

Barley Coordinated Agricultural Project Work Plan FY06 (4/1/07 – 3/31/08)
Timothy Close & Stefano Lonardi, University of California, Riverside

1) Describe the research, education, and outreach activities you are planning for the next year (4/1/07 – 3/31/08)

Research. The implementation of Illumina SNP marker technology in BarleyCAP is preceded by several actions: 1) SNP identification, 2) SNP testing in the oligonucleotide pool assay (OPA), referred to as "pilot" OPAs, 3) data analysis and genetic map construction from the results of pilot OPAs, 4) selection of SNPs for two permanent OPAs. In total, the plan is to have three "pilot" OPAs (PilotOPA1, PilotOPA2, PilotOPA3) from which two permanent OPAs (BarleyOPA1, BarleyOPA2) will be designed. PilotOPA1 tested 1524 SNPs and preceded BarleyCAP in an NSF project of Close and Lonardi in collaboration with colleagues in Germany and Scotland. BarleyCAP supports the production of two additional pilots (PilotOPA2, PilotOPA3) and two permanent OPAs (BarleyOPA1, BarleyOPA2). As of January 5, 2007 PilotOPA2 work had been completed (1536 SNPs tested), BarleyOPA1 had been designed (1536 successful SNPs represented) and PilotOPA3 design was on track for completion before 4/1/07. This leaves only BarleyPilotOPA2 to be designed during the period of 4/1/07 to 3/31/08. To design BarleyOPA2, we will analyze data obtained from PilotOPA3 to determine the technical performance and allele frequencies of 1536 additional SNP markers. This second of the two planned permanent OPAs will then be designed and purchased. The content of BarleyOPA2 will have a somewhat different emphasis from BarleyOPA1. The only two criteria used for SNPs represented on BarleyOPA1 were technical high performance and a minor allele frequency not less than 8%. The design of BarleyOPA2 will adhere to these two criteria for most SNPs, but some SNPs that reside in genes of particular interest to BarleyCAP will be represented even if the minor allele frequency is below 8% or technical performance is somewhat less than the cutoff point applied to BarleyOPA1. Lists of genes of particular biological interest have been provided by Patrick Hayes ("beer genes", including SNPs), Blake Cooper and Nora Lapitan, (malting quality candidate genes from expression studies), Brian Steffenson (*Rpg1* SNPs) and Roger Wise (disease response genes from expression studies). These will complete the scope of PilotOPA3 SNPs and thereby feed material into the design of BarleyOPA2. The first of two permanent OPAs (BarleyOPA1), designed in December 2006, is expected to be synthesized by May 2007 and implemented for genotyping in Shiaoman Chao's lab soon thereafter.

The genotypes that are providing data from PilotOPAs include three doubled haploid mapping populations derived from crosses of Morex/Barke, Oregon Wolfe Barleys and Steptoe/Morex. This enables the iterative evolution of high density SNP-based genetic linkage maps. The genotypes applied to the PilotOPAs also include core US and European germplasm samples. This enables the determination of allele frequency in breeding germplasm and selection of SNPs that will be most informative for the breeding programs served by BarleyCAP. For PilotOPA3, we will again ready the DNA, process raw data obtained from a genotyping lab at University of California, Los Angeles, generate Illumina BeadStudio workspaces and standard summary tables, and create other data tables in formats compatible with mapping programs including JoinMap 4.0 and RECORD. Since the production of genetic linkage maps is part of the iterative process of SNP workspace definition, we will also provide the project with first-draft genetic linkage maps through HarvEST:Barley. We will be engaged with Patrick Hayes to finalize the maps, partnered with collaborators at Scottish Crop Research Institute and elsewhere. Finalized maps will be provided to The Hordeum Toolbox and other outlets including GrainGenes, TIGR, NCBI and Gramene.

We will apply the OPAs to the deconvolution of BAC clone-locus relationships. During year #1 we refined the deconvolution algorithms and experimented with combinatorial pooling. During the period of 4/1/07 through 3/31/08, a new phase of deconvolution will take place. We intend to use a combinatorial method designed for 2197 BACs using 91 pools of 169 BACs each. Locus-BAC relationships will be partially solved by 3/31/08. A new algorithm that exploits the physical map along with the results of pooling is currently being designed and extensively tested on the rice genome. Preliminary results show a significant increase in the accuracy of the deconvolution. New versions of HarvEST:Barley may display locus-BAC relationships. Transfer of such data to other outlets will occur after 3/31/08, once the results have become stable.

