

# Genetic mapping of microsatellite markers around the arcelin bruchid resistance locus in common bean

Matthew W. Blair · Claritza Muñoz · Héctor F. Buendía · José Flower · Juan M. Bueno · César Cardona

Received: 20 November 2009 / Accepted: 2 March 2010 / Published online: 1 April 2010  
© Springer-Verlag 2010

**Abstract** The deployment in common beans (*Phaseolus vulgaris* L.) of arcelin-based bruchid resistance could help reduce post-harvest storage losses to the Mexican bean weevil [(*Zabrotes subfasciatus* (Boheman))]. Arcelin is a member of the arcelin-phytohemagglutinin- $\alpha$ -amylase inhibitor (APA) family of seed proteins, which has been extensively studied but not widely used in bean breeding programs. The purpose of this study was to evaluate microsatellite markers for genetic analysis of arcelin-based bruchid resistance and to determine the orientation of markers and the rate of recombination around the APA locus. A total of 10 previously developed microsatellites and 22 newly developed markers based on a sequenced BAC from the APA locus were screened for polymorphism and of these 15 were mapped with an F<sub>2</sub> population of 157 individuals resulting from a susceptible  $\times$  resistant cross of SEQ1006  $\times$  RAZ106 that segregated for both the arcelin 1 allele and resistance to the bruchid, *Z. subfasciatus*. Microsatellites derived from APA gene sequences were linked within 0.8 cM of each other and were placed relative to the rest of the b04 linkage group. In a comparison of genetic to physical distance on the BAC sequence, recombination was found to be moderate with a ratio of 125 kb/cM, but repressed within the APA locus itself. Several markers

were predicted to be very effective for genetic studies or marker-assisted selection, based on their significant associations with bruchid resistance and on low adult insect emergence and positions flanking the arcelin and phytohemagglutinin genes.

## Introduction

Common bean (*Phaseolus vulgaris* L.) is a widely adapted legume crop that is predominantly grown in the developing world and consumed mostly as dry grain (Broughton et al. 2003). Under these circumstances, common beans are usually stored in the farm by small producers or in market centers and towns by wholesalers or retailers where the beans suffer from a range of post-harvest problems, among which insect pests are the most serious (Cardona 2004). The most prevalent and severest of the storage pests affecting beans are weevils (Schoonhoven 1976). Two weevil species are predominant around the world: namely, *Zabrotes subfasciatus* (Boheman) and *Acanthoscelides obtectus* (Say) (Schoonhoven and Cardona 1980). These cosmopolitan insects belong to the bruchid family (Coleoptera: Bruchidae) and incur damage by reducing grain quality and hence marketability as well as negatively affecting seed germination.

Between the two major species, *Z. subfasciatus* is especially important in warm tropical regions below 1,000 masl (meters above sea level), while *A. obtectus* is the more common species in cooler climates. Another distinction between the two species is that *Z. subfasciatus* is only found in storage, while *A. obtectus* also lays its eggs on bean pods in the field (Schoonhoven 1976). Together, these insect pests are estimated to cause an average of 13% grain loss to bean crops in developing countries (Cardona and

---

Communicated by D. Lightfoot.

---

M. W. Blair (✉)  
International Center for Tropical Agriculture (CIAT),  
A. A. 6713, Cali, Colombia  
e-mail: m.blair@cgiar.org

C. Muñoz · H. F. Buendía · J. Flower · J. M. Bueno · C. Cardona  
International Center for Tropical Agriculture (CIAT),  
1380 N.W. 78th Ave., Miami, FL 33126, USA

Kornegay 1999). Both weevils can be controlled by improved post-harvest storage, beneficial predators or chemical and non-chemical treatment of the grain; however, these methods are not always practical or available to small farmers or wholesalers (Cardona 2004).

Genetic resistance on the other hand, has been found to be one of the best techniques to control weevils (Cardona et al. 1992; Cardona and Kornegay 1999) and is believed to be based on the seed protein arcelin, which was discovered in various wild accessions of common bean from Mexico that were effective in controlling *Z. subfasciatus*, but only somewhat effective or not effective for *A. obtectus* (Romero-Andreas et al. 1986; Osborn et al. 1988; Cardona et al. 1990; Sparvoli and Bollini 1998). Currently, there are seven variants of arcelin in common bean and several arcelin-like genes in closely related species, which differ in the level of resistance they provide against these insects (Osborn et al. 1986; Lioi and Bollini 1989; Santino et al. 1991; Mirkov et al. 1994; Acosta-Gallegos et al. 1998).

Arcelin and related proteins, including alpha amylase inhibitors and phytohemagglutinins, are all members of the APA family of lectins and lectin-related proteins that provide resistance to bruchids through antibiosis by reducing adult emergence, female fertility and insect growth, especially of first and second instar larvae weevils (Cardona et al. 1989). Some lectins from beans and other legumes are also anti-nutritional to humans (Vasconcelos and Oliveira 2004). The APA proteins are encoded by a multi-gene family of linked genes (Osborn et al. 1986) on linkage group b04 that probably evolved from duplication and divergence (Lioi et al. 2003; Kami et al. 2006), and some homologous genes are found in related legumes or other *Phaseolus* species (Chrispeels and Raikhel 1991; Sparvoli et al. 2001).

Structural analysis of the APA locus shows that the phytohemagglutinin genes are often duplicated, pseudogenes can be present for arcelin in arcelin-containing genotypes, and that partial lectin-related sequences are interspersed between some of the other members of the family (Goossens et al. 1995; Lioi et al. 2003) along with retrotransposons (Kami et al. 2006). Meanwhile, genetic analysis shows that the presence of arcelin protein is inherited as a monogenic dominant trait that provides higher resistance to bruchids when in the homozygous state with heterozygous Arc+/Arc− individual seeds less resistant than Arc+/Arc+ seed (Osborn et al. 1986; Kornegay et al. 1993). The precise mode of action of arcelin is not known, but there are indications that the protein is probably indigestible for the *Z. subfasciatus* (Minney et al. 1990). Researchers at the International Center for Tropical Agriculture (CIAT) have used the Arc1 gene widely in their breeding programs to create resistant breeding lines through backcrossing and gene transfer (Cardona et al. 1990); however, no arcelin-derived bruchid-resistant variety has ever been released.

