Theor Appl Genet (2006)
DOI 10.1007/s00122-005-0201-2

ORIGINAL PAPER

James Kami · Valérie Poncet · Valérie Geffroy
Paul Gepts

Development of four phylogenetically-arrayed BAC libraries and sequence of the APA locus in *Phaseolus vulgaris*

Received: 11 September 2005 / Accepted: 30 November 2005

Abstract The APA family of seed proteins consists of three subfamilies, in evolutionary order of hypothesized appearance: phytohaemagglutinins (PHA), α-amylase inhibitors (αAI), and arcelins (ARL). The APA family plays a defensive role against mammalian and insect seed predation in common bean (*Phaseolus vulgaris* L.). The main locus (*APA*) for this gene family is situated on linkage group B4. In order to elucidate the pattern of duplication and diversification at this locus, we developed a BAC library in each of four different *Phaseolus* genotypes that represent presumptive steps in the evolutionary diversification of the APA family. Specifically, BAC libraries were established in one *P. lunatus* (cv. 'Henderson: PHA^+ αAI^- ARL^-) and three *P. vulgaris* accessions (presumed ancestral wild G21245 from northern Peru: PHA^+ αAI^+ ARL^-; Mesoamerican wild G02771: PHA^+ αAI^+ ARL^+; and Mesoamerican breeding line BAT93: PHA^+ αAI^+ ARL^-). The libraries were constructed after *Hin*dIII digestion of high molecular weight DNA, obtained with a novel nuclei isolation procedure. The frequency of empty or cpDNA-sequence-containing clones in all libraries is low (generally $<1\%$). The Henderson, G21245, and G02771 libraries have a 10× genome coverage, whereas the BAT93 library has a 20× coverage to allow further, more detailed genomic analysis of the bean genome. The complete sequence of a 155 kbp-insert clone of the G02771 library revealed six sequences belonging to the APA gene family, including members of the three subfamilies, as hypothesized. The different subfamilies were interspersed with retrotransposon sequences. In addition, other sequences were identified with similarity to chloroplast DNA, a dehydrin gene, and the Arabidopsis flowering D locus. Linkage between the dehydrin gene and the *D1711* RFLP marker identifies a potential syntenic region between parts of common bean linkage group B4 and cowpea linkage group 2

Communicated by F. J. Muehlbauer

J. KamiV. PoncetV. Geffroy · P. Gepts (✉)
Department of Plant Sciences, Section of Crop and Ecosystem Sciences, University of California, Mailstop 1, 1 Shields Avenue, Davis, CA 95616-8780, USA
E-mail: plgepts@ucdavis.edu
Tel.: +1-530-7527743
Fax: +1-530-7524361

Present address: V. Poncet
UMR 1097 Diversité et Génomes des Plantes Cultivées, Institut de Recherche pour le Développement, 911 avenue Agropolis, 34394 Montpellier cedex 5, France

Present address: V. Geffroy
Institut de Biotechnologie des Plantes, Bâtiment 630, Université de Paris-Sud, 91405 Orsay Cedex, France

Introduction

The arcelin–phytohemagglutinin–α-amylase inhibitor (APA) multigene family encodes albumins, one of the two major groups of seed storage proteins in common bean (*Phaseolus vulgaris* L.). The APA family consists of three subfamilies: (1) phytohemagglutinins or seed lectins (PHAs); (2) α-amylase inhibitors (αAI); and (3) arcelins (ARL). Although early research efforts were prompted by the agglutinating properties of PHAs (Pusztai 1991), it is the role of ARL and αAI in plant defense that has provided impetus for more recent investigations in common bean (Osborn et al. 1988; Chrispeels and Raikhel 1991; Suzuki et al. 1993). PHAs are toxic to mammals and birds. Their toxicity is thought to derive from binding to glycoproteins in the intestinal mucosa. αAI proteins inhibit digestive α-amylases in the intestine of mammals and larvae of certain bruchid (weevil) species. ARL proteins affect bruchid species, either because they are poorly digestible (Minney et al. 1990) or are toxic (Osborn et al. 1988). Thus, these three classes of APA proteins differ in their biochemical and cell biology characteristics. In turn,

these characteristics may relate to the resistance they confer to various predators.

Proteins of the three subfamilies are encoded in common bean by intronless genes of 732–819 bp in length, which show >70% nucleotide sequence identity. Dendrograms reveal three clusters or subfamilies of sequences corresponding to the three protein types (Mirkov et al. 1994). Upon sequence alignment, the genes coding for the three subfamilies can be distinguished primarily by three deletions (Rougé et al. 1993). Alpha-amylase inhibitors differ from lectins primarily by the deletion of three short surface loops whereas in ARL, only one of these loops is missing (Mirkov et al. 1994). αAI-like proteins differ from αAI proteins by the presence of an additional 15 bp in the former (Finardi-Filho et al. 1996). With three exceptions, the individual members of the APA family are linked in tandem in a complex locus (*Lec-αAI-Arl* or *APA* locus, henceforth) on linkage group B(ean)4 of the common bean molecular linkage map (Osborn et al. 1986; Chrispeels and Raikhel 1991; Suzuki et al. 1995; Freyre et al. 1998; Gepts 1999). The exceptions concern three other lectin genes [*Lec-2, Lec-3*, and *FRIL* (Flt3 receptor-interacting lectin)], which map to linkage group B7, but are only loosely linked (Nodari et al. 1993; Colucci et al. 1999; P. Gepts, unpublished results). The actual number of sequences, including pseudogenes, located at the *APA* locus is not known. However, four different PHA sequences have been described in different accessions of common bean (Hoffman and Donaldson 1985; Voelker et al. 1987; Mirkov et al. 1994). In common bean, eight DNA sequences have been described for arcelin, also in different accessions (Osborn et al. 1986; John and Long 1990; Mirkov et al. 1994; Goossens et al. 1994; Sparvoli and Bollini 1998; Acosta-Gallegos et al. 1998), five for αAI (Hoffman et al. 1982; Moreno and Chrispeels 1989; Mirkov et al. 1994; Ho et al. 1994; Suzuki et al. 1994), and one for an αAI-like protein (αAIL: Finardi-Filho et al. 1996). It is not known whether these different genes are alleles at the same locus or different sequences belonging to a multigene, complex locus, which has undergone successive duplications (Lioi et al. 2003). Furthermore, some of the differences among accessions may result from the existence of different stages in the structural evolution of the *APA* locus.

To further characterize the structural evolution of the *APA* locus, we have developed a bacterial artificial chromosome (BAC) library in each of four carefully chosen genotypes along an evolutionary lineage in common bean. In this article, we present a comprehensive description of our procedures and results, from the isolation of high-molecular-weight DNA and construction of BAC libraries to the characteristics of the resulting BAC libraries and the sequence of the *APA* locus contained in a 155 kbp BAC clone from accession G02771, which had been hypothesized to contain all three APA protein subfamilies based on existing literature reports.

