Development of microsatellite markers for common bean (*Phaseolus vulgaris* L.) based on screening of non-enriched, small-insert genomic libraries

Matthew W. Blair, Monica Muñoz Torres, Fabio Pedraza, Martha C. Giraldo, Hector F. Buendía, and Natalia Hurtado

Abstract: Microsatellite markers are useful genetic tools for a wide array of genomic analyses although their development is time-consuming and requires the identification of simple sequence repeats (SSRs) from genomic sequences. Screening of non-enriched, small-insert libraries is an effective method of SSR isolation that can give an unbiased picture of motif frequency. Here we adapt high-throughput protocols for the screening of plasmid-based libraries using robotic colony picking and filter preparation. Seven non-enriched genomic libraries from common bean genomic DNA were made by digestion with four frequently cutting restriction enzymes, double digestion with a frequently cutting restriction enzyme and a less frequently cutting restriction enzyme, or sonication. Library quality was compared and three of the small-insert libraries were selected for further analysis. Each library was plated and picked into 384-well plates that were used to create high-density filter arrays of over 18 000 clones each, which were screened with oligonucleotide probes for various SSR motifs. Positive clones were found to have low redundancy. One hundred SSR markers were developed and 80 were tested for polymorphism in a standard parental survey. These microsatellite markers derived from non-SSR-enriched libraries should be useful additions to previous markers developed from enriched libraries.

Key words: DOR364, genomic microsatellites, plasmid clones, simple sequence repeats.

Résumé : Les microsatellites constituent des outils génétiques utiles pour une grande gamme d'analyses génomiques bien que leur développement prenne du temps et nécessite l'identification de séquences simples répétées (SSR) au sein des séquences génomiques. Le criblage de banques à inserts courts sans enrichissement préalable représente une méthode efficace pour isoler des SSR, laquelle peut livrer un portrait de la fréquence des motifs qui soit sans biais. Dans ce travail, les auteurs adaptent des protocoles à haut débit pour robotiser le prélèvement des colonies et la préparation des membranes au cours du criblage de banques de plasmides. Un jeu de sept banques génomiques non-enrichies préparées à partir de l'ADN génomique du haricot fragmenté par sonication ou encore digéré avec quatre enzymes coupant fréquemment ou une double digestion avec ces enzymes et une enzyme coupant moins souvent. La qualité de banques a été comparée et trois des banques à inserts courts ont été retenues pour analyse ultérieure. Chaque banque a été étalée et les colonies disposées dans des plaques à 384 puits, lesquelles ont servi à préparer les membranes à haute densité comprenant plus de 18 000 clones. Ces membranes ont été ensuite criblées avec des sondes oligonucléotidiques pour les différents motifs SSR. Il a été observé que les clones positifs présentaient une faible redondance et 100 marqueurs SSR ont été mis au point dont 80 ont été testés pour le polymorphisme sur un jeu de parents. Ces microsatellites dérivés de banques non-enrichies devraient s'avérer un ajout utile aux marqueurs développés précédemment à partir de banques enrichies.

Mots-clés : DOR364, microsatellites génomiques, clones plasmidiques, séquences simples répétées.

[Traduit par la Rédaction]

Received 26 November 2008. Accepted 25 May 2009. Published on the NRC Research Press Web site at genome.nrc.ca on 22 August 2009.

Corresponding Editor: G. Scoles.

M.W. Blair,[1] M. Muñoz Torres,[2] F. Pedraza,[3] M.C. Giraldo,[4] H.F. Buendía, and N. Hurtado. International Center for Tropical Agriculture (CIAT), A.A. 6713, Cali, Colombia.

[1]Corresponding author (e-mail: m.blair@cgiar.org).
[2]Present address: Department of Biology, Georgetown University, Washington, DC 20057-1229, USA.
[3]Present address: Department of Plant Sciences, North Dakota State University, Fargo, ND 58108-6050, USA.
[4]Present address: Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA.

Introduction

Microsatellites are PCR-based markers that target simple sequence repeat (SSR) loci based on various repeat motifs that are evaluated for size polymorphisms with different electrophoretic systems (Powell et al. 1996). The value of these markers is that they usually detect single loci and are specific to a given place in the genome. They are also well distributed, highly variable, easy to amplify with standard PCR conditions, and readily scored as codominant markers (Morgante and Olivieri 1993; Hancock 1999). Microsatellite markers have been developed for a wide range of plant species and have been useful for analysis of crop genomes (Gupta and Varshney 2000). Furthermore, microsatellites have been found to be well distributed along most chromo-

somes and genomic regions of eukaryotes, making them very appropriate for genetic mapping (Hancock 1999; Tóth et al. 2000). In addition, they can be used in diversity assessment within closely related germplasm and for marker-assisted selection of linked traits (Powell et al. 1996; Blair et al. 2006, 2007).

Various techniques exist for developing microsatellite markers, based on either mining existing sequences or discovering new microsatellite markers from anonymous genomic DNA. Most isolation protocols rely on the availability of DNA libraries, although some non-library techniques exist as well (Zane et al. 2002). When using libraries for microsatellite isolation, it is best if these consist of small-insert bacterial or phage clones because they are easier to screen for SSRs and to sequence if they are shown to be positive. In the development of microsatellites from small-insert libraries, two general approaches have been used: either no enrichment has been practiced, or libraries have been created from SSR-enriched fractions of the genome. Several enrichment procedures have been used to increase the prevalence of SSRs in genomic libraries. One method relies on the selective capture of SSR-containing fragments using oligonucleotides on Hybond charged nylon membranes (Edwards et al. 1996), while others have used streptavidin beads (Dutech et al. 2007).

