Executive Summary

Towards a Whole Genome Sequence of Common Bean, (*Phaseolus vulgaris*): Background, Approaches, Applications

Phillip McClean, North Dakota State University
Steve Cannon, USDA/ARS, Ames, IA
Paul Gepts, Universtiy of California, Davis
Matt Hudson, University of Illinois
Scott Jackson, Purdue University
Dan Rokhsar, DOE-Joint Genome Institute, Walnut Creek, CA
Jeremy Schmutz, DOE-Joint Genome Institute, Stanford, CA
Carroll Vance, USDA/ARS, St. Paul, MN

Why Sequence the Common Bean Genome?

There are three compelling reasons that common bean (*Phaseolus vulgaris* L.) should be sequenced. First, common bean is a crop that provides critical protein and calories world-wide and particularly to many underdeveloped countries in the world. The sequence will provide powerful tools to improve agronomic and nutritional traits so important to maintaining and improving the nutritional status of these individuals. Second, while common bean is $1.2 billion crop in the US, it is losing competitiveness because it does not have the biotechnology tools available to other crops. The sequence will greatly facilitate marker assisted selection, an important genetic alternative to improve the economic competitive of the crop. And third, as a close relative of soybean, cowpea, and pigeonpea, its sequence is importance for the study of the function of genes within this economically important group of legumes.

Common Bean, A Societally Important Crop. Common bean is a major source of calories and protein in many developing countries throughout the world (FAO: http://faostat.fao.org/) providing as much as 15% of the total daily calories and greater than 30% of the daily protein intake per day in numerous countries. In these poorer countries, malnutrition is a serious health issue that is an aggravating factor for diseases such as HIV-AIDS and tuberculosis. Common bean is a rich source of zinc and iron, two micronutrients depleted from individuals with AIDS. Micronutrient rich foods such as bean improve the health status of HIV infected patients. When placed in this perspective, the value of common bean is best seen through its role as a societally important crop that is worthy of aggressive improvement using the tools that can be generated by the availability of a whole-genome sequence.

***P. vulgaris* Competitiveness in the Evolving Agricultural Production Environment**. As dual purpose food and fuel crops become more prevalent, other crops will only find a place in the production system by being economically competitive. It is estimated that the new technologies in the pipeline will increase soybean and corn yields by 3%/year. Therefore, to maintain long term competitiveness, a genetic improvement program in bean must meet or exceed these targets. However common bean does not have any transgenic solutions in hand or in the pipeline. Therefore the only real option available for common bean improvement is the genetic approach that utilizes an aggressive marker assisted selection (MAS) program. The best markers for MAS are the gene itself, or a sequence that resides very close the functional gene.

A whole genome sequence will greatly facilitate the fine mapping of bean and enable the high throughput selection of improved varieties.

***P. vulgaris* and Soybean: Complementary Reference Species for Phaseoleae Research**. The crop value of common bean in the US ($1.2 billion) is of equal to the combined value to all of the other vegetable leguemes (pea, chickpea, peanut, lentil). As the closest crop relative to soybean, it is arguably the best diploid model for soybean. Extensive macrosyntentic relationships also exist between the species. These species also share a number of phenotypes that appear to be under the similar genetic control. A careful reconstruction of the duplicated soybean genome relative to the common bean genome beyond that macrosyntentic level could benefit both bean and soybean by letting each take advantage of knowledge gained from the other species. Clearly a full genome sequence of common bean would accelerate the improvement of these two species.

A Plan to Sequence the *P. vulgaris* Genome

The current mainstream approach to genome sequencing is the whole genome shotgun (WGS) method based on the Sanger technique. Yet, whole genome sequencing appears to be at a tipping point where new approaches, such as the 454 pyrosequencing method, are being adopted by offering a cost advantage along with increased speed and throughput. It is unclear how successful the long (400-500 pb) 454 reads will be in determining the full sequence of relative moderate size genome such as *P. vulgaris*. Therefore we propose a hybrid approach that combines long 454 reads with a modest Sanger sequencing effort. It should be noted that this approach has provided high consensus accuracy for sequencing the gene space of Arabidopsis. For the 454 portion of the project, we propose 22 runs to generate a 15X genome coverage of the 600 Mb genome. This would be supported by another eight runs of various sizes would also be performed. For the Sanger sequencing portion of the project, we propose a modest 4X plasmid end-shotgun sequence along with ~40X clone coverage in 35 kb fosmids. Because of the novel nature of the sequencing proposal, it will be important to have sufficient genetic and physical tools for scaffold assembly and merging the scaffolds into pseudochromosomes. A greater number of BAC-end sequence are needed to increase scaffold size, and additional markers are necessary to develop the high density map to develop pseudochromosome development. Fiinally, a bioinformatics effort is required that will assemble the genome, predict and annotate gene models, and archive the data for long-term access by the research public.

Applications of the Sequence to Legume Research

The following examples highlight how the genome sequence of common bean will be applied:

- Development of a new set of SSR and SNP bean markers for between and within gene pool MAS
- Fine-mapping and candidate gene identification in bean
- Comparative gene discovery in legumes
- Identification of Phaseoleae adaptation genes
- Creation of the next generation training set for legume gene model

Towards a Whole Genome Sequence of Common Bean, (*Phaseolus vulgaris*)

Background, Approaches, Applications

Phillip McClean, North Dakota State University
Steve Cannon, USDA/ARS, Ames, IA
Paul Gepts, Universtiy of California, Davis
Matt Hudson, University of Illinois
Scott Jackson, Purdue University
Dan Rokhsar, DOE-Joint Genome Institute, Walnut Creek, CA
Jeremy Schmutz, DOE-Joint Genome Institute, Stanford, CA
Carroll Vance, USDA/ARS, St. Paul, MN

Version 2.0

I. The Importance of a Whole Genome Sequence of Common Bean

There are three compelling reasons that common bean (*Phaseolus vulgaris* L.) should be sequenced. First, it is an important crop world-wide in many diets. Secondly, the data will provide important research tools to make this $1 billion crop competitive in the rapidly evolving production agriculture environment. And finally, the sequence will have great utility as a companion to the soybean genome sequence for the study species within the economically important Phaseoleae clade of legumes.

Common Bean, A Societally Important Crop. Common bean is a major legume crop with significant nutritional importance. It is a major source of calories and protein source in many developing countries throughout the world (FAO: http://faostat.fao.org/). For countries such as Burundi and Rwanda, with some of the lowest total caloric intakes per day, common bean provides about 15% of the total daily calories. It also provides greater than 30% of the daily protein intake per day in these countries. In these poorer countries, malnutrition is one aggravating factor for AIDS patients. In particular, micronutrient deficiencies and HIV-1 disease progression are associated (Baum et al. 1995). Common bean is a rich source of zinc and iron, two micronutrients depleted from individuals with AIDS (Savarino et al. 1999; Buys et al 2002). Diets containing foods rich in these micronutrients are suggested to benefit the health status of HIV infected patients (ADA 2004; Kruzich et al. 2004) which in turn delays the onset of AIDS. Common bean also contains a protein that inhibits the HIV-1 reverse transcriptase (Wong et al. 2006). Collectively, these features support the importance of common bean as one of the many factors that can address the AIDs problem through improved nutrition. The immediate benefit of improved nutrition will be to improve food security in these countries (Gillespie and Kadiyala 2005). When placed in this perspective, the value of common bean is best seen through its role as a crop worthy of aggressive improvement using the tools that can be generated by the availability of a whole-genome sequence.

