

## Chapter 4

# *Phaseolus vulgaris*: A Diploid Model for Soybean

Phillip E. McClean, Matt Lavin, Paul Gepts, and Scott A. Jackson

## The Importance and Structure of Modern Common Bean Germplasm

### *Social, Economic, and Agronomic Importance of Common Bean*

Along with maize and cassava, common bean is a critical component of diets for many of the developing countries in the world. Beans are an important source of family income and a critical component of the daily diet within African countries where the population is projected to double by 2020. As such, it is the most important edible food legume in the world's diet. It represents 50% of the grain legumes consumed worldwide, and its production is nearly twice that of chickpeas, the second most consumed food legume. Because poverty limits access to animal protein in developing countries, these peoples turn to common bean as a protein source. From a dietary perspective, it accounts for 40%, 31%, and 15% of the daily intake of total protein in some of the least developed countries, such as Burundi, Rwanda, and Uganda, respectively. And even for a major producer like Brazil, 9% of the dietary protein is provided beans.

Common bean is grown in monocultures or as the primary or secondary species in a multicropping system. While the cash value of the crop exceeds \$1 billion in the United States, in many ways it is more important elsewhere. For a very poor country such as Myanmar, bean is its most important agricultural export commodity accounting for 10% of their total export income (<http://faostat.fao.org/faostat/>; verified July 14, 2006). Yield varies significantly, from 638 kg/ha, 671 kg/ha, and 918 kg/ha in Uganda, Rwanda and Burundi, respectively, to 1,944 kg/ha in developed countries. Improving agricultural productivity in Africa and the developing Americas is seen as a means to reverse the trend of increasing poverty and hunger in these regions. Therefore, identifying and minimizing yield limiting factors is an on-going concern for many bean improvement programs. For example, aluminum

---

P.E. McClean

Department of Plant Sciences, North Dakota State University, Fargo, ND 58105, USA  
e-mail: [pPhillip.mcclean@ndsu.edu](mailto:pPhillip.mcclean@ndsu.edu)

toxicity and phosphate deficiency affect yields in both the Americas and Africa and as such improvement programs are focusing on these traits (Beebe et al. 2006; Ochoa et al. 2006). In addition, given the prevalence of bean in the diets of these countries, modifying the nutrient content of common bean to make it a more balanced and nutritious food source is also receiving emphasis. Specifically, zinc content is a focus because of the recognition that this mineral is an important dietary component for those individuals infected with the AIDS virus. Understanding the genetic and genomic aspects of these traits and developing tools for their improvement will accelerate improvement of this important societal crop.

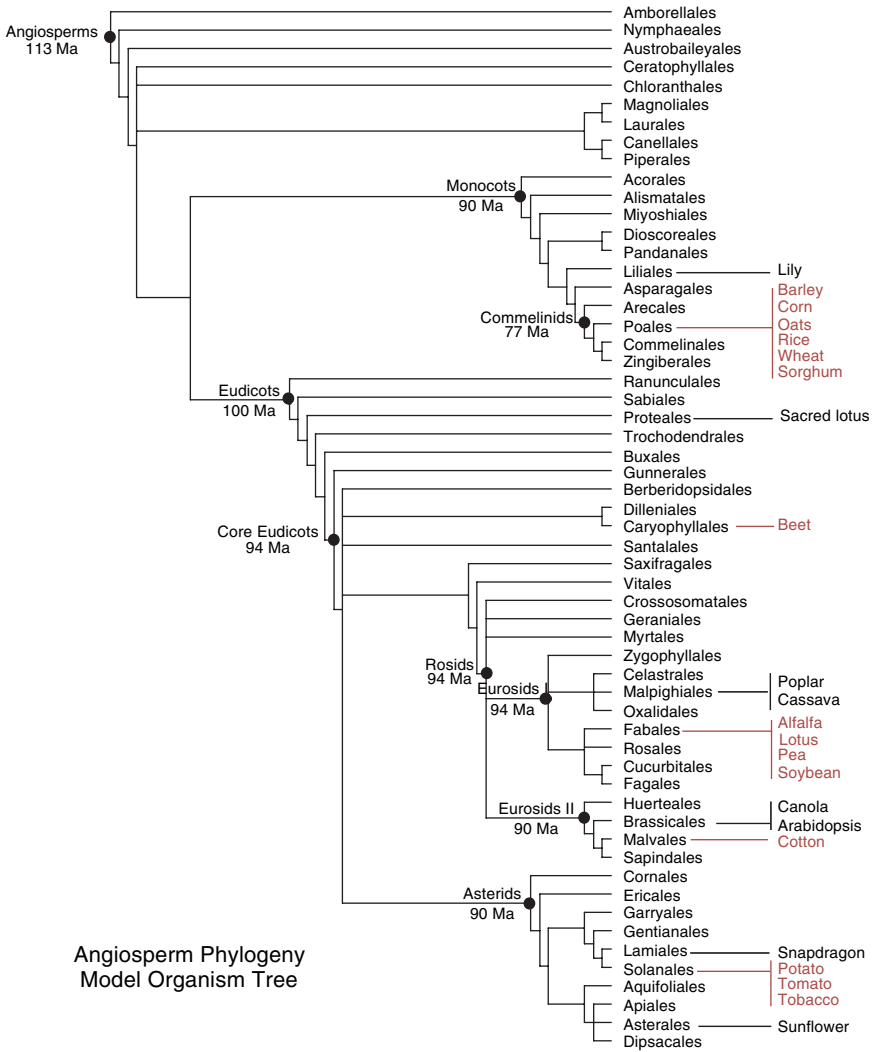
### ***Population Structure of Domesticated Common Bean***

The major subdivisions of wild common bean progenitors are known, and the domesticated gene pools have been defined. Based on phaseolin seed storage protein variation (Gepts and Bliss 1986; Gepts 1990), marker diversity (Becerra Velásquez and Gepts 1994; Koenig and Gepts 1989; Tohme et al. 1996; McClean et al. 2004b), and morphology (Gepts and Debouck 1991), two major gene pools of wild common bean were identified. The Middle American gene pool extends from Mexico through Central America and into Colombia and Venezuela, while the Andean gene pool is found in southern Peru, Chile, Bolivia and Argentina. The two gene pools appear to converge in Colombia (Gepts and Bliss 1986). A third, possibly ancestral gene pool based in southern Ecuador and northern Peru, was described (Debouck et al. 1993; Kami et al. 1995). Two major domestication events appear to have resulted in the Middle American and Andean gene pools (Kaplan and Lynch 1999) that mirror the geographic distribution of the wild progenitors (Gepts 1998; Islam et al. 2002; Blair et al. 2006). These domestication events appear to have arisen from wild beans in the region where the domestication occurred. This domestication event is in contrast to species such as maize (Matsuoka et al. 2002) and soybean (Powell et al. 1996; Xu and Gai 2003) where a derived domesticated population emerged from a single subspecies or related species, respectively. Following domestication, gene pool divergence led to the appearance of three races within the Andean gene pool, Nueva Granada, Peru and Chile (Singh et al. 1991), and four Middle American races, Durango, Jalisco, Mesoamerica, and Guatemala (Singh et al. 1991; Beebe et al. 2000).

### **Phylogenetic Evolution of Legumes and the Relationship Between *Glycine max* and *Phaseolus vulgaris***

#### ***Phylogenetic Relationship of Legumes to Other Angiosperms***

The legume family is well supported by many recent molecular phylogenetic studies as a member of the Eurosoid I clade of dicot flowering plants (Fig. 4.1). Although



**Fig. 4.1** Synopsis of the model organism tree adapted from the Angiosperm Phylogeny web site (Stevens 2001; [www.mobot.org/MOBOT/research/APweb](http://www.mobot.org/MOBOT/research/APweb)). The minimum ages of marked clades in millions of years (Ma) are taken from Crepet et al. (2004), except for the minimum age of the Commelinid clade, which is taken from Magallón and Sanderson (2001). Crepet et al. (2004) take a conservative approach to assigning fossils to extant clades only with unequivocal morphological evidence. Some clades deeper in the phylogeny have equivalent minimum ages to those nested higher in the phylogeny. This is because the fossils used to make these age assignments come from strata with similar minimum ages (*See* also Color Insert)

most of these Eurosid families have an exceptional fossil record, no obvious morphological features characterize them (Crepet et al. 2004). The minimum age of the Eurosid I clade is 94 Ma (Crepet et al. 2004), which is significantly older than the age of the origin of legumes, estimated at about 60 Ma (summarized in Lavin et al. 2005).