The objective of this research was to develop microsatellites for the genetic analysis of arcelin-based bruchid resistance and determine the orientation of the APA locus and the rate of recombination around the Arc1 gene using an F<sub>2</sub> population of 157 individuals. A set of eight microsatellite markers were selected based on their position within APA genes or in the b04 linkage group, while 22 new microsatellite markers were designed from sequences at the APA locus and integrated with these previously developed markers. Finally, we aligned the full set of microsatellites to the sequence for the APA locus to estimate a genetic to physical distance ratio for the region.

## Materials and methods

### Plant material

A genetic mapping population was created from a susceptible × resistant cross consisting of 157 F<sub>2,3</sub> families. The parents used in this cross were SEQ1006 (bruchid susceptible, containing no known arcelin gene) and RAZ106 (bruchid resistant, containing the derived Arc1 gene). SEQ1006 is an advanced breeding line developed for drought tolerance and has pink mottled seed type; while RAZ106 was developed by backcrossing Arc1 into the purple mottled seed class from the Mesoamerican (Mexican) wild bean accession G12882 (Cardona et al. 1990). Both parents were of the Andean large-seeded type and lethal dwarfing (Hannah et al. 2007) was not a problem in the F<sub>1</sub> plants. The population was developed by selfing of the F<sub>1</sub> generation and harvesting the F<sub>2,3</sub> seed from individual F<sub>2</sub> plants. All seed production was carried out at the CIAT headquarters in Palmira, Colombia (3°N latitude, 24°C average yearly temperature).

### Bruchid resistance evaluations

Techniques for insect rearing and testing of genotypes for bruchid resistance were based on Schoonhoven et al. (1983). Briefly, bruchid resistance was evaluated by infesting 15 seeds of each F<sub>2,3</sub> family with nine pairs of adult *Z. subfasciatus* (Boheman) from an insect colony maintained at CIAT where seed and insect pairs were split evenly between three repetitions. Testing was carried out by placing the seed in a mesh-covered, clear plastic vial (9-cm high × 1.7 cm in diameter) and adding the adult insects. To avoid egg laying on the plastic surface of the vial rather than the bean seed coat, the walls of the chambers were previously covered with the rough side of sandpaper (No. 150) facing inward toward the chambers' interior. Experiments were conducted in a rearing chamber at 27°C and 70% RH with adults removed at 5 days after

infestation, upon which data were collected on the number of eggs laid. The number of emerged adults was then evaluated over a 70-day period and subsequently percentage adult emergence (PAE) was calculated based on the total number of adults emerged relative to the number of eggs laid. Averages were determined for the PAE variable and used for quantitative trait analysis.

#### DNA extraction

A standard organic solvent (phenol/chloroform)-based “miniprep” technique from Afanador and Haley (1993) was used for DNA extraction. Briefly, 0.5 g of young trifoliolate leaf tissue from greenhouse-grown plants was ground with a plastic pestle in a total of 800  $\mu$ l of extraction buffer (150 mM Tris-HCl, 15 mM EDTA, 1 M NaCl, 1.5% CTAB, 2%  $\beta$ -mercaptoethanol) and incubated at 65°C for 45 min before adding an equal volume of chloroform:octanol (24:1) mix, shaking for 30 min and precipitating the supernatant with 500  $\mu$ l of isopropanol and 125  $\mu$ l of sodium acetate at -20°C in a new eppendorf tube. The DNA was pelleted by centrifuging at 12,000 rpm for 10 min and cleaned with 500  $\mu$ l of 70% ethanol. The DNA pellets were then dried and resuspended in 150  $\mu$ l of ddH<sub>2</sub>O to a concentration of 50–100 ng/ $\mu$ l. DNA was then diluted to 5 ng/ $\mu$ l for use as template in PCR reactions.

#### Microsatellite selection, design and amplification

Three groups of markers were used for the study. One group of markers was selected because they targeted sequences of the APA gene family (BMd9, BMd15, BMd16, PV-ag004, PV-atct001, PV-atgc001 and PV-atgc002), while another group of markers was selected based on map location and proximity to the arcelin locus (BMd8, BMd26 and BMd30) on linkage group b04 (Yu et al. 2000; Blair et al. 2003). A final set of markers was developed based on a bacterial artificial chromosome (BAC) clone 71F18 sequence for the arcelin locus (Genbank entry DQ323045) from Kami et al. (2006). Within the first group, BMd9 was derived from the gene for D-Lec 2 phytohemagglutinin-L (X06336), BMd15 and PV-atgc001 from an erythroagglutinating phyto-hemagglutinin gene (K03288), BMd16 and PV-atgc002 from a leucoagglutinating phytohemagglutinin gene (K03289), PV-ag004 from a phytohemagglutinin pseudogene (X04660) and PV-atct001 from part of the arcelin gene itself (M68913).

Within the last group, the markers from the BAC sequence were designed with BatchPrimer 3 software (You et al. 2008) based on the FASTA sequence, and the criteria for selection of markers were that the primers pair should target any di-nucleotide, tri-nucleotide or tetra-nucleotide motifs with >5, 4 or 3 repeats, respectively, and have a

melting temperature of near 55°C. Once designed, microsatellites were located with Vector NTI (Invitrogen, Carlsbad, CA, USA) on the BAC sequence and the most evenly spaced markers were selected.

PCR conditions for all the previously developed markers were based on the thermocycling profile described in Blair et al. (2003) and varied only in the annealing temperatures as given in Table 1. Meanwhile, for the newly developed microsatellites, a touchdown profile was used according to the melting temperatures of the primer pairs as indicated in Table 2. MgCl<sub>2</sub> concentration for all the markers was at 1.5 mM. The microsatellites were run in 4% polyacrylamide (29:1 acrylamide : bisacrylamide) gels at 100 constant watts and 45°C constant temperature in Biorad Sequi-Gen GT vertical gel rigs for 1.5 h. Gels were silver-stained with a recirculating tank system as described in Blair et al. (2003). Alleles were identified by comparing to a molecular weight standard based on 10 and 25 bp ladders (Invitrogen).

#### Data analysis

Linkage analysis for the mapping population was based on the software application Mapmaker 2.0 (Lander et al. 1987), and the Kosambi function was used to convert recombination fraction to genetic distance. Group command was carried out with a minimum LOD of 3.0 and a maximum distance of 15 cM. Marker order was determined by multipoint analysis and confirmed with the ‘compare’ and ‘ripple’ commands. Chi-square tests for expected segregation ratios as well as the phenotypic effect of each marker locus on bruchid resistance evaluated with single-point regression analysis were carried out with the software program qGene (Nelson 1997).

