Theor Appl Genet (2003) 107:1362–1374
DOI 10.1007/s00122-003-1398-6

ORIGINAL PAPER

M. W. Blair · F. Pedraza · H. F. Buendia · E. Gaitán-Solís · S. E. Beebe · P. Gepts · J. Tohme

Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.)

Received: 21 May 2003 / Accepted: 2 July 2003 / Published online: 20 September 2003

Abstract A total of 150 microsatellite markers developed for common bean (*Phaseolus vulgaris* L.) were tested for parental polymorphism and used to determine the positions of 100 genetic loci on an integrated genetic map of the species. The value of these single-copy markers was evident in their ability to link two existing RFLP-based genetic maps with a base map developed for the Mesoamerican × Andean population, DOR364 × G19833. Two types of microsatellites were mapped, based respectively on gene-coding and anonymous genomic-sequences. Gene-based microsatellites proved to be less polymorphic (46.3%) than anonymous genomic microsatellites (64.3%) between the parents of two intergenepool crosses. The majority of the microsatellites produced single bands and detected single loci, however four of the gene-based and three of the genomic microsatellites produced consistent double or multiple banding patterns and detected more than one locus. Microsatellite loci were found on each of the 11 chromosomes of common bean, the number per chromosome ranging from 5 to 17 with an average of ten microsatellites each. Total map length for the base map was 1,720 cM and the average chromosome length was 156.4 cM, with an average distance between microsatellite loci of 19.5 cM. The development of new microsatellites from sequences in the Genbank database and the implication of these results for genetic mapping, quantitative trait locus analysis and marker-assisted selection in common bean are described.

Communicated by H.F. Linskens

M. W. Blair (✉) · F. Pedraza · H. F. Buendia · E. Gaitán-Solís · S. E. Beebe · J. Tohme
CIAT—International Center for Tropical Agriculture,
1380 N.W. 78th Ave.Miami, Florida, 33126, USA
e-mail: m.blair@cgiar.org
Tel.: +57-2-4450000
Fax: +57-2-4450073

P. Gepts
Department of Agronomy and Range Science,
University of California, Davis, CA 95616-85150, USA

M. W. Blair
Apartado Aéreo 6713, Cali, Colombia

Introduction

Microsatellites are polymerase chain reaction (PCR)-based markers that have been developed for a wide range of plant species, including many commercial crops. The utility of microsatellites derives from the fact that they detect length polymorphisms at genetic loci that have simple sequence repeats (SSRs) and as a result are highly variable (Morgante and Olivieri 1993; Powell et al. 1996). Although the first simple sequence repeats were primarily developed for temperate crops they have lately become available for additional crops important in the tropics (Mba et al. 2001; Temnykh et al. 2001). Among the grain legumes, many of which are important to developing countries for their nitrogen fixation capacity and valuable source of proteins, microsatellite markers are now available for soybeans (Cregan et al. 1999b), chickpeas (Huttel et al. 1999), cowpeas (Cheng et al. 2001), peanuts (Hopkins et al. 1999) and common beans (Yu et al. 1999, 2000). However with the exception of soybeans the number of microsatellites for most legume species remains small, especially when compared to the grasses, making it difficult to conduct whole-genome or fine-mapping studies.

Microsatellite markers have been developed for plant genomes from both non-coding and coding sequences containing simple repeats (Brown et al. 1996; Temnykh et al. 2001; Schloss et al. 2002). Various techniques exist for discovering new microsatellite markers from the anonymous genomic sequence, all of which rely on the availability of sequence information, DNA libraries and/or enrichment procedures which increase the prevalence of simple sequence repeats (SSRs) in genomic libraries (Edwards et al. 1996). Microsatellite loci that are found with these techniques can be referred to as genomic microsatellites and are usually thought to be from non-coding regions of the genome such as introns or inter-

genic spaces. SSRs have also been found at a predictable frequency in gene-coding regions of higher eukaryotes and recently many microsatellite markers have been developed from expressed sequence tag (EST) database entries or gene sequences (Temnykh et al. 2001; Eujay et al. 2002; Thoquet et al. 2002). Microsatellite loci that are found in gene-coding sequences can be referred to as gene microsatellites.

Common bean is the most important food legume for direct human consumption in the world (Schoonhoven and Voysest 1991). The crop is a major product in international commerce and is produced and consumed by large numbers of the rural and urban poor in Latin America and Africa (Singh 1999). Bean-breeding programs are well established in many developed and developing countries and marker-assisted selection programs have been implemented for the crop (Kelly and Miklas 1998). Furthermore, a tradition of genetic studies in common bean has produced genetic maps that include domestication genes and agronomic traits (Freyre et al. 1998; Gepts 1999). A genetic map of common bean based entirely on microsatellites would be tremendously useful for marker-assisted selection projects as well as further genetic studies. To this end, we have developed a set of genomic microsatellites from enriched libraries of common bean DNA (Gaitán-Solís et al. 2002) and have implemented gene-based microsatellites from Yu et al. (1999, 2000), and from additional searches for SSRs in the Genbank sequence database.

Our principal objective in this study was to map the genomic and gene-based microsatellites developed for common bean in a single mapping population derived from the cross DOR364 × G19833, and to integrate this map with the genetic maps developed by Freyre et al. (1998) and Vallejos et al. (1992). In this genetic-mapping exercise we tested a total of 150 microsatellite markers and located a set of 100 new loci on an integrated map of the species. Given the different sources of the microsatellites we compared the polymorphism rates of markers derived from genes versus those derived from random genomic sequences.

Materials and methods

Populations and DNA extraction

Two populations of recombinant inbred lines (RILs) of *Phaseolus vulgaris* L. were used for this study: the first population was based on the cross DOR364 × G19833 (and therefore will be referred to as the DG population). DOR364, a small red-seeded variety developed by CIAT (International Center for Tropical Agriculture) and released in several countries of Central America (Costa Rica, Honduras, El Salvador, Nicaragua) that belongs to the Mesoamerican genepool of common bean. G19833, a CIAT Germplasm accession with large, yellow and red-mottled seed, is a landrace originally collected in Peru that belongs to the Andean genepool of common bean. A total of 87 RILs were developed for this cross by a modified single-seed descent from the F_2 to the F_9 generation. The plants within the F_9 progeny row were bulked and used for subsequent genetic analysis. The second population was based on 91 RILs from the cross BAT93 × JaloEEP558 (and will be referred to as the BJ population) whose development and origins are described by Freyre et al. (1998). Total genomic DNA for each of the recombinant inbred lines in both populations was isolated from bulked leaf tissues of eight greenhouse-grown plants per line, using a CTAB extraction method as described in Afanador and Hadley (1993).

Source and development of markers

We used three sets of markers in this study: (1) genomic microsatellites developed in this laboratory by Gaitán-Solís et al. (2002); (2) gene-coding microsatellites developed by Yu et al. (1999, 2000), and (3) additional gene-coding and non-coding microsatellites from searches for SSR-containing *Phaseolus* sequences deposited in the Genbank database before July 15, 2001 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=Nucleotide). SSRs were found using the SSR identification tool (SSRIT) that screens for all possible dimeric, trimeric and terameric repeats (http://www.gramene.org/). Only sequences containing a minimum of three tetra-nucleotide, four tri-nucleotide or five di-nucleotide motif repeats were used for primer design. Primers were designed using Primer 3.0 software (http://www.genome.wi.mit.edu/) to produce PCR amplification fragments that were on average 150-bp long, and PCR primers with consistent melting temperatures of 55°C or above and an average length of 20 nucleotides. Primer pairs were checked to make sure that they had similar melting temperatures and did not suffer from palindromes or end-pairing. Some of the primers developed by Yu et al. (2000) were re-designed from the original Genbank sequences to have a more consistent annealing temperature of 47°C and were named with a bean microsatellite-database (BMd) designation. Newly designed primer sequences and their Genbank entry source are given in Table 1.