***P. vulgaris* Competitiveness in the Evolving Agricultural Production Environment**. Competition for acres is requiring new strategies for crops to succeed as viable economic crops in the current agricultural production environment in the US. As dual purpose food and fuel crops become more prevalent, other crops will only find a place in the production system by being economically competitive. Production costs for soybean and corn, two major food and fuel crops, have been lowered (and profits increased) by the introduction of new production traits by using transgenic technology. To be viable and find their niche in the production system, a crop such as common bean, which does not have any transgenic solutions in hand or in the pipeline, must utilize genetic solutions to be successful. For common bean improvement, the genetic approach that utilizes marker assisted selection (MAS) is essential.

Public and private US breeders and geneticists have utilized MAS to develop germplasm and cultivars with improved disease resistances [anthracnose (Ragagnin et al. 2003; Miklas et al. 2003c), *Bean common mosaic virus* (Kelly et al. 1994; Miklas et al. 2002a; Miklas and Kelly 2002), *Bean golden yellow mosaic virus* (Beaver et al. 2005), common bacterial blight (Miklas et al. 2006b; Mutlu et al. 2005), bean rust (Stavely 1998), and white mold (Miklas 2007). This genetic effort has focused on diseases because bean yields are impacted by fungal, viral, and bacterial diseases. For a disease like common bacterial blight, up to 75% of the plants in some fields were infected with the pathogen resulting in yield losses of up to 40% (Saettler 1989). For a disease such as bean rust, yield losses in the US ranged form 30-40%, while 100% losses were observed in other fields (Stavely and Pastor-Corrales 1989). A more detailed study showed that a 1% incidence of the disease reduces yield by 19 kg/ha (Lindgren et al. 1995).

To estimate the economic impact of a concerted molecular breeding effort to improve disease resistance in common bean, and subsequently its competitiveness, economic data related to yield gains must be considered. First, it is estimated that averaged over all diseases, yield is reduced 10% by bean pathogens (Schwartz et al 2005). Again this is just a working estimate compared to the specific impacts described above. Therefore, what follows is a conservative estimate of the benefit that can be reaped by having a full genome sequence available to develop the necessary molecular tools available for genomics assisted breeding.

The best estimate of the production costs for dry bean is available at: http://www.ag.ndsu.edu/pubs/ecguides.html. Using the 2008 crop budgets for southeast North Dakota, and including $24/acre cost for fungicide as additional input cost for disease protection, the return to labor and management is $138/acre. If we were to increase yield 10% by focusing on disease resistance, which in turn would eliminate the fungicide cost, the return would be $189/acre. Nationally an additional return of $51/acre would represent $72 million to US dry bean producers (based on 2008 USDA production estimate of 1.41 million acres). Several points need to be emphasized. First, it is estimated that the new technologies in the pipeline will increase soybean and corn yields by 3%/year. Therefore, to maintain the long-term competitiveness of the crop, a genetic improvement program in bean must meet or exceed the targets in those crops. And to continue this improvement we need the best genetic data; that data is a full genome sequence.

***P. vulgaris* and Soybean: Complementary Reference Species for Phaseoleae Research**. Legumes are important components of our agricultural economy. From an economic stand point the two most important species are soybean (*Glycine max*) and alfalfa

(*Medicago sativa*) with 2007 crop values in the US of $26 billion (rank: #2) and $9 billion (rank: #5), respectively. These two species diverged 54 mya and are members of the two major legume clades, the Phaseoleae (bean, soybean, cowpea) and the IRLC (alfalfa, lentil, pea) (Fig. 1). Among the remaining legume crops, the 2007 value of common bean was $1.2 billion (rank: #17) which was equal to the value of all of the remaining legumes on the USDA crop value list (peanut, chickpea, pea, and lentil).

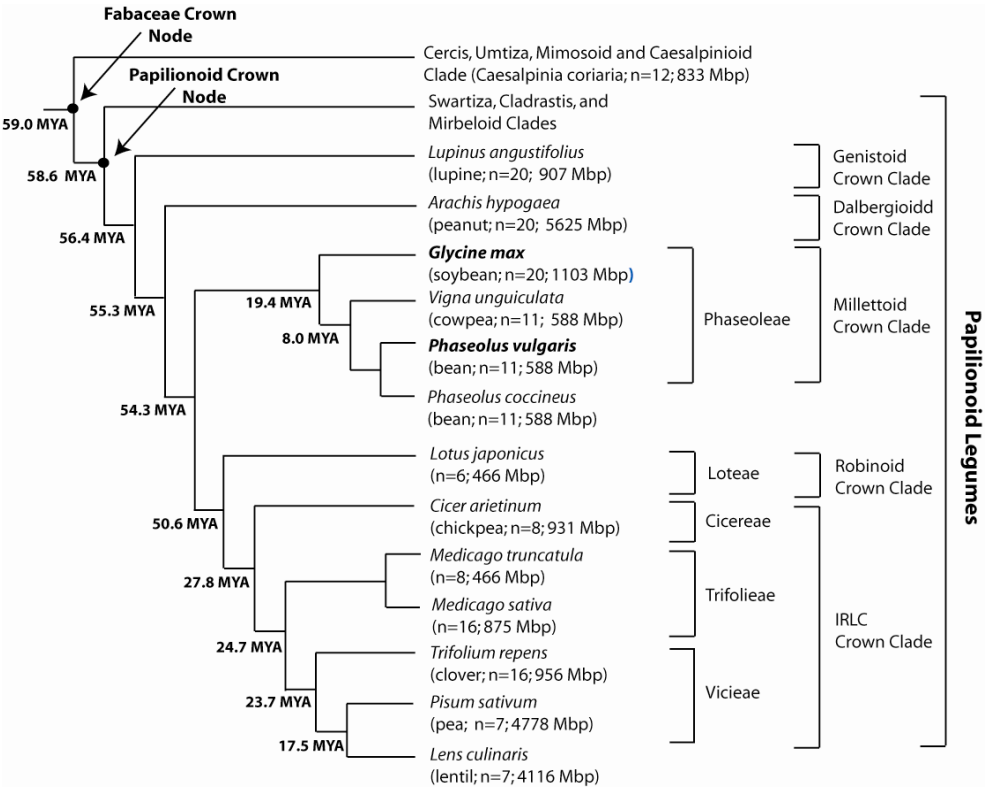

Figure 1. The phylogeny of the economic legumes relative to the others in the Fabaceae. Based on Lavin et al. 2005.

Given the economic importance and phylogenetic and relationship of common bean and soybean, attention has recently focused on common bean as the model diploid species for soybean and Phaseoleae research (McClean et al. 2008). Both species are members of the Phaseoleae clade and evolutionarily are recent relatives (19 MYA; Lavin et al. 2005). Also early macrosyntenic studies found common bean shares the greatest sytneny with soybean (Boutin et al. 1995), and that among the major legumes, these species exhibited the least amount of chromosomal rearrangements (Lee et al. 2001). To further the concept of common bean as a diploid model for soybean, it was important to understand the duplication history of common bean and to further study the macrosynteny between the two species.

To be a diploid model, common bean should not share any event associated with soybean's recent polyploidy history but should show share duplication events that preceded the evolutionary split between the two species dated. Schlueter et al. (2004) and Pfiel et al. (2005) studied the evolutionary history of gene duplicates and discovered a whole-genome duplication event shared between the legumes soybean and *M. truncatula*. This event presumably predated the divergence of the legumes. In addition, a recent duplication in soybean that appears to be

the signature of the polyploid event that led to the modern genome organization of the species was observed. Recently Mamidi et al. (2008) used a phylogenetic reconstruction based on 220 gene families that included the major Phaseoleae species and other legumes to show that the whole genome duplication event was unique to soybean and not shared by *P. vulgaris*.

