Coordinated Agricultural Project (WheatCAP): "Validation, characterization and deployment of QTL for grain yield components in wheat"

Program Area Priorities Addressed: This project will establish a nationally coordinated consortium of public wheat breeders, molecular geneticists, high-throughput genotyping laboratories, database experts and educators focused on increasing wheat yields. Our research will identify and deploy beneficial alleles for genes affecting grain yield components, leading to more productive varieties. In the area of education, this project will train a cohort of 15 PhD students in plant breeding, integrating field, laboratory and bioinformatics skills. This project addresses specific objectives of Program Area Code A1142. In the research area, this project addresses the need to significantly increase grain production in diverse environments by identifying the genes regulating the spike/grain sink dynamics and by developing genetic tools that facilitate breeding of improved grain yield. This objective perfectly overlaps with the overarching IWYP goal of significantly increasing genetic yield potential of wheat. This project also addresses the educational need to train the next generation of plant breeders.

A. INTRODUCTION

The 700 million tons of wheat produced each year provides humanity more than one fifth of the calories and protein [1]. Increases in global wheat production required to feed a growing population are hampered by limited knowledge of the genes controlling wheat yield. The identification of these genes is a necessary first step to understand how their epistatic interactions and interactions with the environment shape the pathways that regulate yield [2].

The tools required for the identification of genes underlying QTL for yield components are now available. The assembly of the non-repetitive regions of the wheat genome provides a valuable reference, whereas genotyping by sequencing (GBS) and exome capture platforms provide tools for the rapid identification of genome-wide DNA polymorphisms. TCAP researchers have also sequenced tetraploid and hexaploid wheat mutant populations and developed a database including >10 million induced mutations in wheat coding regions (registration required: http://dubcovskylab.ucdavis.edu/wheat_blast). These populations, together with genome editing technologies, allow a rapid functional characterization of candidate genes. These tools also provide a unique opportunity to assess phenotypic effects of genes discovered in diploid grass species that are masked by duplicated homoeologs in polyploid wheat [3].

Natural variation of grain yield components can be used to identify candidate genes. The previous CAP grants developed association mapping panels (AM), nested association mapping (NAM) populations, and dedicated bi-parental populations, which have been genotyped and phenotyped in multiple environments. Several QTL for grain yield components have been identified. The application of new genomic tools to these well-characterized populations provides a unique opportunity to clone the causal genes. Wheat is today in a situation similar to that of rice after the release of the first draft genome in 2005, which resulted in an exponential increase in the number of cloned QTL in rice [4].

A long-term constraint to future increases in wheat production in the USA is the limited number of trained plant breeders. In the TCAP we developed the Plant Breeding Training Network that is now available to train a new cohort of plant breeders. Our group has the unique advantage of active public breeding programs within the universities, which is an essential resource to provide students in plant breeding an integrated training, including field and laboratory experiences.

A.1. Selection of leading and participating institutions

Dr. J. Dubcovsky (UCD) was selected to lead this project based on his experience in leading previous WheatCAP (2006-2010) and TCAP projects (2011-2016). He has demonstrated experience in positional cloning of wheat genes and QTL [5-11]. The co-PD, Dr. E. Akhunov (KSU), brings extensive experience in wheat genomics [12] and positional cloning [10]. Dr. Akhunov led the development of the high-throughput SNP and exome capture platforms in the previous TCAP project [13, 14]. Dr. J-L. Jannink developed the T3 database during the last TCAP and will continue its development and expansion in this project. All four USDA genotyping laboratories are part of this project. Dr. J. Cook (MSU) was selected to coordinate the student training based on his experience as a WheatCAP student, and close collaboration with education advisor Dr. J. Sherman (education leader in WheatCAP and TCAP). L. Sandalls (NE) brings expertise in online education environments such as eLibrary and eXtension.

All US public wheat breeding programs were invited to submit a two-page proposal that was evaluated by a committee elected by unanimous vote at the PAG 2016 meeting. Applications from 15 different states were submitted and all were accepted. Expertise on positional cloning differs among programs, but we decided to maximize the number of programs and wheat growing regions that will benefit from these resources. Centralized training workshops will be provided to the PhD students so they can all benefit from the collective group expertise.

A.2. Relevance to stakeholders

Surveys of state wheat-growers associations have repeatedly shown that grain yield is the main priority for wheat growers. Increases in kernel weight are highly correlated with increases in flour yield and will also benefit grain millers. Students from previous wheat CAP projects are working as breeders in different crops benefiting the whole breeding industry. Attached support letters document the high relevance of this proposal to U.S. Wheat Associates, the National Association of Wheat growers (signed by 24 state wheat commissions), the North American Millers Association, the American Bakers' Association and private breeding companies.

A.3. Role of stakeholders in problem identification

The TCAP project included an Industry Liaison Group that received annual reports and sent representatives to the annual TCAP meetings. In the last annual meeting, TCAP researchers and industry liaisons discussed the research priorities for this project. Members of this project are well aware of grower's priorities through their participation in the National Wheat Improvement Committee, which coordinates research priorities with NAWG and US Wheat Associates. Finally, most of the public wheat breeding programs are supported by state growers associations that tax themselves to support wheat breeding efforts. Thus, addressing the needs of the wheat growers is a top priority for the public wheat breeding programs. The focus of this project on yield potential has the enthusiastic support of the wheat industry as documented in the attached letters.

A.4. Preliminary data

***A.4.1. Genotyped and phenotyped populations*:** The previous TCAP (2011-2016) genotyped and phenotyped the core wheat collection at the NSGC, spring and winter wheat AM panels, NAM populations, and dedicated bi-parental populations (Table 1). AM panels of adapted germplasm are based on recombination accumulated throughout the evolutionary history of the populations and are useful to identify QTL and candidate genes for the targeted traits [15]. Many SNP marker-trait associations have been already identified in TCAP panels and bi-parental populations listed in Table 1. For many of these SNPs, high-throughput KASP assays have been

developed through collaborations between the genotyping labs and the TCAP students. GBS pipelines have been already established in the genotyping labs and are now routinely used to characterize germplasm and segregating populations.

Table 1. Available segregating populations (already genotyped and phenotyped).

Population	Genotype	Phenotyping yield	QTL
Spring AM PPD sensitive	90 K	10 env. MT, WA, ID, KS, Sask.	Y
Spring AM PPD insensitive	90 K	10 env. CA (2 env.) & CIMMYT	Y
Spring NAM	90 K-GBS	6 env. CA, MT, WA	Y
NSGC Spring AM	9K	2 env. ID	Y
NSGC Winter AM	9K	2 env. ID	Y
Hard Winter AM	90 K-GBS	12 env. CO, KS, NE, OK and TX	Y
Yield QTL validation panel	90K	7 env. NE	Y
Soft Winter AM	90 K-GBS	18 env. AR, KY, MD, MO, OH, VA	Y
Hard Winter NAM	90 K-GBS	Pending	-
Soft Winter NAM	90 K-GBS	Pending	-
Bi-parental populations	90 K/GBS	http://maswheat.ucdavis.edu/Mapping/.	Y

***A.4.2. Exome capture platform*:** Two wheat capture platforms are available. The first one is focused in low copy number regions [16]. The assay targets 107 Mb of non-redundant low-copy sequence and has been used for the development of a haplotype map, for imputing genotypes, for the characterization of global genetic diversity, and for cross-referencing different SNP datasets. A bioinformatics pipeline was developed to produce genotype calls with >98.9% accuracy [12]. The second capture assay includes 286,800 exons corresponding to 82,500 genes. The final capture design includes 84 Mb and results in average captures of 111.8 Mb in tetraploid wheat and 162.4 Mb in hexaploid wheat. Roughly 70% of the mapped reads are on target. With this tool, wheat genes can be sequenced for the same cost of sequencing one Arabidopsis genome.