Legumes belong to the Fabales, which comprise four families, 754 genera, and 20,055 species (Stevens 2001), or nearly 10% of the Eudicot species diversity (Magallón et al. 1999). Strongly zygomorphic or bilaterally symmetric flowers first appear during the early Tertiary by about 60 Ma in age with the appearance of the Fabales. Wikström et al. (2001) estimated the age of origin of Fabales at 94-89 Ma, and the extant diversification at 79-74 Ma. These age estimates are biased old (Lavin et al. 2005). Regardless, Fabales is an unanticipated group in not being characterized by morphology even though quite strongly supported by molecular phylogenetic analysis (summarized in Stevens 2001). The constituent families of Fabales, as reviewed by Lewis et al. (2005) and Stevens (2001), include the Polygalaceae with 18 genera and 1,045 species having a world wide distribution. Surianiaceae with 5 genera and 8 species are mostly Australian, but *Recchia* is endemic to Mexico and *Suriana maritima* is pantropical. Quillajaceae with one genus (*Quillaja*) and 3 species is endemic to temperate Chile. Fabaceae with 730 genera and 19,400 species (13,855 of which belong to the papilionoid legume clade) is world wide in distribution but with a predilection to deserts, grasslands, savannas, and seasonally dry tropical forests, although certain groups are diverse in tropical wet forests. The relationships among these four families of Fabales are not well supported (Stevens 2001). Polygalaceae and legumes are the only two families of Fabales that produce bilaterally symmetric flowers, but this sister relationship is otherwise tenuously or not at all resolved with molecular phylogenetic analyses. For example, the *rpl22* gene was transferred from the chloroplast to the nucleus in Polygalaceae and Fabaceae, but this condition has not been analyzed for any Quillajaceae and Surianiaceae, as well as many members of Fabaceae. Many papilionoid legumes have lost the *rps16* gene, which is also known to be absent from *Polygala*. Stevens (2001), however, indicates a lack of sampling for these traits from most Fabales.

### ***Phylogenetic Relationships of Legumes to Model Plant Genome Species***

Closely related to the Fabaceae within the Eurosid I angiosperm group are many other economically important plant groups, particularly the Rosales (including *Fragaria*, *Malus*, *Prunus*, *Rubus*, etc.), Cucurbitales (*Citrullus*, *Cucumis*, *Cucurbita*, *Luffa*, etc.), and Fagales (*Fagus*, *Quercus*, *Juglans*, etc.; Fig. 4.1). More distantly related to legumes, yet within this same Eurosid I clade, is the Malpighiales, which includes cassava (Euphorbiaceae) and poplar (Salicaceae). *Arabidopsis* and canola (Brassicaceae), as well as cotton (Malvaceae), belong to the Eurosid II clade that is, in part, sister to the Eurosid I clade. These collective Rosids have a minimum age

of 94 Ma, according to the best fossil evidence (Crepet et al. 2004). As with most molecular phylogenetic analyses, these groups are well supported but not necessarily by obvious morphological characters. Eurosoid I constituents, for example, show a predisposition to fix nitrogen via root-dwelling associates (e.g., actinomycetes, rhizobia, etc.), but this is certainly not a uniform attribute of this clade (Stevens 2001). Legumes are much more distantly related to other model plant groups, such as those that belong to the Asterids (e.g., snapdragon, potato, tomato, tobacco, and sunflower), Caryophyllales (e.g. beet), or Monocots (e.g., lily), including the Comelinids (e.g., the grasses; Fig. 4.1).

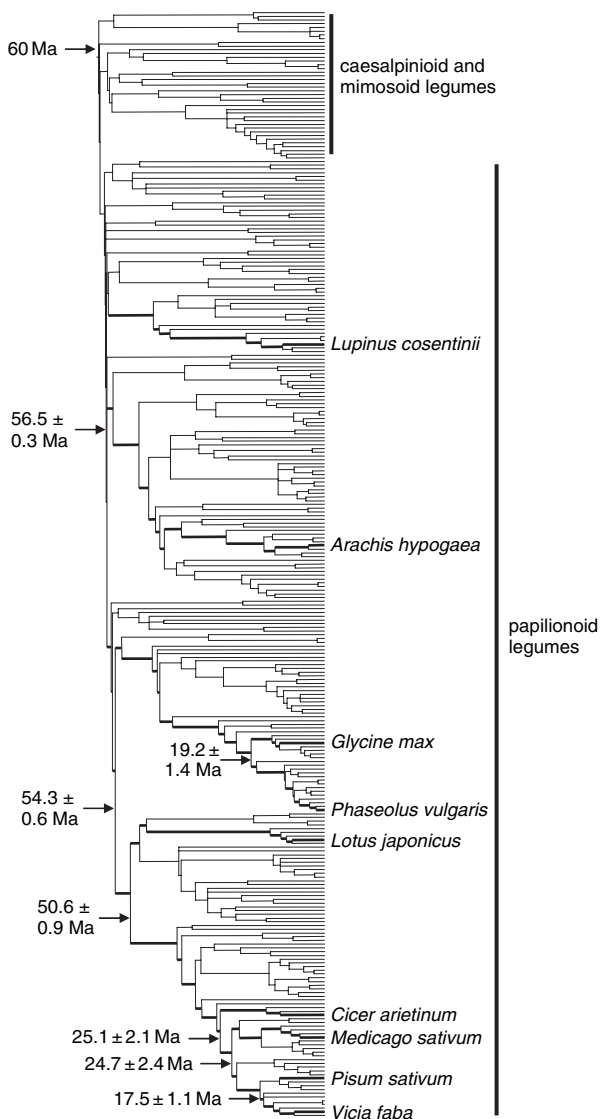
### ***Phylogenetic Relationships of Glycine and Phaseolus to Model Legumes***

Although many economically important species come from the legume family, a small sample of model legume species are illustrated to emphasize the great age that separates the many model and economically important legume species (Fig. 4.2). For example, tracing backwards in time along the genealogical paths that lead to the most recent common ancestor of *Lupinus* or *Arachis* with any of the other domesticated or model legume species (e.g., *G. max* or *Pisum sativum*) reveals that these species evolved along separate lines for at least 56.5 Ma. The early Tertiary age of these lineages is coeval with the Paleocene-Eocene boundary, or about the same time that flowering plants were beginning to dominate terrestrial vegetation. Even seemingly closely related pairs of model legume species, such as *G. max* and *P. vulgaris*, or *P. sativum* and *Vicia faba*, evolved along separate lines for well over 15 Ma. As reference for emphasizing the magnitude of these ages, humans and chimpanzee have been evolving along separate lines for at most 7 Ma (Kumar et al. 2005). In spite of the nearly 20 Ma age that separates *G. max* from *P. vulgaris*, these two species belong to the same group of papilionoid legumes, which is referred to as the core Phaseoleae (Lewis et al. 2005).