Materials and methods

Nuclei isolation

The nuclei were isolated according to a novel procedure, modified from the method of Kuehl (1964). Plants for nuclei isolation were kept in the dark 3–5 days to reduce starch grain content of the cells. Young leaves were flash frozen in liquid nitrogen and stored at −70°C until needed. Nuclei isolation buffer (NIB) was prepared with gum arabic 4%, 10 mM MES, 100 mM sucrose, 10 mM EGTA, and 2 mM $ZnSO_4$. The pH was adjusted to 6.0 and 1 mM spermine, 4 mM spermidine, 4 mM *n*-octanol, 0.5% Triton X-100 and 25 mM *β*-mercaptoethanol were added after adjusting the pH. Gum arabic was purified before use using the following procedure. A 25% (w/v) solution was made by dissolving gum arabic in water overnight. The solution was centrifuged at 20,000*g* for 1 h to remove solid matter. The supernatant was collected and the pH adjusted to 4.6. An equal volume of 95% ethanol was added slowly to the solution to precipitate the gum arabic. After precipitation, the gum was allowed to settle overnight and the ethanol solution was decanted. The gum was redissolved in water to the original volume and the ethanol precipitation repeated. The precipitate was washed several times with 95% ethanol, twice with 100% ethanol and twice with anhydrous ethyl ether. The precipitate was then dried under vacuum. Twenty grams of leaf tissue were ground into a fine powder with a mortar and pestle equilibrated to −70°C. The ground tissue was shaken gently in 200 ml ice-cold NIB on ice for 20 min on an orbital shaker. The solution was filtered through four layers of cheesecloth, then through two layers of Miracloth™ into an ice-cold 250 ml centrifuge bottle. The filtrate was centrifuged for 15 min at 1,000*g* in a 0°C swinging bucket rotor. The supernatant was discarded and the pellet gently resuspended in 50 ml of ice-cold NIB buffer and transferred to a sterile 50 ml centrifuge tube. The solution was centrifuged for 10 min at 25*g* and the supernatant was carefully transferred to a fresh centrifuge tube with a serological pipette. The very soft pellet was discarded. The nuclei were concentrated by centrifugation for 10 min at 1,000*g* (1,800 rpm) and the pellet resuspended in ~45 ml of ice-cold NIB buffer without *β*-mercaptoethanol or Triton X-100. The centrifugation steps were repeated two or three times with ice-cold NIB without Triton X-100, *β*-mercaptoethanol, and *n*-octanol (–T–M–O). The final pellet was resuspended in 1 ml of ice-cold NIB –T–M–O. The quality of the nuclei preparation was checked under a microscope after staining with 1% methyl green/1% acetic acid stain (1 µl of stain/5 µl of nuclei preparation). The nuclei suspension was warmed to 50°C for one min and mixed with an equal volume of melted, 50°C, 1.5% agarose in 10 mM MES 500 mM sucrose, 50 mM EDTA pH 6.0. The solution was drawn into a 3 ml syringe and stored at 4°C until solidified. The top of the syringe was cut off

and the gel sliced into plugs of ~100 µl sections using the graduation markings on the syringe as a guide. The plugs were incubated for 1 h in 30 ml EPS (0.5 M EDTA, 1% sarcosine pH 9.6) with proteinase K (~20 μ/ml) followed by incubation for 48–72 h at 50°C with shaking with two changes of EPS and proteinase K. The plugs were then washed for 1 h with 1 × TE then washed three times for 1 h each with TE with 1 mM PMSF followed by three 1-h washes with TE without PMSF. Plugs were stored in TE at 4°C for use within 6 months. The plugs were washed with 0.5 × TE overnight before use. The quality of the DNA was checked by pulsed field gel electrophoresis (PFGE) using a 1% agarose gel in 0.5× TBE with a 1–40 s switch time at 6 V/cm for 20 h at 12°C.

BAC libraries

The BAC libraries were developed using the method of Peterson et al. (2000), with the following modifications. The agarose plugs were chopped into small fragments with a razor blade and subjected to partial digestion with *Hin*dIII as described, followed by two rounds of size selection. The first size selection was performed with a 0.5 s pulse for 2 h followed by a ramp of 1–40 s for 22 h. The second size selection was a 3–5 s ramp for 24 h. These changes were introduced to allow more time for smaller fragments to migrate further away from the larger fragments of interest. Finally, 20 g/ml BSA was added to the electroelution buffer to block potential non-specific binding sites on the dialysis tubing. The DNA was eluted from the agarose slices for 45 min at 6 V/cm at 4°C followed by reversing the current for 1 min. The DNA was recovered from the dialysis tubes with a wide bore pipet and use directly for ligation. The high molecular weight DNA was ligated into pIndigo-BAC cut with *Hin*d III (Epicentre, Madison WI) with a w/w ratio of 10:1 DNA to vector, in 100 µl volumes with a total of 100 ng DNA and 400 NEB units of T4 Ligase (New England Biolabs, Ipswich, MA) in the supplied buffer at 14°C overnight. The ligation solutions were dialyzed against 0.2× TE buffer for 2–4 h using drop dialysis on 0.025 µm VSWP membranes (Millipore Corp., Bedford, MA). The dialyzed ligations were stored on ice until use. The 2 µl of the ligation were electroporated into 15 µl of DH10B™ ElectroMax cells (Invitrogen, Carlsbad, CA) and plated on LB media containing 12.5 µg/ml chloramphenicol, 25 µM IPTG and 25 µg/ml X-Gal. The plates were incubated at 37°C for 18–24 h and then stored at 4°C. Individual clones were picked into 384-well microtiter plates and replicated using a Q-bot robot (Genetix) in the Genomics Facility of the College of Agricultural and Environmental Sciences at UC Davis. High-density filters were printed in a 4×4 CUGI (Clemson University Genomics Institute) standard pattern and hybridized with ^{33}P-labeled probes. BAC plasmid DNA was isolated by the double acetate method (http://www.genome.ou.edu/BAC_isoln_200ml_ culture.html). BAC ends were sequenced using the ABI Big Dye sequencing method by the sequencing facility in the Division of Biological Sciences at UC Davis. Filters were hybridized with chloroplast DNA clones to assess the number of inserts comprised of chloroplast DNA. Three clones were picked at random from each plate (approximately 400 clones or 0.8% of each library) and used to assess the average insert size. Clone insert size was determined by digestion with *Not*I followed by PFGE using a CHEFII apparatus and a 1% agarose gel in 0.5× TBE at 12°C with a 1–30 s switch time ramped over 14 h at 6 V/cm.

To identify APA-containing clones in the libraries, a probe was generated from PCR-amplified members of the APA family. Conserved regions, identified at the 5′ and 3′ regions of members of the APA family, were used to design a pair of APA primers (upstream, forward primer: 5′ ATGGCTTCCTCCAACTT; downstream, reverse primer: 5′ CAGAGGATGTTGTTGAGG) with cycle conditions of 95°C (2 min), followed by 92°C (30 s), 45°C (30 s), 72°C (30 s) for 35 cycles then 72°C (5 min). The bands between 750 and 900 bp were cut out of a 1.5% agarose gel, purified and labeled with ^{33}P using the random primer method (Feinberg and Vogelstein 1983, 1984). The labeled probe was used without further purification. Filters were hybridized for 16–24 h at 65°C in 7.5% SDS, 5× SSPE and washed three times with 2× SSC, 0.1% SDS at 65°C. X-ray film was exposed to the membranes for 24–72 h at −80°C. Putative positive clones were checked by PCR using the APA primers and Southern hybridization using PCR-generated fragments as a probe. Probes for the phaseolin locus were prepared from PCR fragments around the 15 bp duplication region as described (Kami et al. 1995).