***A.4.3. Sequenced TILLING mutant populations*:** Using a published bioinformatics pipeline [17] and in collaboration with the John Innes Centre, we sequenced 1,209 mutants in hexaploid wheat and 1,536 in tetraploid wheat. We identified 4,300,000 mutations in tetraploid wheat (38 mut./kb) and 7,300,000 mutations in hexaploid wheat (45 mut./kb). This mutation density is sufficient to find truncations in ~75% of the genes and amino acid changes in >98% of the genes. This information is organized in a BLAST searchable database. This tool greatly accelerates validation of candidate genes and studies of gene function in wheat.

***A.4.4. CRISPR-Cas9 implementation in wheat*:** CRISPR-Cas9 technology allows the rapid knock-out of homoeologs. UCD, KSU and OSU have transformation facilities that efficiently transform wheat. CRISPR-Cas9 transgenic wheat plants have been generated at these three facilities. The Akhunov lab developed a wheat codon-optimized version of CAS9 nuclease flanked by two nuclear localization signal peptides and a single guide RNA (sgRNA) sequence fused with the wheat *U6* promoter. Both fragments were subcloned into pHSG399 and the maize ubiquitin (Ubi) promoter and the nopaline synthase terminator (NOS) were used. The CRISPR-Cas9 system was validated by constructs targeting the dwarfing gene *Rht1*.

***A.4.5. The Triticeae Tool box database (T3)*:** Data in the Triticeae Toolbox (T3) currently include 13,725 wheat lines evaluated in 334 phenotypic trials for 147 traits (March 2016). A total of 9,706 lines have genotypic data (164 million data points) and 9,417 have phenotypic data

(439,874 data points). All markers annotated in T3 have been BLASTed against the wheat genome reference to create sequence-based physical consensus maps. A JBrowse tool is now available from T3 to explore the marker's genomic context. Researcher may conduct a GWAS on T3 and simply access gene annotations in the target region and find linked markers. T3 remains the primary portal to data generated over the past five years by TCAP. The information already available in T3 can greatly accelerate the discovery of candidate genes in this project.

***A.4.6. The Plant Breeding Training Network (PBTN)*:** TCAP developed an online training environment to deliver webinars and courses that helped integrate the training of students located in different institutions. Previous CAP educational activities provided us with expertise in the organization of face-to-face workshops, collaboration with the plant breeding industry, organization of student poster sessions, and organization of educational trips to international research centers. PBTN provided valuable education and communication tools and will be used to deliver new content and provide communication support in the proposed project. An external evaluation team concluded that students that participated in the TCAP education program were exposed to more experiences and were more broadly trained than those that did not participate.

B. RATIONALE AND SIGNIFICANCE

The previous WheatCAP and TCAP projects have brought most public wheat breeding programs, molecular genetics researchers and high throughput genotyping laboratories to work synergistically to improve wheat. The initial WheatCAP empowered US public wheat breeding programs to perform QTL mapping studies in relevant breeding populations. This new project will empower breeders to identify the causal genes of important agronomic traits. While private breeding programs focused on maximizing yield, the public wheat breeding programs have a broader mission that includes research and training students. The current projects balances these three objectives by providing 15 graduate students the opportunity to advance our understanding of the genes regulating grain yield, to have access to modern genomic tools, and to contribute to the breeding programs by deploying the beneficial alleles.

The wheat CAP efforts have favorably impacted the wheat industry through the release of >150 improved varieties and germplasm. In 2014, wheat varieties developed by wheat CAP covered 8.5 million acres in the USA with a value of $1.8 billion/year. The continuation of these long-term collaborative projects among public wheat breeding programs is essential to coordinate activities, avoid research duplications among programs, and centralize the storage, organization and analyses of genotypic and phenotypic data generated by the breeding programs. This project will contribute to the development of a central phenotype/genotype database for the US wheat, which we propose expanding to serve as the central database for public breeding programs.

The increased genomics resources and high throughput marker technologies available today have shifted the paradigm in wheat genetics research from finding QTL to identifying the causal genes and SNPs. This information is useful to screen germplasm for novel alleles and to determine precisely which lines in a breeding program have which allele. The identification of causative genes also provides perfect diagnostic markers, enhances the predictive value of the molecular markers, and facilitates the development of varieties combining desirable traits. Cloning the genes controlling grain yield is required to take advantage of new CRISPR technology, 'the new frontier of genome engineering' [18].

As we clone QTL and we describe their characteristics, we are better able to discern QTL signal and reduce false positive rates. This new knowledge will enable us to work in parallel on many QTL. The benefit from cloning these QTL may not come only from their direct incorporation

into new breeding lines but also from initiating a method of breeding that enables us to engineer multiple QTL in parallel. This project will allow us to test this hypothesis. The more precise mapping of the QTL and the identification of the predictive haplotypes will help us to incorporate that information into genomic prediction models. In addition, the improved characterization of these QTL and their interactions with different environments and germplasm will allow breeders to developed better strategies to use them in their breeding programs.

The loss of plant breeding positions in the public sector has resulted in a loss of the infrastructure and expertise that supports plant breeding training, and a loss of a critical mass of students necessary to provide a stimulating learning environment [19-24]. The situation has been exacerbated as the demand for plant breeders in the private sector has increased. A 2010 Delphi study [24] articulates the complexity of training plant breeders, highlighting not only scientific content but also hands-on experience in scientific inquiry and the development of interpersonal skills. Shared training is an appropriate response to these challenges [22]. By combining the resources of multiple public breeding programs, we will cover the expertise required to efficiently train modern plant breeders [25].

Students trained during the previous TCAP represent a small proportion of the current needs for plant breeders in the US, so we expect a high demand for the students trained in this program (see industry support letters). Public wheat breeding programs provide a unique opportunity for training students interested in the field and lab techniques required for modern plant breeding. Proficiency in all subjects important to plant breeding is rarely found in a single institution. The diverse group included in this project and the available PBTN communication tools will provide students access to information in different fields of plant breeding and a rich community of peers. This is expected to reduce student isolation and improve collaboration, two factors known to have a positive impact on learning [26-28]. In summary, this project brings together the wheat breeding, research and educational communities to generate the research, tools, germplasm, and human capital required to accelerate gains in wheat maximum yield potential.

C. OBJECTIVES

Overall hypothesis or goal: The overall goals for this project are the validation, characterization and deployment of QTL for grain yield components in wheat and the training of a new generation of plant breeders.

The specific objectives of the project are to:

1) Characterize a set of 15 QTL for grain yield and identify the underlying genes.

2) Validate the candidate genes using mutants and transgenic approaches.