Enriched libraries have been used fairly frequently in plants, while in other taxa direct library screening methods are favored instead (Zane et al. 2002). Both methods have advantages and disadvantages. One advantage of screening libraries without enrichment is that this procedure can be used to evaluate a wider array of SSR motifs and uncovers the true frequency of repeat types while avoiding biases that can occur with the construction of enriched libraries. Enriched libraries, on the other hand, have the advantage of sometimes producing more positive clones, although clone redundancy has been a problem with most enrichment procedures (Squirrell et al. 2003). A final method that avoids the use of small-insert libraries of either type is the screening of sequence databases (Temnykh et al. 2001); however, this has the limitation of relying on the source of the sequence information and therefore, like enriched libraries, can be biased towards one portion of the genome. For example, expressed sequence tag (EST) or cDNA sequencing has provided a valuable source of microsatellites in various crop species, producing what are commonly known as EST-SSR markers (Varshney et al. 2005); however, this represents only the coding region of the genome, which is thought to have fewer microsatellite loci than non-coding regions.

Common bean (*Phaseolus vulgaris* L.) is an important leguminous crop that originated in the New World tropics (Broughton et al. 2003). The crop is distributed worldwide in temperate, subtropical, and highland tropical environments and there are active breeding programs in most countries of Central, North, and South America as well as central, eastern, and southern Africa. Furthermore, over 75 000 accessions of common bean and close relatives are held in germplasm banks around the world, with the largest collection at the International Center for Tropical Agriculture (CIAT) in Cali, Colombia.

In common bean, two broad categories of microsatellite markers have been developed. The first microsatellites to be developed were gene-based markers produced by mining of database sequences (Yu et al. 1999, 2000; Blair et al. 2003; Guerra-Sanz 2004). Meanwhile, several sets of genomic microsatellites have been developed for common bean from enriched libraries produced from genomic DNA by various SSR capture techniques including bead- or membrane-bound oligonucleotides (Gaitán et al. 2002; Métais et al. 2002; Buso et al. 2006; Benchimol et al. 2007; de Campos et al. 2007). Recently, Hanai et al. (2007) screened 3000 unigenes *in silico* for additional genic microsatellites. To date, however, very few microsatellites have been developed for common bean by direct screening of non-enriched libraries. Only a small group of genomic microsatellites has been developed by direct screening of subcloned bacterial artificial chromosome (BAC) clones specific to an anthracnose resistance locus (Caixeta et al. 2005).

The objective of this research was to develop and screen non-enriched libraries with simple sequence repeat motifs and determine whether these small-insert libraries could be useful sources of genomic microsatellites. The libraries were generated by screening genomic DNA for digestion with a set of frequently cutting enzymes to determine which were most appropriate, then cloning the DNA fragments into a common vector and arraying the clones in high-density filters, which were hybridized with SSR-containing oligonucleotide probes. The screening of non-enriched random genomic and cDNA libraries has allowed us to investigate the frequency with which different microsatellite motifs occur in the common bean genome and to develop a new set of genomic microsatellites for mapping and tagging projects in common bean.

Materials and methods

Plant materials and DNA extraction

Total genomic DNA was extracted for the genotypes DOR364 and G19833 using a CTAB extraction method as described in Afanador and Haley (1993) and then used in cloning experiments at a concentration of 100 ng/μL as measured on a Hoefer DyNA Quant 2000 fluorometer. DOR364 is a variety developed by CIAT and released in several countries of Central America (Costa Rica, Honduras, El Salvador, Nicaragua). It has small red seeds and belongs to the Mesoamerican gene pool of common bean. G19833 is a CIAT germplasm accession with large, yellow and red mottled seed. It is a landrace originally collected in Peru that belongs to the Andean gene pool of common bean.

Library construction

The following library construction techniques were tested: (1) *Alu*I digestion (from DNA of the genotype DOR364), (2) *Hae*III digestion (DOR364), (3) *Rsa*I digestion (DOR364), (4) *Taq*I digestion (DOR364), (5) *Dra*I/*Alu*I digestion (G19833), and (6) *Dra*I/*Hae*III digestion (G19833) with enzymes provided by Promega (Madison, Wisconsin). For these enzyme-digested libraries, bean DNA was digested with the appropriate restriction enzyme(s), size-selected to the range of 0.4–1.2 kb, and ligated into vectors using T4 DNA ligase. In the case of the *Rsa*I library the insert was ligated into the pGEM-T Easy vector from Promega, while for the *Taq*I library a modified pUC19 vector (pJV1) was

used. The other libraries were made in pBluescript (pBS) KS+ vectors. All ligation reactions were transformed into Electromax DH5α *E. coli* cells and plated onto LB agar media containing the appropriate antibiotic (100 mg/L), X-gal, and IPTG. DNA was extracted from 25 white colonies per library by standard alkaline lysis (Sambrook and Russell 2001) procedures using an Autogen robot (Holliston, Massachusetts) and checked for insert presence and size by restriction digestion with appropriate enzymes. The clones from the small-insert libraries with the best insert size (approx. 0.5 kb) and transformation efficiency were plated on square plates for blue–white screening, and a Q-bot robot (Genetix, Boston, Massachusetts) was used to pick and array colonies for each of the libraries. Automatic blue–white screening was used to pick the clones. All the clones were placed into 384-well plate format glycerol stocks, grown overnight, and copied twice into working and master copies of the libraries. The clones were then spotted onto gridded Hybond N+ nylon filters (GE Healthcare Life Sciences, Piscataway, New Jersey) containing 6 fields of eight 384-well plates each, for a total of 18 432 clones per filter. Clones were arrayed on these filters in a double-replicate 4 × 4 pattern with duplicate spots for each address. The 3 selected small-insert libraries were named Pv for the species and DOR for the source genotype (DOR364), namely Pv-DORa for the *Alu*I library, Pv-DORb for the *Hae*III library, and Pv-DORc for the *Rsa*I library.