If *P. vulgaris* is a diploid version of soybean, and soybean has undergone a whole genome duplication, then each *P. vulgaris* locus should map to two loci in soybean. To test this assumption, McClean et al. (in preparation) compared all mapped genes from bean against all of the scaffold sequences from the initial build of soybean genome (Schmutz et al. 2008). As fully expected the blastn scores for the two best hits were low (median = $4x10^{-94}$) and the percent identity was high (median=90.7%). Importantly, a high degree of conserved gene macrosynteny was observed between the full *P. vulgaris* genetic map and the soybean genome build. Fig. 2 shows the synteny for two *P. vulgaris* linkage groups. Several discoveries are important to note. First, as expected for a diploid model, nearly all of the *P. vulgaris* genes are duplicated in soybean. Secondly, note that a single *P. vulgaris* linkage group is syntenic with not two soybean chromosomes, as would be expected for a simple duplication, but rather multiple chromosomes. This implies that following the duplication event, the soybean genome was fractionated and the fragments rearranged to construct the modern soybean genome with 20 chromosomes.

All of the data was collected from these comparisons, and an entire reconstruction of all soybean chromosomes relative to *P. vulgaris* was assembled (Fig. 3). The major theme is that soybean is a duplicated genome that underwent major rearrangements prior to assembly of the modern genome. This conclusion is fully consistent with the original RFLP map of soybean that showed the nearly all loci were duplicated and that the duplication represented a mosaic pattern that would only have arisen following rearrangements (Shoemaker et al. 1996)

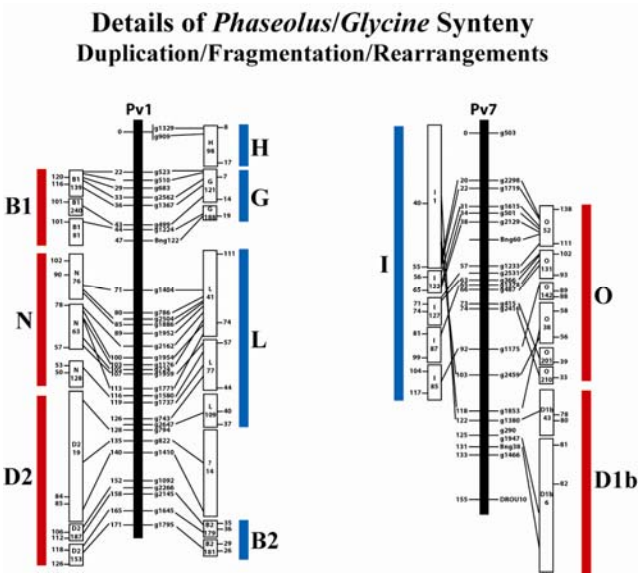

Figure 2. Conserved macrosynteny between the *P. vulgaris* genetic map and the soybean genome build. The genetic order of mapped gene *P. vulgaris* transcripts is in black. The open boxes are soybean scaffold using the nomenclature of the first build (http://www.phytozome.net/soybean). The red and blue bars are a representation of the length of specific soybean chromosomal fragments that map to a particular *P. vulgaris* linkage group.

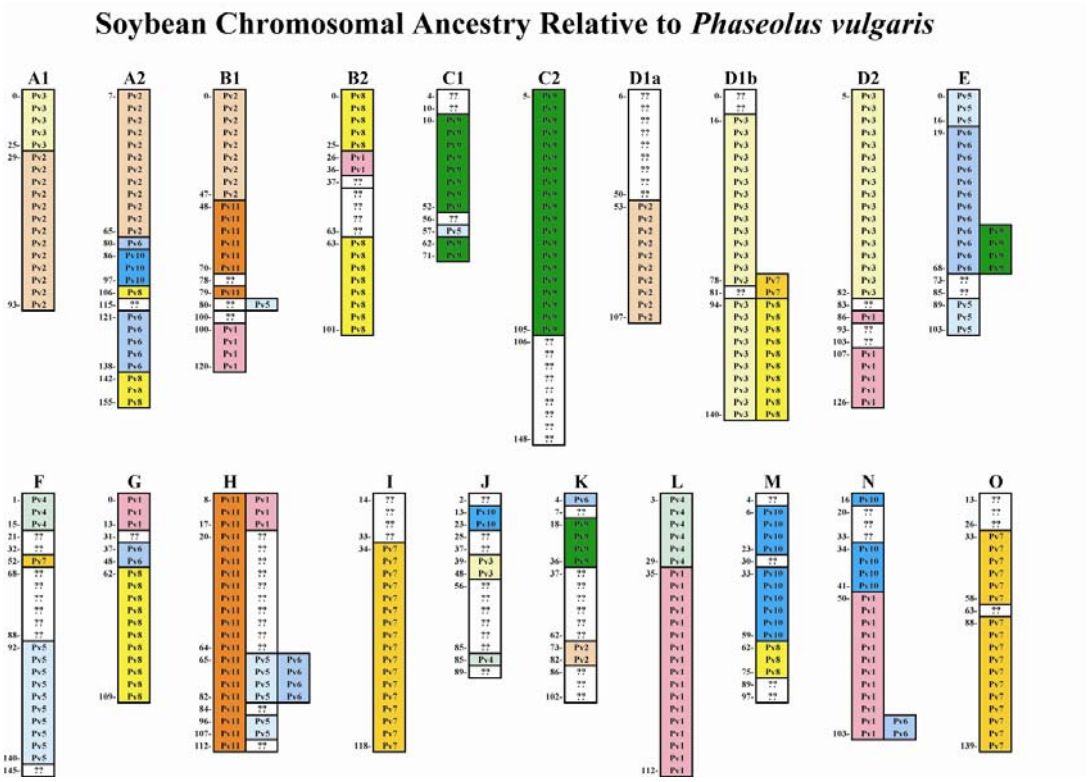

Figure 3. The reconstruction of all soybean chromosomes relative to *P. vulgaris* linkage groups. Blocks of the same color represent the same *P. vulgaris* linkage group.

Can this syntenic research be exploited in a way that can be used to improve soybean? A major gene in the soybean production system is *Dt1*, which controls the determinancy phenotype (Woodworth 1933). Determinate plants are short and don't lodge under the modern production practices. Industry is preparing the next generation of soybean, and the target is to increase yield by 3%/year over the next 20 years. This will make for a heavier plant that could have lodging problems. Therefore understanding the genetic nature of *Dt1* may allow researchers to modify the gene in a manner that shortens the plant further yet still provide the structure necessary to support the yield gains. First, *Dt1* must be identified.

The function of *Dt1* is very similar to the function of the *Fin* locus in common bean, another determinancy locus. *Fin* maps to the center of the Pv1 linkage group. The Gepts lab recently cloned an ortholog of the *Arabidopsis TFL1* determinancy locus and discovered that it cosegregated with *Fin* (Kwak et al. 2008). The comparative mapping described above showed that Pv1 was syntenic with, among others, chromosome L of soybean (Fig. 2). *Dt1* maps to chromosome L, 0.1 cM from the BARC-029975-0675 SNP marker (http://soybeanbreederstoolbox.org/). All gene models 250 kb up and downstream of the marker were collected, and a blastn analysis using the common bean *TFL1* ortholog discovered a candidate gene for the *Dt1* gene within this interval. Sequence analysis is underway to determine the functional difference between the indeterminate and determinate alleles at this locus (Lee and McClean, unpublished).