3) Deploy beneficial alleles in commercial varieties and advanced breeding lines.

4) Develop genomics tools to characterize the regulatory regions of the wheat genome.

5) Train a new cohort of 15 plant breeders, and make the training resources widely available.

D. EXPERIMENTAL METHODS

D.1. Methods for Objective 1

***D.1.1. General strategy for QTL cloning*:** The cloning of a QTL using a map-based approach requires precise mapping of the QTL effect, which can be achieved by reducing the genetic and environmental variability in segregating populations. To reduce the genetic variability among segregating lines we will use heterogeneous inbred families (HIFs) [29]. Once a QTL is

validated in a bi-parental population, individuals heterozygous for the target regions will be selected from advanced generations. For example, in the sixth generation of a SSD population 3.1% of the markers are still in the heterozygous state, allowing identification of heterozygous individuals for selected QTL. We will use more than one HIF since epistatic interactions may result in better detection of the QTL in some genetic backgrounds than others. After selecting heterozygous markers flanking the QTL for the initial HIFs, we will continue to self-pollinate the heterozygous lines to improve isogenicity in the second phase of the high-density mapping. An isogenic background eliminates most of the epistatic interactions and reduces variation within each allelic class, leading to improved statistical power.

We will adjust the number of replications based on the observed variability between and within isogenic lines to maintain high statistical power (>90%). Lines with critical recombination events will be self-pollinated and homozygous recombinant and non-recombinant sister plants will be used to produce seeds for field experiments. To maintain a reasonable number of recombinant lines in the field evaluations, we will use a *two stage high-density mapping strategy*. Based on the distance between the closest flanking markers for each QTL, we will adjust the initial population to generate ~50 lines with recombination events in the targeted region. Precise phenotyping of these recombinant lines and new marker development will be used to delimit the QTL region to sub-cM intervals. In the second phase of high-density mapping we will screen ~6,000 segregating chromosomes (3,000 near isogenic $F_{6:2}$ segregating plants). In regions with normal recombination, this is usually sufficient to generate recombination events delimiting regions with 1-5 candidate genes [5, 6, 8-10]. The flanking markers will be used to delimit the QTL region in the current assembly of the wheat genome and to identify candidate genes.

To saturate the target regions with markers we will use the exome and regulatory capture platforms described in sections A.4.2 and D.4.1.1 to identify a large number of polymorphisms in the selected HIFs. These lines are heterozygous for the target QTL region and homozygous for most of the genome, so segregating polymorphisms will be concentrated in the target region. This will be validated by comparing the sequences of the polymorphic markers with the current wheat genome sequence. To improve the number of mapped reads, we will expand the reference used in each population by re-assembling un-mapped reads from captures from four independent individuals. The use of multiple independent captures reduces the noise generated by off-target captures relative to the signal generated by the on-target captures. The use of this strategy in the Kronos and Cadenza TILLING populations increased the number of mapped reads from 92% in the original reference to 98.2% in the expanded reference. This procedure has the advantage of adding paralogs that might be absent in the original reference.

Once reads are mapped to the improved references, the KSU program will generate a list of polymorphisms in the target regions and deliver them to the respective PhD students. The list of potential candidate genes and putative causal mutations will be prioritized based on their significance in AM panels, known function of the genes in other species and expression profiles [30, 31]. We will integrate information from the MNase assays (section D.4) with polymorphisms identified in regulatory regions to prioritize those with potential effects on gene expression (see section D.2 for methods and D.4.2 for the T3 tools). The genotyping laboratories will collaborate with the PhD students in the development of KASP markers. These markers will be used to genotype critical recombinant lines and delimit the candidate gene regions. Information will be deposited in T3 and integrated with GrainGenes.

***D.1.2. Summary of targeted QTL by program*:** Wheat includes different market classes adapted to diverse environments, providing a unique opportunity to characterize genotype-by-

environment interactions affecting yield components in great detail. Different programs are at different stages of the QTL mapping strategy described above, but all have already validated the selected QTL in bi-parental populations and are working on the development of HIFs to initiate the high resolution maps. One important contribution of this project is the coordination of the QTL cloning targets at a national level, avoiding unnecessary duplications, bringing together diverse expertise, and generating collaborations among programs interested in similar QTL. For example, the QTL for spikelet number targeted by the CA program was also detected in the mapping populations from CO, KS, MN, and MT, providing the opportunity of validation across multiple environments. The QTL prioritized by the different programs are described below.

AR: The PhD student under the supervision of E. Mason will focus on a QTL for grain yield and spike weight on chromosome arm 1AL located between markers IWA6341 and IWA531 (227.4 to 231.4 Mb). This QTL was first identified in the bi-parental mapping population Pioneer brand 26R61 x AGS2000 (PA) evaluated in 12 environments and was confirmed in an association mapping panel including 240 soft red winter wheat (SRWW) genotypes in 8 environments. This QTL (LOD 4.8, P=1.4E^{-05}) explains up to 8% of the variation in grain yield and is associated with increased kernel number per spike. Lines carrying the AGS2000 allele showed on average 116.5 kg ha^{-1} higher grain yield than those carrying the P26R61 allele. Validation of this QTL is in its second season using a population of 360 adapted breeding lines genotyped with GBS and KASP markers. RILs from the PA population heterozygous at IWA6341 and IWA531 will be used for developing HIFs for fine mapping. AR will also collaborate with MT on the 3BL QTL which was also found to increase grain yield and biomass in HRSW.

CA: The UCD PhD student will focus on a QTL for number of spikelets per spike (SPS) located on chromosome arm 7AL (161-163 Mb). This QTL was identified in five environments in two GWAS studies: one including 875 spring wheat lines from the NSGC and the other one including 240 spring / photoperiod-insensitive advanced breeding lines. The QTL was validated in the Berkut x RAC875 NAM population (LOD 3.3, 20% of variation explained, and heritability 0.88). On average, lines carrying the Berkut allele have 2.4 more spikelets per spike than those carrying the RAC875 allele. We have already identified two F_6 lines from this NAM population that are heterozygous for the QTL region and the $F_{6:2}$ are being grown to validate the phenotype and generate >1,000 $F_{6:2}$ HIFs for the high-density map. QTL in the same position have been detected in segregating populations from CO, KS, MN and MT, supporting the value of this QTL as a target for positional cloning. The PhD student at CA will coordinate collaborations among these programs to identify and deploy this gene.

CO: The PhD student at CSU will be jointly supervised by S. Pearce, P. Byrne and S. Haley to characterize a QTL on chromosome 6BL (position 145-149Mb) which affects grain width (h^2=0.66) and thousand kernel weight (h^2=0.586). This QTL was identified from a GWAS using the hard winter wheat association mapping panel (HWWAMP) and was significant in three Great Plains environments (LOD 4.65, R^2 = 4.7%). Lines carrying the beneficial allele exhibit an average increase of 0.059 mm in grain width and 1.12 g per thousand grain weight. This QTL region was validated in the Platte/CO94610 RIL population in two environments. To identify HIFs to map this QTL, we are developing KASP marker to identify lines heterozygous for the peak SNPs within the Platte/CO940610 RIL population, which consists of 186 $F_{5:6}$ lines. This population is also segregating for the 7A QTL targeted by CA, and will be part of ongoing collaborations between CO and CA. The 7AL QTL is also significant in a GWAS performed in the HWWAM panel.

