

Barley Coordinated Agricultural Project Work Plan for Year 4 (4/1/2009-3/30-2010)
Timothy J. Close and Stefano Lonardi, University of California Riverside

1) Describe the research, education, and outreach activities you are planning for Year 4 (4/1/2009-3/30-2010)

Research

Consensus Map. As of September 2008, a SNP-only consensus map produced by the UCR group was composed of 2943 OPA-SNP markers. The number of mapped SNPs was 341 for 1H, 485 for 2H, 475 for 3H, 338 for 4H, 535 for 5H, 352 for 6H and 417 for 7H. For all but 18 of these SNPs, it was also clear which are on the short arm and which are on the long arm, based on results obtained with flow-sorted chromosomes in collaboration with Jaroslav Dolezel and members of his laboratory in Czech Republic, as part of the NSF project of Close and Lonardi. Approximately 400 genes represented by SNPs on BOPA1 and BOPA2 still had an uncertain map position as of September 2008. Results of linkage disequilibrium mapping and the application of BOPA1 and BOPA2 to additional mapping populations will provide data that can resolve many of these unmapped SNPs, as will additional data from flow sorted arms. The UCR group will continue to update the consensus SNP map as new data become available during year 4, and disseminate the new map through Shiaoman Chao and THT, as well as through journal articles.

Linking the physical and genetic maps. The method of deconvolution of BAC-gene relationships developed by the UCR group using BOPA1 and BOPA2 will be extended to include at least two more sets of BAC pools, and possibly as many as six more sets of BAC pools, pending a continued high success rate. Because the BACs that comprised the first set of BAC pools were pre-enriched to include genes represented on BOPA1 and BOPA2, subsequent sets of BAC pools are expected to resolve fewer gene-BAC relationships. BAC pools will be prepared at UC Riverside and pooled BAC DNAs will be sent to Chao for processing on BOPA1 and BOPA2. Deconvolution of gene-BAC relationships will be accomplished at UC Riverside by a new PhD student working with Lonardi and Close. In addition, purely computational deconvolution of gene-BAC relationships will be improved by using existing BAC-gene relationships as ground truth for the training of enhanced deconvolution algorithms.

Support of BarleyCAP genotyping. BarleyCAP breeder lab genotyping data will continue to be produced by Shiaoman Chao in year 4. The UC Riverside group will continue to work with Chao to act as a second source of decisions to make the genotyping calls before dissemination to BarleyCAP. This method of interaction has worked well during years 1, 2 and 3. Programmer Wanamaker at UC Riverside will continue to enhance the HarvEST:Barley software to provide ready access to the details of SNPs, the genetic map and the physical map.

Education

A PhD student in Computer Sciences will be supervised by Lonardi and Close to improve the purely computational gene-BAC deconvolution algorithm and to support gene-BAC deconvolution using BOPA1 and BOPA2. At least one undergraduate student in the Biological Sciences will participate in production of BAC combinatorial pools in the Close laboratory. Materials and algorithms developed and used by this project will be integrated into graduate and undergraduate courses instructed by Close and Lonardi.

Outreach

Close and Lonardi will make presentations about BarleyCAP and the SNP and BAC methods that our project is using. Lonardi also will provide explanations of the genetic and physical mapping innovations and locus-BAC deconvolution algorithms to appropriate audiences. We will provide input and feedback to Peggy Lemaux and Brian Steffenson in the design of brochures, posters, and PowerPoint presentations in relation to the methodologies that we are leading on this project.

Specific outcomes and deliverables that will be accomplished in the first 6 months (4/1 – 9/30/09)

- Update genetic linkage maps by incorporation of data from linkage disequilibrium and new mapping populations
- Release to Shiaoman Chaot and THT new linkage map coordinates for BOPA1 and BOPA2 markers
- Maintain complete SNP linkage maps and query mechanisms through HarvEST:Barley
- Refine MSTMap as a web-based tool to make individual maps
- Refine MergeMap as a web-based tool to make consensus maps
- Analysis of data from BOPA1 and BOPA2 produced by Shiaoman Chao
- Prepare combinatorial pools of Morex barley BACs
- Apply combinatorial pools of BACs to BOPA1 and BOPA2
- Solve BAC-locus relationships for an additional ~500 SNP loci using BOPA1 and BOPA2
- Release the most stable information on BAC-locus relationships to all BarleyCAP participants
- Provide link information to THT team for genetic maps, BACs, the physical map
- Publication of research articles

Specific outcomes and deliverables that will be accomplished in the second 6 months (10/1 – 3/31/10).