This example is important for several reasons. First, it shows a high degree of microsynteny than has been demonstrated for other crops such as rice and wheat. Assuming this microsynteny is highly extensive, then these two species can serve as reciprocal reference

genomes. All that is needed is a carefully reconstruction of the duplicated soybean genome relative to the common bean genome beyond that afforded so far by using common bean markers at a macrosyntentic level. This could benefit soybean by offering a vast array of new alleles (or genes) that can expand the value of this very important economic crop. Clearly a full genome sequence of common bean would accelerate this improvement.

II. Common Bean Genomic Features

Genome size and G+C content. *Phaseolus vulgaris* L. is a diploid species with 11 chromosomes and a medium-sized genome with estimates ranging from 588 (Bennett and Leitch 2005) to 637 Mbp (Aramuganathan and Earle 1991). Recently gene sequence data was collected from over 1000 genes in common bean (McConnell et al. 2007). The G+C content (including data from exons, introns, and 3'-UTR) is 40.1%. This value is consistent with results (39.4%) obtained using the historical chemical approach to measure G+C content (Baxter and Kirk 1969).

Polymorphism. Nucleotide polymorphism between BAT93 and Jalo EEP558, the parents of the community mapping population, was assessed by sequencing 588 gene fragments from each genotype (McConnell et al. 2007). SNPs were detected for 382 (65%) of the genes. Averaged over all of the genes, an SNP was observed every 375 nt. Of these, 42.9% were located in introns, 40.3% in exons, and 17.2% in the 3'-UTR. 65.1% of the SNPs were synonymous. Excluding one nucleotide indels, 22% of the genes contained indels. The vast majority (73.7%) were located in introns, with only 16.8% of the indels in the 3'-UTR.

Repeat structure. RFLP mapping experiments (Vallejos et al. 1992; Freyre et al. 1998) revealed that common bean duplicate loci are rare. McConnell et al. (2007) mapped 300 gene-based loci in common and discovered only 8 tandemly repeated genes and 5 that were repeated between linkage groups. In addition, BAC sequencing revealed that the well known APA gene family exists as tandem duplicates (Kami et al. 2006). Sequencing traditionally large gene families, such as resistance gene analogs (Rivkin et al. 1999) and protein kinases (Vallad et al. 2001) revealed they are moderate in size. Recently a search for duplicate genes among 2941 EST contigs and 7939 EST singletons detected only 222 pairs of duplicate genes (Mamidi et al. 2008).

Several reports describing retrotransposon sequences in common bean have been published (Galindo et al. 2004; Erdmann et al. 2002; Garber et al. 1999) and hybridization experiments (Erdmann et al. 2002) have shown these to represent only a small portion of the genome. One detailed sequence effort discovered short fragments that appear to be remnants of once active elements (Kami et al. 2006). A repeat library based on Recon analysis of 1.36 Mb of DNA is available at http://phaseolus.genomics.purdue.edu/data/pv_gba_recon_repeats.fasta. Four repeat families with 164, 87, 82 and 34 members/family were observed. One is AT rich and is possibly a mite, and a second contains 26S ribosomal RNA genes. The other two families are both gypsy-type retrotransposon families similar to the maize Cinful/Grande/Huck super-family and are estimated to comprise 10% of the bean genome.

III. Available Common Bean Genomic Resources

ESTs. 83,000 common bean ESTs are currently available (Ramirez et al. 2005; Melotto et al. 2005; Thibivilliers et al. unpublished). This EST set was analyzed, and 11,000 contigs and 9,000 singletons were discovered. Multiple *P. vulgaris* genotypes were used to obtain the EST data which in turn facilitates SNP discovery. EST data was collected from seedling shoots [with or without *Colletotrichum lindemuthianum* (anthracnose) infection], seedling leaves, nodules elicited by *Rhizobium tropici*, roots, leaves (three genotypes), and pods. In *P. coccineus*, ESTs were isolated from the suspensor regions in globular-stage embryos six days after pollination (e.g., GenBank: CA916678; http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=27403670). Because of the close relationship between the two species, sequences in *P. vulgaris* can be identified through similarity with *P. coccineus* (Nanni et al. 2005).

BAC libraries. Eleven BAC libraries are available in the genus *Phaseolus*, ten in *P. vulgaris* and one in *P. lunatus* (Gepts et al. 2007). Based on their average insert size, library coverage varies (5-12X). The BAT93 library has a coverage of 20X, in part because it has been designated as the standard genotype for *Phaseolus* genomics (Broughton et al. 2003). The *Phaseolus* BAC libraries are a phylogenetically ordered set useful for evolutionary studies (Gepts et al. 2007). DGD1962 is a wild bean from northern Peru, representing the presumed ancestral gene pool of the species (Debouck et al. 1993; Kami et al. 1995). The remainder of the libraries are representative of the two evolutionary gene pools. G02771 and G12946 are wild Mexican beans of Mesoamerican origin that contain the three subfamilies of the APA seed proteins, which confer resistance to seed weevils. G2333 is a Mexican landrace highly resistant to anthracnose. BAT93, OAC-HR45 and OAC-HR67 are breeding lines and OAC-Rex is a cultivar from the Mesoamerican gene pool. G19833 is an Andean landrace from Peru, whereas Sprite is a bred Andean variety. Using this array of BAC libraries it is possible to study *Phaseolus* genome evolution both prior to and after domestication and analyze phenotypic changes resulting from specific structural modification at the genome level. Single BAC clones have been fully sequenced, one around the *Co-4* locus for resistance to anthracnose (Melotto et al. 2004), and the other around the *APA* locus (Kami et al. 2006). Comparative sequence analysis of similar regions in soybean can address the question of the utility of common bean as a diploid model for soybean.

Genetic maps. Over 25 linkage maps, mostly low density (markers on average every 10 cM), have been developed for common bean (Kelly et al. 2003; Miklas et al. 2006). To maximize molecular polymorphism, the majority of mapping populations were derived from crosses between domesticated parents belonging to the Andean vs. Middle American gene pools. For specialized purposes, some maps were developed by crossing parents within a gene pool but polymorphism was low.

A highly polymorphic core map utilizing a recombinant inbred population from the cross BAT93 x Jalo EEP558 (Nodari et al. 1992) was developed to coalesce the mapping data (Freyre et al. 1998). BAT93 is a breeding line from the Mesoamerican gene pool, and Jalo EEP558 is an Andean cultivar resulting from selection in a Brazilian landrace. The two parents show contrasting resistances to pathogens. Some 600 markers have been mapped directly in this population, including 71 RFLPs, 161 AFLPs, 158 RAPDs, 50 ISSRs, and 200 microsatellites (Freyre et al. 1998; Papa and Gepts 2003; González et al. 2005; Blair et al. 2003; Grisi et al. 2007). Shared markers, principally RFLPs and sequence-tagged markers, are

used to correlate linkage groups among the different maps. Recently, this map was greatly extended by the addition of 300 gene-based markers (McConnell et al. 2007). These will be very useful to correlate the genetic map with the physical map.

Physical maps. Utilizing the G19833 BAC library, a common bean physical map was constructed using High Information Conent Fingerprinting and BAC end sequencing (41,717 BACs, ~9x clone coverage). This physical map assembled into 1,183 contigs (6,385 singletons) and has been anchored with more than 540 markers derived from RFLPs, genes, ESTs and other sequences. Eighty-four of these anchors are genetically mapped and provide linkage between the physical and genetic maps. The physical map is publicly available at http://phaseolus.genomics.purdue.edu/. Initial analysis of BAC end sequences (~62 Mbp) revealed that ~49% of the genome is repetitive and 29% genic.

Bioinformatics. As with other legumes, newly generated information is deposited in the Legume Information System (http://www.comparative-legumes.org/).

