Theor Appl Genet (2010) 121:1103–1116
DOI 10.1007/s00122-010-1375-9

ORIGINAL PAPER

Syntenic relationships among legumes revealed using a gene-based genetic linkage map of common bean (*Phaseolus vulgaris* L.)

Melody McConnell · Sujan Mamidi · Rian Lee · Shireen Chikara · Monica Rossi · Roberto Papa · Phillip McClean

Received: 20 October 2009 / Accepted: 21 May 2010 / Published online: 6 July 2010

Abstract Molecular linkage maps are an important tool for gene discovery and cloning, crop improvement, further genetic studies, studies on diversity and evolutionary history, and cross-species comparisons. Linkage maps differ in both the type of marker and type of population used. In this study, gene-based markers were used for mapping in a recombinant inbred (RI) population of *Phaseolus vulgaris* L. *P. vulgaris*, common dry bean, is an important food source, economic product, and model organism for the legumes. Gene-based markers were developed that corresponded to genes controlling mutant phenotypes in *Arabidopsis thaliana*, genes undergoing selection during domestication in maize, and genes that function in a biochemical pathway in *A. thaliana*. Sequence information, including introns and 3′ UTR, was generated for over 550 genes in the two genotypes of *P. vulgaris*. Over 1,800 single nucleotide polymorphisms and indels were found, 300 of which were screened in the RI population. The resulting LOD 2.0 map is 1,545 cM in length and consists of 275 gene-based and previously mapped core markers. An additional 153 markers that mapped at LOD <1.0 were placed in genetic bins. By screening the parents of other mapping populations, it was determined that the markers were useful for other common Mesoamerican × Andean mapping populations. The location of the mapped genes relative to their homologs in *Arabidopsis thaliana* (At), *Medicago truncatula* (Mt), and *Lotus japonicus* (Lj) were determine by using a tblastx analysis with the current pseudochromosome builds for each of the species. While only short blocks of synteny were observed with At, large-scale macrosyntenic blocks were observed with Mt and Lj. By using Mt and Lj as bridging species, the syntenic relationship between the common bean and peanut was inferred.

Communicated by C. Schön.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1375-9) contains supplementary material, which is available to authorized users.

M. McConnell · S. Mamidi · R. Lee · S. Chikara · P. McClean (✉)
Genomics and Bioinformatics Program,
North Dakota State University, Fargo, ND 58105, USA
e-mail: phillip.mcclean@ndsu.edu

M. McConnell · S. Mamidi · R. Lee · S. Chikara · P. McClean
Department of Plant Sciences,
North Dakota State University, Fargo, ND 58105, USA

M. Rossi · R. Papa
Dipartimento di Scienze Ambientali e delle Produzioni Vegetali,
Università Politecnica delle Marche, Via Brecce Bianche,
60131 Ancona, Italy

Introduction

Common bean, *Phaseolus vulgaris* L. (Pv), is an important food source, not only as staple food and major source of calories and proteins in many developing countries, but also as a healthy addition to any diet. It is a major source of calories and protein in many developing countries throughout the world (FAO: http://faostat.fao.org/faostat). For countries such as Burundi and Rwanda, common bean provides about 15% of the total daily calories and 30% of the daily protein intake. Common bean is a rich source of zinc and iron, two micronutrients depleted in individuals with AIDS (Savarino et al. 1999; Buys et al. 2002). Diets containing foods rich in these micronutrients are suggested to benefit the health status of HIV-infected patients by reducing malnutrition (American Dietetic Association 2004;

Kruzich et al. 2004), and beans are specifically recommended for these patients (South Africa Department of Health 2001). The immediate benefit is that improved nutrition will reduce the food security risk (Gillespie and Kadiyala 2005). When placed in this context, the value of common bean is best seen through its role as a societal crop, and its improvement is of continual concern (Singh 2001).

P. vulgaris originated in the Americas, and a variety of studies over many years indicate that there are two major gene pools that diverged prior to domestication (Gepts 1998). These are generally referred to as the Mesoamerican and Andean gene pools. The Mesoamerican gene pool generally refers to beans found in Colombia, Venezuela, Central America, and Mexico, while the Andean gene pool principally refers to beans found in Peru, Bolivia, and Argentina. These two groups can be distinguished on the basis of morphology (Gepts and Debouck 1991), as well as using molecular markers (e.g. Becerra-Velasquez and Gepts 1994; Tohme et al. 1996; Kwak and Gepts 2009; Rossi et al. 2009). This organization is detected at both the wild and landrace levels (Gepts 1998; Becerra-Velasquez and Gepts 1994). Domestication also appears to have occurred independently in the two gene pools, with the Andean domestication at 4,000 years ago predating the Middle American by 2,000 years (Kaplan and Lynch 1999; Piperno and Dilehay 2008). Sequence diversity among the two gene pools is evident, while diversity is greater among Middle American than Andean genotypes (McClean et al. 2004; McClean and Lee 2007). This diversity has been exploited for mapping purposes.

As with any species, common bean benefits from the availability of molecular linkage maps. These maps are important for discovering genes of agronomic or economic importance, breeding, and studies on diversity and evolutionary history. They have applications in marker-assisted selection and candidate gene cloning. Fine mapping can be performed near genes of interest to allow further and more detailed genetic studies. Candidate genes can be discovered for breeding purposes and for cloning and gene analysis. Linkage maps can also be used to compare the macro- and microsynteny of different species. Molecular maps differ in the number and type of markers used. RFLP, AFLP, and RAPD markers are often used to create molecular maps, and recently, simple sequence repeat (SSR) markers have also become a prominently used type of molecular marker.

A collection of over 25 linkage maps have been developed from crosses both within and between the two common bean gene pools (reviewed in Kelly et al. 2003). Freyre et al. (1998) developed a recombinant inbred (RI) population based on a cross between BAT93, a Middle American cultivar, and Jalo EEP558, an Andean landrace (hereafter referred to as the BJ population) and used it to extend an RFLP map originally based on this same cross (Nodari et al. 1993a). This BJ map was subsequently used to integrate other linkage maps (Freyre et al. 1998). The BJ population is generally considered to be a community-wide mapping population that is used by multiple research groups (Nodari et al. 1993b; Miklas et al. 2000, 2001, 2003; Park et al. 2001; Schneider et al. 2001; Tar'an et al. 2001; McClean et al. 2002; Kelly et al. 2003; Kolkman and Kelly 2003; López et al. 2003; Kelly and Vallejos 2004; Murray et al. 2004; Papa et al. 2005; Román-Avilés and Kelly 2005; Papa et al. 2007).

Gene-based molecular markers are emerging as important tools for molecular mapping. First, maps based on SSR markers found within EST sequences have been constructed for maize (Vigoroux et al. 2002), bread wheat (Gao et al. 2004), and soybean (*Glyciine max*, Gm; Song et al. 2004). More recently, EST data were used to develop gene-based maps exhibiting single nucleotide polymorphisms (SNPs) in sunflower (Lai et al. 2005) and pea (Aubert et al. 2006). These functional maps are important for candidate gene analysis, marker-assisted selection, and synteny studies (Choi et al. 2004).

DNA sequences are available for many crops, including the recently finished full genome sequences of *Arabidopsis thaliana* (At; Arabidopsis Genome Initiative 2000) and rice (International Rice Genome Sequencing Project 2005). However, apart from the ongoing model genome projects for *Medicago truncatula* (Mt) and *Lotus japonicus* (Lj; Young et al. 2005), and the recently completed soybean genome (Schmutz et al. 2010), there is comparatively little sequence data for other legumes, including common bean. Given it phylogenetic position in the Phaseoloids, common bean is also considered a model organism for legume comparative genomics (McClean et al. 2008). Over 2,000 EST contig sequences and more than 5,000 singleton sequences were described from EST sequences generated for common bean (Ramirez et al. 2005). These sequences are a critical tool used here to discover and map polymorphic genes in common bean. Specifically, we harnessed this information to generate sequence information for over 550 genes in *P. vulgaris*, 300 of which were mapped in the BJ population. The utility of these polymorphisms with other common bean mapping populations was also investigated. A genetic-based estimate of linkage disequilibrium was also calculated using data from the parents of those mapping populations. Finally, the map was used to investigate macrosynteny between common bean and *A. thaliana*, *M. truncatula*, and *L. japonicus*. This new gene-based map greatly adds to the genomic resources available for *P. vulgaris* and other legumes.