Sequence analysis of an APA BAC clone

Sequencing of an APA gene family-containing BAC clone (71F18) from accession G02771, selected as described in the Results section, was performed to an average redundancy of 4–5× at the Purdue University Genomics Core Facility (Dubcovsky et al. 2001). Subclone sequences were assembled using the PHRED/PHRAP/CONSED package (http://www.phrap.org; Ewing et al. 1998; Ewing and Green 1998; Gordon et al. 1998). Finishing of the gaps was performed as recommended by the Autofinish function of CONSED. The final assembly was compared to restriction digests of the clone insert to verify the order and size of the contigs. The sequence of BAC clone 71F18 was then screened for open reading frames with GeneScan (http://www.genes.mit.edu/GENSCAN.html), TwinScan (http://www.genes.cs.wustl.edu/), and FGENESH (http://www.softberry.com/; trained with *Nicotiana tabacum*, *Medicago truncatula*, Dicot or Monocot). For the purpose of this analysis, each application of the FGENESH program trained with one of the three types of sequence data was considered as a different program. Regions

that were identified by three or more programs were subjected to BLAST searches of both the nucleotide and translated protein sequences. Those regions that were identified by three or more programs but did not have a significant BLAST result, were labeled as hypothetical proteins. Sequences that had an *e* value greater than e^{-20} in BAC 71F18 were identified by the named sequence with the highest *e* value in GenBank. Putative coding regions were compared to Genbank using BLAST and potential matches aligned with BioEdit and ClustalW. A phylogenetic tree of APA DNA sequences was constructed with BioEdit using the neighbor-joining method of Saitou and Nei (1987) and the UPGMA clustering algorithm. Bootstrap values were calculated with PAUP* (Version 4.0 b10 for 32 bit windows, http://www.lms.si.edu/ PAUP). The sequence of BAC clone 71F18 has been deposited with GenBank (DQ323045).

Results

BAC library construction

Hybridization of genomic DNA digested with several restriction enzymes with a probe made from a gel-purified PCR product of the APA genes showed a single, major band of approximately 170 Kbp when digested with *Nru*I (Fig. 1). The size of the band represents an upper limit for the size of the APA locus. This observation also indicated that the entire locus might be contained in one or at most two BAC clones. BAC libraries were therefore developed for each of four genotypes: BAT93 (domesticated Mesoamerican breeding line of *P. vulgaris*); G02771 (Gentry22274, PI318702; wild Mesoamerican accession of *P. vulgaris*); G21245 (DGD1962; wild, ancestral accession of *P. vulgaris* from northern Peru; Kami et al. 1995); and cv. Henderson (domesticated Mesoamerican cultivar of *P. lunatus*; Gutiérrez Salgado et al. 1995). These four accessions were hypothesized to harbor a different complement of the three subfamilies of APA based on the distribution of the corresponding proteins (Table 1). Henderson was included as an outgroup containing only phytohaemagglutinins (PHA^+; αAI^-; ARL^-). The three other accessions contain αAI, in addition to PHA, but they differ by the presence of ARL: G21245 and BAT93 are PHA^+; αAI^+; ARL^- whereas G02771 is PHA^+; αAI^+; ARL^+.

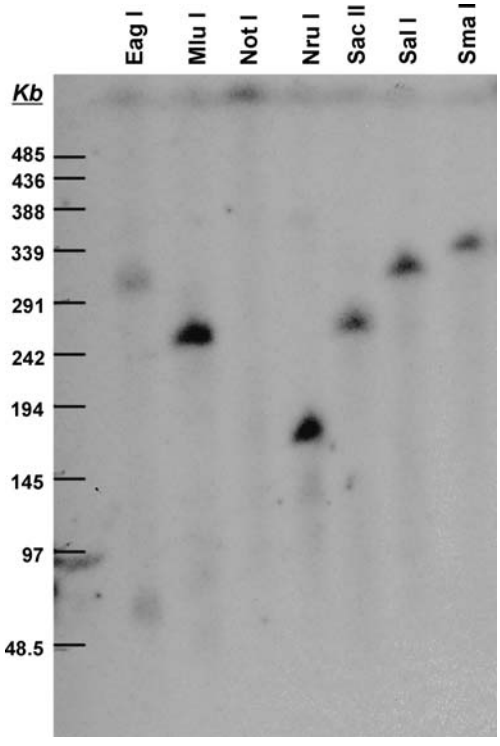

Fig. 1 Estimation of the size of the APA locus by digestion of genomic DNA of *P. vulgaris* with rare-cutting restriction enzymes and agarose electrophoresis. *Left-most lane* molecular weight markers. Note that the *Nru*I digest gives the smallest single fragment of approximately 170 Kbp

Isolation of high quality nuclei from beans using sucrose-based isolation media yielded inconsistent results in our hands, so a novel isolation method based on gum arabic (Kuehl 1964) was developed. The low pH (6.0) of this new extraction method, compared to the 9.4 pH of other methods (Zhang et al. 1995), did not seem to affect the isolation of nuclei from *Phaseolus*. This media, in combination with the differential centrifugation protocol, yielded large quantities of intact nuclei with a very low level of chloroplast contamination (as determined by light microscopy observation). This media worked well with all accessions of *Phaseolus* spp. tested, as well as cotton (data not shown). Under microscope examination, the nuclei appear round, intact and relatively undamaged. Although sometimes significant cell wall debris were observed, they did not appear to have any effect on subsequent procedures. DNA prepared from nuclei isolated using this procedure was of very high

Table 1 Putative phylogenetic distribution of members of the APA protein gene family

Taxonomic level/Gene pool		Lectin (PHA)[a]	αAI[b]	Arcelin
Subfamily Papilionoidea		+	–	–
Genus *Phaseolus*		+	–[c]	–
P. vulgaris complex[d]		+	+	–[e]
P. vulgaris	Andean	+	+	–
	Mesoamerican	+	+	+/–[f]

[a] Pusztai (1991)
[b] As defined by inhibition of porcine pancreas or *Tenebrio molitor* α-amylase inhibitory activity and immunological cross-reactivity with αAI of *P. vulgaris* (Pueyo and Delgado-Salinas 1997)
[c] Two wild species may have cross-reactive polypeptides without detectable activity: *P. microcarpus*, *P. hintonii* (Pueyo and Delgado-Salinas 1997)
[d] *P. vulgaris*, *P. polyanthus*, *P. coccineus*, and *P. acutifolius*
[e] *P. acutifolius* may have arcelin-like proteins but their activity is unknown (Mirkov et al. 1994)
[f] Only a few wild *P. vulgaris* accessions from Mesoamerica have arcelin proteins

molecular weight with a minimal amount of degradation and was easily digested by restriction enzymes.

The changes made to the protocol of Peterson et al. (2000) allowed us to improve the quality of the BAC libraries. In order to reduce the amount of smaller molecular weight DNA fragments following partial digestion, the PFGE run times were increased to allow more time for the small fragments to be resolved from the larger fragments. The result was, generally, a more narrow distribution of insert sizes in the final library with fewer undersized (defined as $<75\%$ of average clone size) clones (data not shown). The four libraries developed here had average insert sizes ranging between 105 and 139 kbp. The frequency of empty clones was below 0.5% for the four libraries and of cpDNA sequences-containing clones between 0.04 and 1.2%. The genome coverage was 8.7×, 12.1×, and 9.5× for the G21245, G02771, and Henderson libraries, respectively. For the BAT93 library, coverage was increased to 20.8× to allow for more detailed analyses in the future (Table 2).