Microsatellite analysis

Polymorphisms between the mapping parents were determined on parental survey gels. Standard microsatellite PCR conditions were used throughout the parental survey, where markers were amplified with a hot start of 92°C for 5 min; then 30 cycles of 92°C denaturing for 1 min; 47°C annealing for 1 min and 72°C extension for 2 min; followed by a 5-min final extension at 72°C. The PCR reaction was carried out in a 20-μl final volume containing 50 ng of genomic DNA, 0.1 μM of each of the forward and reverse primers, 10 mM of Tris-HCl (pH 7.2), 50 mM of KCl, 1.5 to 2.5 mM of $MgCl_2$, depending on the primer combination, 250 mM of total dNTP and 1 unit of *Taq* polymerase. Any primer pairs not amplifying parental DNA under these conditions were not considered further. For mapping the same or similar PCR conditions were used. After amplification, a volume of 5 μl of formamide, containing 0.4% bromophenol blue and 0.25% w/v xylene cyanol FF, was added to each PCR reaction and the mixture was denatured at 92°C for 2 min. From two to four microliters of the mixture were then loaded onto 4% denaturing polyacrylamide (29:1 acrylamide: bis-acrylamide) gels that contained 5 M urea and 0.5 x TBE, and run in Sequi-Gen GT electrophoresis units (Biorad, Hercules, Calif., USA) at a constant power of 120 W. Detection of PCR amplification products was via silver staining according to the Manufacturer's guide (Promega Inc., Madison, Wis., USA) with some modifications, namely repeated use of the staining and developing solutions through a re-circulating tank system. The sizes of the parental alleles were estimated based on 10-bp and 25-bp molecular-weight ladders.

Microsatellites that were polymorphic for the parents of either of the two populations were then amplified on the recombinant inbred line (RIL) individuals. To determine the genotypes of the progenies, alleles were scored based on the parental bands that were amplified as controls along with the RIL individuals. PCR products for all the progeny were multiplexed by four repeated loadings of the polyacrylamide gels. Size ladders, added with the first load,

were used to confirm the allele sizes observed in the parental survey.

Data analysis

Segregation distortion was measured with a chi-square test for an expected 1:1 ratio and segregation data was used to place the microsatellites on the established genetic maps for the DG (map and population derived from the cross DOR364×G19833) and BJ (map and population derived from the cross BAT9×JaloEEP558) populations (Beebe et al. 1998; Freyre et al. 1998). The DG map included 240 RFLP, RAPD, SCAR and AFLP markers described in Beebe et al. (1998), while the BJ map contained 141 markers (segregation data available http://agronomy.ucdavis.edu/gepts/bjril7.htm) described in Freyre et al. (1998). The two maps were linked by common RFLP markers with the map described by Vallejos et al. (1992). Linkage analysis was conducted with the Kosambi mapping function using the software application Mapmaker 2.0 for Windows (Lander et al. 1987). Linkage groups were constructed with the 'group' command and a minimum LOD of 3.0, and a maximum distance of 37.5 cM. A subset of markers with a logarithmic of odd ratio (LOD) of 6.0 and a maximum distance of 15-cM were used to create the framework set. Additional markers were placed to the most-likely interval with the 'try' command and a minimum LOD of 2.0. Marker order determined by multipoint analysis was confirmed with the 'compare' command using a minimum LOD of 4.0.

Results

Source of microsatellite markers

A total of 150 common bean microsatellites were used in this study. Of these, 81 were anonymous genomic or non-coding microsatellites and 69 were gene-derived microsatellites. The genomic microsatellites included all 68 markers from an enriched library as described by Gaitán-Solís et al. (2002) and 13 newly developed non-coding microsatellites from Genbank sequences (Table 1). The gene-coding microsatellites included the 12 markers described by Yu et al. (1999), 44 newly developed markers designed for sequences obtained from database searches of *Phaseolus* genes (Table 1) and an additional 13 Genbank-based microsatellites from Yu et al. (2000) (PV-aaat001, PV-at005, PV-atcc001, PV-atcc002, PV-atcc003, PV-atct001, PV-cca002, PV-ccct001, PV-cct001, PV-gat001, PV-tta001, PV-tttc001 and VA-ag001).

In the Genbank searches, SSRs were found in a range of coding and non-coding sequences. A total of 632 *Phaseolus* sequences were downloaded from the database of which 492 sequences represented partial or full-length cDNAs, ESTs or genes from *Phaseolus vulgaris*, *P. lunatus*, *P. coccineus* and *P. acutifolius*. A total of 130 of these sequences were of multiple alleles of the small subunit ribosomal RNA gene from different *Phaseolus* species, none of which contained SSR motifs. Of the remaining sequences, microsatellite markers were derived from 54 *P. vulgaris* sequences and three *P. coccineus* sequences. The di-nucleotide motif GA/CT was the most common in these sequences (23%); while AT/TA (11%) and GT/AC (7%) motifs were less common (Table 1). A variety of tri-nucleotide repeats were found in these sequences, of which the most common were CAA (9%), TCC (7%) and GAT (7%). Some trinucleotide repeats in gene sequences were associated with amino-acid repeats in the corresponding protein; for example, the hydroxy-proline rich glycoprotein contains a CCA repeat which codes for proline. Among the 44 gene-based microsatellites, simple sequence repeats were predominantly found in open reading frames (ORF) (52%) or in 5′ untranslated regions (UTR) (36%), and less frequently in 3′ UTRs (4%) or introns of genes (2%) (Table 1).

An additional 120 of the individual sequences searched in this study represented genomic RFLP clones (from the Bng series) that have been mapped in common bean by Vallejos et al. (1992). Eleven of these sequences contained SSRs (Table 1). Interestingly, five of the RFLP clones that were positive for SSR sequences contained the di-nucleotide AT repeat which was infrequent in the coding sequences. In addition, two tri-nucleotide motifs (ATT and AGC) which were uncommon in the coding sequences described above were found in these non-coding RFLP sequences. Two additional SSRs were found in a RAPD clone isolated at CIAT (Genbank AY298744) and in a CEN-like sequence (Genbank AY028711) (Table 1).

The highest number of repeats detected in the simple sequence repeats was nine, while the average number of repeats was 5.5 among all the microsatellites identified (Table 1). Genomic microsatellites had a significantly higher (P=0.02) average number of repeats than gene microsatellites (6.2 versus 5.3) in an unpaired t-test. In this study, di- and tri-nucleotide motif containing microsatellites did not have a significantly different average number of repeats.

Polymorphism survey

All the microsatellite markers described above were screened for amplification products and polymorphism in the parents of the DG and BJ populations. Figure 1 shows the amplification pattern of ten of the polymorphic microsatellites on a subset of 24 recombinant inbred-line progenies. No difference in band intensity between the cDNA and genomic derived microsatellites was observed. A majority of the microsatellites produced single bands for the parents; however, four genic (BMd26, BMd28, BMd54 and Pv-ag004) and three genomic microsatellites (BM20, BM150 and BM205) produced double or multiple banding-patterns (1.5 and 2.5% of the total, respectively). In these cases, the different bands were interpreted as separate loci. Twelve microsatellites, of which three were genomic (BM161, BM180 and BMd40) and nine were gene-derived (BMd24, BMd29, BMd34, PV-aaat001, PV-tta001, PV-tttc001, PV-ctt002, PV-at005 and PV-at009), did not amplify under our PCR conditions after two attempts and were not investigated further.

