PROJECT SUMMARY

Instructions:
The summary is limited to 250 words. The names and affiliated organizations of all Project Directors/Principal Investigators (PD/PI) should be listed in additions to the title of the project. The summary should be a self-contained, specific description of the activity to be undertaken and should focus on: overall project goals(s) and supporting objectives; plans to accomplish project goal(s); and relevance of the project to the goals of the program. The importance of a concise, informative Project summary cannot be overemphasized.

Title: Common Bean Coordinated Agricultural Project

PD: McClean, Phillip E.	**Institution**: North Dakota State University
CO-PD: Garden-Robinson, Julie	**Institution**: North Dakota State University
CO-PD: Johnson, Christina	**Institution**: North Dakota State University
CO-PD: Osorno, Juan	**Institution**: North Dakota State University
CO-PD: Slator, Brian	**Institution**: North Dakota State University
CO-PD: Kelly, Jim	**Institution**: Michigan State University
CO-PD: Brick, Mark	**Institution**: Colorado State University
CO-PD: Ryan, Elizabeth	**Institution**: Colorado State University
CO-PD: Thompson, Henry	**Institution**: Colorado State University
CO-PD: Myers, Jim	**Institution**: Oregon State University
CO-PD: Gepts, Paul	**Institution**: University of California, Davis
CO-PD: Urrea, Carlos	**Institution**: University of Nebraska, Lincoln
CO-PD: Grusak, Michael A.	**Institution**: USDA Children's Nutrition Research Center, Houston, TX
CO-PD: Cregan, Perry	**Institution**: USDA Soybean Genomics and Improvement Lab, Beltsville, MD
CO-PD: Hyten, David	**Institution**: USDA Soybean Genomics and Improvement Lab, Beltsville, MD
CO-PD: Cichy, Karen	**Institution**: USDA Sugarbeet and Bean Research Unit, East Lansing, MI.
CO-PD: Porch, Tim	**Institution**: USDA Tropical Agriculture Research Station, Mayaguez, PR
CO-PD: Miklas, Phillip	**Institution**: USDA Vegetable and Forage Crops Research Lab, Prosser, WA

The Common Bean Coordinated Agricultural Project (BeanCAP) will significantly impact the future direction of research by providing new tools and research directions for this important nutritional and commodity crop. The first market-class-specific markers, whose value will extend well beyond the project duration, will be a major outcome affecting all bean research. When genotypic data, generated by using these markers, is coupled with nutritional profiling data, also generated by the project, species-wide and market-class-specific loci affecting the nutritional traits will be discovered. This will set the stage for significant common bean improvements for years to come. All public US bean breeding programs will also be supported by 1) a genotyping program that will aid the discovery of genetic factors controlling traits of local agronomic importance; and 2) conversion of high throughput markers into low cost makers for day-to-day use in breeding programs. The nutrition, genetic, and genomic scientists will coordinate the development of an eXtension Community of Practice that will utilize high-quality animations and other multimedia to educate the general public and educational communities about the biology of nutrition and how genetics/genomics technology assists with the improvement of nutritional traits. The BeanCAP will also initiate a modern plant breeding training program focusing on early career recruitment and practical breeding/genomics training that illustrates, as an example, how the integration of genomic and phenotypic data can be used to improve nutritional traits in plants. This will provide a stream of students interested in filling the plant breeding human resource pool.

PROJECT NARRATIVE

A. Introduction

The Common Bean Coordinated Agricultural Project (**BeanCAP**) will strengthen the bean research, education, and extension communities by focusing on the genetics and genomics aspects of nutrition in this important food crop.

The **research component** will translate current and newly developed genomic information into the first suite of market-class-specific markers with broad utility to improve ***any trait*** of interest. The markers will be used in the first ever nation-wide project to define loci controlling a large collection of nutritionally important traits at both the species and market-class-specific levels. Each breeding program can then design the best strategy for the nutritional improvement of the market classes they work with. A genotyping service will support all public US breeding programs improving traits of local importance. Low cost markers that leverage the information from the high throughput marker screening will be developed for the long-term improvement of common bean. The marker and phenotypic data generated by this project will be organized along with all available genetic map, marker, and loci data into a database that is fully operable with other legume databases.

The **extension component** will develop a new Community of Practice (COP), "**Nutritional Genetics and Genomics: Healthy Foods from the Field to the Table."** The COP will inform educators and the public about the biology, genetics, and genomics of nutrition. High quality, narrated animations will be the primary learning tool.

The **educational component** will design models for the recruitment of new students into the plant breeding discipline. Participants will be a trained in a program that introduces the basics of plant breeding, genetics, and genomics using nutritional genetics and genomics as the theme. The education and extension participants will collate all education materials developed by other plant genomics CAP projects into a newly designed learning environment that integrates those materials with those developed for this project. The specific project objectives are:

- **Objective 1**: Develop high throughput, market-class-specific markers for the predominant common bean market classes produced in the US, convert those markers into breeder-friendly markers, and genotype breeder-defined populations with these markers.
- **Objective 2**: Discover genetic loci associated with nutritional traits that define "healthy beans" by combining genotype and nutritional profile data of association mapping and bi-parental populations.
- **Objective 3**: Integrate common bean phenotypic, genotypic, and molecular marker data with other emerging legume genomic resources into breeder-friendly bioinformatic tools.
- **Objective 4**: Launch the "**Nutritional Genetics and Genomics: Healthy Foods from the Field to the Table"** eXtension Community of Practice (COP). The COP will utilize high-quality animations and other multimedia to highlight the biology and technology associated with the genomic-based improvement of nutritional traits.
- **Objective 5**. Initiate a modern plant breeding training program that focuses on early career recruitment and provides practical training that illustrates how the integration of genomic and phenotypic data can be used to improve nutritional traits in plants.

Genesis of the Project. The BeanCAP team is drawn from a well-organized community of individuals who exhibit a high degree of coordination and cooperation. Team members interact with local commodity groups and industry to seek input on their research direction. These meeting guide the W1150 experiment station project entitled "Exotic Germplasm Conversion and Breeding Common Bean (*Phaseolus vulgaris* L.) for Resistance to Abiotic and Biotic Stresses and to Enhance Nutritional Value". The project recently received the 2009 Excellence in Multistate Research Award by the Western Association of Agricultural Experiment Station directors. (The Statement of Issues and Justification for this project can be viewed at: http://nimss.umd.edu/homepages/home.cfm?trackID=7076). **That white paper provided the impetus for the project**.

BeanCAP participants coordinate with the private sector and the international research community during the biennial Bean Improvement Cooperative (BIC) meeting. BIC is an international organization that provides a focal point for worldwide interactions and is an important venue for private industry to also provide input into the public research effort. **The feedback that we have received from all of these groups informed our decision to focus on the nutritional status of common bean.**

Large, integrated projects such as the BeanCAP can make significant educational impacts beyond the research component. The general public wants to understand health and nutrition yet they need unbiased, peer-reviewed information that is presented in a manner that facilitates learning. Numerous research reports have demonstrated that animations are a superior learning tool (McClean et al. 2005) especially when integrated with other learning materials. **This research has informed our decision to use animations as the primary learning tool for our nutrition oriented COP.**

Multiple reports have described industry's need for plant breeding training. Yet the number of individuals obtaining degrees in plant breeding has been relatively stagnant in recent years (Guner and Wehner, 2003; Bliss, 2006), and current numbers of plant breeding graduates do not meet the number of open plant breeding positions. **This human resource need is the basis for our decision to focus our education efforts on developing models for filling the plant breeding training pipeline.**

B. Rationale and Significance

A significant aspect of the BeanCAP project is the full integration of the research, extension, and education components around the theme of nutrition, genetics, and genomics. This theme was chosen because common bean is a nutritionally important crop. By mobilizing all three components around a single theme, we will 1) **develop research tools** in order to better understand the genetic and genomic factors that define a nutritional crop; 2) **create effective delivery methods** to disseminate this knowledge in a way that assists the public as they make informed decisions regarding biology and health; and 3) **design and implement a training program** for young plant breeders that emphasizes how a crop can be improved nutritionally by using all of the research tools available.

From a **research perspective**, little is known about the genetic and genomic control of important nutritional traits in common bean or any crop. The modern genetic tools developed here will be most effective when we evaluate the expression of the nutritional traits in experiments that consider the full breadth of genotype (market classes) and environmental

(production regions; abiotic stresses) factors important in the US. ***The development of market class specific markers that fully supports all US research programs will be a significant outcome of the BeanCAP***.

For learners to comprehend nutrition and how genomic tools can effectively improve the nutrition of a plant, they must learn basic cellular, molecular and physiological processes. These three-dimensional processes are best learned by visualizations that contain a motion component that promotes a better understanding of interrelationships that involve space and time. The **extension component** of the project will develop a Community of Practice featuring a suite of high-quality animations of important biological processes that are integrated with other animations that teach the methods of modern crop improvement using nutritional crops as an organizing theme. These animations will also be coupled with other educational media designed to enhance the learning experience.

In the current job market, there is a strong demand for plant breeders, especially in the private sector. Yet the number of individuals obtaining degrees in plant breeding has been relatively stagnant in recent years (Frey, 1996; Guner and Wehner, 2003; Trexler et. al., 2005), and the current number of plant breeding graduates does not meet the number of open plant breeding positions. To address this growing need, the **educational activities** will provide early exposure of high school students to plant breeding as a profession, and recruit young undergraduates into a year-round internship program that involves students in the field and lab components of a plant breeding program. The theme of nutrition, genetics and genomics will be prominent in all high school and undergraduate recruitment and training materials.

C. Approach

1. BeanCAP Research Program: Nutritional Genetics and Genomics of Common Bean

a. Background. Common bean, an important US economic crop. Common bean (*Phaseolus vulgaris* L.; 2n = 2x = 22) is in important economic crop in the US where 1.7 million acres of dry and snap beans were planted in 2008. While production of dry bean is primarily located in North Dakota, Michigan, Nebraska, and Colorado (90% of production), the crop is also grown in 15 other states. US dry bean production is primarily restricted to five market classes - pinto, navy, black, red kidney, and great northern (in descending order of importance) - that account for 81% of the US production. Snap bean are grown for the fresh and processing markets on 300,000 acres, primarily in Oregon, Wisconsin, and New York. **Collectively, the value of common bean production at the farm gate in 2008 was $1.5 billion.** This exceeds the value of all other vegetable legumes combined (chickpea, lentil, pea, and peanut). The 16% increase in common bean value in 2008 was one of the largest increases among major crops.

Common bean genome, genetic organization, and diversity. The common bean genome is moderate in size (450-650 MBp/haploid; Bennett and Leitch. 2005). It appears to be a “true” diploid since nearly all loci are single copy (Vallejos et al. 1992; Freyre et al. 1998; McClean et al. 2002) while, the traditionally large families such as resistance gene analogs (Rivkin et al. 1999) and protein kinases (Vallad et al. 2001) are smaller than most crops.

Two major domestication events appear to have resulted in the Middle American and Andean gene pools (Kaplan and Lynch 1999). These are distinguished by biochemical (Gepts and Bliss

1986; McClean et al. 2004a, b) and morphological (Gepts and Debouck 1991) traits. Following domestication, gene pool divergence lead to the development of several races within each of the two major domesticated gene pools (Singh et al. 1991).

US breeders focus on five dry bean classes (pinto, navy, black, Great Northern, kidney) and snap beans for the fresh and processing markets. Each market class is defined by a specific seed size, color, and pattern, traits controlled by many genes (McClean 2002). Crosses are typically made between genotypes within a market class because it is difficult to recover a market class phenotype from inter-market-class crosses This has lead to limited sequence variability within each market class (McClean et al. 2004, 2007). The variability is lowest with the kidney and Andean-derived snap beans and is greater within the Middle American races Durango and Mesoamerican. By contrast, a large amount of variability is observed between genotypes from the two gene pools. This has been exploited extensively during the development of populations used to map important agronomic traits (Miklas et al. 2006).

Bean breeders: early adopters and innovators of molecular breeding. Marker-assisted selection (MAS) has become an essential tool in public bean breeding to introgress and pyramid resistance genes (Kelly et al. 2003; Miklas et al. 2006). Bean breeders pioneered the discovery and application of markers by using advanced generation segregating lines and cis- and/or trans-markers to effectively integrate desirable traits (reviewed in Kelly et al. 2003). Currently, 41 markers linked to qualitative and quantitative loci important for disease resistance and other production traits (http://www.css.msu.edu/bic/PDF/SCAR%20Markers%202007.pdf) are available. Since 2000, over 90% (17/19) of the disease-resistant germplasm released in the US were developed using molecular markers. Since this germplasm is being broadly used by all public breeding programs, the impact of molecular marker breeding on the bean industry will be long lasting. These tools are critical for private breeding companies, as they too become increasingly active users of these molecular markers. (See industry support letters.)

To repeat this success, bean breeders face a number of challenges to improve important agronomic traits. Because of the narrow genetic basis of beans in the US (McClean et al. 1993; Sonnante et al. 1994), bean breeders increasingly use exotic, unadapted germplasm as a source of useful traits. This raises significant breeding challenges because each market class has strict phenotypic characteristics that must be maintained. Molecular markers are needed not only to introgress gene(s) but also to facilitate the recovery of the adapted (domesticated *P. vulgaris*) genetic background. **Yet market-class specific markers are few in number and are needed to facilitate bean improvement.** Most markers were developed one at a time using the cumbersome process of identifying RAPD markers and converting these to SCAR markers (Kelly et al. 2003). **Therefore, a national effort to develop and apply high-throughput marker technology for bean improvement is necessary.**

Common bean, a major nutritional crop. Plants are a rich source of basic sugars, proteins, and lipids essential for our normal growth and development, and supply vitamins and minerals required for many of the basic cellular processes necessary for life. Recent research has focused on phytochemicals, plant-derived molecules that while not essential for human growth and development, positively affect human health. Legumes are considered to be some of the healthiest foods in our diet. Although some provide the full complement of 15 minerals required for human metabolism (Grusak 2002), levels of the minerals vary among crops within the family. **Among legumes, common bean are consumed by humans more than any other crop.**

Health benefits from eating beans are numerous and include reducing the blood cholesterol and sugar levels which in turn prevents or alleviates certain types of cancer, Type 2 diabetes, and cardiovascular diseases (Andersen et al. 1984; Tietyen-Clark, 1986). Diets rich in zinc and iron, two micronutrients abundant in bean, can delay the onset of AIDS (Savarino et al. 1999; Buys et al 2002), and as such, HIV positive patients are encouraged to include beans in their diets (SADH 2001; Jackson et al. 2007; Malete et al. 2007). Bean seeds also contain a protein that inhibits the HIV-1 reverse transcriptase (Wong et al. 2006). Recent research has shown that beans significantly reduce the onset of breast cancer (Thompson et al. 2009), colon cancer (Hangen and Bennink 2002; Feregrino-Perez et al. 2008), and biomarkers for heart disease risk (Winham et al. 2007).

From a crop improvement perspective, it is important to know the range of the natural variation of any nutritional factor. In common bean, colored beans have greater antioxidant activity than white beans (Madhujith et al. 2005), and that among the colored beans, the antioxidant activity of black beans was higher than for yellow-seeded beans (Rocha-Guzman et al. 2007). These experiments are typical not only for common beans but, all crops in general where only a few genotypes or market classes are profiled during an experiment. **Species-wide variation is simply not assessed.**

Obviously, it is foolish to claim any one plant species is a superfood that contains abundant amounts of all the basic nutrients of life and is a rich source of those health boosting phytochemicals. That is why a balanced diet is necessary. But that fact does not preclude the possibility of improving the health value of a crop plant like common bean by 1) **assessing its natural variation** for food nutrients, 2) **discovering the genetic factors** controlling the amount of the nutrients, and 3) **applying modern plant breeding practices** to increase the amount of these nutrients in the food. These three activities define the BeanCAP research component.

b. BeanCAP Research Products

The following research products will significantly improve common bean for many years into the future:

- A suite of market-class specific SNP markers for whole genome screening
- Breeder friendly CAP markers linked to nutritional traits and all known single gene and known QTL in common bean
- Discovery of species-wide and market class specific loci affecting expression of nutritional traits
- Breeder accessible database of phenotypic and genotypic data

c. Research Activities – Specific Experiments

i. SNP marker development **(Cregan, Hyten)**. A 1536 oligonucleotide pool array (OPA) set (Illumina Golden Gate Assay) is currently available. It was developed from ~7000 SNP discovered using a combination of genome resequencing of a large number of common bean loci utilizing soybean gene-specific primers, 454 sequencing of BAT96, and Solexa sequencing of Jalo EEP558 (Cregan et al. 2009). To uncover the level of diversity among and within the six major common bean market classes (pinto, black, navy, great northern, kidney, and snap), an equal distribution of genotypes from each of the market classes will be surveyed using this OPA.

Many previous studies have shown that while between gene pool sequence diversity is substantial, within gene pool, and more importantly, within market class variability is more limited. Therefore, we anticipate that while the initial screening with the 1536 SNPs will reveal polymorphisms among market classes, it will not meet our goal of developing a large collection of within market class markers.

To discover within market class SNPs, we will use the high-throughput Solexa "Paired-End Module" sequencing protocols to collect ~100 bp of sequence data per fragment. The source DNA will be a reduced-representation library developed by cutting genomic DNA with a group of restriction enzymes. Two diverse modern cultivars (based on the above diversity screening) from the pinto, navy, black, great northern, and kidney market classes will be surveyed. Since the snap bean market class is derived from both Middle American and Andean germplasm, a total of four cultivars, two each from the two gene pools, will be sequenced. The SNP data will then be compared with our reference sequence data to discover market-class-specific SNPs. The reference sequence will be constructed from 454 runs of the Andean landrace Jalo EEP558. The current collection of 454 data for this genotype is small. This project will augment that data set with ~1 million reads of about 400 bp in length.

Based on previous sequencing experiments (McClean et al. 2004a, 2004b, 2007), much of the market class diversity is also captured with the Durango (pinto and great northern market classes), Mesoamerican (navy and black market classes), and Nueva Granda (kidney and most snap market classes) races. If also observe stratified diversity based on race, then we will design a 768 OPA set for each race. Alternatively, if SNPs are confined to market class, then 384 OPA sets specific to each market class will be designed. We should mention that the larger OPA is preferred because we will be able to collect a greater number of data points for a smaller cost.

***ii. Discretionary SNP screening; a service for ALL public US bean breeding programs* (Cregan, Hyten, Gepts)**. The project will offer a **genotype screening service**. SNP genotypes will be collected for 8256 individual DNA samples in years 2-4. In general, two types of materials will be screened. First, all members of the association mapping populations used to discover genetic loci associated with nutritional traits will be screened. In addition, individual members of all bi-parental populations developed to confirm nutritional loci within the various market classes will also be genotyped. (Details of the experiments that will use these populations are described below.) Genotypes submitted by interested common bean breeder or geneticists will also be screened. The follow are examples of materials to be screened.

1. All members of the **common bean core collection** housed at the Pullman, WA site of the National Plant Germplasm System (NPGS). This will support future research by providing a genotypic basis for any phenotyping efforts that use the core collection to discover new useful traits of agronomic importance.
2. The **association mapping populations**, consisting of 400 dry bean and 200 snap bean lines, respectively, that will be used for the nutritional genomics experiments. All of these lines will be genotyped, and that data will be used for the statistical analysis to discover loci associated with the nutritional traits under study.
3. All **members of bi-parental populations** used by multiple geneticists to discover disease resistance loci (Miklas and Singh 2007). This will include any internationally developed population that has important utility to the US common bean breeding. This

effort, when coupled with phenotypic data, will enable the discovery of markers more tightly linked to major disease resistance loci.

4. **Breeder contributed genetic materials**. Plant breeders are either evaluating parents for hybrid breeding or screening populations for loci linked or associated with agronomic traits of interest. The genotyping of these populations will greatly increase the efficiency of the efforts to discover important traits. To participate in the program, breeders will agree to collect phenotypic data using local funds and to submit that phenotypic data to the bioinformatics group for inclusion into the program database.

The final decision regarding genotyping will be made by breeders that are members of the Steering Committee (**Kelly, Miklas, Myers, Osorno**). To ensure Golden Gate assays quality control, the **Gepts Lab** (UC, Davis) will isolate DNA from seed provided by breeders or geneticists. That DNA will then be sent to USDA/Beltsville where the Cregan lab will perform the SNP genotyping. The OPA set that will be used for genotyping will depend upon the market class or race of the materials that will be screened.

iii. Conversion of SNP loci into breeder-friendly CAP markers linked to all known agronomic genes in common bean **(McClean)**. Although the SNP technology provides a highly-efficient method of performing whole genome scans for important marker trait associations, it is cost prohibitive to use this technology to screen only a single or a few markers of interest within a breeding program. For that purpose, we will develop cleaved amplified polymorphic (CAP) markers for important regions and traits of interest. To effectively link the SNP discover and the down-stream breeder screening, we will, whenever possible, select SNPs for the OPA set that are part of an inexpensive restriction enzyme site (e.g. *Hin*dIII, *Eco*RI). Given the large number of SNPs that will be generated by the Solexa sequencing effort, we do not foresee any difficult with this type of SNP selection process.

We will first define the location of all known single genes and QTL associated with important agronomic traits using the database tools described below. SNPs that colocalize with these loci will be converted into CAP markers. The first year of the project will focus on disease resistance loci. As nutritional trait loci are discovered, the emphasis will switch to those loci. By the end of the project, our goal is to define at least 15 CAP loci for each common bean chromosome in a race (Durango, Mesoamerican, Nueva Granda) specific manner. This will enable breeders to select the best CAP marker for their particular application based on genetic location and genotyping cost. The **McClean Lab** (NDSU) will develop the CAP markers

iv. Nutritional genetics and genomics.

Year 1: Nutritional profiling the common bean germplasm. In year 1, the project will develop nutritional profiles of 600 genotypes representing all of the major market classes of common bean grown in the US. Because this same collection will also be grown under field conditions at multiple US locations in year 2, only genotypes adapted to US production conditions (maturity, photoperiod) will be considered. To study the maximum potential for each of the traits under consideration, plants will first be grown under controlled greenhouse conditions (Waters and Grusak 2008) at NDSU. Seed (for dry bean) and pods (for snap beans) will then be sent to the testing locations for nutritional profiling. The following set of nutritional traits will be evaluated by the investigators noted in parentheses. (**Note:** Because of cost, some

traits will only be evaluated with the greenhouse materials, whereas other nutritional traits will be profiled with cultivars used in all experiments. That information is listed for each trait.)

1. **Minerals** (Grusak; all experiments). Common bean is a rich source of minerals (Wang et al. 2003; Grusak 2002), and therefore it is important to measure the full range of variation. Samples will be delivered to **USDA Children's Nutrition Research Center** in Houston where data for 16 minerals (P, Ca, K, Mg, Cu, Fe, Zn, Mn, Co, Ni, B, Se, Mo, Na, As, and Cd) will be collected using ICP-MS (inductive coupling plasma mass spectrometry) technology.
2. **Nutritional availability and CaCo-2 cell tests** (Grusak; greenhouse experiment only). Although content of a specific mineral may be high, it is necessary to determine the nutritional availability of that mineral. CaCo-2 cells cultures are typically used for these analyses as a substitute for the much more expensive rat or human studies. These analyses correlate well with whole animal experiments (Fairweather-Tait et al. 2005) and have already been applied for studies in common bean (Hu et al. 2006; Laparra et al. 2008). The analyses will be performed at the **USDA Children's Nutrition Research Center** in Houston.
3. **Antioxidants, phenolics, and anthocyanins** (Brick/Thompson/Ryan; all experiments). The colored components of the various common bean market classes are products of various steps along the anthocyanin biosynthetic pathway and are extracted as phenolics. The molecules themselves have beneficial antioxidant activity (Madhujith et al. 2005; Rocha-Guzman et al. 2007), but variation both within and among various market classes is unknown. The **Cancer Prevention Laboratory at Colorado State University** will perform these analyses.
4. **Soluble/insoluble carbohydrates** (Brick/Thompson/Ryan; greenhouse experiment only). As beans pass through the gut, some carbohydrates are broken down and absorbed, while others are insoluble and pass into the colon. In the colon, they may be broken down into various short-chain fatty acids that decrease the onset of colon cancer (Hangen and Bennik 2002; Feregrino-Perez et al. 2008). Therefore, a crop wide survey at The **Cancer Prevention Laboratory at Colorado State University** will be performed.
5. **Protein, oil, fiber** (Tulmek ; all experiments). Protein, oil and fiber are three basic components that determine a crop's nutritional value. In particular, common bean is recognized as a major protein source, especially in developing countries. The **North Crop Institute at NDSU,** with a long history of seed diagnostics, will evaluate these traits.
6. **Phytate** (Cichy; all experiments). Phytate has been implicated as both a nutritional (Zhou and Erdman 1995; Thompson and Zhang 1991) and anti-nutritional (Oatway et al. 2001) molecule. This important trait will be analyzed by Karen Cichy, a new researcher at the **USDA Sugarbeet and Bean Research Unit** in East Lansing, MI.
7. **Carotenoid, vitamin C, fiber** (Myers). Snap beans have different nutritional features than dry bean. For this crop, the pod rather than the dry seed is the organ of importance. Carotenoid and vitamin C provide antioxidant protection, while fiber reduces risks of cancer, heart disease, diabetes, and obesity. The nutritional traits will be analyzed in common bean pods at **Oregon State University**. The fiber test is different than that performed at NDSU (see above).