Selection of APA and phaseolin sequence-containing BAC clones

An alignment of all *P. vulgaris* mRNA sequences in GenBank for α-amylase inhibitor, arcelin and phytohaemagglutinin was used to develop PCR primers that could amplify all known members of the family. Conserved regions were found at the 5′ and 3′ ends of the aligned sequences, which were used to design PCR primers (see Materials and methods) and amplify members of the APA family from the four genotypes used in the construction of the libraries. The PCR fragments were gel-purified and used as probes for screening the BAC libraries. The number of positive signals for each library is shown in Table 2. There was a relatively good agreement between the number of positive clones and the number of genomes in each library, with the exception of the G02771 library. G02771 is part of a group of wild Mesoamerican beans in which the most recent duplication and expansion of the APA family is thought to have occurred, namely the one generating the arcelin subfamily. To further investigate the APA gene content in G02771, clone 71F18 from the G02771 library was selected for complete sequencing because it was one of the larger-sized clones among the positive clones isolated. The 71F18 clone encompasses the APA locus on linkage group B4 based on the following evidence. Firstly, the only locus comprising all three subfamilies of the APA family is the *APA* locus (Osborn et al. 1986, 1988; Nodari et al. 1993). Secondly, probes derived from the sequenced BAC ends hybridized to clones that did not contain any detectable APA sequences, either through PCR or by direct hybridization. This suggested that the APA locus was contained entirely within this particular clone.

Phaseolin is the major seed storage protein of common bean. It is encoded by a single locus on linkage group B7 (Freyre et al. 1998). Reconstruction experiments show that this locus harbors an estimated six to nine sequences (Talbot et al. 1984). Thus, structurally and functionally, phaseolin provides an interesting comparison with APA. As in the case of the APA locus, there was generally a good agreement between the calculated number of genome equivalents and the number of putative positive clones (Table 2).

APA sequence analysis of BAC clone 71F18

Analysis of the sequence of the G02771 71F18 BAC clone revealed six regions with similarity to APA sequences (Fig. 2) These regions were designated APA-1 to APA-6 based on their consecutive (5′ → 3′) location in the BAC clone (Fig. 3). A tBLASTn search showed that each of these individual sequences most closely resembled one of the previously identified APA sequence subfamilies (Table 3). Thus, at least one representative of each of the major APA gene subfamilies is present in G02771, as predicted. To determine the evolutionary relationships between APA sequences identified in the 71F18 ABC clone, sequence fragments extending from 500 bp upstream of the putative ATG start site to 500 bp downstream of the putative polyadenylation site were used to construct a neighbor-joining tree (Fig. 4). The individual genes, grouped by similarity, fall into the following categories; APA-1 and APA-2 are most similar to the arcelin subfamily, APA-3, APA-4 and APA-5 are most similar to the phytohemagglutinin subfamily; and APA-6 is closest to the α-amylase inhibitor/inhibitor-like subfamily. All of the genes are potentially functional with the exception of the APA-1 gene. This

Table 2 Characteristics of BAC libraries produced in this project

Library	Clones (#)	Size (kbp)	Genome equivalents	Empties (%)	Chloroplast		APA (#)	Phaseolin (#)
					#	%		
P. vulgaris								
BAT93	110,592	125	20.8	<0.5	50	0.05	16	13
G21245	55,296	105	8.7	<0.5	200	0.4	10	11
G02771	55,296	139	12.1	<0.5	49	0.08	38	14
P. lunatus								
Henderson	55,296	~110	9.5	ND	645	1.2	ND	ND

ND not determined

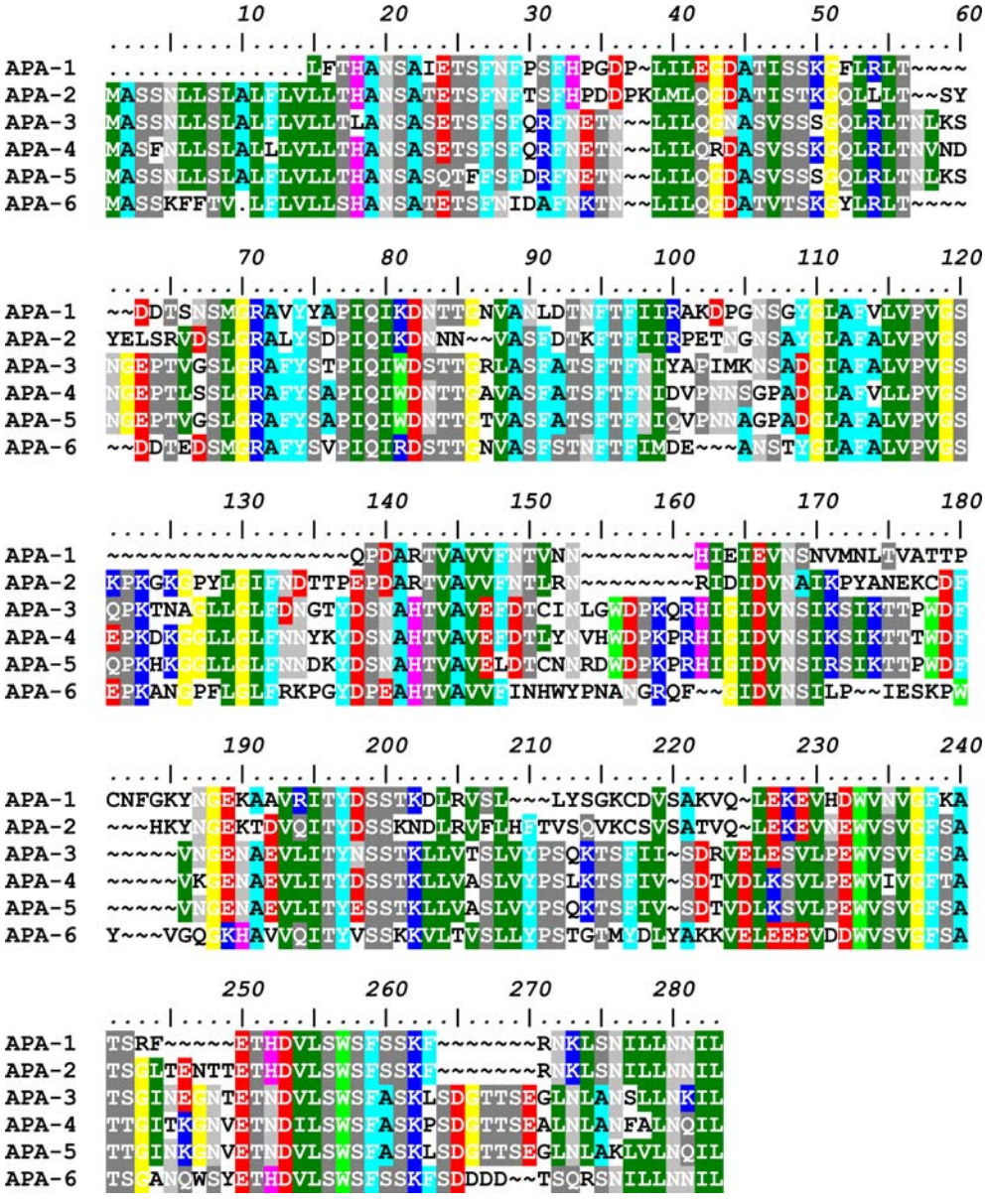

Fig. 2 Protein sequence alignment of the six APA sequences identified in *P. vulgaris* G02771 BAC clone 71F18

member of the arcelin group has a deletion at the 5′ start site of the gene. The deletion results in the loss of the start codon and the first 14 amino acids, which includes the signal peptide region. However, there is an in-frame ATG codon immediately upstream of the deletion; therefore, although this sequence might be a pseudogene, this designation should be confirmed as should the expression of the other APA sequences identified in this BAC clone. The sequence similarity between members of the APA family (Table 4) supports the hypothesis that the gene family arose through gene duplication and divergence.

In addition to the full length APA genes, several other regions of interest were found in the clone (Table 5). There is a small, 51 amino acid sequence that appears to be a truncated arcelin gene. Although the first 51 amino acids show a 58% identity to an arcelin gene, there is an early stop codon and the remaining downstream sequences fail to demonstrate a significant identity to any of the APA genes either at the nucleotide or at the amino acid level. Thus, if this is a pseudogene, it has undergone extensive mutation or rearrangement.