### ***Phaseolus Genomic Tools***

#### ***Cytogenetics of Phaseolus***

Common bean is a diploid species with  $2n = 22$  chromosomes and a medium-sized genome. Estimates of the size of the haploid genome of *P. vulgaris* range from 588 (Bennett and Leitch 1995, 2005) to 637 Mbp (Arumuganathan and Earle 1991). Most of the species of the genus *Phaseolus* have the same chromosome number with the exception of the *P. leptostachyus* Benth. clade, which has  $2n = 20$  chromosomes (Delgado-Salinas et al. 2006). With some exceptions, including soybean, most other members of the Phaseoleae tribe have the same or a similar chromosome number around  $2n = 22$ . These include, in the genus *Vigna*, which is closely related to the



**Fig. 4.2** A phylogeny with branch lengths scaled to time (a chronogram) derived from the analysis of Lavin et al. (2005) in which selected crop species are shown and the lineages leading to these highlighted. Tracing two of these highlighted lineages back in time to where they intersect or coalesce reveals the age of their most recent common ancestor (MRCA). For example, the age of the MRCA of *Glycine* and *Phaseolus* is about 19 Ma, whereas that of *Arachis* and any of the other species shown is about 56.5 Ma. This illustrates that many of the cultivated or model legume species are deeply divergent in time from one another. The minimum age of legumes, 60 Ma, is fixed for the root of the legume clade, and the evidence for this is summarized in Lavin et al. (2005)

genus *Phaseolus*, cowpea [*Vigna unguiculata* (L.) Walp], mung bean [*Vigna radiata* (L.) R. Wilczek], and rice bean [*Vigna umbellata* (Thunb.) Ohwi & H. Ohashi]. In pigeon pea [*Cajanus cajan* (L.) Millsp.], the chromosome number is also  $2n = 22$ , as well as in hyacinth bean [*Lablab purpureus* (L.) Sweet].<sup>1</sup>

Compared to these diploid Phaseoleae species, the genus *Glycine* has undergone genome duplications followed by rearrangements leading to diploidization (XX); accordingly, its chromosome number is  $2n = 40$  [e.g., soybean (*G. max*). Among its wild relative  $2n = 40$  (*G. soja* Siebold & Zucc.), or  $2n = 80$  [*G. tabacina* (Labill.) Benth.], or  $2n = 38, 40, 78$ , and  $80$  (*G. tomentella* Hayata). Within the phaseoloid group, the closest generic ally of *Glycine* is the genus *Teramnus* with a chromosome number of  $2n = 28$  (Doyle and Luckow 2003). Thus, the tetraploid nature of soybean is unusual compared to most of its relatives in the millettoid/phaseoloid clade. This situation has to be taken into consideration when attempting to establish macro- and microsynteny between the genome of soybean and related species. It is also important to consider ancestral polyploidization and diploidization events as the physical map (Shultz et al. 2006) and the soybean genome sequence are assembled.

The mitotic and meiotic metaphase chromosomes of *Phaseolus* are small and relatively undifferentiated, making identification of individual chromosomes difficult until recently (Maréchal 1971). The presence of polytene chromosomes in the embryonic suspensor cells has not provided a solution to chromosome identification because they remain relatively uncondensed compared to the polytene chromosomes of *Drosophila* (Nagl 1974). Nevertheless, fluorescent in situ hybridization (FISH) was used to locate rRNA and phaseolin loci on these polytene chromosomes (Nenno et al. 1994). More recently, fluorescently-labeled pooled RFLP and individual BAC probes (Kami et al. 2006) were used to perform FISH on mitotic chromosomes of common bean. Idiograms of two European cultivars (Saxa and Tschermak's fadenlose Wachs) were established using a combination of double FISH with 5S and 45S rDNA probes, chromosome morphology, and heterochromatin distribution. Similar idiograms were established for three other *Phaseolus* domesticated species, *P. coccineus*, *P. acutifolius*, and *P. lunatus* (Moscone et al. 1999). FISH with pooled fluorescently-labeled RFLP probes mapping to the same linkage group (Vallejos et al. 1992) led to a correlation between the genetic and chromosomal maps in two cultivars, Saxa and Diacol Calima (Pedrosa et al. 2003). More recently, Pedrosa-Harand et al. (2006) demonstrated that the two 5S rRNA loci were conserved across the species. In contrast, the number of 45S rRNA loci varied between two and five in the Middle American gene pool and six and nine in the Andean gene pool. From an evolutionary perspective, ancestral wild beans from northern Peru and Ecuador (Kami et al. 1995) resemble the Middle American gene pool in the number of 45S rRNA loci. Thus, these results reveal a major amplification event in the Andean lineage after its separation from the Middle American lineage. Since hybrids between the two gene pools exhibit a full range of the number of 45S rRNA loci, the number

---

<sup>1</sup> Chromosome numbers: Missouri Botanical Garden Index to Plant Chromosome Numbers – IPCN: <http://mobot.mobot.org/W3T/Search/ipcn.html>

of 45S rRNA loci will not be a reliable reference loci for synteny mapping between common bean and soybean.

### ***Genetic Maps***

The importance of genetic linkage maps in understanding the inheritance of a trait has led bean breeders to develop over 25 linkage maps in different crosses of common bean (Kelly et al. 2003; Miklas et al. 2006). Most are low-density linkage maps with markers on average every 10 cM and incomplete coverage of the genome. In order to maximize polymorphism at the molecular level, the majority of mapping populations were derived from crosses between domesticated parents belonging to the Andean vs. Middle American gene pools. In some cases, maps were developed in crosses involving between parents of the same gene pool or between a domesticated and wild parent.

To correlate the mapping results of these different maps, a core map was established in the recombinant inbred population resulting from the cross BAT93 × Jalo EEP558 (Freyre et al. 1998). BAT93 is a breeding line from the Mesoamerican gene pool, and Jalo EEP558 is an Andean gene pool member resulting from selection in a Brazilian landrace. The two parents show contrasting resistances to pathogens. This population consists of some 75 lines showing a high level of polymorphism (Nodari et al. 1992). Some 600 markers were 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; Mattos de Grisi et al. unpubl. results). In addition, several of the markers, principally RFLPs and sequence-tagged markers, are shared with other maps, allowing a general correlation among linkage groups of different maps. The linkage group numbering adopted by Freyre et al. (1998) has become the standard across linkage mapping studies in common bean.

A new set of gene-based markers is being implemented for genetic mapping in common bean including TRAP (Targeted Region Amplification Polymorphism) and RGA (Resistance-Gene Analogs)-based markers targeting disease resistance genes (Divkin et al., 1999; López et al. 2003; Mutlu et al. 2005; Miklas et al. 2006). In addition, gene-based SNP markers are being mapped at CIAT, and mapping of EST sequences at North Dakota State University and at the University of Saskatchewan will lead to the development of a transcriptional map for beans that could help in the establishment of correlations between candidate genes and specific QTLs (Gepts et al. 2007). A deliberate effort needs to be developed to identify markers that identify homologous sequences in both common bean and soybean to delineate regions of synteny.

### ***Gene Cloning and Sequencing***

As with many species, investigators have studied genes on an individual basis in common bean. One of the earliest examples is the cloning of the phaseolin gene in

which the presence of introns was first demonstrated in plants (Sun et al. 1981). The breadth of research for a species can be estimated in a relative manner by comparing the number of complete coding sequences (CDSs) deposited in GenBank. These sequences represent full gene sequence data that was confirmed either genetically or biochemically. Among the *Fabaceae* species, soybean has the most CDSs sequences with 956 (Table 4.1). By comparison, the number for common bean is 195. This number is only half that of *M. truncatula*, an emerging species for legume research. It is important to note that for each of the species in Table 4.1, the number of CDSs represents an almost doubling of those available just three years ago (McClellan et al. 2004a).

Disease resistance is one area in which gene cloning is progressing using a variety of methods. The historical pathogenesis-related genes were cloned several years ago (Ryder et al. 1987; Walter et al. 1990; Blyden et al. 1991; Margis-Pinheiro et al. 1994). Recently the region surrounding the *I* gene, a major gene involved in bean common mosaic virus resistance, was sequenced, and genes similar to other disease resistance genes were discovered (Vallejos et al. 2006). Recently, a cDNA-AFLP screen defined a number of genes associated with the induction of resistance including those associated with G-protein and ABA signaling (Cadde-Davidson and Jahn 2006). This led the authors to suggest general rather than specific signaling systems may be recruited for the defense response. Genes potentially associated with anthracnose resistance were also recently described (Melotto et al. 2004).