Year 2: Association mapping experiments. Association mapping (AM), an alternative to linkage analysis, uses the natural sequence diversity within a species to define the various loci controlling a complex trait (Jorde 2000; Mackay 2001; Nordborg and Tavaré 2002). As an example, the McClean lab used this approach to discover loci associated with iron deficiency chlorosis in soybean (Wang et al. 2008). Other recent examples of association mapping in plants include fiber quality (Abdurakhmonov et al. 2008), plant height (Brown et al. 2008), and flowering time (Ducrocq et al. 2008; Stracke et al. 2009). Association mapping has also been used to discover genes associated with nutritional traits such as vitamin A content (Harjes et al. 2008), starch content and composition (Wilson et al. 2004), and carotenoid content (Palaisa et al. 2003).

An AM experiment consisting of the 400 dry bean genotypes described will be performed in replicated trials under normal production conditions in **North Dakota** (Osorno), **Michigan** (Kelly), **Nebraska** (Urrea), and **Colorado** (Brick). These sites were chosen because 90% of the US dry bean production comes from these states. For this experiment, standard production trait data (**days to flowering, days to maturity, plant height, lodging, growth habit, 100 seed weight, and yield**) will be collected to determine if any trait is correlated with any nutritional trait. A snap bean AM population, consisting of 200 lines, will be grown in **Oregon** (Myers), a primary snap bean production state in the US. Standard production data (**yield, pod shape, pod color, days to harvest, concentration of set, lodging, and architecture**) will be collected, and correlations with nutritional traits will be calculated. AM statistical analyses will also be performed to discover those loci associated with each nutritional trait. (See details below.)

Year 2: Effects on abiotic stress effects on nutritional traits. Some genetic loci exert their effect regardless of the environmental conditions; whereas others affect the trait only under a specific condition. From a genetic perspective, these can be considered **constitutive QTL** and **adaptive QTL** (Collins et al. 2008), respectively. Distinguishing between these two types of loci is important as plant breeders decide what loci to follow in a marker-assisted breeding program. We are sensitive to the fact that breeding for optimum performance under abiotic stress conditions is now an important consideration of all bean breeding programs. Therefore, it is important that nutritional traits also be compared under drought and normal conditions. Therefore, an AM field trail will be conducted with 100 lines that compares performance under drought and normal conditions. (This number was chosen because it is not practical to perform experiments of this nature with more gentotypes.) Drought will be simulated in **North Dakota** (Osorno), **Michigan** (Kelly), **Nebraska** (Urrea), **Washington** (Miklas), and **Puerto Rico** (Porch) by applying irrigation water at 50 to 70% of normal application for optimum yield. Non-stress trials will be planted at all three locations. The same dry bean agronmic trait data will be collected as described above. Correlations between the agronomic and nutritional trait data will be determined. Statistical analyses will be performed to distinguish between constitutive and adaptive loci.

Years 2-4: Within market-class nutritional trait QTL. While AM mapping is a powerful tool, it is necessary to confirm the loci using a different genetic base such as a bi-parental population. Following the nutrient profiling in the first year, parents will be selected from all market classes that vary for many of the nutritional traits. These parents will be crossed, and F5 recombinant inbred populations will be developed and evaluated in the field. In addition, a full nutrient profile will be conducted. QTL analysis will be performed to confirm those loci

important for different nutrients. The following market classes (with geneticisits in parentheses) will be evaluated: pinto (**Osorno**); navy (**Kelly**); black (**Kelly**); great northern (**Urrea**); kidney (**Osorno**); and snap (**Myers**).

v. Statistical analysis. As described above, AM has become an accepted genetic analysis approach. Coupled with that has been the development of user-friendly software for the analysis of AM experiments, including STRUCTURE (Falush et al 2003; Pritchard et al 2000) and TASSEL (Bradbury et al 2007). **McClean** (NDSU) is fully adept at using the software (Wang et al. 2008) and will provide guidance to the breeder group (**Brick, Kelly, Miklas, Osorno, Porch, Urrea)** that will perform the analysis. The **AM statistical analysis** will include 1) genotypic data analysis; 2) phenotypic data analysis; and 3) association analysis of genotypic and phenotypic data. The SNP data from the AM population will analyzed for subpopulation structure (Q) using STRUCTURE software. The final population subgroups will be determined based on Wilcoxson test (Wang et al. 2008) and breeders' knowledge about these germplasm (Camus-Kulandaivelu et al 2007; Flint-Garcia et al 2005). Because of the importance of relative kinship in controlling for false positives in association analysis (Weber et al 2008; Yu et al 2005), a pair-wise kinship relationship matrix using SPAGeDi (Hardy and Vekemans 2002; Yu et al 2005) will be developed. . The relative kinship information will be fitted as a random term to control polygenic effects in the mixed model analysis. Next, a combined analysis of phenotypic data from the multiple-year multi-location nursery trials will provide critical information on the precision of the field experiment and generate the mean performance for each line. A mixed model framework will be used to evaluate each trait of each line (Bernardo 2002). Data inspection with descriptive statistics as well as residual analysis will be conducted for quality control. The final step involves association analysis of genotypic and phenotypic data. First, model fitting will be conducted to determine the influence of population structure and relative kinship on different traits (Yu et al 2005). Individual SNPs will be tested with the selected model (Yu et al 2005) using the TASSEL software. Given the number of SNPs we will evaluate, our association analysis represents a first genome-wide scan that is critical for future experiments designed to discover causative SNPs.

The **bi-parental statistical analysis** will use composite interval mapping approaches. We will use Q-Gene (Johanes and Nelson 2008), a statistical package specifically designed for such analyses. Permutation tests (Churchill and Doerge 1994) will be performed to determine experiment-wise error rates.

vi. The BeanCAP portal and storage of research data. The BIC WWW site (hosted at Michigan State; http://www.css.msu.edu/bic/) is the primary on-line resource for bean research information. **Kelly**, the current BIC site curator, will continue to provide leadership for this site. For example, this site provides all of the accepted disease screening procedures at http://www.css.msu.edu/bic/ResearchTechniques.cfm. Therefore, it is logical to utilize the BeanCAP portal as the project server (http://www.css.msu.edu/bic/BeanCAP.cfm). The portal will advertise BeanCAP activities and provide a schedule of all of the project activities.

From a research perspective, it is our experience that less-experienced database users prefer pre-designed tables to those generated by the user. As examples, the BeanCAP portal will provide:

- research experimental design and data collection methods
- complete pedigree data for all historical varieties and advanced lines

- passport and performance data for the NPGS common bean core collection
- complete passport, genotypic, and phenotypic data for members of the association mapping population (provided by UC Davis database personnel)
- a continuously updated master table containing complete marker/trait association information including linkage distance, allele effect, and PCR detection conditions (provided by UC Davis database personnel)

The Phaseolus Genes database (http://phaseolusgenes.bioinformatics.ucdavis.edu/search/), housed at UC Davis, will be developed as the long term site curating common bean genetic and genomic research data. All of the historical mapping and QTL data generated by the bean research community prior to this project, and all BeanCAP generated phenotypic and marker information, will be incorporated into the database. The database will also include the soybean genome sequence and synteny references between common bean and soybean. This will allow us to anchor a given common bean genetic location to additional electronically mapped common bean loci. For example, all QTL and AM loci will be located on a cMap linkage group that is linked by sequence to additional common bean loci through a GBrowse soybean intermediate. Data mining queries will allow the user to discover additional markers that might be more tightly linked to a trait of interest. The McClean lab has recently demonstrated the value of leveraging the soybean genome by converting a common bacterial blight QTL, that spanned an entire linkage group (Nodari et al. 1993), into one that spans only 5 cM (Jackson et al. 2009). During the analysis and visualization step, users can freely move around GBrowse, CMap and marker database to discover related genome sequences, genetic marker details, and locus annotations. Custom hyperlinks will be provided to facilitate browsing among different types of information. The analysis and visualization module will lead tonew marker discovery by interacting with various custom tracks of sequence information generated by bioinformatics tools and algorithms that implement the necessary steps for marker discovery. These tracks will be based on BLAST of Phaseolus sequences against soybean sequences and primer development for PCR amplification of candidate markers. Throughout the establishment of this tool, we will remain in touch with the National Center for Genomic Research (Santa Fe, NM) to assure close links with the Legume Information System.

To ensure applied utility, Executive Committee plant breeders (**Brick, Kelly, Osorno, Myers**) will also work with a UC Davis staff to exploit the coming interoperability of cMAP and gBrowse. The immediate goal is to design approaches that allow breeders to select markers for incorporation into their projects. This tool will significantly impact breeding by enabling breeders to discover previously unrealized parental combinations that are genetically diverse at regions of the genome highly associated with important agronomic traits. To ensure a more basic utility, Executive Committee molecular geneticists (**Gepts, Grusak, McClean, Myers**) will provide guidance regarding the development of linkages to candidate genes. The linkages will be based on the annotation of the common bean sequences and their soybean orthologs and the physical location of the gene models relative to known genetic loci. This linkage will provide users a more knowledge-rich information basis from which to choose new useful markers. Myers and McClean (unpublished) have shown this functionality by identifying a putative gene for the “stay-green” phenotype in common bean based on the location of an annotated gene model in soybean that genetically cosegregated with the soybean “stay-green” trait.

2. BeanCAP Extension Program: A Nutrition and Health Community of Practice

a. Background. Visualization and learning. Life science courses now emphasize that cellular life is maintained by molecular and cellular processes. The use of **visualization** is a compelling approach that assists biology educators to teach these molecular and cellular processes. From an educational perspective, visualizations greatly aid student understanding of complex processes because they assist in the conversion of an abstract concept into a concrete visual object that can be mentally manipulated (Sekular and Blake 1985). With well-designed visual tools, students can digest large amounts of information in a relatively short time and construct their own personal vision of a process (Kraidy 2002). Graphical visualizations augment textual information (Mayer 1989) and are most effective when they support content for which the learner has had little or no knowledge (Mayer and Gallini 1990).

Computer **animations** foster long-term learning by calling attention to objects during the early steps of instruction. They are effective for learning processes that change over time by reducing the abstractions associated with the temporal transitions of the process (Gagné 1985; Rieber 1994). The benefit of animations is only realized if it is incorporated into a lesson plan that includes lecture and other learning inputs (Rieber 1990). The dual-coding theory (Paivio 1979, 1991) suggests long term memory requires both verbal and visual cues. As such, when animations are combined with narration, dual-coding learning is shown to be supported (Mayer and Moreno 2002). The most effective visualizations reveal the complexity of the objects involved in the process, illustrate how and where the objects interact, provide a spatial representation of the molecules during the process, and smoothly represent the transitional states the objects undergo during the length of the process. High quality, three-dimensional animations have all of these attributes.

The Virtual Cell development team. The NDSU Virtual Cell research and development team has developed a suite of >16 high quality animations depicting basic molecular and cellular processes (http://vcell.ndsu.edu/animations). **The motivation for these animations is to assist novice and advanced students to learn general principles of molecular and cellular biology.** These animations were featured prominently in the Netwatch section of Science Magazine (December 2, 2005; http://www.sciencemag.org/cgi/reprint/310/5753/1401a.pdf), as well as the NSF Discoveries WWW site (http://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=106839). Over 1,200 users have registered, and more than 300 instructors have created accounts so they can download the animations for local usage in their teaching materials. All of the animations were deposited on YouTube in the past year. The YouTube Virtual Cell Channel (http://www.youtube.com/user/ndsuvirtualcell?ob=1) has >1,600 subscribers, and since uploading, the animations have been viewed more than 600,000 times. At times, the Photosynthesis animation was the third (now fifth) most watched video among the featured Education videos on YouTube. The animations are also widely disseminated at the TeacherTube (http://www.teachertube.com/) and Multimedia Educational Resource for Learning and Online Teaching (MERLOT: http://www.merlot.org/) sites.

In addition to animation development, the project team also researches how animations affect the learning of science. A recent publication from our research group clearly demonstrated that animations improve learning (McClean et al. 2005). Those results are supported by other educational researchers in fields such as cell biology (Stith 2004) and biochemistry (Schonborn and Andernson 2006) and for topics such as cholesterol uptake and apoptosis (O'Day 2008). This research has informed our choice of animations as the learning media for the new COP.

b. The Nutritional Genetics and Genomics Community of Practice: Healthy Foods from the Field to the Table

i. The BeanCAP COP – General Concept. The BeanCAP will launch the "**Nutritional Genetics and Genomics: Healthy Foods from the Field to the Table"** eXtension Community of Practice. The **Community of Interest** consists of two **target audiences**: 1) the **general public** seeking knowledge regarding nutrition, health, genetics, and genomics; and 2) the **educational community** looking for education resources that teach students modern components of biology in a manner that links science to the day-to-day health concerns. We will coordinate closely with the plant breeding component of the BeanCAP project to provide materials that are relevant to their educational efforts.

ii. COP Leaders. The first set of COP leaders will be drawn from the BeanCAP project itself. The leaders are: **NDSU**: McClean (Genetics/Genomics), Garden-Robinson (Nutrition/Health), Osorno (Breeding/Genetics), Johnson (Animation Development); **Colorado State University**: Brick (Breeding/Genetics), Thompson (Health and Disease); **USDA/Children's Nutrition Research Center**: Grusak (Genetics/Nutrition/Physiology); **University of California, Davis**: Gepts (Genetics/Genomics). As other individuals become interested in the material, we will look to adding other leaders.

iii. Development of the COP: Animations, Multimedia Materials, Site Management, Design, and Review.

General design approach. The purpose of the COP is **to provide modern, multimedia resources that educate the public regarding the role that plant genetics and genomics play in improving the nutritional status of plants**. Clearly, the materials must be interdisciplinary and integrate a wide range of topics including plant and human physiology, nutritional aspects of health, molecular and cellular biology of metabolism, and genetics and genomics of plant improvement. These materials must be designed in a manner that provides immediate answers to questions, but are also be part of a learning path that integrates the multiple disciplines.

To ensure consistency, we will adopt the VCell development model and build two interrelated modules simultaneously. The **first pair of modules** will focus on 1) root biology and the role soil chemistry plays on nutrient uptake; and 2) the flow of soil minerals from the root to various parts of plants. The purpose of these modules is to explain a) that important food minerals come from the soil, b) that soil chemistry (e.g. low or high pH) dictates what minerals are taken up, and c) how minerals move through the plant and eventually end up in the edible part of the plant. The **second pair of modules** will focus on a few minerals and show how they move from our digestion system to the various organs in our body, and what health problems arise from mineral deficiencies in our diet. The **third pair of modules** will be a basic lesson in genetics that emphasizes the fact that many genes are involved in the uptake and distribution of the minerals throughout the plant, and that the genome of the plant is required to produce all of the proteins necessary for functional uptake and distribution of minerals throughout the plant. And the **fourth pair of modules**, will focus on how plant breeding utilizes the genetic information for trait (mineral uptake) improvement, and the role of genomics in developing the tools that assist the breeder in trait selection.

This development plan is intended to illustrate a continuum that demonstrates 1) how plant metabolism dictates the nutritional status of a crop; 2) how the plant nutritional status affects human nutrition and directs our choice of foods; and 3) how knowledge of the genes controlling plant nutritional status can be improved using genetics and genomics approaches. It is our initial intention to follow this model, but if inherent education flaws arise, it will be modified.

Animation development cycle. Each module will consist of a narrated animation and supporting multimedia WWW pages. Animation development begins with the development of a storyboard and narrative by project scientists. The artists then use ***Autodesk Maya***, a powerful software environment that supports all aspects of modern animation. For example, it has been used for such recent motion pictures as *Spiderman*, *Ice Age*, and *The Lord of the Rings* trilogy. (The Academy of Motion Pictures Arts and Sciences recently awarded Autodesk an Oscar for scientific and technical achievement for its development of Maya.) **Our project has a working relation with Autodesk, and they provide ten Maya Complete (full commercial version) licenses at no charge.**

To expedite the development process and promote visual consistency, we will reuse, whenever possible, the VCell library that currently contains over 250 molecular and cellular objects. In addition to speeding up the development process, reusing objects provides the students with a degree of familiarity and consistency that aids student learning and provides a visual understanding that all processes are part of a continuum of interrelated steps.

Once the animations are complete, they will be rendered at the NDSU Center for High Performance Computing (CHPC; http://www.ndsu.edu/chpc/) Bewoulf cluster that consists of 64 dual-processor computers. Computation time is balanced based on user needs, and our animations are typically rendered in 18-36 hours. Following rendering, lead in stills and voiceover narrations are added to the animations. Currently, it takes an artist about six to eight weeks to develop a completed animation, and an additional one week to implement the WWW pages used to deliver the learning module. Given that our budget provides for one full-time and one half-time artist, we safely estimate that we will be able to produce **4-6 animations** and their associated WWW pages **per year**. We thus anticipate that we will develop **between 16 and 24 modules during the BeanCAP project**.

Alternative animation formats. All animations are rendered at a high quality 640x480 resolution and saved in the .wmv file format because it is the easiest format for embedding the animations in a WWW page. The internet connection of some instructors is too slow to download these large size files. Some instructors wish to embed the animations in their learning modules, such as MS PowerPoint presentations. To meet these needs, all animations will be offered in alternate formats (.avi, .mov, and .mpeg4) rom the VCell server at http://vcell.ndsu.nodak.edu/animations/downloads. The user will register so we can keep track of usage. All stills and narrations will be available from the same download site.

Supporting multimedia materials. We will adopt the VCell model of offering still images of the animation as supporting materials. A selection of images that focus the learner on important learning points will be extracted from the rendered animation, and information bubbles that feature the key knowledge points of the animation will be added. One set will be developed for the general public at the ninth grade level. A second (and more comprehensive) set will be developed for the college freshman to junior level.

COP site activities. During the first 12 months, we will develop an extensive list of **Frequently Asked Questions (FAQ)**. These will be drawn from our knowledge of the field, questions found in textbooks, and those we receive in the classroom or from our colleagues in

other fields. As the COP goes public, we will begin to collate questions and expand the FAQ as appropriate. The leaders will share the FAQ and their answers among themselves to ensure accuracy.

Prior to the launch date, the content portion of the Wiki site will be created. A chart will be created that will suggest possible learning paths. The content will be organized around the animations. Users will select a topic, link to a page where they will have the choice of either directly reviewing 1) the animation; 2) stills; 3) the animation narrative; or 4) learning objectives. All of these topics will be interlinked to allow the user multiple learning paths.

Given the time consuming nature of animation development, the **peer review** process will be **ongoing**. Reviews will be provided initially at the time the storyboard and narration are drafted. This is a critical point because it will dictate the visual nature of the animation objects as well as the interactions and timing of interactions among objects. The National Standards for Science Learning for grades 9-12 (http://www.nap.edu/openbook.php?record_id=4962&page=106) that relate to "The cell", "matter, energy, and organization in living systems", and "interdependence of organisms" will be considered during development. That will be important as we offer these learning modules to the grade 9-12 audience. The next review will consider the visual nature of the objects to ensure their appearance is appropriate for the target audience. When an animation is completed, it will be reviewed to ensure the interactions are smooth and not confusing. Finally, the lead-in materials added in post-production will be reviewed to ensure they are engaging and spur the interest of the learner. All of the **reviews** will be performed **by project leaders**.

iv. Launching and Marketing the COP. Because of their extensive experience with animation development and dissemination, the NDSU members of the COP will serve as the marketing team. The obvious first venue is educators. The large database of VCell Animation users and those registered with the YouTube channel will be contacted within six months of the start of the project. The announcement will describe the nature of the site, the resources available, and the utility of the site as an educational venue. In the second year of the project, we will make a presentation at the Plant and Animal Genome Conference. The goal will be expose educators to the value of the materials that we are developing. Similar presentations will be made by leaders who attend other meetings, for example, the Crop Science Society of America.

v. Managing the COP. Project leaders will provide overall governance of the COP. **McClean and Garden-Robinsion** will take the initial responsibility for direct leadership at the onset and will be assisted by other leaders. **Deb Pankow** (NDSU), the past chair of the Financial Security for All COP, will act as our COP mentor. (See attached commitment letter.) A monthly organizational meeting will be held either via conference call or video conference. Progress in all aspects of the COP will be reviewed and specific assignments will be made to leadership. Christina Johnson will be the direct contact person with the eXtension organization. She handles all animation design, still development, and interactions with youtube.com and teachertube.com for the VCell Animation project. She will be directly responsible for uploading all project information and ensuring it meets eXtension organizational and formatting requirements. She will also work directly with eXtension to ensure the COP site design complies with eXtension guidelines.

III. BeanCAP Education Program: Engaging Students to Broaden the Human Resource Base for the Plant Breeding Industry

Although there is a **strong demand** for plant breeders, especially in the private sector, the **applicant pool** has **decreased** in recent years (Frey, 1996; Guner and Wehner, 2003; Trexler et. al., 2005). Importantly, the current numbers of plant breeding graduates do not meet the number of open plant breeding positions. The gap between demand and supply of plant breeders has been caused by several factors including decreased funding, increased retirements of experienced breeders, and lack of interest from students. Another factor is the emphasis made at universities towards training in molecular biology and other new technologies rather than conventional plant breeding. Yet, plant breeding remains a vibrant, multi-disciplinary science characterized by its ability to reinvent itself by absorbing and using novel scientific findings and approaches such as those offered by the science of genomics (Gepts and Hancock, 2006). Therefore, as the training arm of the plant breeding, universities need to engage students earlier and expose them to the breadth of activities in which a plant breeder is engaged.

From an industry perspective, new hires lack the ability to walk the field and observe important phenotypic differences, and to conduct field experiments with the optimum management practices and statistical design to distinguish genotypic differences (Bliss, 2006; Fehr, 2007). These are essential skills for an individual to be a successful plant breeder. Simultaneously, the breeder should be able to sit with the molecular geneticist and discuss the integration of classical and molecular genetics in a manner that more efficiently reduces the time to new cultivar releases. Therefore, the key to being a successful plant breeder today resides in the ability to integrate both the field breeding techniques and the molecular tools. In addition, the plant breeder needs excellent communication and teamwork skills because plant breeding is a team-oriented effort involving professionals from different disciplines and locations (Baenzinger, 2006). We believe exposing students to all of these aspects of plant breeding will provide them with the necessary information to make an informed career decision.

The BeanCAP will organize three activities in order to assure a flow in the human resource pool for plant breeding. These activities will target students from different education levels ranging from high school students to undergraduate, and graduate students in some instances. We believe that early exposure of high school and undergraduate students to plant breeding is the best way to generate interest in pursuing a career in this area of agriculture. Undergraduate students often only learn about plant breeding in their junior and/or senior year. By this time, most of them have already chosen another career. **Therefore, direct high school and undergraduate experiential learning activities and recruitment are our main activities of the BeanCAP education component.** Other CAP projects have focused on training graduate students whom are already involved and/or pursuing careers in plant breeding. Therefore, **our plan is to show the opportunities that plant breeding offer by bringing exposing new students to the discipline**.

***i. Objectives and Activities*:** The BeanCAP education program has two main objectives:

1. Early exposure of high school students to encourage them to pursue of careers in plant breeding.
2. Hands-on training of undergraduate students at the interface of plant breeding, genetics, and genomics.

Presentations, field/greenhouse trips, and hands-on training will expose students to the general functions of plant breeding. This includes the science of plant breeding (what a plant breeder does), Mendelian genetics, breeding for nutritional traits, qualitative vs. quantitative trait breeding, and marker-assisted selection. Potential students will be invited to participate in summer and year-round internships that demonstrate the applied and basic activities involved in a breeding program. These activities will directly expose many high school students to the day-to-day activities of plant breeding and genetics across the four years of the project. The collaborators of these education activities are: **Brick** (CSU), **Kelly** (MSU), **Osorno** (NDSU), and **Urrea** (UNL). **Drs. Osorno and Urrea will lead this aspect of the project**.

ii. Early Exposure High School Activities. Early exposure activities will target high school students. Participants from each university will visit on an annual basis at least three local and regional high schools to discuss the science of plant breeding. Secondly, follow-up visits by these high school students will be arranged. Our **goal** is to reach at least **1,200 high school** students (75 students per year per location) over the 4 years of the project. Additionally, one interested and motivated high school student will be recruited annually into the breeding program to serve an apprenticeship.