## Results

#### Microsatellite polymorphisms

As a first step in determining the suitability of the microsatellites derived from or near to the APA locus, the full set of 32 markers (10 from previous publications and 22 developed for this study) was evaluated for amplification and polymorphism among the mapping population parents. A total of 28 of these microsatellites were found to amplify well in the parental survey, with 8 of these from Blair et al. (2003) and Yu et al. (2000), both at the APA locus and those near it, and 20 from the newly developed microsatellites (Tables 1, 2). The markers that did not amplify well from previous publications were BMd16 and PV-atgc002, leaving BMd9, BMd15, PV-ag004, PV-atct001 and PV-atgc001 as the best options for previously developed markers from the APA locus. Meanwhile, of the newly

**Table 1** Previously published microsatellite markers near or in the APA locus of common bean that were used for genetic mapping in the SEQ1006 × RAZ106 population

Marker	Reference	Motif	Associated gene (GenBank entry)	Expected allele size	$T_a$	Amplification pattern <sup>a</sup>	Parental polym. <sup>b</sup>
BMd8	Blair et al. (2003)	(CT) <sub>7</sub>	Gibberellin 20-oxidase (U70530)	176	55	Single	C <sup>c</sup>
BMd9	Blair et al. (2003)	(CATG) <sub>4</sub>	D-Lec 2 Phytohemagglutinin-L gene (X06336)	135	45	Multiple	D <sup>c</sup>
BMd15	Blair et al. (2003)	(AG) <sub>6</sub>	Erythroagglutinating phyto-hemagglutinin gene (K03288)	166	50	Single	D <sup>c</sup>
BMd16	Blair et al. (2003)	(CATG) <sub>4</sub>	Leucoagglutinating phytohemagglutinin gene (K03289)	136	50	NA	NA
BMd26	Blair et al. (2003)	(GAT) <sub>6</sub>	Alpha galactosidase (U12927)	141	50	Single	C <sup>c</sup>
BMd30	Blair et al. (2003)	(TTAA) <sub>3</sub>	Cytosolic glutamine synthetase (U32582)	134	50	Single	M
PV-ag004	Yu et al. (2000)	(AG) <sub>8</sub>	Phytohemagglutinin pseudogene (X04660)	201	55	Multiple	D <sup>c</sup>
PV-atct001	Yu et al. (2000)	(ATCT) <sub>3</sub>	Arcelin gene (M68913)	193	50	Multiple	D <sup>c</sup>
PV-atgc001	Yu et al. (2000)	(ATGC) <sub>4</sub>	Erythroagglutinating phyto-hemagglutinin gene (K03288)	126	55	Multiple	D
PV-atgc002	Yu et al. (2000)	(ATGC) <sub>4</sub>	Leucoagglutinating phytohemagglutinin gene (K03289)	144	55	NA	NA

$T_a$  annealing temperatures in °C used in PCR reaction, C co-dominant polymorphic, D dominant polymorphic, M monomorphic, NA no amplification

<sup>a</sup> Amplification pattern refers to number of bands

<sup>b</sup> Polymorphism in the parental survey

<sup>c</sup> Indicates if marker had clear amplification product in the population and was used for mapping

designed markers, only BMarc11 and BMarc19 were unsuccessful. Of the markers from outside the APA locus, BMd8, BMd26 and BMd30 amplified consistently.

Polymorphism was evident in 6 of the previously reported markers and 16 of the newly designed markers that amplified well when comparing arcelin-containing versus susceptible parents. The only monomorphic markers were BMd30 and BMarc1, BMarc2, BMarc18 and BMarc20. Therefore, the polymorphism rate for these microsatellites was 85.7 and 80.0%, respectively, for these two sets of markers overall.

Of the markers described above, several from the phytohemagglutinin genes in the middle of the APA locus (genes APA3, 4, 5 and 6) gave multiple rather than single bands. These included the markers PV-atct001, PV-atgc001, PV-ag004 and BMarc9, which were all interpreted as dominant loci. The remaining markers produced unique bands; however, some markers were co-dominant (BMarc7, BMarc16, BMarc17 and BMarc21) producing two different alleles for the resistant and susceptible genotypes, while other markers (BMarc3, BMarc4, BMarc5, BMarc6, BMarc8, BMarc9, BMarc10, BMarc12, BMarc14 and BMarc15) were missing a band in the susceptible genotype and were considered to represent dominant loci as well. Three markers (BMd9, BMd15 and BMarc13) had bands present in the susceptible, but not in the resistant, genotype with the first two useful for mapping (Table 3). Both markers that were outside the APA locus (BMd8 and BMd26) were co-dominant.

Among the newly designed markers, the microsatellites were mostly based on di-nucleotide repeat motifs (11). Meanwhile, fewer tri-nucleotide (4) and tetra-nucleotide (5) targeting microsatellites were developed. One marker

(BMarc21) was part of the coding sequence of the 2-PPa gene, while others were from non-coding sequences. These additional markers targeted a wider range of motifs and were better distributed across the APA locus than the established BMd and PV markers, which mainly targeted tetra-nucleotide repeats (6) with few di-nucleotide- (3) or tri-nucleotide- (1) based markers. In addition, the new markers were more randomly distributed across the physical length of the APA locus based on the sequence for this region from Kami et al. (2006), as shown in Fig. 1.

The locations of the microsatellite markers on the BAC clone were determined according to Blastn comparisons in the case of the markers from Yu et al. (2000) and Blair et al. (2003), or based on their sequence positions in the case of the new markers. When considering the first group of markers, Pv-atct001 was adjacent to the gene APA 1 (a pseudogene for arcelin) and the markers BMd9, BMd15 and Pv-ag004 were located between the loci APA3, APA4, APA5 and APA6 according to the gene nomenclature of Kami et al. (2006). Meanwhile, among the new markers, BMarc4, BMarc5, BMarc10 and BMarc12 were located near the APA1/APA2 (arcelin) genes along with PV-atct001, while the markers BMarc-14, BMarc15, BMarc16 and BMarc17 were near the APA3/APA4/APA5 and APA6 (PHA) genes along with Pv-ag004 according to the annotation from Kami et al. (2006). In contrast, BMarc22 was near the FL-D gene.