Library screening

The 3 selected libraries were screened for microsatellites with a filter hybridization approach. Six SSR motif oligonucleotide probes were used to screen the filters for each library, with 2 probes that targeted dinucleotide repeat motifs, namely $(CA)_{20}$ and $(GA)_{20}$, combined together, and 4 probes that targeted trinucleotide repeat motifs, namely $(AAT)_{14}$, $(CAG)_{14}$, $(CAA)_{14}$, and $(ACG)_{14}$, combined together. In each case, the sequence within the parentheses indicates the repeat and the number outside the parentheses indicates the number of copies of that repeat. Hybridization consisted of end-labeling the SSR motif probes with [γ-^{32}P]ATP and T4 DNA kinase, and hybridizing these probes to the DNA contained on the filters with the protocol of Edwards et al. (1996). Briefly, the filters were pre-hybridized in 100 mL of hybridization buffer for 4–6 h at 60 °C. Meanwhile, 10 pmol of probe was end-labeled with 1 μL of T4 polynucleotide kinase (7 units/μL) and 5 μL of [γ-^{32}P]ATP in a total reaction volume of 20 μL that was incubated at 37 °C for 80 min and stopped at 65 °C for 20 min. The labeled oligonucleotide was added directly to the pre-hybridizing filters and incubated at 60 °C for 12 h. After the hybridization step, the filters were washed twice at 60 °C with 6× SSC, 0.1% SDS for 5 min each. Longer washes were used when signal was intense (>100 000 cycles per minute). The filters were blotted dry, covered with plastic wrap, and arranged faceup in cassettes along with 3 sheets of X-ray film. The films were taped to each other so that they would not shift during the 0 to 8 day exposure in a –80 °C freezer. Films were developed after exposure intervals of 6, 24, and 72 h to identify high-, medium-, and low-signal positive clones, respectively. Filters were reused for sequential screening of different oligonucleotide repeats by stripping them after each use. Stripping consisted of washing the filters twice at room temperature in 100 mmol/L NaOH, 10 mmol/L EDTA, 0.1% SDS, followed by a 5× SSPE rinse (0.75 mol/L NaCl, 0.05 mol/L NaH_2PO_4, 5 mmol/L EDTA adjusted to pH 7.4 with 10 N NaOH) for 10 min and storage in a 4 °C refrigerator.

Clone identification and sequencing

Positive clones were identified by which filter they were on, which field within the filter they were in, and what address they had within the field. Each filter contained 6 fields and each field contained the equivalent of eight 384-well plates' worth of clones, for a total of 48 plates per filter. Clones could be identified by their position in the double-replicate 4 × 4 pattern found at each grid axis in the address system. Only double-spotted clones were selected. Any spots for which the replicate did not hybridize were considered false positives and were not selected. Putative SSR-containing clones were picked from their appropriate position in the 384-well plate format glycerol stocks. The positive clones were sequenced initially from one end of the insert by dideoxy sequencing at the Clemson University Genomics Institute using T7 or T3 high-temperature primers. The sequences were searched for vector segments to check for insert integrity and were screened for quality with the program Sequencher v. 4.6 (GeneCodes Corp., Miami, Florida). In addition, they were evaluated for homology to the nr/nt nucleotide database sequences at http://blast.ncbi.nlm.nih.gov/Blast.cgi using nucleotide BLAST and BLASTx searches.

Primer design and microsatellite testing

SSRs were found in the sequenced clones using the SSR Identification Tool (SSRIT), which screens for all possible dimeric, trimeric, and tetrameric repeats (Temnykh et al. 2001; available from http://www.gramene.org/db/markers/ssrtool). Primers were designed using Primer 3.0 (Rozen and Skaletsky 2000; available from http://www.broadinstitute.org/genome_software/other/primer3.html) to have consistent melting temperatures of 55 °C or higher and an average length of 20 nucleotides and to produce PCR amplification fragments that were, on average, 150 bp long. Primer pairs were checked to make sure that they had similar melting temperatures and did not suffer from palindromes or end pairing. New markers were tested with DOR364 and G19833 genomic DNA, as these were the genotypes used to create the libraries. 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 was carried out in a 20 μL final volume containing 50 ng of genomic DNA, 0.1 mmol/L of both the forward and reverse primers, 10 mmol/L Tris-HCl (pH 7.2), 50 mmol/L KCl, 1.5 to 2.5 mmol/L $MgCl_2$, depending on the primer combination, 250 mmol/L of total dNTPs, and 1 unit of *Taq* polymerase. Any primer pairs not amplifying parental DNA under these conditions were not considered further, while those that produced a band for each of the parents were used for diversity assessment with 18 genotypes from survey I as described in Blair et al. (2006), including 7 genotypes from the Andean gene pool

and 11 from the Mesoamerican gene pool, with a total of 15 cultivated and 3 wild accessions, these last genotypes representing Argentinean, Colombian, and Mexican wild diversity. Microsatellites were detected on silver-stained polyacrylamide gels. The PCR product was mixed with 5 μL of formamide containing 0.4% bromophenol blue and 0.25% (*w*/*v*) xylene cyanol FF and denatured at 92 °C for 2 min; 2 μL was loaded onto a 4% denaturing polyacrylamide (29:1 acrylamide:bis-acrylamide) gel containing 5 mol/L urea and 0.5× TBE and run in a Sequi-Gen GT electrophoresis unit (Bio-Rad, Hercules, California) at a constant power of 120 W. PCR amplification products were detected by silver staining according to Blair et al. (2003).

