

PROJECT DESCRIPTION

Introduction

The overall theme of the barley Coordinated Agricultural Project (CAP) is to integrate and utilize state-of-the-art genomic tools and approaches in plant breeding programs, thereby facilitating the development of superior cultivars. More specifically, we will integrate genetic and trait data to gain access to the genes that control important biological, agronomic, and quality traits. We will use single nucleotide polymorphic (SNP) markers to conduct high-throughput mapping in barley. A densely saturated genetic map will provide markers for the breeding programs, initiation points for map-based cloning, and scaffolds for genome sequencing. Our goal is to genetically map approximately 3,000 SNP markers utilizing approximately 3 SNPs/gene. The mapped SNPs will be integrated with the emerging BAC-based physical map and Barley1 GeneChip expression data. We will use the SNP markers to genotype 3,840 lines from ten barley-breeding programs, short pedigrees from each breeding program, and a collection of diverse barley genotypes. The SNP genotyping will also be used to understand linkage disequilibrium and the haplotype blocks in current and historic barley breeding germplasm. We will also collect data on the germplasm for a set of important traits and create a large phenotypic database based on the Genomic Diversity and Phenotype Data Model (GDPDM). The genotype and trait data will be combined to conduct association genetics-based studies to detect marker-trait associations. Detecting marker-trait associations using breeding germplasm will lead directly to the implementation of marker-assisted selection (MAS) to manipulate important traits and develop improved varieties. The project website “The *Hordeum* Toolbox” will contain all trait and SNP data compiled by the CAP; a user-friendly “QTL Miner” statistical package for analyzing the genotype-trait data; an interactive viewer to sort and examine SNP data, haplotypes and related germplasm; and the site will be integrated with the barley genetic map, physical map and expression databases. The barley CAP will be fully integrated with the educational and outreach missions of the participating institutions.

GOALS AND DELIVERABLES:

- Obj. 1. **High-Throughput Marker Development** (Close, Hayes) - Develop an integrated 3,000 SNP-based genetic/physical/expression map of the barley genome.
- Obj. 2. **Worldwide Web Access** (Dickerson, Wise) - Develop integrated web portal and database for efficient access to map, gene expression, and trait data.
- Obj. 3. **Genes and Traits** (Smith, Hayes) - Access economically important genes using association genetics and diverse breeding germplasm.
- Obj. 4. **Superior Germplasm** (Smith, Hayes) - Develop elite barley germplasm using high-throughput marker-assisted selection.
- Obj. 5. **Education and Outreach** (Steffenson, Lemaux) - Educate students, scientists, farmers, processors, and consumers in use of genomics technology for gene discovery and crop improvement.

Background and Literature review

Barley genomics resources. Barley (*Hordeum vulgare* L.) is a large-genome monocot (~5,000 Mb; Arumuganathan and Earle, 1991) and a true diploid member of the Triticeae. The genetic map is approximately 1,250 cM and contains over 4,000 molecular markers, 250 morphological markers, and approximately 700 QTL (URL 1; URL 2 [see Website References

below]; Franckowiak, 1997; Kleinhofs and Graner, 2001; Ramsay et al., 2000; Thiel et al., 2003). As of May 20, 2005, there were 419,087 barley ESTs (*Hordeum vulgare* subsp. *vulgare* and *spontaneum* combined) in GenBank. A 6.3X Morex barley BAC library (Yu et al., 2000) is used worldwide. To initiate physical mapping of the genome, an NSF Plant Genome grant was awarded to T. Close to capillary fingerprint 60,000 gene-containing Morex BAC clones and to tie this information to barley unigenes, Barley1 GeneChip content, and the HarvEST interface (URL 3). Functional genomics resources include a structured, mutagenized population for TILLING (Caldwell et al., 2004) and transgenic barley carrying the maize Ac-Ds transposons (Cooper et al., 2004). To enable large-scale RNA profiling, the 22K Barley1 GeneChip probe array was developed through cooperation of the barley community (Close et al., 2004).

To deliver the power of barley genomics to the desktop of geneticists and breeders, the BarleyBase and HarvEST:Barley databases have been developed. BarleyBase (URL 4; Shen et al., 2005; Tang et al., 2005) is an on-line interactive database for raw and normalized expression data for cereal GeneChips. BarleyBase and its successor, PLEXdb (Plant Expression Database; URL 5) is funded through an USDA-NRI grant to J. Dickerson and R. Wise and supports an array of comparative gene expression analyses for the Barley1, wheat, rice, maize, and other plant GeneChips. Controlled vocabularies, gene and trait ontologies, as well as interconnecting links with PlantGDB (URL 6), Gramene (URL 7) and GrainGenes (URL 2) make queries as seamless as possible. HarvEST:Barley (URL 3) served as the source of Barley1 GeneChip content and the EST assemblies and unigene BLAST annotations that are now also shown in BarleyBase. The development of HarvEST:Barley software was initially supported through an USDA-NRI grant to T. Close and co-workers. Additional funds to add BAC clone information, genetic map displays and web operability were provided by the NSF Plant Genome Research Program.

Association genetics. Traditionally, genes controlling a trait have been mapped in segregating populations developed from crosses among inbred lines (Jannink et al., 2001). This approach is limited by 1) the need to generate a population segregating for each trait that is to be mapped, and 2) the number of informative meioses occurring in the progeny population. A more powerful approach for mapping economically important traits in a large collection of germplasm is association mapping (Thornsberry et al., 2001). This method of analysis is based on linkage disequilibrium (LD) or the non-independence of alleles in a population (Gaut and Long, 2003). Association mapping circumvents the need for constructing genetic mapping populations for each trait of interest and instead utilizes all of the recombinations that have occurred throughout the evolutionary history of the plant. Thus, association mapping is very robust because it considers many more informative meioses than traditional mapping.

The degree of LD in a population will determine the resolution of mapping that is possible (whole genome versus intragenic scans). In wild barley, *Hordeum vulgare* ssp. *spontaneum*, intralocus LD decays within 300 bp at rates similar to maize (Morell et al., 2005; Rafalski and Morgante, 2004). In cultivated barley, LD is observed from 5-10 cM (R. Waugh, personal communication; J-L. Jannink, unpublished results; Kraakman et al., 2004; Condon and Smith, 2005), which is a level that should be useful in whole genome scans for marker-trait associations.

Association genetics has been a standard approach to improve the mapping resolution of human disease genes (Jorde, 2000), but it has not been widely used in plants. In maize, significant associations of quantitative traits with candidate genes have been reported (e.g., Thornsberry et al., 2001; Wilson et al., 2004), and in barley (Kraakman et al., 2004) significant associations were identified with yield and yield stability with a limited number of anonymous

AFLP markers. A key feature of association mapping is the need to determine population structure, because mixtures of populations can lead to spurious associations (Pritchard et al., 2000). Association genetics approaches have used molecular markers to determine the probability that an individual belongs to a known population group or subgroup. In breeding germplasm, however, pedigree records already provide information on population structure and that information can be used to account for structure in models to conduct association mapping.

Mixed-model approaches for genetic evaluation have incorporated both phenotypic and pedigree data and have been successfully used in animals (Henderson, 1984) and in plants (Bernardo, 1996). The mixed-model approach can be extended to incorporate genomic data (Kennedy et al., 1992; Bernardo, 1998), leading to association genetics via a mixed-model approach. Association genetics via a mixed-model approach, which incorporates phenotypic, pedigree, and genomic data, was useful for gene discovery in complex structured populations of maize (Parrisieux and Bernardo, 2004) and wheat (Arbelbide and Bernardo, 2006). Simulation studies have confirmed that gene mapping through a mixed-model approach is useful in both cross- and self-pollinated crops (Yu et al., 2005; Arbelbide et al., 2006).

Association mapping via a mixed-model approach has four advantages that make it particularly useful in the context of a plant-breeding program (Parrisieux and Bernardo, 2004). First, it exploits large mapping populations so that the power to detect genes is increased and the estimates of gene location become more precise. Second, genotypes, inbreds in the case of barley, are evaluated in multiple, diverse environments. The use of many environments permits the sampling of a sufficient set of genotype x environment interactions so that the results are applicable across a wide range of future environments. Third, the inbreds tested typically comprise a wide sample of the germplasm and genetic backgrounds that are relevant for breeding. Fourth, the pedigree and trait data used for association genetics via a mixed-model approach are those that are routinely generated in the context of the breeding program. This permits a greater leverage of resources for variety development, gene discovery, and MAS.

Marker-assisted selection (MAS) for crop improvement. In their seminal paper, Lande and Thompson (1990) showed how DNA marker information could improve estimation of genetic values for the purpose of selection. MAS has been employed primarily by either selecting for desirable alleles at target regions in early generation populations to increase the frequency of the desired genes or through the use of marker-assisted backcrossing by selecting for donor alleles at the target region (foreground selection) and recipient alleles for the rest of the genome (background selection). MAS using these approaches has been successful for large effect QTL or traits controlled by single genes. In barley, MAS has been successfully employed to select for single QTL alleles (e.g., Coventry et al., 2003; Collins et al., 2003) as well as to pyramid several QTL to enhance disease resistance (Castro et al., 2003a; 2003b). The effectiveness and efficiency of MAS is dependent on several factors including: the degree to which desirable alleles are linked to undesirable alleles (linkage drag), the stability of the effect of target alleles at QTL across different environments and genetic backgrounds, the availability of high-throughput markers that flank the region of interest, and the relative cost of genotyping versus phenotyping.