Genetic and physical mapping of the microsatellite markers

A total of 15 of the microsatellites surrounding the arcelin gene gave clear segregation patterns for genetic mapping

**Table 2** Newly developed microsatellite markers derived from the BAC sequence of clone 71F18 (GenBank entry DQ323045)

Marker	Forward primer	$T_m$ For	Reverse primer	$T_m$ Rev	BAC position	Motif	Expected allele size	TD	Parental polymor. <sup>a</sup>
BMarc1	GCTAGGCTTCTCTATTCA	52.2	AAGAAAAACCTACCTTTGG	50.3	14078–14240	(CT) <sub>5</sub>	162	55–45	M
BMarc2	AGGGAAGGAAAGAAAAAGA	50.7	AACCTTCTCTACCTCAACC	53.7	15599–15785	(GCT) <sub>5</sub>	186	56–46	M
BMarc3	TGAGAGGAGTCAAAAAACCAT	51.3	GAAAGACCTACACGCAACTTT	52.8	25473–25757	(AT) <sub>5</sub>	284	56–46	D
BMarc4	TTATCACCATCACCCAAAGTT	51.4	CCTTCCGTGTTGTTGTGTT	51.8	39303–39455	(ATAC) <sub>3</sub>	152	56–46	D <sup>b</sup>
BMarc5	TCTGTGTTATGTGGTATCAG	49.6	GCCTAATGCAAGTTATCCAA	50.5	42458–42737	(TTA) <sub>9</sub>	279	55–45	D <sup>b</sup>
BMarc6	CCGATCAATACATAGTATGCTG	50.7	CAACCCACCAATTTACTCT	51.9	45308–45486	(ATGA) <sub>3</sub>	178	56–46	D
BMarc7	GTCTTCTGCAGGTACCATCT	54.0	CCAAAGAACTATTGGGATCA	49.6	49306–49457	(CCGA) <sub>3</sub>	151	55–45	C
BMarc8	CGTGAGACGAGTCAATAACC	52.9	GAAAGACCTACACGCAACTTT	52.8	69383–69675	(AT) <sub>7</sub>	292	58–48	D
BMarc9	TTCGAGGTGACGTAAGAATT	51.2	TCAACAAGGATGGAGAAGAT	50.8	69774–69963	(AAT) <sub>7</sub>	189	56–46	D
BMarc10	CGAGCATCGAATACCTTTAC	50.8	GGATCACTCTTCTCTTCTTCT	51.6	64136–64311	(CGAT) <sub>3</sub>	175	56–46	D <sup>b</sup>
BMarc11	TAAATCGCCACAGAAGGA	50.7	GGGGGTTTTCTTGTAAAG	50.9	77734–77927	(GT) <sub>5</sub>	193	56–46	NA
BMarc12	GAAAACGATGCACACAATC	49.9	TATGTTGTACCATCGGTTGA	51.0	75269–75468	(ATAA) <sub>3</sub>	199	55–45	D <sup>b</sup>
BMarc13	TGTAGGTACCATTCGAGTCC	53.4	CACAAAATTGTTCTCCATCG	49.1	84872–85024	(GACC) <sub>3</sub>	152	54–44	D
BMarc14	GGAAGACACATGTGAAGAGG	52.9	CTTCCCTCACATTTCCACTG	52.6	88292–88454	(GCAT) <sub>5</sub>	162	58–48	D <sup>b</sup>
BMarc15	AAGAGGGCTAGGGAGAGTAA	54.2	GCCATCGTTGCTTAAATTTCT	50.7	91408–91578	(CT) <sub>6</sub>	170	56–46	D <sup>b</sup>
BMarc16	TCATGCATTAACCATCACTC	50.2	GCCATCGTTGCTTAAATTTCT	50.7	97416–97560	(CT) <sub>7</sub>	144	55–45	C <sup>b</sup>
BMarc17	CTTCTGAGCTGAAGTAAAACG	51.3	AGTGAGACAGAAAGGCAAAAT	51.8	113764–114010	(AATT) <sub>3</sub>	246	56–46	C <sup>b</sup>
BMarc18	CTAGCCAGACCTCAGACATC	54.3	GCATGAATCTTCTCGTGACT	52.4	125636–125796	(GA) <sub>5</sub>	160	57–47	M
BMarc19	TATGTCCTACGAGGGTTGAT	52.4	CCCGTTGAAAAGATCAAAGT	50.3	136669–136833	(AT) <sub>12</sub>	164	55–45	NA
BMarc20	AGCTGAAGAACGAGGACAT	53.5	GAGAAGCACTTCAATCGTCAAT	52.4	139381–139527	(CG) <sub>5</sub>	146	57–47	M
BMarc21	AGGACACATGGACATGGATA	53.0	AAGAAGGGGATGAAAATGTCT	51.1	141812–142037	(AT) <sub>11</sub>	225	56–46	C <sup>b</sup>
BMarc22	GAGTTTTGGGGAGACATTG	50.9	ATTTGAGAGCGAAAAGATCAC	50.4	152436–152580	(GGC) <sub>6</sub>	144	55–45	C <sup>b</sup>

TD touchdown profile with upper and lower annealing temperatures in °C, C co-dominant polymorphic, D dominant polymorphic, M monomorphic, NA no amplification

<sup>a</sup> Polymorphism parental survey

<sup>b</sup> Indicates if marker had clear amplification product in the population and was used for mapping



**Table 3** Phenotypic variance (RSq) and significance (*F* statistic, LOD and *P* value) of the association between *Zabrotes subfasciatus* bruchid resistance as measured by percentage of adult emergence (PAE) and markers in the APA region of linkage group b04 in the SEQ1006 × RAZ106 common bean population along with segregation ratio and distance on genetic map