IV. Background to Sequencing the *P. vulgaris* Genome

The current mainstream approach to genome sequencing is the whole genome shotgun (WGS) method. Randomly sheared DNA is subcloned, and a combination of 2 kb and 6 kb plasmid and 35kb fosmid clones are sequenced from both ends using the standard Sanger technology. In addition, a 1X collection of BAC-end sequence (BES) read is also generated using the Sanger technique. Paired-end plasmid capillary sequence data forms the core of the assembly. The relatively small contigs that assembly are then merged using fosmid end and BES data to form larger scaffolds and superscaffolds. Finally, physical and genetic map data is used to merge superscaffolds into pseudochromosomes. This WGS approach was used for the soybean (Schmutz et al. 2008) and grape (FIPCGGC 2007; Velasco et al. 2007) genome sequencing projects.

Although duplicating this approach would seem logical, whole genome sequencing appears to be at a tipping point where new approaches, such as the 454 pyrosequencing method, may be offering a cost advantage along with increased speed and throughput. Importantly, long (400-500 pb) 454 reads have very recently become available, so it is important to consider what role that might play in future sequence projects. While the 454 approach is fast and will generate a large amount of data, the read lengths are similar to that used in the early stages of the human genome project. It was very difficult to assemble the genome with reads of these lengths, and assembly only began to become possible as read lengths increased (600-700 bases) and paired-end reads were introduced. Therefore, given the unknowns associated with assembling 454 data for a relatively moderate size genome such as *P. vulgaris*, it would be sensible to consider other alternatives.

One alternative would be to combine 454 reads with a modest Sanger sequencing effort. This hybrid approach is currently being tested, but it has only been successful for microorganisms (Goldberg et al. 2007). The major unresolved questions with the methodology are: 1) How much Sanger sequencing is necessary to ensure the 454 data will assemble? 2) What auxillary data is needed to provide a quality assembly that can be placed into pseudochromosomes? 3) Will the cost of a combined 454/Sanger sequencing approach be sufficiently lower to forego the well tested WGS method?

Current estimates from simulations suggest that combining a 16x 454 dataset with 2x Sanger coverage is not sufficient for assembly, and that assembly produces larger scaffolds

with 3X-4X Sanger coverage with a larger proportion of fosmids than normally used. While scaffolds can be assembled they are limited in size unless significant genetic marker data is available. Additionally, a significant BES data set will increase the size of the final scaffolds and help organize the pseudochromosomes.

Finally, it is important to note that the soybean genome sequence along with an ordered syntenic relationship between *G. max* and *P. vulgaris* should be very helpful in ordering and orienting *P. vulgaris* scaffolds. For example, the two-to-one relationship between *G. max* and *P. vulgaris* was used to "electronically" place 90 EST contigs within a 40 cM region of Pv7. A subset of these contigs were genetically mapped, and the genetic and electronic maps of the region were completely equivalent (Lee et al. in preparation).

V. Activities Required for an Ordered Genome Sequence of *P. vulgaris*

Mapping tools. In conjunction with the hybrid sequencing project, it will be necessary to increase the number of markers to aid assembly into pseudochromosomes. Gene-based markers developed by resequencing of selected loci based on EST sequence data is the preferred method. This approach is chosen because it will be possible to select an optimum set of loci based on the electronic map of ESTs. The goal of the gene-based mapping component is to add ~120 markers per linkage group for a total of an additional ~1400 markers. We estimate that this number of well-selected markers, together with excellent synteny with soybean, will be sufficient to order and orient most large scaffolds. (**Total cost**: $400,000 for sequencing, mapping, and personnel)

BAC-end sequencing. BAC-end sequences are critical for assembly. A critical component of the success for the soybean project was the availability of 380,000 BES from three different libraries. Given that the genome of *P. vulgaris* is half that of soybean, approximately half the number of sequences would be appropriate. Since we already have 80,000 sequences from one library, an additional set of 100,000 sequences from a separate library would be sufficient. (**Total cost**: $150,000)

454 sequencing. We are basing our estimate of sequence needs on the assumption of 400 bp 454 reads. This is at the low end of the current technology, and tests in a particpating lab have shown these to be accurate. In addition, we are assuming each run would generate one million 400 bp reads. For the 454 portion of the project, we would suggest that 22 runs would generate 15X genome coverage of the 600 Mb genome. In addition, another eight runs of various sizes would also be performed. [**Total costs:** $450,000 (based on the current price of $15,000/run which includes materials, labor, and amortization)].

Sanger sequencing. For the Sanger sequencing portion of the project, we propose a modest 4X plasmid end-shotgun sequence along with ~40X clone coverage in 35 kb fosmids (total: ~600,000 end-reads). The cost of the 4X coverage of a 600 Mb genome is $2.4 million, while the fosmid sequencing would be about $500,000. (**Total cost**: $2.9 million)

Assembly, annotation, and data archiving. We would propose to collaborate with one of the national sequencing centers for assembly, gene modeling, and annotation. Funds would be required for personnel and computational time. In addition, funds would be necessary to create a public archive of the data. Again, it would appropriate to use one of the current genome database infrastructure. [(**Total cost**: $650,000 (based on $500,000 and annotation; $150,000 archiving)]

VI. Applications of the Common Bean Sequence to Legume Research

The *Phaseolus* sequence will have a major impact on legume research. One area is crop improvement, a field that continually relies upon advances in genetics. The increasing volume of sequence information provides a wealth of information that will be used to: 1) increase the number of traditional markers such as simple sequence repeats (SSRs), and 2) support the implementation of high throughput markers such as single nucleotide polymorphisms (SNPs). The markers will significantly increase the accuracy of genetic prediction and in turn will increase the precision of marker assisted selection (MAS). The markers, along with sequence data, will be important resources for the implementation of population genetic approaches to discover genes essential for the plant's success in modern cropping systems. Finally, the sequence data will impact other genome projects. A few examples of how the common bean sequence will affect these endeavors follow.

Molecular marker development in common bean. Marker assisted selection (MAS) has a long history in common bean research that is second to no other plant species (Kelly et al. 2003, Miklas et al. 2006). Most of the markers are sequence tagged sites (STS) derived from RAPD markers. Recently, a small collection of SSR markers were developed (Blair et al. 2003) and are just now being used. Both of these markers classes are limited at times because of the lack of diversity within the various market classes (black, kidney, navy, pinto, snap, etc.) from which they are developed. Furthermore, given the complex genetics that controls seed coat color and pattern (McClean et al. 2002), the major factors along with seed size that define a market class, breeders are reluctant to make inter-market class crosses. Therefore, there is a need to expand the common bean molecular marker collection to broaden the impact of MAS.

The first step will be to scan the common bean sequence for new SSR markers. Techniques such as those described by Lawson and Zhang (2006) can uncover a great number of SSRs. For example, this procedure recognized 298,829 SSR markers in rice, in comparison to the 18,828 originally reported for the rice genome (IRSGP 2005). Similar data from the bean sequence will facilitate the development of additional SSR markers that will be of value to US bean research community.

Single nucleotide polymorphisms (SNPs) are the ultimate definition of genotypic variation. Recently McConnell et al. (2006) collected over 600 Kb of gene-based sequence data and discovered that common bean genome contains one SNP per 170 nt. Based on a genome size estimate of about 600 megabases, common bean would have about 3.6 million SNP. This data was utilized to develop a low density SNP markers map (McConnell et al. 2007; Lee et al., unpublished). The genome sequence data will be used as a reference to discover new SNPs by comparing it with the more than 83,000 EST that are available. This is possible because none of the EST datasets were developed using the genotype that will be sequenced (BAT93). The first application will be the development of a high density SNP map. This will involve targeted sequencing of a panel of phenotypically diverse parents to uncover gene diversity in exons, introns, and the 3'-UTR. SNP discovery will lead to new high throughput technologies that will accelerate the genetic improvement of common bean.