Materials and methods

Selection of genes for mapping

P. vulgaris genes were chosen based on their homology to *A. thaliana* genes controlling mutant phenotypes (Meinke et al. 2003), biochemical pathway genes, genes found to be under selection during maize domestication (Wright et al. 2005), possible legume microsynteny genes (Mudge et al. 2005), and random *A. thaliana* genes. Contig EST sequences, developed by Ramirez et al. (2005), were used as a query in a blastx (Altschul et al. 1997) analysis of a database consisting of Release 5.0 of the *A. thaliana* gene models (http://www.tigr.org; 10 June 2004). Alignments were examined to determine the extent of the 3′ UTR in each contig, and only those contig sequences with a 3′ UTR estimated to be greater than about 150 nucleotides (nt) were selected. The rationale for this decision is that the 3′ UTR may contain higher diversity, which will not only increase the chance of finding polymorphisms, but also ensure that only one member of a gene family is amplified and mapped. The 5′ primer was designed to a region about 550 nt upstream of the 3′ primer. Sequences with homology to genes encoding ribosomal RNAs and histones, along with those having 100% identity with the top hit, were discarded. Some "high *e* value" genes were also chosen, which had only a poor or no match in the *A. thaliana* genome.

Primer design, plant materials, PCR amplification, and sequence analysis

Primers were designed using the on-line version of the Primer 3 software (Rozen and Skaletsky 2000; http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and purchased from Integrated DNA Technologies (http://www.idtdna.com). All primer sequences are available on request. The primers were used to amplify fragments from genotypes BAT93 and Jalo EEP558 DNA. These genotypes were chosen because they were the parents of an important mapping population in common bean used for linkage mapping here. DNA for the reaction was isolated from greenhouse-grown common bean plants by a CTAB extraction followed by isopropanol precipitation (Doyle and Doyle 1987; as modified by Brady et al. 1998). PCR amplification was performed using a final DNA concentration of ~1.2 ng/μl, 0.5 μM each of the forward and reverse primer, 0.25 mM of each dNTP, and a 1× PCR buffer with 1.5 mM $MgCl_2$ in a volume of 30 μl. One unit of Taq polymerase was added to the reaction. The cycling conditions were 45 cycles with the following parameters: 94°C for 20 s, the primer annealing temperature for 20 s, and 72°C for 2 min. This was preceded by an initial denaturation at 94°C for 3 min and a final extension at 72°C for 7 min. The annealing temperature varied between 52 and 58°C depending on previous results and the calculated T_m of the primer. Amplification success was determined by agarose gel electrophoresis, and the remainder of the amplicon was purified for successful amplifications. The samples were quantified via agarose gel electrophoresis and then sequenced in both directions using the Beckman Coulter CEQ™ DTCS Quick Start kit. Following the sequencing reaction, the fragments were purified using ethanol precipitation per manufacturer's recommendation. The sequences were read by capillary electrophoresis in a Beckman CEQ 8000 genetic analysis system. The sequences were manually edited, and a consensus sequence was developed from the forward and reverse sequence data using the Staden package (Staden 1996). The BAT93 and Jalo EEP558 sequences were compared with each other and with the contig sequences to discover SNP and indel polymorphisms. These were confirmed by checking the original trace files, and then the sequence location and type of each polymorphism was recorded. All sequences are available in GenBank (accession numbers DX823090-DX824144 and ED510083-ED544198).

Polymorphism analysis, genotyping, and genetic mapping

The BAT93 × Jalo EEP558 recombinant inbred population (BJ population; 75 lines) was chosen for mapping because of its high degree of polymorphism based on RFLP, SSR, AFLP, and RAPD segregation data (Freyre et al. 1998; Papa et al. 2007; Rossi et al. 2009). Polymorphisms were detected using one of several PCR-based amplification protocols. These include: insertion/deletion (indel) events; microsatellite sequence; CAPS markers (Konieczny and Ausubel 1993; http://pgrc.ipk-gatersleben.de/snp2caps/); dCAPS (Neff et al. 2002; http://helix.wustl.edu/dcaps/dcaps.html); and allele-specific markers (Drenkard et al. 2000; http://pga.mgh.harvard.edu/cgi-bin/snap3/websnaper3.cgi). All primers, and where necessary, corresponding restriction enzymes used for each gene locus are described in Supplementary Table 1. BAT93 and Jalo EEP558 parents were initially screened to define the specific polymorphism, and subsequently the entire population was screened. The concentrations of reagents and the cycling protocol used were the same as those used to amplify the original fragment. Amplification products were scored directly, or following restriction digestion using 1 unit of the appropriate restriction enzyme and a 1× concentration of the recommended restriction digestion buffer using agarose gel electrophoresis. MapMaker (Lander et al. 1987) was used to construct the genetic linkage map. First, the group command was used, with varying parameters, to assign the markers to linkage groups. Segregation data from

previously mapped markers (Freyre et al. 1998) were combined with the current data to correlate linkage groups and to provide a framework map. Each linkage group was constructed using the order command or using the compare and try commands to build an LOD 2.0 map. These LOD values were confirmed using the ripple command with a window of five markers. The orientation of the map is consistent with the recently established standard (Pedrosa-Harand et al. 2008).

To determine the utility of each marker locus for other common bean mapping populations, the parents of seven other commonly used mapping populations were tested for polymorphisms. The screening protocols used were the same as above for screening and genotyping the BJ RI population. Those parents are: Bunsi, Newport, PC50, Xan159, Benton, NY 6020-4, Dorado, G19833, Xan176, A55, G122, Belneb RR-1, Montcalm, and California DRR 82.

Linkage disequilibrium

Using the polymorphism data for BAT93 and Jalo EEP558 and the parents of the seven other mapping populations, the decay of intergenic LD with genetic distance was measured as described in Remington et al. (2001). The estimates of LD were calculated by using the squared allele frequency correlations (r^2; Weir 1990) between pairs of polymorphic loci that are on the same chromosome using PowerMarker (Liu and Muse 2005). The expected decay of LD was modeled using a recombination drift model as described in Remington et al. (2001). This non-linear regression equation was fitted using PROC NLIN in SAS 9.1.3® (SAS Institute 1999).

Macrosynteny between common bean and *A. thaliana*, *M. truncatula*, and *L. japonicus*

The sequence of the contig used for primer design and mapping was used as a query in a tblastx analysis against a database consisting of the most recent pseudochromosome assembly of *A. thaliana* (http://www.tigr.org), *M. truncatula* (http://www.medicago.org/genome/), and *L. japonicus* (http://www.kazusa.or.jp/lotus/). The results were filtered by only retaining those hits in which the *e* value was less than 1e-20, and the length of region of the query relative to the hit was 150 nucleotides. Dot blots comparing the physical location of the three model plant species and the genetic location of common bean sequences were developed using Microsoft Excel. Macrosyntenic blocks were defined as those in which four genes in consecutive genetic order in common bean corresponded to four gene models in consecutive physical order in the model species.

Results

Rate of fragment amplification and sequencing success

A series of culling steps were used to settle on a set of gene-based fragments for the development of the gene-based genetic map. Those are outlined in Fig. 1. The analysis began with 2,686 *P. vulgaris* contig sequences generated by Ramirez et al. (2005). BLAST analysis determined that 2,394 (89%) of the contigs had an *A. thaliana* homolog at an *e* value less than 1e-20, while 2,094 (78%) contained a homolog at an *e* value less than 1e-30. As much as 61% of the genes at 1e-20 or less (1,458) had more than 20 bases of 3′ UTR according to their alignment with the *A. thaliana* sequences (Fig. 1).