Polymorphism rates for the DG and BJ populations were 65.4 and 63.2% for the genomic microsatellites, and

Table 1 Characteristics of 57 common bean microsatellite (BM) markers designed from coding and non-coding sequences in the Genbank database

Marker	Genbank entry	Description	Predicted location in genea	Species	Motif	Max. repeat	Expected product size	Forward/reverse primer
Coding								
BMd-1	X96999	Ypr10	5′ UTR	*P. vulgaris*	AT	9	165	*CAAATCGCAACACCTCACAA* *GTCGGAGCCATCATCTGTTT*
BMd-2	U18349	Phaseolin G-box binding protein	ORF	*P. vulgaris*	CGG	8	106	*AGCGACAGCAAGAGAACCTC* *CAACAAACGGTGATTGACCA*
BMd-3	U34754	Cellulase	Intron	*P. vulgaris*	AT	8	223	*TGTTTCTTCCTTATGGTTAGGTTG* *GTATCCTCCGATCAAATTCACCT*
BMd-4	AF190462	9-cis-epoxycarotenoid dioxygenase	ORF	*P. vulgaris*	ACC	4	146	*CTCCACTACCCCAAACAGTACC* *TTTGAGGAAATGGTTGTTTGGT*
BMd-5	AJ132438	GA 2-oxidase	5′ UTR	*P. coccineus*	CAA	5	122	*CGTGGACTTGAATGGTTTCAG* *TCCTTACCCTGTTCTGCTTCTC*
BMd-6	Z99957	Moldavian encoding legumain-like proteinase prec.	ORF	*P. vulgaris*	TCC	5	122	*CATCGAATGCCCAAGAGAATA* *CTCACTGTCTTCCATCCAAGC*
BMd-7	X13596	Glycine-rich cell wall protein	ORF	*P. vulgaris*	TGG	5	166	*GGATATGGTGGTGATCAAGGA* *CATACCCAATGCCATGTTCTC*
BMd-8	U70530	Gibberellin 20-oxidase	5′ UTR	*P. vulgaris*	CT	7	176	*TTCATCCTCTCTCCCGAACTT* *CTTTTGTGGCTGAGACATGGT*
BMd-9	X06336	Phytohemagglutinin-L (PHA-L)	5′ UTR	*P. vulgaris*	CATG	4	135	*TATGACACCACTGGCCATACA* *CACTGCGACATGAGAGAAAGA*
BMd-10	X13329	Glutamine synthetase beta subunit	5′ UTR	*P. vulgaris*	GA	8	139	*GCTCACGTACGAGTTGAATCTCAG* *ATCTGAGAGCAGCGACATGGTAG*
BMd-13	AF204210	Llipoxygenase	ORF	*P. vulgaris*	GA	5	194	*TCATGCCTGAGAAAGGGTCT* *CCCTGCATCAGAAGTCCAAT*
BMd-14	AF190462	9-cis-epoxycarotenoid dioxygenase	3′ UTR	*P. vulgaris*	TA	5	186	*TCGCTCCAAGTGCATGTAAG* *TGGGGAAACAAAACAAGAAGA*
BMd-15	K03288	Phytohemagglutinin (PHA-E)	5′ UTR	*P. vulgaris*	AG	6	166	*TTGCCATCGTTGCTTAATTG* *TTGGAGGAAGCCATGTATGC*
BMd-16	K03289	Phytohemagglutinin (PHA-L)	5′ UTR	*P. vulgaris*	CATG	4	136	*ATGACACCACTGGCCATACA* *GCACTGCGACATGAGAGAAA*
BMd-17	U77935	DnaJ-like protein	ORF	*P. vulgaris*	CGCCAC	6	116	*GTTAGATCCCGCCCAATAGTC* *AGATAGGAAGGGCGTGGTTT*
BMd-18	X59469	Chalcone synthase	5′ UTR	*P. vulgaris*	TGAA	3	156	*AAAGTTGGACGCACTGTGATT* *TCGTGAGGTAGGAGTTTGGTG*
BMd-19	X79722	plsB	ORF	*P. vulgaris*	CCT	5	154	*GCCAACCACATTCTTCCCTAC* *GGAAGCGAGGCAGTTATCTTT*
BMd-20	X74919	Endochitinase	3′ UTR	*P. vulgaris*	TA	5	123	*GTTGCCACCGGTGATAATCT* *GTGAGGCAAGAAGCCTTCAA*
BMd-21	AJ297395	Class I chitinase	ORF	*P. vulgaris*	CCT	6	190	*GGCTCCACCATCGACTACTG* *TGCGATGTTTGAGCATTTGT*
BMd-22	X74403	Cyclophilin	5′ UTR	*P. vulgaris*	TC	6	121	*GGTCACTTCCGGAGCATTC* *CGGGAAATGGAAGTCACAGT*
BMd-23	E04043	Phenylalanine ammonia lyase	ORF	*P. vulgaris*	GA	5	127	*GGCTTGGTCCTCTCATTGAA* *TGGAAATTACCACCATGCAA*
BMd-24	M13968	Chitinase	5′ UTR	*P. vulgaris*	GA	5	122	*CACCTTATCATTTAGAGGAAA* *CACTGCTCTCCGTAGCTTCC*
BMd-25	L47221	Eukaryotic initiation factor 5	ORF	*P. vulgaris*	GAT	6	118	*GCAGATCGCCTACTCACAAA* *CGTTGACGAGAAGCATCAAG*

Table 1 (continued)

