

**Barley Coordinated Agricultural Project Work Plan FY08 (4/1/08 – 3/31/09) Objective 2:  
THT Development**

**Julie A. Dickerson, Roger Wise, Iowa State University**

**1) Describe the research, education, and outreach activities you are planning for the next year (4/1/08 – 3/29/09)**

***THT Development (Dickerson, Wise)***

The redesigned website will continue to be tested by the different Barley CAP user communities to gather comments on web site performance and ease of use. The site and database will also be systematically assessed by the THT team with respect to different user categories to make sure that all basic functionality needs have been met. Ongoing informal assessments with the CAP project director and CAP breeding program coordinator will aid this process.

***Data Visualization and Exploration:*** This effort has been reprioritized to FY2009 while database redesign and website development is taking place. Current visualizations allow display of maps and plots of data points across an experiment. The Germinate SNP JAVA viewing tools (GVT) from the Scottish Crop Research Institute will be made available for use by the BarleyCAP project and THT will develop a download application to provide data for this application. Automated clustering and search routines will be added to allow users to search for similar data sets across a range of variable types including categorical, numerical, and SNP calls. For phenotypic data, interactive data plotting and outlier detection tools will be added to aid users in sorting through datasets. Display of genetic maps will be improved by implementing CMAP tools for map comparisons.

***File format development.*** The THT team will continue to verify that all traits are fully defined with value ranges and “reasonable” inputs given for developing a robust data verification process. This task is a small effort at this time.

***Data upload.***

Data from the 2007 cropping seasons needs to be standardized against the list of existing traits for review and uploaded to THT as it is curated. SNP datasets for 2007 and the mapping populations will be added to the database as well as the newly developed maps.

***Data download.***

All datasets will be downloadable as separate spreadsheets (complete) and as combined data sets including different trials. Support for programs such as QTLMiner, GVT, Tassel will be made available to THT users at the web service and input file level as appropriate.

***Data Integration***

The main task is to continue developing a structure for integrating information from a variety of partner databases including GrainGenes, WebHarvEST, PLEXdb, and Germinate. This will involve web services and the development of common viewing tools. Our plan is to show sample interactions at the 2009 CAP meeting in January and gather user comments on the current THT web portal, then schedule a set of web meeting(s) with the different databases to discuss coordination.

***Education and Outreach:***

The main outreach component will be to make THT publicly accessible on the web and to publicize it to the Triticeae, as well as the broader plant breeding community. A publication on the website and the types of data will be submitted to Nucleic Acids Research.

**2) List specific outcomes and deliverables that will be accomplished in the first 6 months (4/1 – 9/30). These will be used as benchmarks for your bi-annual progress report.**

- The Germinate SNP viewing tools will be adapted for use by the BarleyCAP project and will be made available for use.
- All curated phenotypic datafiles with defined traits for 2007 will be uploaded to THT.
- THT use case assessment report to see if functional design goals have been met. Any deficiencies in will be corrected.
- A webservices interface for THT from the TASSEL program will be provided.

**3) List specific outcomes and deliverables that will be accomplished in the second 6 months (10/1 – 3/31). These will be used as benchmarks for the bi-annual progress report.**

- Users will be able to display and manipulate SNP data on the THT website.
- Data clustering algorithms for traits including categorical and numerical methods for selecting datasets will be added to the THT website.
- Continued development of display and analysis tools on the THT website.

**Barley Coordinated Agricultural Project Year 3 Progress Report  
Objective 2: THT Development  
FY08(4/1/08 – 9/30/08)  
Julie A. Dickerson, Roger Wise, Iowa State University**

**1) Research, education, and outreach activities**

***THT Development (Dickerson, Wise)***

The production version website will be tested by the Barley CAP user community to gather comments on web site performance and ease of use. The initial design prompted a restructuring of the underlying data schema this spring. The upload and query scripts are now being further developed. Once a complete data set for CAP FY 2006 (year 1) has been provided to THT, the site and database will be reassessed by the THT team with respect to the different user categories to make sure that all basic functionality needs have been met. Ongoing informal assessments with the CAP project director and CAP breeding program coordinator have aided this process. All curated phenotypic datafiles with defined traits have been uploaded to THT. This data can be reviewed by searching by year and breeding program. The user is then presented with a summarized list of all data available for that program and year. Each dataset can be viewed individually and downloaded.

***Data Visualization and Exploration:*** This effort has been reprioritized while database redesign and website development is taking place. The Germinate SNP JAVA viewing tools from the Scottish Crop Research Institute will be made available for use by the BarleyCAP project and THT will develop a download application to provide data for this application. Automated clustering and search routines will be added to allow users to search for similar data sets across a range of variable types including categorical, numerical, and SNP calls. Visualization tools for

SNP data such as CMAP or adaptations from the HapMap project will be added. Different methods for accessing the database such as searches using Google Earth or Google Maps will be explored.

***File format development.*** The input data file formats were revised and cleaned up in May after a meeting with the CAP data curator, Jennifer Kling. The existing data files were updated to the new formats. As new traits with different characteristics (time series, arbitrary scales) are added to the database, trait definitions are needed for THT. This is an on-going effort between THT and the CAP data curator. Traits that have been problematic have been the disease traits. The THT team will continue to verify that all traits are fully defined with value ranges and “reasonable” inputs given for developing a robust data verification process.

#### ***Data download.***

Data from the 2006-7 cropping seasons has been standardized against the list of proposed traits for review and upload to THT. The data was crosschecked to ensure that the line names correspond to previous pedigree and trait data entered by the curator. The calls, certainties, and data coordinates are stored in the database and the raw data from the Illumina Bead arrays from 2006 and 2007 CAP lines and the CapCore lines has been provided by the Close and Chao labs. The 2006 and 2007 data has been uploaded to THT. The raw data files and sample sheets will be compressed and be made available for download on THT.

#### ***Data Integration***

The main task is to continue developing a structure for integrating information from a variety of partner databases including GrainGenes, Gramene, WebHarVEST, PLEXdb, and Germinate. This may involve web services and the development of common viewing tools. The updated data schema will be made available to the Cornell group to allow them to develop webservices for an interface to the TASSEL program. Our plan is to discuss proposed interactions at the 2009 CAP meeting in January and gather user comments on the updated THT web portal, then schedule a set of web meeting(s) with the different databases to discuss coordination.

#### ***Education and Outreach:***

The main outreach component will be to make THT publicly accessible on the web and to publicize it to the Triticeae, as well as the broader plant breeding community. A publication on the website and the types of data will be submitted to Nucleic Acids Research.

## **2) Specific outcomes and deliverables accomplished in FY08 (4/1/08 – 09/30/08).**

- Website was evaluated and then redesigned to make it more accessible to the plant breeding community.
- THT schema was altered to accommodate experiments that cross breeding programs.
- Finalized plans for submission of breeder data from Jennifer Kling using a spreadsheet interface. Developed system for pedigree data entry.
- Entered all curated phenotype data from 2006 into THT.
- Entered CAPCore, BOPA1 2006 and 2007 data into THT.
- Entered genetic map data, markers and their synonyms, etc into THT.
- Added to a list of measured traits, their definitions, and their units.