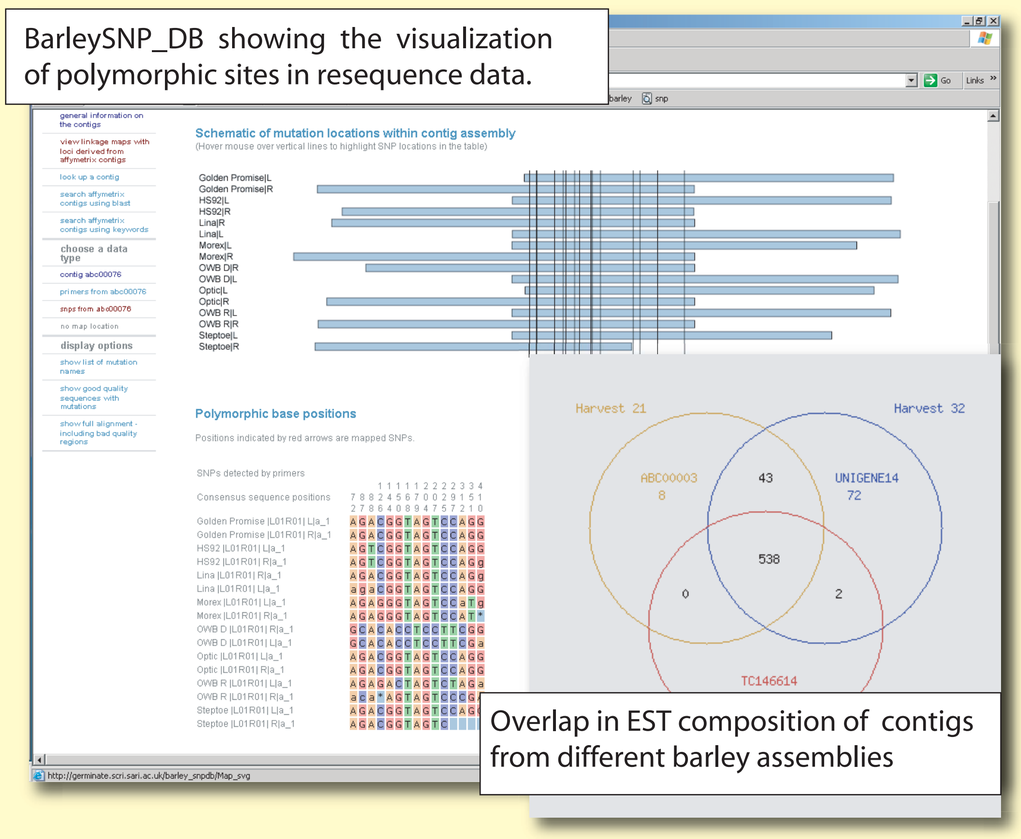
The Visualization & Analysis of Barley SNPs

David Marshall, Linda Milne, Daniel Nwankwo, Andy Flavell*, Luke Ramsay and Robbie Waugh

Scottish Crop Research Institute, and the *University of Dundee at SCRI, Invergowrie, Dundee, DD2 5DA, Scotland, UK (David.Marshall@scri.ac.uk)