Marker	Genbank entry	Description	Predicted location in genea	Species	Motif	Max. repeat	Expected product size	Forward/reverse primer
BMd-26	U12927	Alpha galactosidase	ORF	*P. vulgaris*	GAT	6	141	*CTTGCCTTGTGCTTCCTTCT* *TCCATTCCCAACCAAGTTTC*
BMd-27	M18093	Hydroxyproline-rich glycoprotein	ORF	*P. vulgaris*	CCA	6	109	*GGACCCACCATCACCATAAC* *TGGTGGAGGTGGAGATTTGT*
BMd-28	Z30345	Nodulin	ORF	*P. vulgaris*	GT	4	151	*TGCATCAACTTTAGGAGCTTG* *TCTTGTCTTATCAGCAGGTGGA*
BMd-29	U54703	Dehydrin	ORF	*P. vulgaris*	GAA	5	159	*CTTCACCGATCTGACAGCAG* *TTTCTCCACTGGAACACTCG*
BMd-30	U32582	Cytosolic glutamine synthetase	5′ UTR	*P. vulgaris*	TTAA	3	134	*CAGCAAATGCAACGCTAAGA* *GGTTGAATTTTGAAACCCTGA*
BMd-31	U28645	Embryo-specific acidic transcriptional activator	ORF	*P. vulgaris*	TGG	6	161	*TGAAGAGGATCGCAAGGTTC* *AGCCGAAACACTGTCCTTGT*
BMd-32	X85804	Plasma membrane H+ ATPase	5′ UTR	*P. vulgaris*	CTT	4	150	*ACACCCTTCATCTCCCTCAT* *ACCCATGTTGGATGTTGGAT*
BMd-34	AF078082	Receptor-like protein kinase homolog	ORF	*P. vulgaris*	GAT	4	119	*ATCGCCGTTGTCGTTTCTA* *TCGAATTGCAATGACTGAGAA*
BMd-35	AX008671	GA 2-oxidase gene	5′ UTR	*P. coccineus*	CAA	5	128	*TCTCTTCCTTACCCTGTTCTGC* *GCGTGGACTTGAATGGTTTC*
BMd-37	AY007525	Putative pyridoxine biosynthetic enzyme	5′ UTR	*P. vulgaris*	AC	8	134	*GGCACGAGCAACAATCCTT* *CCATCATAGAGGGCAACCAC*
BMd-45	AF293023	IAA-protein conjugate	ORF	*P. vulgaris*	AG	5	129	*GGTTGGGAAGCCTCATACAG* *ATCTTCGACCCACCTTGCT*
BMd-46	AF293404	Suspensor-specific protein	ORF	*P. coccineus*	TCT	4	158	*GGCTGACAACAACTCTGCAC* *CTGGCATAGGTTGCTCCTTC*
BMd-47	AF350505	bZip transcription factor	ORF	*P. vulgaris*	AT	5	150	*ACCTGGTCCCTCAAACCAAT* *CAATGGAGCACCAAAGATCA*
BMd-48	AF088188	Beta-glucan binding protein	ORF	*P. vulgaris*	CT	5	131	*CCCCACCAACTCTTTCTTCC* *CAGAATTGACTTGGCGAGAA*
BMd-49	AX008671	n.a.	5′ UTR	*P. vulgaris*	AAC	5	95	*TCTCTTCCTTACCCTGTTCTGC* *GGTTCAATGCTGGCTGAGA*
BMd-50	AF285172	Senescence-associated receptor-like protein kinase	ORF	*P. vulgaris*	AAC	4	124	*TGGTGAGAGAAGGACAATAGCA* *GCCGCTTGTGACGTTTATTT*
BMd-51	AF128454	Gamma-glutamylcysteine synthetase precursor	5′ UTR	*P. vulgaris*	CT	5	116	*CGCCAATTCTTCAACCCTAA* *GTAGTTCGCCCGAGGACTG*
BMd-52	AF325187	Putative POL3-like reverse transcriptase	Na	*P. coccineus*	ATT	4	151	*TCTTGGTGCGCAGAAAGTTA* *AAGGCTTTGTTTTGATTAAGGTT*
BMd-53	AF324244	Translation initiation factor 2	ORF	*P. vulgaris*	GTA	5	105	*TGCTGACCAAGGAAATTCAG* *GGAGGAGGCTTAAGCACAAA*
BMd-54	AJ297395	Putative class I chitinase	ORF	*P. vulgaris*	CCT	6	154	*GGCTCCACCATCGACTACTG* *GAATGAGGGCGCTAAGATCA*
BMd-55	AF142702	*Phaseolus coccineus* maturase-like protein	Na	*P. coccineus*	AT	5	188	*GTCCGTTCATGGGTTTGACT* *TCGAGATCTACGGAGGAGTTC*
Non-coding								
BMd-11	AZ044955	Bng95/F Common bean genomic clone	Na	*P. vulgaris*	GA	8	161	*GCTCAACATTCCAGAGGCTAA* *TCAAACCTACATAAATAAAACAAAACA*
BMd-12	AZ044945	Bng225/R Common bean genomic clone	Na	*P. vulgaris*	AGC	7	167	*CATCAACAAGGACAGCCTCA* *GCAGCTGGCGGGTAAAACAG*

Table 1 (continued)

Marker	Genbank entry	Description	Predicted location in genea	Species	Motif	Max. repeat	Expected product size	Forward/reverse primer
BMd-33	AZ301561	Bng91/R common bean genomic clone	Na	*P. vulgaris*	ATT	9	110	*TACGCTGTGATGCATGGTTT* *CCTGAAAGTGCAGAGTGGTG*
BMd-36	AY298744	RAPD clone isolated at CIAT	Na	*P. vulgaris*	TA	8	164	*CATAACATCGAAGCCTCACAGT* *ACGTGCGTACGAATACTCAGTC*
BMd-38	AZ301538	Bng26/R common bean genomic clone	Na	*P. vulgaris*	GA	5	178	*GCGTTTCCATGAATCAATCC* *AAATTCCGAACCCGTGAACT*
BMd-39	AZ301615	Bng27/F common bean genomic clone	Na	*P. vulgaris*	CCA	4	126	*CACCGAGGAGTTTCTGTGAA* *TTAAGTCCTTGCAGCCCTTC*
BMd-40	AZ301605	Bng38/F common bean genomic clone	Na	*P. vulgaris*	AT	6	197	*AACCTTCTTGCGCTGATCTC* *TAGTGGCCATTCCTCGATCT*
BMd-41	AZ301561	Bng91/R common bean genomic clone	Na	*P. vulgaris*	ATT	9	250	*CAGTAAATATTGGCGTGGATGA* *TGAAAGTGCAGAGTGGTGGA*
BMd-42	AZ301511	Bng105/R common bean genomic clone	Na	*P. vulgaris*	AT	5	149	*TCATAGAAGATTTGTGGAAGCA* *TGAGACACGTACGAGGCTGTAT*
BMd-43	AZ301513	Bng112/R common bean genomic clone	Na	*P. vulgaris*	CCT	5	176	*CAGCATCAAGAAGACCCAAG* *CAGCACCACTATGGGAGGAC*
BMd-44	AZ301573	Bng125/F common bean genomic clone	Na	*P. vulgaris*	AG	5	135	*GGCAGCTTACTAACCCGAAA* *TTCCTTCCCCTTTCTTCTCC*
BMd-56	AY028711	CEN-like sequence II	Na	*P. vulgaris*	AT	5	193	*AATGCGTGAGCATGATTAAGG* *TCATCTGTCAGCACCAAACC*
BMd-57	AZ301498	Bng68/F common bean genomic clone	Na	*P. vulgaris*	TA	5	140	*GAGGCAAAGAAGCTATTGAA* *TTTTTAATTCCTGTGAATTGTTT*

[a] Predicted location in gene, ORF = open reading frame, UTR = untranslated region (either 5′ or 3′ end), Na = not applicable in case of the non-coding region

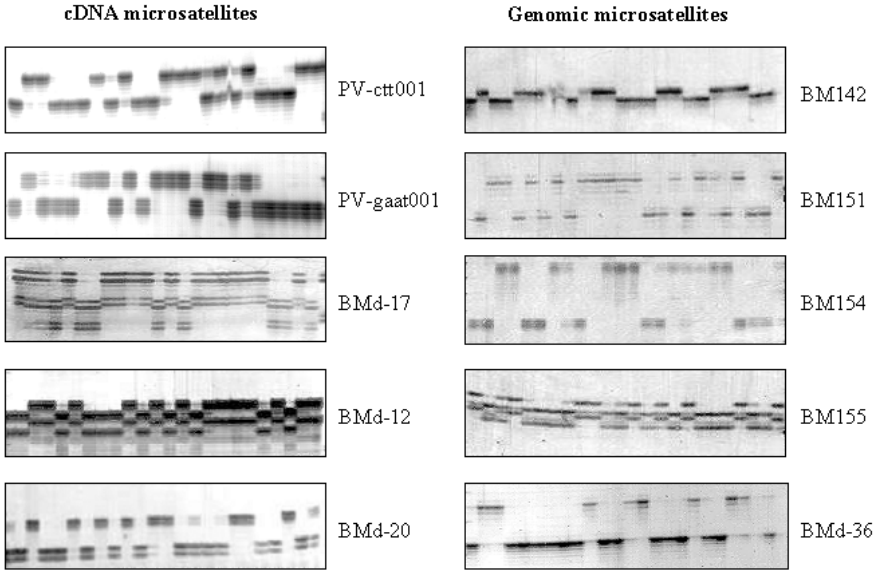

Fig. 1 Segregation pattern of mapped bean microsatellites (*BM*) from both cDNA and genomic sequences in a recombinant inbred-line population (DOR364 × G19833) showing the microsatellite amplification product as analyzed on silver-stained polyacrylamide gels

46.3 and 46.2% for the genic microsatellites, respectively. Overall the percentage of polymorphism between the parents of both populations was very similar: a total of 84 out of the 150 microsatellites tested for the parents of the DG population were polymorphic (56.0%), while a total of 68 out of the 122 microsatellites tested for the parents of the BJ population were polymorphic (55.7%).