Despite the successes of MAS in barley, there are several barriers to its widespread adoption in breeding programs. Linkage with undesirable traits has limited the use of MAS for *Fusarium* head blight (FHB) resistance in barley. Major QTLs on chromosome 2H and 6H have been identified; however, the resistance alleles are linked to late heading and high grain protein concentration, respectively (de la Peña et al., 1999; Mesfin et al., 2003; Canci et al., 2003). Fine

mapping and additional marker saturation are necessary to break up these linkages and permit MAS. Some QTL have exhibited variable effects across different genetic backgrounds (e.g., Bilgic et al., 2005) and therefore better information about the effect of target alleles in breeding germplasm is necessary prior to MAS. Finally, effective MAS strategies require that large numbers of breeding lines be genotyped over a short period of time at a low cost. PCR-based markers have potential use in MAS; however, the current density is insufficient to permit breeders to identify sufficient numbers of informative markers for any region of the genome.

In addition to approaches that impose selection for specific alleles at marker loci, there are MAS strategies that estimate breeding value based on a combination of marker and phenotypic data. In this case, a marker score is calculated based on the effects of alleles that it carries at marker loci for each line that is under selection. The marker score is then combined with the phenotype of the line into an index that accounts for the trait heritability. Theoretically, this index better estimates the line's true genetic value than the phenotype alone. Simulation studies, however, have shown that this approach to MAS is more efficient than phenotypic selection only under conditions (low trait heritability and low number of loci affecting the trait) that are not likely to hold for most quantitative traits. The reason for the poor gain in efficiency has been dubbed the "catch-22 of MAS" (Holland, 2004): effective MAS requires accurate estimates of QTL effects, which in turn requires accurate phenotypic data. But in the presence of accurate phenotypic data, phenotypic selection itself is effective (Bernardo, 2001; Moreau, 2000).

A critical objective of the barley CAP is to overcome the MAS catch-22 by providing accurate estimates of QTL effects through LD mapping. The key difference between QTL effect estimates from traditional linkage mapping versus from LD mapping is that in the former, the estimates apply to only a single segregating population resulting from a biparental cross, whereas in the latter case the estimates apply (as implemented here) across all U.S. barley breeding programs. Indeed, the broader inference enabled by the LD approach proposed here means that the large number of phenotypic observations present in the database can substitute for phenotypic observations on experimental lines to which MAS will be applied.

Barley breeding programs and traits of interest. During the past year as part of the planning process for this proposal, invitations were extended to all public and private U.S. barley-breeding programs. The ten CAP participating programs constitute a large portion of the barley improvement research community in the U.S. and service the entire spectrum of barley growing regions and market end-uses. The University of Minnesota (K. Smith; UM), North Dakota State University (R. Horsley, J. Franckowiak), and Busch Agricultural Resources, Inc. (B. Cooper; BARI) breeding programs service the Upper Midwest production area and emphasize spring malting barley and to a lesser extent feed barley. Efforts are directed to both six- and two-rowed varieties with the greater emphasis on six-rowed. The primary traits that are important for this region are yield, lodging, disease resistance (FHB, net blotch, Septoria speckled leaf blotch, common root rot, and spot blotch), and malting quality. The expansion of barley production into central and western North Dakota have placed increased emphasis on grain protein concentration, plant height, and adaptation to more arid environments. The Oregon State University (P. Hayes; OSU) program concentrates on winter barley varieties for malting. The USDA-Aberdeen, Idaho (D. Obert) and Utah State University (D. Roche) programs have efforts on both two- and six-rowed spring/winter barley for both malting and feed. The Washington State University (S. Ullrich) program develops spring two- and six-rowed barley varieties for malt, feed and food uses, including hullless and waxy types. In addition to quality, yield and

resistance to lodging, stripe rust, root rots, Hessian fly, and Russian wheat aphid are emphasized. The Montana State University (T. Blake) program develops spring two-rowed barley varieties for both malt and feed uses. The important traits for this region are yield (including water-limiting conditions), lodging resistance, disease resistance (stripe rust, barley yellow dwarf virus, and net blotch), and malting and feed quality. The USDA-Aberdeen program breeds low phytate barley for feed in animals as well as aquaculture. The Virginia Tech (C. Griffey) program breeds for the Southeastern U.S. and focuses on winter barley for feed, food, and ethanol production. The important traits for this region are yield, test weight, seed quality, lodging resistance, disease resistance (leaf rust, net blotch, and powdery mildew), and hull adherence.

Integration of genomics resources into barley breeding programs. Molecular markers are fully integrated into many U.S. barley-breeding programs. Two examples are described here. Fusarium head blight (FHB) is a disastrous disease problem in barley resulting in reduced yield and malting/feed quality (McMullen et al., 1997). Mapping and utilizing QTL for FHB resistance has been pioneered at the UM and MAS for resistance is currently a major emphasis of the UM barley-breeding program (de la Peña et al., 1999; Mesfin et al., 2003; Canci et al., 2004). Barley stripe rust can be a devastating disease, particularly in the Pacific Northwest of the U.S. Fifteen years ago, U.S. barley researchers established a partnership with the ICARDA/CIMMYT program in Mexico to mine genes determining quantitative resistance to stripe rust and other diseases. This led to the mapping of multiple resistance genes (Chen et al., 1994; Toojinda et al., 1998); pyramiding of quantitative resistance genes (Castro et al., 2003a); pyramiding quantitative and qualitative resistance genes (Castro et al., 2003b); and the release of varieties developed via MAS (Hayes et al., 2000; Hayes et al., 2003a). These efforts have established the foundation for the current effort to characterize genes determining quantitative resistance, agronomic performance and malting quality traits in terms of structure, function, and mechanism (Hayes et al., 2003b). These examples show that U.S. barley breeders are in a position to directly utilize marker-trait information in the development of new cultivars.

Genotyping centers. To facilitate MAS in the small grains, the USDA-ARS has established four genotyping centers. These centers are located in Manhattan, KS; Pullman, WA; Fargo, ND; and Raleigh, NC. The Fargo genotyping center, led by S. Chao, is an integral part of the barley CAP. Chao will establish the Illumina-based SNP mapping system in her lab and conduct SNP mapping and genotyping for the project. Thus, the barley CAP will leverage the investment in the genotyping centers and provide a long-term structural resource, in the form of Illumina-based SNP detection system, for future SNP genotyping in the small grains.

Rationale

The substantial resources of barley genomics have not been integrated to improve the efficiency of barley breeding and provide geneticists and breeders with seamless access to genes controlling economically and biologically important traits. ***The keys to this project*** are high-throughput SNP genotyping of U.S. barley breeding germplasm and leveraging the tremendous amount of trait data routinely collected by breeders to map QTLs for use in MAS. Thus, the most practical and tractable outcomes are a comprehensive understanding of genetic diversity within the U.S. breeding germplasm, improved tools for manipulating traits through high-throughput MAS, enhanced germplasm exchange among breeding programs, and improved barley varieties. QTL mapping and introgression have focused on, and been effective with, maximum variance structured populations: e.g. disease resistant x disease susceptible. These "allele" mining efforts continue and are the subject of endeavors supported by other agencies:

e.g., FHB (U.S. Wheat and Barley Scab Initiative); stripe rust (USDA-NRI); disease resistance genes from wild barley (U.S. Barley Genome Project; USBGP). However, only a fraction of the sources of new alleles used in breeding programs are the subjects of bi-parental mapping studies. **The current project is unique** in that it addresses a more pressing need for barley and as a model for all other crops - identifying genes controlling economically important phenotypes in breeding material. Thus, our approach to mine breeding germplasm for novel alleles and genes will complement traditional mapping efforts and tremendously enhance MAS-based breeding approaches. To ensure that this information is readily accessible to all barley researchers, we will develop the “The *Hordeum* Toolbox” (THT) web portal. THT will contain all genotype and trait data generated in this project and will provide the platform for continued addition of genetic and phenotypic data for association genetics studies and breeding objectives beyond the granting period. THT will also be integrated with barley genetic maps, physical maps, and expression data with the ultimate goal of establishing a model for the application of genomics to crop improvement. To enable plant scientists to conduct association genetics, we will develop the user-friendly “QTL Miner” software. Our education plan will provide the next generation of plant breeders with the newest translational genomics technologies, statistical tools and breeding approaches. Extension of the project information and tools will be multifaceted including: development of THT, release of the “QTL Miner” software, and talks/workshops at scientific and industry meetings. Finally, in addition to developing a coordinated barley improvement program, we will develop and integrate an approximately 3,000 SNP map with the physical map and expression databases to lay the foundation for future barley genomics, including more efficient map-based cloning, and whole genome physical mapping and sequencing.