From these 1,458 genes, 1,046 were selected for primer design and analysis. This included 62 genes that exhibit a mutant *A. thaliana* phenotype (Meinke et al. 2003), 148 putative biochemical pathway genes, five genes that were determined to be undergoing selection during maize evolution (Yamasaki et al. 2005), and 801 randomly chosen genes. Of these, BAT93 and Jalo EEP558 amplification products were obtained for 730 (70%) of the sequences (Fig. 1). The average fragment size was about 550 base pairs, although many fragments were 1,500 base pairs or longer. Useful sequence data for both BAT93 and Jalo

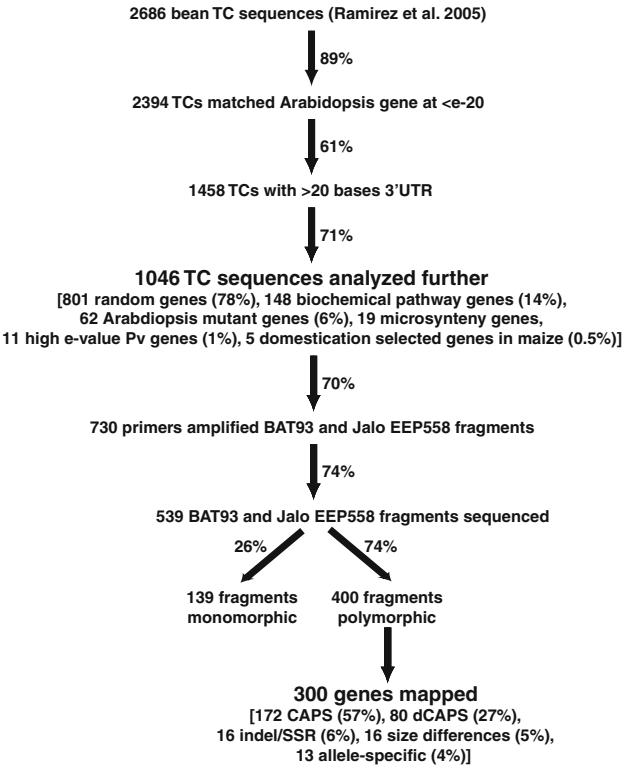

Fig. 1 Steps of gene selection and percentages of successful gene selection at each step

EEP558 were obtained for 534 fragments. This represents 51% of the original primer pairs.

Of the 534 gene fragments, 395 (74%) were polymorphic between BAT93 and Jalo EEP558, and 139 (26%) were monomorphic. Of the 395 polymorphic genes, 300 were mapped: 16 by size difference, 19 with all-specific size polymorphisms, 172 by CAPS, 80 by dCAPS, and 13 with allele-specific primers based on SNPs.

Characterization of gene sequence data

Approximately, 593 kb of sequence data was collected for 395 polymorphic gene fragments. The average sequence length is 527 base pairs. By comparing the genomic sequence with EST-derived contig sequences, 143 introns were found. In addition, partial intron data were generated at the 5′ or 3′ ends for some of the sequences. The average length of the internal introns is 137 nucleotides, while the range is from 37 to 778 nucleotides.

Among the polymorphic genes, 382 SNPs and 95 contained indels. Eighty genes contained one-base indels only, which were not included in the analysis or mapping because of the possibility that they may be sequencing errors. In all, 1,812 polymorphisms were found, which included 1,580 SNPs and 130 indels. Among the SNP-containing genes, the average number of SNPs was 4.14 per gene fragment, while among all genes sequenced the average was 2.9 SNPs per gene. As much as 1.37 indels/gene were observed for the indel-containing genes, while the average was 0.24 indels/gene for all sequenced genes; 46% of all polymorphisms were located in the introns, 37% in exons, and 17% in the 3′ UTR. Among SNPs, there was a similar distribution with 43% located in introns, 40% in exons, and 17% in the 3′ UTR. Among indels, however, 74% were located in introns, 18% in the 3′ UTR, and only 8% in exons (Table 1).

Of the SNPs, 49% were transitions and 51% transversions. Of the SNPs in exons, 54% were transitions and 46% transversions, while SNPs in the introns and 3′ UTR were 47% transitions and 53% transversions. In exons, 341 synonymous (65%) and 183 nonsynonymous SNPs (35%) were detected. Of the nonsynonymous SNPs, 61% were transversions, and 62% of the synonymous SNPs were transitions (Table 1).

Table 1 Statistics on the frequency, location, and type of polymorphisms found in all sequenced common bean genes

Polymorphism statistics	Value
Number of polymorphisms	1,812
Polymorphism frequency (nucleotides/polymorphism)	163.66
Number of SNPs	1,580
Number of genes containing SNPs	382
SNPs per gene	2.9
SNP frequency (nucleotides/SNP)	375.39
Number of indels	130
Number of genes containing indels (excluding 1 nt indel)	95
Number of indels per gene	0.24
Indel frequency (nucleotides/indel)	2281.23
Percentage of polymorphisms in introns	46.3%
Percentage of polymorphisms in exons	36.6%
Percentage of polymorphisms in UTR	17.2%
Percentage of SNPs in introns	42.9%
Percentage of SNPs in exons	40.3%
Percentage of SNPs in UTR	16.8%
Percentage of indels in introns	73.7%
Percentage of indels in exons	7.9%
Percentage of indels in UTR	18.4%
Percentage of nonsynonymous SNPs	34.9%
Percentage of synonymous SNPs	65.1%
Percentage of nonsynonymous SNPs that are transitions	39.3%
Percentage of nonsynonymous SNPs that are transversions	60.7%
Percentage of synonymous SNPs that are transitions	61.6%
Percentage of synonymous SNPs that are transversions	38.4%

Variation among parents of common bean mapping populations

Based on the sequence data, forward and reverse primers were developed for 300 gene-based loci and used to amplify PCR amplicons. Some of the amplicons showed a size difference between BAT93 and Jalo EEP558 based on indels in one of the genotypes. Other primers were developed that also resulted in a size difference in the amplicon because of the presence of an SSR sequence. CAPs polymorphisms (Konieczny and Ausubel 1993) were detected by digesting the amplicon with a restriction enzyme. dCAPS polymorphisms (Neff et al. 2002) were developed by introducing a mismatch in one of the primers in a manner that created a scorable restriction enzyme site in the amplicon. Full information on the primers, enzyme, and polymorphism for each locus is presented in Supplemental Table 1.

To determine the applicability of these primers to score polymorphisms in other common bean mapping populations, the parents of those populations were screened using the diagnostic procedures that revealed the BAT93 and Jalo EEP558 polymorphisms. The degree of polymorphism varied among the parents. The Middle American × Andean crosses used were Dorado × XAN176 (DX), A55 × G122 (AG), Belneb RR-1 × A55 (BA), and Dorado × G19833 (DG) (Table 2). DX was polymorphic for 10% of the genes, AG for 74%, BA for 38%, and DG for 72%. The parents of three Andean × Andean crosses were evaluated: PC50 × XAN159 (PX), Montcalm × California DRK 82 (MC), and

Table 2 Proportion of the genes used for mapping that are polymorphic among parents of several common bean mapping populations

	Population							
	BNe	PX	BN	MC	DG	DX	AG	BA
Number that are monomorphic	212	166	197	225	34	248	69	42
Number that are polymorphic	52	50	68	20	86	28	197	26
Total number	264	216	265	245	120	276	266	68
Percentage of polymorphic	19.7	23.2	25.7	8.2	71.7	10.1	74.1	38.2

BNe Bunsi/Newport (Mesoamerican × Mesoamerican), *PX* PC50/Xan159 (Andean × Andean), *BN* Benton/NY 6020-4 (Andean × Andean), *MC* Montcalm/California DRK 82 (Andean × Andean), *DG* Dorado/G19833 (Mesoamerican × Andean), *DX* Dorado/Xan176 (Mesoamerican × Mesoamerican), *AG* A55/G122 G19833 (Mesoamerican × Andean), *BA* Belneb RR-1/A55 (Mesoamerican × Mesoamerican)

Benton × NY 6020-4 (BN). PX was polymorphic for 23% of the genes, MC for 8%, and BN for 26%. Bunsi × Newport (BNe), a Middle American × Middle American cross, was polymorphic for 20% of the genes (Table 2). The data for BA was based on only 68 successfully screened genes out of the 300, and DG on 120, while each of the other populations were successfully screened on 216 or more of the polymorphic genes.