ID: The ID PhD student will focus on a major QTL for grain yield identified on chromosome arm 1BL (*Q.Gy.ui-1B.2*) in the RIL population from the cross 'Rio Blanco' (PI 531244) x 'IDO444' (PI 578278). Rio Blanco is a high-yielding HWW wheat under irrigation and IDO444 is a HRW wheat with improved grain yield under rainfed conditions. The RIL population included 338 F_5 lines, of which 159 lines have been advanced to $F_{8:10}$. *Q.Gy.ui-1B.2* was detected in six testing environments and a preliminary characterization has been published [32]. *Q.Gy.ui-1B.2* accounts for up to 22% variation for grain yield. Average grain yields of the RIL lines carrying the IDO444 allele were 123 - 330 kg ha^{-1} higher than those from the lines carrying the Rio Blanco allele. Two F_5 lines heterozygous for the QTL region were identified and will be used to generate $F_{5:2}$ populations to validate the QTL and generate a high-density map.

KS: The KS PhD student will focus on the deployment and dissection of a QTL for grain yield identified in the QTL Validation Panel on chromosome 2DL (Peak: *IWA8562*, position: 111.02 Mb). Yield-related meta-QTL in this region have been identified [32], supporting the value of this QTL. Two RIL populations, Overland/Overley and Lyman/Overley, are available for the high-density mapping. These three varieties are important commercial cultivars in KS, NE and SD. On average, inbred lines with the Overley allele yield 7% above lines carrying the Overland/Lyman allele. To develop HIF in both photoperiod-sensitive and –insensitive backgrounds (to test in different locations), F_3 RILs of Overland/Overley and F_4 RILs of Lyman/Overley are being screened to identify families homozygous for *Ppd-B1a* and *Ppd-B1b* and heterozygous for the 2DL target region (*BS00043082-IWA8562-IWB987*). Progeny of these individuals will be grown to generate >1,000 $F_{4:2}$ HIFs for each of the *Ppd-B1* alleles for high density mapping. The 2nd phase of the high-density map will use more advanced $F_{5:2}$ HIFs.

MI: The MI PhD student will be advised by Dr. E. Olson. A D genome nested association mapping population was produced from direct hybridization of seven *Aegilops tauschii* accessions with the hexaploid wheat genotype, KS05HW14. A grain yield QTL was identified in the distal region of chromosome arm 2DL from *Ae. tauschii* accession TA1615 (162.9 cM in reference D genome) in 420 RILs tested in KS and MI. This QTL explains 7% of phenotypic variation (P=3.5x10^{-04}) in grain yield in both KS and MI. Heritability for grain yield in this trial was 0.67. LD in the QTL region extends from 208.2 to 209.1 Mb on the *Ae. tauschii* reference pseudomolecule. A set of 14 RILs have been identified that are heterozygous across the 2DL QTL region and are being self-pollinated to develop HIFs for high resolution mapping. Collaboration has been established with SD and yield testing will be done in MI and SD.

MN: The MN Ph.D. student working with Dr. J. Anderson will focus on the seed weight QTL *QTkw.mna-2A* residing on the short arm of chromosome 2A [34]. This QTL significantly increased thousand kernel weight (h^2 = 0.90) in two of the three environments tested by an average of 0.90 g (LOD 4.1, R^2 8.5%) and also explained 7.9% of the variation in kernel diameter. This RIL population (139 lines) was generated from the cross of breeding lines MN98550-5 and MN99394-1 developed with Wheat CAP support [35]. RILs heterogeneous for the target region flanked by *Xgwm339 – barc311* (11.9 cM) have been identified. Parents of this population have been genotyped with the Illumina 90K platform. SNP polymorphisms identified near this QTL region will be valuable to characterize recombination events between flanking markers in the HIF.

MT: The MSU PhD student will be jointly advised by Drs. Talbert and Budak to verify, clone and deploy a QTL for productive tillers on chromosome 6B linked to SSR markers *gwm58* and *gwm88* and SNP markers *BS00076397_51* and *bobwhite_c42198_254*. This QTL, designated *QTn.mst-6B*, was identified in three RIL populations, with the cultivar Reeder contributing the

favorable allele. This favorable allele is present in only 7 out of the 220 elite lines in the Spring Wheat AM panel. *QTn.mst-6B* explained ~10% of the variation in productive tiller number (LOD >5) and was associated with yield increases in the tested environments [33]. SSR markers *gwm58* and *gwm88* were used to select heterozygous F_5 individuals from the three populations. Seeds from these individuals were screened to identify homozygous for the two alleles. Field experiments using these plants confirmed that the favorable allele is associated with a high initial tiller formation in all environments. However, these early tillers matured into productive tillers only in high-moisture environments, contributing to environmental plasticity [34]. Heterozygous progeny of the HIFs will be used to generate $F_{6:2}$ populations for fine-mapping, identification of recombinants and cloning of the causal gene.

NC: The PhD student working with Paul Murphy's winter wheat breeding program will coordinate efforts to clone and deploy a QTL affecting thousand kernel weight located on chromosome arm 6AL. This QTL was identified in the Massey x MPV 57 F_5-derived RIL mapping population (LOD = 3.5, 16.5% of variation explained), with the favorable allele present in the Massey parent. Five F_5 lines from the original mapping population that are heterozygous in the QTL region were already identified. Homozygous NIL pairs derived from these RILs were planted in replicated rows in the field at Raleigh, NC during fall 2015 to confirm QTL effects. The lines are also being grown to generate a large population for high density mapping. A QTL for kernel weight in a similar position was identified in a TCAP panel consisting of 250 elite spring wheat lines by researchers at WSU.

ND: The ND PhD student will work with J. Faris and S. Baenziger (UNL) on the cloning of a QTL for grain yield on chromosome 3A, derived from the HRWW variety Wichita (WI). This QTL has been characterized and validated in multiple experiments [35, 36] using recombinant inbred chromosome lines (RICLs) derived from crossing the winter wheat variety Cheyenne (CNN) with a CNN(WI 3A) chromosome substitution line. The QTL has been narrowed to a 12 cM region and WI alleles at this QTL lead to an average increase of 66 kg ha^{-1}. Because HIFs do not exist for a RICL population, and these lines are already isogenic for all chromosomes except 3A, we will cross a RICL harboring WI alleles at the QTL and CNN alleles for the remainder of the chromosome to CNN to develop a segregating F_2 population for high-resolution mapping. The PhD student will be based in the laboratory of J. Faris (ND), but will work with S. Baenziger's breeding program to phenotype critical lines in Nebraska where the QTL is known to be expressed. The student will also work to deploy the QTL into ND varieties.

NY: The PhD student at Cornell will target a seed length and weight QTL on chromosome 5BS in the 41-43 cM interval in the SynOpDH population. This QTL was identified in a 150 winter wheat association mapping population, the original ITMI population and the reconstructed SynOpDH population (LOD 6.1, R^2=14%). The putative orthologous QTL for seed weight, length and wide was identified in the colinear interval of chromosome arm 5AS same interval (seed weight LOD = 5.3, R^2=9%). Flanking markers were mapped in the 1,400 SynOpRIL population and heterozygous RILs were identified in 12 lines for the 5BS QTL and in 33 lines for the 5AS QTL. These heterozygous lines will be used to generate >1,000 $F_{5:2}$ HIFs for each region. NY and MN programs will collaborate on the cloning of the 5BS QTL, each using their very different populations. NY will work separately on the 5AS orthologous QTL.