Research into other important traits for common bean productivity is also identifying genes and their involvement in the expression of the trait. Recent research that focused on drought and other water relations: (1) identified specific ABA 8'-hydroxylases genes involved in the response to and recovery from water stress

**Table 4.1** Summary GenBank statistics for the Fabaceae family and specific species (verified December 20, 2006)

	Total <sup>a</sup>	EST <sup>b</sup>	CDS <sup>c</sup>	Complete CDS <sup>d</sup>	Genomic survey <sup>e</sup>
Fabaceae	1, 467, 205	887, 838	9, 519	3, 239	545, 653
<i>Cicer spp.</i>	2, 119	297	451	16	0
<i>Glycine max</i>	645, 347	359, 435	2, 137	956	281, 031
<i>Lotus japonicus</i>	199, 018	150, 631	305	163	46, 569
<i>Medicago truncatula</i>	397, 575	225, 129	481	251	168, 809
<i>Phaseolus vulgaris</i>	27, 222	22, 666	579	158	2, 980
<i>Phaseolus spp.</i>	48, 465	43, 534	715	195	2, 980
<i>Pisum sativum</i>	6, 058	3, 594	1, 114	485	154
<i>Vicia spp.</i>	1, 616	1	418	37	686
<i>Vigna spp.</i>	1, 657	466	406	182	89

<sup>a</sup> PubMed nucleotide query: species [orgn]

<sup>b</sup> PubMed nucleotide query: species [orgn] est NOT cds

<sup>c</sup> PubMed nucleotide query: species [orgn] cds NOT est NOT plastid NOT chloroplast NOT mitochondria

<sup>d</sup> Pubmed nucleotide query: species [orgn] complete cds NOT est NOT plastid NOT chloroplast NOT mitochondria

<sup>e</sup> Pubmed Genome Sequence Survey (GSS) query: species [orgn]

(Yang and Zeevaert 2006); (2) showed that specific bZIP transcription factors are specifically expressed in the root and only during water stress (Rodriguez-Uribe and O'Connell 2006); and (3) revealed the differential expression of aquaporins (Aroca et al. 2006) during the stress. Another example is the cloning of several phototropin genes that encode receptor-type protein kinases and the demonstration that they undergo a change in phosphorylation state during the phototropism response (Inoue et al. 2005).

Insect resistance is a key to preventing seed deterioration caused by insect predation during storage. Studies have identified members of the lectin gene family as key to the resistance. This family has been extensively studied, and key members with regards to resistance, the arcelins, phytohemmagglutinins, and the amylase inhibitor family members are encoded by a block of genes co-located on linkage group B4 (Freyre et al. 1998; Kami et al. 2006). Sequence comparisons revealed that the evolutionary history of the *Phaseolus* arcelins (Lioi et al. 2003) and phytohemmagglutinins (Hoffman and Donaldson 1985) includes duplications and diversification of function. The recent cloning and sequencing of 156 kb genomic region around the APA family in common bean will further extend our understanding of the roles of these genes in seed predation (Kami et al. 2006). From a comparative genomics perspective, this genomic data can be compared to the emerging genome sequence of soybean to further understand the effects of duplication and diploidization on the structure of the soybean genome.

## ***EST Collections***

In recent years, the number of expressed sequence tags (ESTs) in *Phaseolus* has increased markedly (compare Table 4.1 here and Table 4.1 in McClean et al. 2004a). To date some 45,000 ESTs have been developed in *Phaseolus*. These include 25,000 in *P. vulgaris* (Ramirez et al. 2005; Melotto et al. 2005) and 20,000 in the closely related species *P. coccineus*. Three *P. vulgaris* genotypes were used to obtain these ESTs: two of Mesoamerican origin (Negro Jamapa and SEL1308) and one of Andean origin (G19833). Tissues sampled include seedling shoots [with or without *Colletotrichum lindemuthianum* (anthracnose) infection], seedling leaves, nodules elicited by *Rhizobium tropici* strain CIAT899, 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). Gene indices for the *P. vulgaris* ESTs were recently developed ([http://biocomp.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=p\\_vulgaris](http://biocomp.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=p_vulgaris); [http://www.plantgdb.org/download/download.php?dir=/Sequence/ESTcontig/Phaseolus\\_vulgaris](http://www.plantgdb.org/download/download.php?dir=/Sequence/ESTcontig/Phaseolus_vulgaris)).

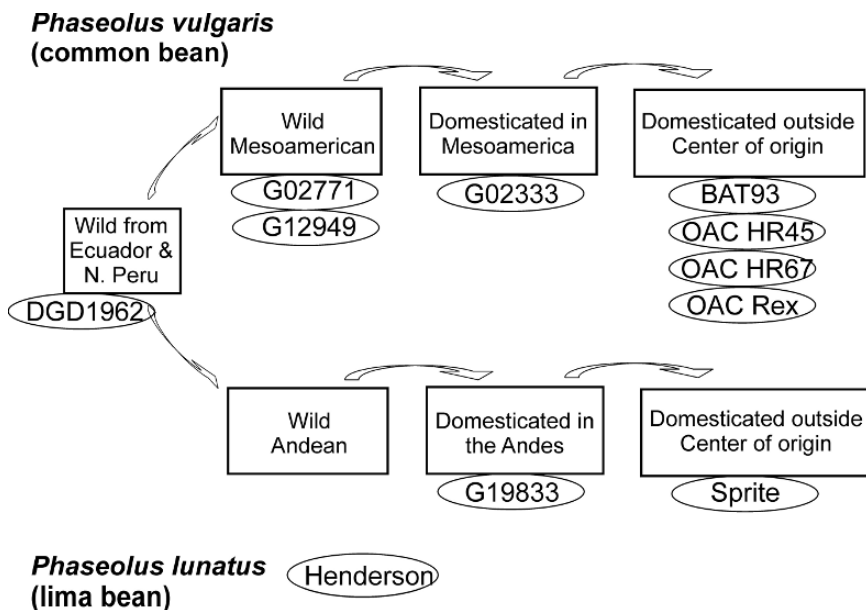
Additional EST resources in common bean (for example, from different flower, pod, and seed developmental stages, organs, and tissues) would be particularly use-

ful because these are or lead to the harvested organs. A comparison of expression patterns between *P. vulgaris* and *G. max* would also be useful to gauge the effect of polyploidization/diploidization.

### ***BAC Constructions and Their Applications***

Eleven BAC libraries have been constructed in the genus *Phaseolus*, ten in *P. vulgaris* and one in *P. lunatus* (Gepts et al. 2007). To increase the concentration of high-molecular weight DNA available for cloning, Kami et al. (2006) developed an extraction procedure based on a novel cell nuclei isolation procedure. The majority of the libraries were developed after restriction digestion with *Hind*III, although one library results from *Eco*RI digestion and two from *Bam*HI digestion. Most libraries have a coverage of 5–12x genome, based on their average insert size. 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 unique phylogenetically ordered set useful for evolutionary studies (Fig. 4.3) as each represents a key genotype in the intra-specific evolution and domestication of common bean (Gepts et al. 2007). DGD1962 is a wild bean from northern Peru, representing the presumed ancestral



**Fig. 4.3** Phylogenetic and genealogical distribution of BAC libraries in *Phaseolus* spp. Boxes represent different segments of the *P. vulgaris* gene pools and the general direction of their evolution. Names surrounded by ellipses are the genotypes in which BAC libraries have been established (from Gepts et al. 2007)

gene pool of the species (Debouck et al. 1993; Kami et al. 1995). The remainder of the libraries is distributed in the two evolutionary lineages that were domesticated. In the Mesoamerican lineage, G02771 and G12946 are wild Mexican beans 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 and OAC-HR45 and OAC-HR67 are breeding lines and OAC-Rex is a cultivar from the Mesoamerican gene pool. In the Andean gene pool, G19833 is a landrace from Peru, whereas Sprite is a bred variety. Thus, using this array of BAC libraries it is possible to study overall structural evolution of the genome in *Phaseolus* both prior and after domestication. It is also possible to analyze phenotypic changes resulting from specific structural modification at the genome level. Single BAC clones were 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.