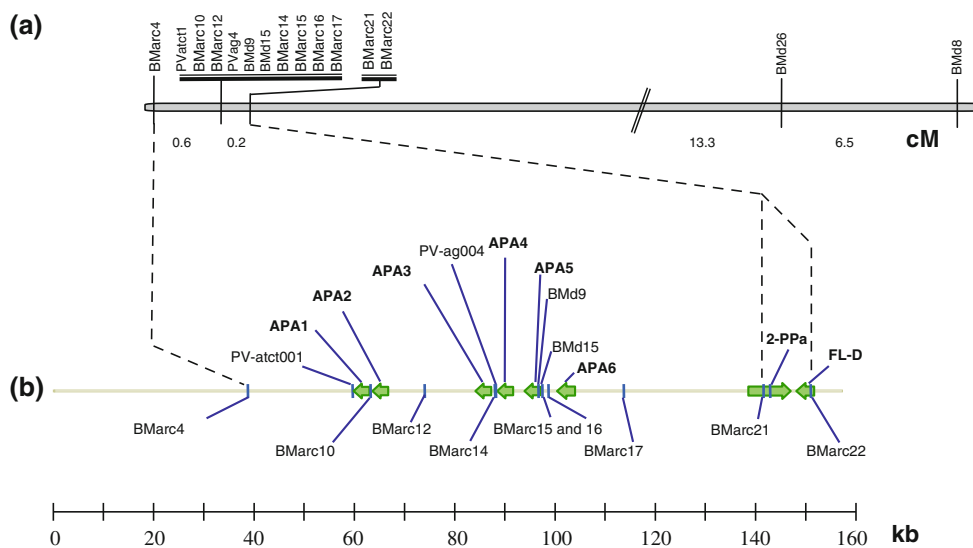
Marker name	Segregation ratio	Distance (cM)	Single-point regression			
			<i>F</i>	RSq	LOD	<i>P</i>
BMarc4	1:3	0.0	737.1	0.8328	58.25	0.0000
PV-atct001	1:3	0.6	283.4	0.6662	34.31	0.0000
BMarc10	1:3	0.6	709.5	0.8274	57.22	0.0000
BMarc12	1:3	0.6	460.6	0.7568	46.05	0.0000
BMd15	3:1	0.6	84.5	0.3683	14.66	0.0000
BMd9	3:1	0.6	59.1	0.2855	10.95	0.0000
PV-ag004	1:3	0.6	569.7	0.8027	50.05	0.0000
BMarc14	1:3	0.6	858.0	0.8529	62.42	0.0000
BMarc15	1:3	0.6	630.8	0.8100	54.09	0.0000
BMarc16	1:2:1	0.6	267.8	0.6441	33.65	0.0000
BMarc17	1:2:1	0.6	317.0	0.6817	37.29	0.0000
BMarc21	1:2:1	0.8	356.5	0.7067	39.95	0.0000
BMarc22	1:3	0.8	852.7	0.8529	62.02	0.0000
BMd26	1:2:1	14.1	96.1	0.3937	16.3	0.0000
BMd8	1:2:1	20.6	47.4	0.2425	9.05	0.0000

based on polymorphism across the mapping population. Of these, six were from Blair et al. (2003) and Yu et al. (2000) and nine were newly designed markers from the BAC sequences (BMarc4, BMarc10, BMarc12, BMarc14,

BMarc15, BMarc16, BMarc17, BMarc21 and BMarc22). Segregation data were used to produce the genetic map in Fig. 1, which shows the positions of the microsatellite markers aligned with the physical map for the region.

Tight clustering was observed for 13 markers related to the APA locus, all located in 0.8-cM interval between BMarc4 and BMarc21/BMarc22. Among these markers, BMarc4 separated from a cluster of ten co-segregating markers (BMarc10, BMarc12, BMarc14, BMarc15, BMarc16, BMarc17, BMd9, BMd15, PV-atct001 and PV-ag004), as did the markers BMarc21 and BMarc22, which co-segregated by themselves. Meanwhile, the two markers that were not part of the APA gene sequences, BMd26 and BMd8, were located at 13.3 and 19.8 cM, respectively, from the last of the APA locus markers, BMarc21 and BMarc22. Segregation distortion was not evident for any of the markers analyzed in the population, with dominant markers fitting a 3:1 or 1:3 ratio and co-dominant markers fitting a 1:2:1 ratio based on Chi-square tests.

The alignment of the physical and genetic maps for the APA locus was used to determine the physical to genetic distance ratios likely to exist for the region. For example, given that the majority of the APA locus (from BMarc4 to BMarc22) covered approximately 100 kb of physical space on the BAC within 0.8 cM on the genetic map, the ratio was estimated to be 125 kb/cM on average for the region. This varied for subsections of the locus, especially between BMarc4 and PVatct001 where a 0.6-cM interval covered 20 kb compared to the interval between BMarc17 and BMarc21 where 0.2 cM covered 27 kb, resulting in a physical



**Fig. 1** Genetic (a) and physical (b) maps for the region surrounding the arcelin locus on linkage group b04 of common bean based on microsatellite linkage analysis in the SEQ1006 × RAZ106 population (as described in the text) and sequence and gene annotation for BAC 71F18 from GenBank entry DQ323045. Scale shown for physical map

(in kilobase) with genetic distance (in CentiMorgans) indicated in marker intervals. Established microsatellite markers (PV or BMd) are shown above the physical map, while new microsatellite markers (BMarc) are shown below the physical map

to genetic distance ratios of 33 and 135 kb/cM, respectively. No recombination was found within the region between PV-atct001 and BMarc17, which is specifically the region containing the APA gene family covering a physical distance of 73 kb. These results indicated that recombination was uneven around the APA locus.

#### Marker association with bruchid resistance

Phenotypic data for the population showed a highly significant association of the markers in the region of the APA locus with bruchid resistance as measured by the percentage of *Z. subfasciatus* adult emergence (PAE). Single-point regression analysis was carried out with the PAE values and both the APA locus and flanking markers. The first observation was that in all cases, the resistance was associated with the wild-derived RAZ allele for ARC1 rather than the susceptible allele from the SEQ parent.

Secondly, the most significant markers were in the cluster of markers near the APA1 and AP2 loci with  $R^2$  values between 0.76 and 0.83 and LOD scores between 46.1 and 58.3, followed by those near the APA3 and APA5 markers with similar  $R^2$  and LOD values. Meanwhile, the distantly linked markers, BMd26 and BMd8, had successively lower  $R^2$  values and LOD scores. Specifically, these latter two markers at 13.5 and 20.0 cM from the BMarc17 marker had  $R^2$  values of 0.39 and 0.24, respectively. Therefore, the greater the distance of the marker from the arcelin locus, the lower was the phenotypic variance explained by that marker.

Finally, the genotypes containing two copies of the Arc1 locus from the RAZ source had as little as 6.4% adult bruchid emergence compared to those genotypes that were homozygous for the SEQ parental allele for this locus, which had as much as 95.9% adult bruchid emergence. Individuals that were heterozygous for the APA locus had on average 25% adult bruchid emergence, suggesting partial dominance.