Genetic mapping of microsatellites

A total of 100 new microsatellite loci were placed on the two genetic maps (78 on the DG population and 22 on the BJ population) during this study (Fig. 2). To generate a more reliable map for the DG population we removed any RAPD or AFLP markers that did not fit at a LOD>2.5, resulting in a map of with a total of 246 markers, of which 78 were microsatellites, 48 were RFLPs, 102 were RAPDs and 18 were AFLPs. For the BJ population, microsatellites were placed relative to 71 markers but were shown at their position in the map presented by Yu et al. (2000). Microsatellite loci were found on each of the 11 chromosomes of the species and each chromosome was tagged with at least five or more microsatellites. Two chromosomes, b02d and b04b, had a relatively greater number of microsatellites placed on them, with 17 and 13 markers respectively; while the average number of microsatellite loci per chromosome was ten. A total of four microsatellite loci remained unlinked in the DG map (BM158, BM187, BM202 and BM212); however additional mapping allowed BM158 to be placed on the BJ map.

The total cumulative map length for the DG population was 1,720 cM with an average chromosome length of 156.4 cM. The average distance between microsatellite loci in this map was 19.5 cM; however, the distribution of loci was variable and several large gaps between microsatellites remained on the map. The actual distance between microsatellites ranged from loci that were co-segregating to those that were separated by over-100 cM and many intervening AFLP, RFLP or RAPD markers (as between Pv-ag3 and BM53 on b01). The largest gaps between microsatellite markers remained on chromosomes b01h, b07a, b08f and b11j. Chromosomes b05e, b08f and b09k had good coverage of microsatellite markers at one end but not at the other end, while chromosomes b02d, b03c and b06g were sparsely covered at both ends. On chromosome b10i only the lower half of the linkage group contained microsatellites.

Among the markers that presented multiple bands, duplicate loci could be mapped for BMd28 and BM205. The two BMd28 loci could be mapped in the DG population on the same chromosome (b05e) but at a distance of 17 cM, while the two BM205 loci mapped to separate chromosomes (B07a and b11j). Two other microsatellites, Pv-ag004 and BMd26, consistently produced two bands of which only one could be mapped in either population. The two loci for Pv-ag004 co-segregate in another mapping population (data not shown) and are presumed to be very tightly linked (under 1 cM). Meanwhile the markers BMd9, BMd15 and BMd16, the three microsatellites aside from Pv-ag004 that were also designed for related but different phytohemagglutinin gene sequences, amplified single bands but were found to co-segregate, indicating that this gene family clusters at a single location in the genome on chromosome b04b. Except for this example, the gene-based microsatellites were better distributed than the genomic microsatellites. Several clusters of genomic microsatellites were found on every chromosome except b06g and b08f which were the linkage groups with the fewest microsatellites on them. The majority of the clustered markers were microsatellites developed by Gaitán-Solís et al. (2002), including the microsatellites BM197, BM181, BM98, GAT54 and BM159 (on chromosome b03c), and BM150, BM201, BM185 and BM210 (on chromosome b07a).

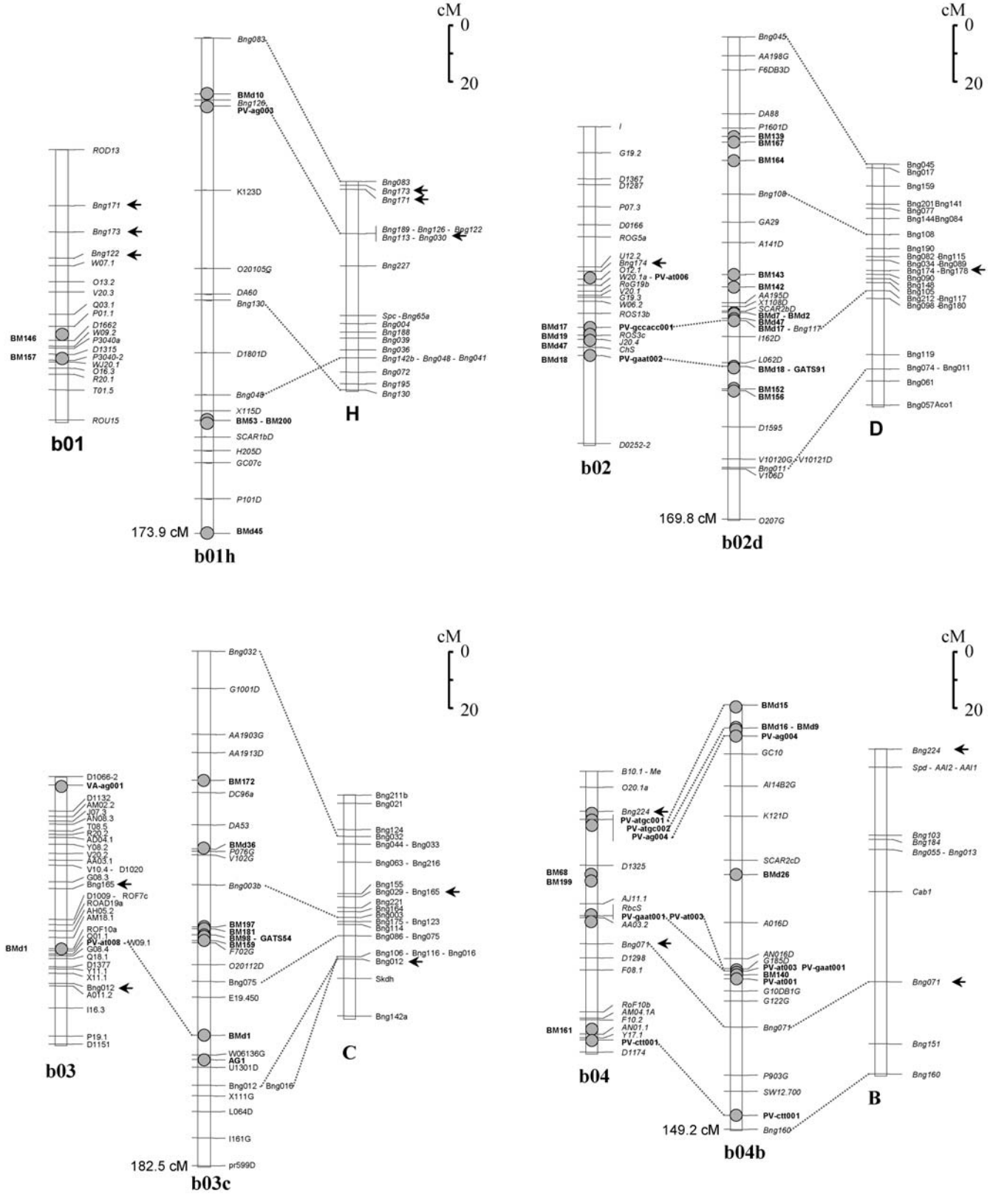

Fig. 2 Comparative genetic maps for the DOR364 × G19833 (chromosome designations b01h through b11j) population relative to two reference maps published by Freyre et al. (1998) (b01 through b11) and Vallejos et al. (1992) (*A through K*). Microsatellite markers (*in bold*) are shown as *circles* on the first two maps based on current mapping results and those from Yu et al, 1999. *Dotted lines* between all three maps indicate equivalent genetic markers. *Arrows* indicate common RFLP markers between the two reference maps