Individual high schools will be contacted and interested science teachers will be identified that are willing to host a class visit. Given the connection that rural students have with agriculture, rural schools will be one of the target clientele. The visit by the breeder will be general in nature and focus on genetics as a science and plant breeding as an applied genetics career. These will be generally, 30-45 minutes in length.

Following the in-class presentation, visits by high school classes to a breeding program will be arranged. These visits will emphasize key concepts of plant breeding such as genetic variability and the importance of selection. The need for crops with better nutritional quality will be a featured example. Classroom materials will be developed for school teachers to take back to the school where they can demonstrate important concepts such as the importance of water and light for plant development. This activity would help teachers have more general discussions of water, agriculture and society.

At all schools, one motivated and interested student will be brought into the program. Where possible, students will enroll in a program to receive college credit for their internship. For example, Huskers Horizons is a program offered through the College of Agricultural Sciences and Natural Resources at UNL that offers high school seniors college credit for activities like those we are planning. High school students from a Scottsbluff area high school working with Dr. Urrea will take part in this program. We will look for similar opportunities at the other locations.

Promotional materials such as brochures, flyers, and posters describing plant breeding will be developed for the high school audience (**Osorno and Urrea leads**). A brochure describing the science of plant breeding will be prepared for distribution during high school visits and campus activities. The brochure will be question oriented and answer questions such as: 1) What does a plant breeder do on a daily basis? 2) What is the role of in plant breeding? 3) How are genetic concepts applied to solving several production problems? 4) What roles have famous plant breeders and geneticists played in solving society problems? The main goal is to show students the career possibilities and excite them about a career in plant breeding. All materials generated

will be publicly available in a website of general access so they can be used in recruiting activities and workshops made by other groups and institutions in the future.

iii. Early Undergraduate Activities. Summer and year-round internships will be offered by all participant universities each year. A total of eight students per year (two per campus), will have an 11 week summer internship opportunity to gain hands-on experience in a US dry bean breeding program. These will come for the local universities. In the case of UNL, students attending Scottsbluff Community College will be recruited.

The **summer program** will be announced during the fall and spring semesters. In some cases, students in the program may register for a 1-3 intern credits. **This experience will focus on field** (whole plant phenotyping, field operations, production and agronomy, experimental design, disease evaluation), **and lab** (nutrition profiling, DNA isolation, molecular marker screening and detection, analysis of marker data) **aspects of plant breeding**. Students will also visit other breeding programs. For example, NDSU has 13 breeding programs that the students can learn from. These activities will be designed to be informative to the undergraduate students and are not intended to duplicate the graduate training offered by the other CAP projects. At the same time, though, we will leverage the more advanced materials developed by the other CAP projects by collating all of the learning materials and hosting them on the BeanCAP WWW site. This will not only serve our students, but provide additional coordination for all CAP educational activities.

Students will be admitted into this program based on their potential to be recruited for M.S. or Ph.D. programs at one of the participating institutions. All the applicants should have at least a 2.8 GPA. An informational seminar about research opportunities in this program will be offered each fall semester. Candidates will apply in fall semester and are interviewed early in the spring of each year. Students will be required to report on their experiences in a departmental seminar or other appropriate forum. The goal is to motivate these students to pursue a graduate degree in plant breeding.

Undergraduate students (sophomore or early juniors) that express an interest in plant breeding will be recruited into **school year internships** at each university. They will work part-time during the fall and spring semesters (10 hours per week). **These students will participate in a student-centered research project that couples breeding activities such as selection, field and greenhouse trials, nutritional screening with molecular marker, and genomics-assisted breeding techniques.** This will give them an exposure to all aspects of a breeding program. At each campus, a ½ time Ph.D. student (**on non-BeanCAP funds**) will supervise the interns' activities and act as mentors. Faculty will participate actively and interact with the interns on a regular basis through group meetings and mentoring sessions.

iv. Assessment and Tracking Participant Accomplishments. The success of the educational component of the project will be measured based on both quantitative and qualitative evidence. A data base of all student participants will be maintained and contact information will be periodically updated. Exit interviews will be conducted at the conclusion of each project activity. Tracking of participants will continue throughout the project as well as after the termination of the project. The database will be used as a way to periodically contact ex-participants and document what impact participation in the project has had on recruitment and education of plant breeders.

Bibliography & References Cited

Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, et al. (2008) Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genomics 92:478-487.

Andersen JW, Story L, Sieling B, Chen W-JL, Petro MS, Story J (1984). Hypocholesterolemic effects of oat-bran or bean intake for hypercholesterolemic men. Am. J. Clinical Nutri. 40:1146-1155.

Baenziger, P.S. 2006. Plant breeding training in the U.S.A. HortScience 41:28–29.

Bernardo R (2002) Breeding for Quantitative Traits in Plants Stemma Press, Woodbury, Minnesota.

Bliss, F. 2006. Plant breeding in the private sector of North America. Hortscience 41:45-47.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) Tassel: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2644-2635.

Brown PJ, Rooney WL, Franks C, Kresovich S (2008) Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics 180:269-637.

Buys H, Hendricks M, Eley B, Hussey G (2002) The role of nutrition and micronutrients in paediatric HIV infrection. SADJ 57:454-456.

Camus-Kulandaivelu L, Veyrieras J, Gouesnard B, Charcosset A, Manicacci D (2007) Evaluating the reliability of structure outputs in case of relatedness between individuals. Crop Science 47:887 - 890.

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971.

Collins NC, Tardieu F, Tuberosa R (2008) Quantitative trait loci and crop performance under abiotic stress: where do we stand. Plant Physiology 147:469-486.

Cregan PB, Buso GSC, Souza T, Choi I-Y, Hyten DL, Song Q, Pastor-Corrales MA (2009) Single nucleotide polymorphism (SNP) DNA marker discovery in common bean. Abstracts of Plant and Animal Genomes Conference, San Diego, CA, Jan 10-14, 2009. http://www.intl-pag.org/17/abstracts/W40_PAGXVII_301.html (validated 24 March 2009).

Ducrocq S, Madur D, Veyrieras J-b, Camus-Kulandaivelu L, Kloiber-Maitz, Presterl T, Ouzunova M, Manicacci D, Charcosset A (2008) Key impact of Vgt1 on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. Genetics 178:2433-2437.

Fairweather-Tait S, Lynch S, Hotz C, Hurrell R, Abrahamse L, Beebe S, et al. (2005) The usefulness of in vitro models to predict the bioavailability of iron and zinc: a consensus statement from Harvest-Plus expert consultation. International Journal of Vitamin Nutrition Research 75:371-374.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567-87.

Flint-Garcia SA, Thuillet AC, J. Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES. 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054-64.

Fehr, S. 2007. An endangered species? Seed World. May:4-5.

Feregrino-Perez AA, Berumen LC, Garcia-Alcocer G, Guevara-Gonzalez RG, Ramos-gomez M, Reynoso-Camach R, Acosta-Gallegos JA, Loarca-Pina G (2008) Composition and chemopreventive effect of polysacchrides from common bean (Phaseolus vulgaris L.) on azoxymethane-induced colon cancer. Journal of Agricultural and Food Chemistry 56:8737-8744.

Frey, K. 1996. National plant breeding study—I. Human and financial resources devoted to plant breeding research and development in the United States in 1994. Iowa Agric. Home Econ. Spec. Rep. 98. Iowa State Univ. Ames. Also available online at: http://www.ers.usda.gov/Data/PlantBreeding/Plant%20Breeding.pdf

Freyre R, Skroch P, Geffroy V, Adam-Blondon A-F, Shirmohamadali A, Johnson W, Llaca V, Nodari R, Pereira P, Tsai S-M, Tohme J, Dron M, Nienhuis J, Vallejos C, Gepts P (1998) Towards an integrated linkage map of common bean. 4. Development of a core map and alignment of RFLP maps. Theor Appl Genet 97: 847-856

Gagné, RM (1985). The conditions of learning (4th ed.). New York: Holt, Rinehart, & Winston.

Gelin JR, Forster S, Grafton KF, McClean PE, Rojas-Cifuentes GA (2007) Analysis of seed zinc and other minerals in a recombinant inbred population of navy bean (*Phaseolus vulgaris* L.). Crop Sci 47:1361-1366

Gepts, P. (1990) Biochemical evidence bearing on the domestication of *Phaseolus* (Fabaceae) beans. Econ. Bot. 44, 22-38.

Gepts P (1998) Origin and evolution of common bean: past events and recent trends. HortScience 33:1124-1130

Gepts, P., Bliss, F.A. (1986) Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ. Bot. 40, 469-478.

Gepts, P, Debouck, D. 1991. Origin, domestication, and evolution of common bean (*Phaseolus vulgaris* L.). *In*: A. van Schoonhoven and O. Voyest (Eds.) *Common Beans: Research for Crop Improvement*. C.A.B. Intl., Wallingford, UK and CIAT, Cali, Colombia, pp. 7-53.

Gepts, P., and J. Hancock. 2006. The future of plant breeding. Crop Sci. 46:1630-1634.

Grusak MA (2002) Enhancing mineral content in plant food products. Journal of the American College of Nutrition 21:178S-183S.

Guner, N., and T.C. Wehner. 2003. Survey of U.S. Land-Grant Universities for Training of Plant Breeding Students. Crop Sci. 43:1938-1944.

Hangen L, Bennink MR (2002) Consumption of black beans and navy beans (*Phaseolus vulgaris*) reduced azoxymethance-induced colon cancer. Nutr Cancer 44:60-65.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Eco. Notes 2:618-620.

Harjes CE, rocheford TR, Bai L, Brutnell TP, Kandianis CB, et al. (2008) Natural genetic variation in *Lycopene Epsilon Cylcase* tapped for maize biofortification. Science 319:330-333.

Hu Y, Cheng Z, Heller L, Karsnoff S, Glahn RP, Welch RM (2006) Kaempferol in red and pinto bean seed (Phaseolus vulgaris L.) coats inhibits iron bioavailability using an invitro digestion/human Caco-2 cell model. Journal of Agricultural and Food Chemistry 54:9254-9261.

Jackson J, L. Malete, M. Nnyepi, L. Mokgatlhe, J. Makhanda, G. Anabwani, P. Lyoka, M. Bennink, L. Weatherspoon (2007) Integration of Food and Nutrition into HIV and AIDS Programmes in Botswana. Presented at the European Nutrition Conference, Paris, France, July, 2007.

Jackson SA, Schlueter J, McClean P, Gepts P, Blair M, Vallejos E (2009) Genomics in common bean: leverging the soybean genome for bean improvement. Abstracts of Plant and Animal Genomes Conference, San Diego, CA, Jan 10-14, 2009. http://www.intl-pag.org/17/abstracts/W29_PAGXVII_235.html (validated 24 March 2009)

Joehanes R, Nelson JC (2008) QGene 4.0, and extensible Java QTL-analysis platform. Bioinformatics 24:2788-2789.

Jorde LB (2000) Linkage disequilibrium and the search for complex disease genes. Genome Res 10:1435-1444

Kaplan, L., and Lynch, T.F. (1999) *Phaseolus* (Fabaceae) in archaeology: AMS radiocarbon dates and their significance for pre-Columbian agriculture. Econ. Bot. 53, 261-272.

Kelly JD, Gepts P, Miklas PN, Coyne DP (2003) Tagging and mapping genes and QTL and molecular marker-assited selection for traits of economic importance in bean and cowpea. Field Crops Res 32:135-154

Kraidy, U (2002). Digital media and education: cognitive impact of information visualization. Journal of Educational Media, 27, 95-106.

Laparra JM, Glahn RP, Miller DD (2008) Bioaccessibility of phenols in common beans (Phaseolus vulgaris L.) and iron (Fe) availability to Caco-2 cells. Journal of Agricultural and Food Chemistry 56:10999-11005.

Mackay TFC (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35: 303-339

Madhujith T, Naczk M, Shahidi F. (2005) Antioxidant activity of common beans (Phaseolus vulgaris L.). Journal of Food Lipids 11:220-233.

Malete L, Jackson J, Mokgathe L, Nnyepi M, Makhanda J, Lyoka GP, Anabwani G, Bennink M, Weatherspoon L (2007) Integrating Nutrition into the Care of HIV+ Children : The Impact of Food Based Intervention on Growth, Cognition and Motor Development of 6-14 Year Old Children in Botwana. Presented at the World Food Programme (WFP) Workshop on Advanced Nutrition Training: Measuring and Interpreting Malnutrition and Mortality, Livingstone, Zambia, November, 2007.

Mayer, RE (1989). Systematic thinking fostered by illustrations in scientific text. Journal of Educational Psychology, 81, 240-246.

Mayer, RE, and Gallini, JK (1990). When is an illustration worth ten thousand words? Journal of Educational Psychology, 82, 715-726.

Mayer, RE, and Moreno, R (2002). Animation as an aid to multimedia learning. Educational Psychology Review, 14, 87-99.

McClean, P, Johnson, C, Rogers, R, Daniels, L, Reber, J, Slator, BM, Terpstra, J, and White, A (2005). Molecular and cellular biology animations: development and impact on student learning. Cell Biology Education, 4, 169-179.

McClean PE, Lee RK (2007) Genetic architecture of chalcone isomerase non-coding regions in common bean (*Phaseolus vulgaris* L.) Genome 50:203-214.

McClean, P.E., Kami, J., Gepts, P. (2004a) Genomics and genetic diversity in common bean. *In*: R.F. Wilson, H.T. Stalker, and E.C. Brummer (Eds.) *Legume Crop Genomics*. AOCS Press, Champaign, Illinois, USA, pp. 60-82.

McClean, P.E., Lee, R.K. Miklas, P.N. (2004b) Sequence diversity analysis of dihydroflavanol reductase intron 1 in common bean. Genome 47, 266-280.

McClean PE, Lee RK, Otto C, Gepts P, Bassett,MJ (2002) Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean *(Phaseolus vulgaris* L.) J Heredity 93:148-152.

McClean PE, Myers JR, Hammond JJ (1993) Coefficient of parentage and cluster analysis of North American dry bean cultivars. Crop Science 33:190-197

Miklas PM, Kelly JD, Beebe SE, Blair MW (2006) Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. Euphytica 147:105-131

Miklas PN, Singh SP (2007) Common Bean. *In* Genome Mapping and Molecular Breeding in Plants, Volume 3, Pulses, Sugar, and Tuber Crops. C. Kole (ed). Springer-Verlag, Berlin, Heidelberg.

Nodari RO, Tsai SM, Guzman P, Gilbertson RL, Gepts P (1993) Towards an integrated linkage map of common bean. III. Mapping genetic factors conrolling host-bacteria interactions. Genetics 134:341-350

Nordborg M, and Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet.18:83-90

Oatway l, Vasanthan T, Helm J (2001) Phytic acid. Food Review International 17:419-431.

O'Day D (2008) Using animations to teach biology: past and future research on the attributes that underline pedagogically sound animations. The American Biology Teacher 70, 274-278.

Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. The Plant Cell 15:1795-1806.

Paivio, A (1979). Imagery and verbal processes. Hillsdale, NJ: Lawrence Erlbaum Associates.

Paivio, A (1991). Dual coding theory: Retrospect and current status. Canadian Journal of Psychology, 45, 255-287.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-59.

Rieber, LP (1990). Using animation in science instruction with young children. Journal of Educational Psychology, 82, 135-140.

Rieber, LP (1994). Computers, Graphics, and Learning. Brown and Benchmark.

Rocha-Guzman NE, Gallegos-Infante JA, Gonzalez-Laredo RF, Preza AM, Lerma Y, Ibarra-Perez FJ. (2007) Antioxidant activity in cotyledon of black and yellow common beans (*Phaseolus vulgaris* L.) Research Journal of Biological Sciences 2:112-117.

SADH (South Africa Department of Health) (2001) South African national guidelines on nutrition for people living with TB, HIV/AIDS and other debilitating diseases. Ministry of Health, Pretoria, South Africa. http://www.sahealthinfo.org/nutrition/sanational.htm (accessed 1/26/2007).

Savarino A, Pescarmona GP, Boelaert JR (1999) Iron metabolishm and HIV infection: reciprocal interactions with potentially harmful consequences? Cell Biochem Funct 17:279-287

Schlecht NW, Slator B, and McClean P (2009). Usage of a web-based factorial experiment testing system (FETS). Proceedings of Midwest Instructional Computing Symposium, April 17-18, 2009, in press.

Schonborn KJ, Anderson TR (2006) The importance of visual literacy in the education of biochemists. Biochemistry and Molecular Biology Education. 34:92-102.

Sekular, R and Blake, R. (1985). Perception. New York: Alfred A. Knopf.

Singh, S.P., Gepts, P., DeBouck ,D.G. (1991). Races of common bean (*Phaseolus vulgaris*, Fabaceae). Econ.Bot. 45, 379-386.

Singh SP, Teran H, Lema J, Webster DM, Strausbaugh CA, Miklas PN, Schwartz HF, Brick MA (2007) Seventy-five years of breeding dry bean of the Western USA. Crop Sci 47:981-989

Singh SP, Westermann DT (2002) A single dominant gene controlling resistance to soil zinc

deficiency in common bean. Crop Sci. 42: 1071-1074
Sonnante G, Stockton T, Nodari RO, Becerra Velasques VL, Gepts P (1994) Evolution of genetic diversity during the domestication of common bean ((*Phaseolus vulgaris* L.). Theor Appl Genet 89:629-635
Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. The Plant Journal 3: 739-744
Stith, BJ (2004). Use of animation in teaching cell biology. Cell Biology Education, 3, 181-188.
Stracke S, Haseneyer G, Veyrieras, Geiger HH, Sauer S, Graner A, Piepho H-P (2009) Association mapping reveals gene action and interactions in the determination of flowering time in barley. Theoretical and Applied Genetics 118:259-273.
Thompson LU, Zhang L (1991) Phytic acid and minerals: effect on early markers of risk for mammary and colon carcinogenesis. Carcinogenesis 12:2041-2045.
Thompson MD, Brick MA, McGinly JN, Thompson HJ. 2009. Chemical composition and mammary cancer inhibitory activity of dry bean. Crop Sci 49:179-186.
Tietyen-Clark J. (1986) The benefits of bean. p. 1. *In* J.W. Andersen (ed.) Diabetes and Nutrition News. HCF Diabetes Foundation, Lexington, U.S.A.
Trexler, G., A. Acquaye, K.J. Frey, and A.M. Thro. 2005. Public Plant Breeding Resources in the US: Study Results for the year 2001. Available at: http://www.csrees.usda.gov/nea/plants/part/pbgg_part_study.html
Vallad G, Rivkin M, Vallejos CE, McClean PE (2001) Cloning and homology modeling of a Pto-like protein kinase family of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 103:1046-1058.
Vallejos CE, Sakiyama NS, Chase CD (1992) A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics 131:733-740.
Wang TL, Domoney C, Hedley CL, Casey R, Grusak MA (2003) Can we improve the nutritional quality of legume seeds? Plant Physiology 131:886-891.
Wang J, McClean P, Lee R, Goos RJ, and Helms T (2008) Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. Theoretical and Applied Genetics 116:777-787.
Waters BM, Grusak MA (2008) Quantitative trait locus mapping for seed mineral concentratioins in two *Arabidopsis thaliana* recombinant inbred populations. New Phytologist 179:1033-1047.
Weber A, Clark RM, Vaughn L, Sánchez-Gonzalez JDJ, JYu J, Yandell BS, Bradbury P, Doebley JF (2008) Major regulatory genes in maize contribute to standing variation in Teosinte (*Zea mays* ssp. parviglumis). Genetics (in press).
Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler IV ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. The Plant Cell 16:2719-2733.
Winham DM, Hutchins AM, Johnston CS (2007) Pinto bean consumption reduces biomarkers for heart disease risk. Journal of the American College of Nutrition 26:243-249.
Wong JH, Zhang XQ, Wang HX, Ng TB (2006) A mitogenic defensin from white cloud bean (*Phaseolus vulgaris*). Peptides 27, 2075-2081.
Yu J, Arbelbide M, Bernardo R (2005) Power of in silico QTL mapping from phenotypic, pedigree, and marker data in a hybrid breeding program. Theor Appl Genet 110:1061-7.
Zhou JR, Erdman JW (1995) Phytic acid in health and disease. CRC Critical Reviews in Food Science and Nutrition. 35:495-508.

(a) Key Personnel

The following table describes the roles of the PD, the Co-PDs, and other service providers. The table is organized around the major activities of the BeanCAP Project. These are also noted in the Project Narrative.

Participant (Location)	**Role and responsibility**
Phillip McClean (NDSU)	Project administrator CAP marker development
Perry Cregan and David Hyten (USDA/Beltsville)	Solexa sequencing within market classes 454 sequencing of Jalo EEP558 SNP marker development Illumina Golden Gate Assays
Mike Grusak (USDA/Houston)	Mineral analysis CaCo-2 cell tests
Mark Brick (CSU) Elizabeth Ryan (CSU) Henry Thompson (CSU)	Antioxidant tests Phenolics tests Anthocyanins test Soluble/insoluble fiber determination
Mehmet Tulbek (NDSU)*	Protein analysis Fiber analysis Oil analysis
Jim Kelly (MSU) Karen Cichy (USDA/East Lansing)**	Phytate test
Jim Myers (OSU)	Carotenoid test (snaps only) Vitamin C test (snaps only) Fiber (snaps only; different than CSU test)
Mark Brick (CSU)[1,5] Jim Kelly (MSU)[1,2,4,5] Phil Miklas (USDA/Prosser)[2,5] Jim Myers (OSU)[3,4,5] Juan Osorno (NDSU)[1,2,4,5] Tim Porch (USDA/Mayaguez)[2,5] Carlos Urrea (UNL)[1,2,4,5]	[1]Dry bean association mapping field test [2]Dry bean abiotic stress field test [3]Snap bean association mapping test [4]Bi-parental population field test [5]Statistical analysis
Paul Gepts (UC, Davis)	DNA analysis cMAP implementation gBrowse implementation Electronic marker selection

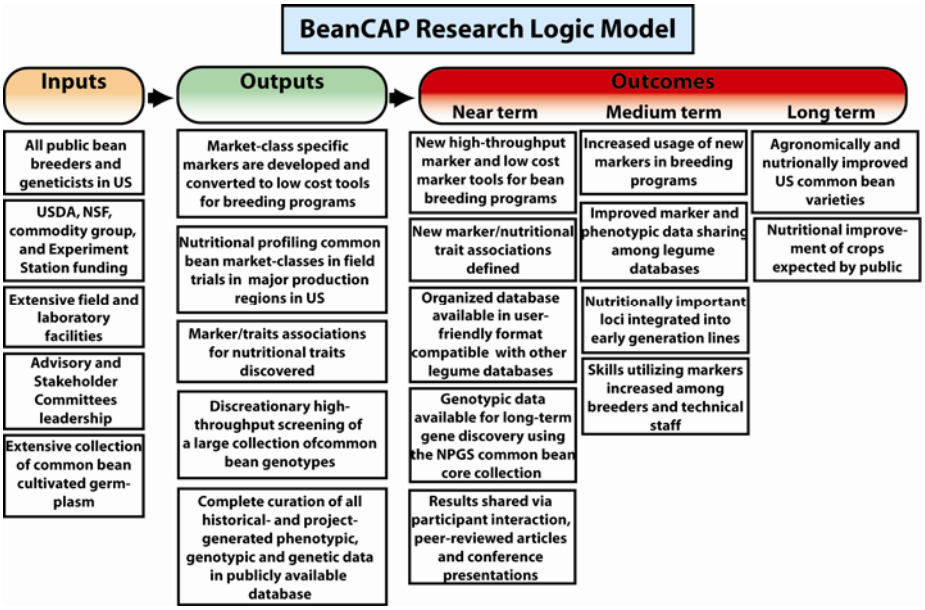
Mark Brick (CSU)[1,2,3] Julie Garden-Robinson (NDSU)[1,2,3,4] Paul Gepts (UC, Davis)[1,2,3] Mike Grusak (USDA/Houston)[1,2,3] Christina Johnson (NDSU)[1,2,3,4] Phil McClean (NDSU)[1,2,3,4] Juan Osorno (NDSU)[1,2,3] Deb Pankow (NDSU)[5] Henry Thompson (CSU)[1,2,3]	[1]COP leaders [2]COP FAQ development [3]COP experts [4]COP launch [5]COP mentor
Christina Johnson (NDSU) Phil McClean (NDSU) Brian Slator (NDSU)	Animation development
Mark Brick (CSU) Jim Kelly (MSU) Juan Osorno (NDSU) Carlos Urrea (UNL)	High school visits Undergraduate recruitment and training Plant breeding education materials Plant breeding recruitment materials

*Not listed as Co-PD. See letter of cooperation in "Documentation of Collaboration."
**Dr. Karen Cichy will begin at the USDA/Sugarbeet and Bean Research Unit, East Lansing, MI in the summer of 2009.