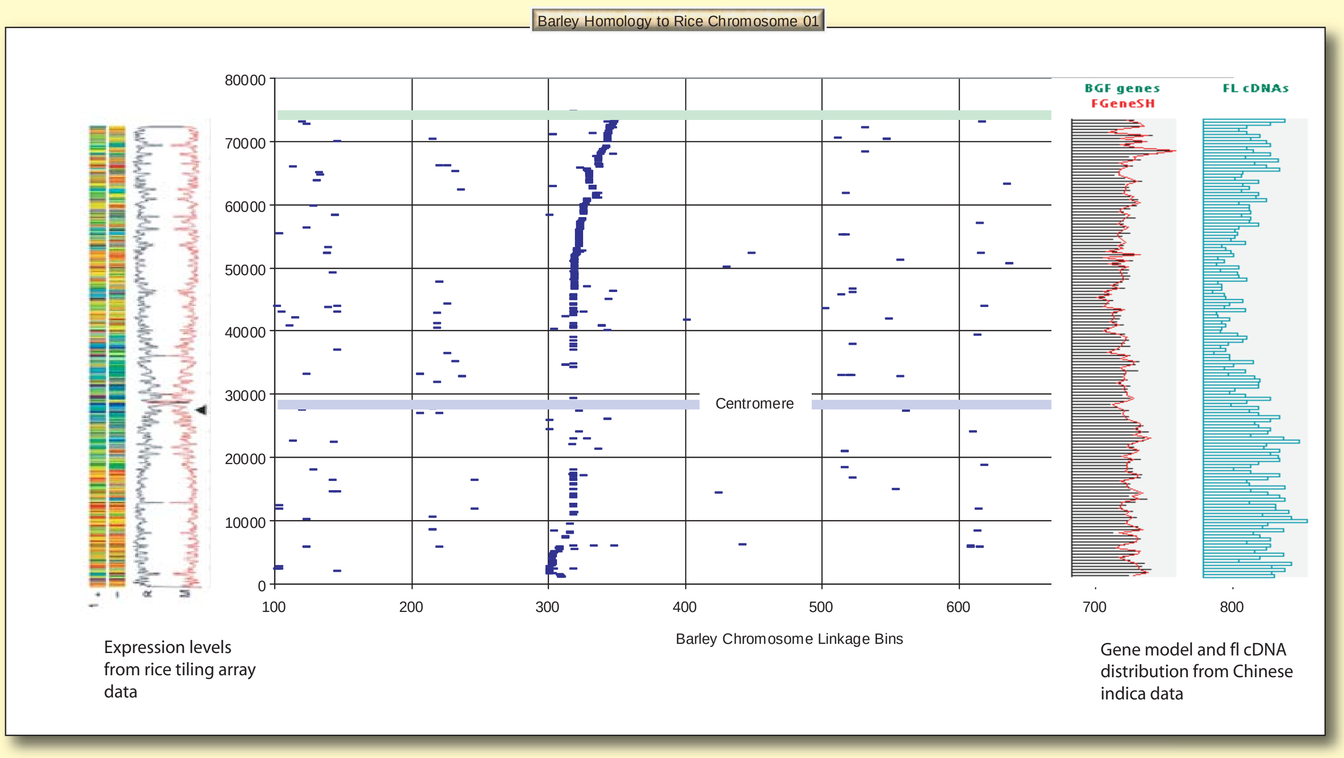
http://bioinf.scri.ac.uk

Through a collaboration between the barley team at SCRI, Tim Close's group at UCR and Andreas Graner's group at IPK, Gatersleben, we have developed a high throughput genotyping platform for barley based on the Illumina Golden Gate assay. We have developed 3 x 1536 OPA SNP assays and have map and diversity data from the first two OPAs. The sheer scale and quality of the data that this approach is generating is setting us a major challenge to develop and implement the required range of informatics resources and tools capable of storing, analysing and visualizing this SNP. The major components so far developed for our system are: BarleySNP_DB, a new comparative mapping tool, the Germinate Database, GVT (a genotype visualisation tool) and PVT (a comparable tool for the visualization of genotype data in a pedigree context.



BarleySNP_DB showing the visualization of polymorphic sites in resequence data.

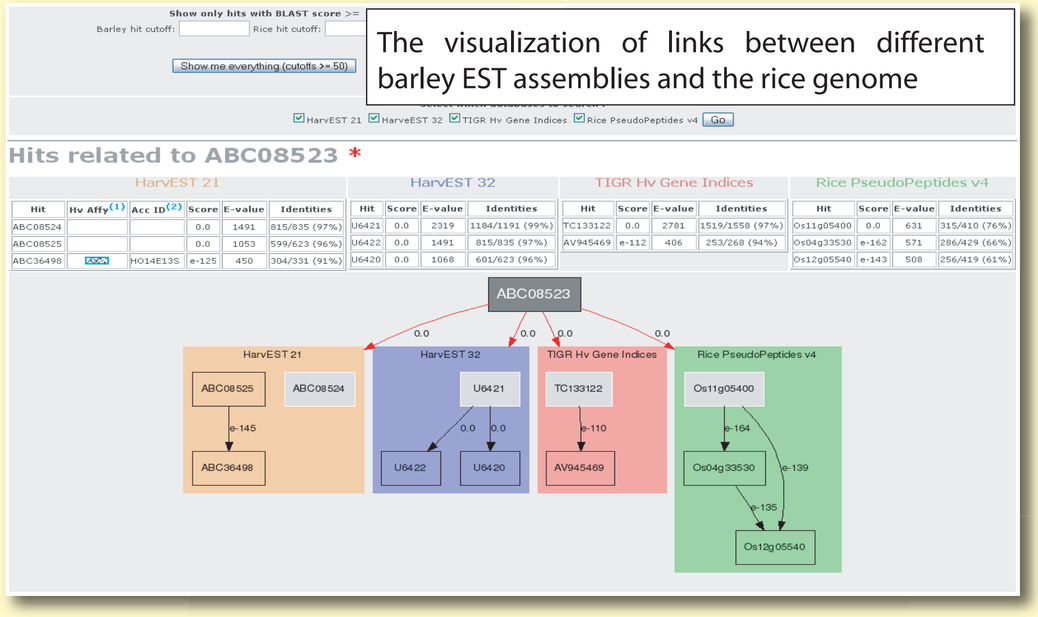
Overlap in EST composition of contigs from different barley assemblies