## Discussion

#### Distinctive microsatellites found in the APA region

A set of highly polymorphic markers for the arcelin gene and other parts of the APA locus was designed and tested for this study. Out of the 32 markers evaluated, 28 amplified well and of these 22 were polymorphic between the resistant and susceptible parents, equivalent to an overall polymorphism rate of 78.6%. The reason that the polymorphism rate was so high was because we targeted a large number of different simple sequence repeat types and compared the introgression from a wild (non-cultivated)

Mesoamerican gene pool genotype used to breed into the resistant RAZ106 line against the cultivated Andean background line SEQ1006. In addition, the majority of the markers were from genomic regions outside the protein-coding sequences found at the APA locus, although some were closely linked either upstream or downstream of several of the members of the APA gene family.

The region of the APA locus appears to be especially rich in microsatellite sequences. For example, a total of 48 simple sequence repeat loci were found in the BAC 71F18 sequences, though only 22 were selected for development. Since that BAC is approximately 157-kb long, the frequency of microsatellite loci was one every 3.3 kb and the new microsatellites were designed to be one every 7 kb on average. This may be higher than the frequency throughout the rest of the genome as estimated in an evaluation of non-enriched small-insert libraries of common bean by Blair et al. (2009). The high frequency of microsatellites in the region could be associated with the retrotransposons found among some of the APA genes. It would be interesting to evaluate additional genomic sequences from various cultivated and wild beans with or without arcelin for the number of SSR motifs around the APA locus. With the sequences analyzed here or with additional sequences for the region, it is to be expected that only a certain proportion of microsatellite loci would be useful for marker development due to the sequence context around the repeat motif. In this study, some SSR that appeared to be good for marker design did not make good microsatellites due to poor PCR amplification. This occurred in 2 of the 22 markers designed for the APA locus, with 17 of the remaining markers polymorphic between arcelin-containing resistant versus susceptible genotypes.

Another interesting point was the presence of multiple bands for some of the markers that probably indicate that the markers detected duplicated loci. This was to be expected for some of the phytohemagglutinin-based markers (BMd9, BMd15, BMd16 and PV-ag004), because this protein is known to be encoded by a gene family with repeated or highly similar multiple copy sequences (Gossens et al. 1995; Lioi et al. 2003; Kami et al. 2006). In addition, multiple copies of arcelin-like sequences and pseudogenes have been found in different genotypes as was evident for APA1 (pseudogene) and APA2 (arcelin). Multiple banding was also found when testing new markers for the BAC sequence from the BMarc series where we found similar SSRs and flanking sequences at multiple sites within the BAC. This was notable for the markers BMarc7, BMarc8, BMarc10, BMarc15 and BMarc16, all of which are located within the APA3 to APA6 interval and were associated with repeated sequences.

It is also interesting to note the large number of microsatellites in the arcelin and phytohemagglutinin members

of the APA family, but the lack of microsatellites in the alpha amylase gene. On further distinguishing the microsatellites from this region, a majority of the repeat motifs found within this central part of the APA locus seem to be based on tetra-nucleotide repeats, rather than di or tri-nucleotide repeats, and may also be associated with the retrotransposons found there or with activity of retroelements.

Another noticeable feature of markers from the regions was that while some of the markers were co-dominant, principally those outside the APA locus; many were dominant, especially those near the APA locus or representing the APA gene family members. This may reflect the fact that genes in the region could be present in the resistant arcelin-containing allele, but absent in the cultivated background. For example, the evaluation of the parents showed dominant bands for most of the markers within the immediate region of the APA genes, except for PV-atct001 and co-dominant bands only for flanking markers, BMarc16, BMarc17, BMarc21 and BMarc22, as well as the distantly linked markers, BMd8 and BMd26. The existence of many dominant markers within the APA locus that amplify bands only in the RAZ106 parent or segregants with the RAZ106 allele may indicate that insertion/deletion events have occurred at the APA locus. Both Mirkov et al. (1994) and Lioi et al. (2003) found the duplication of certain APA genes in resistant genotypes versus loss of specific APA genes in susceptible genotypes through sequencing of APA alleles. Few other cases of missing genes have been investigated in common bean, but these are found frequently on comparing alleles of other species.

#### Linkage relationships of microsatellites at the APA locus

Genetic mapping of the microsatellites confirmed that most of the markers from the APA locus were closely linked and elucidated the map order of the other microsatellites in the region. The close genetic linkage of BMd9, BMd15, PV-ag004, PV-atct001 and six of the newly developed BAC-derived markers (BMarc10, BMarc12, BMarc14, BMarc15, BMarc16 and BMarc17) was not surprising given that they were linked physically between the APA1 and APA6 genes. Meanwhile, the flanking marker BMarc4 was found 20 kb and 0.6 cM away from the APA1 gene, while markers BMarc21 and BMarc22 co-segregated and were found 0.2 cM away from the APA6 locus though they were separated by a physical distance of approximately 9 kb.

Apart from the genetic mapping per se, we were able to determine the probable orientation of locus sequenced by Kami et al. (2006) by comparing our genetic map with the physical map based on the BAC71F18 sequence. Alignment to the mapping carried out by Blair et al. (2003) shows that APA genes are likely to be oriented with BMarc4 toward the distal end of the linkage group, and

BMarc21 and BMarc22 toward the centromere of the corresponding chromosome. This can be concluded from the placement of the linked and proximal markers, BMd26 and BMd8, to the same side of the APA locus as BMarc21 and BMarc22. The position of BMd8 and BMd26, according to Blair et al. (2003), was in the middle of the linkage group compared to the markers BMd9, BMd15, PV-ag004 and PV-atct001, which were at the end of the linkage group.

#### Physical to genetic distance ratio at the APA locus

The alignment of the physical and genetic maps presented here for the APA locus was used to determine the physical to genetic distance ratios likely to exist in the region. An average of 125 kb/cM was found in the region, but varied from 33 to 135 kb/cM depending on the interval and no recombination or evidence of crossing over was found directly within the interval from APA1 to APA6.

Variability in crossing over can be due to chromosome context and structure in euchromatin and gene-dense regions outside the centromeres versus heterochromatin or gene-poor regions, especially those near centromeres (Lichten and Goldman 1995; Mezard 2006). Furthermore, recombination can be high within tandem repeated gene families (Meyers et al. 2005) and perhaps at seed protein loci (Spielmeyer et al. 2000), although we did not find this to be the case in our mapping of the APA locus.

The amount of recombination at the APA locus could be influenced by the genetic background cross analyzed in this study. In this regard, both the resistant and susceptible lines used to develop the mapping population were Andeans with type I growth habit and red mottled seed, so that they may share alleles across the genome except at and around the APA locus. This may increase recombination outside the APA locus, but limit it within the APA locus itself.