Of the BJ mapped genes, 29 were monomorphic or did not produce data among all of these mapping parents. The maximum number of parental pairs for which any one gene was polymorphic was five out seven possible pairs. Three genes were polymorphic in five of the sets of parents, 33 were polymorphic in four or more, 41 in three or more, and 93 in two or more. The remaining 111 genes were polymorphic in only one set of parents.

Gene-based linkage map

The 300 gene-based markers were mapped using a core marker set from the BJ map (Freyre et al. 1998) as a guide for the assignment of linkage group number and orientation. The genetic linkage map (LOD 2.0) is displayed in Fig. 2 and is 1545.5 cM in length. A total of 275 markers were mapped at LOD 2.0. These consisted of 199 gene-based markers, 59 core markers, and 17 other markers. Supplemental Table 2 provides full details for all mapped loci. That table also includes the bin location for 153 marker loci that did not meet the LOD 2.0 criteria. All totaled 428 markers could be placed on the genetic map. This includes 300 gene-based markers, 103 core markers, and 24 other markers. Supplementary Table 2 also identifies the *Arabidopsis* homolog and annotation for each marker and the other bean populations in which they are polymorphic.

Linkage group Pv1 is the longest linkage group at 231.6 cM, while Pv10 is the shortest at 69.6 cM. The number of markers on each linkage group ranges from 15 on Pv9 to 33 on Pv1 and Pv3. The average number of markers on each linkage group is 25, and the average number of gene-based markers on each linkage group is 18. Linkage group Pv1 has the greatest number of gene-based markers, 28, while Pv4 has the least with 7.

For rapid mapping of genes to a linkage group, we selected a core set of five markers for each linkage group. These were chosen because of their relatively even distribution and ease of use and low cost (i.e. inexpensive restriction enzyme or variation in the initial PCR fragment). These markers are shown in Table 3 and full details are provided in Supplemental Tables 1 and 2.

Linkage disequilibrium

An important consideration for the application of association mapping with any species is the decay in linkage disequilibrium (LD) over a specific genetic or physical distance. We used data from the polymorphism analysis to evaluate the genome-wide intrachromosomal LD decay. The r^2 value was calculated among all loci within each linkage group. This statistic was chosen because it allows us to determine the degree to which a marker is correlated with a trait of interest (Flint-Garcia et al. 2003). The r^2 value was regressed onto the genetic distance based on the linkage map. Only those gene-based loci on the LOD 2.0 map were included. A non-linear regression model was then fit to the data. Typically, the genetic distance at which r^2 decays to 0.1–0.2 is considered to be the extent of LD in a species (Zhu et al. 2008). From the non-linear regression line in Fig. 3, we calculated the following distances for various r^2 values: $r^2 = 0.20$, 6 cM, $r^2 = 0.10$, 12 cM, $r^2 = 0.05$, 27 cM, and $r^2 = 0.025$, 45 cM.

Macrosynteny with model species

We employed a stringent cutoff to ensure any syntenic blocks between common bean and three model organisms, At, Lj, and Mt, were based on homologous sequences of high similarity. Therefore, we only considered those tblastx hits with an *e* value <1e-20 and a sequence overlap of 150 nt. In addition, we required four matches, in consecutive physical and genetic order (excluding duplicated loci found on other chromosomes). No syntenic blocks were observed between Arabidopsis and common bean with our criteria (Fig. 4a). Yet, close examination revealed several instances of microsynteny where common bean loci cosegregated or were colocalized in the same mapping bin with At loci that were in a physical linear order. For example, g985, g2596, g292, and g2507 are all localized to a bin

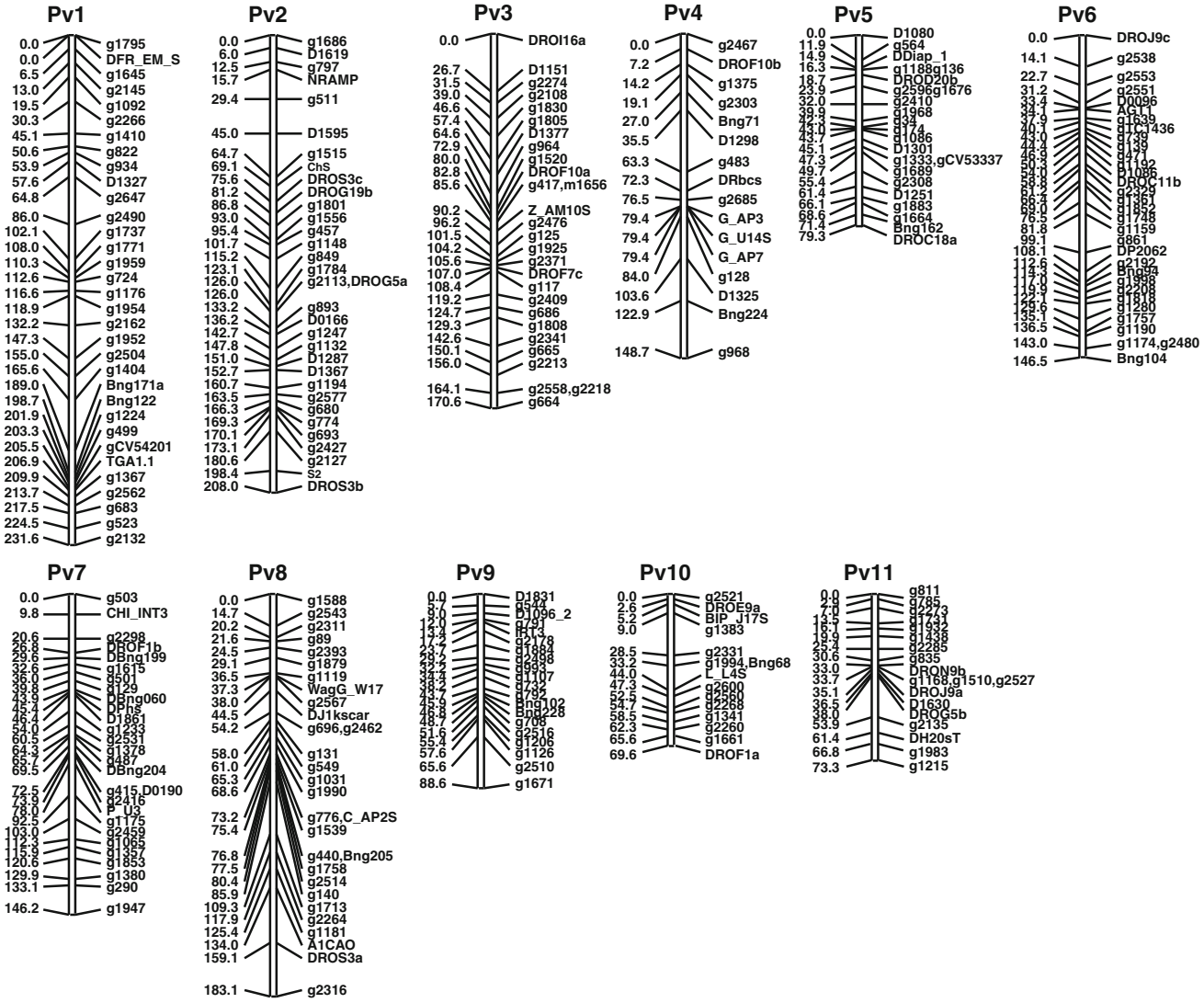

Fig. 2 Gene-based linkage map of common bean. Marker names, including both previously mapped core markers and new gene-based markers, are shown to the *right* of each linkage group. Cumulative map length is shown to the *left* of each marker. Each of the markers was mapped at LOD >2.0. More information on these markers and other markers placed in genetic bins defined by LOD values <2.0 can be found in Supplementary Table 2

region between 23.8 and 33.3 cM on Pv5. The best At hits for these loci were found consecutively between 8,838,118 and 13,663,794 on At 4.