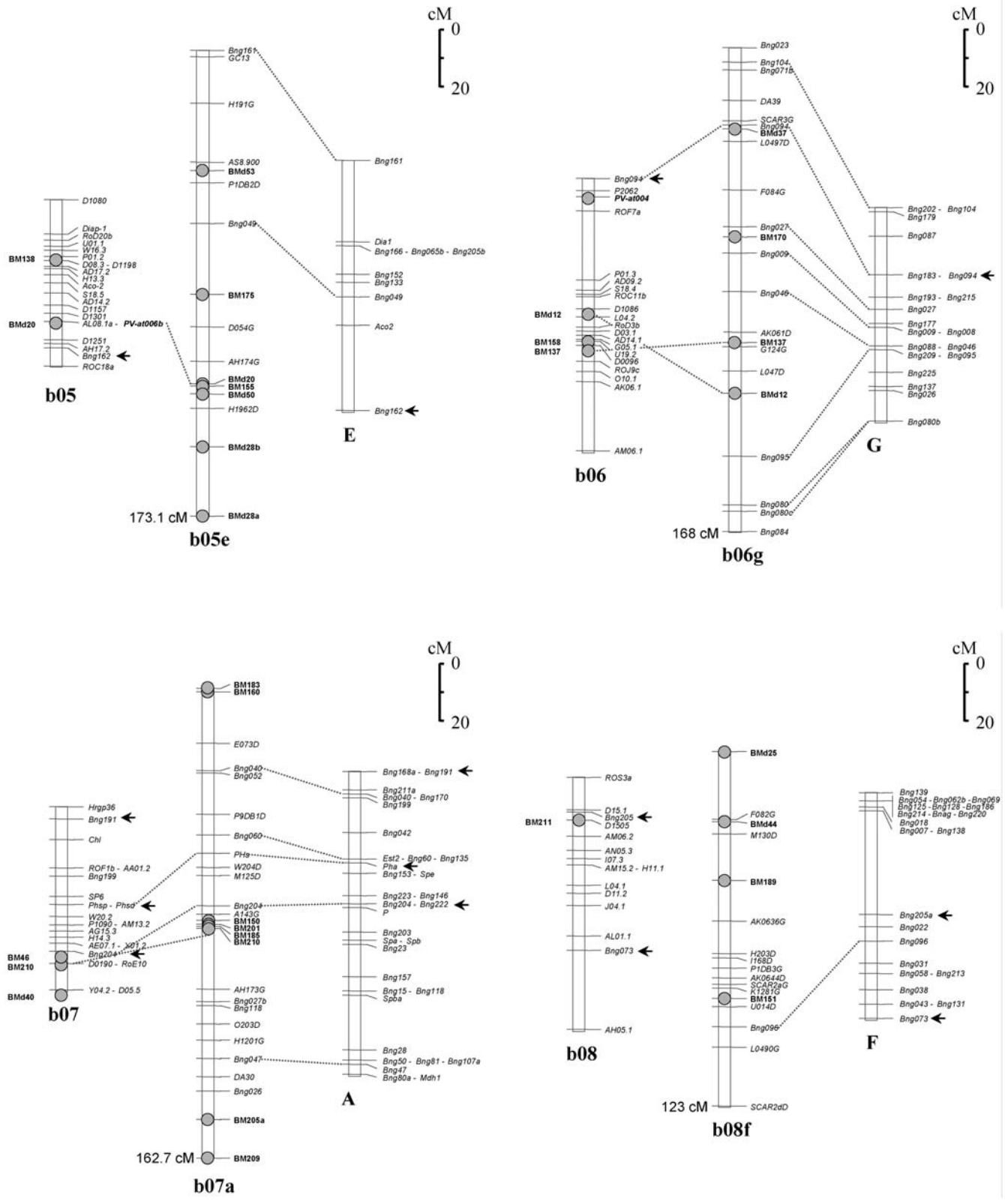

Fig. 2 (continued)