- Update genetic linkage maps by incorporation of data from linkage disequilibrium and new mapping populations
- Release to Shiaoman Chaot and THT new linkage map coordinates for BOPA1 and BOPA2 markers
- Maintain complete SNP linkage maps and query mechanisms through HarvEST:Barley
- Refine MSTMap as a web-based tool to make individual maps
- Refine MergeMap as a web-based tool to make consensus maps
- Analysis of data from BOPA1 and BOPA2 produced by Shiaoman Chao
- Prepare combinatorial pools of Morex barley BACs
- Apply combinatorial pools of BACs to BOPA1 and BOPA2
- Solve BAC-locus relationships for an additional ~500 SNP loci using BOPA1 and BOPA2
- Release the most stable information on BAC-locus relationships to all BarleyCAP participants
- Provide link information to THT team for genetic maps, BACs, the physical map
- Publication of research articles

Barley Coordinated Agricultural Project Biannual Progress Report
FY08 (4/1/08 – 9/30/08)
Timothy Close and Stefano Lonardi, University of California, Riverside

1) Describe the research, education, and outreach activities you completed in the first half of FY08 (4/1/08 – 9/30/08)

Research.

Production OPA #2 (BarleyOPA2; BOPA2). Following the UCR group's completion of the design of BOPA2 from POPA1, POPA2 and POPA3 SNPs, the BOPA2 design was transmitted to Illumina and several BarleyCAP and AGUOEB participants. A two month delay in processing the order took place due to the complex nature of Illumina dealing with a consortium of four buyers (Muehlbauer, Hicks, Hayes, Close). This was followed by delays in synthesis by Illumina due to a shortfall capacity caused by a surge in the demand for Illumina products. BOPA2 arrived in the Close lab at UCR on July 22. By July 29 all intended recipients had received their expected portions of BOPA2: 51 tubes to Shiaoman Chao at the USDA genotyping lab in Fargo, 39 tubes to Kevin Hicks and Gary Muehlbauer at University of Minnesota, 12 tubes to Joe DeYoung at University of California Los Angeles as a swap for 12 BOPA1 tubes, and 3 tubes of BOPA2 retained by the Close lab at UC Riverside. The availability of BOPA2 fulfilled the objective of having two complete OPAs representing more than 3000 SNPs.

Consensus Map. As of 16 April 2008, the SNP-only consensus map produced by the UCR group was composed of 2943 mapped OPA-SNP markers, with no conflicts of marker order between any pair of individual maps or between the consensus map and any of four DH-derived individual maps. The method used by the UCR team to produce the consensus map involved production of ordered bins of markers from each DH population, followed by production of a directed graph, followed by calculation of a non-conflicted linear map representing one possible solution of the directed graph. The method was developed mainly by UCR PhD student Yonghui Wu under the guidance of Stefano Lonardi. The directed acyclic graphs are available through the HarvEST web interface (www.harvest-web.org). The MxB map contains 1647 SNPs in 443 bins, the OWB map contains 1562 SNPs in 509 bins and the SxM map contains 1267 SNPs in 396 bins. The consensus map contains 1069 markers from POPA1, 1038 from POPA2 and 836 from POPA3. The number of mapped SNPs is 341 for 1H, 485 for 2H, 475 for 3H, 338 for 4H, 535 for 5H, 352 for 6H and 417 for 7H. For all but 46 of these SNPs, it is also clear which are on the short arm and which are on the long arm, based on results obtained with flow-sorted chromosomes in collaboration with Jaroslav Dolezel and members of his laboratory in Czech Republic, as part of the NSF project of Close and Lonardi. A comprehensive compilation of synonymous marker names, HarvEST:Barley unigene numbers, map location, rice gene orthologs and Affymetrix Barley1 probe sets matching the SNP loci was disseminated to BarleyCAP participants on April 16. All data from POPA1, POPA2 and POPA3 were made available by FTP to Julie Dickerson for integration into THT.