Common bean, Lj, and Mt exhibited distinct syntenic blocks that suggested evolutionary relationships between the three legume species. Syntenic blocks were observed between common bean and each of the Lj chromosomes (Fig. 4b). Conversely, common bean syntenic blocks from all linkage groups, except Pv4, were observed in Lj. Distinct evolutionary relationships were observed. The structure of Pv3 is represented in its near entirety by blocks of Lj2 and Lj4, while a large portion of Pv8 consists of blocks from Lj1 and Lj2. Conversely, Lj2 is defined by syntenic blocks from Pv3 and Pv8, Lj4 can be defined by Pv3 and Pv5 blocks, and Lj5 structure is evolutionarily related mostly to Pv 5 and a short block of Pv9.

From the perspective of common bean and Mt, common bean shares syntenic blocks with all Mt chromosomes except Mt6, whereas Mt shares synteny with all common bean chromosomes except Pv4 (Fig. 4c). Of the common bean chromosomes, Pv1 is defined by Mt3 and Mt7 blocks, Pv2 by Mt4 and Mt5 blocks, Pv3 by two Mt5 blocks as well as a Mt8 block, and Pv6 by Mt2 and Mt5 blocks. The strongest signals of chromosome structure for Mt were observed for Mt2 (based on Pv5 and Pv6 blocks) and Mt5 (based on Pv2 and Pv8).

Table 3 A core subset of mapped gene-based markers

Chromosome	Core markers
Pv1	g683, g499, g1404, g1176, g1410
Pv2	g2127, g693, g1247, g1784, g1515
Pv3	g2218, g1808, g2654, g1805, g2108
Pv4	g118, g755, g2595, g2685, g1375
Pv5	g2308, g1086, g2410, g1395, g136
Pv6	g2480, g1818, g861, g1852, g139
Pv7	g503, g1615, g366, g2416, g1357
Pv8	g135, g131, g1758, g580, g1713
Pv9	g860, g634, g1126, g1107, g544
Pv10	g2521, g1383, g2600, g1341, g1661
Pv11	g785, g1731, g2307, g2527, g188

These 55 markers, 5 from each linkage group, are spaced 15–35 cM apart, as evenly as possible. Almost all markers are mapped at LOD >3.0, and all are very easily scored. Where possible, CAPs markers were chosen because the polymorphism could be detected using common and inexpensive enzymes, while other markers could be distinguished by the size of their PCR product

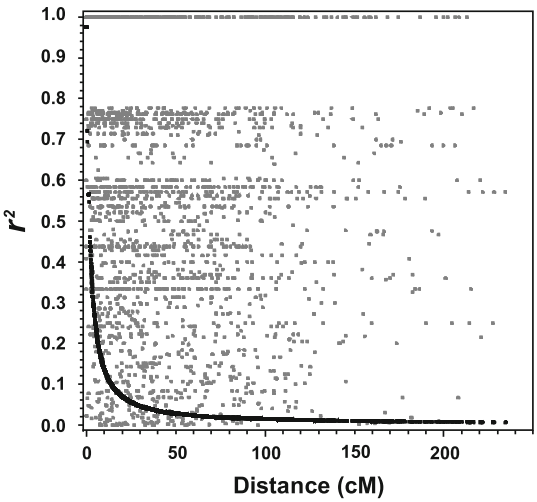

Fig. 3 Linkage disequilibrium decay within the common bean genome

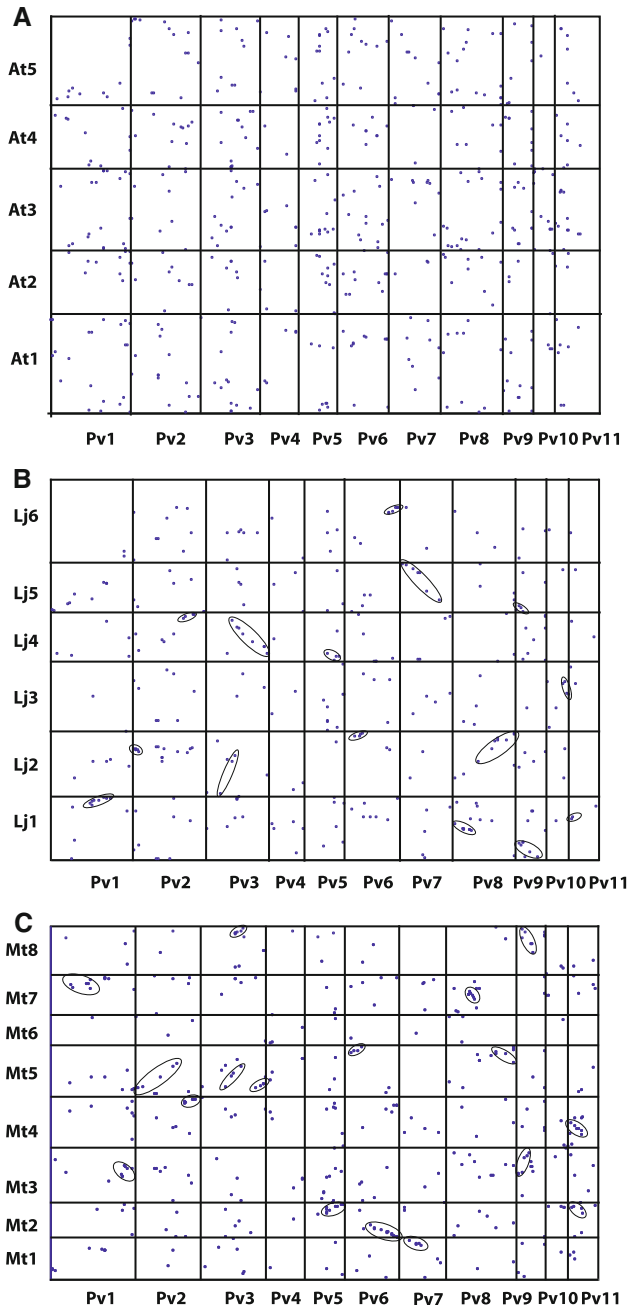

Fig. 4 Dot plot displays of syntenic regions between common bean and *A. thaliana* (**a**), *L. japonicus* (**b**), and *M. truncatula* (**c**) (see Supplementary Tables 3 and 4 for details of the *L. japonicus* and *M. truncatula* synteny analysis)

Discussion

Issues related to gene-based marker development and implementation

As much as 71% of the primers successfully amplified a product. This rate is lower than that observed by Lai et al. (2005); Gao et al. (2004), and Zhu et al. (2003b) for sunflower, wheat, and soybean, respectively. One possible reason for failed amplification is that one of the primers may have been designed over an intron/exon boundary. Alternatively, the primers may target a large genomic region consisting of multiple or large introns that are difficult to amplify using our conditions. The latter is plausible because, although most of the fragments successfully amplified were 500–600 bases long, many were 1,000–1,500 bases or longer, and these often experienced intermittent amplification or required multiple attempts to obtain a successful amplification product. Weak amplification and poor product recovery were the primary reasons that we were not able to obtain reliable sequences for 30% of the

amplified fragments. Overall, this method (from gene selection to polymorphism analysis) was about 35% effective in finding possible polymorphic gene-based markers to map in common bean.