Results

Library construction and screening

A total of 6 enzyme-digested libraries were created for this research. Table 1 shows the average insert size of each library, the percentage of clones with inserts, and the transformation efficiency of each ligation reaction. All of the single-enzyme libraries were found to be of adequate transformation efficiency (10^5 to 10^7 clones / μg of plasmid DNA) and the percentage of clones with missing inserts was found to be low (0% to 29%) except in the *Taq*I library, which was not processed further for this reason. However, even in this library the number of clones with missing inserts may have been overestimated, since read-through of the β-galactosidase gene at the multiple cloning sites was possible given the small size (0.5 to 0.7 kb) of the cloned DNA fragments, especially if the insert was from a protein-encoding region. This was borne out by the observation of light blue rather than dark blue colonies, which may have contained inserts but were counted as clones with missing inserts.

Average insert size was an important criterion by which we decided to select several of the libraries for further screening. The 3 small-insert genomic libraries with average insert sizes closest to 0.5 kb were the libraries made with *Alu*I, *Hae*III, and *Rsa*I (frequently cutting restriction enzymes). This insert size was determined to be favorable for microsatellite discovery because it would be more likely than a smaller average insert size to provide flanking sequences for primer design, while larger insert size libraries were likely to be more difficult to screen for SSR motifs. Double digestion with *Dra*I and either of the frequently cutting enzymes *Alu*I or *Hae*III produced a smaller average insert size (0.3 and 0.4 kb, respectively), which is why these libraries were not evaluated further. In addition, a final library was made from sonicated DNA (of DOR364) but was not evaluated further because of lower transformation efficiency (10^4) and a smaller percentage of clones with inserts (56%), which might have been due to the difficulty of cloning blunt-end DNA from the sonication procedure into the pBS vector.

From the 3 libraries chosen for further analysis, a total of 55 296 clones were arrayed onto high-density filters by a Q-bot robot. These libraries each represented a total of approximately 9.2 Mb of DNA, which is equivalent to 0.014 genome equivalents, considering that the genome size of common bean is 650 Mb. Taken together, the 3 libraries contain approximately 27 Mb of bean DNA, which should be equivalent to 4% of the total genome. This coverage was deemed to be adequate for assessing the frequency of microsatellites detectable by hybridization-based screening, so the filter sets for each library were screened with SSR-containing probes.

In the process of hybridization-based screening, we were interested in calculating the overall frequency of di- and trinucleotide repeats; therefore, these types of probes were hybridized separately. Table 2 shows the number of positive clones found in each hybridization experiment with each of the libraries and with each set of oligonucleotide probes. A total of 282 clones hybridized with the dinucleotide probes, while 135 clones hybridized with the trinucleotide probes, giving a total of 417 positive hits across the 3 libraries. For those repeats screened for, this number of positive clones would be equivalent to a positive hit rate of 0.75% or 1 microsatellite every 66 kb of sequence. All 3 libraries produced around the same number of positive hits, from 131 to 147, but the most productive library overall was the *Hae*III library. The ideal exposure time for the hybridized high-density filters was found to be from 24 to 72 h, producing 43% and 46% of the positive clones, respectively. Meanwhile, a short exposure time of only 6 h produced merely 11% of the positive hits in the radiographic system used for these experiments, while exposures of 24 h or longer produced 89% of the positive clones. Trinucleotide probes had noticeably lower signal intensity than dinucleotide probes, as reflected in the time required to expose the positive clones.

Marker development and testing

After the hybridization screening step, the positive clones from each of the libraries were isolated according to their addresses on the high-density filters and sequenced by Sanger sequencing. Sequencing results confirmed that the majority of the hybridizing clones from each library contained simple sequence repeats and provided information on what repeat motif and what repeat length these sequences had. The sequencing results confirmed that clones containing dinucleotide repeats were more common than other types of clones, confirming the earlier results based solely on hybridization. Among the clones that were confirmed to have dinucleotide repeats, the CA motif was 50% more common than the GA motif; meanwhile, among the clones confirmed to have trinucleotide repeats, there was a mix of motifs. Clone redundancy was not a problem in the libraries, as no redundant clones were identified among the sequenced positive hits.

To evaluate the utility of the SSRs discovered through this library screening procedure, SSR-containing clone sequences where the repeat was flanked on both sides by high-quality sequence were used to design microsatellite primers for the development of new markers for common bean (Table 3). Newly designed markers were designated with the prefix BM, starting at BM213, the number subsequent to the genomic SSR markers developed by Gaitán et al. (2002). A total of 100 microsatellite primer pairs were made for the small-insert clones, with most from dinucleotide repeat containing clones (66 markers), some from trinucleotide repeat containing clones (29), and a few from tetranucleotide repeat containing clones (5). Many of the di- or trinucleotide-based markers (44%) targeted compound SSR motifs. Among the markers based on trinucleotide repeats, various motifs were

Table 1. Small-insert libraries prepared for microsatellite screening in common bean.

