

Integration of Experimental Plant Data to Develop Novel Visualizations for Genetic Analysis of Quantitative Traits

Paul Shaw^{1,2}, David Marshall¹, Martin Graham² and Jessie Kennedy²

¹Scottish Crop Research Institute, Invergowrie, Dundee, Scotland DD2 5DA.
²School of Computing, Edinburgh Napier University, Edinburgh, Scotland



Introduction

Many commercially important agronomic traits such as yield and malting quality are under the control of one or more genes plus environmental interactions (see figure 1). These environmental interactions include the effects of climate change on the environment. Food security is a critical issue and a key question is how plant breeding can adapt to maintain this security through periods of changing climatic and environmental conditions. The integration, analysis and novel visualization of genetic and phenotypic data will allow better decisions to be made as to which crop varieties grow best in varying conditions resulting in increased yield, quality, and profitability for farmers



Figure 1 - Plant quantitative traits
 Plant height is a trait whose data follows a normal distribution. This image shows experimental barley (*Hordeum vulgare*) plots at SCRI.

A data warehouse system has been developed in order to store and allow integration of the large number of data types that are necessary in such experiments. In addition basic tools for biological data visualization, representation and data export have been developed. These will form the foundations of subsequent work.

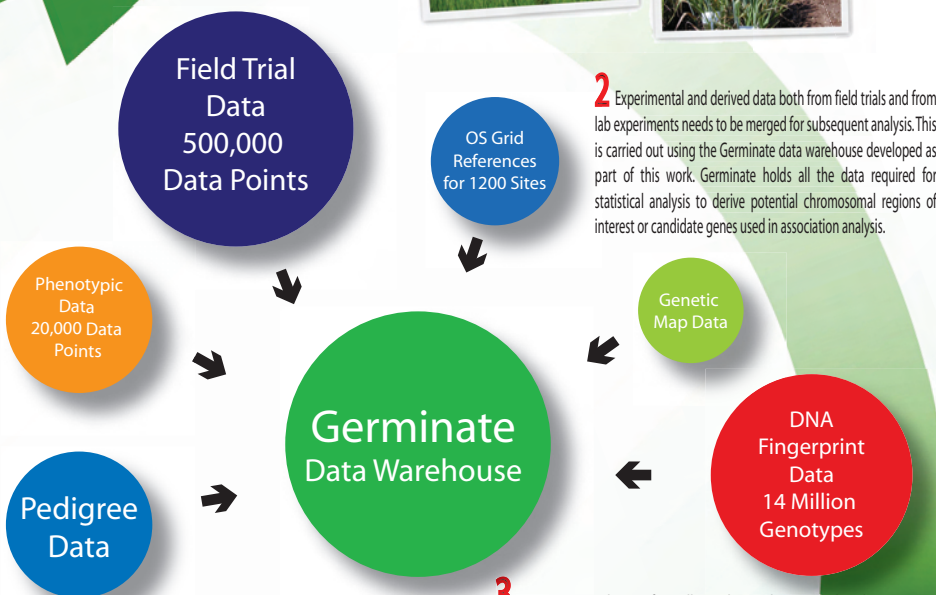
By interacting with geneticists and plant breeders we can now begin to address specific biological problems that are associated with the types and volumes of data that are currently being generated by current high-throughput technologies.

Identifying Potential Genes Responsible for Heritable Characteristics Workflow

1 Selected plant lines are grown and phenotyped based on both morphological and physiological / biochemical characteristics. Examples of these being plant height, b-glucan concentrations, and grain yield. Each of these has applications to varying industries such as food, animal feed production and brewing.



6 Cycle starts again...



2 Experimental and derived data both from field trials and from lab experiments needs to be merged for subsequent analysis. This is carried out using the Germinate data warehouse developed as part of this work. Germinate holds all the data required for statistical analysis to derive potential chromosomal regions of interest or candidate genes used in association analysis.

3 Germinate web-interface allows data to be merged and offers standard data views and export options for other analysis programs. Germinate also has direct integration with tools such as SCRI's Flapjack for graphical genotyping. Future development of visualisation tools may be web or standalone application based.

5 Analysis results are used to select lines which may be grown or crossed in subsequent years based on detected desirable characteristics at the DNA level.

4 Exported data is subject to statistical analysis using applications such as Genstat or R. Analysed data is fed back into the Germinate system in case it is required again in the future. All data is tagged as a completed dataset that includes any additional information on how the original data was analysed.