### ***Emerging Procedures: TILLING and Transformation***

As defined by McCallum et al. (2000), TILLING (or Targeted Induced Local Lesions IN Genomes) is a tool to identify artificially-induced mutants in specific genes of interest. This reverse genetics approach does not require an effective transformation system. In common bean, a TILLING platform is being developed by W. Broughton, P. Lariguet, T. Porch, and M.W. Blair (unpubl. data and Gepts et al. 2007) in the genotype BAT93, mentioned before. Mutagen (EMS) concentration and the resulting lethality have been determined. Currently, some 1,500 M2 families have been produced, and 900 of these were advanced to the M3 stage. This TILLING platform will benefit from future sequencing efforts in common bean and other species, such as soybean by providing reference sequences of loci of interest. In turn, these sequences will be used to design gene-specific primers.

Common bean is reputed to be a species recalcitrant to transformation. However, recent progress has increased the feasibility of transformation as a tool in *Phaseolus* (Veltcheva et al. 2005). In tepary bean (*Phaseolus acutifolius* A. Gray), *Agrobacterium* transformation is relatively efficient and was used to test gene function (De Clercq et al. 2002; Zambre et al. 2005). Although transformation of *P. vulgaris* appears to be more difficult, three avenues have been pursued. Biolistics was used to obtain transgenic common bean plants (Aragão et al. 1998, 1999, and 2002), some of which are undergoing field testing in Brazil (F. Aragão personal communication). Liu et al. (2005) developed a protocol based on the combination of sonication and vacuum infiltration to transform common bean with a Late Embryogenesis Abundant gene conferring abiotic stress tolerance. Although the efficiency remains to be improved, this method looks promising. Recently, Estrada-Navarrete et al. (2006) modified a protocol originally developed in soybean for efficient and robust *Agrobacterium rhizogenes* transformation of common bean. The availability

of transgenic roots in *Phaseolus* will provide a new method for overexpressing or suppressing endogenous genes, especially those involved in root biology and root-microbe interactions.

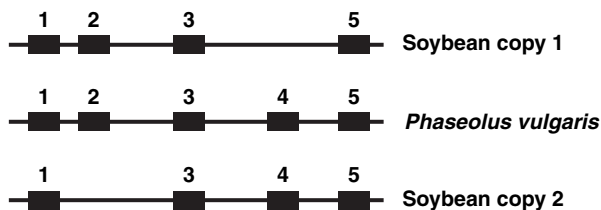
## Application of *Phaseolus* Genomic Tools to *Glycine*

Soybean and common bean are valuable legumes in their own regard, and share the feature of being important nitrogen fixing species. They also share an evolutionary lineage that has the potential to foster a research synergy that will benefit both species. *G. max* has a relatively large sized genome of  $\sim 1,100$  Mbp organized in 20 chromosome pairs, whereas, *P. vulgaris* has a genome nearly half the size of soybean organized in 11 chromosome pairs. Taxonomically, *P. vulgaris* is closely related to soybean, separated by 19.2 Ma (Lavin et al. 2005), possibly around the time of the last major duplication event in soybean (Schlueter et al. 2004). Genetically, *P. vulgaris* appears to be a true diploid, whereas, soybean is genetically/genomically complicated by multiple rounds of duplication/polyploidization (Blanc and Wolfe 2004; Schlueter et al. 2004). Thus, given the close evolutionary proximity of these two species, it may be possible to exploit the simple genome of *Phaseolus* to understand the organization and evolution of the duplicated soybean genome. To that end, a physical map of *Phaseolus* is being created from fingerprinted BAC clones from the G19833 genotype. This BAC-based physical map, along with a BAC end sequence (BES) database, will be used to help resolve duplication events in soybean as well as to provide a platform for mapping and gene cloning in *Phaseolus*. *Phaseolus*, along with *M. truncatula*, *Lotus japonicus*, and soybean provide a framework within legumes for highly detailed evolutionary analyses of domestication and polyploidization.

## Use of Secondary Species to Assist with Genome Analysis

Soybean has a complicated genome due to multiple rounds of duplication and/or polyploidization (Shoemaker et al. 1996; Blanc and Wolfe 2004; Pagel et al. 2004; Schlueter et al. 2004; Walling et al. 2005). Since *Phaseolus* is diploid, to the extent that we can determine without genome sequencing, and given its phylogenetic proximity to soybean (Doyle and Luckow 2003), it provides an out-group to evolutionarily and structurally help define duplications within the soybean genome that occurred after the divergence of *Phaseolus* and soybean (Fig. 4.4). Shared duplications that occurred prior to the divergence of *Phaseolus* and soybean may be used to compare orthologous regions from either *M. truncatula* or *L. japonicus* to determine the ancestral legume genome structure.

Multiple sequence alignments of orthologous regions from related genomes has proven useful to understand genome dynamics such as gene movement/loss/gain and repeat accumulation/differentiation/loss (Thomas et al. 2003; Lai et al. 2004;



**Fig. 4.4** Sequence alignment of a duplicated region in soybean to the orthologous region from *Phaseolus vulgaris*. The putative ancestral state can be determined and gene loss can be seen in both soybean copies

Ma et al. 2005). Comparative sequence alignments were extremely useful to confirm gene structure and predict regulatory elements as seen in comparisons of human to other mammals and among yeast species (Flint et al. 2001; Kellis et al. 2003). These algorithms benefit from having multiple species at varying levels of separation from the genome of interest in order to provide useful annotation. Closely related genomes can be useful to understand local and recent gene movement, whereas, more distantly related genomes can be used to predict conserved non-coding sequences (CNS). For instance, a fugu-human comparison revealed almost 1,400 CNS and 23/25 tested CNS showed significant enhancer activity (Woolfe et al. 2005).

### ***Defining Gene Models Based on Genes of Different Evolutionary Distance***

Annotation of sequenced genomes uses a variety of approaches including *ab initio* prediction using sequence features (Salamov and Solovyev 2000), sequence identity to expressed sequence tags (ESTs) (Wei and Brent 2006) and comparative sequence alignments (Flicek et al. 2003). The assumption is that non-coding sequences should not be selectively constrained across evolutionary time, therefore, conserved sequences between distantly related species should hold some functional significance, often being coding or regulatory. Comparisons of human sequences to various mammals helped to define gene/gene structure and regulatory elements (Woolfe et al. 2005). In addition, using carefully placed species on the evolutionary tree, inferences can be made about conserved non-coding sequences that may play significant roles in gene function (Margulies et al. 2003; Margulies and Green 2003; Margulies et al. 2005). For example, plants species separated by 16–50 MY were useful for such comparisons (Guo and Moose 2003; Inada et al. 2003). The approximate evolutionary distance between *Phaseolus* and soybean is 19.2 Ma (Lavin et al. 2005). Sequence information from *Phaseolus*, combined with sequence from *M. truncatula* and/or *L. japonicus* (~50 Ma) will provide an evolutionary framework in the legumes for determining gene structure, organization and to make hypotheses on gene function.

### ***Assisting with Final Assembly of Sequenced Genome***

Very few genomes are likely to be sequenced using a clone-by-clone approach as was done for *Arabidopsis* (AGI 2000), rice (IRGSP 2005), *C. elegans* (CESC 1998) and a few others. The advent of shotgun sequencing of entire eukaryotic genomes altered the approach by which most future genomes will be sequenced. Shotgun sequences of whole genomes are assembled as scaffolds of sequence contigs with many holes due to repeats and/or lack of sequences coverage and occasional mis-assemblies due to recent duplications (Adams et al. 2000; Myers et al. 2000; Venter et al. 2001; Tuskan et al. 2006). Yet this approach still provides most of the gene models of the species. Most genomes will likely be sequenced using hybrid approaches, most commonly a mixture of shotgun sequencing and limited physical mapping.