A number of fragments of different sizes were found that have sequence similarity with retrotransposons. These fragments range from short coding regions that appear to encode truncated portions of the Gag-Pol polyprotein to the largest fragment of 914 amino acids that includes both the reverse transcriptase and the Rnase H domains. Interestingly, this sequence similarity is only recognizable at the amino acid level as the nucleotide sequence failed to yield any significant matches in GenBank. It is also interesting to note that the three major APA groups are separated by retrotransposon elements or fragments.

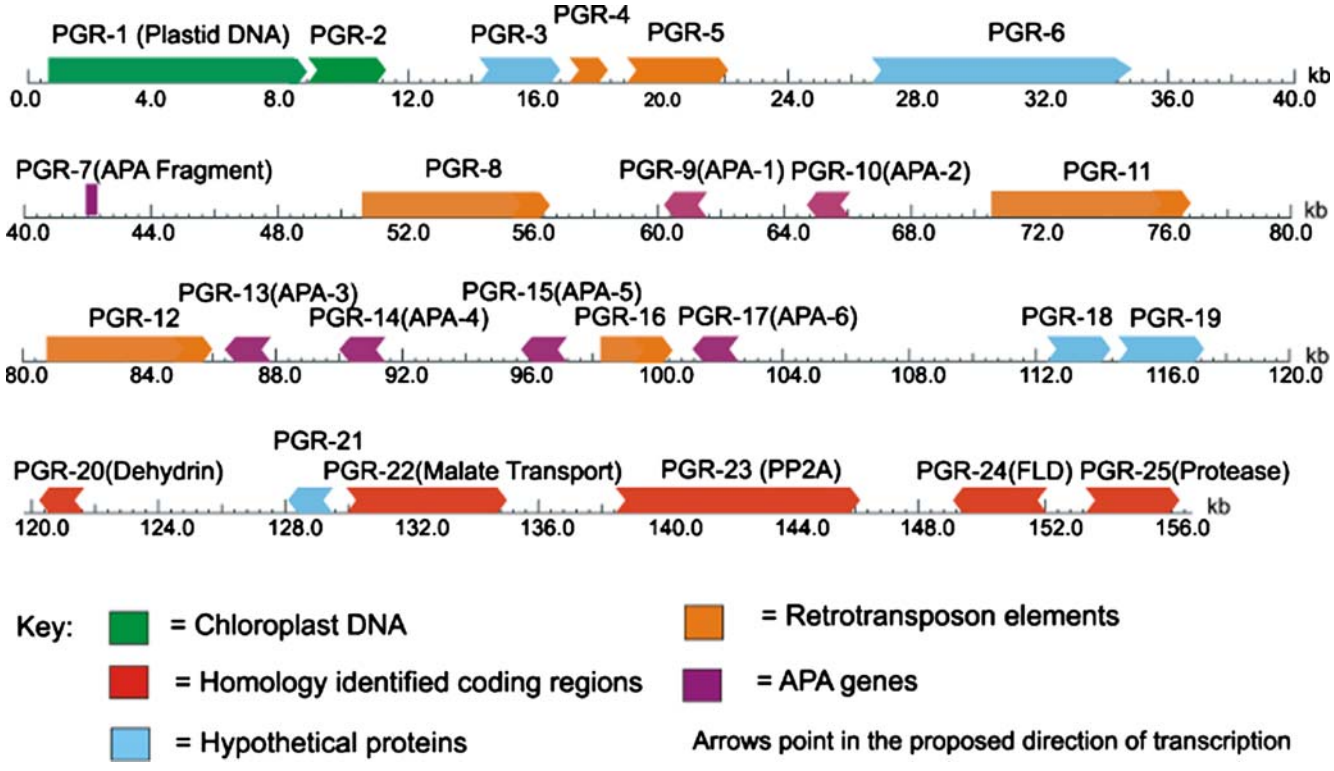

Fig. 3 Schematic of the distribution of putative genes identified in *P. vulgaris* G02771 BAC clone 71F18

Table 3 Results of a tBLASTn sequence similarity search of GenBank for each of the APA sequences identified in G02771 BAC clone 71F18

71F18 sequence	Most similar sequence in Genbank		
	Identifier	Gene	*e* value
APA-1	AF255723	*P. vulgaris* ARL5-IV pseudogene	0.0 (DNA)
APA-2	AF193029	*P. vulgaris* arcelin 5c	0.0 (DNA)
APA-3	PVU439715	*P. vulgaris* G12849 lec4-B17 gene	0.0 (Peptide)
APA-4	PVU439714	*P. vulgaris* G24591 PHA-E gene	0.0 (Peptide)
APA-5	U10349	*P. vulgaris* G02771 Arcelin 5 phytohemagglutinin	0.0 (Peptide)
APA-6	D49828	*P. vulgaris* Ofuku-5 mRNA for α-amylase inhibitor-like protein	0.0 (Peptide)

Other sequences in BAC clone 71F18

One of the more significant regions is the apparent presence of a plastid DNA region. This region spans approximately 8,100 bp with ~95% sequence identity to published sequences for the soybean, Arabidopsis and rice chloroplast genomes (Table 5). The sequences correspond to the soybean chloroplast DNA for rps12, rps7, 16S rRNA, tRNA-Val, NADH dehydrogenase and ORF 143. BLAST searches with this region matched the same nuclear-encoded plastid region as found in the Arabidopsis and rice genomes as well as genomic elements in *Lotus japonicus* (data not shown). Although there are some gaps between the nuclear-encoded plastid sequences and those from the chloroplast genome, the relative gene order is maintained. The high degree of sequence similarity of this plastid fragment to known chloroplast sequences implies that this is a recent insertion (Noutsos et al. 2005). Indeed, the overall degree of sequence conservation (Table 5) is higher between the plastid insertion and the chloroplast genome, about 90% at the nucleotide level, than between the APA genes themselves (Table 4).

Another coding region of interest includes a putative dehydrin gene. In cowpea [*Vigna unguiculata* (L) Walp.], this gene is expressed under dehydration stress conditions and is implicated in drought tolerance as well as freezing tolerance during emergence (Ismail et al. 1999; Hall 2004). The dehydrin gene of cowpea and the APA region of *P. vulgaris* are both linked to the same *P. vulgaris* RFLP marker D1711a on linkage groups 2 and B4 of cowpea and common bean, respectively (Nodari et al. 1993; Ouédraogo et al. 2002). A region with sequence similarity to a protein phosphorylase 2 A (PP2A) regulatory subunit was found downstream of the APA gene cluster. This gene shows 90% amino acid identity