The various mapping methods differ in their utility. CAPS, dCAPS, size, and SSR primers are codominant markers, while allele-specific markers, whether based on indels or SNPs, are dominant. SSR, dCAPS, allele-specific, and some of the size difference markers required new primer design, while CAPS and a size difference based on original primers did not. A total of 523 different attempts were needed to discover the 300 mappable polymorphisms, because if a polymorphism method failed for a gene, a different method was tried on the same polymorphism, or a different polymorphism was selected in the same gene until the gene was either successfully mapped or it was determined that it was unlikely to be successfully mapped. For example, if a CAPS method failed, the same polymorphism was not tested again. Among the original 505 marker assays attempts were 23 fragment size differences, 255 CAPS, 169 dCAPS, 18 allele-specific indels, and 40 allele-specific SNPs. The first attempt for the various methods was successful for 80% of the genes that were eventually mapped. This represents 46% of all polymorphism detection attempts. The most successful were CAPS, where 58% were successful the first time, and the least successful were allele-specific SNPs, where only 18% were successful. About 4% of all primers tried succeeded but were not useful for mapping because greater than 11% of the population (our cutoff value) had missing data. Forty percent (208) completely failed, either because no product was obtained (especially in the case of dCAPS) or the two parents were monomorphic for the amplified product.

A small number of the primer pairs amplified fragments only BAT93 or Jalo EEP558 DNA. If this is a consistent difference, perhaps these genes could be mapped in the future as dominant markers.

BAT93 and Jalo EEP558 polymorphism

Among the genes for which sequence data were available for both BAT93 and Jalo EEP558, 73% were polymorphic. This percentage was higher than that found for other legume species (Choi et al. 2007). One reason for the higher level of polymorphism could be related to the strong differentiation between the Andean and Mesoamerican gene pools represented by two parental genotypes assayed. Indeed, in our case we found a higher polymorphism rate in bean than with sunflower (64%, Lai et al. 2005), and a much higher rate than observed in pine (40–52%, Komulainen et al. 2003, cited in Lai et al. 2005), bread wheat (30%, Gao et al. 2004), and soybean (19%, Zhang et al. 2004, cited in Lai et al. 2005). The rate of one SNP for every 375 bases on average was more frequent that the one SNP for every 562 bases Zhu et al. (2003a) discovered in soybean. However, that study only focused on coding regions and indels were not included in the analysis.

Of the polymorphisms, 10% were located in introns, which were presumably under less selection pressure. A small percentage of the polymorphisms were located in the 3′ UTR, but this may simply reflect the fact that less 3′ UTR sequence was available for each gene. However, it is interesting that 37% of all polymorphisms were found in exons, including 183 non-synonymous SNPs and 9 indels. Four of these nine indels are multiples of three nucleotides, but the other five are not. They include indels 4, 7, 8, and 10 nucleotides long. This may indicate that the selective pressure in these exons is not as strong as to preclude genetic change in the relatively short time since the two gene pools diverged. The ratio of synonymous to nonsynonymous SNPs was about 1.9, while Zhu et al. (2003b) found approximately equal numbers of synonymous and nonsynonymous SNPs in soybean. Interestingly, Ching et al. (2002) found a transitions/transversions ratio of 1.53 in maize, while in this study it was very close to 1.0. Their data, however, was based on only 18 loci.

A gene-based linkage map of common bean

The core markers (Freyre et al. 1998) aided in the development of the gene-based map. Compared to that map, the gene-based map is longer in length. Not all core markers were included in the LOD 2.0 map. This may account for some of the differences between the two maps since an increase in markers usually corresponds to an increase in linkage distance of the map. Given that core markers were included in this map, other core markers could be positioned relative to other loci on the gene-based map.

Pv1 and Pv7 each contain a region of extensive segregation distortion. For most genes in the distorted region on Pv1, 64–68% of the mapping population contained the Jalo EEP558 allele, whereas for genes in the distorted region on Pv7, 55–88% of the population contained the BAT93 allele. These two regions of distortion were also discovered by Freyre et al. (1998), and a region of distortion on Pv1 was observed by Vallejos et al. (1992), also with an excess of the Andean allele. These distortions may be due to meiotic drive or to a translocation event between the two gene pools.

Duplicate mapped genes (those that have their best match to the same *A. thaliana* gene) were noted in the *P. vulgaris* genome, including some tandem duplications located on Pv2, Pv3, Pv6, Pv7, and Pv9, and one group of three loci that was found on Pv6. These are duplications in consecutive map order within a few cM of each other. The sequences available showed 93–99% identity to each other.

Four pairs of duplicate genes were found on different linkage groups. These sequences showed 79–88% identity to each other. Three genes matching the same *A. thaliana* histone gene were also found in three separate parts of the genome, on Pv2, Pv3, and Pv6, with 76–83% identity. In both the duplicate and tandem duplicate categories, some loci were not reciprocal hits, although their best hit was to the same *A. thaliana* gene. This may simply be because the amplified product was from a different region of each homolog.

Qualitative and QTL gene mapping and subsequent functional gene identification are goals of all crop improvement programs. Therefore, placing these loci relative to genes on a gene-based map may identify such functional genes. A number of QTL have been mapped in the BJ population (Nodari et al. 1993b; Koinange et al. 1996; Yu et al. 1998; Miklas et al. 2000, 2001, 2003; Park et al. 2001; Schneider et al. 2001; Tar'an et al. 2001; McClean et al. 2002; Kelly et al. 2003; Kolkman and Kelly 2003; López et al. 2003; Kelly and Vallejos 2004; Murray et al. 2004; Papa et al. 2005; Román-Avilés and Kelly 2005; Papa et al. 2007). As further analysis is performed, these loci can be more accurately mapped, and as the common bean genome sequence becomes available and the gene space better defined, specific functional genes may be discovered. One example of interest suggests a possible connection exists on Pv1, where a days to flowering QTL (Koinange et al. 1996) is located in approximately the same area as g822. Although the QTL interval is large, g822 appears to fall within the interval. That fact that g822 is homologous to the *A. thaliana* gene FPF1, a factor involved in regulating flowering time, is a compelling discovery worth further investigation. In general, the map and corresponding QTL may assist common bean and other legume researchers with the discovery of important genes, including genes associated with known QTL, as well as pursuits such as marker-assisted selection, diversity studies, and additional comparisons of both the sequences and genetic and physical positions of genes in evolutionarily related species.

Middle American/Andean diversity

Each BAT93 and Jalo EEP558 polymorphism could conceivably be considered a polymorphism between the two Mesoamerican and Andean subpopulations of *P. vulgaris*. By screening parents from other mapping populations, it is possible to determine if the polymorphisms are in fact Middle American/Andean polymorphisms or simply specific to the BAT93 and Jalo EEP558 comparison. This screening was also done to assess the relative levels of polymorphism in each mapping population. The results did not show that Middle American × Andean crosses necessarily have greater polymorphism, at least among this group of SNPs and indels. Instead, the inter-gene pool crosses varied greatly, although the intra-gene pool crosses for the most part had fewer polymorphisms. This demonstrates that the variability in *P. vulgaris* is such that, while many polymorphisms may be extrapolated to the subspecies level, many others may be specific to a single smaller lineage.

Linkage disequilibrium

The best marker is the actual gene, which controls a specific phenotype. While the goal of marker-assisted breeding is to discover such genes, they are difficult to identify. The alternative is to discover a marker that is linked to the functional gene. Association mapping (AM) is considered in a number of plant species as a complementary approach to the traditional biparental linkage mapping method (Zhu et al. 2008). The extent to which AM protocols will successfully discover linked loci depends on the degree of linkage disequilibrium (LD) within the population. While utilizing LD for these types of analyses, other factors such as population size and structure and the effect size of the QTL must also be considered.

Two loci are said to be in LD if one can infer the genotype at a second locus based on the genotype at a first locus. There are several computational methods that statistically describe this association. The most relevant LD statistic is r^2 because it allows us to determine the degree to which a marker is correlated with a trait of interest (Flint-Garcia et al. 2003). To display the genome-wide level of LD, r^2 is calculated for pairs of mapped loci within a population. The LD estimate we calculated here was based on a limited number of genotypes representing the full extent of variation in common bean, including representatives of both the Mesoamerican and Andean gene pools as well as races within each gene pool. Typically, the genetic distance at which r^2 decays to 0.1–0.2 is considered to be the extent of LD in a species (Zhu et al. 2008). Here, this occurs between 6 cM ($r^2 = 0.2$) and 12 cm ($r^2 = 0.1$). This is similar to the level of LD detected by Rossi et al. (2009) using AFLP markers on a population consisting of wild and domesticated common bean genotypes.