OK: The PhD student working with Dr. L. Yan at OSU will focus on *QYld.osu-1BS*, a major QTL for grain yield that was discovered in a population of 282 doubled haploid (DH) lines generated from 'Duster' and 'Billlings', two winter wheat cultivars released in recent years. This population has been tested in the field at the Stillwater Agronomy Research Station in 2014 and

2015 and genotyped using 2,358 GBS markers. *QYld.osu-1BS* explained 23% and 24% of the total phenotypic variation in grain yield in 2014 and 2015, respectively. The Duster *QYld.osu-1BS* allele increased yield by 16% in 2014 and 23% in 2015, relative to the Billings allele. *QYld.osu-1BS* is located in the distal region of the short arm of chromosome 1B and does not come from the 1RS translocation (flanked by genes TPT and Di19). The OK PhD will characterize the *QYld.osu-1BS*, identify the gene underlying the QTL in different isogenic backgrounds and heterozygous plants, and deploy this gene into novel winter wheat varieties. NILs and HIFs for fine mapping are being generated using Buster siblings and DH lines to backcross with different lines with various genetic backgrounds.

SD: The SDSU PhD student working with S. Sehgal's winter wheat breeding program will coordinate efforts to fine map, clone and deploy a grain yield QTL from *Aegilops tauschii* transferred to hexaploid wheat (KS05HW14). This QTL is located on chromosome arm 5DL at position 161 cM on the *Ae. tauschii* genetic map, and 147-149 Mb on CS pseudomolecule. This QTL was identified in the Nested Association Mapping (DNAM) population (n=1,200). *Ae. tauschii* accession TA1662 contributes the beneficial allele. This QTL explains 4% of the variation in grain yield (P= 5.8E^{-04}, heritability 0.67) and was identified in three environments. This population is being grown in nine locations. We have identified 12 $BC_2F_{4:5}$ lines from the DNAM population that are still heterozygous for the QTL region and will be grown to generate >1,000 $F_{5:2}$ HIFs for high-density mapping and cloning of the causal gene.

TX: The TX PhD student will work on grain yield QTL *Qgy.tamu.6AL*. A set of 217 $F_{6:7}$ to $F_{6:8}$ RILs derived from the cross of CO 960293-2/TAM 111 were tested in eight environments across Texas, Kansas, Colorado, and Idaho. TAM 111 is currently the most widely planted HRWW cultivar in the US and its derivatives have been used in many wheat breeding programs. The 6AL QTL, associated with SNP markers *IWA2192* (191.3 Mb) and *IWA38* (192.3 Mb) explains 5.4% of the variation in grain yield and has an additive effect of 58 kg /ha (-logP = 4.8). Heterozygous plants in the region will be identified in F_5 and HIFs will be developed.

WA: The WSU PhD student will be co-advised by Drs. M. Pumphrey and A. Carter. The project will validate and clone a QTL associated with kernels per spike, spike length, and kernel weight located on chromosome arm 4AL between SNP markers *IWB21713* and *IWB28330.* This QTL was identified in a population of 180 RILs from the cross 'Kelse' x 'Scarlet' evaluated in four environments. This QTL explains 13% and 26 % of the variation (R^2) in kernels per spike, for early and late planting, respectively; and 12% for kernel weight. This QTL region was also associated with differences in flowering senescence, peduncle length and chlorophyll content. F_4 seeds are currently being screened using SNP markers spanning the QTL region to identify heterozygous families. HIFs will be used to generate $F_{4:2}$ for the first phase of the high-density map and more advanced $F_{5:2}$ for the final map and cloning.

Potential pitfalls: If some of the HIF populations fail to show a clear discrimination between alleles, we have already identified alternative QTL targets, which are segregating in several of the selected populations. To minimize the probability of epistatic interactions masking the effect of the target QTL we will develop multiple independent HIFs families for each QTL.

***D.1.3. Deployment of favorable alleles in CIMMYT germplasm*:** From the second year of the project HIF lines homozygous for alternative yield QTL alleles will be tested in Obregon to evaluate the value of these alleles for the CIMMYT program. Homozygous BC_3F_2 NILs differing for the presence and absence of the target QTL (or linked marker) will be bulked and their phenotype evaluated. These lines will then be incorporated into CIMMYT yield trails through

the IWYP hub in Obregon. Activities will be coordinated by co-PI M. Reynolds in collaboration with CIMMYT wheat breeders R. Singh and K. Ammar. QTL identified in winter wheat or photoperiod sensitive wheats will be transferred to two spring advanced lines from CIMMYT though four backcrosses before testing. A haplotype analysis of the SNPs in the QTL region will be used to select the most appropriate CIMMYT line with a contrasting haplotype. If possible we will prioritize the hexaploid varieties Reedling and Kingbird that are already being used in other IWYP projects. This will facilitate future studies of epistatic interactions among different genes affecting grain yield. During the fourth year of the project the PhD students will visit the Obregon Station in CIMMYT to see their trials and attend a short training course.

D.2. Methods for Objective 2

To prioritize candidate genes identified in the targeted regions for validation, we will integrate information from model species, wheat gene expression databases, and whole-genome imputation methods based on genotypic and phenotypic data generated in the TCAP project.

Information from other species: The function of the genes in the target regions will be annotated using gene ontologies and published information from other species. Transcription factors or genes associated with grain or spike development will be prioritized.

Expression profiles: Wheat Exp [30] and expVIP [31] will be used to characterize the expression profiles of individual wheat homoeologs from different target genes. These tools provide access to published RNA-Seq studies including different tissues at different developmental stages [37] and detailed analyses of grain development [38]. RNA-Seq studies of early stages of spike and grain will be available before the start of this project through the collaboration of two coPIs (Pearce and Dubcovsky) in a separate IWYP funded project (PD C. Uauy, UK). Genes differentially regulated during grain or spike development will be prioritized. Expression of these genes will be compared between the NIL for the target QTL.

Variation in regulatory regions and regions with high sensitivity to nuclease: In this project we will generate a catalog of SNP variation in the regulatory regions of the wheat genome and will perform differential nuclease sensitivity profiling of wheat chromatin (see D.4). These data will be integrated with gene networks and functional annotation of genes in the targeted QTL regions to prioritize candidates for functional validation. Tools for visualization and genomic data integration will be developed in the T3 database.

Natural variation: Captures from exome and regulatory regions will generate an extensive catalogue of polymorphisms between parental lines in each mapping population. We will determine the predicted effect of polymorphisms in the coding regions using SIFT [39] and tools developed by T3 (section D.4.2). We will prioritize polymorphisms that generate premature stop codons, modify splicing sites, or generate amino acid changes in conserved protein positions. For the regulatory regions we will prioritize polymorphisms in known regulatory motifs, regions identified by nuclease sensitivity profiling, and conserved regions identified by phylogenetic footprinting [40].

