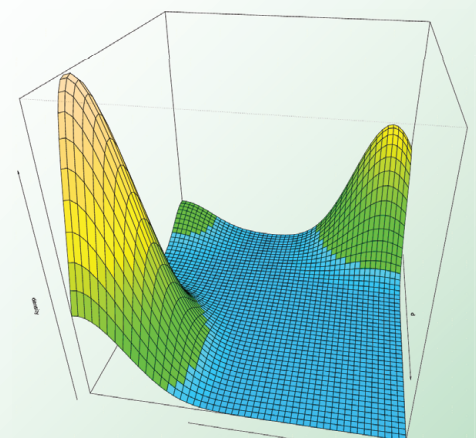

Data Workflow
Germinate stores raw experimental data and allows users to export in a format suited to particular analysis programs.

R Kernel Density Plot
Shows the density of XY signal data from Illumina SNP assays. These images are generated on the fly from data held in a Germinate based database.

http://bioinf.scri.ac.uk/germinate