(b) Logic Model

Research Logic Model

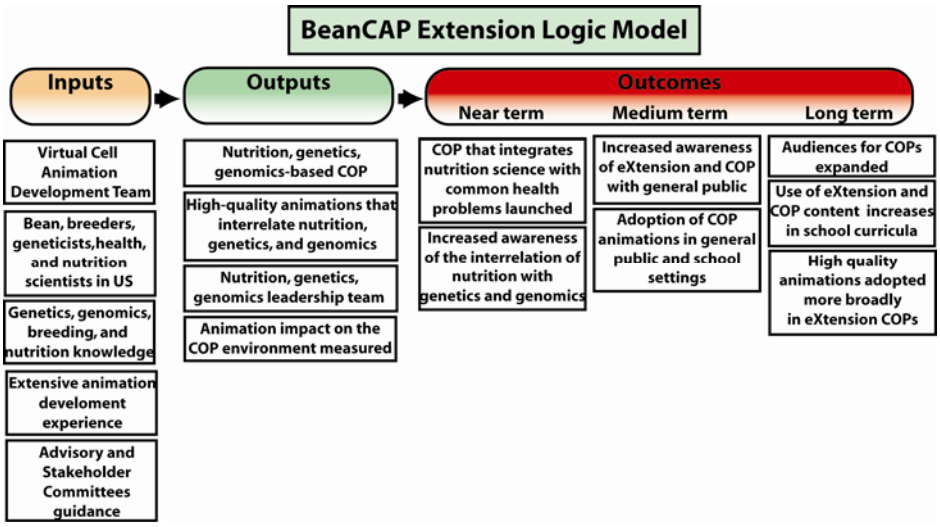
The current situation is that common bean is considered a major nutritional crop. **Improvement of nutritional traits is limited** by the fact that loci controlling a large collection of nutritional traits have not been discovered in a systematic species-wide or market-class-specific manner. Also, large scale, species-wide nutritional profiling is lacking, and the tools for genome-wide screening for important nutritional loci are not available. To systematically change the nutritional status of common bean, discovering the specific nutritional trait/loci associations is a **priority**. Genome-wide tools that have utility for all, not just nutritional traits, should be developed and converted into low cost breeder markers to support this priority. This data must be shared for national and global improvement. To accomplish this, **we assume that** 1) within market-class (or race) variation is sufficient to discover within market class loci that can improve all (or at least most) market classes; 2) species-wide profiling will uncover new loci with major impact on nutritional trait expression; and 3) addressing such a broad suite of nutritional traits is best approached using community-wide human and technical resources.



Extension Logic Model

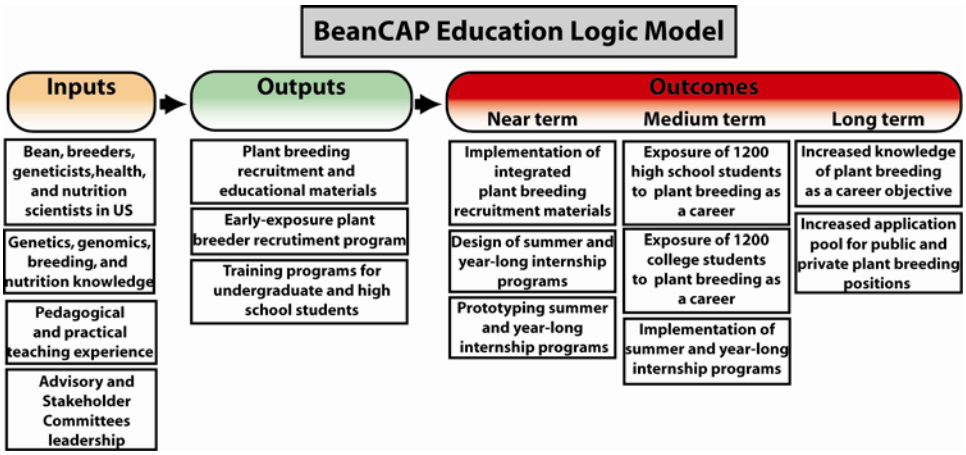
The current situation is that the relationships between nutrition, genetics, and genomics are not understood by the general public. **Changing this knowledge gap is limited** by the fact that effective learning tools that capture the dynamic nature of the interactions and cover the multiple biological aspects of nutrition are not available. To remedy this situation, the development of learning modules that capture the interrelationships are a **priority**. Animations that capture the dynamic nature of nutrition should be developed and coupled with a broad-reaching learning environment such as an eXtension Community of Practice. To ensure the public's awareness of the interrelationship increases, the COP learning modules must be shared

both nationally and globally. To accomplish this, **we assume that** 1) the public is interested in this topic; and 2) animations are the most appropriate learning media.



Education Logic Model

The current situation is that human resource pool for plant breeding falls far short of meeting the needs of the profession. **Increasing the pool is limited** by the fact that fewer and fewer students are knowledgeable about the profession and hands-on early experiences are not widely available. Developing early exposure programs are a **priority** to address the shortfall. Face-to-face interactions and breeder-guided training that focus on a single improvement goal, such as nutritional status, would support this priority. These programs must be shared to have a national or global impact. To accomplish this, **we assume that** 1) face-to-face interactions that focus on the daily aspects of a profession is an effective initial recruitment approach; and 2) extensive hands-on training is an effective tool that transitions interest into commitment.



(c) Management Plan

1. Administrative Groups. The direction of the BeanCAP project will be guided by four committees. The **Advisory Committee** will provide all guidance to the project. The **Executive Committee** will provide operational leadership of the project. The **Industry Stakeholders Committee** will provide guidance from an industry perspective. The **International Stakeholders Committee** will provide a mechanism to introduce BeanCAP outputs throughout the world, while ensuring duplication is minimized and avenues of cooperation remain strong. Committee members and their expertise follow.

Advisory Committee members are: **Fred Bliss**, long time bean geneticists and former head of plant improvement at Seminis, Inc.; **Charlie Brummer**, forgage molecular geneticist, University of Georgia; **Chuck Hibberd**, director, Purdue Extension Service; and **David Sleper**, plant breeder and author of "Breeding Field Crops", University of Missouri.

Executive Committee members are: **Paul Gepts**, molecular geneticist, UC, Davis; **Julie Garden-Robinson**, human food and nutrition extension specialist, North Dakota State University; **Mike Grusak**, human nutrition and plant physiologist, USDA, Houston; **Jim Kelly**, dry bean breeder, Michigan State University; **Phil Miklas**, bean molecular geneticist, USDA/Prossser, WA; **Jim Myers**, snap bean breeder, Oregon State University; and **Juan Osorno**, genetics and breeder educator, and dry bean breeder, North Dakota State University.

Industry Stakeholders Committee members are**: Tom Grebb**, Central Bean owner, and member of US Dry Bean Council, Qunicy, WA; **Ken Kmecik,** snap bean breeder for Seminis, Inc.; and **John Ryapati**, dry bean breeder, Archer Daniels Midland.

International Stakeholders Committee members are: **Kirsten Bett**, bean breeder/geneticist, Univ. of Saskatchewan; **Horacio Guzman**, bean nutrition expert, INIFAP, Mexico; **Susan Nchimbi Msolla**, bean breeder, Tanzania; **Federico Sanchez**, bean molecular genetics, UNAM, Mexico; **oe Tohme**, bean biotechnology lead, CIAT, Colombia; and **Rajeev Varshney**, legume genomics, ICRISAT, India..

2. Management Roles and Responsibilities. The BeanCAP reporting lines are graphically presented in Fig. 1. **Co-PDs** will provide one-page quarterly reports to program leads highlighting progress and problems. **Program leads** will review the reports to ensure that the Co-PDs activities are being achieved according to the project timeline. During monthly conference calls with the Executive Committee, the **Project Director** (McClean) will summarizes the reports and discuss other project issues with an emphasize on project activities that are working in an integrative manner and how further integration can be enhanced in the future. The Executive Committee provides direction to the Project Director who in turn provides guidance to the program leads. The **Advisory Committee** reviews all reports and provides guidance to the Executive Committee in the form of semi-annual reports. The **Stakeholder Committees** inform the Executive Committee and the Advisory Committee on issues relating to the outputs of the three programs.

The program **Evaluator** (to be hired upon funding) will prepare an annual report for the Advisory and Executive committee based on the BeanCAP Formative Evaluation Plan (Appendix Table 1). To assess progress and impact of the project, four major questions, each with multiple sub-questions, were developed. These questions focus on the research, extension, and education components of the project, as well as the integration of the components. Data will be collected from a combination on questionnaires and interviews. Interviewees will include appropriate project personnel, Stakeholder and Advisory Committees, plant breeding education interns, and on-line users of the newly developed Community of Practice.

Management Plan

The Advisory Committee will meet with the Program Director and Executive Committee at the annual Plant and Animal Genome Conference to receive additional detailed input. The Advisory Committee will summarize overall progress based on project timeline (**Table 2**) and suggest improvements in an annual report that will be delivered to the Executive Committee and the USDA/Plant Genome, Genetics, and Breeding program officer. The Advisory Committee will ensure that the project is integrative by highlighting points of contact and information exchange between the three programs in the annual report.

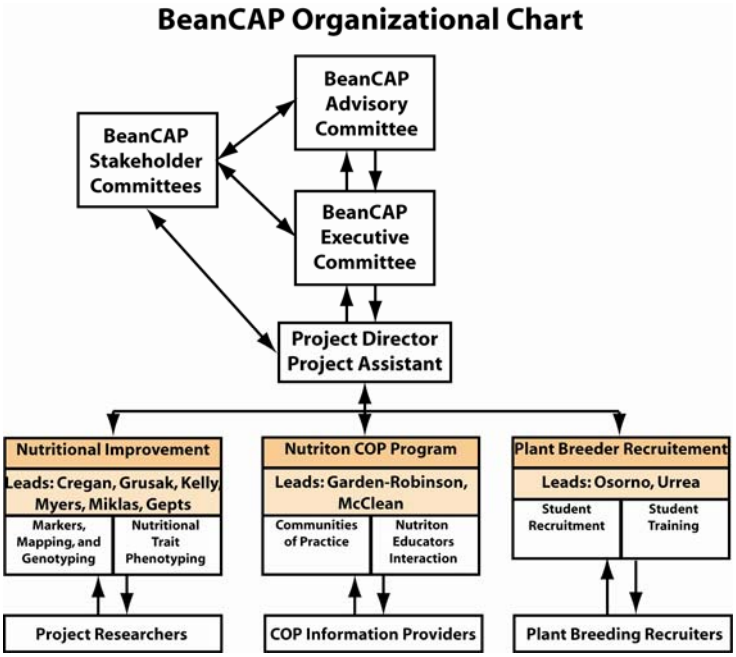


Figure 1. The BeanCAP Organizational Chart that depicts the reporting lines among the various participants and administrative groups.

Table 2. BeanCAP Project Timetable

	Date							
Program/Activity	**9/09**	**3/10**	**9/10**	**3/11**	**9/11**	**3/ 12**	**9/12**	**3/13**
Research Program								
SNP development	**X**	**X**						
Genotyping		**X**	**X**	**X**	**X**	**X**	**X**	**X**
Nutritional phenotyping	**X**	**X**	**X**	**X**	**X**			**X**
Marker associations/QTL discovery		**X**	**X**	**X**	**X**	**X**	**X**	**X**
Drought and heat stress study	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Database development	**X**							
Database delivery		**X**	**X**	**X**	**X**	**X**	**X**	**X**
Extension Program								
COP development	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Stakeholders communication	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Education Program								
Year-round research internships	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
High school recruitment	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Plant breeding literature development	**X**	**X**						
Hands-on program development	**X**	**X**	**X**	**X**				

3. Discretionary genotyping funding. Intrinsic to this project is a large scale, high-throughput genotyping service. The service will provide the full common bean improvement community (project as well as non-project PD/Co-PDs) an opportunity to collect data that would be of value to full research community. The materials to be screened will be determined annually, and those materials determined to be of broad utility for common bean improvement will be chosen. All individuals requesting the service must agree to submit the genotypic, and all associated phenotypic data, into the project database within six months of the service.

4. Modifying the Project. Reviews by the Executive Committee and the Advisory Committee and Evaluator reports will be used to evaluate progress of each Co-PD. That person's activities will be modified if it is determined by Executive Committee that the initial plan for that activity was overly aggressive. The Executive Committee reserves the right to halt a Co-PDs activities for neglect of responsibility. Those funds would be redirected to another Co-PD or another individual will be recruited to assume that project role.

It is understood by the BeanCAP management team that science and technology progresses often in unforeseen directions. If the Executive Committee and Advisory Committee determine that a new technology or scientific design would improve the efficiency of the project or significantly increase the value of the outputs, the project will be modified by reassigning Co-PD activities or redirecting funds to another Co-PD.

5. Managing Intellectual Property (IP). IP issues will be resolved among those researchers involved in the discovery using the IP policies in place at the appropriate universities.

6. Publications and Data Release. All Publications will be written in a timely manner and appear in established peer-reviewed journals. Authorship will be properly ascribed to individuals associated with a specific BeanCAP research activity at time of publication. The BeanCAP project supports the concept of open access to all research results. We will share our research results and education and extension materials in a timely manner to all stakeholders by placing them in public databases with direct access via the BeanCAP portal (http://www.css.msu.edu/bic/BeanCAP.cfm). The raw results will be incorporated into a database hosted at University of California, Davis. The mapping data will be displayed using cMAP, while sequence related data will be presented using gBrowse. The legume research database organizers have agreed upon these two applications for continuity between species data. Continuity will enable future data integration that will significantly leverage all data.

7. Sustainability. The BeanCAP will be sustained as the genetics and genomics tools are adopted by the bean research and extension community and stakeholders throughout the United States. That will result in continued long-term support from our federal, regional, local, commodity groups, and private funding sources. Related efforts such as those of the W 1150 regional research and extension committee will incorporate BeanCAP advances into its ongoing efforts and future strategic plans to genetically improve beans, and use the BeanCAP successes and wealth of information to interest and recruit highly promising students into our genetic improvement programs. The BIC will continue to provide the infrastructure as a portal for the BeanCAP outputs. The extensive data will continue to be maintained by the UC, Davis database. The National Plant Germplasm System will continue to function as depository for unique genotypes and/or germplasm identified by the program.

Common Bean Coordinated Agricultural Project
Project #: 2009-01929

BeanCAP Members' Response to Panel Review Questions

Answers provided by:

Phillip McClean, Karen Cichy, Perry Cregan, Julie Garden-Robinson, Paul Gepts, Mike Grusak, David Hyten, , Jim Kelly, Jim Myers, Juan Osorno, Mehmet Tulbek

Research

1) If the common bean genome sequence becomes available, how will that data be fully leveraged and integrated within BeanCAP?

Dr. Scott Jackson, Purdue University, was recently informed that the common bean sequencing project he submitted to the 2009 AFRI Plant Genome, Genetics, and Breeding program (A Sequence Map of the Common Bean Genome for Bean Improvement) was selected for funding. Dr. Phil McClean, the BeanCAP PD is also a co-PD on that project. Therefore, it is not a question of if, but rather when the sequence data will become available. The most relevant data from that project will initially be the assembled sequence. The timetable for that project shows assembly will begin at nine-months and continue through month 24. Our tentative plan is for the assembled sequence to be integrated into the project GBrowse database at UC Davis (initially funded by the Kirkhouse Trust of the UK to assist African bean breeders and now to be funded by the BeanCAP project) as a separate tract at increments to be determined (possibly months 12 and 24). The current GBrowse interface already includes the soybean genome sequence obtained from phytozome.net and the sequence-based markers developed and provided by Phil McClean (e.g., marker g1307: http://phaseolusgenes.bioinformatics.ucdavis.edu/search/markers/?ALL=G1307&format=html). Furthermore, these markers are linked to a CMap display.

All of the SNP-based loci will also be included as a separate GBrowse track in the BeanCAP genome database. This will depict the distribution of the SNPs relative to euchromatic and pericentromeric heterochromatic regions. These regions will be determined by the distribution of repeat sequences as currently implement in the soybean database at http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01/. This is important information because it will allow the project to determine if the markers are preferentially distributed in one or the other regions. If resources in the BeanCAP are available, new SNP markers will be developed that provide a more even distribution relative to the physical assembly.

Later on in the BeanCAP project, a track will be included that shows the physical location of the SNP loci significantly associated with the many traits we will be studying. That will be valuable information for any future bean researcher looking to discover the causative gene associated with a specific trait. For example, if the SNP is located in a euchromatic region, it is most likely that many more genes can be considered as candidates than if the trait maps to the (most likely) gene

poor heterochromatic region. The researcher can then decide which trait to pursue for further study. The most likely choice will be the trait that maps to the gene-rich region.

Furthermore, all markers developed in this project and others published during the life of this project and beyond) will be included in a marker database (current status at http://abckt.bioinformatics.ucdavis.edu).

The development of breeder-friendly markers will also be aided by the sequence. The genotype selected for sequencing is G19883, an Andean landrace. Conversely, nearly all of the ESTs available were derived from genotypes of Mesoamerican origin. The variation between the Andean genome sequence and Mesoamerican EST sequences will maximize the opportunity for polymorphism discovery. Therefore, by comparing the EST sequences with the genome assembly, additional polymorphism can be discovered that can subsequently be used for CAPs marker development.

Finally, all of the currently available BAC-end sequences (~88,000), and those generated in the sequencing project, will be added as a GBrowse track. These can also serve as another source of markers in the future, not only for BeanCAP researchers, but for any other common bean researcher.

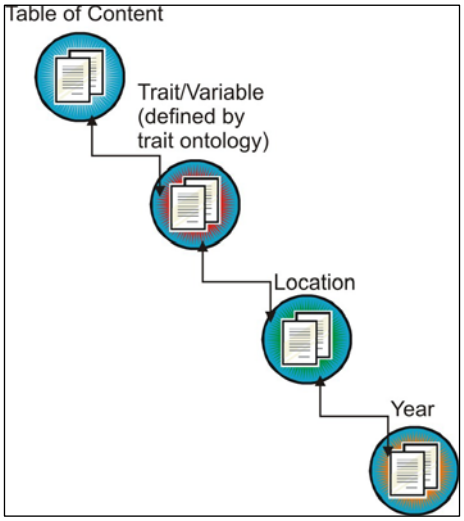
2) Describe in more detail how phenotypic data would be coordinated and quality checked. Specifically, how will multilocation phenotypic data be collected, standardized, curated and uploaded to the database?

There are two types of data: agronomic and nutritional. The agronomic data will be collected by researchers according to common protocols already in place and utilized for the Coordinated Dry Bean Nursery. Protocols used by Dr. Jim Myers will be established as the standard for the snap bean production traits. These protocols will be posted in the BeanCAP portal on the Bean Improvement Cooperative web site (http://www.css.msu.edu/bic/) for the entire community to refer to.

Dr. Michael Grusak (USDA-ARS Children's Nutrition Research Center; Houston, TX) will serve as the curator for all the phenotypic data and will review all incoming datasets prior to transfer to UC-Davis for inclusion in one of the databases. Multi-location data for specific nutritional components will utilize the same units of measure (e.g., µg/g dry weight or mg/g dry weight for mineral concentrations; mg/100 g fresh weight for snap bean ascorbate; percent protein for seeds; grams for 100 seed weight; etc.) or will be standardized to a fixed set of assigned values (e.g., values for pod shape or pod color). Grusak will also review each dataset to ensure that the range of values is consistent with existing published values for bean seed or snap bean pods. Timelines will be established between the Executive Committee and each 'phenotype' investigator for the completion of datasets each year. As curator, Grusak will monitor progress of these analyses and dataset compilations and as a member of the Executive Committee he will report the status of all analyses (progress and/or delays) to the Committee.

3) How will variation in phenotypes be incorporated into the database?

Within the phenotype database, the different data sets will be arranged in a searchable, hyperlinked hierarchical classification at the following levels: Table of Content – Trait/Variable – Location – Year (see Fig. below) . The different data sets will be downloadable from a web page.



4) Have the phenotypic traits been organized into trait-ontologies?

We had not considered trait ontology at the time of original submission. Since this question has been posed, we searched the various databases and found the trait ontologies developed by Gramene (http://www.gramene.org/) were the most advanced. Therefore, we have developed the following ontologies for all of the traits that for which we will be collecting data. Below is the current version of the schema.

Trait ontology
- Quality
  - Seed quality
    - Seed composition
      - Mineral content
        - Zinc, Iron, etc.
      - Protein content
        - Total protein content
      - Carbohydrate content
        - Total carbohydrate content
        - Soluble carbohydrate content
        - Insoluble carbohydrate content
      - Oil content
        - Total oil content

- Fiber content
  - Total fiber content
- Antioxidant content
  - Total phenolics
  - Anthocyanins
- Phytate content
  - Total phytate content
- Carotenoid content
  - Total cartenoid content
- Vitamin C
  - Vitamin C content

Nutritional availability
- Ca-Co-2 cell test
  - Mineral availability
    - Zinc, iron, etc.

Yield
- Pod yield
  - Number pods per plant
- Seed yield
  - Kilograms per acre
- Seed weight
  - 100-seed weight

Growth and development
- Flower development
  - Days to flowering
- Pod development
  - Pod shape
  - Pod color
  - Concentration of set

Seed development
- Days to maturity

Anatomy and morphology
- Shoot anatomy and morphology
  - Plant height
  - Shoot habit

Stress trait
- Abiotic stress
  - Lodging incidence

We wish to emphasize that is this only a **draft ontology**. Since one of the goals of the legume community is to have interoperability between species specific databases, we intend to work with soybase.org (the soybase database group at USDA/ARS, Iowa State) and the Legume Information System (Santa Fe) to develop a more comprehensive ontology that fosters interoperability. Other legume databases will be contacted to learn of their activities in this area, and it is our intention to work with those also.

It should be noted that the USDA/National Plant Germplasm System/GRIN system contains a large collection of *P. vulgaris* genotypes. GRIN has collected phenotypic data for 60 traits in 10 categories. It is our intention to work with Dr. Molly Welsh, *Phaseolus* curator, to ensure the ontology system is compatible with their data. This is important because we plan to SNP genotype the 442 members of the *P. vulgaris* USDA core collection.

5) No methods are described for obtaining "nutritional profiles". The panel recognized the participating investigators as accomplished in obtaining the data. However the database resource could be expandable beyond the scope of CAP PDs. Describe what steps will be taken to permit quality control should similar data be obtained for subsequent populations by other institutions.

Because of space limitations in the original grant proposal, we were unable to include all of the methods to be used. Below, we list various methods that will be employed and specifically address the issue of quality control and transferability of methods to other investigators. It is envisioned that a set of standard operating procedures (SOPs) will be developed and written for each of these assays; these will be listed on a public database for other bean researchers to access. When standard materials are required (for calibration purposes), efforts will also be made to make these available for other researchers. A set of 'check' cultivars will also be established (pending variation seen in our studies) that will be recommended for other researchers to include in their studies and nutrient profile analyses. The combination of SOPs, standard calibration materials, and 'check' cultivars should ensure good comparability of data added to the database for subsequent populations by other institutions.