So, how does this calculation affect AM experimental design in common bean? Using the Illumina GoldenGate assay, 1,536 SNP marker loci can be screened simultaneously. Typically, only loci with minor allele frequencies >0.1 are used to discover marker/trait associations, because rarer alleles do not have sufficient phenotypic effects in an AM population to be discovered by this technique. For the populations we used to develop the r^2-based LD plot, 85% of the loci had a minor allele frequency greater than 0.1. Assuming this value holds for our experiment, ~1,300 loci from the GoldenGate assay will prove useful for AM mapping. Given that the genetic distance of the common bean

genome is ~1,500 cM, we will have a marker density of, on average, one SNP every 1.2 cM. This is clearly less than the 6–12 cM distance at which LD decays to a point of equilibrium. Therefore, it seems that a 1,536 GoldenGate assay will be sufficient to uncover significant associations using AM methods. However, as shown by Rossi et al. (2009), LD in common bean is largely affected by population structure and may drastically decay within gene pools and even more within races.

Macrosynteny between common bean and plant species

Whole genome duplications, chromosomal rearrangements, and gene loss are observed across many taxonomic hierarchies in plants. These events tend to break ancestral syntenic relationships as chromosomal changes accumulate within a lineage. For example, two or three whole genome duplications and subsequent diploidization events appear to be part of the At evolutionary history (Vision et al. 2000; Simillion et al. 2002; Bowers et al. 2003). Of these, one major duplication event post-dated the divergence of dicots from monocots, while the most recent event occurred in the Eurosid II after the At lineage diverged from the Malvaceae (Bowers et al. 2003). Similarily, a major duplication event appears to have occurred early in the legume lineage ~55 MYA (Mamidi and McClean, unpublished data). Subsequently, individual duplication events have occurred in the Gm, Lj, Mt, and Pv lineages (Blanc and Wolfe 2004; Schlueter et al. 2004; Mamidi et al. submitted). Given these duplication, gene rearrangement, and gene loss events occurring in the At and legume lineages, it is not surprising that we were only able to detect short syntenic blocks over centimorgan distances in Pv and a few megabase distances in At. Others have also observed only short microsyntenic stretches between At other legumes including Gm, Lj, and Mt (Grant et al. 2000; Kevei et al. 2005; Zhu et al. 2003a). In contrast, the physical structure of the At genome exhibits extensive macrosynteny with the genetic linkage map of a recent relative, *Brassica napus* (Parkin et al. 2005). This is presumably because polyploidization and rearrangements since the divergence of the two species are limited.

As shown recently by Mamidi et al. (submitted), the legumes studied here share a common duplication event early in the history of evolutionary legumes. It was also observed that each of these species exhibits duplications that are lineage specific. Given the fact that chromosome breakage and rearrangement appears to be coincident with duplication events, it is not surprising that fragmented macrosynteny is observed between Pv, Lj, and Mt. For both species, examples of chromosomes sharing distinct syntenic blocks from different Pv chromosomes were observed. Some chromosomes, such as Lj5 and Mt1, are syntenic across nearly a full Pv chromosomes (Pv7). In other cases, two disparate chromosomal segments of Mt2 and Mt5 account for the near full length of Pv6. These syntenic relationships are clear evidence of chromosome breakage and rearrangement in these legume lineages.

Recently, Hougaard et al. (2008) evaluated Pv synteny with Lj and Mt using a set of legume anchor markers. That set of markers is limited relative to the number used here (65 vs. 300), and the cutoff used to declare synteny was less stringent (1e-7 vs. 1e-20). Even when considering the differences in experimental protocol, the syntenic relations between Pv, Mt, and Lj uncovered in that research are consistent with what we report here. A similar experiment comparing a peanut linkage map derived by mapping legume anchor markers also revealed syntenic relationships between that species and the two model legumes (Bertioli et al. 2009). In essence, those results can serve as a control to determine that the syntenic signals that we observed are indeed true signals. For example, we observed that fragments of Lj5 and Mt1 were both syntenic to the same region on Pv7. Those two chromosomes were also syntenic with peanut chromosome Ar9, a result that supports our observations. By analogy, those results also strongly suggest that Pv7 and Ar9 would also share some degree of synteny. Using all of the reciprocal syntenic regions between Lj, Mt, and Pv, and peanut reported here and in Bertioli et al. (2009), the following Pv and peanut chromosomal relationships can be inferred: Pv1, Ar6 and Ar8; Pv2 and Ar3; Pv3, Ar5, and Ar7; Pv6 and Ar4; Pv7 and Ar9; and Pv11, Ar3, and Ar8.

The number of tools for legume genetics and genomics are increasing. The advent of gene-based markers for other species will facilitate additional syntenic comparisons within this important family. As additional full genome sequences become available for other legume species, more accurate syntenic relationships can be drawn. Collectively, we can then generate a more accurate legume species circle than those currently available (Choi et al. 2004). With that refined data set, comparative genomics in legumes will be more robust in nature.

Acknowledgments This project was funded by the USDA Cooperative State Research, Education and Extension Service: National Research Initiative, Plant Genome Program. We would also like to thank Dr. Paul Gepts for supplying us with genotype information on the markers from the core BAT93 × Jalo EEP558 RI linkage map.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389

American Dietetic Association (2004) Position of the American Dietetic Association and Dietitians of Canada. Nutrition intervention in the care of persons with the human immunodeficiency virus infection. J Am Diet Assoc 104:1425–1441

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Aubert G, Morin J, Jacquin F, Loridon K, Quillet MC, Petit A, Rameau C, Lejune-Henaut I, Huguet T, Burstin J (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. Theor Appl Genet 112:1024–1041

Becerra-Velasquez VL, Gepts P (1994) RFLP diversity of common bean (*Phaseolus vulgaris)* in its centres of origin. Genome 37:256–263

Bertioli DJ, Moretzsohn MC, Madsen LH, Leal-Bertioli SCM, Guimaraes PM, Hougaard BK, Fredslund J, Nielsen AM, Sato S, Tabata S, Cannon SB, Stougaard J (2009) An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. BMC Genomics 10:45

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667–1678

Bowers JE, Chapman BA, Rong J, Paterson AM (2003) Unravelling angiosperm evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438

Brady L, Basset MJ, McClean PE (1998) Molecular markers associated with *T* and *Z,* two genes controlling partly colored seed coat patterns in common bean. Crop Sci 38:1073–1075

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

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3:19

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

Choi I-Y, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon M-S, Hwang E-Y, Yi S-I, Young ND, Shoemaker RC, van Tassell CP, Specht JE, Cregan PB (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:685–696

Doyle JJ, Doyle JL (1987) A rapid DNA extraction procedure for small quantities of fresh leaf material. Phytochem Bulletin 19:11–15

Drenkard E, Richter BG, Rozen S, Stutius LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis*. Plant Phys 124:1483–1492

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

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

Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392–1400

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

Gepts P, Debouck DG (1991) Origin, domestication, and evolution of the common bean (*Phaseolus vulgaris* L.). In: Voysest O, Van Schoonhoven A (eds) Common beans: research for crop improvement. CAB Intern, Wallingford, Oxon

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

Grant D, Cregan P, Shoemaker RC (2000) Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proc Natl Acad Sci USA 97:4168–4173

Hougaard BK, Madsen LH, Sandal N, de Carvalho Moretzsohn M, Fredslund J, Schauser L, Nielsen AM, Rohde T, Sato S, Tabata S, Bertioli DJ, Stougaard J (2008) Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula*, and *Arachis*. Genetics 179:2299–2312