In soybean, a physical map is being constructed for the genotype Williams 82, and a shotgun sequence is targeted for completion by 2008 (Jackson et al. 2006). The physical map with its sequence tag connectors (STCs) will be used, inasmuch as possible, to help assemble sequence contigs into sensible scaffolds. In conjunction with this, a physical map of *Phaseolus*, with its own STC database, will be used to (1) check assembly, (2) resolve duplicated regions by comparison to a 'diploid' sister species, and (3) possibly span gaps in the soybean sequence/physical map. In rice, ~ 35 gaps (excluding telomere/centromere gaps) existed in the sequence map at completion (2005), but sister species, in the same genus, were found that contained BAC clones that putatively spanned the gaps (Wing et al. 2005). This is an example of how a closely related out-group such as *Phaseolus* can be used to facilitate genome sequencing/assembly in soybean.

### ***Identifying Adaptation Genes Within the Phaseolaceae***

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 process and necessitated the large scale sequencing 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 contain 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. Syntenic map data (Choi et al. 2004) provides a reference point from which orthologs between two legume species can be identified.

Orthologous genes found to have undergone selection by the common bean and soybean pairwise comparison can then be compared with a Galegoid 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 Phaseoleae lineage. Likewise, if the comparison to one of the Galegoid 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. Such comparisons, though, would require significant data from a member of the Rosales or Cucurbitales, data which is currently not available in significant depth.

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 1,095 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 two socially and economically important legumes, we simply need the resources and will to collect common bean sequence to the same depth as soybean.

## References

- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D. et al. (2000) The genome sequence of the *Drosophila melanogaster*. *Science* 287, 2185–2195.
- AGI (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.

- Aragão, F., Ribeiro, S., Barros, L., Brasileiro, A., Maxwell, D., Rech, E. and Faria, J. (1998) Transgenic beans (*Phaseolus vulgaris* L.) engineered to express viral antisense RNAs show delayed and attenuated symptoms to bean golden mosaic geminivirus. *Mol. Breed.* 4, 491–499.
- Aragão, F.J.L., Barros, L.M.G., de Sousa, M.V., Grossi de Sa, M.F., Almeida, E.R.P., Gander, E.S. and Rech, E.L. (1999) Expression of a methionine-rich storage albumin from the Brazil nut (*Bertholletia excelsa* H.B.K., Lecythidaceae) in transgenic bean plants (*Phaseolus vulgaris* L., Fabaceae). *Genetics & Mol. Biol.* 22, 445–449.
- Aragão, F.J.L., Vianna, G.R., Albino, M.M.C. and Rech, E.L. (2002) Transgenic dry bean tolerant to the herbicide glufosinate ammonium. *Crop Sci.* 42, 1298–1302.
- Aroca, R., Ferrante, A., Vernieri, P. and Chrispeels, M.J. (2006) Drought, abscisic acid and transpiration rate effects on the regulation of PIP aquaporin gene expression and abundance in *Phaseolus vulgaris* plants. *Ann. Bot.* 98, 1301–1310.
- Arumuganathan, K. and Earle, D.E. (1991) Nuclear DNA content of some important plant species. *Plant Mol. Biol. Report.* 9, 208–218.
- Becerra Velásquez, V.L., and Gepts, P. (1994) RFLP diversity in common bean (*Phaseolus vulgaris* L.). *Genome* 37, 256–263.
- Beebe, S.E., Rojas-Pierce, M., Yan, X., Blair, M.W., Pedraza, F., Muñoz, F., Tohme, J. and Lynch, J.P. (2006) Quantitative trait loci for root architecture traits correlated with phosphorus acquisition in common bean. *Crop Sci.* 46, 413–423.
- Beebe, S.E., Skroch, P.W., Thome, J., Duque, M.C., Pedraza, F., Nienhuis, J. (2000) Structure of genetic diversity among common bean landraces of middle American origin based on correspondence analysis of RAPD. *Crop Sci.* 40, 264–273.
- Bennett, M. and Leitch, I. (1995) Nuclear DNA amounts in angiosperms. *Ann. Bot.* 76, 113–176.
- Bennett, M. and Leitch, I. (2005) Plant DNA C-Values database. Release 4.0 <http://www.rbgekew.org.uk/cval/database1.html>.
- Blanc, G., and Wolfe, K.H. (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 17, 1667–1678.
- Blair, M.W., Giraldo, M.C., Buendia, H.F., Tovar, E., Duque, M.C., and Beebe, S.E. (2006) Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.) *Theor. Appl. Genet.* 113, 100–109.
- Blair, M.W., Pedraza, F., Buendia, H.F., Gaitán-Solís, E., Beebe, S.E., Gepts, P., and Tohme, J. (2003) Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.) *Theor. Appl. Genet.* 107, 1362–1374.
- Blyden, E.R., Doerner, P.W., Lamb, C.J. and Dixon, R.A. (1991) Sequence analysis of a chalcone isomerase cDNA of *Phaseolus vulgaris* L. *Plant Mol. Biol.* 16, 167–169.
- Broughton, W.J., Hernandez, G., Blair, M., Beebe, S., Gepts, P. and Vanderleyden, J. (2003) Beans (*Phaseolus* spp.) - model food legumes. *Plant and Soil* 252, 55–128.
- Cadle-Davidson, M. and Jahn, M. (2006) Differential gene expression in *Phaseolus vulgaris* I locus NILs challenged with bean common mosaic virus. *Theor. Appl. Genet.* 112, 1452–1457.
- CECIS (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018.
- Choi, S.S., and Lahn, B.T. (2003) Adaptive evolution of MRG, a neuron-specific gene family implicated in nociception. *Genome Res.* 13, 2252–2259.
- Choi, H.K., Mun, J.H., Kin, D.J., Zhu, H., Baek, J.M., Mudge, J., Roe, B., Ellis, N., Doyle, J., Kiss, G.B., Young, N.D., and Cook, D.R. (2004) Estimating genome conservation between crop and model legume species. *Proc. Natl. Acad. Sci., USA* 101, 15289–15294.
- Crepet, W.L., Nixon, K.C., and Gandolfo, M.A. (2004) Fossil evidence and phylogeny: the age of major angiosperm clades based on mesofossil and macrofossil evidence from Cretaceous deposits. *Amer. J. Bot.* 91, 1666–1682.
- CSAC (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69–87.
- Debouck, D.G., Toro, O., Paredes, O.M., Johnson, W.C., and Gepts, P. (1993) Genetic diversity and ecological distribution of *Phaseolus vulgaris* in northwestern South America. *Econ. Bot.* 47, 408–423.