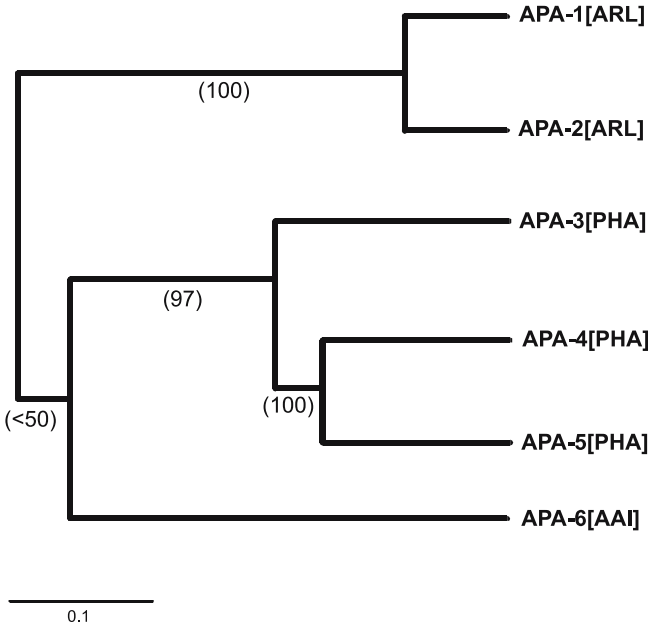

Fig. 4 Neighbor-joining tree of APA sequence diversity contained in BAC clone 71F18 isolated from *P. vulgaris* G02771. *Values at nodes* bootstrap values. *AAI* α-amylase inhibitor; *ARL* arcelin; *PHA* phytohaemagglutinin

to a sequence from *Medicago sativa* (Toth et al. 2000). The PP2A gene family has been implicated in the regulation and control of a number of critical cellular functions such as cell cycle control and cellular morphology and development. Although the precise function of this regulatory subunit is not known at this time, other PP2A regulatory subunits have been shown to be involved in signal transduction cascades. An additional coding region of interest is PGR-24. Analysis of the peptide encoded by this sequence revealed both a SWIRM domain (for SWI3p, Rsc8p, and Moira; Aravind and Lyer 2002) and an amino oxidase domain. This peptide shows a strong sequence similarity to the Flowering Locus D gene from Arabidopsis with 54% identity at the amino acid level. In Arabidopsis, this gene has been implicated in the control of flowering through chromatin remodeling by histone acetylation (He et al. 2003).

Discussion

The four libraries developed here are of high quality especially with regard to the large insert size and the very low frequency of empty and chloroplast DNA clones (Table 2). Phylogenetically, they complement two existing BAC libraries developed in Andean bean genotypes (cv. Sprite; Vanhouten and Mackenzie 1999) and G19833 (M. Blair, personal communication) and a BAC library developed in another Mesoamerican wild bean (G12946), which also contains arcelin (F. Sparvoli, personal communication). The development of seven phylogenetically arrayed BAC libraries in *Phaseolus* spp. represents a first among crop plants to our knowledge. These seven BAC libraries will provide an important tool to follow the evolution at the structural level of complex loci such as the APA and phaseolin loci, as well as disease resistance gene clusters such as the B4 disease resistance gene cluster, which includes resistance specificities against anthracnose [*Colletotrichum lindemuthianum* (Sacc. & Magn.) Briosi & Cav.] and rust [*Uromyces appendiculatus* (Pers.) Unger] (Geffroy et al. 2000; Kelly et al. 2003; Ferrier-Cana et al. 2005), and the *I* resistance gene cluster on linkage group B2 (Vallejos et al. 2004). The size of the APA and phaseolin loci (of the order of 50–200 kbp; present data and Llaca et al. 1996) requires large-insert libraries such as BAC libraries to include the entire locus in one to a few inserts. Our prior experience with the phaseolin locus showed that restriction digestion mapping has resolution limitations due to the duplicative nature of the loci resulting in multiple fragments with the same size (Llaca et al. 1996). Hence, large clone inserts that span loci are highly desirable to fully ascertain locus structure. BAC clones or contigs are also potentially useful to isolate marker sequences at known physical distances to determine local levels of linkage disequilibrium, for example, around domestication genes, which have been mapped in common bean (Koinange et al. 1996). In addition, the coverage of the BAT93 library was increased to double that of the other libraries (Table 2) as a first step towards the development of physical map and sequencing of the bean genome, in line with recent recommendations of the Phaseomics Initiative (Broughton et al. 2003) and the conference on Cross-legume Advances Through Genomics (Gepts et al. 2005). Further progress is needed in the sequencing of other legumes, such as soybean and *Medicago truncatula*, to examine microsynteny with these species (Young et al. 2005).

The order and sequence similarity of the members of the APA gene family in the 71F18 sequence supports our original supposition that the G02771 accession contains all three subfamilies. Because the lectin and lectin-like sequences are linked and show a high degree of sequence similarity (50–80% at the amino acid level), they are believed to have arisen by successive duplications of an ancestral gene (Osborn et al. 1986; Chrispeels and Raikhel 1991; Mirkov et al. 1994; Lioi et al. 2003). The proposed order of expansion of the gene family appears

Table 4 Sequence identity (%) between members of the APA family

		Nucleotide					
		APA-1	APA-2	APA-3	APA-4	APA-5	APA-6
Peptide	APA-1	–	73	60	60	62	67
	APA-2	42	–	71	72	72	75
	APA-3	29	28	–	90	92	74
	APA-4	27	28	63	–	93	74
	APA-5	30	31	71	74	–	74
	APA-6	38	36	35	32	32	–

Table 5 Putative gene regions (PGR) in BAC clone 71F18 of *P. vulgaris* accession G02771

PGR	Position (bp)	Direction[a]	DNA sequence similarity	Size (aa)	BLAST result	Organism	*e* value[b]	Genbank Accession No.
1	734–8,677	F	96% (4989/5227)		Soybean plastid DNA	*Glycine max*	0.0 d	CHGMX
2	8,892–10,914	F	94% (1636/1742)		Chloroplast DNA	*Lotus corniculatus*	0.0 d	AP002983.1
3	14,401–16,659	F	No Significant Match	689	No Significant Match			
4	17,621–18,229	F	No Significant Match	202	gag-pol polyprotein (Integrase core domain)	*P. vulgaris*	$6e^{-44}$ p	AAR13317.1
5	19,267–22,085	F	No Significant Match	914	gag-pol polyprotein (RT/RnaseH domain)	*P. vulgaris*	0.0 p	AAR13317.1
6	27,559–34,916	F	No Significant Match	671	No Significant Match			
7	41,372–41,527	F	97% (78/80)	51	Alpha-amylase inhibitor like protein	*P. vulgaris*	e^{-31} p	PVU10352
8	51,322–56,489	F	No Significant Match	1,120	Putative retrotransposon	*P. vulgaris*	e^{-153} p	AY341443.1
9	60,517–61,185	C	100% (666/666)	222	ARL5-IV pseudogene	*P. vulgaris*	0.0 d	AF255723
10	65,139–65,921	C	100% (780/780)	260	Arcelin 5c	*P. vulgaris*	0.0 d	AF193029
11	70,599–73,597	F	No Significant Match	846	Putative Retrotransposon	*Medicago truncatula*	$4e^{-128}$ p	AC148609.23
12	81,902–85,541	F	No Significant Match	440	Putative Retrotransposon fragment	*Vitis vinifera*	$2e^{-17}$ p	AB111100.1
13	86,714–87,542	C	99% (820/825)	275	G12849 lectin (lec4-B17 gene)	*P. vulgaris*	0.0 p	PVU439715
14	90,115–90,943	C	98% (813/825)	275	Pha-E gene for phytohemagglutinin	*P. vulgaris*	0.0 p	PVU439714
15	96,080–96,907	C	99% (822/825)	275	Pdlec1 for phytohemagglutinin	*P. vulgaris*	0.0 p	PVPDLEC1
16	98,435–10,0173	F	No Significant Match	408	Unknown repetitive element	*P. vulgaris*	7e−61 d	AY341443.1
17	101,634–102,422	C	99% (785/786)	262	Alpha-amylase inhibitor like protein	*P. vulgaris*	0.0 p	D49828.1
18	112,309–113,344	F	No Significant Match	191	Unknown protein	*Lotus corniculatus* var. *japonicus*	$2e^{-45}$ p	AP006422.1
19	114,928–117,606	F	No Significant Match	195	Unknown Protein			
20	120,501–121,208	C	89% (1201/1355)	259	CPRD22 protein, Dehydrin	*Vigna unguiculata*	e^{-94} p	D83972.1
21	128,408–128,974	C	82% (201/245)	188	Clone:LjT33E11, TM0310b	*Lotus corniculatus* var. *japonicus*	$4e^{-29}$ p	AP006422.1
22	130,857–134,987	F	No Significant Match	415	Putative malate transport protein	*Arabidopsis thaliana*	$4e^{-132}$ p	CNS0A4R2
23	138,632–148,159	F	91% (1369/1512)	560	Ser/Thr specific protein phosphatase 2PPa	*Medicago truncatula*	0.0 p	AF196287.1
24	148,201–151,827	C	67% (478/704) p	783	Lz-0 flowering locus D (FLD) gene	*Arabidopsis thaliana*	0.0 p	AY849997.1
25	152,924–155,450	F	No Significant Match	326	Aspartyl protease family protein	*Arabidopsis thaliana*	e^{-112} p	NP_564539.1