Obj. 1. High-Throughput Marker Development (Close, Hayes) - Develop an integrated 3,000 SNP-based genetic/physical/expression map of the barley genome. There are over 4,000 molecular markers mapped on the various barley mapping populations. However, the utility of these markers in breeding programs is restricted due to the lack of high-throughput technologies and insufficient polymorphism in breeding germplasm. Thus, our aim is to develop a saturated genetic map with SNP markers for haplotype-block analysis, QTL identification through association genetics, and MAS for barley improvement. SNP markers have the advantage of being numerous, high throughput and easily utilized by breeding programs through the genotyping centers. Our goal in this objective is to identify and map approximately 3,000 SNP markers with 3 SNPs/gene. To enable integration of the SNP map and emerging physical map with transcript accumulation data from the Barley1 GeneChip platform, only those SNPs that have sequences represented on the GeneChip will be chosen.

The number of genes and the number of SNPs/gene chosen (3,072 SNPs derived from 1,000-1,500 genes) to map for this objective and screen the germplasm (**Obj. 3**) was determined based on the observations that haplotype structure and LD is maintained up to distances of 5-10 cM (Kraakman et al., 2004). This number of SNPs will ensure an average of about 2 SNPs/cM on the approximately 1,250 cM barley genetic map (Kleinhofs et al., 1993). Therefore, the genome coverage will provide the ability to identify haplotype blocks that underlie the patterns of variation in this elite germplasm. We will use Illumina’s BeadArray™ platform technology for the SNP genotyping because it will generate a large amount of data quickly and economically.

Identification of SNPs using barley EST sequence information in the HarvEST:Barley database. The HarvEST:Barley database (URL 3) displays several barley EST CAP3 (Huang and Madan, 1999) sequence assemblies made for different purposes. The current version, HarvEST:Barley 1.35, displays four assemblies: 21, 25, 31 and 32, and provides a cross-

referencing tool between them. Assembly 32 uses CAP3 settings that are appropriate for SNP discovery and is the basis of electronic SNP (eSNP) calling in Close's NSF project entitled, "Coupling Expressed Sequences and Bacterial Artificial Chromosome Resources to Access the Barley Genome". A pairwise comparison of the eight barley genotypes that provided the largest number of ESTs yielded a total of 10,985 eSNPs among 3,334 assembly 32 unigenes. Of these, approximately 1,000 SNPs for abiotic-stress-related genes, one SNP per gene, have been represented on a pilot Illumina Oligo Pool Assay (OPA) in Close's NSF project. The remaining eSNPs then define the minimum number of additional SNPs that can be utilized in an expanded SNP effort such as this proposed project. To enable integration with Barley1 GeneChip expression data, only those SNPs that have sequences represented on the Barley1 GeneChip will be used for SNP mapping. Close's eSNP calling method has a validity rate of >90% as determined by direct sequencing of PCR-derived amplicons. This very high validity rate is accomplished by applying several criteria to discard questionable data including the following.

- Only bases with a phred quality value of 25 or higher are considered
- Positions that contain any disagreement between two reads from each end of the same clone are excluded
- Positions near the ends of the sequences are ignored
- Positions adjacent to low-quality patches are ignored
- A polymorphism must have support from at least 2 ESTs from each of two genotypes and must represent at least 90% of the sequences available from each respective genotype

In the proposed project, Close will leverage expertise gained in his NSF project to design the content of the Illumina OPAs, and to relate these eSNPs to the barley physical map and EST databases. This will include the generation of a new eSNP list, the output of data in standard Illumina design format, interaction with the Illumina team to determine a list of SNPs that have a high likelihood of success, and relational databasing to facilitate decision-making and prioritization by the participants in this project. HarvEST:Barley will continue to be a public source of full details of the underlying EST assemblies and provide project participants with complete displays of eSNP details. HarvEST:Barley and the Morex barley physical map database (URL 8) will soon (supported by current grants to Close from USDA and NSF) allow users to query SNPs, Morex BAC clone addresses, Genbank numbers, regions of rice synteny and barley genetic map positions. This information will be interlinked to trait, QTL and expression data through THT web portal (**Obj. 2**).

In addition to the eSNPs defined by Close, SNP discovery solely by genomic amplicon sequencing (dSNPs) by collaborators R. Waugh at the Scottish Crop Research Institute (SCRI) and A. Graner at the Institute of Plant Genetics and Crop Plant Research (IPK) identified 565 and 217 unigenes containing SNPs, respectively. The complete eSNP and dSNP compilation was used for an Illumina OPA designed to detect 1,536 SNPs. The cooperative plan between these groups is to genotype 96 lines each from Steptoe x Morex, Barke x Morex, and Oregon Wolfe Barley (OWB) populations. Also 96 different cultivars, landraces, and other elite lines will be examined. The OPA was recently received by the N. Friemer laboratory at the University of California, Los Angeles with the results expected in early July. Close will coordinate the incorporation of these SNPs into the barley CAP project.

Illumina SNP technology. Illumina has developed a cost-effective SNP-based genotyping method based on a 1,536-SNP multiplex assay protocol that utilizes the BeadStation 500G, which is a suitable instrument for core facilities (URL 9). This method utilizes hybridization of

allele-specific primers to template genomic DNA, extension and ligation, signal amplification by PCR, and finally signal detection using fiber optic bundles in the BeadArray™ platform. BeadStation 500G can easily handle hundreds of genotypes per day and comes with an advanced Laboratory Information Management System (LIMS) for error-free sample tracking and facile integration with relational databases such as HarvEST:Barley and THT (**Obj. 2**). The Illumina method has an accuracy rate comparable to high-performance PCR-based genotyping assays and has been widely tested and proven in the Human HapMap project (URL 10). Two and possibly three 1,536-SNP Illumina OPAs will be used to achieve the goal of 3,000 mapped SNPs.

Linkage map construction. We will map approximately 3,000 SNPs using the Oregon Wolfe Barley (OWB) mapping population as our primary vehicle. Hayes will coordinate this effort. The OWB population shows very high polymorphism rates and the linkage map contains multiple types of markers, including morphological, RFLP, SSR and AFLP, as well as STS markers (Cooper et al., 2004). The OWB map thus provides an efficient mapping vehicle and direct connectivity to the standard barley mapping populations and the barley BIN map (URL 11). Additional information on the OWB population is available at URL 1 and URL 2. When no polymorphism is found in the OWB parents, we will use two alternative mapping populations: Steptoe x Morex and Barke x Morex. SNP marker data will be generated using the Illumina bead technology as described above. Our principal tool for linkage map construction will be JoinMap (VanOoijen and Voorrips, 2001), with periodic cross-checking with G-Mendel (Holloway and Knapp, 1994). We will update maps frequently, and post supporting data sets at THT web portal (**Obj. 2**).

Deconvolution of gene-BAC relationship. To continue to integrate the genetic and physical map in barley for future map-based cloning of genes controlling traits mapped in **Obj. 3**, we will integrate the SNP map with the emerging BAC-based physical map. One objective of the NSF Plant Genome Research project entitled “Coupling Expressed Sequences and Bacterial Artificial Chromosome Resources to Access the Barley Genome” (2003-2007, Close PI, S Lonardi co-PI) is to screen the 6.3X genome coverage Morex barley BAC library with at least 12,600 overgo probes to identify most of the gene-bearing BAC clones, expected to be about 60,000 among the total number of 313,344 in the entire library. In general, the approach is to use pools of 192 probes to identify groups of gene-bearing BACs, assemble the BACs into contigs by fingerprinting and assembly, define a minimal set of BACs, then deconvolute the probe-BAC relationships for only 1,000 genes related to abiotic stress. New information from this project is continuously merged with probe-BAC relational information available from worldwide collaborators. The overgo probes were designed essentially as described in Zheng et al. (2004) from a barley unigene dataset, so we have a good idea of the relationships between overgo probes and the genes that they will detect. However, since probes are grouped in pools during hybridization, determining the relationship between probes and BACs requires additional effort. Close and Lonardi’s NSF project will utilize an Illumina OPA representing approximately 1,500 stress-related genes and ~96 multidimensional sub-pools of ~4,000 BAC clones carrying stress-related genes to deconvolute the relationships between a target 1,000 abiotic stress genes and BAC clones. In this proposed CAP project, Close and Lonardi will use the new OPAs to deconvolute additional BAC-gene relationships, but with sub-pools from a considerably larger set of ~15,000 BAC clones representing the remainder of the gene-positive BAC clones.

Specific Outcomes from Objective 1:

- Map positions of 3,000 informative SNP markers for U.S. barley germplasm

- SNP map integrated with BAC-based physical map

Obj. 2. Worldwide Web Access (Dickerson, Wise) - Develop integrated web portal and database for efficient access to map, gene expression, and trait data. Our objective is to develop The *Hordeum* Toolbox (THT), as an integrated public resource for discovery and dissemination of large-scale SNP data sets as they apply to genetic, phenotypic, and trait data. THT will provide an essential resource to serve as a bridge from SNP markers to traits through association to transcript (and eventually protein and metabolite) profiling. The underlying premise is that these molecular level phenotypes manifest as traits that plant breeders select for. At present, there is no resource in the community that fulfills the role of integrating the new and rapidly expanding SNP data sets with traditional phenotypic data, structural genomics, and gene expression profile data sets. THT will serve this function by integrating multiple data sets within and across germplasm and providing a single site to access, analyze, and disseminate all the data currently available for comprehensive association genetics and functional genomics studies. Our goal for THT is to provide the electronic framework for such community efforts. Once the data model has been established in THT, SNP data, pedigree data, and phenotypic data from breeding germplasm will be uploaded to the database on a regular basis. Long-term maintenance of THT will be supported with additional funds from source such as the U.S. Barley Genome Project, American Malting Barley Association, U.S. Wheat and Barley Scab Initiative and state barley grower groups. All of these sources fund research for barley improvement.