We will also take advantage of other biparental populations and AM panels segregating for the same QTL. Natural variation among these lines will be used to identify historic recombination events to dissect the target regions. The TCAP-NAM populations have already proven useful for this purpose. The T3 database team will develop tools to facilitate mining the available phenotypic and genotypic data, and to integrate these data with genomics information (D.4.2).

TILLING mutants: An available database including >10 million EMS mutations in tetraploid and hexaploid wheat (see preliminary results A.4.3) will be used to identify mutations in each homoeolog. For validation purposes, it is easier to use the tetraploid mutant population, because a null-mutant can be generated by a simple cross between homoeologs. Mutants will be backcrossed twice to the non-mutagenized Kronos to reduce mutation load. The hexaploid TILLING population will be used if appropriate mutations are not identified in tetraploid wheat or when loss-of-function mutations generate the desired allele (e.g. mutation of the negative growth regulator *TaGW2* that increases wheat grain size [3]).

Genome editing and complementation: When mutations are not found or when closely linked paralogs need to be mutagenized, we will use CRISPR-Cas9 genome editing to generate knock-out mutants of the target genes. CRISPR-Cas9 technology in wheat is available in KSU, OSU, and UCD. This is a good option to generate mutations in multiple homoeologs. The UCD transformation facility has developed a protocol to transform the tetraploid Kronos line used to generate the TILLING population (efficiency >5%). This can be used to validate candidate genes by complementation of the respective Kronos TILLING mutant. The KSU and UCD transformation facilities will provide individual positional cloning programs support for transformation of hexaploid and tetraploid wheat, respectively. We have budgeted centralized support for four independent targets (~5 independent events per target gene). Dr. Akhunov is presenting an independent grant to optimize CRISPR-Cas9 for wheat. If funded, this project will be coordinated with the TCAP grant to avoid any duplication (see collaboration letter).

D.3. Methods for Objective 3

In collaboration with the local wheat breeder, each PhD student will select at least one QTL to deploy into his/her home breeding program. Using backcrossing or forward breeding strategies the selected allele will be introgressed into commercial varieties or advanced breeding lines. The student can select the QTL from his/her cloning project or from previously published studies [41-51]. The PhD student can also select a mutant with known positive effects on wheat yield components. For example, the mutant for *TaGw2* (wheat ortholog of rice GW2) is a well-studied example of a loss-of function mutation in a negative regulator of growth, which results in significant increases in grain size in both tetraploid and hexaploid wheat [3]. This gene is being targeted by several breeding programs. Alternatively, students can pursue the rapid pyramiding of multiple QTL for yield components using genomic selection approaches. Multiples cycles of crossing and selection will be used to combine multiple QTL and to analyze combined effects in field experiments. Progress in spring and winter programs will proceed at different speeds.

The PhD students will work in close collaboration with the wheat breeders in their programs to select recurrent parents and strategies for forward breeding approaches. The students will collaborate with the regional genotyping labs in the development of high-throughput markers and in the screens of the large populations required to implement forward breeding MAS strategies.

PhD students will also collaborate with CIMMYT to test their isogenic lines in Obregon. Each student is expected to visit CIMMYT at least once during the project. PhD students working with spring wheat lines (photoperiod insensitive) will be able to test their isogenic lines directly at Obregon. Students working on winter wheat will transfer their QTL to one spring advanced line from CIMMYT-ESWYT panel through backcrossing (see section D.1.3).

A SNP-based haplotype analysis will be used to select CIMMYT recurrent parents with haplotypes different from the ones found in the QTL donor. This selection will be based on the genotyping of 96 lines, including 15 parental lines from this project, 32 lines used to develop the

NAM populations in TCAP, and 46 lines from the ESWYT panel. Genotyping will be done using five to ten KASP assays for each targeted QTL region. These KASP assays will be developed based on the SNP identified by capture in coding and regulatory regions.

D.4. Methods for Objective 4

D.4.1. New genomics tools: The previous exome capture platforms cover the protein coding regions but not the adjacent regulatory sequences [52]. In this project we will characterize sequence variation in the regulatory regions using a dedicated capture platform. We will also characterize chromatin structure in these regulatory regions using a nuclease sensitivity assay. These results will be integrated with RNA-Seq data [30, 38] and with candidate genes identified in this study to characterize gene regulatory networks controlling yield-related traits.

D.4.1.1. Capture of regulatory regions: Using the last version of the wheat genome annotation the KSU group will develop a capture assay targeting gene regulatory regions (henceforth, RegCap assay). Up to 1.5 kb of genomic regions upstream and 0.5 kb downstream from wheat gene models, introns not included in the previous exon capture and regions including miRNA precursors will be selected to design a 100 Mb Nimblegen SeqCap EZ capture assay. A panel of 240 genetically diverse wheat lines, including parental lines of the different HIF populations used in the positional cloning projects and lines previously genotyped using the 107 Mb exome capture assay [12], will be re-sequenced using the RegCap assay. Genomic libraries for RegCap will be prepared at the KSU genomics facility following the Biomek FXP Laboratory Automation Workstation protocol. Equimolar quantities of 12 barcoded genomic library (assessed by qPCR), will be pooled before hybridization. Sequence data will be generated for each pool to obtain ~30x coverage per diploid genome. A published bioinformatics pipeline [12] will be applied for variant calling. Regulatory SNP variation will be mapped to the latest version of wheat genome assembly and will be made available through the T3 and Ensembl Plants databases. Captures will be performed using this platform at KSU for parental lines of the 15 HIF populations. The combination of RegCap with the previous exon capture will provide wheat researchers a comprehensive list of polymorphisms in coding and regulatory regions.

D.4.1.2. Nuclease Sensitivity Analyses of Wheat Chromatin Structure: To delimit critical sequences within these wheat regulatory regions, we will develop and perform Micrococcal Nuclease (MNase) sensitivity assays [53]. This assay provides information about occupancy of *cis*-regulatory elements by *trans*-acting factors and of nucleosome positioning near these regulatory regions [54]. By mapping a wide range of fragment sizes, the genomic distributions of both nucleosomes and numerous non-histone proteins can be assessed using a single sequenced sample, making the MNase method especially cost-effective [54]. Studies in maize have shown that genome regions hyper-sensitive to MNase digestion are enriched in 5’ and 3’ ends of genes, are associated with gene expression levels, and overlap with conserved noncoding sequences [53]. SNPs located in the MNase hyper-sensitive regions explain a significant fraction of the phenotypic variance of multiple maize traits (E. Buckler, PAG 2016).

Samples will be prepared at UC Davis from Kronos and Cadenza at four early stages of spike and carpel development using available protocols. Tissues and time points were selected based on their relevance to the QTL selected in this project, and on the availability of RNA-Seq data from an independent IWYP project directed by C. Uauy (John Innes, UK). The protocol for nuclei extractions and MNase sensitivity profiling has been already tested at KSU using two levels of MNase. MNase digested DNA will be used to prepare Illumina genomic libraries [53, 55]. As a control, we will include naked DNA samples of Kronos and Chinese Spring. Sequence

reads will be mapped to the wheat reference genome and nucleosome occupancy peaks will be identified using the F-seq program [56]. To gain insights into dynamics around regions occupied by transcription factors and nucleosomes, V-plots will be generated as described before [57].