Education. We have two PhD students working on algorithms for the deconvolution of locus-BAC relationships and the construction of an improved physical map, which is essential in order to improve the deconvolution.

Outreach. Close will make presentations about BarleyCAP and the SNP methods that our project is using. Lonardi will be on call to explain the locus-BAC deconvolution algorithms to appropriate audiences. We will provide input and feedback to Peggy Lemaux in the design of brochures, posters, and PowerPoint presentations in

relation to the methodologies that we are leading on this project.

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.

- Generate data from PilotOPA3 for three mapping populations, 96 US and 96 European breeding genotypes
- Update genetic linkage map using PilotOPA3 data
- Provide compiled linkage map through HarvEST:Barley
- Release all PilotOPA mapping data to THT and other public outlets
- Transfer all BarleyCAP PilotOPA raw data and summary tables to THT
- Assist with the production of a workspace from BarleyOPA1
- Use results from PilotOPA3 to begin the design of BarleyOPA2
- Apply combinatorial pools of 2197 BACs to re-synthesized PilotOPA1

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.

- Close the design of BarleyOPA2
- Organize a buyer group to purchase BarleyOPA2
- Assist with the production of a workspace from BarleyOPA2
- Show the results of BAC-SNP deconvolution in HarvEST:Barley and through <http://phymap.ucdavis.edu:8080/barley/>
- Provide deep link information to THT developers to help the THT portal link to information on BACs, the physical map, BAC-SNPs, and genetic maps

Barley Coordinated Agricultural Project Progress Report FY06 (4/1/06 – 3/31/07)

Objective 1

Progress Report for UC Riverside (4/1/06 – 3/31/07)

Timothy Close and Stefano Lonardi, University of California, Riverside (UCR)

Research.

The design was completed for the first of two project pilot oligonucleotide pool assays (OPAs). This pilot OPA (PilotOPA2) was ordered, synthesized, received and used at the University of California, Los Angeles genotyping lab (Joe DeYoung) to generate data from 480 DNA samples. These included 96 samples for each of three mapping populations: 1) Morex x Barke - DNA provided by Nils Stein at Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany; 2) Oregon Wolfe Barley, DNA produced at UCR; and 3) Steptoe x Morex, DNA produced at UCR. Also included were 96 European germplasm selections, DNA provided by Nils Rostoks at Scottish Crop Research Institute (SCRI, Dundee, Scotland) and Nils Stein, and 96 US germplasm selections, DNA provided by BarleyCAP partner Patrick Hayes (Oregon State University). The name "PilotOPA1" refers to a pilot OPA that was developed and implemented prior to BarleyCAP in an NSF-funded project of Close and Lonardi in collaboration with colleagues at SCRI and IPK. GenCall workspaces were made for the PilotOPA2 data and transmitted in September 2006 to Patrick Hayes, SCRI and IPK. Data tables were exported from the PilotOPA2 GenCall workspace and provided to BarleyCAP participants Patrick Hayes for pedigree analysis and collation of Oregon Wolfe Barley marker data, Julie Dickerson and Roger Wise for familiarization with data structures, and Jean-Luc Jannink for an assessment of the genotyping power of the SNPs on PilotOPA2. Close met with Dickerson and Wise in September to provide instruction on the nature of OPA data. PilotOPA2 added 843 high quality SNP markers to an evolving SNP-based barley genetic linkage map, taking the total to about 1810 when combined with high quality mapped loci from PilotOPA1. HarvEST:Barley (<http://harvest.ucr.edu> and www.harvest-web.org) provides a portal to the OPA-based genetic maps during the current map development phase.

Nils Rostoks (SCRI) worked with Close in California for two weeks in 2006 to begin the design of PilotOPA3, identifying approximately 900 SNPs that cover genes not addressed by PilotOPA1 and PilotOPA2. Completion of PilotOPA3 design was postponed until after the January 2007 Plant and Animal Genome meeting to review the results from PilotOPA2 and provide additional opportunities for input from BarleyCAP participants, the Advisory Board and other members of the barley community. PilotOPA3 is being designed with extra attention to high priority genes advocated by project participants including Blake Cooper and Nora Lapitan (malting quality candidates), Patrick Hayes (malting quality candidates), Brian Steffenson (*Rpg1*), and Roger Wise and Gary Muehlbauer (disease responsive). In February 2007 Close received ~94,000 new EST trace files from Nils Stein (IPK) and Kazuhiro Sato (Okuyama University, Japan). These new sequences were integrated with previously available sequences by UCR programmer Steve Wanamaker to create a new HarvEST:Barley EST assembly, and to derive an extended list of SNPs. Annotations of ESTs and unigene consensus sequences were updated with BLASTs against TIGR's version 5 rice gene models. The infusion of new ESTs resulted in more than 1000 additional unigenes with good SNPs. This simplified the task of PilotOPA3 design, as it created choices of how to fill PilotOPA3 other than by focusing multiple SNPs on some hundreds of genes. PilotOPA3 is on track for finalization in April 2007. The new surplus of SNP-bearing genes also sets the stage for PilotOPA4, which is outside the scope of BarleyCAP but is now under discussion in the worldwide barley community.