A. Dry Bean Nutritional Quality Tests

1. ICP-OES analysis of tissue mineral concentrations (Grusak Lab; Houston, TX)

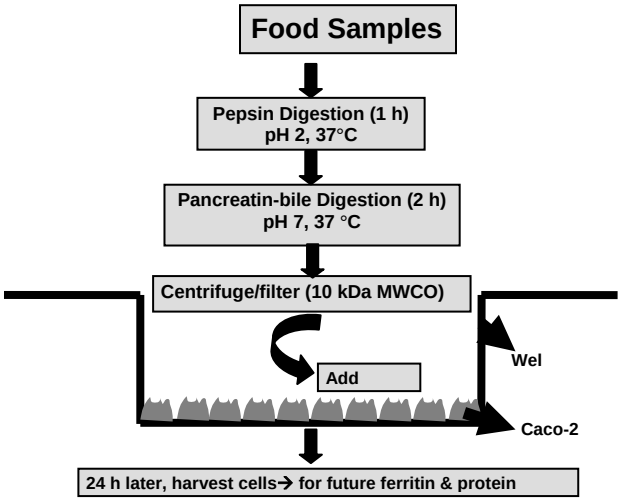
Mineral nutrient concentrations will be analyzed in seed or pod (snap bean) samples, using ICP-OES (inductively coupled plasma optical emission spectroscopy; CIROS ICP Model FCE12; Spectro, Kleve, Germany) methodology. Samples shipped to the USDA-ARS Children's Nutrition Research Center (Houston, TX) from various collaborators will be dried at 60ºC to a constant dry weight, and then ground and homogenized to a fine powder using a stainless steel grinder. Collaborators will be instructed on how to collect soil-free, clean samples, and will be provided with recommendations for shipping samples to the Houston lab (to minimize contamination). Once tissues are ground, a minimum of two subsamples of approximately 0.25g each will be weighed for acid digestion. Replicate subsamples are digested with concentrated nitric acid for 1.5 h at 125ºC, then with concentrated nitric acid: hydrogen peroxide (3:4) for 2 h, then are taken to dryness at 200ºC. Samples are run in blocks of 56 tubes, which will include a standard and blank in each run. All blanks will include nitric and hydrogen peroxide but no tissues (any measurable mineral levels in blanks will subsequently be subtracted from sample values); certified tomato leaf standards (National Bureau of Standards, Standard Reference Material 1568a) will be run to assess instrument calibration accuracy. Certified standards will be used each morning to calibrate the ICP-OES and a 0.3x elemental standard will be run each 15 samples to ensure instrument calibration stability over the course of each day. Elements to be analyzed will include: Ca, Mg, K, P, Na, Fe, Zn, Cu, Mn, Ni, Se, Co, Mo, and Cd. Final values

are calculated as µg or mg element per gram dry weight of seed or snap bean sample (e.g., µg Fe/g dry weight).

The ICP-OES instrument in the Grusak Lab is operational and available for use with the BeanCAP project. The Grusak Lab maintains a yearly service contract with the manufacturer, which includes semi-annual service calls to assess instrument operation and to provide preventive maintenance on appropriate components. A fully trained technician is currently running all samples on the ICP-OES and he will be continuing to provide service for this project. The Quality Control measures noted above are easily transferable to other labs. Standard reference material and elemental standards are readily available to other researchers.

2. In vitro digestion/Caco-2 cell method (Grusak/Etcheverry Lab; Houston, TX)

Caco-2 cells (passages 30-35) grown on collagen-treated 6-well plates will be used in the experiments. Studies will be conducted 16 days post-seeding, when the cells are fully differentiated. An in vitro simulation of the human digestion will be conducted with the food samples. First the samples will be brought to pH 2 with 1 N HCl. A volume of pepsin (0.8 grams pepsin dissolved in 40 ml 0.1 N HCl) will be added to all samples to simulate the digestion that takes place in the stomach. The samples will be transferred to a 37ºC incubator for 1 hr. Following the peptic digestion, the pH of the samples will be brought up to 5.5 with 1M $NaHCO_3$ and pancreatin/bile (0.2 g pancreatin and 1 g bile dissolved in 192 ml of 0.1 M $NaHCO_3$) will be added and the pH re-adjusted to 7 with 1 M $NaHCO_3$. The samples will be returned to the 37ºC incubator for 2 hrs. Following the intestinal digestion, the entire contents of the digests will be transferred to Millipore Amicon Ultra-15 centrifugal filter tubes with a 10,000 daltons MWCO (molecular weight cut off). Following centrifugation at 3750 RPM for 15 min, the filtrate will be collected and 1.5 mL placed on the cells. Following 24 hrs of exposure to the food digests, the cells will be harvested with 2 mL H_2O. Cell protein and ferritin will be determined: protein is determined using a Bio-Rad DC protein assay kit, based on the Lowry assay and Caco-2 cell ferritin formation is assessed with a one-stage sandwich immunoradiometric assay. Ferritin formation by Caco-2 cells is an indicator of iron uptake, and thus iron bioavailability from foods. There are 2 fully trained technicians in the lab who are involved with the preparation of samples and maintenance of the cell culture.



3. Protein, oil, fiber (Tulmek Lab, NDSU)

Protein, oil, fiber, and moisture content will be determined using a Perten Diode Array 7200 analyzer. Part of the evaluation process will be the development of a standard curve for common bean. The actual preparation of the materials for analysis will be performed using AACC Method 39-21. The procedure follows.

Protein, oil, or moisture content determination in soybeans may be ascertained by near-infrared reflectance or transmittance through intact seed either at discrete nonadjoining wavelengths or in a continuous wavelength region. Measurement is based upon a calibration to a suitable standard method (protein by Method **46-11A**, oil by Method **30-25**, and moisture by Method **44-15A**). Nearinfrared absorptions, attributable to combination or overtone vibrational frequencies of NH, CH, OH, and CO, are the primary bases of protein and oil content models.

Apparatus

1. Near-infrared instrument.
2. Sieves 10/64- × 3/4-in. (4- × 19-mm) oblong-hole and 8/64-in. (3-mm) roundhole, plus a pan.

Procedure

Calibration

1. Follow manufacturer's recommended daily check procedure before starting tests.
2. Select at least 80 samples of commodity to be tested, covering full range of protein, oil, and moisture that can be expected to be encountered in future testing. A uniform distribution of samples throughout entire range of protein, oil, and moisture is imperative. It is also important to try to develop a sample matrix in which oil, protein, and moisture levels are not correlated with each other.
3. Clean samples by hand sieving over 10/64- × 3/4-in. oblong-hole sieve on top and 8/64-in. round-hole sieve on bottom. Handpick soybeans remaining on each sieve so that only clean soybeans and splits remain, and combine material from top of both sieves. Discard fine material passing through 8/64-in. sieve and material removed by handpicking.
4. Analyze samples and determine calibration according to manufacturer's recommendations. Use reference results from replicate standard laboratory tests (protein, Method **46-11A**; oil, Method **30-25**; moisture, Method **44-15A**) on all samples, at least in duplicate.
5. Select 30 additional samples covering full range of constituents, and test to determine slope and bias.
6. If necessary, correct for slope and bias according to instruction manual.

Routine tests

1. Before tests each day, follow manufacturer's recommended daily check procedures.
2. Clean and analyze each sample in same way as calibration samples, using same type of grinder.

Fiber content will also be determined using a Perten near-infrared analyzer. Part of the process will the development of a standard curve for common bean. The actual preparation of the

materials for analysis will be performed using AACC Method 32-10. This is along procedure and will not be duplicated here. The user can refer to the procedure at the following URL:

http://www.aaccnet.org/ApprovedMethods/toc.htm

4. Phytic acid measurement (Cichy Lab, East Lansing, MI)

Phytic acid will be measured in seed samples using a modified Wade reagent colormetric assay as described in Gao et al. (2007, Crop Sci 47:1797). This method consists of a simple extraction, followed by a colorimetric assay based on the following: phytic acid binds to iron in the solution, reducing the amount of iron that reacts with the colorimetric agent. The first step will be to freeze dry seed samples and grind to a fine powder using stainless steel balls. Next, 2.4% hydrochloric acid will be added to samples and they will be shaken for 16 hr. Then samples will be transferred to tubes containing NaCl. Samples will be shaken at 350 rpm for 20 min and held at 4°C for 60 min. Samples will be centrifuged at 1000g for 20 min. Next the supernatant will be collected and diluted 25x.with ddH2O. The diluted samples will be mixed with modified Wade reagent (0.03% FeCl3·6H20 + 0.3% sulfosalicylic acid). Samples will be vortexed and centrifuged at 1000 g for 10 min. Absorbance will be read at 500 nm on a spectrophotometer. Sodium phytate (available from Sigma) will be used as a standard and seed phytic acid concentration and content will be calculated.

All equipment to conduct this method, including freeze drier, seed grinder, centrifuge, and spectrophotometer is available in the Cichy lab. A technician is available to run samples. To insure accuracy and reproducibility of this method a preliminary study will be conducted. In addition, a project currently underway, supported by the American Oil Chemists Society, lead by Victor Raboy, is evaluating the accuracy and reproducibility of this method for analysis of seed phytate in cereals grains and legumes and *Phaseolus* will be included in that study.

B. Snap Bean Nutritional Quality Tests

1. Snap collection and preservation

Fresh pods of snap bean accessions will be sampled at harvest maturity by hand picking all the pods on plants in 3 m row and sorting pods using a rotary grader. Harvest maturity will be determined when the length of the middle seed in a mature pod is equal to or greater than 1 cm (Silbernagel, 1979), which is equivalent to 50% 1-4 sieve of full sieve cultivars. The pods will be sorted into as many as six grades. For full sized snap beans, five sieve (9.7 – 10.9 mm diameter) sub samples will be prepared. For whole class beans three (7.5 – 8.5 mm) or four (8.5-9.7 mm) sieve size will be used as appropriate for the cultivar (some whole bean cultivars will never produce pods in the four sieve class). Two sieve pods (5.8 – 7.5 mm) will be used for extra fine cultivars. For flat- or oval-podded cultivars, we will rely mainly on seed length to determine maturity. These types will be sorted by hand.

Separate replicated samples will be preserved for mineral, flavonoid, carotenoid, vitamin C and fiber analysis. Samples for all but minerals and fiber will be collected on ice and frozen and stored in a -80C freezer until analysis. Flavonoid samples will be shipped to CSU on dry ice.

Samples for mineral analysis will be dried at 60C until constant in weight and shipped to the USDA-Houston collaborator. Samples for fiber analysis will be washed, run through a snipper to remove the beak and pedicle from the pod then blanched and frozen and stored at -20C until analysis.

2. Carotenoids content

The method presented here is a modification of existing methods, principally those of Khachick et a1. (1992) and Tonucci et a1. (1995). This is a reverse phase HPLC method using gradient chromatography and a photo-diode array detector which does not require roto-evaporation or sample partitioning. The protocol was published in Jones (2000, MS Thesis, Oregon State University), and will be calibrated using the method of Khachick et al. (1992).

Standards. β-apo-8'-carotenal (all-trans) will be obtained from Sigma-Aldrich Chemical company for use as the internal standard. all-*trans*-lycopene, all-*trans*-β-carotene, prolycopene, α-carotene, δ-carotene, ζ-carotene, neurosporene, lutein, and *E/Z*-phytoene standards will be purchased from CaroteNature (Lupsingen, Switzerland).

Extraction. Frozen bean pods, internal standard β-apo-8'-carotenal (0.0416 mg/g tomato FW) and sodium carbonate (10% of fresh weight) will be then homogenized in a Waring blender immersed in an ice bath. All the extractions will be completed under gold lamps (Sylvania F40/GO) to reduce light-degradation. Extraction solvent will be maintained at 0C. A 4 g subsample of the homogenate will be added to a Thomas Instruments (Philadelphia, PA) teflon pestle and glass extraction tube along with approximately 5 ml THF. This solution is macerated and filtered with a Buchner funnel and Whatman #1 filter paper. The solid material is repeatedly re-extracted with THF in the Thomas tube until the filtrate is colorless, resulting in a volume of approximately 25 mI. Final volume is measured with a pipette and recorded for quantitation. One ml of aliquot of the sample is then combined with one ml of eluent A and vortexed.

Preparation for injection. Samples will be filtered with 0.2 um 13 mm diameter nylon syringe filter (Alltech Inc. Deerfield, IL) and 50 ul will be were injected into the HPLC system.

HPLC. HPLC will be performed with a Beckman model 334 gradient liquid chromatograph (Beckman Instruments Inc., Berkeley CA), Waters 991 Photo-diode array detector (Waters Inc., Milford MA). Analysis was carried out using a reverse phase Spheri-5 RP18 column (220m x 4.6mm, 5 Perkin Elmer Brownlee, NOIWalk,CT) coupled to a ODS-5S guard column (3.0 cm x 4.6 mm, Bio-Rad, Richmond, CA). The Mobile phase consists of eluent A) 85% acetonitrile, 10% methanol, 2.5% hexane and 2.5% dichloromethane and eluent B) 45% acetonitrile, 10% methanol, 22.5% hexane and 22.5% dichloromethane. The chromatographic conditions will be a flow rate of 0.7 ml per min, 0-10 min 100% isocratic eluent A, 10-40 min 0-100% eluent B linear gradient, 40-45 min 100% eluent B. The system was returned to 1000/0 eluent A in a linear gradient over 5 minutes and allowed to re-equilibrate for at least 10 minutes.

Peak detection and integration. HPLC runs will be monitored from 250-650nm with the Waters PDA detector. Peak identification of phytofluene, neurosporene, zeta-carotene, gamma-carotene, lycopene, p-carotene and lutein will be based on matching spectra and retention times of

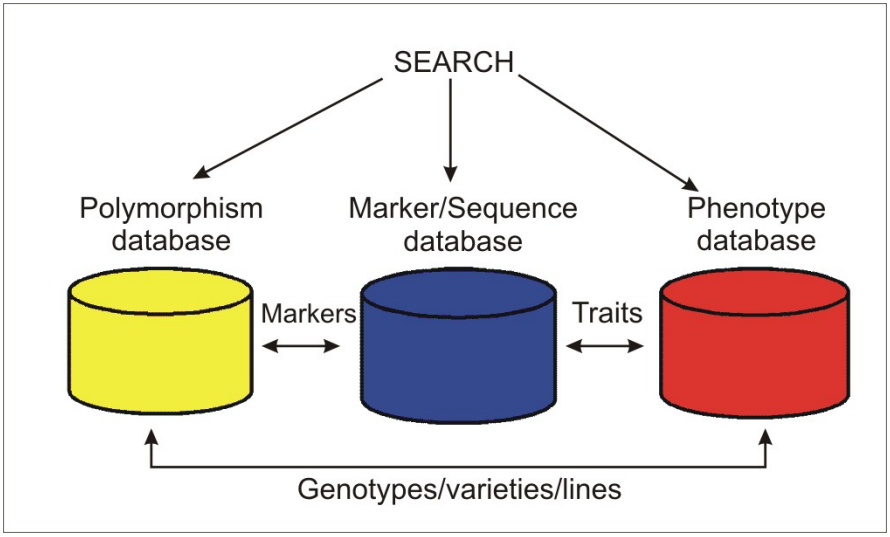
reference standards. Identification of other detected carotenoids was based upon published spectra, lambda max and proportional retention times (Khachik et al., 1992, J Agric Food Chem, 40:390-398). Quantitation will be performed on lycopene, B-carotene, and phytofluene based upon calibration curves generated from reference standards at the beginning and end of each session. All carotenoids will be quantified based on integrated peak area at their respective lambda max in a maxplot-derived channel using Waters Millenium software version 2.0 (Waters Inc. Milford, MA). Results will be reported in mg/100 g FW. The HPLC apparatus is maintained in the Department of Horticulture and the green bean breeding program has the technical expertise to conduct the analysis.

3. Vitamin C and fiber content

Established AOAC procedures for the analysis of vitamin C and fiber in vegetables will be used (Horowitz and Latimer 2005, Official Methods of Analysis of AOAC). Vitamin C analysis will be reported as mg/100 g FW and will be supervised by Bob Durst of the Linus Pauling Institute on the Oregon State University Campus. Fiber data will be presented as % total pod weight. Fiber analysis will be conducted by Brian Yorgey, Department of Food Science and Technology, Oregon State University.

6) How will phenotypic data and genotypic data be integrated when it is curated in separate databases? How will the availability of a scaffold sequence for Bean affect this integration?

We envision three databases. In addition to the marker/sequence database, there will be a polymorphism database and a phenotype database.



The three databases will be searchable via a text string search, which will bring up all instances of the string in the three databases. The databases will also be reciprocally hyperlinked as shown in the figure just above. Marker hyperlinks between the Polymorphism and Marker/Sequence databases will allow a quick assessment of the map location in one direction and the level of

polymorphism and abundance of polymorphism data in the other. Hyperlinks for traits between the Marker/Sequence and Phenotype databases will allow a retrieval of markers/QTLs associated with the trait in one direction and an assessment of the phenotypic data supporting an association or linkage between a gene or QTL and a phenotype. Finally, the hyperlinks between the Polymorphism and Phenotype database will allow a determination of which lines have been both genotyped and phenotyped.

7) What features of a marker and/or database will make theses outcomes from BeanCAP "breeder friendly"?

We are defining a "breeder-friendly marker" as one that is associated with an agronomic trait of interest and is inexpensive to use to follow a single trait. The SNP OPA sets are very valuable because they can serve as a genome-wide resource that can be used to quickly discover linked marker loci. But this is an expensive proposition. By leveraging the size of our project, we are able to purchase enough reagents to screen 96 genotypes for ~$7,000. Although that translates into $0.05 per data point ($7,000/96 genotypes/1536 loci), it is unlikely that a public breeding program can afford to spend $7,000 to discover a single marker. From a national project perspective, this cost justifies a unified effort to develop high-throughput SNP genotyping resources that can be applied to many traits across the various market classes.

But most breeders work on only a few of the market classes and are primarily concerned with a subset of the agronomic traits. The breeder would rather use funds dedicated to marker screening to follow those few loci in their program. These markers would be used to screen advanced generation breeding lines at various stages of the breeding program. Ideally, the markers can be applied at the local level by project personnel or a molecular genetics cooperator. CAP markers that distinguish different allelic state relevant to the specific program are an ideal tool. First, they only require the basic skill set that most technicians and plant breeding graduate students possess: the ability to amplify DNA and the ability to cut the DNA amplicon with a restriction enzyme. Cost is also a major consideration. The cost of a CAP data point in the McClean lab is $0.13 (excluding labor). Therefore, if a breeder is using the marker to follow the integration of a trait in 100 lines, that would only cost them $13.00 in reagents for screening. Given that most programs have dedicated technicians and graduate students, that is a cost that a pubic breeding program can afford.

The "breeder friendliness" of the markers will also be represented in the database itself by the addition of practical information for each marker that facilitates its application, such as primer sequences, annealing temperature (Tm), expected amplicon size, mapping information, synteny information, and links to the polymorphism database. We envision the user enters the database by navigating to a search page. Here they can provide key words either of their chose or from an index on that page. From there all entries relative to the keyword would be displayed. The user can then choose the appropriate link. From there, it will be possible to link either to the sequence, marker, or phenotype databases. If the user selects a phenotype, then they would also be able to migrate from that page to the sequence and marker databases.

8) How many genotypes each of pinto, black, navy, great north, kidney, and snap bean will be selected for the preliminary genotyping and how will these be selected?

Our initial SNP screening will use a broad spectrum of genotypes within the major market classes grown in US. This includes pinto, black, navy, great northern, kidney and snap beans. It is our intention to develop a broad collection of 196 genotypes representing these market classes. Because this diversity analysis will set the stage for subsequent experiments relevant to US public bean breeding programs, it is important the population represent the diversity found within the genotypic backgrounds most often used by public US plant breeders. Therefore we will rely upon the breeders to nominate those modern genotypes that are integral to their bean improvement programs. The number of genotypes that will be chosen from each market class is a somewhat tricky question. While it may be desirable to balance the number of genotypes between market classes, we don't want to survey many of the same genetic backgrounds within one market class at the expense of sampling the variation in another market class. (**Note**: This is a change relative to what we described in our initial proposal where we planned on equal representation of the different market classes.) From our perspective, it is better to first ensure that the genetic variability for each market class is represented. We take this view because this will form the basis of SNP selection, especially as it pertains to within race or market class variability. Furthermore, we need to be sure that enough genotypes within a market class are scored so we can subsequently select the two genotypes from each market class that will be characterized using the modified reduced representation Solexa "Paired-End Module" sequencing protocol described below.

We do not anticipate that the initial set of genotypes nominated by the breeders will equal 196. To fill out the remainder of the population, we will coordinate with our International Stakeholders committee. They will be asked to nominate genotypes that are widely grown in their countries. Where possible we will preferentially select those genotypes that have previously shown adaptation to US production systems and environments. By including these genotypes in the variability analysis, our US public breeders will be given a broader understanding of the variation available from these international sources relative to the variability available to them within the US.

We would like to present a related topic relative to the initial genotyping. To date, the degree of linkage disequilibrium (LD) decay relative to genetic distance can be considered a preliminary estimate. We presented that data in the appendix to the original proposal. That study was based on 16 genotypes and ~300 loci and is certainly not as comprehensive as the estimates we can obtain using the diversity data based on 196 genotypes and 1536 genotypes. Since all of the SNP loci will be genetically mapped using an extend Bat96 x Jalo EEP558 population, we will be able to next develop several estimates of LD decay at different levels of granularity. First, we can perform a species-wide estimate using the full data set. But what is more important from a breeding perspective is the decay estimates within the two gene pools (Mesoamerican vs. Andean) as well as within each race or market class. To obtain those estimates we can use appropriate subsets of data to generate those estimates. Finally, it will be important to study the decay on a linkage group by linkage group basis. This will inform our later decisions as we discover the genetic location of the loci associated with specific nutritional and agronomic traits.

9) Defend the choice of reduced representation DNA sequencing rather than normalized mRNA.

Firstly, it would be helpful to provide more detail about the combined Roche 454 GS FLX Titanium and Illumina GA sequencing system which is proposed for SNP discovery. The initial step is the digestion of genomic DNA with a combination of three blunt-end restriction endonucleases. Size selection is used to isolate the 450-550 bp fragment fraction from the resulting digest. This library is sequenced using the Roche 454. A similar 450-550 bp fraction is collected from an alternative genotype. This library is subdivided into three and each sub-library is digested with one restriction enzyme. The three digests are combined and the 160-180 bp fraction is isolated for paired-end Illumina GA sequencing (100 bp/end). Similar 160-180 bp libraries will be created and sequenced from a number of alternative genotypes. The Illumina GA reads are then aligned to the Roche 454 reads using MAQ software (http://maq.sourceforge.net) which is freely available. The MAQ software allows the discovery of SNPs between the reference 454 sequence and the Illumina reads that are mapped to the 454 reference. In addition SNPs can be discovered among the Illumina GA reads from the alternative genotypes.

As concerns the question of genomic DNA-derived reduced representation versus a normalized cDNA library for purposes of SNP discovery, there are two reasons for choosing the former. The first is that sequence diversity is likely to be relatively low within the Durango, Mesoamerican and Nueva Granada races and even lower within the market classes in each race: pinto and great northern in the Durango race, navy and black in the Mesoamerican and kidney and most snap beans in the Nueva Granada race. Because of the lower sequence diversity in genic sequence we propose not to limit the search for SNPs to only coding sequence but rather focus on non-coding DNA as well as the fraction of the genomic-derived reduced representation that is coding sequence. A second reason for the choice of the genomic DNA-derived reduced representation library is the error that occurs in reverse transcription used in the synthesis of cDNA.

A REVISED STATEGY FOR SNP DISCOVERY: The amount of sequence that can be accumulated in a single paired-end run on the Illumina GA is rapidly increasing. Illumina indicates that in the latter part of 2009 more than 50 billion bases of sequence can be obtained in a single run on the Illumina GA (http://www.illumina.com/downloads/AGBT_brochure_FINAL.pdf). Thus, in one run, 43.75 billion bases of data can be obtained from the seven non-control lanes on the Illumina Sequencing Flow Cell which is a 68X coverage of the common bean genome. Given the increased sequencing capacity, an alternative procedure for rapid SNP discovery is proposed. Rather than use reduced representation libraries we will first create a whole genome sheared and size selected library of 500 bp fragments of Jalo EEP558. This library will be sequenced with the Roche 454 GS FLX and should obtain approximately 0.63X genome coverage. This will be followed by the construction of sheared genomic DNA libraries size selected for the 160-180 bp fraction. The 160-180 bp libraries will be created from four genotypes of each market class (pinto, great northern, navy, black, red and white kidney, and snap). Barcoding will be used to identify the four genotypes from each market class. Each market class would then be run in one of the seven non-control lanes on the Illumina Sequencing Flow Cell. Two flow cells will be run

and should provide approximately 4.9X genome coverage of each of the 28 genotypes. These sequence data will be analyzed for SNP discovery using MAQ software with alignment to the Roche 454 "reference sequence" of Jalo EEP558. The analysis of the 28 genotypes in addition to the Jalo EEP558 reference will permit the assessment of the polymorphism of SNPs both within and between market classes.

10) How will SNPs or other polymorphisms be named and how will this naming system allow integration with the physical map and the eventual genome sequence?