Linking the physical and genetic maps. The UCR team sent to Shiaoman Chao in July the same combinatorial pools of gene-bearing Morex minimal tiling path BAC DNAs that they applied to BOPA1 in 2007. Chao then applied these DNA samples to BOPA2 in September and transmitted the data to the UCR group. Prasanna Bhat at UC Riverside, working with former PhD student Yonghui Wu (graduated August 2008) then solved the gene-BAC relationships. The results were somewhat better than anticipated with 447 BOPA2 SNPs being unambiguously related to 636 BACs. This raised the totals to 1150 BOPA1 or BOPA2 SNPs anchored to 1319

unique BACs. Because most of these SNPs have been positioned on the genetic linkage map, this provides nearly 200 BACs anchored to the genetic linkage map. A minimal tiling path of barley gene-bearing BACs was refined and finalized by UCR PhD student Serdar Bozdag using some improvements to the standard FPC method, providing a list of 13,182 MTP BACs which were rearranged by Clemson University Genomics Institute and sent to UC Riverside in September. This material forms the basis of additional combinatorial pools that will be applied to BOPA1 and BOPA2 to increase the density of anchored BACs. The vision of a well populated SNP map tied to the barley physical map has become an expanding reality.

Support of BarleyCAP genotyping. The 2007 BarleyCAP breeder lab data from BOPA2 were received from Shiaoan Chao in September. Prasanna Bhat and others in the UC Riverside group refined the genotype calls somewhat after receiving an excellent draft of genotyped calls from Chao and returned the calls to Chao for distribution to BarleyCAP breeders and THT. Programmer Wanamaker added a zoom-in function and BAC addresses to the HarvEST:Barley software.

Education. Two Computer Sciences students and one post-doc worked on the project during the reporting period. One PhD student, Serdar Bozdag (graduated August 2008, now at NIH), continued to improve algorithms for FPC assembly using a compartmentalized approach and calculation of the minimal tiling path of gene-bearing BAC clones using integer linear programming. The other PhD student, Yonghui Wu (graduated August 2008, now at Google), created a web interface for the mapping algorithm which he developed to compute a genetic map as a solution of the minimum spanning tree of a graph and continue to refine and create a web interface for his algorithm to compute a consensus map from a directed acyclic graph derived from multiple maps. Post-doc Prasanna Bhat continued to have the lead role in handling the genotyping and BAC-gene deconvolution data using the Illumina BeadStudio workspace and locally developed scripts for data processing. Close and Bhat communicated with programmer Steve Wanamaker to streamline the workflow between HarvEST, BeadStudio, the mapping algorithm and the HarvEST web interface for DAGs.

Outreach. Close summarized the BarleyCAP genotyping strategy at the annual meeting of the Tropical Legumes I project in Dakar, Senegal in July, a minisymposium on plant genomics at the IPK in Gatersleben Germany in July, and briefly at the Annual Meeting of the Generation Challenge Program in Bangkok, Thailand in September. Yonghui Wu presented "On the accurate construction of consensus genetic maps" in CSB2008 at Stanford University in August. Serdar Bozdag provided a presentation on "Computing the minimal tiling path from a physical map by integer linear programming" in the WABI2008 Workshop on Algorithms in Bioinformatics at Karlsruhe Germany in September.

2) List specific outcomes and deliverables accomplished in the second half of FY08 (4/1 – 9/30)

- Finished design of BarleyOPA2 (BOPA2)
- Received and distributed BOPA2
- Propagated 2943-SNP map and supporting information throughout BarleyCAP to support genotyping for BarleyCAP breeder data
- Provided all POPA1, POPA2, POPA3 raw data and Illumina workspaces to THT
- Upgraded HarvEST:Barley map displays with zoom-in and BAC addresses
- Improved the genetic linkage mapping algorithms

- Improved BAC assembly and minimal tiling path algorithms
- Related 447 additional SNPs to BACs using BOPA2 data
- Made five presentations in July, August and September