Comparative Candidate Gene Discovery in Legumes. The best marker for marker-assisted selection is the gene itself. Obviously this requires a concerted effort at positional cloning which is a major effort for all plant species. Yet, with a whole genome sequence, gene

models predicted from the sequence, and extensive mapping resources, it is possible to place a gene within a small interval which contains only tens of genes. Genetically, these genes, when used as linked markers, will either cosegregate in typical breeding population, or map close enough for selection purposes. Remember, the most efficient use of markers is during early generation testing when the breeder wants to eliminate any potential phenotypic off types. Later in the breeding process, the breeder will practice traditional phenotypic selection to ensure the MAS did not carry through an unwanted recombinant.

The key to defining this interval of interest is fine-mapping. Geneticists are now creating larger population to define linkages that are less than 1 cM for major qualitative genes, or to narrow the confidence interval for quantitative trait loci (QTL). With this fine detailed mapping, it will then be possible to discover candidate genes for qualitative traits in legumes using the approach described above for the *Dt1* locus in soybean.

For quantitative traits, a comparative candidate gene approach using common phenotypic regions is also amenable. A case in point is the region in common bean on Pv7 that contains major QTL for both common bacterial blight and white mold [(for a comprehensive review of QTL for these and other disease loci see Miklas et al. (2006)]. The region on Pv7 to which these QTL map is syntentic with soybean linkage group O (Fig. 2). A scan of corresponding gene models syntentic to the 40 cM common bean interval that contains these QTL revealed leucine-rich repeat gene models reminiscent of classic disease resistance genes. Given that any common bean resistance gene analog is more similar to a soybean sequence than to any common bean paralog (Rivkin et al. 1999) suggests a direct lineage of these resistance genes. Disease resistance analogs are also known to map in this region which suggests a function for these sequences (Mutlu et al 2006). Therefore these soybean genes are potential candidate genes. This comparative approach will be that much more powerful with the ability to use gene models derived from common bean itself.

Identifying Adaptation Genes within the Phaseoleae. Recently, random genome-wide scans identified genes positively associated with domestication and/or agronomic productivity (Wright et al. 2005; Yamasaki et al. 2005). These scans were costly because they did not have *a priori* information regarding genes that might be undergoing the selection. This necessitated a large scale resequencing of over 1,000 genes from multiple genotypes. Utilizing pairwise sequence data obtained from comparing common bean and soybean, it may be possible to narrow the candidates necessary for this discovery process. Once these domestication or agronomic productivity genes are discovered, breeding populations developed by introgressing wild germplasm with beneficial alleles into an improved variety can be screened using high-throughput technologies to select lines containing a high proportion of essential domestication, agronomic and improvement alleles. An outline of how to leverage soybean and common bean sequences to identify these adaptation loci follows.

By comparing the coding sequence between common bean and soybean orthologous genes, it may be possible to discover genes undergoing purifying or positive selection within the Phaseoleae lineage. The classical method to detect these two forms of selection is to measure the K_a/K_s ratio. K_a is the non-synonymous substitution rate, and K_s is the synonymous substitution rate. If the $K_a/K_s << 1.0$, then the gene is assumed to be undergoing purifying selection to eliminate deleterious mutations, whereas if the K_a/K_s ratio is $>> 1.0$, then the gene is considered to be undergoing positive, and possibly adaptive, selection. Key to these comparisons is to ensure that orthologs are compared. The syntenic map described above

provides a reference point from which orthologs between common bean and soybean orthologs can be identified.

Orthologous genes found to have undergone selection can then be compared with a IRLC species such as *M. truncatula*, *L. japonicus*, or *P. sativum*. If a comparison at this level does not indicate selection, then the significant K_a/K_s ratio for the common bean/soybean comparison would mean that the gene may be important in the evolution of the Phaeoleae lineage. Likewise, if the comparison to one of the IRLC genes is still significant, a similar comparison to an ortholog from an outgroup would allow us to determine if it is important to the legume lineage.

Genes undergoing positive selection deserve a more detailed analysis because they may encode functional changes that drove evolution of a specific taxonomic lineage. With genome-wide sequence data for legumes, similar calculations can be performed to identify genes important to a specific lineage. Additionally, a sliding-window calculation of the K_a/K_s ratio across the gene can identify specific regions of the gene that were strongly affected by selection (Choi and Lahn 2003).

The polymorphism data generated by these K_a/K_s studies can then be used in a manner described by Wright et al. (2005) and Yamasaki et al. (2005). The advantage, though, is that this genome-wide K_a/K_s survey will act as a prefilter to identify candidate genes in the common bean/soybean lineage. For example, a genome-wide scan identified 13,454 human-chimpanzee orthologs, of which 585 had a K_a/K_s ratio greater than 1 (CSAC 2005). These are logical candidate genes for further studies of adaptation in that lineage. By comparison, Yamasaki et al. (2005) prescreened 1095 sequences and identified eight candidates for maize adaptation. Using rice as a reference for the number of genes (IRGSP 2005), this prescreen only considered 3% of the genes. Clearly, it would be more efficient to use a genome-wide approach than random searches through a subset of the genome to select genes to study adaptation. To apply this approach to the study of adaptation to these socially (bean) and economically important (soybean) legumes, we simply need the resources to collect common bean sequence to the same depth as soybean.

Applications to other legumes. Cowpea (*Vigna unguiculata*), a sister species to common bean within the Phaseoleae (Fig. 1), is a critical species in the diets of developing countries, especially in Africa (Singh 2005). Although its world wide production (4 million metric tons) trails common bean (18 million metric tons) as an edible legume (FAO: http://faostat.fao.org/), it is an important dietary component in Western Africa. The two species exhibit a high degree of synteny and their molecular probes can be readily be shared (Boutin et al. 1995; Choi et al. 2004). This is mostly likely a result of the recent shared ancestry (8.0 MYA; Lavin et al. 2005). This divergence time is similar to that found between the closely related human and chimpanzee genomes (CSAC 2005). The fact that genomic tools can be shared between these two species suggests the same should also be true for cowpea and common bean. First, as demonstrated in cereals (Zhang et al. 2005; Bossolini et al. 2006), SSR loci in common bean could serve as a source for markers in cowpea. The common bean genome database can be searched for sequences homologous to cowpea markers linked to a trait of interest. The established synteny relationship will ensure the match is to the correct bean chromosome. Then the cowpea researcher can walk in either direction and develop PCR primers more tightly linked to their trait of interest. As further resources are developed, these tightly linked markers can then be used to study the genome organization of cowpea candidate

genes as has been down in the cereals (Colasanti et al. 2006; Quarrie et al. 2006; Singh et al. 2006; Faure et al. 2007) and crucifers (Mayerhofer et al. 2005; Parkin et al. 2005).

The sequencing of *M. truncatula* is well underway and its completion is in sight (Young et al. 2005). As is the case with common bean, *M. truncatula* also contains a simple diploid genome. These two species last shared a common ancestor some 54 MYA. An interesting question is the degree to which each of these two species can share genomic resources. For example, it will be interesting to determine the extent of microsynteny between the two species. By combining the sequence data from Phaseoleae (common bean, soybean), with that of the Loteae (*Lotus*) and Trifoleae (*M. truncatula*) we will be able to develop an in depth analysis of legume evolution and compare its pattern with that of other lineages (Brassicaceae, for example) to see if general trends can be discovered that underlie angiosperm evolution. From a more practical genomic perspective, the common bean and *Medicago* sequence should be useful in defining legume gene models that will assist with annotation.