The question of how markers, in this case SNPs, will be named is an important one particularly in light of the increasing ability to collect large quantities of DNA sequence data. This leads to the rapid accumulation of the SNPs discovered by two or more researchers each with a different name. To avoid this situation, dbSNP, the National Center for Biotechnology Information (NCBI) SNP database uses two different SNP nomenclatures. When a SNP is submitted to dbSNP it receives a submitted SNP ID number (ss#). The submitted SNP and its flanking sequence is aligned with the existing sequence for the species (if there is any) to determine the contig on which it resides. If the Submitted SNP is the first SNP discovered at a given position, it then receives a reference SNP number or RefSNP ID number (rs#). Subsequently submitted SNPs that go to the same position receive the same RefSNP ID number and are then part of that particular "reference SNP cluster". So, it is possible to have many millions of SNPs submitted with millions of ss# but these may only represent a few hundred thousand unique SNPs or RefSNPs. Thus, the NCBI dbSNP provides a universal system to eliminate duplicate SNPs and to preserve data.

Relative to the question of naming SNPs in common bean we provide the following information based upon our experience in soybean SNP discovery and naming: When the USDA Beltsville submitted our first set of SNPs to dbSNP back in 2003 they were assigned ss# and rs#'s. These can be seen at: http://www.ncbi.nlm.nih.gov/SNP/snp_viewBatch.cgi?sbid=5633. More recently the USDA Beltsville submitted well over 18,000 new soybean SNPs which received ss#'s and should soon also receive rs#'s. In the case of this recent submission a naming convention was used which included the "submitter handle", BARC, for Beltsville Agricultural Research Center, followed by a 6 digit number which designated the sequence tagged site (STS) in which the SNP was discovered. This was followed by a five digit number which designated the specific SNP. Because these SNPs were discovered via the Sanger re-sequencing of from six to nine soybean genotypes, the information on each STS was also submitted to the NCBI uniSTS database. The important point is that once a SNP is submitted to dbSNP and receives the ss#, and ultimately the rs#, the rs# can be used as the designation for the SNP. All information associated with the SNP is then permanently available to all. dbSNP can be readily searched for individual SNPs as well as for batches of SNPs via the ss# or the rs# at http://www.ncbi.nlm.nih.gov/projects/SNP/. Because the SNPs that will be discovered in common bean will be done using a combination of Roche 454 and Illumina sequence by synthesis technology rather than Sanger re-sequencing, the sequence variants will be submitted directly to dbSNP without the need to submit to uniSTS. We plan to name common bean SNPs "Pv-" followed by a six digit number. Once these are submitted to dbSNP and receive a RefSNP number, the rs# would be the preferred designation for each SNP.

11) What pitfalls are envisioned for the Association Mapping?

Association mapping (AM) is multistep process each with its own considerations. The **first step** is to identify a population that represents the diversity in which you are trying to discover an association between a trait of interest and a genome-wide collection of markers (or a set of candidate genes). **Secondly**, it is important to choose a set of markers that represent much of the diversity in the population. Early on, these were generally SSR markers, but with the advent of large national and international sequencing efforts, attention is now focused on SNP markers because of the high throughput capacity. Alternatively, researchers have chosen a specific set of candidate genes because of their putative function in the target species or observed role in the particular phenotype in another species. Once the marker set is chosen, the population is genotyped. **Concurrently** with genotyping, the population must be phenotyped. The actual data select depends upon the trait(s) of interest. **Thirdly**, the phenotypic and genotypic data is analyzed using a number of procedures to uncover significant associations. **Fourthly**, it is important to confirm the association using another AM population or a traditional bi-parental population. Ultimately, the goal is to determine the causative variant for the phenotypic difference. Therefore, the **final** step in the process is to focus intently on the confirmed region of association and try to determine what feature of a candidate gene is actually responsible for the phenotypic variation. This project will be performing the first four steps described above. We will discuss the technical issues associated with each.

Our first interest is to discover loci associated with nutritional quality across the cultivated portion of the common bean species (*P. vulgaris*). Given that both dry bean and snap beans are the major types of beans, we have developed AM populations for each. The dry bean population will contain representatives of the major market classes grown in the US (pinto, navy, black, great northern and kidney; see the answer to question #8 as to how these will be selected). The major concern with this type of population is that a specific trait may be associated with the population structure and not a particular gene. For humans, this is a particular concern when the allelic frequencies are not equal in the two phenotypic states (disease and non-disease). The classic case is where a HLA haplotype was originally discovered that associated with diabetes in Pima Indians. Further analysis demonstrated that the association was actually a product of population stratification based on the degree of European ancestry in the two phenotypic classes. We believe structure will be a minimal problem and can be handled statistically in our analyses for two reasons. First, we will be using our genotypic data along with multiple procedures to ensure that population structure is accounted for during our analysis. In particular, we will perform STRUCTURE analysis (which defines subpopulations that are in Hardy-Weinberg and linkage equilibrium) along with kinship analyses (SPAGeDi software). Both of these factors will be incorporated into a mixed linear model used to discover markers associated with the trait of interest. In addition, we plan to compare these results with one where structure is estimated using the standard principal component analytical procedure. The structure/kinship and principal component analyses will be compared using the Bayesian Information Content statistic to determine which approach best explains the phenotypic variation for each trait. Secondly, population structure is a concern in humans because two phenotypic classes are defined: cases (those with a disease) and controls (those without the disease). We anticipate, based on a number of studies of quality traits in common bean, that our data sets will be distributed in a continuous nature. This will allow us to use the procedures initially develop by the Buckler lab

to deal with quantitatively inherited traits in plants (Thornsberry et al., 2001, Nat Genet 28:286). These procedures have more recently been modified, and the mixed-linear model approach has now become the standard (Yu et al. 2006, Nat Genetic 38:203). More recently, Kang et al. (2008, Genetics 178:709) described a new algorithm to perform the mixed linear model analysis. This algorithm performed 40x faster than that implemented in SAS and TASSEL. That approach is offered as a WWW application on their lab's web server. We will test our data before committing to using it, given that we have been successful with TASSEL in the past.

We will use SNP developed by the Cregan lab for our genotyping. Currently one 1536 OPA set is under study and this project will add another OPA set of a similar size. These will provide a very large data set both for population structure analysis and marker/trait association testing. Given the robustness of the Illumina Golden Gate Assay and the experience of the Cregan lab to perform the analysis, we feel this is an ideal choice. The only downside will be the potential lack of variation for a large set of the markers. The many studies with various molecular marker types have shown that marker polymorphism is robust among all of the common beans. A recent study (Lee et al. unpublished) showed that 15% of the SNP markers developed by the McClean lab were polymorphic among two representative parents within the Durango race. If this value is representative, then we anticipate that ~450 SNP loci should be polymorphic within this major market class (pinto, great northern beans). This should be a sufficient number of markers (~ 40 per linkage group) for population structure determination and marker/trait association discovery.

Phenotypic data is the most essential component of any analysis. We have developed a consortia consisting of field researchers from the major bean producing regions of the US that are highly experienced with phenotypic common bean genotypes. To facilitate accurate data collection, standardized procedures for general agronomic trait data were established to ensure consistent data results from the multiple cooperators taking part in the Cooperative Dry Bean Nursery (CDBN). All of the field trials will be performed with researchers with many years of collecting the CDBN data. Similarly we decided to centralize the collection of the nutritional quality data for all trials. Field grown seed will distributed to all of the testing locations by the field researchers, and highly experienced quality testers will collect the data. This will ensure consistency of data results.

Using the genotypic and phenotypic data, we will perform statistical analyses to discover significant marker/trait associations. The critical issue is to set a *P* value that protects against Type I errors. Since we are performing multiple tests of association (the 1536 – 3072 markers) on the same phenotypic data, we must protect against random associations due to multiple testing. The most conservative method is the Bonferroni correction which sets the *P* value based on the number of independent tests. In our case, the *P* value would be $1/1536 = 6.5 \times 10^{-4}$ to $1/3072 = 3.3 \times 10^{-4}$. Recently, Storey and Tibshirani (PNAS, 2003, 100:9440) established an approach to establishing significance levels for genome-wide studies similar to what we are proposing. This is based on the concept of false discovery rate and provides a balance between Type 1 (false positives) and Type II (false negative) errors. With this procedure, each marker is given a *q* value (essential equivalent to a *p* value) except it expresses significance in terms of the false discovery rather than the false positive rate. It is our intention to apply the FDR approach as we select those marker loci that appear to be significantly associated with a given trait of interest.

The last issue to consider is confirmation of putative marker/trait associations. We are planning on developing bi-parental populations between the most extreme parents for a number or market classes and a number of traits. Clearly, we will not be able to cover all of the potential traits. Therefore, we will focus on the strong associations and develop these populations from crosses between parents with the most extreme phenotypes. Alternatively, individual investigators may choose to apply a candidate gene approach based on the location of the associated marker and genes that are defined within the interval. Those gene positions will be the result of the genome sequencing project of common bean.

In summary, we are aware of the potential pitfalls associated with AM, but we believe our design will provide a robust data set that will inform breeders and geneticists decisions regarding the best strategy for improving nutritional traits in common bean.

12) Given the proposed population sizes and likely sub-structuring, what is the proportion of genetic variation that a QTL must explain in order to be detected?

The abundance of diversity research has shown that the cultivated form of common bean is divided into the Mesoamerican and Andean gene pools, and that each of these gene pools is further divided into races. This was revealed recently (Kwak and Gepts, 2009, TAG 118:979) using SSR diversity data that was analyzed by the STRUCTURE software often used to reveal population structuring for AM experiments. Therefore, we fully agree with the premise of this question that sub-structuring will be observed with our AM population. One relevant question is the amount of variation that will be accounted for by the population structure. We simply will not know the answer to this question before we complete the experiment. For example, although the mean percentage of variation explained for 60 maize traits was ~9%, and the range of values was 0.1 - 35.0%.

At this point, we simply do not know if the variation for any of the nutritional traits will be affected by population structure. A study of Fe, Zn, P, and S seed content for both Mesoamerican and Andean genotypes revealed equivalent ranges within each of the two genepools (Welch et al, 2000, J Ag Food Chem 43:3576). This would suggest that at least for these nutritional traits, the element contents within the seed would not necessarily be affected by population structure. Of course, we will need to await the full results from our far more extensive survey to determine if this trend holds up for all of our traits within our AM population.

As to the question regarding the size of the QTL that can be detected, we again refer to the work of the Buckler lab (Zhu et al. 2008, The Plant Genome 1:5). Since the amount of variation with an AM population is inherently greater than that found within a biparental population, a large number of individuals are needed to detect a QTL of similar size. This is especially the case if population structure confounds the estimate. While accounting for structure controls for false positives, it still reduces the amount of variation accounted for by the marker. Therefore, we plan to perform our analyses both with the entire AM population and the subpopulations. Given that the members of the AM population would only represent three races (and potentially three subpopulations), we will attempt to balance, to some degree, the number of genotypes within the

three races. This would therefore lead to the similar (and larger) power of detection within subpopulation because the LD within each will be greater and structure effects would be minimized. But as with other questions here, we simply do not know the answer *a priori*. It will take our experimentation to provide the first experimental answers.

13) What is known about the heritability of the various nutritional traits that will be examined?

There is not much information on estimates of heritability for nutritional traits in common bean. Most recent studies have been more focused on estimating the QTL effects than in estimating heritabilities *per se* of these traits. In other crops such as rice, narrow-sense heritability for Fe grain content is 0.43. However, the large difference between estimates of narrow-sense and broad-sense heritability (88%) further confirmed the effect of both the additive and the non-additive type gene action controlling the high Fe concentration (Gregorio et al., 2009, Breeding for micronutrient enriched rice. *In* Bañuelos G.S. and Lin Z.Q (eds.) Development and uses of biofortified agricultural products. CRC Press, Boca Raton, FL. p. 171-180). One of the few reports of heritabilities in nutritional traits in bean is the one from Kelly and Bliss (1975, Crop Sci 15:753). They reported a narrow-sense heritability that ranged from 0.32 to 0.61 for total protein and from 0.52 to 0.87 for available methionine.

The two most studied micronutrients in common bean are Zn and Fe given that seed content of these elements may not be sufficient for human consumption if no other source is available (Blair et al., 2009a). In the case of Zn for example, one study reported narrow and broad-sense heritabilities of 0.84 and 0.82, respectively (Cichy et al., 2005). The high narrow-sense heritability estimate suggests that additive effects are a major component of the total genetic variance and therefore, genetic progress could be accomplished. In the same way, a single gene has been found to control soil Zn deficiency (Singh and Westernmann, 2002, Crop Sci. 42:1071), and results suggest that Zn-efficient genotypes possessed a greater concentration of seed-Zn than Zn-inefficient genotypes (Moraghan and Grafton, 1999, Soil Sci. Soc. Am. J. 63:918). These results also suggest high heritability for Zn content. In addition, several major QTLs have been reported for Zn and Fe (Blair et al., 2009, http://www.biokemi.org/biozoom/issues/525/articles/2397; Gelin et al., 2007, Crop Sci. 47:1361), which suggest that genetic improvement for these elements is achievable.

The BeanCAP would provide an excellentl opportunity to obtain estimates of heritabilities for each one of the nutrients evaluated given the amount of phenotypic data collected across different environments. The International Center for Tropical Agriculture (CIAT) has been collecting phenotypic and genotypic data for Zn and Fe across several years and locations. They are currently using this dataset to obtain reliable estimates of heritability (S. Beebe, personal communication). Preliminary results are showing high phenotypic correlation among environments, which may suggest high heritabilities for these two elements.

Heritability estimates for other nutrients in common bean could not be found, but an effort will be made within the BeanCAP to find previous studies that may address this in detail.

14) How will field locations and replicates be handled in the AM analysis?

It is our plan to develop a phenotypic scoring system that incorporates the location mean and standard deviation into each phenotypic data point. First, the dependent variable will be the mean across replications at a particular location. That mean will be subtracted from the individual observation and that value will be divided by the standard deviation of the location. This is the approach used in a recent study of genetic architecture traits in teosinte (Weber et al., 2008; Genetics 180:1221) and similar to a procedure used recently in other studies.

15) Will the market-class-specific markers provide enough density for effective association mapping?

As stated above, an analysis of the within market class variation for the Durango market class, found that 15% of the SNP based markers were polymorphic. For the set of 3096 markers, this translates into ~450 markers that would be useful within a race (=market class). We also plan to develop races specific marker sets. Those sets would consist of 768 markers each for the Durango (pinto, great northern), Mesoamerican (navy, black), and Nueva Granada (kidney) races. These markers will also be useful for snap beans since they are either of Mesoamerican or Nueva Granada origin. This density is greater than used with other recent AM experiments in plants [iron deficiency chlorosis in soybean (Wang et al. 2008, TAG 116:777)); fiber quality (Abdurakhmonov et al. 2008, Genomics 92:478); plant height (Brown et al. 2008, Genetics 180:269); flowering time (Ducrocq et al. 2008, Genetics 178:2433; Stracke et al. 2009, TAG 118:259); vitamin A content (Harjes et al. 2008, Science 319:330), starch content and composition (Wilson et al. 2004, Plant Cell 16:2719); and carotenoid content (Palaisa et al. 2003, Plant Cell 15:1795)].

16) What is the accuracy of SNP calling and the call rate for Bean genotypes on the Illumina OPA?

Based upon our experience in soybean which is an ancient tetraploid with a highly complex genome, we are quite confident that the Illumina GoldenGate assay system will function well in the less complex common bean genome. This is based upon the analysis of three OPAs in soybean, one with 384 SNPs and two with 1536 SNPs. Each OPA was used to genotype three recombinant inbred line mapping populations as well as a diverse set of 96 soybean landraces from the USDA Soybean Germplasm Collection as well as a highly diverse set of elite cultivars representing the diversity of North American soybean germplasm. In the case of the first 384 OPA, the rate of successful assays was 89% (342 of the 384 loci). SNP calls were obtained for more than 99% of the genotypes analyzed. These results were described in Hyten et al. 2008. Theor. Appl. Genet. 116:945-952. Essentially, identical results were obtained for the two sets of 1536 GoldenGate soybean OPAs. In both cases, successful assays were obtained for 90% of the SNPs and based upon the analysis of the same sets of genotypes described above, unambiguous genotype calls were obtained for 99% of the genotypes. One virtue of the GoldenGate assay is its sensitivity to copy number. Thus, in the case of soybean it became clear that certain of the SNPs in the OPAs were in duplicated regions. However, because only one of the two duplicated regions contained the SNP, the assay could distinguish between those genotypes that were homozygous for the SNP i.e., all four copies of the locus contained allele 1 versus when the SNP was heterozygous at the position at which the SNP was segregating i.e., three copies of allele 1

vs. one copy of allele two; and the case when the SNP was homozygous for allele 2 at the position at which the SNP was segregating i.e., two copies of allele 1 vs. two copies of allele 2. Because of the lower complexity of the genome we anticipate at least a 90% successful call rate with the common bean SNPs assayed using the Illumina GoldenGate Assay.

17) How will missing genotypic data be handled in the AM analysis?

It is our intention to handle missing genotypic and phenotypic data separately. We are certain that we will have missing loci scores for genotypes for the AM population. In the past, we have simply treated these as missing data. But as datasets become larger, and the amount of marker data within a certain region becomes denser, it is possible to impute the allelic state at a specific locus. Imputation uses the linkage disequilibrium structure around the region with the missing data to determine the most likely allelic state at the missing locus. Critical to this analysis is the designation of a reference population. At this point, it is difficult to determine the exact manner in which reference data set we will used for the imputation. For example, we will need to test if we wish to use the SNP data for the entire population or if we would rather simply use the data from within a race (or market class). We intend to use the fastPHASE software (http://stephenslab.uchicago.edu/software.html) developed by the Stephen's lab at the University of Chicago for these determinations. Any imputed SNP loci that are found to be associated with one of the traits, will be scored to ensure the SNP call is correct. (See Jannink et al. 2009, Plant Genome 2:11 for an imputation study using barley along with the fastPHASE software.)

So far, the McClean lab has dropped missing phenotypic data from their AM analysis of soybean iron deficiency chlorosis (Wang et al. TAG 116:777). For example, if data for a replication was missing, the phenotypic mean for the genotype was simply based on the mean of the remainder of the replications at a particular location. With these experiments, each location contained four replications. Therefore, good phenotypic mean estimates could be based on the remaining three replications. But for the AM experiments described here, the phenotypic mean will only be based on two replications. Therefore, it will be important to impute the missing phenotypic data. To this point, we have not implemented this method. Yet we are aware of a number of recent papers (Niu et al, 2005, Genetics 169:1021; Guo et al., 2008 BMC Genetics 9:82) in which missing phenotypic data is imputed. We plan on using those as guidance as we consider the missing phenotypic data problem.

Education and Extension

18) How will you increase the involvement of non-rural, community college, and minority students?

The education component is focused at four land grant universities: North Dakota State University (NDSU), Michigan State University (MSU), University of Nebraska-Lincoln (UNL), and Colorado State University (CSU). Two of these campuses (MSU and CSU) are located in largely urban states with population centers in excess of 500,000. The Michigan State University College of Agriculture and Natural Resources Office of Diversity and Pluralism and the MSU Office of Academic and Student Affairs have established multiple, comprehensive initiatives to

attract and support students and staff pursuing careers in food, agriculture and natural resources. The program is directed toward underserved student populations in urban areas in the North Central region. The MSU Professional Development Diversity Endowment is available to provide support for professional development programs for undergraduate students and graduate students for summer internships. In addition MSU also has a Pre-college Leadership Program for Native Americas and a Multicultural Apprentice Program that brings students to campus during the summer to work with a researcher. Fort Collins, CO, with more than 100,000 people, but because of its proximity with the capital (Denver), is sometimes considered as a suburb of Denver. Therefore, we assume that non-rural students will be exposed to the plant breeding activities at least in these two institutions in MI and CO.

Members of the project team will work with existing outreach networks to reach diverse populations within and beyond each university. At North Dakota State University, for example, we will work with the university's Tribal College Liaison to reach students at Tribal colleges. Garden-Robinson is a co-PI on an existing grant with the goal of reaching underserved and minority students with food systems/safety education, so she will help coordinate efforts between the grant projects to avoid duplication in efforts. In addition, Juan Osorno is actively involved with a summer internship program at NDSU for students from the University of Puerto Rico-Mayaguez (UPRM). Some of these students will be exposed to several aspects of the BeanCAP at the same time they are doing their summer internships at different breeding projects within the department. A new chapter of MANNRS (Minorities in Agriculture, Natural Resources, and Related Sciences) has been recently started at NDSU. Juan Osorno was involved in the startup of this chapter during 2008. Therefore, efforts will be made to recruit students already involved in this group.

As mentioned in the project narrative, Carlos Urrea at UNL will be working directly with Scottsbluff Community College students since the main campus at Lincoln is 400 miles away from the Scottsbluff station, where the dry bean breeding program and the main production region is located.

Members of the project team will work with existing outreach networks to reach diverse populations within and beyond each university. At North Dakota State University, for example, we will work with the NDSU Tribal College Liaison to reach students at Tribal colleges. Garden- Robinson is a co-PI on an existing grant with the goal of reaching underserved and minority students with food systems/safety education, so she will help coordinate efforts between the grant projects to avoid duplication in efforts between the grant projects.

19) What other learning materials will be integrated with the animations?

In addition to the animations, the VCell animation project provides a suite of stills that are drawn from each animation. These stills are selected because they highlight a key point that is emphasized in the animation. These are provided in what we call the "First Look" and "Advanced Look" levels. The "First Look" is general overview that is intended for the learner who is first encountering the topic. The "Advanced Look" focuses on details that are normally covered in a higher level course. The approach will be modified for the COP. Instead, we will provide stills aimed at the "General Public" and "Educational Community". The "General

Public" stills will be provided at the 6th to 9th grade levels and will be presented in a manner that engages the user by focusing on the features of the animation that are related to biology and human health. The "Educational Community" stills will focus more on the details of the process and will include more scientific terms that support the learning from a curriculum or standards-based perspective at the 9th to 12th grade levels. For example, our first animation will highlight root biology and the role soil chemistry plays in nutrient uptake. The "General Public" stills will provide images of the various root components and how they developed over the life cycle of the plant. Here the purpose is to convey the concept that multiple interacting components are necessary to get minerals into a plant. In contrast, the "Educational Community" stills will provide more details of root biology and will emphasize the concept of state change necessary to convert a nutrient from the unavailable to available state.

We also plan on developing quizzes that the general public, students or teachers can complete on line or download in the Adobe .pdf format. These again will focus on the main concepts supported by the animations. The answers to the quiz will also be available on-line and short explanations of the correct answer will be provide in case a learner get an answer wrong.

All of the learning materials will be organized in a manner that encourages a systematic learning process. After the learner logs in to the COP, the student will first be presented a **list of learning objectives**. The list will highlight the key points that the student will learn in the module. From here, the student can choose among multiple learning components. The component list will be ordered in what we believe will be the best learning order, but the student can choose whatever path they feel best suits their learning style. One component will be **stand-alone images** of the various steps in the biological process. These will be delivered as a slideshow. The next component will be the **animation** itself. Following this, the student will have access to a **subtitled animation** with the narration below in a subtitle box. The learner can also read a **text version of the animation narration**. This will let the learner see the main points highlighted in the animation.

The last component of each application will be a **multiple choice quiz** designed to allow the student to reflect on what they have learned. A key feature of the test will be its linkage to the various learning components. For example, when the student answers a question correctly, they will be presented a screen that provides a brief explanation of the subtopic addressed by the question. The screen will also have links to sections of the various learning components that address the specific subtopic. In a similar vein, if the student answers incorrectly, links will send the student to relevant sections in the learning materials that will help offer a correct answer.

In addition to providing materials to accompany the animations on the Co-P, consumer-friendly nutrition education materials about beans, derived from credible sources such as the "Meat & Beans" group information from www.mypyramid.gov, will be made available on the CoP. "Nutrition Facts" labels of bean varieties will be created using Nutriform software to help consumers understand the contribution of beans in the diet in a familiar format to them. To encourage the use of beans in the diet, health benefits based on published research, information about incorporating beans in the diet and recipes with nutrition analysis also will be featured on the CoP.

20) How will the CoP interact and publish animations and learning materials to eXtension?

The CoP personnel will develop learning modules. These modules will be animation centric, and additionally, contain still images, written scripts of the animation narrative, and self-guided quizzes. Once they are completed and tested locally, they will be uploaded into eXtension. These animations will also contain subtitles to ensure broad accessibility. Finally, the animations will be developed in such a manner that download time is minimized. Once these objectives are met, they will be published into eXtension.

21) Who is going to fill out the application for the CoP to eXtension?