[a] *F* forward; *C* complementary
[b] *d* DNA similarity; *p* Protein similarity

to be a duplication from the phytohaemagglutinin core genes to generate the amylase inhibitor gene in one direction, followed by a more recent duplication in the other direction to form one of the arcelin genes. In addition to these primary duplications, there appears to be a series of more recent duplications of the arcelin gene to two members and of the phytohaemagglutinin genes to add either one or two additional members to this gene sub-family as suggested by the neighbor-joining tree (Fig. 2). Based on the distribution of the three gene subfamilies and their protein products among *Phaseolus* species, as currently inferred from surveys (summarized in Table 1), we hypothesized the following evolutionary scenario for this multigene family in the genus *Phaseolus*. The ancestral sequences of the APA multigene family are those coding for lectin proteins, which are distributed throughout the legume family and therefore preexisted the genus *Phaseolus*. After appearance of the genus *Phaseolus*, the αAI subfamily originated, immediately prior or during the divergence of the *P. vulgaris* complex as defined by Delgado-Salinas et al. (1999). Thus, *P. vulgaris, P. coccineus, P. dumosus* (previously *P. polyanthus*), and *P. acutifolius* have active αAI proteins in their seeds (Pueyo and Delgado-Salinas 1997) in addition to lectins. The relatedness of these species as indicated by the presence of active αAI proteins is consistent with other systematic data (Freytag and Debouck 2002; Llaca et al. 1994; Delgado-Salinas et al. 1999). The arcelin subfamily originated presumably after dispersal of the wild progenitor of *P. vulgaris* from what is now Ecuador and northern Peru into Mexico (Kami et al. 1995) as arcelin proteins have been found only among some wild populations of *P. vulgaris* in Mexico and not in other wild or domesticated accessions of the species either in Mexico or in the Andes, the two centers of domestication of the crop (Romero-Andreas et al. 1986; Osborn et al. 1988).

The APA multigene family therefore presents an interesting case of a rapidly diversifying gene family. Nevertheless, the processes underlying this diversification are not well understood, specifically with regard to the role of unequal crossing-over and functional diversification. Based on studies of individual members of the APA family, it appear that no single gene product is sufficient in providing complete protection against insect and mammalian predation. However, by duplication and divergence, the APA family has evolved multiple functionalities that complement each other. Small changes in the tertiary structure of the protein would allow the plant to affect multiple targets in a predator's digestive system (e.g., Santimone et al. 2004). This evolutionary strategy not only expands the number of potentially susceptible predators, but greatly reduces the rate at which a predator might evolve tolerance to this form of defense by requiring the predator to undergo several different rounds of mutation to develop tolerance. The unique genealogical structure of common bean, with an extant ancestor giving rise to two geographically distinct evolutionary lineages (Andean vs. Mesoamerica; Gepts 1998), each of which was domesticated provides an excellent model to examine the effect of differential selection pressures in the two lineages. Furthermore, the evolution of the APA locus can be contrasted with that of other complex loci in common bean and other species, such as disease resistance gene clusters (e.g., Melotto et al. 2004).

Acknowledgments BAC sequencing was performed at the Purdue Genomics Core Facility by Phillip San Miguel and Paul Parker. The project was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number #00–01972.

References

Acosta-Gallegos J, Quintero C, Vargas J, Toro O, Tohme J, Cardona C (1998) A new variant of arcelin in wild common bean, *Phaseolus vulgaris* L., from southern Mexico. Genet Res Crop Evol 45:235–242

Aravind L, Iyer LM (2002) The SWIRM domain: a conserved module found in chromosomal proteins points to novel chromatin-modifying activities. Genome Biol 3: research0039.0031–0039.0037

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

Chrispeels MJ, Raikhel NV (1991) Lectins, lectin genes, and their role in plant defense. Plant Cell 3:1–9

Colucci G, Moore J, Feldman M, Chrispeels M (1999) cDNA cloning of FRIL, a lectin from *Dolichos lablab*, that preserves hematopoietic progenitors in suspension culture. Proc Nat Acad Sci USA 96:646–650

Delgado-Salinas A, Turley T, Richman A, Lavin M (1999) Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (Fabaceae). Syst Bot 24:438–460

Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiology 125:1342–1353

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186–194

Ewing B, Hillier L, Wendl M, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:9–13

Feinberg AP, Vogelstein B (1984)A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity (addendum). Anal Biochem 137:266–267

Ferrier-Cana E, Macadre C, Sevignac M, David P, Langin T, Geffroy V (2005) Distinct post-transcriptional modifications result into seven alternative transcripts of the CC-NBS-LRR gene *JA1tr* of *Phaseolus vulgaris*. Theor Appl Genet 110:895–905

Finardi-Filho F, Mirkov T, Chrispeels M (1996) A putative precursor protein in the evolution of the bean α-amylase inhibitor. Phytochemistry 43:57–62

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

Freytag GF, Debouck DG (2002) Taxonomy, distribution, and ecology of the genus *Phaseolus* (Leguminosae—Papilionoideae) in North America, Mexico and Central America, edn. Botanical Research Institute of Texas, Forth Worth, TX

Geffroy V, Sévignac M, De Oliveira J, Fouilloux G, Skroch P, Thoquet P, Gepts P, Langin T, Dron M (2000) Inheritance of partial resistance against *Colletotrichum lindemuthianum* in *Phaseolus vulgaris* and co-localization of QTL with genes involved in specific resistance. Mol Plant-Microbe Interact 13:287–296

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

Gepts P (1999) Development of an integrated genetic linkage map in common bean (*Phaseolus vulgaris* L.) and its use. In: S. Singh (ed) Bean breeding for the 21st century. Kluwer, Dordrecht, pp 53–91, 389–400

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (2005) Legumes as a model plant family. Genomics for food and feed Report of the Cross-Legume Advances through Genomics Conference. Plant Physiol 137:1228–1235

Goossens A, Geremia R, Bauw G, Van Montagu M, Angenon G (1994) Isolation and characterisation of arcelin-5 proteins and cDNAs. Eur J Biochem 225:787–795

Gordon D, Abajian C, Green P (1998) Consed: A graphical tool for sequence finishing. Genome Res 8:195–202

Gutiérrez Salgado A, Gepts P, Debouck D (1995) Evidence for two gene pools of the lima bean, *Phaseolus lunatus* L., in the Americas. Genet Res Crop Evol 42:15–22

Hall AE (2004) Breeding for adaptation to drought and heat in cowpea. Eur J Agron 21:447–454