VII. Project Support

This white paper is a community-wide effort with input and support from many individuals. There names and affiliations follow.

Name	**Institution**
Kirstin Bett	University of Saskatchewan
Fred Bliss	Seminis
Mark Brick	Colorado State University
William Broughton	Université de Genève, Switzerland
Tim Courneya	Northarvest
Perry Cregan	USDA/ARS, Beltsville, MD
Jeff Doyle	Cornell University
Rubella Goswami	North Dakota State University
Tom Grebb	National Dry Bean Council and Central Bean Company
Georgina Hernández	Centro de Ciencias Genómicas, UNAM, México
Khwaja Hossain	Mayville State Universityl
Venu Kalavacharla	Delaware State University
Hans Kandel	North Dakota State University
Jim Kelly	Michigan State University
Miguel Lara	Centro de Ciencias Genómicas, UNAM, México
Matt Lavin	Montana State University
Phil Miklas	USDA/ARS, Prosser, WA
Jim Myers	Oregon State University
Jim Nienhuis	University of Wisconsin
Tom Osborne	Seminis
Juan Osorno	North Dakota State University
Roberto Papa	University of Ancona, Italy
Tim Porch	USDA/ARS, Puerto Rico
Carlos Urrea	University of Nebraska
Federico Sanchez	Centro de Investigación sobre Ingenieria Genética y Biotecnología, UNAM, Mexico

Howard Schwartz	Colorado State University
Randy Shoemaker	USDA/ARS, Ames, IA
Gary Stacey	University of Missourii
Jim Steadman	University of Nebraska
Jim Specht	University of Nebraska
Alyson Thornton	Harris Moran Seed
Molly Welsh	USDA/ARS, Pullman, WA

VIII. References

ADA (2004) Position of the American Dietetic Association and dietitians of Canada: nutrition intervention in care of persons with human immunodeficiency virus infection. JADA 104: 1425-1441

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Pl Mol Biol Rept 3:208-218

Baum MK, Shor-Posner G, Lu Y, Rosner B, Sauberlich HE, Fletcher MA, Szapocznik J, Eisdorfer C, Buring JE, Hennekens CH (1995) Micronutrients and HIV-1 disease progression. AIDS 9:1051-1056

Baxter,R. Kirk JTO (1969) Base composition from chloroplasts and nuclei of *Phaseolus vulgaris*. Nature 222:272-273

Beaver JS, Osorno JM, Ferwerda FH, Munoz Perea CG (2005) Registration of bean golden yellow mosaic virus resistant germplasm PR9771-3-1, PR0247-49 and PR0157-4-1. Crop Sci 45:2126-2127

Bennett M, Leitch I (2005) Plant DNA C-Values database. Release 4.0 http://www.rbgkew.org.uk/cval/database1.html

Blair MW, Pedraza F, Buendia HF, Gaitán-Solís E, Beebe SE, Gepts P, Tohme J (2003) Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.) Theor Appl Genet 107:1362-1374

Bossolini E, Drattinger SG, Keller B (2006) Development of simple sequence repeat markers specific for the Lr34 resistance region of wheat using sequence information from rice and *Aegilops tauschii*. Theor Appl Genet 113:1049-1062

Boutin SR, Young ND, Olson TC, Yu Z-H, Shoemaker RC, Vallejos CE (1995) Genome conservation among three legume genera detected with DNA markers. Genome 38:928-937.

Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.) – model food legumes. Plant and Soil 252:55-128

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

Choi SS, Lahn BT (2003) Adaptive evolution of MRG, a neuron-specific gene family implicated in nociception. Genome Res 13:2252-2259

Choi HK, Mun JH, Kin DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR. (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci, USA 101:15289-15294

Colasanti J, Tremblay R, Wong AY, Conveva V, Kozaki A, Mable BK (2006) The maize INDETERMINATE1 flowering time regulator defines a highly conserved zinc finger protein family in higher plants. BMC Genomics 19:7:158

CSAC (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69-87

Debouck DG, Toro O, Paredes OM, Johnson WC, Gepts P (1993) Genetic diversity and ecological distribution of *Phaseolus vulgaris* in northwestern South America. Econ. Bot. 47:408-423

Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Phys. 131;900-910

Erdmann PM, Lee RK, Bassett MJ, McClean PE (2002) A molecular marker tightly linked to *P*, a gene required for flower and seedcoat color in common bean (*Phaseolus vulgaris* L.) contains the *Ty3*/gypsy retrotransposon *Tpv3g*. Genome 45:728-736

Faure S, Higgins J, Turner AS, Laurie DA (2007) The FLOWERING LOCUS T-like gene family in barley (*Hordeum vulgare*). Genetics 176:599-609

F-IPCGGC (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 440:463-467

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

Galindo LM, Gaitan-Solis E, Baccam P, Tohme J (2004) Isolation and characterization of RNase LTR sequences of Ty1-*copia* retrotransposons in common bean (*Phaseolus vulgaris* L.). Genome 47:84-95

Garber K, Bilic I, Pusch O, Tohme J, Bachmair D (1999) The Tpv2 family of retrotransposons of *Phaseolus vulgaris*: structure, integration characteristics, and use for genotype classification. Pl Mol Biol 39:797-807

Gepts P, Aragão F, de Barros E, Blair MW, Brondani R, BroughtonWJ, Galasso I, Hernández G, Kami J, Lariguet P, McClean, P, Melotto M, Miklas P, Pauls P, Pedrosa-Harand A, Porch T, Sánchez F, Sparvoli F, Yu K (2008) Genomics of *Phaseolus* beans, a major source of dietary protein and micronutrients in the Tropics. *In*: P.H. Moore and R. Ming (Eds), *Genomics of Tropical Crop Plants*. Springer, Berlin, pp 113-143

Gillespie S, Kadiyala S (2005) HIV/AIDS and food and nutrition security: from evidence to action. International Food Policy Research Institute, Washington, DC

Goldberg SM, Johnson J, Busam D, Feldblyum, Ferriera S, Friedman R, Halpern A, Khouri H, Kravitz SA, Lauro FM, Li K, Rogers Y-H, Strausberg R, Sutton G, Tallin L, Thomas T, Venter E, Frazier M, Venter JC (2006) A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. Proc Natl Acad Sci USA 103:11240-11245

González,A, Wong A, Delgado-Salinas A, Papa R, Gepts P (2005) Assessment of inter simple sequence repeat markers to differentiate sympatric wild and domesticated populations of common bean (*Phaseolus vulgaris* L.). Crop Sci 35:606-615

Grisi MC, Blair MW, Gepts P, Brondani C, Pereira PA, Brondani RP (2007) Genetic mapping of a new set of microsatellite makers in a reference common bean (*Phaseolus vulgaris*) population BAT93 x Jalo EEP558. Genet Mol Res 6: 691-706

IRSGP (2005) The map-based sequence of the rice genome. Nature 436, 793-800

Kami J, Becerra Velásquez B, Debouck DG, Gepts P (1995) Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. Proc Natl Acad.Sci U.S.A 92:1101-1104

Kami J, Poncet V, Geffroy V, Gepts P (2006) Development of four phylogenetically-arrayed BAC libraries and sequence of the APA locus in *Phaseolus vulgaris*. Theor Appl Genet 112:987 - 998

Kelly JD, Hosfield GL, Varner GV, Uebersax MA, Haley SD, Taylor J (1994) Registration of 'Raven' black bean. Crop Sci 34:1406-1407