Subcloning method	Restriction enzyme	Bean genotype	Vector	Insert size (kb)	Clones with insert (%)	Transformation efficiency	Library name
Enzyme digestion, ligation	*Alu*I	DOR364	pBS	0.5	71	1.02×10^5	PV-DORa
	*Hae*III	DOR364	pBS	0.5	75	3.86×10^5	PV-DORb
	*Rsa*I	DOR364	pGEM	0.5	100	7.53×10^5	PV-DORc
	*Taq*I	DOR364	pJV1	0.7	35	5.15×10^5	—
	*Dra*I/*Alu*I	G19833	pBS	0.3	80	1.73×10^7	—
	*Dra*I/*Hae*III	G19833	pBS	0.4	50	1.61×10^7	—
Sonication (blunt ending), ligation	—	DOR364	pBS	0.5	56	7.56×10^4	—

Table 2. Number of positive clones detected by hybridization of each library, at each exposure time and with each set of oligonucleotide probes.

		No. of positive clones			
Probe type	Exposure time (h)	*Alu*I library	*Hae*III library	*Rsa*I library	Total
Dinucleotide	6	6	2	29	37
	24	78	57	11	146
	72	25	46	28	99
	Total	109	105	68	282
Trinucleotide	6	1	0	6	7
	24	21	10	4	35
	72	8	32	53	93
	Total	30	42	63	135
Overall total		139	147	131	417

represented, as would be expected based on the hybridization probes used, while among the markers based on dinucleotide repeats, more were for AC motifs (31) than for AG (27) motifs. The "nucleotide blast" results are reported in Table S1; most similarities occurred between BM marker sequences and other SSR-containing sequences from common bean such as IAC86, IAC109, IAC119, PvBR20, and FJ30.[5] Only 3 BM markers (BM216, BM245, and BM301) may be related to expressed gene sequences based on the "blastx" search results and an *E* value of 10^{-30} as a threshold for similarity. Three other markers (BM234, BM265, and BM311) were similar to intergenic ribosomal DNA sequences from *Phaseolus coccineus* based on "blastn", while 10 sequences were related to either common bean (BM228, BM264, BM274, BM284, and BM294) or soybean (BM239, BM240, BM255, BM224, and BM303) genomic sequences. The common bean genomic sequences with most similarity to the 5 markers listed above were from a single BAC from linkage group b04 (GenBank accession No. DQ323045.1). Finally, 3 markers (BM245, BM254, and BM293) may be from untranslated regions or intronic regions of the genes for the proteins 9-cis-epoxycarotenoid dioxygenase (NCED1), phosphatidylinositol 3-kinase 1 (PI3K1), and plasma membrane H^+-ATPase, respectively.

A polymorphism survey with 18 genetic mapping parents was carried out with the first 80 BM markers and showed that the newly developed markers had PIC values ranging from 0.08 to 0.90, with 2 markers producing multiple-banding patterns and 2 that did not amplify all genotypes and were not used for PIC estimates. The distribution of PIC values for those markers that were polymorphic is shown in Fig. 1, and the average PIC values were 0.45 for dinucleotide-containing markers and 0.32 for trinucleotide-containing markers. The highest PIC values, from 0.70 to 0.90, were for the dinucleotide markers BM221, BM239, BM238, BM275, BM218, BM278, and BM236, while the highest for the trinucleotide markers were for BM287 and BM240, with PIC values of 0.64 and 0.67, respectively.

Discussion

As part of a program to develop microsatellite markers for common bean and to analyze their frequency in the genome, we created a set of 7 non-enriched, small-insert genomic libraries and screened them for di- and trinucleotide SSR motifs. The libraries created with 3 of the frequently cutting restriction enzymes (*Alu*I, *Hae*III, *Rsa*I) proved to be more useful (owing to the GC content of the restriction site, lack of methylation sensitivity, appropriate insert size, and transformation frequency) than the libraries created with sonication, with the other frequent cutter (*Taq*I), or with double digestion using *Alu*I or *Hae*III and a less frequently cutting restriction enzyme (*Dra*I).

[5] Supplementary data for this article are available on the journal Web site (http://genome.nrc.ca) or may be purchased from the Depository of Unpublished Data, Document Delivery, CISTI, National Research Council Canada, Building M-55, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada. DUD 5254. For more information on obtaining material refer to http://cisti-icist.nrc-cnrc.gc.ca/eng/ibp/cisti/collection/unpublished-data.html.

Table 3. Microsatellite markers developed from the small-insert libraries.