Establish infrastructure for housing SNP, phenotypic, and trait data and integrating with structural genomics and parallel expression data. THT will store all the SNP and trait data, view SNPs, haplotypes, and sort genotypes. All of these data will be cross-referenced to the Barley1 GeneChip contigs with integration to the physical and genetic maps and expression data. By genotyping diverse and common breeding lines in addition to mapping the same SNP markers on three populations, a target sequence (i.e., mapped gene) can be associated with traits, phenotypes, or expression.

THT will leverage existing federally funded bioinformatics work to develop an integrated information source. For example, germplasm, polymorphism, and trait data will be displayed using the Genomic Diversity and Phenotype Data Model (GDPDM; URL 12) its GBROWSE compatible-viewer being used for the Panzea project web site (URL 13). This tool is an interactive web-based viewer for displaying pre-computed multiple sequence alignments. The viewer provides a number of general features such as querying for alignments, setting global parameters and generating text dumps of sequences. The viewer also provides a set of features for assisting nucleotide polymorphism analysis. Varying positions are highlighted and displayed along with statistics. Quality scores can be displayed and poor quality sequences or segments can be filtered out. Sequences with low similarity to other sequences in the alignment can be removed. The viewer's functional features and look-and-feel can be customized to fit individual web site needs through a simple configuration file. The GDPDM schema is built on top of an open-source MySQL database. We will also make THT Genomic Diversity and Phenotype Connection (GDPC) enabled, providing access to genomic diversity data such as SNPs, SSRs, sequences, etc. and phenotypic data that will be collected in field, genetic, or physiological experiments (Casstevens and Buckler, 2004). GDPC simplifies access to genomic diversity and phenotype data by retrieving data from one or more data sources and by allowing researchers to analyze integrated data in a standard format.

We will also port the fully curated data to Gramene and Graingenes (see D. Ware and V. Carollo letters) on a semi-annual basis. A key advantage of using the Panzea software described

above is that it is already developed for SNP and trait data. This software is also being used by various wheat and rice projects that are also being ported to Gramene as well. All map data will be tied to the Barley1 GeneChip expression data through sequence alignments in BarleyBase.

Integrate SNP and gene expression data in a user-friendly web portal. The Barley CAP project will build on the team's existing successes with HarvEST:Barley and BarleyBase to facilitate comparative SNP data analysis as it applies to genetic, phenotypic, and trait data. The Barley1 GeneChip now serves as a reference set of 20,000 gene sequences from which active projects in genome-wide SNP development are being launched (Rostoks et al., 2005). Thus, every SNP marker will be mapped back to the sequences on the Barley1 GeneChip (via VMATCH technology) as a permanent sequence based link to the rapidly accumulating expression profile data sets.

Develop a user-friendly web portal with links to Hordeum Information Resources. The barley CAP project will maintain a project web portal that will provide access to a variety of *Hordeum* information resources as well as THT. THT will be designed to meet the needs of plant biologists for analysis of SNP, gene expression, and trait data by using controlled gene and plant ontologies to describe experimental conditions. The web portal will also contain links to software applications for association genetics (see "QTL Miner" in **Obj. 3**). Interconnecting links to plant genomic resources such as PlantGDB (Dong et al., 2004), Gramene (Ware et al., 2002), and GrainGenes (Matthews et al., 2003) facilitate access to contig alignments, oligo probe information, and a variety of gene function annotation tools from the NCBI, PlantGDB, TIGR, TAIR, or Rice genome databases.

The barley CAP project will develop mirror sites to make the data available internationally for broad bioinformatics studies. We also plan to make THT compatible with community efforts for data sharing standards such as GDBC (mentioned above) and will implement web services such as BioMoby (Wilkinson et al., 2003) to make data more accessible. We will work with the community to define the aspects of data analysis and retrieval that require detailed description. Additionally, we will explore methods of associating metadata such as experiment descriptions and analysis results with specific services.

Data management. All phenotypic, SNP marker, and pedigree data (**Obj. 3**) will be collected by the central data coordinator (J. Kling at Oregon State Univ.). Each breeding line and testing location will be assigned a unique designator, and all data collected will be recorded on Excel spreadsheets "pre-loaded" with breeding line and location designators. The central coordinator will compile the data into a single spreadsheet that will be used to upload the data to THT. Similarly, the SNP genotype data will use the same designators prior to upload to THT.

Specific Outcomes from Objective 2:

- Web-accessible pedigree, trait, SNP database integrated with gene expression datasets
- Software for SNP analysis, haplotype, and germplasm analysis and sorting
- THT development as a prototype database for other organisms

Obj. 3. Genes and Traits (Smith, Hayes) - Access economically important genes using association genetics and diverse breeding germplasm. The main goal of this objective is to utilize the database structure developed in **Obj. 2**, along with the SNP and trait data generated in this objective, to map QTL for important traits using association genetics. Many QTL have been mapped in barley (see URL 1). However, the utility of many of these QTL is limited because they are either fixed in the breeding germplasm or sufficient marker polymorphism is not

available for MAS. In this objective, we will conduct high-throughput SNP genotyping to identify SNPs associated with QTL in breeding germplasm and thus directly usable for barley improvement. Once QTL are identified, they will be posted on GrainGenes. Access to this information will facilitate MAS and germplasm exchange, and accelerate variety development.

Germplasm. Each of the ten breeding programs will identify 96 lines to include in the database in each year of the grant period for a total of 3,840 lines. These lines will be selected from a stage of the breeding program in which inbred lines are at the F₄ generation or beyond and trials are conducted in multiple locations with replication, but also at a stage early enough in the breeding process to capture significant genetic variation for the traits of interest. Data generated from the original set of 96 lines that continue on in the breeding program and are evaluated in subsequent years will be added to the database to increase the number of locations for each genotype.

All of the participating breeding programs use some form of pedigree breeding. Therefore, accurate pedigree information will exist for each of the breeding lines included in the study. The pedigree information for each line will be assembled in the THT database for later use to model population structure in mixed models for subsequent association mapping.

In addition to the breeding germplasm set, we have identified over 50 varieties and genetic stocks that will be genotyped with the same set of SNP markers. This set of genotypes includes important varieties from Canadian, European, and Australian breeding programs and mapping parents and genetic stocks that have been used in previous genetic studies. Including genotype data from these lines will enable direct comparisons between QTL identified in Barley CAP research and previous and ongoing international research efforts.

High-throughput SNP genotyping. Each breeding program will provide lyophilized leaf tissue from each breeding line in a 96 deep-well plate to the USDA genotyping Center at Fargo, ND directed by Chao. DNA will be isolated according to standard procedures in the Fargo genotyping center (URL 14). SNP data for each genotype will be obtained using the Illumina technology (see **Obj. 1**). Two 1,536-SNP Illumina OPAs, from the mapped SNPs, will be used for genotyping. Using Illumina technology (URL 8) we can process at least 3,000 SNPs on 384 plant genotypes per week, well within the capacity to carry out this project.

Barley haplotype-block structure. Large-scale SNP genotyping in humans has led to the identification of a haplotype-block structure of linkage disequilibrium (Gabriel et al., 2002): discrete genome blocks can be identified such that LD among SNP pairs within blocks greatly exceeds that between blocks. The significance of this block structure is that it allows haplotype tagging SNPs to be identified that capture a high fraction of allelic information with a greatly reduced genotyping requirement (Kruglyak, 1999; HapMap Consortium, 2003). Given the level of LD in U.S. barley germplasm (Condon and Smith, 2005), the SNP density proposed here will allow us to determine haplotype structure along the barley genome. To determine the haplotype-block structure, we will apply the LD partitioning methods pioneered in Gabriel et al. (2002) and extended by Ding et al. (2005) and Takeuchi et al. (2005). Determining the haplotype-block structure is an essential first step for determining haplotype-trait associations. P. Bradbury and J-L. Jannink will take the lead in this analysis.

Phenotypic data. Based on a survey of the ten participating breeding programs, we have identified and prioritized a list of traits that will be measured and deposited in the phenotype database. There are three main criteria for including a trait for subsequent association mapping studies. **First**, all of the traits included were identified by breeders as priorities in breeding or in

some cases important for new end-uses in barley. **Second**, there was reasonable evidence that phenotypic variation for each of the traits existed within the breeding germplasm. **Third**, there must be a sound rationale for understanding the genetics of the trait as it relates to MAS. For this consideration, some traits are very appropriate for MAS themselves due to their importance, lack of success using traditional breeding methods, or cost in phenotypic screening. Other traits such as plant height or heading date are important for all breeding programs; however, they themselves would not be targets for MAS since they can be easily scored with little cost. The rationale for including traits like these is that they may be associated with traits that are targets for MAS, such as disease resistance; therefore, understanding interactions among traits will lead to more effective MAS strategies.