Expression levels from rice tiling array data

Gene model and fl cDNA distribution from Chinese indica data

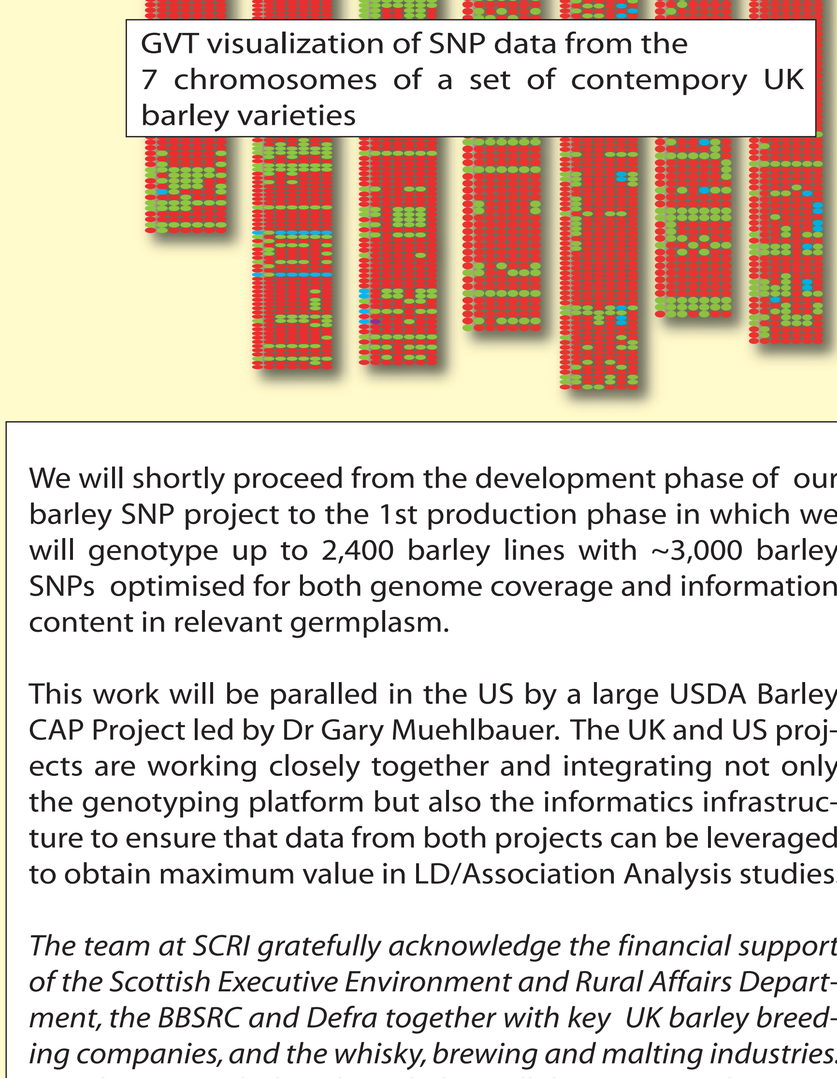
Comparative map between the barley genome and the TIGR rice chromosome 01 pseudomolecule showing the excellent conservation of synteny exception in the the gene poor retroelement rich centromeric region