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Kaplan L, Lynch TF (1999) *Phaseolus* (Fabaceae) in archeology: AMS radiocarbon dates and their significance for pre-Colombian agriculture. Econ Bot 53:261–272

Kelly JD, Vallejos VA (2004) A comprehensive review of the major genes conditioning resistance to anthracnose in common bean. HortScience 39:1196–1207

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

Kevei Z, Seres A, Kerest A, Kao P, Kiss P, Toth G, Endre G, Kiss GB (2005) Significant microsynteny with new evolutionary highlights is detected between *Arabidopsis* and legume model plants despite the lack of macrosynteny. Mol Gen Genomics 274:644–657

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

Kolkman JM, Kelly JD (2003) QTL conferring resistance and avoidance to white mold in common bean. Crop Sci 43:539–548

Komulainen P, Brown GR, Mikknoen M, Karhu A, Garcia-Gil MR, O'Malley D, Lee B, Neale DB, Savolainen O (2003) Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. Theor Appl Genet 107:667–678

Konieczny A, Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4:403–410

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

Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). Theor Appl Genet 118:979–992

Lai Z, Livingstone K, Zou Y, Church A, Knapp SJ, Andrews J, Rieseberg LH (2005) Identification and mapping of SNPs from ESTs in sunflower. Theor Appl Genet 111:1532–1544

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

López CE, Acosta IF, Jara C, Pedraza F, Gaitán-Solís E, Gallego G, Beebe S, Tohme J (2003) Identifying resistance gene analogs associated with resistances to different pathogens in common bean. Phytopathology 93:88–95

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

McClean PE, Lee RK, Otto C, Gepts P, Bassett MJ (2002) Molecular and phenotypic mapping of genes controlling seed coat pattern

and color in common bean (*Phaseolus vulgaris* L.). J Hered 93:148–152
McClean PE, Lee RK, Miklas PM (2004) Sequence diversity analysis of dihydroflavonol 4-reductase intron 1 in common bean. Genome 47:266–280
McClean PE, Lavin M, Gepts P, Jackson SA (2008) *Phaseolus vulgaris* L.: a diploid model for soybean. In: Stacey G (ed) Genomics of soybean. Springer Science + Business Media, LLC, New York, pp 55–76
Meinke DW, Meinke LK, Showalter TC, Schissel AM, Mueller LA, Tzafrir I (2003) A sequence-based map of *Arabidopsis* genes with mutant phenotypes. Plant Physiol 131:409–418
Miklas PN, Stone V, Daly MJ, Stavely JR, Steadman JR, Bassett MJ, Delorme R, Beaver JS (2000) Bacterial, fungal, and viral disease resistance loci mapped in a recombinant inbred common bean population ('Dorado'/XAN 176). J Am Soc HortSci 125:476–481
Miklas PN, Johnson WC, Delorme R, Riley RH, Gepts P (2001) Inheritance and QTL analysis of physiological resistance to white mold in common bean G122. Crop Sci 41:309–315
Miklas PN, Delorme R, Riley RH (2003) Identification of QTL conditioning resistance to white mold in a snap bean population. J Am Soc HortSci 128:564–570
Mudge J, Cannon SB, Kalo P, Oldroyd GED, Roe BA, Town CD, Young ND (2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. BMC Plant Biol 5:15
Murray JD, Michaels TE, Cardona C, Schaafsma AW, Pauls KP (2004) Quantitative trait loci for leafhopper (*Empoascafabe* and *Empoascakraemeri*) resistance and seed weight in common bean. Plant Breed 123:474–479
Neff MM, Turk E, Kalishman M (2002) Web-based primer design for single nucleotide polymorphism analysis. Trends Genet 18:613–615
Nodari RO, Tsai SM, Gilbertson RL, Gepts P (1993a) Towards an integrated linkage map of common bean. 2. Development of an RFLP-based linkage map. Theor Appl Genet 85:513–520
Nodari RO, Tsai SM, Guzman P, Gilbertson RL, Gepts P (1993b) Towards an integrated linkage map of common bean III: Mapping genetic factors controlling host–bacteria interactions. Genetics 134:341–350
Papa R, Acosta J, Delgado-Salinas A, Gepts P (2005) A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. Theor Appl Genet 111:1147–1158
Papa R, Bellucci E, Rossi M, Leonardi S, Rau D, Gepts P, Nanni L et al (2007) Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. Ann Bot 100:1039–1051
Park SO, Coyne DP, Steadman JR, Skroch PW (2001) Mapping of QTL for resistance to white mold diseases in common bean. Crop Sci 41:1253–1262
Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171:765–781
Pedrosa-Harand A, Porch T, Gepts P (2008) Standard nomenclature or common bean chromosomes and linkage groups. Annu Rept Bean Improv Coop 51:106–107
Piperno DR, Dilehay TD (2008) Starch grains on human teeth reveal early broad crop diet in northern Peru. Proc Natl Acad Sci USA 105:19622–19627
Ramirez M, Graham MA, Blanco-Lopez L, Silvente S, Medrano-Soto A, Blair MW, Hernandez G, Vance CP, Lara M (2005) Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. Plant Physiol 137:1211–1227
Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484
Román-Avilés B, Kelly JD (2005) Identification of QTL conditioning resistance to *Fusarium* root rot in *Phaseolus vulgaris* L. Crop Sci 45:1881–1890
Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, Papa P (2009) Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. Evol Appl (online) doi:10.1111/j.1752-4571.2009.00082.x
Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386
SAS Institute (1999) The SAS system for windows, version 9.13. SAS Institute, Cary
Savarino A, Pescarmona GP, Boelaert JR (1999) Iron metabolism and HIV infection: reciprocal interactions with potentially harmful consequences? Cell Biochem Funct 17:279–287
Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC (2004) Mining EST databases to resolve evolutionary events in major crop species. Genome 47:868–876
Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183
Schneider KA, Grafton KF, Kelly JD (2001) QTL analysis of resistance to *Fusarium* root rot in bean. Crop Sci 41:535–542
Simillion C, Vandepoele K, Van Montagu MC, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 99:13627–13632
Singh SP (2001) Broadening the genetic base of common bean cultivars: a review. Crop Sci 41:1659–1675
Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128
South Africa Department of Health (2001) South African national guidelines on nutrition for people living with TB, HIV/AIDS and other debilitating diseases. Ministry of Health, Pretoria, South Africa. http://www.sahealthinfo.org/nutrition/sanational.htm (accessed 1/26/2007)
Staden R (1996) The Staden sequence analysis package. Mol Biotechnol 5:233–241
Tar'an B, Michaels TE, Pauls KP (2001) Mapping genetic factors affecting the reaction to *Xanthomonas axonopodis* pv phaseoli in *Phaseolus vulgaris* L. under field conditions. Genome 44:1046–1056
Tohme J, Orlando Gonzalez D, Beebe S, Duque MC (1996) AFLP analysis of gene pools of a wild bean core collection. Crop Sci 36:1375–1384
Vallejos CE, Sakiyama NS, Chase CD (1992) A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics 131:733–740
Vigoroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc Natl Acad Sci USA 99:9650–9655
Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. Science 290:2114–2117

Weir BS (1990) Genetic data analysis. Sinauer Publications, Sunderland

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

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

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 Phys 137:1174–1181

Yu ZH, Stall RE, Vallejos CE (1998) Detection of genes for resistance to common bacterial blight of beans. Crop Sci 38:1290–1296

Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY, Chen SY (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theor Appl Genet 108:1131–1139

Zhu H, Kim D-J, Maek J-M, Choi H-K, Ellis LC, Kuester H, McCrombie MW, Pend H-M, Cook DR (2003a) Syntenic relationships between *Medicago truncatula* and *Arabidopsis* reveal extensive divergence of genome organization. Plant Physiol 131:1018–1026

Zhu YL, Song QJ, Hyten DL, van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003b) Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134

Zhu H, Choi H-K, Cook DR, Shoemaker RC (2005) Bridging model and crop legumes through comparative genomics. Plant Physiol 137:1189–1196

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20