The design of PilotOPA1 was conveyed to Shiaoman Chao so that it could be re-synthesized and utilized in the North Dakota genotyping lab to test barley DNA preparation methods and establish standard operating procedures. The re-synthesized PilotOPA1 also provides an opportunity to generate genotype calls from the same US core germplasm samples that were genotyped using PilotOPA2. The first set of results was produced by Shiaoman Chao in December 2006. Close and Chao jointly examined the data, created a workspace and concluded that the results are of excellent quality.

The first permanent OPA, BarleyOPA1, was designed by December 7, 2006 by choosing only technically high performing SNPs from PilotOPA1 and PilotOPA2. The content of BarleyOPA1 was further limited to loci that have a major allele homozygote frequency not more than 92% among cultivated germplasm and parents of mapping populations. Of these 1536 loci, 1314 have been genetically mapped and 222 have not been genetically mapped. Following the Plant and Animal Genome meeting in January 2007, Close coordinated a worldwide barley community purchase of BarleyOPA1 sufficient to process 20,064 DNA samples, including all of the samples foreseen from BarleyCAP and many more. The order was placed in early March with an expected

delivery date in April 2007. The design of BarleyOPA1 was provided to Julie Dickerson for integration into THT, and to others in the barley community.

On the computational side, Lonardi, Close and graduate student Wu have been working on the BAC-unigene deconvolution problem. We designed a new algorithm that is capable of producing high accuracy deconvolution even in the presence of a weak pooling design, i.e., when probe pools include 200-300 probes. The algorithm compensates for the decrease of information in the hybridization data by taking advantage of a physical map of the BAC clones. In a simulation carried out on the rice genome, we showed that the right combination of combinatorial pooling and our algorithm not only dramatically reduce the number of pools required, but also deconvolutes BAC-gene relationships with almost perfect accuracy. This has been summarized in a manuscript submitted in February to one of the top computational biology conferences ISMB/ECCB 2007.

Education. Two Computer Sciences PhD students and two post-docs. One of these PhD students, Serdar Bozdag, has been improving FPC assembly and the minimal tiling path of gene-bearing BAC clones. The other PhD student, Yonghui Wu, has been working on combinatorial algorithms for the deconvolution of BAC-SNP relationships. Post-doc Prasanna Bhat has assisted with genotyping and routinely produced genetic linkage maps from OPA data as part of an iterative workflow between genetic map production and manual supervision of the GenCall/BeadStudio workspaces. Close, Bhat and post-doc Jan Svensson interacted with Programmer Steve Wanamaker, co-PI Lonardi and graduate students Bozdag and Wu to streamline the workflow between HarVEST, GenCall/BeadStudio, JoinMap, RECORD and an evolving in-house mapping algorithm.

Outreach. Close made oral presentations on BarleyCAP at: 1) the RiceCAP meeting in Houston, Texas in February 2006, 2) the USDA Western Regional Research Center in Albany, California in March 2006, 3) a legume workshop in Asilomar, California in August 2006 (also a poster), 4) the annual Society of Italian Agricultural Genetics Congress in Ischia, Italy in September 2006, 5) the Plant and Animal Genome meeting Genomics Assisted Breeding Workshop in January 2007, and 6) UC Berkeley in February 2007. Close and Lonardi provided input and feedback to Peggy Lemaux and Barbara Alonso on preparation of BarleyCAP posters.

Specific outcomes and deliverables accomplished in FY06 (4/1/06 – 3/31/07)

- Designed PilotOPA2 by April and produced data by August
- Genetic linkage map with 843 new SNP markers by December
- Designed BarleyOPA1 by December, synthesis order placed in March
- Algorithm for BAC-SNP deconvolution improved to consider assembly and hybridizations
- Presentations in February, March, August, September, January, February