- DeClercq, J., Zambre, M., Van Montagu, M., Dillen, W. and Angenon, G. (2002) An optimized *Agrobacterium*-mediated transformation procedure for *Phaseolus acutifolius* A. Gray. *Pl. Cell Reports* 21, 333–340.
- Delgado-Salinas, A., Bibler, R. and Lavin, M. (2006) Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. *Syst. Bot.* 56, in press.
- Doyle, J.J., and Luckow, M.A. (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Phys.* 131, 900–910.
- Estrada-Navarrete, G., Alvarado-Affantranger, X., Olivares, J.-E., Díaz-Camino, C., Santana, O., Murillo, E., Guillén, G., Sánchez-Guevara, N., Acosta, J., Quinto, C., Li, D., Gresshoff, P.M., and Sánchez, F. (2006) *Agrobacterium rhizogenes*-transformation of the *Phaseolus* spp.: A tool for functional genomics. *Mol. Plant Microbe Interact.* 19, 1385–1393.
- Flicke, P., Keibler, E. Hu, P. Korf, I., and Brent, M.R. (2003) Leveraging the mouse genome for gene prediction in human: from whole-genome shotgun reads to a global synteny map. *Genome Res.* 13, 46–54.
- Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R.J., Hardison, R., Miller, W., Philipson, S., Tan-Un, K.C., McMorrow, T., Frampton, J., Alter, B.P., Frischauf, A.M., and Higgs, D.R. (2001) Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. *Hum. Mol. Genet.* 10, 371–382.
- Freyre, R., Skroch, P.W., Geffroy, V., Adam-Blondon, A.F., Shirmohamadali, A., Johnson, W.C., Llaca, V., Nodari, R.O., Pereira, P.A., Tsai, S.M., Tohme, J., Dron, M., Nienhuis, J., Vallejos, C.E., and 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.
- 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. *HortSci.* 33, 1124–1130.
- Gepts, P., Aragão, F., de Barros, E., Blair, M.W., Brondani, R., Broughton, W.J., 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. and Yu, K. (2007) 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, in press.
- Gepts, P., and Bliss, F.A. (1986) Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. *Econ. Bot.* 40, 469–478.
- Gepts, P. and 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.
- González, A., Wong, A., Delgado-Salinas, A., Papa, R., and 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.
- Guo, H., and Moose, S.P. (2003). Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. *Plant Cell* 15, 1143–1158.
- Hoffman, L.M., and Donaldson, D.D. (1985) Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. *EMBO J.* 4, 883–889.
- Inada, D.C., Bashir, A., Lee, C., Thomas, B.C., Ko, C., Goff, S.A., and Freeling, M. (2003). Conserved noncoding sequences in the grasses. *Genome Res.* 13, 2030–2041.
- Inoue, S., Konoshita, T., and Simazaki, K. (2005) Possible involvement of phototropins in leaf movement of kidney bean in response to blue light. *Plant Physiol.* 138, 1994–2004.
- IRGSP (2005) The map-based sequence of the rice genome. *Nature* 436, 793–800.
- Islam, F.M.A., Basford, K.E., Jara, C., Redden, R.L., and Beebe, S. (2002) Seed compositional and disease resistance differences among gene pools in cultivated common bean. *Genet. Resour. Crop. Evol.* 49, 285–293.

- Jackson, S.A., Rokshar, D., Stacey, G., Shoemaker, R.C., Schmutz, J., and Grimwood, J. (2006) Toward a reference sequence of the soybean genome: a multiagency effort. *Crop Sci.* 46, S-55–61.
- Kami, J., Becerra Velásquez, B., Debouck, D.G., and 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. and 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.
- 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.
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423, 242–254.
- Kelly, J.D., Gepts, P., Miklas, P.N. and Coyne, D.P. (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.
- Koenig, R., and Gepts, P. (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. *Theor. Appl. Genet.* 78, 809–817.
- Kumar, S., Filipowski, A., Swarna, V., Walker A., and Hedges, B.S. (2005) Placing confidence limits on the molecular age of the human–chimpanzee divergence. *Proc. Natl. Acad. Sci., U.S.A.* 102, 18842–18847.
- Lai, J., Ma, J., Swigonova, Z., Ramakrishna, W., Linton, E., Llaca, V., Tanyolac, B., Park, Y.J., Jenong, O.Y., Bennetzen, J.L., and Messing, J. (2004) Gene loss and movement in the maize genome. *Genome Res.* 14, 1924–1931.
- Lavin, M., Herendeen, P.S., and Wojciechowski, M.F. (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *System. Biol.* 54, 530–549.
- Lewis, G., Schrire, B., Mackinder, B., and Lock, M. (2005) *Legumes of the World*. Royal Botanic Gardens, Kew Publishing; London.
- Lioi, L., Sparvoli, F., Galasso, I., Lanave, C., and Bollini, R. (2003) Lectin-related resistance factors against bruchids evolved through a number of duplication events. *Theor. Appl. Genet.* 107, 814–822.
- Liu, Z.C., Park, B.J., Kanno, A., and Kameya, T. (2005) The novel use of a combination of sonication and vacuum infiltration in *Agrobacterium*-mediated transformation of kidney bean (*Phaseolus vulgaris* L.) with lea gene. *Mol. Breed.* 16, 189–197.
- López, C.E., Acosta, I.F., Jara, C., Pedraza, F., Gaitán-Solís, E., Gallego, G., Beebe, S. and Tohme, J. (2003) Identifying resistance gene analogs associated with resistances to different pathogens in common bean. *Phytopathology* 93, 88–95.
- Ma, J., SanMiguel, P., Lai, J., Messing, J., and Bennetzen, J.L. (2005) DNA rearrangement in orthologous *orp* regions of the maize, rice, and sorghum genomes. *Genetics* 170, 1209–1220.
- Magallón, S., Crane, P.R., and Herendeen, P.S. (1999) Phylogenetic pattern, diversity, and diversification of eudicots. *Ann. Mo. Bot. Gard.* 86, 297–372.
- Magallón, S., and Sanderson, M.J. (2001) Absolute diversification rates in angiosperms. *Evolution* 55, 1762–1780.
- Maréchal, R. (1971) Observations sur quelques hybrides dans le genre *Phaseolus*. II. Les phénomènes méiotiques. *Bull. Rech. Agron. Gembloux* 6, 461–489.
- Margis-Pinheiro, M., Martivet, J. and Burkard, G. (1994) Bean class IV chitinase gene: structure, developmental expression and induction by heat stress *Plant Sci.* 98, 163–173.
- Margulies, E.H., and Green, E.D. (2003) Detecting highly conserved regions of the human genome by multispecies sequence comparison. *Cold Spring Harb. Symp. Quant. Biol.* 68, 255–263.
- Margulies, E.H., Blanchette, M., Haussler, D., and Green, E.D. (2003) Identification and characterization of multi-species conserved sequences. *Genome Res.* 13, 2507–2518.

- Margulies, E.H., Vinson, J.P., Miller, W., Jaffe, D.B., Lindblad-Toh, K., Chang, J.L., Green, E.D., Lander, E.S., Mullikin, J.C., and Clamp, M. (2005) An initial strategy for the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing. *Proc. Natl. Acad. Sci., USA*. 102, 4795–4800.
- Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler IV, E., and Doebley, J. (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci., USA*. 99, 6080–6084.
- McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Pl. Physiol.* 123, 439–442.
- McClean, P.E., Kami, J., and 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. and Miklas, P.N. (2004b) Sequence diversity analysis of dihydroflavonol reductase intron 1 in common bean. *Genome* 47, 266–280.
- Melotto, M., Coelho, M.F., Pedrosa-Harand, A., Kelly, J.D. and Camargo, L.E.A. (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, C.B., Bruschi, A.G., and Camargo, L.E. (2005) Comparative bioinformatic analysis of genes expressed in common bean (*Phaseolus vulgaris* L.) seedlings. *Genome*, 48, 562–570.
- Miklas, P.N., Kelly, J.D., Beebe, S.E. and Blair, M.W. (2006) Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. *Euphytica* 147, 106–131.
- Moscone, E.A., Klein, F., Lambrou, M., Fuchs, J. and Schweizer, D. (1999) Quantitative karyotyping and dual-color FISH mapping of 5S and 18S–25S rDNA probes in the cultivated *Phaseolus* species (Leguminosae). *Genome* 42, 1224–1233.
- Mutlu, N., Miklas, P.N. and Coyne, D.P. (2005) Resistance gene analog polymorphism (RGAP) markers co-localize with disease resistance genes and QTL in common bean. *Mol. Breed.* 17, 127–135.
- Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., *et al.* (2000) A whole-genome assembly of *Drosophila*. *Science* 287, 2196–2204.
- Nagl, W. (1974) The *Phaseolus* suspensor and its polytene chromosomes. *Z. Pflanzenphysiol.* 73, 1–44.
- Nanni, L., Losa, A., Bellucci, E., Kater, M., Gepts, P. and Papa, R. (2005) Identification and molecular diversity of a genomic sequence similar to SHATTERPROOF (SHP1) in *Phaseolus vulgaris* L. Abstract, Plant and Animal Genome XIII, Jan. 15–19, 2005: [http://www.intl-pag.org/pag/13/abstracts/PAG13\\_P470.html](http://www.intl-pag.org/pag/13/abstracts/PAG13_P470.html)
- Nonno, M., Schumann, K. and Nagl, W. (1994) Detection of rRNA and phaseolin genes on polytene chromosomes of *Phaseolus coccineus* by fluorescence *in situ* hybridization after pepsin treatment. *Genome* 37, 1018–1021.
- Nodari, R.O., Koinange, E.M.K., Kelly, J.D., and 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.
- Ochoa, I.E., Blair, M.W., and Lynch, J.P. (2006) QTL analysis of adventitious root formation in common bean under contrasting phosphorus availability. *Crop Sci.* 46, 1609–1621.
- Pagel, J., Walling, J.G., Young, N.D., Shoemaker, R.C., and Jackson, S.A. (2004) Segmental duplications within the *Glycine max* genome revealed by fluorescence *in situ* hybridization of bacterial artificial chromosomes. *Genome* 47, 764–768.
- Papa, R., and 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.
- Pedrosa, A., Vallejos, C.E., Bachmair, A., and Schweizer, D. (2003) Integration of common bean (*Phaseolus vulgaris* L.) linkage and chromosomal maps. *Theor. Appl. Genet.* 106, 205–212.