MNase hyper-sensitive and hyper-resistant regions will be added to the annotation of the wheat genome and made available through the T3 and Ensembl Plants databases. This information will be integrated with RNA-Seq data, and with the polymorphism identified in the regulatory-capture to help the positional cloning projects identify candidate SNPs, and to gain a better understanding of the gene networks regulating spike and grain development.

D.4.1.3. CRISPR-Cas9 transformation technologies to support candidate gene validation: Constructs with a functional wheat-codon optimized CAS9 under the control of maize ubiquitin promoter (Ubi), and guiding RNA under the control of wheat U6 promoter (see Prelim. Results) will be used to introduce knock-out mutations in the coding of candidate genes using the described procedure [58]. Constructs will contain the *bar* gene for transgenic plant selection. Cultivars Fielder and Bobwhite are routinely transformed at the KSU facility using a biolistics approach. KSU is in process of acquiring the license to perform transformation using an *Agrobacterium* protocol developed by Japan Tobacco International. This license is available at the UC Davis transformation facility and is routinely used to generate tetraploid and hexaploid transgenic wheat plants (efficiency >5%). The UCD facility will be used until the KSU *Agrobacterium* protocol is established. For each candidate gene, constructs with 2 to 3 guiding RNAs will be created. Each tiller of a transgenic plant will be tested for the presence of CAS9 by PCR; CAS9-positive tillers will be tested for the expression of CAS9 and gRNA by RT-PCR.

D.4.2. New informatics tools: To manage the large amount of genotypic and phenotypic information generated in this project, we will take advantage of "The Triticeae Toolbox" (T3) database developed as part of the TCAP. T3 will be improved to facilitate the identification and validation of candidate genes and to link current phenotypic, genotypic and genomics data.

QTL candidate regions often span many genes and polymorphisms. Final identification of the correct gene can be accelerated through candidate prioritization based on multiple sources of genomic information. SNP effect predictions are available for wheat based on coding sequence perturbation [59] and Genomic Evolutionary Rate Profiling based on comparative sequence alignment and conservation [60]. For regulatory regions, MNase (or DNase I) hyper-sensitive sites have been shown to be strongly enriched for QTL signals [61]. MNase sensitivity assays are proposed as part of this project.

The previous TCAP has generated enormous amount of data to identify genomic regions enriched for QTL signal: T3 now stores 334 phenotyping trials comprising some 440,000 measurements on over 9,400 lines, a trove of data unrivaled in the public sector. T3 will provide the tools to take advantage of these data sets by *i*) imputing genotyping datasets up to full set of exome polymorphisms [12]; *ii*) implementing GWAS on each phenotyping trial and trait while accounting for population structure using fixed principal component and random genomic relationship matrix factors; and *iii*) performing meta-analyses to identify genomic regions enriched and depleted in QTL signal [62]. These analyses would inform a tool in which the user would input the QTL support interval and T3 would then produce a putative set of causal SNPs [63] based on this data and on prior trait associations of the SNPs over the database as a whole. This information should stimulate hypothesis development in the pursuit of yield QTL. Furthermore, ongoing identification and cloning of QTLs in this proposal will serve to validate (or reject) whole-database analyses as proposed here.

This information will be linked to a JBrowse track already available in the TILLING mutant database, which displays 10 million EMS mutations in wheat coding regions and predicts the effect of the induced mutations. This database includes truncations or amino acid changes in ~98% of the genes. We will also develop a JBrowse track showing local recombination rate based on recombination events tallied in wheat NAM populations developed during TCAP. Recombination provides a known mechanism to uncouple deleterious from favorable variants thereby increasing the effectiveness of selection at removing deleterious variants [64]. T3 will enable researchers to easily identify and leverage recombination hot- and cold- spots within the QTL support intervals. We will also link genes identified in the candidate gene region with new tools that discriminate expression among wheat homoeologs [30, 31]. Gene expression is a powerful tool for candidate prioritization. We will develop an interface page that will allow a user to go from a marker or genomic position of interest, identify neighboring genes in the reference sequence, and display their relative expression across user-selected tissues.

D.5. Methods for Objective 5

Modern plant breeders need skills in field experimental design, marker-assisted selection and genomics/bioinformatics [21, 24, 65]. This project will provide support and integrated training in these areas to 15 PhD students. PhD students will have a field-based component and will be trained in experimental design and bioinformatics through online conferences and courses, face-to-face workshops, student seminars, and student discussion workshops. Each student will lead a QTL dissection project, which will provide training in genetic studies, marker development and integration of genomic resources to gene identification and validation. Each student will also have a QTL deployment project to give them the opportunity to work in close collaboration with breeders and genotyping labs. Students will see firsthand the challenges and rewards of transferring value from research to commercial varieties. Finally, the interaction with CIMMYT will provide students a global vision of plant breeding. The simultaneous training of a cohort of 15 PhD students will enhance the opportunities for collaboration and teach students the value of team work to solve complex problems. As in the previous TCAP, graduate students not specifically funded by the project will also be invited to participate in the educational activities.

***D.5.1. Face-to-face workshops*:** Students funded by the project will have the obligation to attend one workshop organized by the project each year. Courses during the first years are focused on the skills and knowledge required for their research project whereas courses during the last years are focused on expanding the scope of their training and generating networking opportunities required to enhance their chances to find jobs in plant breeding. Planned courses include:

Year 1. UC-Davis (responsible J. Dubcovsky): Positional cloning basic strategies and tools.

Year 2. KSU (responsible E. Akhunov): Candidate gene validation and bioinformatics tools.

Year 3: Cornell (responsible J.L. Jannink). Utilization of database resources.

Year 4: CIMMYT (responsible M. Reynolds). International centers and IWYP program.

Year 5: Private/Public: Syngenta, LimaGrain, opportunity for networking and finding jobs.

These courses will include lectures, workshops, bioinformatics labs and discussions of student projects. To broaden the impact of these face-to-face workshops, materials will be later reorganized by the education team into self-paced online courses. These PBTN self-paced courses will greatly increase the educational impact of the proposed face-to-face workshops and will benefit students and breeding professionals around the world (as in previous TCAP).

***D.5.2. On-line training*:** The online environments created during TCAP will be maintained and administered by L. Sandall at UNL. Online content will be made available through PBTN and eXtension. UNL is a premium eXtension member and has free access to the eXtension system. UNL is currently delivering several courses through eXtension, demonstrating the experience required for distributing online course materials. PhD students funded by this project will have the obligation to attend and complete one on-line course per year. The planned courses include:

Year 1: T3 resources and tools (late in year 1).

Year 2: Bioinformatics tools.

Year 3: Mutagenesis and CRISPR-Cas9 complemented by an advanced mentorship course.

Year 4: Data analysis and scientific writing. This course will be organized to help students publishing their results. We will form small writing groups to improve written communication.

Year 5: Genetic data analysis for breeding using ASReml: Linear mixed models, variance modeling, breeding values, genetic values, multivariate models, etc. (Based on a class and book put together by Fikret Isik and others at NCSU, https://faculty.cnr.ncsu.edu/fikretisik/breedingbook/

The education coordinator will have on-line meetings with groups of students working in the same region or in similar topics. The educational coordinator will monitor progress and collect information about the educational needs of the students. An important objective for creating this online environment is to decrease student isolation, improve collaboration, and provide access to technical expertise in the multiple subjects required by modern plant breeders.