Marker	GenBank acc. No.	Left primer	Right primer	Product size (bp)	Motif	T_a (°C)	PIC[a]
BM214	GQ149513	TTTGACAAAGCAGCTCCAG	AATTAGAACCTCCTTTAGATACCAA	294	$(TAA)_4$	52	na
BM215	GQ149514	TCTCCTCGCGCCGTTC	CAAGAGACCCAGACCCAGAC	187	$(TCG)_4$	57	na
BM216	GQ149515	GTGGTACGCCAAACCTTCG	ACCACCACACTTCCACCAAT	148	$(GT)_6$	56	0.477
BM217	GQ149516	GCGTATGGCACTCTGAAGGT	GCCCGTGCTTGAGGATTAC	111	$(GT)_5$	55	0.305
BM218	GQ149517	GAGTGGGGAACGAAGCATAA	GAAGTGCATTGCTGAGGTGA	103	$(GA)_{15}$	55	0.794
BM219	GQ149518	GCGTATGGCACTCTGAAGGT	GCCCGTGCTTGAGGATTAC	111	$(TG)_2(TA)_2(TG)_6$	55	0.375
BM220	GQ149519	AGTGACGAACAAGGGACTGG	TCAAGTCAGCCAGCAAGAAA	150	$(GA)_4(CA)_3(CA)_3$	55	0.477
BM221	GQ149520	CCCCTCAGGTTTGGACTCTT	GGAATGAGAGAGAAAGCATGG	149	$(CT)_{21}(CT)_3$	55	0.734
BM222	GQ149521	CCCCTCAGGTTTGGACTCTT	GAGGCTGTTGTAGGAATTGG	249	$(CT)_4$	55	0.477
BM223	GQ149522	TTTGGGTCGTTTACCAGTATCA	TTTATGCCAAAGAGTCATTTCAA	167	$(TA)_4(TA)_7$	54	0.000
BM224	GQ149523	GAGGACCAAACACCGGAAC	CCTCCAGGACAGATGGTGAG	121	$(GGT)_3(GGA)_3$	56	0.375
BM225	GQ149524	GAGCCTGGAAGGAAAAGAAGA	CGCGCAATTAACCCTCACTA	169	$(GCT)_5(GAG)_2$	56	mc
BM226	GQ149525	GGGTGTGAGCGTGTTCATC	ACTTCTCGTCAATCCCAACG	156	$(GGT)_4(GGC)_2$	55	0.375
BM227	GQ149526	CTTGTTCACCCAAACCCTGA	GTTCGCCAATGCTACCTACG	122	$(CTG)_5$	56	0.117
BM228	GQ149527	GTTATGCCCAGAAGCAGAGG	GCCCCCAAATGAGAAACCTA	120	$(AG)_4(CA)_4$	56	0.227
BM229	GQ149528	TGCCCGTGCTTTAGGATTAC	AAAAATCTGTTCACCCAGGA	138	$(CA)_5$	54	0.581
BM230	GQ149529	GCGTATGGCACTCTGAAGGT	TAGTTGCTTGCTCGTGCTTC	121	$(GT)_7$	55	0.000
BM231	GQ149530	TGCCCGTGCTTTAGGATTAG	TGTTCACCCACGGTATCTGA	135	$(CA)_5(TA)_2(CA)_2$	55	0.664
BM232	GQ149531	TGCAGAAGATGGACCAAAGA	GCAAAACATATCGTGCAAGTG	112	$(AC)_3(CA)_4$	54	0.445
BM233	GQ149532	AACCAGTTGCCTGTCCAACA	AGGAAGGACGGCTTTATGCT	177	$(CA)_5(CCA)_2$	56	0.375
BM234	GQ149533	CACCGAATCCGAGAGAACTG	AGTGTTTCCTTCCGGGCTTC	146	$(GCTT)_5$	57	0.414
BM235	GQ149534	GAGTGTTGCACCGTCGAGA	GGTCCTTGCTTGGTGATAGG	145	$(GCA)_4(GCA)_4$	55	0.239
BM236	GQ149535	GCTGAAAAGAGGAGGTCGAA	GGACTTGCCAGAAGAACTGC	130	$(GA)_{25}$	55	0.904
BM237	GQ149536	CTGCAGCTCAAACAGGGC	GCAATACCGCCATGAGAGAT	113	$(GA)_{26}$	55	0.638
BM238	GQ149537	GGTCCCTGATTGAAAAACTAAA	GCAAAACTTTTAGCAATCTTACA	150	$(GA)_{14}$	52	0.744
BM239	GQ149538	CTGCTACTACTCCCACTACTTCA	ATGTAAGCCATTCCCTCTTC	113	$(TC)_{12}$	53	0.737
BM240	GQ149539	CAGCAGAAACAGCAGCAGAA	ATGGAGAGGGTAGCCGATGT	138	$(CCA)_7$	55	0.670
BM241	GQ149540	TGGCACTCTGAAGGTGGTAA	TGGAACCTTGGACAAATTGAG	147	$(TG)_6$	55	0.321
BM242	GQ149541	TCAAGTCAGCCAGCAAGAAA	GCTGAGAATGTTGGGAACG	226	$(TG)_2(TG)_5$	55	0.000
BM243	GQ149542	TGTTCTATTTTCTGACTCCCTCTC	TGTCGACAAAGCAATCAGAC	218	$(TA)_5$	53	0.141
BM244	GQ149543	GCGTATGGCACTCTGAAGGT	TAGTTGCTTGCTCGTGCTTC	119	$(TG)_6$	55	0.000
BM245	GQ149544	ACCCCAAACAGTACCAACCA	CCACCATATCCAAGCCTGTC	186	$(CCA)_5(CCA)_2$	55	0.375
BM246	GQ149545	AGAAACGCTTTGGTTGCTTG	TGGCACTCTGAAGGTGGTAA	122	$(CA)_5(CA)_2$	55	0.141
BM247	GQ149546	CCAACTTCTAAAGCGCGTGT	CACGGTATGCGTTTGGACTA	126	$(TTA)_5$	55	0.099
BM248	GQ149547	ATTCACCTCCGTCTCTGAGC	TCGTGCTTCAAGGTCACGTA	129	$(TG)_5$	55	0.000
BM249	GQ149548	CACGAACTTCATTCATCACTTC	TCTCGGCCATAGGAGTTACC	159	$(TC)_4(TC)_5$	53	0.239
BM250	GQ149549	TTCTTCATCCGACCCAACTC	TGTGTTAGATTCGGGAGTTTCA	165	$(CA)_6$	55	0.000
BM251	GQ149550	AACATTTCCAGCCGAATCAT	TTCAGGAACGATCTCCTTTG	199	$(TAA)_7(TTA)_4$	54	0.141
BM252	GQ149551	CTGGGCGTAGTCCAAGTTGT	GGTACCTGCAAAAATATCAAGC	110	$(CT)_7(CT)_4$	54	0.178
BM253	GQ149552	GGAATCATTTCAACCCCATAAG	ATTCCGCGAAGAAAATAAGG	150	$(CT)_5$	54	0.327
BM254	GQ149553	TCTGTTTGGAAAACTCTATGCTCTC	GATGTTTGTGGAGATCCAAACT	193	$(CTG)_3(CTG)_4$	54	0.374
BM255	GQ149554	GGAATCATTTCAACCCCATAAG	ATTCCGCGAAGAAAATAAGG	150	$(CAC)_6(CA)_3(CT)_5$	54	0.327

Table 3 (*continued*).