Furthermore, the possibility of indels between different alleles of the APA locus might limit crossing over to adjacent regions outside the APA genes themselves in resistant  $\times$  susceptible crosses as observed in this study. We can conclude that recombination within the APA locus appears to be repressed in a population derived from genotypes present and absent in arcelin, while recombination around the APA locus appears to be somewhat above average in the same population compared to the whole genome of common bean where genetic to physical map distances are higher than 350 kb/cM according to Vallejos et al. (2006).

#### Microsatellites as markers for arcelin genes

While the most important objective of this work was to evaluate the genetics of the APA locus as described in the previous paragraphs, the markers identified in this study



could facilitate the pyramiding of bruchid resistance with other biotic and abiotic stress resistances and the further study of Arc1. The selection of the best microsatellites from this study to use in further analysis of bruchid resistance would depend mainly on: (1) their association with bruchid resistance with those closest to arcelin and with high  $R^2$  values being ideal and (2) on their ease of use, with co-dominant markers generally better for selection than dominant ones (Miklas et al. 2006). In this sense, the most reliable markers would be the co-dominant markers from the APA region such as BMarc16 or BMarc17, although dominant markers such as PV-atct001, BMd9, BMd15, BMarc10, BMarc14 and BMarc15 or flanking markers still with highly significant association with bruchid resistance such as BMarc4 and BMarc21 or BMAarc22 could also be used.

Finally, the wide range of markers within and between each of the APA genes could be useful for selection of recombinants within the APA locus itself, and it would be interesting to see if the block of APA family genes near the Arc1 gene could be broken up during introgression by multiple backcrosses, inbreeding and recombination. This process would be useful for the creation of genetic stocks with or without certain APA genes, while selection with the flanking markers would be useful for marker-assisted backcrossing into a wide range of cultivated backgrounds, which could be used to create arcelin-containing lines for each commercial class of Andean or Mesoamerican beans. Isolines would be useful for confirming the high level of resistance provided by the Arc1 gene found in this study and could expand on results from Nishizawa et al. (2007) where an Arc1 isolate was more resistant than Arc2, Arc3, Arc4 and Arc5 isolines. Utility of the Arc1 gene against *Z. subfasciatus* agrees with the early evaluations of the arcelin alleles from Osborn et al. (1988), Cardona et al. (1990) and Kornegay et al. (1993). In contrast, Arc5 allele may be less useful based on artificial feeding from Goossens et al. (2000), and both the Arc1 and Arc5 alleles appear not to be effective in transgenic experiments with tepary bean transformation conducted by Zambre et al. (2005).

Despite their utility in the analysis of Arc1, we should point out that the markers developed here may be more useful for this arcelin gene than for other gene variants. Specifically, polymorphisms near Arc1 may not be found for other Arc genes given the duplications and deletions evident at the APA locus of different genotypes (Lioi et al. 2003). Overall, the bruchid-resistance markers developed here are unique because most previous DNA markers have been developed for disease resistance, while very few are available for the analysis of insect resistance (Frei et al. 2005; Blair et al. 2006). Insect resistance is notably more difficult to evaluate than disease resistance due to the high chances for escape in field or artificial testing and the need

to maintain live colonies, specialized staff and appropriate infestation conditions (Yencho et al. 2000). Furthermore, the fact that we were able to find markers that flanked each side of the arcelin gene, several polymorphic markers very near the arcelin gene itself and various other APA gene family members means that they can be used for diagnosing these genes in various different cross combinations.

**Acknowledgments** We are grateful to Sergio Prieto for initiating work on some of these markers and to Agobardo Hoyos and Alcides Hincapie for field support. This work was partially financed by the Generation Challenge Program and by CIAT core funds.

## References

- Acosta-Gallegos JA, Quintero C, Vargas DJ, Toro O, Tohme J, Cardona MC (1998) A new variant of arcelin in wild common bean *Phaseolus vulgaris* L. from southern Mexico. *Crop Evol* 45:235–242
- Afanador L, Haley S (1993) Adoption of a “mini-prep” DNA extraction method for RAPD’s marker analysis in common bean *Phaseolus vulgaris*. *Ann Rep Bean Improv Coop* 36:10–11
- Blair MW, Pedraza F, Buendía HF, Gaitán-Solís E, Beebe SE, Gepts P, Tohme J (2003) Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 107:1362–1374
- Blair MW, Muñoz C, Garza R, Cardona C (2006) Molecular mapping of genes for resistance to the bean pod weevil (*Apion godmani* Wagner) in common bean. *Theor Appl Genet* 112:913–923
- Blair MW, Muñoz M, Pedraza F, Giraldo MC, Buendía HF, Hurtado N (2009) Development of microsatellite markers for common bean (*Phaseolus vulgaris* L.) based on screening of non-enriched small insert genomic libraries. *Genome* 52:772–782
- Broughton WJ, Hernandez G, Blair MW, Beebe SE, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.)—model food legumes. *Plant Soil* 252:55–128
- Cardona C (2004) Common beans: Latin America. In: Hodges R, Farrell (eds) *Crop post-harvest: science and technology*. Durables, vol 2. Blackwell, Oxford, pp 145–150
- Cardona C, Kornegay J (1999) Bean germplasm resources for insect resistance. In: Clement S, Quisenberry S (eds) *Global plant genetic resources for insect-resistant crops*. CRC Press, Boca Raton
- Cardona C, Posso CE, Kornegay J, Valor J, Serrano M (1989) Antibiosis effects of wild dry bean accessions on the Mexican bean weevil and the bean weevil (Coleoptera: Bruchidae). *J Econ Entomol* 82:310–315
- Cardona C, Kornegay J, Posso C, Morales F, Ramirez H (1990) Comparative value of four Arcelin variants in the development of dry bean lines resistant to the Mexican bean weevil. *Entomol Exp Appl* 56:197–206
- Cardona C, Dick K, Posso CE, Ampofo K, Nadhy SM (1992) Resistance of a common bean (*Phaseolus vulgaris* L.) cultivar to post-harvest infestation by *Zabrotes subfasciatus* (Bohemian) (Coleoptera: Bruchidae). II. Storage tests. *Trop Pest Manag* 38:173–175
- Chrispeels MJ, Raikhel NV (1991) Lectins, lectins genes and their role in plant defense. *Lectin Rev* 1:183–194
- Frei A, Blair MW, Cardona C, Beebe SE, Gu H, Dorn S (2005) QTL mapping of resistance to *Thrips palmi* Karny in common bean (*Phaseolus vulgaris* L.). *Crop Sci* 45:379–387
- Goossens A, Ardiles Diaz W, De Keyser A, Van Montagu M, Angenon G (1995) Nucleotide sequence of an arcelin 5-I genomic clone from wild *Phaseolus vulgaris* (accession no. Z50202) (PGR 95–075). *Plant Physiol* 109:722