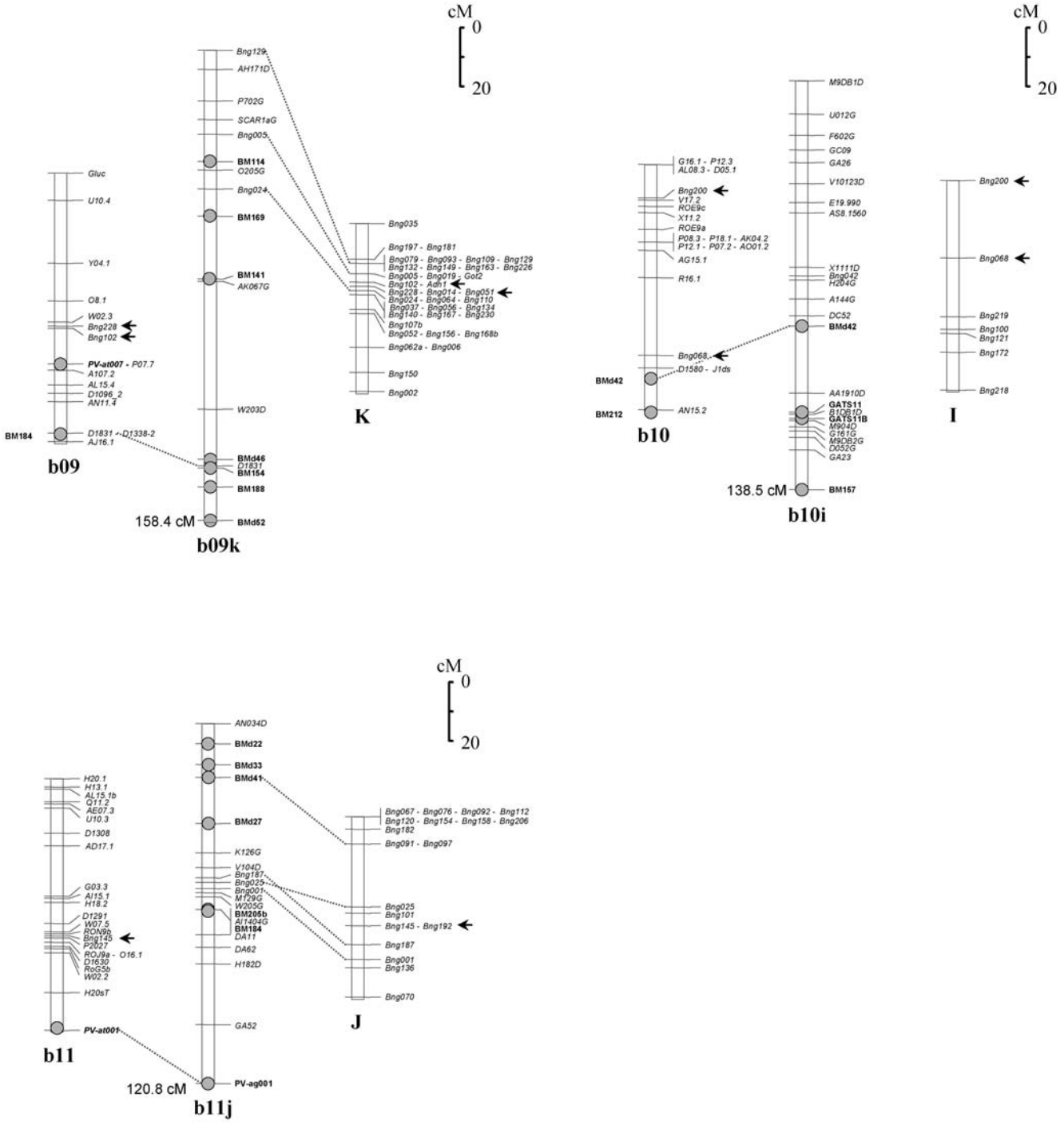

Fig. 2 (continued)

Heterozygosity and segregation distortion

As co-dominant markers the microsatellites were useful for uncovering the amount of residual heterozygosity remaining in the RIL populations and for evaluating segregation distortion occurring in the populations. The majority of RI lines in both populations were homozygous for most microsatellite loci; however, heterozygosity was higher in the DG population (averaging 4.1% across all 78 mapped microsatellites) than in the BJ population (averaging 0.2% across all 22 mapped microsatellites). Meanwhile, significant segregation distortion (P<0.05 for chi-square tests) was evident for 23% (18/78) of the microsatellite loci in the DG population and for 9% (2/22) in the BJ population. These tended to be in regions of significant segregation distortion for other linked co-dominant or dominant markers as well. Using a criterion of finding more than one linked marker with significantly

distorted segregation, distortion for the Andean alleles in the DG population occurred on parts of chromosomes b01h, b02d, B03c and b08 (data not shown). The Mesoamerican allele was predominant on regions of chromosomes b05e, b09k, b10i and b11j. In the BJ populations the Andean allele was more common than the Mesoamerican allele for the microsatellite BM157, and two RFLPs in the equivalent region in chromosome b01 as in the DG population.

Discussion

Due to their high to moderate polymorphism rates, co-dominance and predominantly single-copy nature, the genomic and gene-derived microsatellites were useful for genetic mapping of the two common-bean populations, and a total of 100 new microsatellite loci could be located on the genetic map of common bean using these inter-genepool populations (Fig. 2). These loci are integrated with the 15 microsatellite markers previously mapped on seven chromosomes on the BJ population (Yu et al. 2000). Therefore this study brings to a total of 115, the microsatellite loci located on the common bean genetic map, and provides coverage for every chromosome in the genome with from five to 20 markers each.

The genomic distribution of microsatellite markers in this study tended to be random; however, some large gaps between microsatellites occurred and certain chromosomes contained more microsatellite loci than others. This was probably not due to insufficient polymorphism since many other markers occurred in the intervening intervals between microsatellites. One exception to this may have occurred in the middle of chromosome b01h, a region that suffers from low diversity in other crosses (unpublished data, this laboratory) and is known to be associated with the domestication syndrome (Koinange et al. 1996). Additionally, many of the linkage-group ends were not covered by microsatellite markers, perhaps reflecting a bias against telomeres or a bias towards internal locations.

Greater clustering occurred with the genomic microsatellites developed from enriched libraries, but less so for the new Genbank-derived genic or genomic microsatellites. In previous studies, gene-based microsatellites generally do not cluster (Scott et al. 2000; Cordeiro et al. 2001), while some genomic microsatellites especially those associated with retrotransposons do (Ramsay et al. 1999). The non-random distribution of microsatellites could also be due to bias for certain sequences that occur during marker development, especially from enriched libraries. Clustering of microsatellites has been observed for markers developed from genomic libraries in several but not all crops (Areshchenkova and Ganal 1999; Ramsay et al. 2000; Tang et al. 2002). In one case, markers detecting long di-nucleotide repeats derived from enriched libraries were clustered around centromeric regions (Jones et al. 2002).

To fill present gaps in the common-bean genetic map, it will be useful to saturate the map with additional microsatellites. Techniques that target microsatellites to certain regions (Cregan et al. 1999a), or that take advantage of other SSR motifs that might be better distributed in the genome (Jones et al. 2002), would be useful to develop additional markers for this effort.

As single-locus markers, the microsatellites in this study were specific to a given place in the genome and this allowed them to be used for comparative mapping across populations. A total of 14 microsatellites were reciprocally mapped across both the DG and BJ populations, including four microsatellites mapped on the BJ population by Yu et al. (2000) (Pv-ctt001, Pv-gaat001, Pv-at003 and Pv-ag004) that were mapped in the DG populations; four of the new microsatellites (BMd1, BMd12, BM137 and BM210) that were mapped on both the DG and the BJ population; and six additional BMd microsatellite markers that were equivalent to Genbank sequences mapped on the BJ population by Yu et al. (2000) (BMd17=Pv-gccacc001, BMd18=Pv-gaat002, BMd1=Pv-at008, BMd15=Pv-atgc001, BMd16=Pv-atgc002 and BMd20=Pvat006b) and that were placed on the DG population (Fig. 2). This comparative mapping showed the consistency of microsatellite location on both populations: with all the microsatellites mapping to the same individual chromosome and equivalent map locations in each of the populations. Relatively little map expansion occurred with the inclusion of the microsatellite markers in the DG map relative to the framework RFLP and RAPD maps for this population.

Mapping in two populations was useful for placing microsatellites that were monomorphic in one or the other of the populations. Eleven BM and three BMd microsatellites that could not be mapped in the DG population due to monomorphism were mapped in the BJ population. One microsatellite (Pv-ag003) that could not be mapped on the BJ populations by Yu et al. (2000) due to monomorphism was polymorphic for DOR364 and G19833, and could be placed on the DG map on chromosome 1. In contrast, the microsatellite Pv-at001, listed as monomorphic however, mapped to chromosome 11 by Yu et al. (2000) and was mapped in the DG population to chromosome 4. Meanwhile we mapped Pv-ag001 to the equivalent location on chromosome 11 shown for Pv-at001 in Yu et al. (2000) possibly indicating previous misidentification of this locus.

In combination with the RFLP markers, the co-mapped microsatellites allowed us to link the DG map with those of Freyre et al. (1998) and Vallejos et al. (1992). Comparative mapping across all three genetic maps allowed us to determine the identity and orientation of each linkage group in comparison with the BJ and UF maps, and to obtain a more-accurate position for each of the microsatellites (Fig. 2). The number of links between the DG and BJ maps was fewer than those between the DG and UF maps due to the large number of RFLP clones that were mapped in this second group of populations; however, the microsatellites created more linkages across the first group of populations: BJ and DG maps were linked by a total of 20 jointly mapped markers consisting

of both five RFLP clones (D1595, D1831, Bng071, Bng094 and Bng204), one biochemical marker (Phs) and the 14 microsatellites described above (an average of 1.8 markers per chromosome). Meanwhile the DG and UF maps were linked by a total of 38 jointly mapped RFLP- or RFLP-derived markers (an average of 3.5 markers per chromosome).