***D.5.3.* Educational travel:** The project will assign each PhD student travel support to attend workshops and PAG meetings. Additional funding will be available through the Education coordinator to supplement travel needs. PhD students will have the obligation to attend PAG project meetings and present posters and oral presentations when requested. PhD students will also be encouraged to attend ASA-CSSA meetings and present their results to increase their networking opportunities. Finally, all students will travel to CIMMYT for a short course in the fourth year of the project. This will teach students the international dimensions of plant breeding and provide opportunities to build international collaborations. CIMMYT will provide hands on phenotyping experience and an introduction to international plant breeding.

***D.5.4.* Participation in graduate group meetings:** Each year at PAG a one-day meeting will be dedicated to presentations by the graduate students and group discussion of the different projects.

***D.5.5.* Presentations in National Scientific meetings:** Students will be encouraged to present posters in PAG and CSSA meetings and to participate in other forums related to Plant Breeding such as the National Association of Plant Breeders.

***D.5.6.* Industry internships:** To enhance students' understanding of the plant breeding industry, we will provide them opportunities to conduct short internships in industry. The educational coordinator will work with industry representatives and graduate students to coordinate internships during the fourth and fifth years of the project (see attached letters).

E. OVERSIGHT AND EVALUATION

E.1. Oversight research activities. Each participant will be required to submit a yearly progress report and work plan that will be reviewed by the project directors and the executive committee. The Scientific Advisory Board and the Industry Liaison Committee (see Management plan) will receive the reports detailing progress for each milestone. The PD will provide USDA a written

report including board recommendations and actions taken by the project. Specific milestones of the positional cloning project include:

1.-Mendelization of the QTL by using isogenic materials and highly replicated field trials (Y1).
2.-Development of a sub-cM map based on >1,000 HIFs (Y2).
3.- Delimitation of the candidate gene region and identification of potential candidates (Y3).
4.- Validation of the candidate genes using mutant and transgenic approaches (Y4).
5.- Publication of the QTL dissection results (Y5).

For the QTL deployment sub-project the milestones will include:

1.- Selection of recurrent parents and development of crossing blocks for forward breeding (Y1).
2.- Delineating detailed plan for QTL deployment (Y2).
3.- Development of BC_4F_2 lines for preliminary yield testing.
4.- High-throughput genotyping of advanced breeding lines (Y3).
5.- Development of BC_5F_2 and BC_4F_3 lines for yield testing (Y4).
6.- Germplasm releases of isogenic lines (Y5).

E.2. Oversight educational activities: Oversight of the student's progress will be provided by the educational coordinator in collaboration with the Project Director. Oversight for the educational activities of the project will be complemented by an advisory committee composed of nationally recognized education and plant breeding experts (see 'Management Plan' and 'Advisory Appendix'). The previous Educational Coordinator of TCAP has agreed to participate in an oversight capacity. The advisory committee will receive annual reports from the education coordinator and provide recommendation to the education coordinators and coPDs.

The education coordinator and education assistant (DeAnna Crow who was in charge of the same activities in the previous TCAP) will conduct surveys and interviews to evaluate the baseline education status of incoming PhD students, measure their progress, and obtain feedback to assess the effectiveness of the education program. We will use the same evaluation tools created by the outside evaluators of the TCAP. This will allow us to obtain consistent long-term data about collaborative training of plant breeders. Surveys and focus sessions will also be used to assess student interest and morale. PhD students supported by this project will have to sign a document describing their responsibilities (e.g. active participation in online and face-to-face activities) that will clearly indicate that continued support will be contingent upon satisfactory evaluation by the education coordinator of their level of participation in the project's educational activities.

F. EXPECTED OUTCOMES

The expected research outcomes of this project include: *i*) well characterized QTL for grain yield components in wheat, *ii*) identification of genes that control wheat yield, *iii*) deployment of recessive alleles previously masked by gene redundancy in polyploid wheat, *iv*) new technologies and strategies to identify and deploy genes in wheat, *v*) genotypic and phenotypic information for a large number of breeding lines organized in a database that serves wheat breeders, *vi*) perfect markers for genes regulating grain yield components, *vii*) a database integrating germplasm, genotypes, phenotypes and genomics information, *viii*) a new capture platform of wheat regulatory regions, *ix*) information about critical regions in the promoters of a large number of wheat genes, *x*) wheat varieties with improved yield potential.

The education outcomes of this project include *i*) fifteen new plant breeders trained in plant breeding and molecular genetics, *ii*) an improved plant breeding community of learning (PBTN). *iii*) broader and collaborative educational experiences for plant breeding PhD students.

G. DISSEMINATION PLAN

Results from this project will be disseminated in two ways. The first one is by presenting germplasm and varieties with alleles for increased yield. Growers will be able to see in their fields the benefits of this research. Isogenic lines carrying the selected alleles will be also showcased in field days and demonstration trials to show growers and industry concrete examples of the value generated by this research project. Results from comparative yield trials will be presented in local growers meetings, wheat commissions, and local agronomic journals to directly reach the growers. In addition, results will be presented in publications in peer reviewed scientific journals and in national and international scientific conferences. Students will be encouraged to disseminate their work in posters, conferences and field day presentations.

H. TIMELINE

Obj.	Tasks and milestones*	Y1	Y2	Y3	Y4	Y5
	Research					
1	Validation of QTL in isogenic backgrounds	x	x			
1	Development of high density mapping populations		x	x		
1	Identification of candidate region and candidate genes			x	x	
2	Validation by expression and natural variation		x	x		
2	Validation by mutants and CRISPR-Cas9			x	x	x
2	Validation by transgenic complementation		x	x	x	x
3	Selection of backcrossing parents and BC_1	x				
3	Initiation of crossing block for forward breeding and F_2	x				
3	Development of $BC_{2\text{-}3}$ and F_3		x			
3	Development of BC_4, BC_4F_2 and F_4			x		
3	Development of $BC_4F_{3\text{-}4}$ and testing			x	x	x
3	Development of F_5 and evaluation of headrows				x	x
4.1	Development of capture regulatory regions	x	x			
4.1	Captures of parental HIFs & genotyping by sequencing		x	x		
4.1	Nuclease studies of ocupancy regulatory regions				x	x
4.2	T3 tools for candidate gene identification development	x	x	x		
4.2	T3 tools for genomics and validation			x	x	x
	Education					
5	Participation in scientific meetings	x	x	x	x	x
5	Workshops on positional cloning	x	x			
5	Workshops on bioinformatics			x		
5	Trip to CIMMYT				x	
5	Internships in industry			x	x	x

* Crosses in winter programs will take an additional month to satisfy vernalization requirements.

Personnel responsible for the different tasks are listed in the attached **'Key Personnel'**. The role of each participant and of the advisory boards and industry liaison committees are listed in the requested '**Management Plan'**, '**Advisory board Appendix**' and '**Budget Management Plan and Timeline Appendix**'. A separate '**Data Management Plan**' describes plans to manage and disseminate data generated by the project.