Kelly JD, Gepts P, Miklas PN, Coyne DP (2003) Tagging and mapping of genes and QTL and molecular marker-assisted selection for traits of economic importance in bean and cowpea. Field Crops Res 82:135–154

Kruzich LA, Marquis GS, Carriguiry AL, Wilson CM, Stephensen CB (2004) US youths in the early stages of HIV diseas hae low intakes of some micronutrients important for optimal immune function. J Am Diet Assoc 104, 1095-1101

Kwak M, Velasco,D, Gepts P (2008) Mapping homologous sequences for determinancy and photoperiod sensistivity in common bean (*Phaseolus vulgaris*). J Hered 99:283-291

Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biology, 7:RI4

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. System Biol 54:530-549

Lee JM, Grant D, Vallejos CE, Shoemaker RC (2001) Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resole genome evolution events among legumes. Theo. Appl Genet 103:765-773

Lindgren DT, Eskridge KM, Steadman JR, Schaaf DM (1995) A model for dry bean yield loss due to rust. HortTechnology 5:35-37

Mamidi S, Lee RK, Terpstra J, Schlueter JA, Dixon P, Shoemaker RC, Lavin M, McClean PE (2008) Investigating gene duplications events in legumes using EST sequence data. Plant and Animal Genome XVI, page xx

McClean PE, Lavin M, Gepts P, Jackson,SA (2008) *Phaseolus vulgaris*: a diploid model for soybean. *In* G. Stacey (Eds), *Soybean Genomics*. Springer, Berlin,pp 55-78

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

McConnell MD, Mamidi S, Lee RK, McClean PE (2006) DNA sequence polymorphisms among common bean genes. Annual Report of Bean Improvement Cooperative 49, 143-144

McConnell MD, Mamidi S, Rossi M, Lee RK, and McClean PE (2007) A gene-based linkage map of common bean (*Phaseolus vulgaris* L.). Plant and Animal Genome XV Abstracts. p. xx

Melotto M, Coelho MF, Pedrosa-Harand A, Kelly JD, Camargo LEA (2004) The anthracnose resistance locus *Co-4* of common bean is located on chromosome 3 and contains putative disease resistance-related genes. Theor Appl Genet 109:690-699

Melotto M, Monteiro-Vitorello CB, Bruschi AG, Camargo LE (2005) Comparative bioinformatic analysis of genes expressed in common bean (*Phaseolus vulgaris* L.) seedlings. Genome 48:562-570

Miklas PN, Kelly JD (2002) Registration of two cranberry bean germplasm lines resistant to bean common mosaic and necrosis potyviruses: USCR-7 and USCR-9. Crop Sci 42:673-674

Miklas PN, Kelly JD, Singh SP (2003) Registratioin of anthracnose-resistant pinto bean germplasm line USPT-ANT-1. Crop Sci 43:1889-1890
Miklas PN, Kelly JD, Beebe SE, Blair MW (2006) Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. Euphytica 147:106-131
Miklas PN, Smith JR, Singh SP (2006b) Registration of common bacterial blight resistant dark red kidney bean germplasm line USDK-CBB-15. Crop Sci 46:1005-1007
Mutlu N, Miklas PM, Coyne DP (2006) Resistance gene analog polymorphism (RGAP) markers co-localize with disease resistance genes and QTL in common bean. Mol Breed 17:127-135
Mutlu N, Miklas PN, Steadman JR, Vidaver AV, Lindgren D, Reiser J, Pastor-Corrales MA (2005) Registration of ponto bean germplasm line ABCP-8 with resistance to common bacterial blight. Crop Sci 45:806
Nanni L, Losa A, Bellucci E, Kater M, Gepts P, Papa R (2005) Identification and molecular diversity of a genomic sequence similar to SHATTERPROOF (SHP1) in Phaseolus vulgaris L. Plant and Animal Genome XV Abstracts. p. xx
Nodari RO, Koinange EMK, Kelly JD, Gepts P (1992) Towards an integrated linkage map of common bean. I. Development of genomic DNA probes and levels of restriction fragment length polymorphism. Theor Appl Genet 84:186-192
Papa R, Gepts P (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. Theor Appl Genet 106:239-250
Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171:765-781
Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in relation to taxon divergence: a phylogentic analysis in legumes using 39 gene families. Sys. Biol 54:441-454
Quarrie S, Pekic Quarrie S, Radosevic R, Rancic D, Kaminska A, Barnes JD, Leverington M, Ceoloni C, Dodiq D (2006) Dissecting a wheat QTL for yield present ina range of environments: from the QTL to candidate genes. J Exp Bot 57:2627-2637
Raganin VA, Sanglard DA, de Souza TLPO, Moreira MA, de Barros EG (2003) Simultaneous transfer of resistance genes for rust, anthracnose, and anglular leaf stot to cultivar Perola assisted by molecular markers. Annu Rev Bean Improv Coop 46:159-160
Ramirez M, Graham MA, Blanco-Lopez L, Silvente S, Medrano-Soto A, Blair MW, Hernandez G, Vance CP, Lara,M. (2005) Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. Pl Physio. 137, 1211-1227
Rivkin M., Vallejos CE, McClean PE (1999) Disease-resistance related sequences in common bean. Genome 42, 41-47
Salamov AA, Solovyev VV (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. Genome Res 10:516-522
Saettler AW (1989) Common bacterial blight, in HF Schwartz and MA Pastor-Corrales (eds) Bean Production Problems in the Tropics, CIAT, Cali, Colombia, pp 261-284
Savarino A, Pescarmona GP, Boelaert JR (1999) Iron metabolishm and HIV infection: reciprocal interactions with potentially harmful consequences? Cell Biochem Funct 17:279-287

Schlueter JA, Dixon MP, Granger C, Shoemaker RC (2003) Mining the EST databases to resolve evolutionary events in major crop species. Genome 47, 868-876

Schmutz JJ, SGSC (2008) Progress on sequencing the soybean genome. Plant and Animal Genome XVI, page xx

Schwartz HF, Steadman JR, Hall R, Forster RL. 2005 Compendium of Bean Diseases, 2nd edition, APS Press, St. Paul, MN. Shoemaker, R.C., Polzin, K., Labate, J., Specht, J., Brummer, E.C., Olson, T., Young, N., Concibido, V., Wilcox, J., Tamulonis, J.P., Kochert, G., Boerma, H.R. (1996) Genome duplication in soybean (*Glycine* subgenus *Soja*). Genetics 144, 329-338

Singh BB (2005) Cowpea [*Vigna unguiculata* (L.) Walp. In: R.J. Singh. and P.P. Jauhar (Eds.), *Genetic Resources, Chromosome Engineering and Crop Improvement. Volume 1,* CRC Press, Boca Raton, FL, USA, pp. 117-162

Singh NK, Dalal V, Batra K, Singh BK, Chitra G., et al. (2006) Single-copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes. Funct. Integr. Genomics 7, 17-35

Stavely JR (1998) Recombination of two major dominnat rust resistance genes that are tightly linked in repulsion. Annu Rev Bean Improv Coop 41:17-18

Stavely JR, Pastor-Corrales MA (1989) Rust, in HF Schwartz and MA Pastor-Corrales (eds) Bean Production Problems in the Tropics, CIAT, Cali, Colombia, pp 159-194

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

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, et al. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One 12:e1326

Wong JH, Zhang XQ, Wang HX, Ng TB (2006) A mitogenic defensin from white cloud bean (*Phaseolus vulgaris*). Peptides 27, 2075-2081

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005). The effects of artificial selection on the maize genome. Science 308:1310-1314

Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley,JF, Gau, BS, McMullen MD (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell 17:2859-2872

Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. Theor. Appl Genet 111:677-687