He Y, Michaels SD, Amasino RM (2003) Regulation of flowering time by histone acetylation in Arabidopsis. Science 302:1751–1754

Ho MF, Yin X, Finardi Filho F, Lajolo F, Whitaker JR (1994) Naturally occurring α-amylase inhibitors: structure/function relationships. In: Yada R, Jackman LR, Smith J (eds) Protein structure-function relationships in food Bishopbriggs. Blackie Academic and Professional, Glasgow, pp 89–119

Hoffman LM, Donaldson DD (1985) Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. EMBO J 4:883–889

Hoffman LM, Ma Y, Barker RF (1982) Molecular cloning of *Phaseolus vulgaris* lectin mRNA and use of cDNA as a probe to estimate lectin transcript levels in various tissues. Nucl Acids Res 10:7819–7828

Ismail AM, Hall AE, Close TJ (1999) Allelic variation of a dehydrin gene cosegregates with chilling tolerance during seedling emergence. Proc Nat Acad Sci USA 96:13566–13570

John ME, Long CM (1990) Sequence analysis of arcelin 2, a lectin-like plant protein. Gene 86:171–176

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

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

Koinange EMK, Singh SP, Gepts P (1996) Genetic control of the domestication syndrome in common-bean. Crop Sci 36:1037–1045

Kuehl R (1964) Isolation of plant nuclei. Z Naturforsch, B: Anorg Chem, Org Chem, Biochem Biophys Biol 198:525–534

Lioi L, Sparvoli F, Galasso I, Lanave C, Bollini R (2003) Lectin-related resistance factors against bruchids evolved through a number of duplication events. Theor Appl Genet 107:814–822

Llaca V, Delgado Salinas A, Gepts P (1994) Chloroplast DNA as an evolutionary marker in the *Phaseolus vulgaris* complex. Theor Appl Genet 88:646–652

Llaca V, Gepts P (1996) Pulsed field gel electrophoresis analysis of the phaseolin locus region in *Phaseolus vulgaris*. Genome 39:722–729

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

Minney B, Gatehouse A, Dobie P, Dendy J, Cardona C, Gatehouse A (1990) Biochemical bases of seed resistance to *Zabrotes subfasciatus* (bean weevil) in *Phaseolus vulgaris* (common bean)—a mechanism for arcelin toxicity. J Insect Phys 36:757

Mirkov TE, Wahlstrom JM, Hagiwara K, Finardi-Filho F, Kjemtrup S, Chrispeels MJ (1994) Evolutionary relationships among proteins in the phytohemagglutinin-arcelin-alpha-amylase inhibitor family of the common bean and its relatives. Plant Mol Biol 26:1103–1113

Moreno J, Chrispeels MJ (1989) A lectin gene encodes the α-amylase inhibitor of the common bean. Proc Natl Acad Sci USA 86:7885–7889

Nodari RO, Tsai SM, Gilbertson RL, Gepts P (1993) Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. Theor Appl Genet 85:513–520

Noutsos C, Richly E, Leister D (2005) Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Res 15:616–628

Osborn TC, Alexander DC, Sun SSM, Cardona C, Bliss FA (1988) Insecticidal activity and lectin homology of arcelin seed protein. Science 240:207–210

Osborn TC, Blake T, Gepts P, Bliss FA (1986) Bean arcelin. 2. Genetic variation, inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris*. Theor Appl Genet 71:847–855

Ouédraogo JT, Gowda BS, Jean M, Close TJ, Ehlers JD, Hall AE, Gillaspie AG, Roberts PA, Ismail AM, Bruening G, Gepts P, Timko MP, Belzile FJ (2002) An improved genetic linkage map for cowpea (*Vigna unguiculata* L.) combining AFLP, RFLP, RAPD, biochemical markers, and biological resistance traits. Genome 45:175–188

Peterson DG, Tomkins JP, Frisch DA, Wing RA, Paterson AH (2000) Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. J Agric Genomics 5: (http://www.ncgr.org/research/jag)

Pueyo JJ, Delgado-Salinas A (1997) Presence of α-amylase inhibitor in some members of the subtribe Phaseolinae (Phaseoleae: Fabaceae). Am J Bot 84:79–84

Pusztai A (1991) Plant lectins. Cambridge University Press, Cambridge, p 263

Romero-Andreas J, Yandell BS, Bliss FA (1986) Bean arcelin. 1. Inheritance of a novel seed protein of *Phaseolus vulgaris* L. and its effect on seed composition. Theor Appl Genet 72:123–128

Rougé P, Barre A, Causse H, Chatelain C, Porthe G (1993) Arcelin andalpha-amylase inhibitor from the seeds of common bean (*Phaseolus vulgaris* L.) are truncated lectins. Biochem Syst Evol 21:696–703

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Santimone M, Koukiekolo R, Moreau Y, Le Berre V, Rougé P, Marchis-Mouren G, Desseaux V (2004) Porcine pancreatic α-amylase inhibition by the kidney bean (*Phaseolus vulgaris*) inhibitor (α-AI1) and the structural changes in the α-amylase inhibitor complex. Biochim Biophys Acta 1696:181–190

Sparvoli F, Bollini R (1998) Arcelin in wild bean (*Phaseolus vulgaris* L.) seeds: sequence of arcelin 6 shows it is a member of the arcelins 1 and 2 subfamily. Genet Res Crop Evol 45:383–388

Suzuki K, Ishimoto M, Kikuchi F, Kitamura K (1993) Growth inhibitory effect of an alpha-amylase inhibitor from the wild common bean resistant to the Mexican bean weevil (*Zabrotes subfasciatus*). Japan J Breed 43:257–265

Suzuki K, Ishimoto M, Kitamura K (1994) cDNA sequence and deduced primary structure of an α-amylase inhibitor from a bruchid resistant wild common bean. Biochim Biophys Acta 1206:289–291

Suzuki K, Ishimoto M, Iwanaga M, Kikuchi F, Kitamura K (1995) Inheritanceof seed α-amylase inhibitor in the common bean and genetic relationships toarcelin. Theor Appl Genet 90:762–766

Talbot D, Adang M, Slightom J, Hall T.C. (1984) Size and organization of a multigene family encoding phaseolin, the major seed storage protein of *Phaseolus vulgaris* L Mol. Gen Genet 198:42–49

Toth EC, Vissi E, Kovacs I, Szoke A, Arino J, Gergely P, Dudits D, Dombradi V (2000) Protein phosphatase 2A holoenzyme and its subunits from *Medicago sativa*. Plant Mol Biol 43:527–536

Vallejos CE, Astua-Monge G, Jones V, Plyler T, Sakiyama N, Vanhoutten W, Mackenzie S (2004) Abstract 25003: Genetic and molecular characterization of the *I* locus of *Phaseolus vulgaris*. American Society of Plant Biologists Annual Meeting July 24–28, 2004, Lake Buena Vista, FL. http://www.abstracts.aspb.org/pb2004/public/M12/7715.html

Vanhouten W, Mackenzie S (1999) Construction and characterization of a common bean bacterial artificial chromosome library. Plant Molec Biol 40:977–983

Voelker T, Sturm A, Chrispeels M (1987) Differences in expression between seed lectin alleles obtained from normal and lectin-deficient beans are maintained in transgenic tobacco. EMBO J 6:3571–3577

Young ND, Cannon SB, Sato S, Kim D, Cook DR, Town CD, Roe BA, Tabata S (2005) Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiol 137:1174–1181

Zhang HB, Zhao X, Ding X, Paterson AH, Wing RA (1995) Preparation of megabase-size DNA from plant nuclei. Plant J 7:175–184