To formally become an eXtension CoP is a two step process. First a pre-application is completed and reviewed. If it passes this step, then a full application will be submitted. The actual name on the application will be Dr. Julie Garden-Robinson. She will be assisted by Ms. Christina Johnson, the animation and WWW development lead for the VCell Animation project, who will provide much of the technical information necessary to complete the CoP eXtension application. We have chosen Ms. Johnson for this task because she is highly skilled technically, not only with animation development, but WWW design and WWW interactivity. She also has a full understanding of content creation tools and indexing procedure needed for content discover by search engines such as Google and bing.com. In addition, other members of the BeanCAP project will provide initial assistance in the developing the FAQs, Ask An Expert materials, and the educational principles that will underline module development.

22) Will any funds be directed to developing the eXtension CoP?

Yes, project funds will be spent to develop the "**Nutritional Genetics and Genomics: Healthy Foods from the Field to the Table"** eXtension Community of Practice. Specifically, funds are allocated for the lead animator (Ms. Christina Johnson), an animation development assistant, and hardware and software. The direct costs associated with these activities are $303,384 ($435,356 including indirect costs). Funding for the personnel funding is ongoing for each year of the project. The hardware and software funds are only for the first year. These funds are included in the NDSU budget.

23) Who will be responsible for training and/or uploading materials into the eXtension wiki?

Ms. Christina Johnson will be responsible for publishing all materials in eXtension. These will only be published after they have under internal review by members of the CoP.

24) Please describe the peer review system to be used by BeanCAP for materials on eXtension, and how this system could be expanded as the CoP expands.

The BeanCAP has a broad collection of educator and researchers who are experts in nutrition, genetics, genomics, and plant breeding. In addition, the fields of plant physiology and human health are expert areas for some of the project personnel. Our initial review will be performed within the BeanCAP project. Additionally, Dr. Julie Garden-Robinson is well connected to the

human health field by her NDSU Extension appointment. Dr. Garden-Robinson will recruit outside personnel from the universities in which the BeanCAP project is representing. Our current plan is to recruit two individuals to review our storyboards. These will provide the project with a broader learning perspective.

This grant team represents an interdisciplinary, multistate network of experts who can recommend peers to become part of the peer review process and Ask an Expert network. In keeping with the principles of eXtension, the BeanCAP CoP will provide credible expertise (Ask an Expert), reliable answers based on sound research (FAQs) and trustworthy, field-tested data. In the planning phase for the development of this CoP, we will learn from the experiences of others who have developed CoPs and coordinated peer review systems. We will work with two mentors at North Dakota State University, Debb Pankow, Ph.D., who led the development of the Financial Security for All CoP, and Charles Stoltenow, DVM, who led the development of the Agrosecurity and Flooding CoP. Through the development of this proposal, we have begun to learn from their experiences as we plan for a network of experts to approach as the process begins and as the program expands. As an initial step, the team will put together a plan to identify and recruit experts for the CoP.

General Questions:

25) What would the specific CAP deliverables be at the ends of years 1, 2, 3, and 4?

The BeanCAP deliverables at the end of each year of the project presented here based on the five project objectives.

Objective 1: Develop high throughput, market-class-specific markers for the predominant common bean market classes produced in the US, convert those markers into breeder-friendly markers, and genotype breeder-defined populations with these markers.

> **Year 1: SNP diversity data** for a collection of 196 common bean genotypes. **Solexa sequence data** for two genotype each for the pinto, navy, black, great northern, kidney, and snap bean market classes. **SNPs** defined for each market class or each race (depending upon the results of the diversity screening and Solexa sequencing). **125 CAP loci** developed for each common bean chromosome in a race (Durango, Mesoamerican, Nueva Granada) specific manner.
>
> **Year 2: SNP data** for 2752 genotypes nominated by breeders and geneticist. **125 CAP loci** developed for each common bean chromosome in a race (Durango, Mesoamerican, Nueva Granada) specific manner.
>
> **Year 3: SNP data** for 2752 genotypes nominated by breeders and geneticist. **125 CAP loci** developed for each common bean chromosome in a race (Durango, Mesoamerican, Nueva Granada) specific manner.

Year 4: **SNP data** for 2752 genotypes nominated by breeders and geneticist. **125 CAP loci** developed for each common bean chromosome in a race (Durango, Mesoamerican, Nueva Granada) specific manner.

Objective 2: Discover genetic loci associated with nutritional traits that define "healthy beans" by combining genotype and nutritional profile data of association mapping and bi-parental populations.

Year 1: **Nutritional and agronomic performance data** for 400 dry bean and 200 snap bean grown under controlled greenhouse conditions.

Year 2: **Nutritional and agronomic performance data** for 400 dry bean (grown at four locations) and 200 snap bean (grown at one location). **Nutritional and agronomic performance data** for 100 dry bean grown at four locations under water stress conditions.

Year 3: **Marker loci associated with nutritional and agronomic traits** defined on a location-by-location and over all locations using the association mapping population..

Year 4: **QTL associated with nutritional and agronomic traits** discovered using bi-parental populations and standard QTL detection procedures.

Objective 3: Integrate common bean phenotypic, genotypic, and molecular marker data with other emerging legume genomic resources into breeder-friendly bioinformatic tools.

Year 1: Establish BeanCAP WWW site; link it through the Bean Improvement WWW site. **Enrich the Phaseolus Genes database** by incorporating all of the **historical mapping and QTL data** generated by the bean research community prior to this project, and where possible **link those markers** to available common sequence data

Year 2: Enrich the Phaseolus Genes database by incorporating sequence information from the common bean sequencing project. **Establish relationships with other legume database** to determine feasibility of **interoperability among the databases**. **Establish principles** that enable breeders to select appropriate markers for genetic and breeding purposes.

Year 3: Enrich the Phaseolus Genes database by incorporating sequence information from the common bean sequencing project. **Candidate gene discovery** based on AM mapping data and colinearity between common bean and soybean. **Enrich the Phaseolus Genes database** by incorporating all BeanCAP project phenotypic and marker information into the database. **Implement practices and protocols** that enable breeders to select appropriate

markers for genetic and breeding purposes. **FAQs and How-To** that describe the utility and practices of the database **developed.**

Year 4: Enrich the Phaseolus Genes database by incorporating sequence information from the common bean sequencing project. Additional **candidate gene discovery** based on AM mapping data and colinearity between common bean and soybean. **FAQs and How-To** that describe the utility and practices of the database **tested and improved.**

Objective 4: Launch the "**Nutritional Genetics and Genomics: Healthy Foods from the Field to the Table"** eXtension Community of Practice (COP). The COP will utilize high-quality animations and other multimedia to highlight the biology and technology associated with the genomic-based improvement of nutritional traits.

Year 1: COP will be **launched.** An extensive list of **Frequently Asked Questions (FAQ)** will be developed. **Animations** focused on the themes of 1) root biology and the role soil chemistry plays on nutrient uptake; and 2) the flow of soil minerals from the root to various parts of plants. **Supporting still images and quizzes** will be developed that support the learning points that are the focus of the animations.

Year 2: Frequently Asked Questions (FAQ) will be modified and extended. **Animations** that focus on a few of minerals and show how they move from our digestion system to the various organs in our body, and what health problems arise from mineral deficiencies in our diet. **Supporting still images and quizzes** will be developed that support the learning points that are the focus of the animations.

Year 3: Frequently Asked Questions (FAQ) will be modified and extended. **Animations** that provide a basic lesson in genetics by emphasizing the fact that many genes are involved in the uptake and distribution of the minerals throughout the plant, and that the genome of the plant is required to produce all of the proteins necessary for functional uptake and distribution of minerals throughout the plant. **Supporting still images and quizzes** will be developed that support the learning points that are the focus of the animations.

Year 4: Frequently Asked Questions (FAQ) will be modified and extended. **Animations** that emphasize how plant breeding utilizes the genetic information for trait improvement (mineral uptake) and the role that genomics in developing the tools that assist the breeder in trait selection. **Supporting still images and quizzes** will be developed that support the learning points that are the focus of the animations.

Objective 5. Initiate a modern plant breeding training program that focuses on early career recruitment and provides practical training that illustrates how the integration of genomic and phenotypic data can be used to improve nutritional traits in plants.

Year 1: Promotional materials such as brochures, flyers, and posters describing plant breeding will be developed for the high school audience. **Collating and delivery** of advanced learning materials from other CAP projects on the BeanCAP WWW site. **Curriculum** for summer and year-round internships will be developed. **Summer and year-round internships** will be offered by all participant universities each year. **Intern program database established** to track students from their college experience to graduate school to a plant breeding career. **Contacts** established with local high schools. **High school visits** completed with a target of 300 students (75 students at each of 4 locations). **Visits to breeding programs** at each of the four locations by local high school students.

Year 2: Summer and year-round internships will be offered by all participant universities each year. **High school visits** completed with a target of 300 students (75 students at each of 4 locations). **Visits to breeding programs** at each of the four locations by local high school students.

Year 3: Summer and year-round internships will be offered by all participant universities each year. **High school visits** completed with a target of 300 students (75 students at each of 4 locations). **Visits to breeding programs** at each of the four locations by local high school students.

Year 4: Summer and year-round internships will be offered by all participant universities each year. **High school visits** completed with a target of 300 students (75 students at each of 4 locations). **Visits to breeding programs** at each of the four locations by local high school students.

26) What innovative opportunities might underrepresented groups have, including minority serving institutions, to fully participate in the CAP? How would that be accomplished?

The BeanCAP is open to participation by any common bean researcher. Currently, the only bean researcher in a minority serving school is Dr. Venu Kalavacharla at Delaware State University. His research team of students is focusing on the molecular genetics of common bean disease resistance. His students are also a broad spectrum of underrepresented groups. Dr. Kalavacharla will be extended an invitation to participate in the project.

Our animation COP project will also be extended to minority serving schools. We intend to compile a list of faculty from both traditional and minority serving universities with an interest in molecular genetics and/or nutrition and inform them about the COP. We will urge them to join the COP and to use the learning tools in any of their learning environments.

As noted in the response to question 18, through our partnerships with several universities and their outreach to diverse audiences, we will serve a broad-based audience. The primary mechanism will be the use of the BeanCAP eXtension CoP. Partnering universities will provide links to it from their Web sites. The educational materials will be readily available on the CoP.

27) Has any part of this proposal been submitted to another agency? If so, what is the status of that submission? What other funding resources are available to support the CAP goals and objectives? What is the commitment from those sources?

The entire proposal has not been submitted to any other funding agency. However, it is correct to infer that the educational component of the BeanCAP has some similar activities with another proposasl recently submitted by NDSU (Osorno, McClean, Kelly, Brick) and MSU (Kelly) to the Plant Breeding and Education Program (AFRI program 10.310). That program requires an education component were internships and direct experimental activities are highly encouraged. Therefore, the target group is the same: high school students and undergraduate students. However, the research activities are different since these students will be mainly involved with aspects of research on drought tolerance, abiotic stress, and sustainability. In addition, MSU has a similar AFRI proposal largely targeted at graduate training and internships in breeding for abiotic stresses in the U.S and overseas. USDA-TARS in Mayaguez, Puerto Rico (Dr. Timothy Porch) is also involved in activities in both the research and education components of both projects. We believe this is an excellent opportunity to strengthen both programs since we can increase the number of students involved in all the recruiting activities. As a result, this complementary program would increase the outreach to students and also the probabilities of getting more students interested in pursuing careers in plant breeding.

28) How might the scope and scale of the CAP change with a 10% reduction in total budget? 20% reduction? Specifically, how could the budget be streamlined? How might the scope and scale of the CAP change with a 10% increase in total budget? 20% increase?

The BeanCAP is a tightly integrated project built around the theme of nutritional genetics and genomics of common bean. In addition, the project is bound by the program requirements that no more than 2/3 of the budget be allocated to one of the three target areas. Given those constraints, and after careful study of our budget, the easiest way to handle a **10% cut** would be to **eliminate the genotyping program** for a savings $485,886 [$441,714 direct cost + $44,171 (=10% USDA indirect cost rate)]. To reach a **20%** project cut, we would still need to eliminate the genotyping program. In addition **eliminating the genotyping program**, we would **cut the plant breeding education program in half** for a reduction of $235,897 [$184,000 direct costs + $51,897 (standard university rate of 22% of total project costs)]. In addition, we would **drop the animation assistant and reduce the salary of the lead animation developer** by 10%. This would lead to a 1/3 reduction in animation and COP activities. The total savings for this reduction would be $78,217 [$61,009 direct costs + $17,208 indirect costs (standard university IDC rate of 22% of total project costs)]. These changes would severely impact the scope and impact of the project.

A **10%** ($400,000) or **20%** ($800,000) **increase** in the budget would certainly be welcome for the project. It is our opinion that these funds would be best expended on **additional nutritional**

screening. For example, this would enable the Colorado State University research team to look for specific components of the bean that reduce cancer risks. (See Crop Sci, 2009, 49:179 for preliminary research showing the positive benefits of beans on the reduction of mammary cancer.) Once that study is completed, it would possible to go back to the original association mapping population, score it for the factor, and use the already available genotyping data to discover loci associated with elevated levels of that particular molecule. This is just one example. The executive committee would coordinate with the steering committee on the selection of the best set of experiments to pursue.

29) Given that an integrated CAP project must budget sufficient resources to carry out the proposed set of research, extension and education activities, with no more than two-thirds of a project's budget being allocated to a single knowledge area, please breakdown the budget and clarify what % goes for research, teaching and extension?

In some regards, it is challenging to calculate these percentages in a manner that truly reflects the operational split of the funds. For example, in what category should the indirect costs be charged? At some universities, these go to activities that would not be truly BeanCAP project oriented. We also feel that in addition to the research, extension, and education components of the research, funds are allocated to project management. In this category, we would include the 1) PD Assistant, 2) office supplies, 3) travel of executive and advisory committee for management of the project, 4) evaluator, and 4) the subaward indirect charges that would be collected by NDSU.

Therefore, we have calculated the percentage of each category based on the direct costs for the research, extension, and education components **and** based on total project costs (including indirect costs). The table below outlines those percentages.

	Direct costs only		**Total project costs**	
Category	**Budget**	**Percentage**	**Budget**	**Percentage**
Research	$1,971,788	65.5%	$1,971,788	49.3%
Extension	$ 481,963	16.0%	$ 481,963	12.0%
Education	$ 356,614	11.9%	$ 356,614	8.9%
Management	$ 288,382	6.6%	$ 288,382	5.0%
Indirects	NA	NA	$ 952,188	23.8%

As the table shows, regardless of the manner in which each individual percentage is calculated no one category is greater than 2/3 of the total budget.

Common Bean Coordinated Agricultural Project
Project #: 2009-01929

BeanCAP Members' Response to Recommendations of Reverse Site Visit Team

Answers provided by:

Phillip McClean, Perry Cregan, Julie Garden-Robinson, Paul Gepts, Mike Grusak, David Hyten, Jim Kelly, Jim Myers, Juan Osorno

RESEARCH

1. We recommend that the Bean CAP team develop a set of criteria for inclusion in the panel. Further we suggest that the panel should represent diversity in common bean breeding programs and be structured to address specific articulated hypothesis.

Plant breeders and geneticists will nominate genotypes for inclusion in the diversity and association genotype panels. Those criteria and their rationale that will be considered when making the final selection follow.

1. The genotype must be adapted to the US production regions. This is especially important for the association pane. We want to ensure that we will be able to collect usable agronomic data and generate sufficient seed for nutritional analysis.
2. The dry bean genotypes will be chosen from the pinto, navy, black, great northern, and kidney market classes. These market classes are the major ones grown in the US and represent the three major races within the two distinct gene pools.
3. The snap bean varieties will represent a range of pod and horticultural types and will also be derived from the Mesoamerican and Andean gene pools. These are major gene pools from which modern snap beans are derived.
4. Preference will be given to modern genotypes. Breeders typically use modern genotypes in their crosses, therefore the association, and to a lesser degree diversity panel, will primarily sample that modern diversity.
5. The panel will not include any genotypes for which restrictions on crossing with other materials are in place. It would not be reasonable to sample allelic diversity among lines that cannot be used by public breeders in their breeding programs.
6. A subset of the diversity panel will include a limited set of older cultivars and landraces from which many modern genotypes in the Durango, Mesoamerican, and Nueva Granada races were derived. Major genetic changes have been incorporated into modern cultivars and it is critical for breeders to understand the extent to which the original germplasm background is retained in newer varieties. In contrast, limited genetic progress has been made in modern cultivars in the Nueva Granada race.

Two plant breeders and two geneticists will make the final decision regarding the diversity and association panels.

Careful selection of genotypes will allow the BeanCAP to address several hypotheses of importance to the bean breeding community as well as the plant genetics community. The first null hypothesis is that the modern breeding genotypes have equivalent diversity as older genetic

materials. As background, major changes in plant architecture traits were incorporated in races Mesoamerica and Durango during the 1980-1990s, allowing the release of several cultivars suitable for direct harvest in the recent years. This expanded the genetic diversity of these market classes compared to the diversity present prior to that period. However once the value of that architectural improvement was recognized, diversity was reduced in these classes as public and private breeders sought to exploit and maintain those 'introgressed' traits. So market class diversity is not a static component – it expands from time to time but likewise diversity is reduced until a point when progress is limited and new germplasm is needed to drive future improvements. Several studies have looked at genome-wide diversity of traditional genetic materials. All of these studies conclude that the Andean genepool is less diverse than the Mesoamerican genepool. Furthermore, the race Mesoamerican genotypes are more diverse than those from race Durango. The genotype collection described above for the diversity panel analysis will allow us to measure diversity among modern genotypes during the first year of the project. From a plant breeding perspective, this will inform the bean breeding community regarding the effectiveness of their focus upon modern genotypes. We may (or may not) discover a narrowing of the germplasm base which in turn should inform the breeding community regarding parental selection decisions in the future.

The second hypothesis regards linkage disequilibrium (LD). Specifically is LD distributed equally among: 1) the various gene pools; 2) within each market class or race; 3) among various linkage groups. The null hypothesis in each case will be that LD is equivalent. The extent of LD will also affect our choice of the number of genotypes for our association panel. If for example, we observe greater LD within the race Nueva Granada, then we would tend to include fewer genotypes of that race in association panel. The comparison of LD among the various linkage groups may point us to regions of the genome that may have been impacted to a greater extent by the domestication process.

The third hypothesis we will be able to test using our diversity panel is that specific regions of the genome have indeed undergone selection during the development of modern cultivars from the landrace form of the species. The most informative papers in this regard are Yamasaki et al. (2005, Plant Cell 17:2859) and Wright et al. (2005, Science 308:1310). The experiments we plan will compare allelic diversity among landraces and modern cultivars. If we see specific regions of the genome that are less diverse in the cultivars than landraces, then those regions could contain loci critical for the modern cultivar phenotype. An exciting prospect is the use of the annotate bean genome sequence to uncover potential candidate genes (and their allelic states) that make a cultivar well suited for our modern production conditions. By comparing the results for the various races, we can also determine whether the same regions of the genome associated with the "modern cultivar" phenotype within each of the races and/or whether regions controlling quality traits of older genotypes have been retained in modern cultivars.

2. We recommend that formatted data sheets for these traits be developed in conjunction with the database curator(s), such that consistent file formats and units are provided to Dr. Grusak for quality control and then to the database.

Phenotypic analysis of agronomic traits will follow standard procedures already in place and currently being used by the bean breeding community (e.g., in the Cooperative Dry Bean Nursery [CDBN] National Trials). All procedures are available on the Bean Improvement Cooperative (BIC) website at: http://www.css.msu.edu/bic/ResearchTechniques.cfm for dry bean

traits. We apologize for providing an incorrect web link in the previous response to the Review Panel.

Although all the procedures can be found at the link above, we provide the following definitions for the sake of summarizing some of the parameters that will be measured:

Dry Bean Traits

1. Early Vigor (EV): Scored on a 1 to 9 scale, where 1=excellent and 9= very poor, within the first three weeks after emergence.
2. Days to Flower (DF): Actual number of days from planting to when approximately 50% plants in a plot have at least one opened flower.
3. Days to Maturity (DF): Actual number of days from planting to when approximately 50% of plants in a plot have at least one dry pod.
4. Plant Height (PH): Recorded in cm from the base of the plant (soil surface) to the top node bearing at least one dry pod with seed.
5. Growth Habit (GH): Recorded during flowering and verified when crop is senescing as type I= determinate erect or upright, II= indeterminate erect, and III= indeterminate prostrate.
6. Lodging (LG): Scored at harvest on a 1 to 5 scale, where 1 =100% plants standing erect, and 5= 100% plants flat on the ground.
7. Pod Clearance (PC): Recorded at harvest as % pods on plants not touching the ground or in contact with the soil surface.
8. Biomass Yield (BY): Total plant dry weight recorded at 16% moisture and rounded up to the nearest whole number.
9. Seed Yield (SY): Recorded in pounds per acre at 16% moisture and rounded up to the nearest whole number.
10. Harvest Index (HI): The ratio of SY/BY expressed in % BY at 16% moisture.
11. Weight of 100 Seeds (SW): Weight of 100 randomly taken undamaged seeds recorded in grams at 16% moisture.
12. Appearance Desirability (AD): An aggregate value for seed size, shape, color, and brilliance for the respective market class scored on a 1 to 9 scale, where 1= excellent and 9= commercially unacceptable.

We are in the process of developing a similar list of traits and trait scales for the following snap bean phenotypes: **yield, pod shape, pod color, days to harvest, concentration of set, lodging, and architecture**. Dr. Myers, OSU will lead this development.

All collaborators will be instructed to use the same data collection procedures and units of measure (as detailed at the web sites above). Before the first year's planting, we will arrange a conference call between the PD (McClean), the Phenotype Database Curator (Grusak), and all the field site collaborators to discuss data collection and reporting procedures, timelines for data reporting, general expectations, potential problems, etc. With respect to the collection of trait data, all contributors will provide datasets in excel spreadsheet format using a standardized assignment of rows and trait columns. Pre-organized files will be provided to the collaborators and will include the genotypes chosen for the Association Mapping Panel. As part of the dataset file, collaborators will also provide site-specific information on plot location (latitude and longitude), environment (total rainfall, supplemental irrigation, soil type), fertilization (minerals, application rates), and disease control (pesticide applications) using standardized units of measure.

3. We recommend that some of the budget devoted to CAPS marker development might be better re-allocated to enhance other components of the proposal (including population development and validation of marker-trait association).

The funds reduced from the CAP development program have been reallocated to education activities. See the answer to recommendation 10 in this research section for more details.

4. We recommend that the Bean CAP expand their advisory committee to include an individual with expertise in next-generation sequencing and marker systems and an expert in education/extension.

The PD for the common bean sequence project, **Dr. Scott Jackson**, Purdue University, has been recruited as a member of the advisory committee. The addition of Dr. Jackson to the committee will ensure that the sequence information is incorporated into the BeanCAP in a timely manner. Since a significant portion of the sequencing project involves massively parallel next generation sequencing, Dr. Jackson will also bring that expertise to the committee. As currently constituted, Dr. Charles Hibberd (Purdue University), Dr. Charles Brummer (University of Georgia), Dr. David Sleper (University of Missouri), and Dr. Fred Bliss (Senior Director, R&D Special Projects, Seminis Vegetable Seeds) are members of the committee. **Dr. Hibberd** is the director of the Purdue University Extension, and therefore will provide our Extension expertise. A major goal of **Dr. Brummer's** program is the application of molecular markers to alfalfa improvement. Selected recent publications in the field plant molecular genetics include: Robins, et al (2008 Crop Sci. 48:1780); Cruz et al. (2007, Euphytica 153:43); Robins et al (2007 Crop Sci 47:1); Robins et al (2007 Crop Sci 47:11); Xiong et al. (2007, Mol Breed 18:327); Cruz et al (2006, Euphytica 152:339); Xiong et al (2006, Mol Breed 18:327). We feel Dr. Brummer will fulfill the role of an expert in marker systems. **Dr. Sleper** is a soybean breeder whose project focuses developing soybean varieties with improved food qualities. This focus is highly complementary to our project goals. Dr. Sleper is also the author of the textbook "Breeding Field Crops" and teaches the Plant Breeding and Genetics course at the University of Missouri. He has been extensively involved in plant breeder training during his entire career. Finally, **Dr. Bliss** is a long time plant breeder and recently he was the director of research for Seminis Vegetable Seeds. In these positions he was closely involved in training plant breeders and directing plant breeding programs in a large company. He has also published in the area of plant breeding training (Crop Sci, 2008, 47(S3):S250; HortSci, 2006, 41:45). We believe the collective knowledge of plant breeding training and education will be provided by Dr. Sleper and Dr. Bliss. Lastly, we will also add an individual with expertise in **bioinformatics, genomics, and translational genomics** to the advisor committee.