The phenotypic data for the breeding germplasm will come from a combination of breeding trials and collaborative screening experiments. We will take advantage of the large amount of data that will be collected in the normal course of breeding and funded through individual breeding programs. Each year we will generate a database matrix consisting of 960 breeding lines by 38 traits. A third dimension of the matrix will be the environments/years in which the trait is measured. Some of the traits will be evaluated for all of the breeding lines while others will be evaluated for an appropriate subset of lines.

Agronomic traits. Seven agronomic traits will be evaluated for all of the lines in standard breeding program trials (yield, height, heading date, lodging, percent plump grain, test weight, grain protein concentration). In addition, the winter breeding programs will evaluate winter-hardiness. All of the data collected from breeding yield trials will be unbalanced. Therefore, to assess the effect of environment on performance of breeding lines, all of the spring breeding programs will include the checks Robust, Harrington, and Baronesse and the winter breeding programs will include Strider and 88Ab536. In addition, five sites will grow all the breeding lines. Two sites in Montana, managed by T. Blake, will evaluate lines in a drought stressed and unstressed environment. B. Cooper will plant one site in Ft. Collins, CO and another winter site in Yuma, Arizona. J. Franckowiak will examine various morphological traits that are associated with yield and in some cases disease resistance (kernels per spike, tiller per unit area, awn length, spike density, leaf width, leaf length, peduncle length, and hull adherence), as well as document the effects of photoperiod on the breeding germplasm. D. Roche will examine traits related to harvest index (above-ground dry matter, grain dry matter, number of fertile tillers, sink size per fertile tiller, 1000 kernel weight, number of seeds per spike) in a fifth location in Utah.

Malting quality. The USDA Cereal Crops Research Unit (M. Schmitt) in Madison, Wisconsin will conduct analysis of nine malting quality traits (kernel weight, barley color, malt extract, wort color, wort protein, soluble protein/total protein, diastatic power, alpha amylase, and malt beta-glucan). The BARI (B. Cooper) breeding program will generate similar data from their own analysis lab.

Disease resistance. To standardize disease-screening protocols and generate a balanced dataset, we will set up several centralized experiments to quantify disease reaction on all or a relevant subset of the breeding lines. The breeding lines and appropriate resistant and susceptible checks will be evaluated for reaction to eight diseases. L. Jackson will screen net blotch, stripe rust, barley yellow dwarf and scald in the field on all lines each year. R. Dill-Macky will also screen net blotch in the seedling stage in the greenhouse. B. Steffenson will screen the germplasm for reaction to Septoria speckled leaf blotch and spot blotch in the field. C. Griffey will screen reaction to three leaf rust pathogens. K. Smith, R. Dill-Macky, B.

Steffenson, R. Horsley, S. Neate, and B. Cooper will screen FHB on the Midwest subset (384 lines) in five field nurseries (St. Paul, MN, Fargo, ND, Osnabrock, ND, Amenia, ND, and a winter nursery Hangzhou, China).

Food Quality. The physical and chemical attributes of barley grains largely determine the processing, product quality, and potential health-related attributes of barley for its food uses. B-Y. Baik will measure grain hardness, starch amylose content, phenolic compound content and polyphenol oxidase activity. M. Wise will measure barley beta-glucan content.

Phenotyping quality control. Collection of high quality phenotypic data is essential for gene mapping research. To ensure that high quality data is obtained from the wide range of experiments proposed, each researcher will assess the quality of the experiment for which they are responsible. Only experimental data that is of high quality as determined by check performance, coefficient of variation, and examination of MSE will be sent to the central trait coordinator. Specific information about the experiment, such as ANOVA results, check performance, field or greenhouse environmental conditions will be included as an annotation to the experiment in the trait database.

Statistical analysis and “QTL Miner” software development. The tools developed in this objective will expand THT by adding association mapping functions and functions that use a combination of THT and user data to more accurately estimate the genetic values of experimental lines for selection purposes. A mixed-model approach will be used to implement these functions. This approach has been effective for gene discovery from phenotypic, pedigree, and marker data routinely generated in breeding programs (Parisseaux and Bernardo, 2004; Yu et al., 2005; Arbelbide et al., 2006; Arbelbide and Bernardo, 2006). The mixed-model approach accounts for complex pedigree relationships among lines in a breeding program as well as the unbalanced nature of phenotypic data in breeding programs. The analysis comprises three steps: (i) estimation of genetic and non-genetic variance components; (ii) single-marker analysis at a relaxed significance level to identify putative SNPs for the trait; and (iii) multiple-marker analysis at a stringent significance level to identify a final set of SNPs associated with the trait and to estimate the quantitative effects of SNP alleles. Software will be developed and implemented by R. Bernardo, referred to as “QTL Miner” to (i) extract the needed phenotypic, pedigree, and SNP data from THT and (ii) to conduct mixed-model analyses. Observed phenotypes are represented as: $y = X\beta + M\alpha + Zg + e$ where β are environmental effects and X relates the phenotypes to the environments where they were observed, α are genetic effects associated with SNPs and M relates experimental lines to their SNP genotypes, g are experimental line genetic effects not associated with SNPs and Z relates phenotypes to the experimental lines on which they were observed, and e are residual errors. The predicted genotypic values that result from this analysis are: $\hat{y} = M\alpha + Zg$. Given the estimation of variance components that takes place in the analysis, step (i) above, the analysis automatically weights the estimated SNP-associated and non-associated genetic effects to most accurately predict genetic values. If the user combines THT data with data from experimental lines that have SNP data but no phenotypic data, genotypic values are predicted by $\hat{y} = M\alpha$. Then, the estimation of the SNP-associated effects entering into α derives entirely from THT data. The estimation of genotypic values prior to expending resources in phenotyping can guide the choice of lines to increase and evaluate in the field. Simulation studies have shown this an effective use of MAS (Hospital et al., 1997).

The pedigree, phenotypic, and genotypic data present in THT will be organized by dataset. For any given analysis, the user will determine which THT datasets to include and will also have the option of incorporating datasets outside of THT. Users may wish to include or exclude specific datasets based upon their perception of the data quality, or their desire to restrict / expand the inference space to which the analysis applies. The inclusion of datasets outside of THT will allow for the analysis of proprietary or pre-publication data, as well as the application of MAS to datasets that are as yet insufficiently complete for submission to THT.

We will post summaries of QTL mapping results on GrainGenes using a new QTL resource that will be developed in the next year (see V. Carollo letter). GrainGenes curates QTL information and therefore is the logical place to put our data. Each QTL described on the page will include a description of the data sets that were used to identify the QTL and the specific SNP marker or haplotype associated with the trait. Researchers will be able relate QTL back to the gene from which the SNP was designed, physical map location, and gene expression database. Breeders will be able to use the marker information to look for polymorphisms in their own germplasm and design appropriate crosses (**Obj. 4**). The Genotyping Centers will be fully equipped to facilitate MAS for cooperating breeding programs.

Validation. Validating marker-trait associations will be accomplished in a number of ways. J. Kling will coordinate this activity. First, it will be possible to compare to previous mapping studies based on map position. SNPs identified by “QTL Miner” will be mapped on existing mapping populations since the SNPs will be screened on a large set of mapping parents. The most relevant means to validate associations, however, will be in populations developed by the breeding programs themselves. Many of the breeding lines included in the association genetics studies will have been, or will be, used as parents in breeding. Identifying the appropriate breeding populations segregating for SNPs linked to traits of interest will provide opportunities to validate these SNPs. In some cases it may be desirable to employ the SNP in MAS and validate the marker-trait association simultaneously. The following approach will be adopted in each breeding program:

1. Identify breeding populations created prior to CAP start in which the parents carried divergent SNP alleles at the identified loci. Progeny from these crosses should be at the F₄ stage but not yet phenotypically evaluated as increased F₄-derived families.
2. Genotype the F₄ plants. Identify plants that carry the favorable SNP alleles. Increase the seed of those plants. Also, increase the seed of a randomly-sampled control group of plants.
3. Evaluate MAS plants and control group and test the hypothesis that MAS identifies genotypes with higher genetic value.

Specific Outcomes from Objective 3:

- Web-based database of the location of QTL for important traits
- Haplotypes that identify desirable alleles for important traits
- Haplotype block structure of U.S. barley
- Comprehensive SNP genotypes of breeding lines to facilitate MAS and germplasm exchange
- Candidate genes for QTL
- “QTL Miner” software to facilitate mixed-model analysis of breeding program data

Obj. 4. Superior Germplasm (Smith, Hayes) - Develop elite barley germplasm using high-throughput marker-assisted selection. As specific SNP markers or haplotypes are identified that are associated with desirable QTL alleles, they will be employed in MAS by the

participating breeding programs. Identification of QTL in breeding germplasm ensures that trait-associated SNP markers will be segregating in existing populations, opening two avenues for MAS.