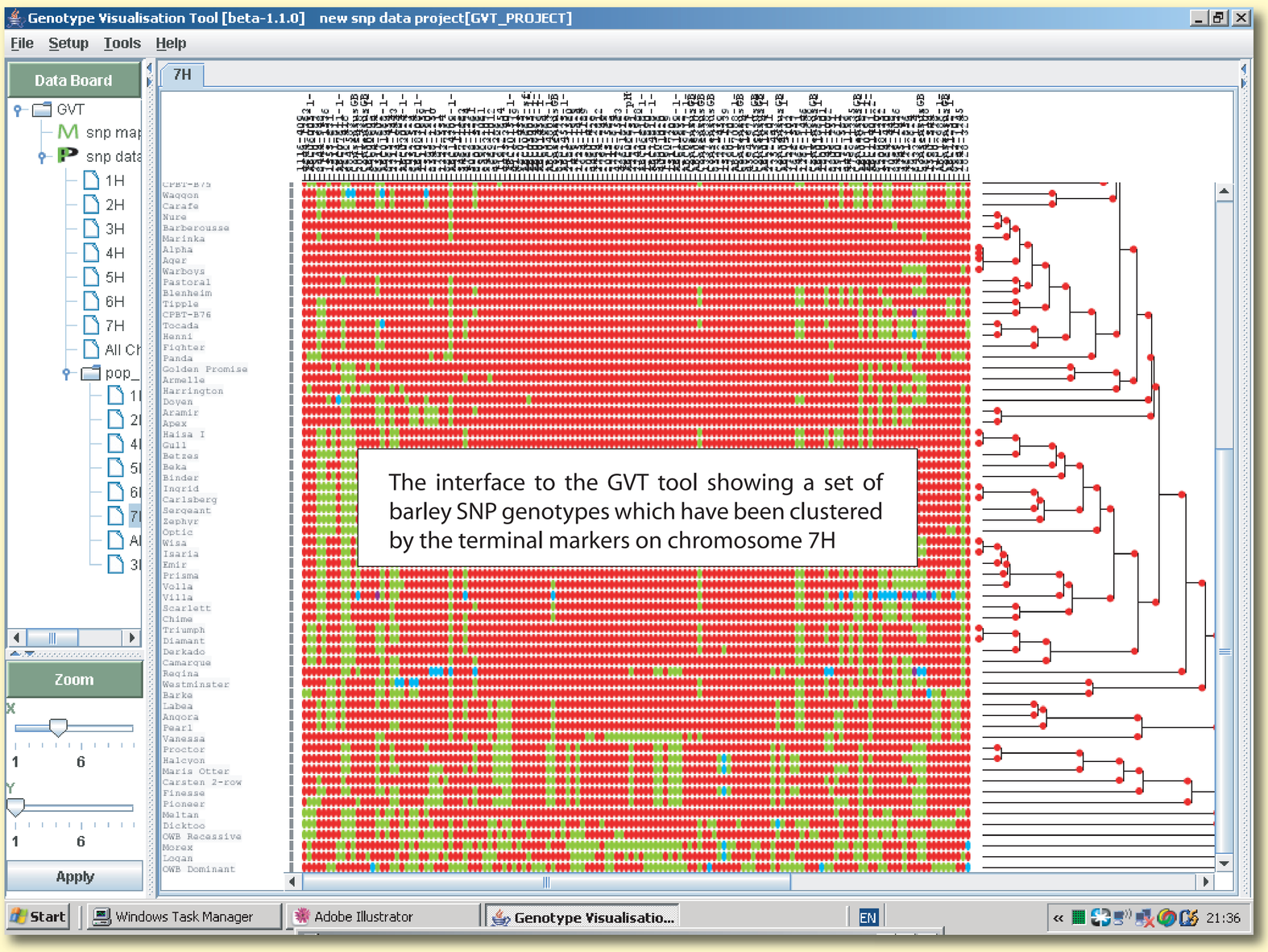
The visualization of links between different barley EST assemblies and the rice genome

Currently, through a combination of SNP and SFP (Single Feature Polymorphism) markers we have mapped in excess of 4,000 sequence based markers in barley. This has enabled us not only to build a high resolution gene map of barley but also to generate a unique high resolution comparative map between the barley and rice genomes which is yielding new insights into the organisation of the barley genome in terms of the distribution of recombination events and genes. This in turn is helping us to design map-based gene isolation approaches and also to understand their limited utility in certain regions of the barley genome.

The GVT program, the first of a series of planned Java-based software tools, enables us to visualize SNP and SSR data and to cluster, sort or query our database or barley lines based on genotype or phenotype data. GVT can access data from sources ranging from a simple spreadsheet to a Germinate database. Currently we are optimising the SOAP-based webservices link to GVT.



GVT visualization of SNP data from the 7 chromosomes of a set of contempory UK barley varieties



The interface to the GVT tool showing a set of barley SNP genotypes which have been clustered by the terminal markers on chromosome 7H

We will shortly proceed from the development phase of our barley SNP project to the 1st production phase in which we will genotype up to 2,400 barley lines with ~3,000 barley SNPs optimised for both genome coverage and information content in relevant germplasm.

This work will be paralled in the US by a large USDA Barley CAP Project led by Dr Gary Muehlbauer. The UK and US projects are working closely together and integrating not only the genotyping platform but also the informatics infrastructure to ensure that data from both projects can be leveraged to obtain maximum value in LD/Association Analysis studies.

The team at SCRI gratefully acknowledge the financial support of the Scottish Executive Environment and Rural Affairs Department, the BBSRC and Defra together with key UK barley breeding companies, and the whisky, brewing and malting industries. We also particularly acknowledge collaborations with Dr Tim Close of UCR, the barley team at IPK Gatersleben and members of the USDA Barley CAP Project.