Beyond these members, we will invite **Dr. Linda Beaver** (University of Puerto Rico) to coordinate with Dr. Osorno and Dr. Urrea. These two individuals are the leads for the education portion of the project. Dr. Beaver has been actively involved in recruiting UPR students for internship and graduate programs in the US. In addition, Dr. Beaver has been doing recruiting at the high school level in Puerto Rico. She will be a valuable collaborator as we try to broaden our applicant pool for the BeanCAP education program.

5. We recommend that the Bean CAP team meet with the advisory committee on a regular basis, seek written input from the advisory committee, and respond to the advisory committee's criticism and advice (at least on an annual basis).

On p. 2 of the Management Plan provided with the original proposal, a plan for the Executive Committee, the PD, and the Advisory Committee to meet at the annual meeting Plant and Animal Genome conference held each January in San Diego is described. It is our intention to follow this plan. The NDSU budget contains funds to support travel of both the Advisory and Executive Committee to that meeting.

6. We recommend that the validation of associations be strengthened by including a clear description of population development and validation.

Validation of associations requires that an independent population be screened phenotypically and subsequently be genotyped For example, the McClean lab discovered two loci associated with iron deficiency chlorosis (IDC) by screening two entirely different population of breeding lines (Wang et al. 2008, TAG 116:777). Subsequently, two additional populations were recently genotyped with 1536 SNPs and the original marker/trait associations were confirmed and additional loci were observed (primarily due to greater genome saturation).

Our original plan was to develop bi-parental populations for validation, but further discussion has led us to alter our plans. We now intend to validate the loci by evaluating new populations in year four of the project. This will allow us to perform our statistical analysis in year three to discover the original associations for which testing will be performed in year four.

Because we plan to use the same number of lines proposed for the bi-parental populations, the cost of these experiments remains the same. Here are the details.

Unlike the original discovery phase, where lines representing all market classes were evaluated in four locations, the validation experiments will evaluate specific market class lines in the location where their production is greatest. This will allow us to increase the number of lines in each market class for the validation. Instead of having ~50 lines from each market class (the average in the original AM population) we plan to have ~four times that number in the validation population. The following table details the locations where these lines will be grown.

Location	Market class	No. of lines
North Dakota State Univ	Pinto	200
	Kidney	200
Michigan State Univ	Navy	200
	Black	200
Univ of Nebraska	Great Northern	200
Oregon State Univ	Snap beans	100

Each population will consist of include cultivars and heirloom varieties not used in the original population. In addition, the population will include advanced generation breeding lines that represent the most current breeding materials. (The value of using advanced generation lines is outlined Wang et al. 2008, TAG 116:777). These genotypic sources will ensure we are evaluating a larger number of recombination events for each market class.

These populations will be grown under standard production conditions at each location. DNA will be isolated and shipped to the USDA/Beltsville lab for genotyping. Statistical analyses will be performed as described in the original proposal and in our response to questions posed prior to the reverse site visit.

7. We strongly recommend moving this aim to Year 1 to insure the timely establishment of efficient interactions, and to define synergies among complementary projects early on.

Among individual groups working on legume databases, an informal agreement has been reached where crop specific databases are built and maintained by experts from each crop. Shared platforms, such as GBrowse and cMAP, would be used by each database to promote interoperatability between each database. At the time of funding, Dr Gepts will initiate discussions with the Soybase and the Legume Information Service that will focus on issues of interoperability and how best to leverage database resources in a manner that facilitates improvement of all legumes. One aspect will focus on the development of a common trait ontology in order to define commonalities with legumes, both crop and model legumes.

8. We recommend a re-consideration of the candidate gene approach.

The candidate gene approach is certainly valid if there is *a priori* reason to focus on genes with a specific pathway. To date though, this has not been a fruitful approach for genes presumably involved in mineral concentration in legumes. (No examples exist for common bean.) For example, O'Rourke et al. (2009, PAG XVII; http://www.intl-pag.org/17/abstracts/W40_PAGXVII_306.html) selected all genes from the 1.01 build of the soybean genome known to be involved in the iron uptake, translocation, and metabolism pathways. The physical locations of these genes were compared to the physical locations of SSR markers that define QTL associated with iron deficiency chlorosis in soybean. None of the SSR markers co-located with the "candidate genes".

The BeanCAP has an extensive list of phenotypes that will be evaluated. Since so little is known about the potential genes controlling these many traits, it would be difficult to predict what genes should be selected for resequencing and subsequent inclusion in the second OPA set. Furthermore, the important "nutrition" genes might not be pathway genes but rather transcription factors that control pathway genes. It would be very challenging to select among the many transcription factor families for resequencing and SNP development.

A more fruitful approach from our perspective would be to leverage the genetic results from the association analysis and the emerging common bean gene model set with the soybean genome models and assembly. Once strong marker/trait association signals are confirmed from several locations (or over all locations), a scan of common bean and soybean gene models in an interval around the SNP marker will be made. From these gene models, we would select any candidate gene (transcription factor, pathway gene, for example) that from our scientific judgment may be responsible for the phenotypic variation. Next, we would take individuals at the extremes of the phenotypic continuum and sequence the gene in each. From here, we would look for any potential causative variant (non-synonymous mutations; indels). The full population would then be analyzed, and an association analysis will be performed to see if the strength of the association increases, as we would expect if the candidate gene is affecting the phenotype. This

has been demonstrated as an effective approach to candidate gene discovery in humans (Yeo et al. 2003, Hum Mol Genet 12:561; Cohen et al. 2004. Science 305:869) and is worthy of testing in plants.

9. We recommend that the researchers reduce the total number of genotypes (28 = 4 genotypes x 7 market classes) which will be sequenced to a smaller number of thoughtfully chosen genotypes in each market class.

The Review Team recommended that the next-generation sequence analysis for SNP discovery be reduced from the analysis of 28 genotypes (4 genotypes/market class x 7 market classes) to 14 (2 genotypes/market class x 7 market classes). This seems to be a reasonable suggestion. Two carefully selected genotypes per market class should allow successful market class-specific SNP discovery and will require the purchase of only one, rather than two, Illumina Sequencing Flow Cells. Thus, we propose to follow the recommendation of the Review Panel and sequence only two genotypes per market class.

An alternative approach suggested by the Review Team was the creation of reduced-representation libraries via digestion of genomic DNA with methylation-sensitive restriction enzymes which favor sequencing the unmethylated portion of the genome. We believe this is an excellent alternative to the whole genome shotgun approach for the next-generation sequence analysis for SNP discovery. It will provide sequence data of genomic DNA, and it is better targeted to DNA with higher gene content. We propose to follow the suggestion of the Review Panel and use reduced representation libraries derived via digestion with methylation-sensitive restriction enzymes for library construction followed by next-generation sequencing and SNP discovery.

10. Some of the budget savings from sequencing should be redirected to education and extension objectives.

Rather than redirect the funds from sequencing, a better use of our resources would be to redirect funds from the reduced CAP development effort into education efforts. The following table shows how we plan to reduce the CAPs marker development.

Budget category	Original budget	Revised budget	
Year 1:			
CAPs supplies	$20,000	$10,000	
CAPS personnel	$13,000	$ 8,000	
Year 2:			
CAPs supplies	$15,000	$10,000	
CAPS personnel	$13,000	$ 8,000	
Year 3:			
CAPs supplies	$10,000	$10,000	
CAPS personnel	$13,000	$ 8,000	
Year 4:			
CAPs supplies	$10,000	$10,000	
CAPS personnel	$13,000	$ 8,000	
Total CAP budget			
CAPs supplies	$ 55000	$40,000	
CAPS personnel	$ 52,000	$32,000	
Total Budget	$107,000	$77,000	
Total reallocation			$35,000

A total of $35,000 will be moved to the education activities (See Table above.) These funds will reallocated to the education program in the following manner.

a. Education leaders travel: In year one, $10,000 is being allocated for the travel of education leaders to conferences where they can advertise and represent the education portion of the BeanCAP. During the second year of the project, $5,000 is being provided for these same activities. The total support for this travel is **$15,000**. Given that the project will be maturing, we are reducing the funding in this second year. Given that the training program will be fully operational in years three and four, we do not believe it necessary to fund this travel in those years.

b. Intern travel: To ensure we provide opportunities for students from throughout the US to participate in the intern program, we feel it is necessary to cover the interns travel cost to and from the work site once they have accepted the position. Therefore, we have allocated $5,000 per year (total $**20,000**) to support that travel.

11. The panel recommends that the Bean CAP researchers generate (or contract this step out to a service lab with experience) two normalized expression libraries from two diverged genotypes in the Snap Bean market class and to use GAII sequencing of those libraries.

The Review Panel recommended an alternative approach to SNP discovery that would rely on the next-generation sequence analysis of normalized expression libraries of two snap bean lines. The Illumina GoldenGate assay that will be used for SNP detection relies on hybridization of oligos to **genomic DNA**. In the design phase, Illumina requests that the user provide 60 bp of sequence on either side of the SNP as well as the two SNP alleles. Using this information Illumina selects appropriate allele-specific and locus specific oligos to optimize the GoldenGate assay. Following oligo design, an experiment that evaluates a germplasm collection is used to

determine the conversion rate. That rate is the % of the OPA primer sets that are useful for SNP detection. The only report we could find that described OPA primer development based on a normalized library was with lettuce where researchers obtained a 70% conversion rate (Matvienko et al. 2009, PAG XVII Abstracts; http://www.intl-pag.org/pag/17/abstracts/P01_PAGXVII_042.html). (Since the lettuce source sequence data was from cDNA, the lower conversion rate might be the result of primer that spans an intron. This would not be an issue with sequence data obtained from genomic data.) By contrast, a conversion rate of 89% was obtained from resequencing genomic DNA from soybean (Hyten et al. 2008; TAG 116:945) and 84% using a reduced representation soybean library (Hyten and Cregan, personnel communication). We anticipate the conversion rate for common bean to be around 90% because we used SNPs that were predicted with three or more Solexa reads whereas the soybean SNPs relied on a single Solexa read. It should be noted, though, that the panel recommendation of developing a reduced representation genomic library from methyl-filtrated DNA would theoretically increase the number of genes that will be defined. This hypothesis can be tested by comparing the sequences from the reduced representation library with the recently developed library based on resequencing of random genomic loci in common bean. That result should inform researchers working on other crops.

12. We recommend that the Bean CAP team restate deliverables using more comprehensive and conservative estimates.

The following is a revised list of BeanCAP objectives. It was unclear from the recommendations as to whether we should simply restate the marker deliverables or all of them. We have modified all of them.

Objective 1: Develop high throughput, market-class-specific markers for the predominant common bean market classes produced in the US, convert those markers into breeder-friendly markers, and genotype breeder-defined populations with these markers.

Year 1:

- **SNP diversity data** for a collection of 196 common bean genotypes.
- **Solexa sequence data** for two genotypes from each of the following market classes: pinto, navy, black, Great Northern, kidney, and snap beans.
- **SNPs** defined for each race and OPA set developed
- **66 CAP loci** developed; two per chromosome for each common bean race (2 CAP loci x 11 linkage groups x 3 races)

Year 2:

- **SNP data** for ~2000 genotypes nominated by breeders and geneticists
- **66 CAP loci** developed; two per chromosome for each common bean race (2 CAP loci x 11 linkage groups x 3 races)

Year 3:

- **SNP data** for ~2000 genotypes nominated by breeders and geneticist.
- **~20 CAP marker loci** developed for breeder defined phenotypes either discovered in the project or defined by breeders

Year 4:

- **SNP data** for ~2000 genotypes nominated by breeders and geneticist.
- **~20 CAP marker loci** developed for breeder defined phenotypes either discovered in the project or defined by breeders

Objective 2: Discover genetic loci associated with nutritional traits that define "healthy beans" by combining genotype and nutritional profile data of association mapping and bi-parental populations.

Year 1:

- **Nutritional and agronomic performance data** for ~300 dry bean and ~150 snap bean grown under controlled greenhouse conditions
- **Grow out** lines for association mapping

Year 2:

- **Nutritional and agronomic performance data** for ~300 dry bean (grown at four locations) and ~150 snap bean (grown at one location)
- **Nutritional and agronomic performance data** for ~100 dry bean grown at four locations under water stress conditions

Year 3:

- **Marker loci associated with nutritional and agronomic traits** defined on a location-by-location and over all locations using data from the association mapping population

Year 4:

- **Confirmation of QTL associated with nutritional and agronomic traits** discovered using market class specific association mapping populations grown at one location, each, for the pinto, navy, black, Great Northern, kidney, and snap bean market classes

Objective 3: Integrate common bean phenotypic, genotypic, and molecular marker data with other emerging legume genomic resources into breeder-friendly bioinformatic tools.

Year 1:

- **Establish the Phaseolus Genes database**; link it through the Bean Improvement WWW site
- **Begin enriching the Phaseolus Genes database** by incorporating **historical mapping and QTL data** generated by the bean research community prior to this project, and where possible, **link those markers** to available common sequence data

- **Begin establishing relationships with other legume database** to determine feasibility of **interoperability among the databases**.

Year 2

- **Begin incorporating sequence information** from the common bean sequencing project into **the Phaseolus Genes database**
- **Complete** incorporating **historical mapping and QTL data** generated by the bean research community, prior to this project, into the database, and where possible, **link those markers** to available common sequence data
- **Establish principles** that enable breeders to select appropriate markers for genetic and breeding purposes.

Year 3:

- **Continue incorporating sequence information** from the common bean sequencing project into **the Phaseolus Genes database**
- **Test candidate gene discovery** based on AM mapping data and colinearity between common bean and soybean.
- **Enrich the Phaseolus Genes database** by incorporating all BeanCAP project phenotypic and marker information into the database.
- **Implement practices and protocols** that enable breeders to select appropriate markers for genetic and breeding purposes.
- **Develop FAQs and How-Tos** that describe the utility and practices of the database**.**

Year 4:

- **Complete incorporating sequence information** from the common bean sequencing project.into **the Phaseolus Genes database**
- Additional **candidate gene discovery** based on AM mapping data and colinearity between common bean and soybean
- **Test and improve FAQs and How-Tos** that describe the utility and practices of the database**.**

Objective 4: Launch the "**Nutritional Genetics and Genomics: Healthy Foods from the Field to the Table"** WWW presence that uses high-quality animations and other multimedia to highlight the biology and technology associated with the genomic-based improvement of nutritional traits.

Year 1:

- **BeanCAP WWW site** will be launched.
- **Ning** learning community will be established and advertised to the Phaseolus community
- **Animations** focused on the themes of 1) root biology and the role soil chemistry plays on nutrient uptake; and 2) the flow of soil minerals from the root to various parts of plants will be developed. These will be delivered from the BeanCAP WWW site, the Ning site, and via Youtube.

- **Supporting still images and quizzes** will be developed that support the learning points that are the focus of the animations.
- The **Moodle** learning site will be established. Learning materials related to plant breeding as a career will be entered into the site.
- **Advertise the BeanCAP** at the biennial **Bean Improvement Cooperative** meeting.

Year 2:

- **Frequently Asked Questions (FAQ)** will be developed and populated on the Ning site.
- **Animations** that focus on a few of minerals and show how they move from our digestion system to the various organs in our body, and what health problems arise from mineral deficiencies in our diet. These will be delivered from the BeanCAP WWW site, the Ning site, and via Youtube.
- **Supporting still images and quizzes** will be developed that support the learning points that are the focus of the animations.
- **MS Powerpoints**, with audio, that describe plant breeding as a career will be developed and distributed via the BeanCAP WWW and the Ning site.
- A **Moodle** learning module that outlines the principles of plant breeding will be developed.
- The **Ning** site will be updated with announcements related to the project and links to learning materials.

Year 3:

- **Frequently Asked Questions (FAQ)** will be modified and extended.
- **Animations** that provide a basic lesson in genetics by emphasizing the fact that many genes are involved in the uptake and distribution of the minerals throughout the plant, and that the genome of the plant is required to produce all of the proteins necessary for functional uptake and distribution of minerals throughout the plant. These will be delivered from the BeanCAP WWW site, the Ning site, and via Youtube.
- **Supporting still images and quizzes** will be developed that support the learning points that are the focus of the animations.
- A **Moodle** learning module that teaches the day-to-day practice of plant breeding will be developed.
- The **Ning** site will be updated with announcements related to the project and links to learning materials.

Year 4:

- **Frequently Asked Questions (FAQ)** will be modified and extended.
- **Animations** that emphasize how plant breeding utilizes the genetic information for trait improvement (mineral uptake) and the role that genomics in developing the tools that assist the breeder in trait selection. These will be delivered from the BeanCAP WWW site, the Ning site, and via Youtube.
- **Supporting still images and quizzes** will be developed that support the learning points that are the focus of the animations.

- A **Moodle** learning module that teaches the scientific intersection of plant breeding, genetics, and genomics will be developed.
- The **Ning** site will be updated with announcements related to the project and links to learning materials.

Objective 5. Initiate a modern plant breeding training program that focuses on early career recruitment and provides practical training that illustrates how the integration of genomic and phenotypic data can be used to improve nutritional traits in plants.

Year 1:

- **Promotional materials** such as brochures, flyers, and posters describing plant breeding will be developed for the high school audience.
- **Collating and delivery** of advanced learning materials from other CAP projects on the BeanCAP WWW site.
- **Curriculum** for summer and year-round internships will be developed.
- **Summer and year-round internships** will be offered by each participating university. 8 total (2 per site x 4 sites)
- **Intern program database established** to track students from their college experience to graduate school to a plant breeding career.
- **Contacts** established with local high schools.
- **High school visits** completed with a target of 300 students (75 students at each of 4 locations).
- **Visits to breeding programs** at each of the four locations by local high school students.
- Leaders will **attend national agricultural conferences to announce the BeanCAP education program**.

Year 2:

- **Summer and year-round internships** will be offered by each participating university. 8 total (2 per site x 4 sites)
- **High school visits** completed with a target of 300 students (75 students at each of 4 locations).
- **Visits to breeding programs** at each of the four locations by local high school students.
- Leaders will **attend national agricultural conferences to announce the BeanCAP education program**.

Year 3:

- **Summer and year-round internships** will be offered by each participating university. 8 total (2 per site x 4 sites)
- **High school visits** completed with a target of 300 students (75 students at each of 4 locations).
- **Visits to breeding programs** at each of the four locations by local high school students.

Year 4:

- **Summer and year-round internships** will be offered by each participating university. 8 total (2 per site x 4 sites)
- **High school visits** completed with a target of 300 students (75 students at each of 4 locations).
- **Visits to breeding programs** at each of the four locations by local high school students.

13. We recommend that the Bean CAP team engage in an ongoing dialog concerning how marker-assisted-selection can be more fully integrated into breeding programs.

We agree that is a critical activity that is necessary for the project to have significant impact. As a first step, we will hold a meeting at the 2009 Bean Improvement Cooperative (BIC) Biennial meeting that will be held October 25-28, 2009. The first purpose of the meeting will be to describe the BeanCAP project with emphasis on the genotyping service and the marker tools that will be developed. From this meeting, we will recruit several interested breeders, both public and private, to interact with the marker development leads (Cregan, Grusak, Kelly, Myers, Miklas, Gepts; see Management Plan for organizational structure). These individuals will provide input regarding 1) the utility of the marker data that is being generated, 2) how the data will be organized in the marker database, and 2) how a breeder would utilize the database for marker selection. Beyond data organization, we will evaluate that process by which markers are adopted by a breeding program. Particular emphasis will be placed on 1) the traits emphasized (disease? agronomic?) in a program, 2) the utility of a marker in the trait selection process, and 3) the financial resources available for marker-assisted selection.

Of particular interest is how private companies are utilizing markers and marker-assisted selection. It is generally acknowledged that major plant breeding companies have marker capabilities well beyond what is available to public programs. We will attempt to develop a dialog with one or more major companies in an attempt to collect a list of “best practices”. We will then evaluate these relative to resources available to public programs and determine which would be amenable to a public program. These will then be reported to the bean community at a subsequent BIC meeting.

Extension/Education

1. We recommend that the Bean CAP team include a more comprehensive plan to establish a CoP.

After extensive discussion within the Extension group of the BeanCAP and following consultation with Dr. Mike Newman (Extension member of the site visit panel), we have altered our plans with regards to the Community of Practice. Rather than developing a COP, our resources will be spent on creating several entry points for the BeanCAP extension information. We also intend to interweave a number of the educational materials into these

The BeanCAP WWW site will collect all of the various portals, but they will also be accessible from distinct WWW addresses.

The Web site will have some static features and some interactive features:

1. The first portal will support **social networking** using the Ning technology. The **Ning** site will allow for networking among professional plant breeders, molecular biologists, the research team and others interested in being part of the community. The Ning social networking tool allows the posting of research results, videos, photos, wall posts and has the capability to allow for threaded discussions and sharing of research results. Ning supports the development of niche social networks that are membership centric. An important feature is that Ning does not allow adult networks. This is an important consideration because we may have high school students visiting the network. The network will be hosted on the Ning network and can be accessed through http://www.ning.com/.

2. The successful delivery of our VCell animations via **youtube.com** will be replicated for this project. We will develop a youtube channel that will provide a single entry point for all of the animations. We will advertise these animations initially via the NDSU Virtual Cell channel where >1,600 individuals have registered. The **Ning** site also will provide a direct link to the channel. In addition, links will be provided via the **Moodle** modules (see below) as well at the BeanCAP site.

3. Online module(s) will be designed for high school students to learn about careers in plant breeding. We will use **MS Powerpoint** presentations that link audio with visual components. We chose this solution, rather than enhanced podcasts, because the latter only can be delivered via iTunes. If new technologies emerge that allow us to deliver these modules via the WWW, we will test their feasibility and adopt them if we feel it is a satisfactory solution. These presentations will be available in static form as downloadable written materials in .pdf format. All of the materials will be available from the education component of the BeanCAP site.

4. Supplementary teaching materials (quizzes, etc) will be created and posted within a "**Moodle learning management system**" for use by registered teachers and other users of the Web site. We anticipate that this solution will be teacher centric and used to teach the basic concepts of genetics, genomics, breeding, and nutrition. Our animations will be interwoven in the Moodle course. The Moodle application will be installed on an NDSU server.

5. Questions and answers for the general public will be offered via an **"Ask an Expert" widget.** Here, the user will pose specific questions, and a member of the BeanCAP team will provide the answer. This will be supplemented by a **FAQ section developed by the group**.

6. The BeanCAP WWW site will offer links to relevant research-based information. The target audience here will be professionals, high school students and consumers. The BeanCAP will filter the sites to ensure that our target audiences receive the best information based on the current knowledge base.

2. We recommend that the Bean CAP team develop a marketing and outreach plan to promote proposed educational/extension activities.

We fully support these recommendations and will plan to implement those during the course of the project. Here are a few highlights.

a. We have already budgeted funds for the development of the publicity materials.

b. As stated above, we will be spending $15,000 in the first two years for travel to various meetings to promote our educational program. (These funds came from a reduction in the CAP marker development effort.)

c. We describe workshops for users of our research results in the response to recommendation 13 in the research section above. In addition, we will work with the North Dakota Dry Bean Council to offer a session that highlights the program during their annual Bean Day meetings held during January in Fargo, ND each year. Once developed, we plan to have bean researchers in other states offer a similar session at appropriate bean growers/producers meeting.

d. The original proposal outlined high school and internship programs. Those programs are fully funded and will be carried out throughout the duration of the project.

e. Based on experience with our student interns, we will develop a short-course that emphasizes genomics as a research field and plant breeding as a career. This will be implemented as a Moodle course.

Other

We were asked to show in Table form the research, education, split to ensure that no one category received more than 67% of the budget. The table below shows the split by location and activity area and that no one area received greater than 67%

Location	Research	Education	Extension
Colorado State Univ	$ 312,315	$128,027	$ 0
Michigan State Univ	$ 213,160	$117,949	$ 0
NDSU	$ 371,609	$194,870	$388,953
Oregon State Univ	$ 75,641	$ 0	$ 0
Univ CA, Davis	$ 549,691	$ 0	$ 0
Univ Nebraska, Lincoln	$ 46,154	$117,949	$ 0
USDA/Beltsville	$ 727,356	$ 0	$ 0
USDA/Houston	$ 148,500	$ 0	$ 0
USDA/Mayaquez	$ 13,200	$ 0	$ 0
USDA/Prosser	$ 30,800	$ 0	$ 0
Prorated management	$ 187,942	$187,942	$187,942
Total activity area	$2,676,368	$746,737	$576,895
Percentage total budget	67%	19%	14%