A “traditional” approach to MAS involves designing crosses to pyramid identified alleles into single lines (Servin et al. 2004). The GrainGenes QTL resource described above will display all potential targets for MAS in barley. Breeders will utilize existing populations or create new populations if necessary to implement MAS. This is routine work for the breeders on this project, and we see no obstacles for its implementation. A “QTL Miner” approach to MAS will involve applying the “QTL Miner” statistical models to untested or little-tested lines to estimate, and select upon, their genetic values for quantitative traits. Lines to be evaluated for selection will be typed at the relevant SNP and their genetic values estimated as described above using $\hat{y} = M\alpha$ or $\hat{y} = M\alpha + Zg$ depending on whether no or some phenotypic data is available for them. While the “traditional” approach serves to select among lines within a cross, the “QTL Miner” approach also discriminates between crosses because different crosses may be fixed for different (desirable or not) SNP alleles, which will be reflected in the \hat{y} values estimated for lines within each cross. In addition, while the “traditional” approach characterizes alleles qualitatively (desirable or not), the “QTL Miner” approach assigns a quantitative value to each allele, potentially allowing for more subtle selection discrimination.

We propose a two-tiered approach in which traditional MAS will begin as soon as robust QTL are identified. This activity will generate the necessary SNP genotypes for “QTL Miner” MAS while also validating QTL as described above. J-L. Jannink will use data generated from K. Smith’s MAS program to ask whether “QTL Miner” and “traditional” MAS would have made the same selections and to assess whether the additional operational complexity of “QTL Miner” MAS is warranted. Findings will be presented to the CAP breeders who will choose the approach most suited to their programs. Thus, the CAP will be able to initiate MAS during the funding cycle and extend these activities with iterative improvement past the end of the grant through collaboration with the USDA-funded genotyping centers.

In all cases, SNP data for MAS will be conducted in collaboration with Chao at the Fargo genotyping center as per the USDA mandate. SNPs associated with traits will be converted to PCR-based markers that individual breeders can handle and the information on map location, trait, primers and PCR conditions will be posted on THT and the QTL resource on GrainGenes.

Specific Outcomes from Objective 4:

- Implementation of MAS approaches to develop improved barley germplasm
- Assessment of the effectiveness of two approaches to MAS
- Improved barley germplasm
- Breeder-friendly markers posted on THT and GrainGenes

Obj. 5. Education and Outreach (Steffenson, Lemaux) - Educate students, scientists, farmers, processors, and consumers in use of genomics technology for gene discovery and crop improvement. Given the simple diploid nature of the barley genome and the extensive genomics resources, barley will be among the first crops to make the shift from MAS based on QTL detection in experimental bi-parental populations to MAS based on population-wide LD in breeding populations as proposed here. Barley consequently is a model system for breeding methods within the new paradigm and we propose in the barley CAP an education plan that will allow students and breeders of other species to benefit from the insights we gain over the course of the proposed research.

Education plan. Our education plan includes undergraduate, graduate student, and postdoctoral training to the maximum extent possible. Our recruiting efforts will emphasize women and other groups under-represented in science, including Native Americans, African-Americans, and Hispanics. We will advertise in widely-read publications, including *Science*, and *Chronicle of Higher Education* and in specialized web sites such as the American Indian Science and Engineering Society (URL 15), American Association for University Women (URL 16), and Association for Women in Science (URL 17). Posters and flyers announcing positions will be mailed to agronomy, genetics, computer science, biological sciences, plant pathology and statistics departments at four-year colleges and universities with 25% or greater black or Hispanic student populations.

The CAP project will provide many opportunities for student and postdoctoral projects: SNP development and mapping, bioinformatics, association genetics, linkage disequilibrium, haplotype analysis, marker-assisted selection, and software and database development. Barley CAP participants will coordinate research projects such as not to duplicate effort and provide a clear publication path for all trainees. Undergraduates will participate in summer internships, honors projects, and student symposia related to translational genomics at their institution. Graduate students and postdocs will participate in the same activities, but will also mentor undergraduate students. All trainees will participate in outreach activities by creating displays and brochures for meetings, developing websites, and attending grower/industry meetings.

Outreach to plant researchers outside barley CAP. Database and software development are of primary importance to our outreach and extension activities to plant researchers. In addition to integrating past barley research as described above, THT will provide one-stop access to all data and results from CAP research. Through “QTL Miner” software, the barley CAP will make association genetics more accessible. Currently, only highly trained statisticians have the ability to conduct association genetics studies. “QTL Miner” will include a user-friendly manual and tutorials complete with example data sets for association genetics studies in any self-pollinated crop. Periodic postings highlighting updates in these databases and software applications will be posted on GrainGenes and other websites frequented by plant scientists.

The barley CAP will conduct workshops at the Intl. Triticeae Mapping Initiative Workshops, the American Phytopathological Society meeting, the Plant and Animal Genome meeting, and the American Society of Agronomy meeting. There, breeders and geneticists will obtain up-to-date information on the genomics tools for barley improvement and tutorials on applying the tools. The workshops will also provide information on how to organize and coordinate a community of breeders and geneticists to conduct community-based gene discovery.

In the second and fourth years of the project, co-PIs Jannink and Smith will offer a multi-day short course on current approaches to MAS. Participants (students, faculty, and industry breeders) should have some training in traditional phenotype-based breeding and want to acquire association-based MAS approaches. The syllabus will include fundamentals of QTL identification, distinction between linkage and association mapping, lessons learned from QTL research and implications for MAS, and breeding plans that optimally include MAS.

Outreach to barley researchers, end users and the general public. To communicate recent results of practical importance to barley researchers and end users, members of the barley CAP will present talks at winter grower meetings, the Barley Improvement Conference, the American Society of Brewing Chemists meetings, the Barley Food Stakeholder's Meeting, and the North American Barley Researchers Workshop. Our target audience is growers, industry and public

barley breeders, pathologists, food scientists, and malting and brewing industry representatives. Our goals are to provide barley end users with information on the outcomes of the project.

We will conduct outreach specifically focused on the general public. We will develop a link on THT that will describe in lay terms the progress of the barley CAP, and descriptions and figures of genetics/genomics, breeding and biotechnology and how they relate to crop improvement. Links to other sites describing biotechnology and crop improvement will be established. Outreach will also be in the form of a dialog with audiences at field days, the American Malting Barley Association Technical Committee and Executive Board meetings, and state fairs. We will develop an interactive display (e.g., barley spikes, seed, used Barley1 Affymetrix GeneChips, barley products) with brochures and a poster for the general public.

Specific Outcomes from Objective 5:

- Graduates and post-graduates educated in translating genomics to crop improvement
- User-friendly “QTL Miner” software, user guide and tutorials released for general use
- Workshops at national meetings that describe the outcomes of the barley CAP
- Enhanced linkages with barley end users on the applied outcomes of the project
- Short courses on applied association-based MAS breeding
- A general audience link on THT, brochures, and posters on genomics and crop improvement

Management plan. Our management group includes a scientific advisory board, a stakeholder advisory board, executive committee, an overall project director, and team leaders for each objective (Figure 1). To increase the efficiency and effectiveness of the CAP, coordination with the U.S. Barley Genome Project (USBGP)-funded research projects will be a priority. P. Hayes is the coordinator of the USBGP thus ensuring appropriate coordination. Team leaders will be responsible for managing each of the individual objectives and for meeting the timeline for the goals of each objective (Table 1). Key personnel assigned to specific tasks within each objective are also shown in Table 1. Muehlbauer is the overall director of the project. A quarterly conference call between the executive committee and the overall director will be conducted to determine progress and set quarterly goals. To help coordinate the day-to-day operations, a 1/4-time project assistant will work with Muehlbauer. Scientific and stakeholder advisory boards have been formed. The scientific advisory board consists of R. Waugh (SCRI), D. Mather (U. of Adelaide), B. Rosnagel (U. of Saskatchewan), D. Marshall (SCRI), and S. Leong (USDA-ARS, Madison, WI) (see letters of intent). The stakeholder advisory board consists of M. Davis (American Malting Barley Association), M. Zutz (Minnesota barley grower), D. Ruark (Washington State barley grower), G. Hanning (Anheuser-Busch) and S. Chan (Rahr Malting) (see letters of intent). Both advisory boards will attend a yearly workshop prior to the Plant and Animal Genome Conference in San Diego, where summaries of research areas will be presented. After the workshop, the boards will meet with the project participants to provide feedback and evaluation of progress.

Data release and intellectual property. The general plan is for all data to be publicly available with “no strings attached”. The data will be housed in our publicly available THT database. All other databases used in this project are committed to making their data publicly available to researchers worldwide. All of the data and results generated from this project will be available in the public domain immediately after verification and publication.

Maintaining international cooperation. One of the strengths of the barley genomics community is the strong international collaborative spirit. An example of international

collaboration that is highly pertinent to this project is with R. Waugh (SCRI) and A. Graner (IPK-Gatersleben). Currently, Waugh, Graner, and Close are collaborating to detect and map SNP markers in barley. Waugh has also initiated a large-scale association genetics project focused on European germplasm. These projects are similar to the barley CAP objectives presented here. Our project will leverage these efforts to develop a more saturated barley genetic map and to conduct association genetics on germplasm of importance to U.S. barley growers. Our plan is for Close to coordinate with Graner and Waugh to include SNPs from their project with a set we will map for a total of 3,000 SNP markers. Thus, we are well positioned to collaborate on this project on the SNP mapping and association genetics aspects of the project.