The synteny between maps as reflected by the map order of jointly mapped markers was conserved except for segments on chromosome 1 and chromosome 11 which differed between the DG and UF maps. The synteny between the DG and other two maps provides further evidence that no major rearrangements have occurred in the common-bean genome (Freyre et al. 1998).

Comparative mapping showed that segregation distortion occurred in some equivalent positions in both the DG and BJ genetic maps. Freyre et al. (1998) found preferential transmission of the Andean allele on chromosomes 1 and 3, and of the Mesoamerican allele on chromosomes 2, 6, 7 and 10. In this study the prevalent allele came from the same genepool for chromosomes 1, 3, 6 and 10 but from different genepools for chromosome 2. The common pattern of segregation distortion between the two genetic maps probably reflects the location of genes for incompatibility or adaptation that influenced selection in both populations. Given that both populations were developed by single-seed descent from Mesoamerican × Andean genepool-crosses (Beebe et al. 1998; Freyre et al. 1998), the pattern of segregation distortion may reflect similar selection histories. Patterns of segregation distortion across multiple populations have previously been associated with sterility genes (Xu et al. 1997). Segregation distortion in the DG population also occurred on additional linkage groups to those identified by Freyre et al. (1998), possibly reflecting some unique genes or gene clusters that influenced selection in this population.

The majority of the microsatellites, mapped in both the DG and BJ populations, produced single bands and detected single loci; however, four of the gene-based and three of the genomic microsatellites (6.0% and 3.7% of the total, respectively) produced consistent multiple banding pattern and some of these detected duplicate loci. The placement of these loci across the genome for the gene microsatellites, may reflect the distribution of duplicated gene families. The majority of genomic microsatellites tend to be single-copy in other organisms as well (Jones et al. 2002).

Several other differences were also seen between genomic and gene microsatellites. Gene microsatellites proved to be less-polymorphic than genomic microsatellites in both inter-genepool populations, although the overall polymorphism rates were similar among the two populations. Gene microsatellites have been seen to be less-polymorphic than genomic microsatellites in other crops as well (Cho et al. 2000; Schloss et al. 2002). Non-amplification was more frequent in gene microsatellites compared to genomic microsatellites, perhaps due to intervening intron sequences which may not have been considered in the design of primer pairs from EST sequences.

The current set of microsatellite markers described in this study provides the basis for anchoring and aligning genetic maps one to each other based solely on PCR-based markers, something that previously was done with single-copy RFLP markers or by tentatively associating RAPD bands found in different populations. Therefore, the microsatellites make ideal second-generation markers for the whole genome-analysis important for gene tagging and quantitative trait loci studies. The microsatellites mapped during the course of this research will also be invaluable for marker-assisted selection because they are simple to analyze, specific for single genes of interest and diagnostic in most crosses due to their high level of polymorphism. The mapped microsatellites can also provide a good set from which to chose markers for studies of genetic diversity in common bean. They are also amenable to the evaluation of a large number of fixed or segregating individuals and to high throughput marker analysis.

Acknowledgements We appreciate help from Patricia Zamorano, the parental survey from Martha Giraldo and Alma Viviana Gonzalez, and seed preparation from Agobardo Hoyos and Margarita Mauro.

References

Afanador LK, Hadley SD (1993) Adoption of a mini-prep DNA extraction method for RAPD marker analysis in common bean. Bean Improv Coop 35:10–11

Areshchenkova T, Ganal MW (1999) Long tomato microsatellites are predominantly associated with centromeric regions. Genome 42:536–544

Beebe SE, Pedraza F, Rojas M, Gutierrez J, Tohme J (1998) A genetic map combining RFLP, RAPD, SCAR and AFLP markers. Bean Improv Coop 41:95–96

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench] Theor Appl Genet 93:190–198

Cheng-Dao L, Fatokun CA, Ubi B, Singh BB, Scoles GJ (2001) Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. Crop Sci 41:189–197

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park J, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross-transferable to erianthus and sorghum. Plant Sci 160:1115–1123

Cregan PB, Mudge J, Fickus EW, Marek LF, Danesh D, Denny R, Shoemaker RC, Matthews BF, Jarvik T, Young ND (1999a) Targeted isolation of simple-sequence repeat markers through the use of bacterial artificial chromosomes. Theor Appl Genet 98:919–928

Cregan, PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, Vantoai TT, Lohnes DG, Chung J, Specht JE (1999b) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Edwards KJ, Baker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques 20:759–760

Eujay I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Freyre R, Skroch PW, Geffory V, Adam-Blondon AF, Shirmohamadali A, Johnson WC, Llaca V, Nodari RO, Periera 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

Gaitán-Solís E, Duque MC, Edwards KJ, Tohme J (2002) Microsatellite repeats in common bean (*Phaseolus vulgaris*): isolation, characterization, and cross-Species amplification in *Phaseolus* ssp. Crop Sci 42:2128–2136

Gepts P (1999) Development of an integrated linkage map. In: Singh SP (ed) Common bean improvement for the twenty first century. Kluwer, Dordrecht, pp 389–400

Hopkins MS, Casa AM, Wang T, Mitchell SE, Dean RE, Kochert GD Kresovich S (1999) Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. Crop Sci 39:1243–1247

Huttel B, Winter P, Weising K, Choumane W, Weigand F, Kahl G (1999) Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Genome 42:210–217

Jones ES, Dupall MO, Dumsday JL, Hughes LJ, Forster JW (2002) An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 102:405–415

Kelly JD, Miklas NP (1998) The role of RAPD markers in breeding for disease resistance in common bean. Mol Breed 4:1–11

Koinange EMK, Singh SP, Gepts P (1996) Genetic control of the domestication syndrome in common-bean. Crop Sci 36:1037–145

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

Mba, REC, Stephenson P, Edwards K, Melzer S, Nkumbira J, Gullberg U, Apel A, Gale M, Tohme J, Fregene M (2001) Simple sequence repeat (SSR) marker survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of casava. Theor Appl Genet 102:21–31

Morgante, M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Ramsay L, Macaulay M, Cardle L, Morgante M, Degliivanissevich S, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J 17:415–425

Ramsay L, Macaulay M, Degliivanissevich S, Maclean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste K, Ganal M, Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997–2005

Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers RE, Paterson AH, Kresovich S (2002) Characterization of RFLP probe sequences for gene discovery and SSR development in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 105:912–920

Schoonhoven A, Voysest O (1991) Common beans: research for crop improvement. Redwood Press, Melssham, Wiltshire, UK

Scott KD, Eggler P, Seaton G, Rossetto EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Singh SP (1999) Production and utilization. In: Singh SP (ed) Common bean improvement for the twenty-first century. Kluwer, Dordrecht, pp 53–91

Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124–1136

Temnykh S, DeClerck, Lukashova A, Lipovich L, Cartinhour S, McCouch SR (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.). Frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

Thoquet P, Ghérardi M, Journet EP, Kereszt A, Ané JM, Prosperi JM, Huguet T (2002) The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. BMC Plant Biol 2:1

Vallejos CE, Sakiyama NE, Chase CD (1992) A molecular-marker based linkage map of *Phaseolus vulgaris* L. Genetics 131:733–740

Xu Y, Zhu L, Xiao J, Huang N, McCouch SR (1997) Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled-haploid, and recombinant inbred populations in rice (*Oryza sativa* L.) Theor Appl Genet 253:535–545

Yu K, Park SJ, Poysa V (1999) Abundance and variation of microsatellite DNA secuences in beans *(Phaseolus* and *Vigna*). Genome 42:27–34

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.). Journal of Heredity 91:429–434