- Pedrosa-Harand, A., Almeida, C.C.S., Mosiolek, M., Blair, M.W., Schweizer, D. and Guerra, M. (2006) Extensive ribosomal DNA amplification during Andean common bean (*Phaseolus vulgaris* L.) evolution. *Theor. Appl. Genet.* 112, 924–933.
- Powell, W., Morgante, M., Doyle, J.J., McNicol, J.W., Tingey, S.V., and Rafalski, A.J. 1996. Genepool variation in genus *Glycine* subgenus soja revealed by polymorphic nuclear and chloroplast microsatellites. *Genetics* 144, 793–803.
- Ramirez, M., Graham, M.A., Blanco-Lopez, L., Silvente, S., Medrano-Soto, A., Blair, M.W., Hernandez, G., Vance, C.P. and Lara, M. (2005) Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. *Pl. Physiol.* 137, 1211–1227.
- Rivkin, M.I., Vallejos, C.E., and McClean, P.E. 1999. Disease-resistance related sequences in common bean. *Genome* 42, 41–47.
- Rodriguez-Urbe, L. and O’Connell, M.A. (2006) A root-specific bZIP transcription factor is responsive to water deficit stress in tepary bean (*Phaseolus acutifolius*) and common bean (*P. vulgaris*). *Plant Sci.* 171, 300–307.
- Ryder, T.B., Hedrick, S.A., Bell, J.N., Liang, X.W., Clouse, S.D. and Lamb, C.J. (1987) Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. *Mol. Gen. Genet.* 210, 219–233.
- Salamov, A.A., and Solovyev, V.V. (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* 10, 516–522.
- Schlueter, J.A., Dixon, P., Granger, C., Grant, D., Clark, L., Doyle, J.J., and Shoemaker, R.C. (2004) Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47, 868–876.
- Shoemaker, R.C., Polzin, K., Labate, J., Specht, J., Brummer, E.C., Olson, T.K., Young, N., Concibido, V., Wilcox, J., Tamulonis, J.P., Kochert, G., and Boerma, H.R. (1996) Genome duplications in soybean (*Glycine* subgenus *soja*). *Genetics* 144, 329–338.
- Shultz, J.L., Kurunam, D., Shopinski, K., Iqbal, M.J., Kazi, S., Zobrist, K., Bashir, R., Yaegashi, S., Lavu, N., Afzal, A.J., Yesudas, C.R., Kassem, M.A., Wu, C.C., Zhang, H.B., Town, C.D., Meksem, K. and Lightfoot, D.A. (2006) The Soybean Genome Database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of *Glycine max*. *Nucl. Acids Res.* 34, D758–D765.
- Singh, S.P., Gepts, P., and DeBouck, D.G. (1991). Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ.Bot.* 45, 379–386.
- Stevens, P.F. (2001) Angiosperm Phylogeny Website. Version 7, December 2006 and continuously updated (<http://www.mobot.org/MOBOT/research/APweb/>).
- Sun, S.M., Slightom, J.L., and Hall, T.C. (1981) Intervening sequences in a plant gene – comparison of the partial sequence of the cDNA and genomic DNA of French bean phaseolin. *Nature* 289, 37–41.
- Thomas, J.W., Touchman, J.W., Blakesley, R.W., Bouffard, G.G., Beckstrom-Sternberg, S.M., *et al.* (2003) Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 424, 788–793.
- Tohme, J., González, D.O., Beebe, S., and Duque, M.C. (1996) AFLP analysis of gene pools of a wild bean core collection. *Crop Sci.* 36, 1375–1384.
- Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604.
- Vallejos, C.E., Astua-Monge, G., Jones, V., Plyler, T.R., Sakiyama, N.S., and Mackenzie, S.A. (2006) Genetic and molecular characterization of the *I* locus of *Phaseolus vulgaris*. *Genetics* 172, 1229–1242.
- Vallejos, C.E., Sakiyama, N.S., and Chase, C.D. (1992) A molecular marker-based linkage map of *Phaseolus vulgaris* L. *Genetics* 131, 733–740.
- Veltcheva, M., Svetleva, D., Petkova, S. and Perl, A. (2005) *In vitro* regeneration and genetic transformation of common bean (*Phaseolus vulgaris* L.) - problems and progress. *Scientia Horticulturae* 107, 2–10.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351.

- Walling, J.G., Pires, J.C., and Jackson, S.A. (2005) Preparation of samples for comparative studies of plant chromosomes using *in situ* hybridization methods. *Methods in Enzymology* 395, 442–459.
- Walter, M.H., Liu, J.W., Grand, C., Lamb, C.J. and Hess, D. (1990) Bean pathogenesis-related (PR) proteins deduced from elicitor-induced transcripts are members of a ubiquitous new class of conserved PR proteins including pollen allergens. *Mol. Gen. Genet.* 222, 353–360.
- Wei, C., and Brent, M.R. (2006) Using ESTs to improve the accuracy of *de novo* gene prediction. *BMC Bioinformatics* 7, 327.
- Wikström, N., Savlainen, V., and Chase, M.W. (2001) Evolution of the angiosperms: calibrating the family tree. *Proc. Royal Soc. London, B, Biol. Sci.* 268, 2211–2220.
- Wing, R.A., Ammiraju, J.S., Luo, M., Kim, H. Yu, Y. Kudrna, D., Goicoechea, J.L., Wang, W., Nelson, W., Rao, K., Brar, D., Mackill, D.J., Han, B., Soderlund, C., Stein, L., SanMiguel, P., and Jackson, S. (2005) The *Oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol. Biol.* 59, 53–62.
- Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.f., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y.J., Cooke, J.E., and Elgar, G. (2005) Highly expressed non-coding sequences are associated with vertebrate development. *PLoS Biol.* 3, e7.
- Wright, S.I., Bi, I.V., Schroeder, S.G., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S. (2005). The effects of artificial selection on the maize genome. *Science* 308, 1310–1314.
- Xu, D.H. and Gai, Y. (2003) Genetic diversity of wild and cultivated soybeans growing in China revealed by RAPD analysis. *Plant Breeding* 122, 503–506.
- Yamasaki, M., Tenaillon, M.I., Bi, I.V., Schroeder, S.G., Sanchez-Villeda, H., Doebley, J.F., Gaut, B.S., McMullen, M.D. (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* 17,2859–2872.
- Yang, S.H., and Zeevaart, J.A. (2006) Expression of ABA 8'-hydroxylases in relation to leaf water relations and seed development in bean. *Plant J.* 47, 675–686.
- Zambre, M., Goossens, A., Cardona, C., Van Montagu, M., Terry, N. and Angenon, G. (2005) A reproducible genetic transformation system for cultivated *Phaseolus acutifolius* (tepary bean) and its use to assess the role of arcelins in resistance to the Mexican bean weevil. *Theor. Appl. Genet.* 110, 914–924.