Marker	GenBank acc. No.	Left primer	Right primer	Product size (bp)	Motif	T_a (°C)	PIC[a]
BM256	GQ149555	ACCACTGCGCACAGACTCA	GTGCGTTCACGTTCTCCAC	129	$(GTC)_4$	56	0.413
BM257	GQ149556	GTATCTGAGCACCGCGTATG	TGCCCGTGCTTTAAGGTTAC	123	$(TG)_5$	55	0.077
BM258	GQ149557	GTCGCACGTACTGCAACGTA	ATCCCGTGTCAACCCTGTTA	175	$(AG)_6(AG)_5$	56	0.099
BM259	GQ149558	GATTCGGGTTTGATGGTGTT	CCACAAACCCCTTAGTCCAA	148	$(GTT)_4(GTT)_3$	55	0.375
BM260	GQ149559	GCGTATGGCACTCTGAAGGT	TAGTTGCTTGCTCGTGCTTC	119	$(TG)_3(GT)_3(TC)_5$	55	0.000
BM261	GQ149560	CCGATTGTTTATGGCAGTAGAG	TGGATGTAGCTCTGGAAAGGA	126	$(GT)_4(GA)_7$	54	0.000
BM262	GQ149561	TGCCCGTGCTTTAAGGTTAC	GTATCTGAGCACCGCGTATG	123	$(CA)_5$	55	0.000
BM263	GQ149562	GTCCACCCACGTATCTGAGC	TGCCCGTGATTGAGGATTAC	133	$(TG)_2(TG)_5$	56	0.000
BM264	GQ149563	CCACAGAACCACCAAATGAG	AAGCATAGCAGGGTTCTGACA	175	$(GT)_4(CT)_2(CT)_4$	54	0.000
BM265	GQ149564	CACCGAATCCGAGAGAACTG	AAGGCATCCTCGTGTGGTT	175	$(GCTT)_6$	56	0.661
BM266	GQ149565	CTTGCCCGTGCTTTAAGGT	GTCCACCCTCGTATCTGAGC	134	$(CA)_5(CA)_2$	55	0.000
BM267	GQ149566	AAACTGTCCACCCACGTCTC	CTTGCTCGTGCTTCAAGGTT	141	$(GT)_2(GT)_5$	55	0.000
BM268	GQ149567	GCACTCTGATGGGTGGTAAA	GAAACTCTCTGGTTGCTTGC	120	$(GT)_5$	53	0.141
BM269	GQ149568	TCTGAAGGGTGGTAAAAGAACT	GAATGTCGGAAAGAGGGTTG	157	$(GT)_2(GT)_5$	53	0.000
BM270	GQ149569	AACTGTTCACCTCCGTCTCTGA	GGTTACGAACAAGGGACTGG	125	$(TG)_3(TG)_5(TC)_5$	55	0.000
BM271	GQ149570	TGGCACTCTGAAGGTGGTAA	GAAACTCTTTGGTGGCTTGC	121	$(GT)_2(GT)_5$	55	0.239
BM272	GQ149571	GGGAGCTTTAACAAAGGAGCA	AGGGTGCAAATGTCAAAACG	128	$(GA)_{13}$	56	0.661
BM273	GQ149572	TCCACCTACGTCTCTGAGCA	TTTAAATAGCGCGGGGAAT	205	$(TG)_5$	55	0.000
BM274	GQ149573	AGATCACGGAACCACCAATG	GGACACGCACAGACACTCAC	215	$(CT)_5$	56	0.375
BM275	GQ149574	GAGTGGGGAACGAAGCATAA	GTGCATTGCTGAGGTGAGAA	100	$(AG)_{16}$	55	0.787
BM276	GQ149575	AAGAGATCCTTTGCTCATGTG	GGGGAGGAAGGTTGACC	348	$(AT)_{18}$	52	mc
BM277	GQ149576	AGAACCGGCGTTAAAAACTG	GAAGGTTGCCCGATAGTCAG	249	$(CG)_4(TC)_5$	54	0.592
BM278	GQ149577	CCAAGGTACATCTCAAGCAAA	TCAATCACATACATCACATATAATTCA	253	$(AT)_{18}$	53	0.825
BM279	GQ149578	ATCACCGCCTCCTTTCTCTT	AAGAGAAGCCCTGGATTTGG	161	$(TC)_5(GA)_3$	55	0.000
BM280	GQ149579	GAATCCCCCACCAGAATTG	AGCCGGTGCCACAGTATAAC	152	$(CCA)_4$	55	0.000
BM281	GQ149580	TCCGCGAACGGGATCA	AGACATGAGATGAGGCCGATG	246	$(CTG)_6$	58	0.077
BM282	GQ149581	TGAGTAAGATAATGAATAAAGGCTTC	CCACAAAATTCCTTCAAAAA	250	$(TC)_6$	51	0.178
BM283	GQ149582	TTCCTCTACTACTGAACCCATCG	ACAAGAGGAGAGCCAGAACTTTC	162	$(GCT)_5$	55	0.389
BM284	GQ149583	TGGCACTGATGCCGTTATT	ATACGGTGCGTGTGAGTGTG	215	$(AG)_9$	55	0.670
BM285	GQ149584	TGCATACGTTGTGGGTTGTT	AAGGCGAGGAGAACGAAAAT	151	$(TTC)_4$	55	0.573
BM286	GQ149585	TCAAGTCAGCCAGCAAGAAA	TTGCGAACAAGTGACTGGAG	156	$(TG)_9$	55	0.141
BM287	GQ149586	ATGCACCACAAGGGTTGTCT	AGCCAGGTTCAACCCTACAA	234	$(TCC)_7$	55	0.641
BM288	GQ149587	ACAGCAATGCCGGAATAGAA	CACTTTGGGGGTGGTAAAAG	226	$(AC)_5$	55	0.372
BM289	GQ149588	GCCATCCCTACTCCTAACAGC	CTTGAAGCCACGTGAGTCAA	386	$(TTA)_{28}(TTA)_7(TTA)_5$	55	0.000
BM290	GQ149589	GGCCTTTGTTAGCACTTGGA	CAGGTCTCCTCCATTCATTTTG	181	$(GA)_9$	56	0.000
BM291	GQ149590	ACCACACCTCTCAACCAAGG	GGAAGAGTGTGCTGGAGGAA	163	$(GCT)_5$	55	0.099
BM292	GQ149591	ACACACTCCCTCGCAAAATG	GGTGTGGAGGATTTGGACAC	241	$(CCG)_4(ACC)_4$	56	0.375
BM293	GQ149592	CCACACATGCTGCTGCTACT	TCTCACAGTGCTCCCTTCCT	120	$(ATG)_6$	55	0.000
BM294	GQ149593	TGCGACAACATACCTCCAGA	CGCGGTCCAGATAAAGGTG	122	$(GT)_2(GT)_4(GT)_2$	56	—
BM295	GQ149594	TTTTCAAGCATCCGGTCACT	CACATCTGACTTGAGCGTTAGAG	96	$(TC)_3(TC)_2$	55	—
BM296	GQ149595	CCCTTGCTCCCTATTCTTCC	AGATTGCGGAGATGGCTCT	131	$(CCT)_2(CCT)_2(AC)_4$	55	—
BM297	GQ149596	TCGTTCGTCCTTTTCTCACC	CGTGCGAAAATGAAACAGTG	227	$(CT)_7(CT)_6$	55	—