Table 1. Project timetable.

Obj	Key personnel, Location	Task	Year 1	Year 2	Year 3	Year 4	Beyond CAP
1a	Close, UCR Hayes, OSU	SNP development and mapping	● →				
1b	Close, UCR Lonardi, UCR	Integrate SNP markers and physical map		● →			
2	Dickerson, ISU Wise, USDA-ISU	Database development and maintenance	● →				→
3a	Smith, UMN; Hayes, OSU; Chao, USDA	Trait mapping, SNP genotyping and phenotyping	● →				
3b	Kling, OSU	Data curation	● →				
3c	Bernardo, UMN	Develop QTL Miner software	● →				
3d	Bradbury, USDA Jannink, ISU	Haplotype block structure analysis		● →			
3e	Kling, OSU	QTL Validation			● →		
4	Smith, UMN; Hayes, OSU; Chao, USDA	Marker-assisted selection				● →	
5	Lemaux, UCB Steffenson, UMN	Education, extension and outreach	● →				

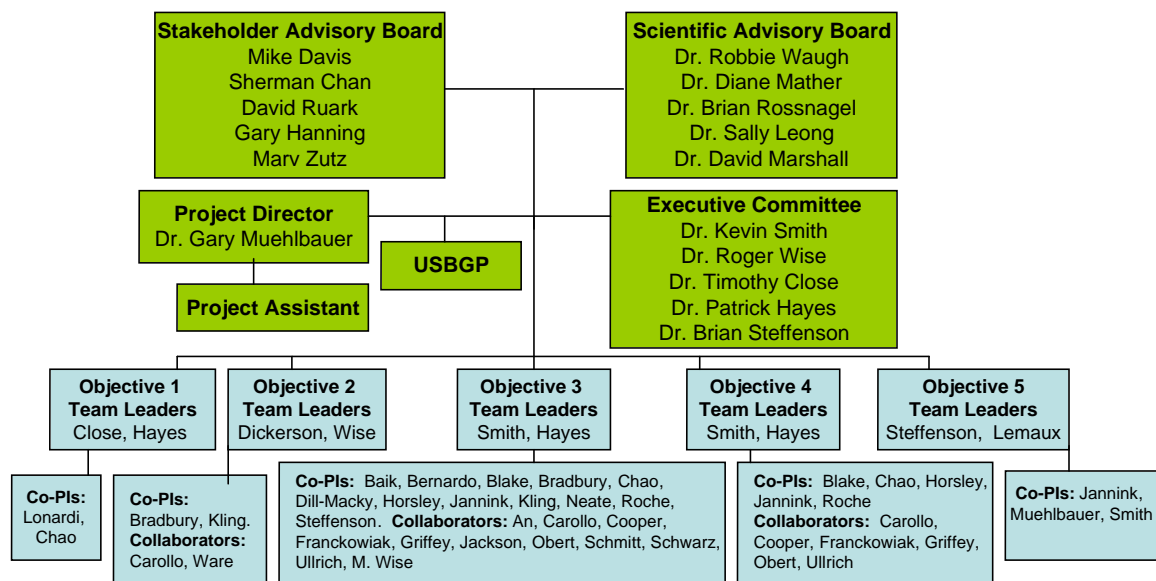


Figure 1. Barley CAP management structure. USBGP, U.S. barley genome project.

References

- Arbelbide, M., J. Yu and R. Bernardo. 2006. Power of *in silico* QTL mapping from phenotypic, pedigree and genomic data for self-pollinated crops. *Theor. Appl. Genet.* (submitted).
- Arbelbide, M. and R. Bernardo. 2006. *In silico* mapping of QTL for kernel hardness and dough strength in bread wheat. *Theor. Appl. Genet.* (submitted).
- Arumuganathan, K. and E.D. Earle. 1991. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9:208-218.
- Bernardo, R. 1996. Best linear unbiased prediction of maize single-cross performance. *Crop Sci.* 36:50-56.
- Bernardo, R. 1998. Predicting the performance of untested single crosses: Trait and marker data. p. 117-127 In K.R. Lamkey and J.E. Staub (eds.) *Concepts and breeding of heterosis in crop plants.* Publ. 25, Crop Sci. Soc. Am., Madison, WI.
- Bernardo, R. 2001. What if we knew all the genes for a quantitative trait in hybrid crops? *Crop Sci.* 41:1-4.
- Bilgic, H., B.J. Steffenson, and P. Hayes. 2005. Differential expression of seedling and adult plant resistance to spot blotch in different genetic backgrounds of barley. *Theor. Appl. Genet.* (in press).
- Caldwell, D.G., N. McCallum, P. Shaw, G.J. Muehlbauer, D.F. Marshall and R. Waugh. 2004. A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L.). *The Plant J.* 40:143-150.
- Canci, P.C., L.M. Nduulu, G.J. Muehlbauer, R. Dill-Macky, D.C. Rasmusson and K.P. Smith. 2004. Validation of quantitative trait loci for fusarium head blight and kernel discoloration resistance in barley. *Mol. Breed.* 14:91-104.
- Casstevens, T.M. and E.S. Buckler. 2004. GDPC: connecting researchers with multiple integrated data sources. *Bioinformatics* 20:2839-2840.
- Castro, A. X. Chen, A.E. Corey, T. Filichkina, P. M. Hayes, C. Mundt, K. Richardson, S. Sandoval-Islas, and H. Vivar. 2003a. Pyramiding and validation of Quantitative Trait Locus (QTL) alleles determining resistance to barley stripe rust: effects on adult plant resistance. *Crop Sci.* 43:2234-2239.
- Castro, A. F. Capettini, A. Corey, T. Filichkia, P.M. Hayes, A. Kleinhofs, D. Kudrna, K. Richardson, S. Sandoval-Islas, C. Rossi, and H. Vivar. 2003b. Mapping and pyramiding of qualitative and quantitative resistance to stripe rust in barley. *Theor. Appl. Genet.* 107:922-930.
- Chen, F., D. Prehn, P.M. Hayes, D. Mulrooney, A. Corey, and H. Vivar. 1994. Mapping genes for resistance to barley stripe rust (*Puccinia striiformis* f. sp. *hordei*). *Theor. Appl. Genet.* 88: 215-219.
- Close, T.A., S.I. Wanamaker, R.A. Caldo, S.M. Turner, D.A. Ashlock, J.A. Dickerson, R.A. Wing, G.J. Muehlbauer, A. Kleinhofs and R.P. Wise. 2004. A new resource for cereal genomics: 22K barley GeneChip comes of age. *Plant Physiol.* 134:960-968.
- Collins, H.M., J.F. Panozzo, S.J. Logue, S.P. Jeffries, and A.R. Barr. 2003. Mapping and validation of chromosome regions associated with high malt extract in barley (*Hordeum vulgare* L.). *Austr. J. Agric. Res.* 54:1223-1240.
- Condon, F. and K.P. Smith. 2005. Linkage disequilibrium and marker-trait associations within six-rowed barley breeding germplasm. In: *Plant and Animal Genome XIII Abstracts*, San Diego, CA. Jan 15-19, 2005, p. 153.