- Goossens A, Quintero C, Dillen W, De Rycke R, Valor JF, De Clercq J, Van Montagu M, Cardona C, Angenon G (2000) Analysis of bruchid resistance in the wild common bean accession G02771: no evidence for insecticidal activity of arcelin 5. *J Exp Bot* 51:1229–1236
- Hannah MA, Krämer KM, Geffroy V, Kopka J, Blair MW, Erban A, Vallejos CE, Heyer AG, Sanders FET, Millner PA, Pilbeam DJ (2007) The DL gene system in common bean (*Phaseolus vulgaris* L.) causes programmed root death due to a shoot-derived inhibitory signal. *New Phytol* 176:537–549
- Kami J, Poncet V, Geffroy V, Gepts P (2006) Development of four phylogenetically-arrayed BAC libraries and sequence of the APA locus in *Phaseolus vulgaris*. *Theor Appl Genet* 112:987–998
- Kornegay J, Cardona C, Posso CE (1993) Inheritance of resistance to Mexican bean weevil in common bean, determined by bioassay and biochemical tests. *Crop Sci* 33:589–594
- Lander ES, Green P, Abrahamson J, Barlow A, Daly M, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174–181
- Lichten M, Goldman ASH (1995) Meiotic recombination hotspots. *Ann Rev Genet* 29:423–444
- Lioi L, Bollini R (1989) Identification of a new arcelin variant in wild bean seeds. *Ann Rep Bean Improv Coop* 32:28–29
- 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
- Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. *Curr Opin Plant Biol* 8:129–134
- Mezard C (2006) Meiotic recombination hotspots in plants. *Biochem Soc Trans* 34:531–534
- Miklas PN, Kelly JD, Beebe SE, Blair MW (2006) Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. *Euphytica* 147:105–131
- Minney BHP, Gatehouse AMR, Dobie P, Dendy J, Cardona C, Gatehouse JA (1990) Biochemical bases of seed resistance to *Zabrotes subfasciatus* (bean weevil) in *Phaseolus vulgaris* (common bean): a mechanism for arcelin toxicity. *J Insect Physiol* 36:757–767
- 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
- Nelson JC (1997) QGENE: software for marker-based genomic analysis and breeding. *Mol Breed* 3:229–235
- Nishizawa K, Teraishi M, Utsami S, Ishimoto M (2007) Assessment of the importance of  $\alpha$ -amylase inhibitor 2 in bruchid resistance of wild common bean. *Theor Appl Genet* 114:755–764
- Osborn TC, Blake T, Gepts P, Bliss F (1986) Bean arcelin. Part 2. Genetic variation, inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris* L. *Theor Appl Genet* 71:847–855
- Osborn TC, Alexander DC, Sun S, Cardona C, Bliss F (1988) Insecticidal activity and lectin homology of arcelin seed protein. *Science* 240:207–210
- Romero-Andreas J, Yandell B, Bliss F (1986) Bean arcelin Part 1. Inheritance of a novel seed protein of *Phaseolus vulgaris* L. and its effect on seed composition. *Theor Appl Genet* 72:123–128
- Santino S, Valesina L, Lioi A, Vitale A, Bollini R (1991) Bean (*Phaseolus vulgaris* L.) seed lectins: a novel electrophoresis variant of arcelin. *Plant Physiol* 10:7–11
- Schoonhoven A (1976) Pests of stored beans and their economic importance in Latin America. In: Proceedings of the symposium on tropical stored product entomology, 15th international congress of entomology, 19–27 August 1976, Washington DC, Entomological Society of America, College Park, MD, pp 691–698
- Schoonhoven A, Cardona C (1980) Insects and other bean pest in Latin America. Storage insects. In: Schwartz HF, Galvez GE (eds) Bean production problems. Centro Internacional de Agricultura Tropical CIAT, Cali–Colombia, pp 363–412
- Schoonhoven A, Cardona C, Valor J (1983) Resistance to the bean weevil and the Mexican bean weevil (Coleoptera: Bruchidae) in non-cultivated common bean accessions. *J Econ Entomol* 76:1255–1259
- Sparvoli F, Bollini R (1998) Arcelin in wild bean (*Phaseolus vulgaris* L.) seeds: sequence of arcelin 6 shows it is a member of the arcelin 1 and 2 subfamily. *Genet Res Crop Evol* 45:383–388
- Sparvoli F, Lanave C, Satucci A, Bollini R, Lioi L (2001) Lectin and lectin related proteins in lima bean (*Phaseolus lunatus* L.) seeds: biochemical and evolutionary studies. *Plant Mol Biol* 45:587–597
- Spielmeier W, Moullet O, Laroche A, Lagudah ES (2000) Highly recombinogenic regions at seed storage protein loci on chromosome 1DS of *Aegilops tauschii*, the D-genome donor of wheat. *Genetics* 155:361–367
- Vallejos CE, Astua-Monge G, Jones V, Plyler TR, Sakiyama NS, Mackenzie SA (2006) Genetic and molecular characterization of the *I* locus of *Phaseolus vulgaris*. *Genetics* 172:1229–1242
- Vasconcelos IM, Oliveira JTA (2004) Antinutritional properties of plant lectins. *Toxicon* 44:385–403
- Yencho GC, Cohen MB, Byrne PF (2000) Applications of tagging and mapping insect resistance loci in plants. *Ann Rev Entomol* 45:393–422
- You MF, Huo NY, Gu QY, Luo M, Ma Y, Hane D, Lazo RG, Dvorak J, Anderson D (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9:253–266
- Yu K, Park SJ, Poysa V, Gepts P (2000) Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). *J Hered* 91:429–434
- Zambre M, Goossens A, Cardona C, Van Montagu M, Terryn N, Angenon G (2005) A reproducible genetic transformation system for cultivated *Phaseolus acutifolius* (teparty bean) and its use to assess the role of arcelin in resistance to the Mexican bean weevil. *Theor Appl Genet* 110:914–924