- Cooper L., L. Marquez-Cedillo, J. Singh, A. K. Sturbaum, S. Zhang, V. Edwards, K. Johnson, A. Kleinhofs, S. Rangel, V. Corolla, P. Bregitzer, P.G. Lemaux, and P.M. Hayes. 2004. Mapping Ds insertions in barley using a sequence-based approach. *Mol. Gen. Genomics* 272: 181–193.
- Coventry, S.J., H.M Collins, A.R. Barr, S.P Jeffries, K.J. Chalmers, S.J. Logue, and P. Langridge. 2003. Use of putative QTLs and structural genes in marker assisted selection for diastatic power in malting barley (*Hordeum vulgare* L.). *Aust. J. Agric. Res.* 54:1241-1250.
- de la Peña, R.C., K. Smith, F. Capettini, G.J. Muehlbauer, M. Gallo-Meagher, R. Dill-Macky, D.A. Somers and D.C. Rasmusson. 1999. Quantitative trait loci associated with resistance to *Fusarium* head blight and kernel discoloration in barley. *Theor. Appl. Genet.* 99:561-569.
- Ding, K., K. Zhou, J. Zhang, J. Knight, X. Zhang and Y. Shen. 2005. The effect of haplotype-block definitions on inference of haplotype-block structure and htSNPs selection. *Mol. Biol. Evol.* 22:148-159.
- Dong, Q., S.D. Schlueter and V. Brendel. 2004. PlantGDB, plant genome database and analysis tools. *Nucl. Acids. Res.* 32: D354-359.
- Franckowiak, J. 1997. Revised linkage maps for morphological markers in barley, *Hordeum vulgare*. *Barley Genetics Newsletter* 26:9-21.
- Gabriel, S.B., S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S.N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E.S. Lander, M.J. Daly, and D. Altshuler. 2002. The structure of haplotype blocks in the human genome. *Science* 296:2225-2229.
- Gaut, B.S., and Long, A.D. 2003. The lowdown on linkage disequilibrium. *The Plant Cell* 15:1502-1506.
- Hayes, P.M., A.E. Corey, R. Dovel, R. Karow, C. Mundt, K. Rhinart, and H. Vivar. 2000. Registration of Orca barley. *Crop Sci.* 40:849-851.
- Hayes, P.M., A.E. Corey, C.Mundt, T. Toojinda, and H. Vivar. 2003a. Registration of Tango barley. *Crop Sci.*: 729-731.
- Hayes, P.M., A. Castro, L. Marquez-Cedillo, A. Corey, C. Henson, B.L. Jones, J. Kling, D. Mather, I. Matus, C. Rossi, and K. Sato. 2003b. Genetic Diversity for Quantitatively Inherited Agronomic and Malting Quality Traits. *In* R. Von Bothmer, H. Knupffer, T. van Hintum, and K. Sato (ed.). *Diversity in Barley*. Elsevier Science Publishers, Amsterdam.
- Henderson, C.R. 1984. Applications of linear models in animal breeding. Univ. of Guelph Press, Guelph, Canada.
- Holland, J.B. 2004. Implementation of molecular markers for quantitative traits in a breeding program - challenges and opportunities. Proceedings of the 4th International Crop Science Congress, Brisbane, Australia.
- Holloway, J. and S. Knapp. 2001. GMendel. <http://cropandsoil.oregonstate.edu/G-mendel/Default.htm>.
- Hospital, F., L. Moreau, F., Lacoudre, A. Charcosset and A. Gallais. 1997. More on the efficiency of marker-assisted selection. *Theor. Appl. Genet.* 95: 1181-1189.
- Huang, X. and A. Madan. 1999. CAP3: A DNA Sequence Assembly Program. *Genome Res* 9:868-877.
- Jannink, J-L., M.C.A.M. Bink and R.C. Jansen. 2001. Using complex plant pedigrees to map valuable genes. *Trends in Plant Science* 6:337-342.
- Jorde, L.B. 2000. Linkage disequilibrium and the search for complex disease genes. *Genome Res.* 10:1435-1444.

- Kennedy, B.W., M Quinton, and J.A.M. van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. *J. Anim. Sci.* 70:2000–2012.
- Kleinhofs, A., A. Kilian, M.A. Saghai Maroof, R.M. Biyashev, P. Hayes, F.Q. Chen, N. Lapitan, A. Fenwick, T.K. Blake, V. Kanazin, E. Ananiev, L. Dahleen, D. Kudrna, J. Bollinger, S.J. Knapp, B. Liu, M. Sorrells, M. Heun, J.D. Franckowiak, D. Hoffman, R. Skadsen, and B.J. Steffenson. 1993. A molecular, isozyme and morphological map of barley (*Hordeum vulgare*) genome. *Theor. Appl. Genet.* 86:705-712.
- Kleinhofs, A. and A. Graner. 2001. An integrated map of the barley genome. In: R.L. Phillips and I.K. Vasil (Eds.), *DNA-Based Markers in Plants*, Kluwer Academic Publishers.
- Kraakman, A.T.W., R.E. Niks, P.M.M.M. Van den Berg, P. Stam, and F.A. Van Eeuwijk. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168:435-446.
- Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 22:139-144.
- Lande, R. and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743-756.
- Matthews, D. E., V. L. Carollo, G. R. Lazo and O. D. Anderson. 2003. GrainGenes, the genome database for small-grain crops. *Nucl. Acids. Res.* 31: 183-186.
- McMullen, M, R. Jones and D. Gallenberg. 1997. Scab of wheat and barley: a re-emerging disease of devastating impact. *Plant Dis.* 81:1340-1348.
- Mesfin, A., K.P. Smith, R. Dill-Macky, C.K. Evans, R. Waugh, C.D. Gustus and G.J. Muehlbauer. 2003. Quantitative trait loci for Fusarium head blight resistance in barley in a two-rowed by six-rowed population. *Crop Sci.* 43:307-318.
- Moreau, L., S. Lemarie, A. Charcosset and A. Gallais. 2000. Economic efficiency of one cycle of marker-assisted selection. *Crop Sci.* 40: 329-337.
- Morell, P.L., D.M. Toleno, K.E. Lundy and M.T. Clegg. 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proc. Natl. Acad. Sci. USA* 102:2442-2447.
- Parisseaux, B. and R. Bernardo. 2004. *In silico* mapping of quantitative trait loci in maize. *Theor. Appl. Genet.* 109:508-514.
- Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Rafalski, A. and M. Morgante. 2004. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics* 20:103-111.
- Ramsay, L., M. Macaulay, S.D. Ivanissevich, K. MacLean, L. Cardle, J. Fuller, K.J. Edwards, S. Tuveesson, M. Morgante, A. Massari, E. Maestri, N. Marmiroli, T. Sjakste, M. Ganal, W. Powell and R. Waugh. 2000. A simple sequence repeat-based linkage map of barley. *Genetics* 156:1997-2005.
- Rostoks, N., J.O. Borevitz, P.E. Hedley, J. Russell, S. Mudie, J. Morris, L. Cardle, D.F. Marshall and R. Waugh. 2005. Single-feature polymorphism discovery in the barley transcriptome. *Genome Biology* 6:R54.
- Servin, B., O.C. Martin, M. Mézard and F. Hospital. 2004. Toward a theory of marker-assisted gene pyramiding. *Genetics* 168:513-523.
- Shen, L., J. Gong, R.A. Caldo, D. Nettleton, D. Cook, R.P. Wise and J.A. Dickerson. 2005. BarleyBase--an expression profiling database for plant genomics. *Nucl. Acids Res.* 33:D614-618.

- Takeuchi, F., K. Yanai, T. Morii, Y. Ishinaga, K. Taniguchi-Yanai, S. Nagano and N. Kato. 2005. Linkage disequilibrium grouping of single nucleotide polymorphisms (SNPs) reflecting haplotype phylogeny for efficient selection of tag SNPs. *Genetics* 170:291-304.
- Tang, X., L. Shen and J.A. Dickerson. 2005. BarleyExpress: a web-based submission tool for enriched microarray database annotations. *Bioinformatics* 21:399-401.
- Thiel, T., W. Michalek, R.K. Varshney and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106:411-422.
- Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S. Nielsen, D., and Buckler, E.S. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. *Nat. Genet.* 28:286-289.
- Toojinda T., E. Baird, A. Booth, L. Broers, P. Hayes, W. Powell, W. Thomas, H. Vivar, and G. Young. 1998. Introgression of quantitative trait loci (QTLs) determining stripe rust resistance in barley: an example of marker-assisted line development. *Theor. Appl. Genet.* 96:123-131.
- Van Ooijen, J.W. and R.E. Voorrips. 2001. JoinMap 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands.
- Ware, D., P. Jaiswal, J. Ni, X. Pan, K. Chang, K. Clark, L. Teytelman, S. Schmidt, W. Zhao, S. Cartinhour, et al. 2002. Gramene: a resource for comparative grass genomics. *Nucl. Acids. Res.* 30: 103-105.
- Wilkinson, M. D., D. Gessler, A. Farmer and L. Stein. 2003. The BioMOBY project explores open-source, simple, extensible protocols for enabling biological database interoperability. *Proceedings Virtual Conference on Genomics and Bioinformatics* 3.
- Wilson, L.M., S.R. Whitt, A.M. Ibáñez, T.R. Rocheford, M.M. Goodman and E.S. Buckler IV. 2004. Dissection of maize kernel composition and starch production by candidate gene association. *The Plant Cell* 16:2719-2733.
- Yu, J., M. Arbelvide and R. Bernardo. 2005. Power of *in silico* QTL mapping from phenotypic, pedigree and marker data in a hybrid breeding program. *Theor. Appl. Genet.* (in press).
- Yu, Y., J.P Tomkins, R. Waugh, D.A. Frisch, D. Kudrna, A. Kleinhofs, R.S. Brueggeman, G.J. Muehlbauer, R.P. Wise and R.A. Wing. 2000. A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. *Theor. Appl. Genet.* 101:1093-1099.
- Zheng, J., T. Close, T. Jiang and S. Lonardi. 2004. Efficient selection of unique and popular oligos for large EST databases. *Bioinformatics* 20:2101-2112.

URL References

1. <http://barleyworld.org/>
2. <http://wheat.pw.usda.gov/>
3. <http://harvest.ucr.edu/>
4. <http://barleybase.org/>
5. <http://plexdb.org/>
6. <http://www.plantgdb.org/>
7. <http://gramene.org>
8. <http://phymap.ucdavis.edu:8080/barley/index.jsp>
9. http://www.illumina.com/Products/prod_snp_ilmn
10. <http://www.hapmap.org/>
11. <http://barleygenomics.wsu.edu/>

12. <http://www.maizegenetics.net/gdpdm>
13. <http://www.panzea.org/>
14. http://maswheat.ucdavis.edu/protocols/general_protocols/DNA_extraction_003.htm
15. <http://aises.org/index.html>
16. <http://www.aauw.org/>
17. <http://www.cs.yale.edu/homes/tap/sci-women-groups.html>