Table 3 (*concluded*).

Marker	GenBank acc. No.	Left primer	Right primer	Motif	Product size (bp)	T_a (°C)	PIC[a]
BM298	GQ149597	CCCGAATCTAACACAACATGG	AGAACGTGGGTTTGGAGTTG	$(GCC)_3(GCC)_2$	162	55	—
BM299	GQ149598	TCTCTCAAACCCCTCTCACAA	GTTGATCTCTCCCTTCCCAAT	$(TA)_4$	122	55	—
BM300	GQ149599	GCTATTTTGGGGGAAATGGT	TACAACCACGCTTCTTGTCC	$(AC)_3(AC)_3$	185	54	—
BM301	GQ149600	CCATTTTGAAAGAGGAATGC	CCAAAAGCAAGGCTTCTGTA	$(AAT)_3$	275	53	—
BM302	GQ149601	AGGAAAACCCTTTGCCATTC	TCTAGGATTCCGCGAAGAAA	$(CA)_5$	190	55	—
BM303	GQ149602	CGCAGACGATGAGAGAGGA	AAAATCTCCGCCGATGGAAC	$(GTC)_5$	118	55	—
BM304	GQ149603	TTGAAGATGCAGACGGGAAT	GGGAACACCAAACCAAGAGA	$(TA)_3(TA)_2$	139	55	—
BM305	GQ149604	TAGTTGCTTGCTCGTGCTTC	GCGTATGGCACTCTGAAGGT	$(CA)_4$	115	55	—
BM306	GQ149605	GGCCTGTCAGCCCTCAAGT	AATTATTTTGGTCTCTAAAAATTGGTT	$(GTTG)_3$	150	55	—
BM307	GQ149606	TGACATCATCGCAACGTTAGT	CTGAAGAGTGAGTAAGTGAG	$(CACT)_5$	175	48	—
BM308	GQ149607	ACACTCCGACGCTTCAAGTC	CAAAATGTAGGCAGCCATGT	$(TC)_5$	175	55	—
BM309	GQ149608	TCTTGCTTACGCGTGGACTA	TCTGGGAGGTGAGGTGAGG	$(CT)_5$	151	55	—
BM310	GQ149609	ACCATGCCCAAAGTGGTAAA	ATTGCACCCAGAGAGCAGAG	$(CT)_2(CT)_3(CT)_8$	158	55	—
BM311	GQ149610	CACCGAATCCGAGAGAACTG	TACGAAAGGCATCCTCGTGT	$(TGCT)_3$	169	56	—
BM312	GQ149611	AATGCAGTGGTGGCTTCC	GGAAAAGAAAACCCCTTTGA	$(GGT)_3$	150	54	—
BM313	GQ149612	CCTGATCCCCCTCAAGAAACA	TGCTTACGCGTGGACTAACA	$(CA)_4(CA)_3$	128	55	—

[a]PIC values were calculated as discussed in the text for 18 genotypes of *Phaseolus vulgaris* as listed in survey I from Blair et al. (2006) for the first 80 SSR markers, of which 2 were found to produce multi-copy banding (mc) and 2 had trouble amplifying all genotypes (na).

Fig. 1. Frequency of polymorphism information content (PIC) values for dinucleotide (Di-nt.) and trinucleotide (Tri-nt.) based markers.

Although size selection in the range of 0.4 to 1.2 kb was used after digestion, the use of double digestion created more small fragments, which were less useful in microsatellite development. Meanwhile, *Taq*I digestion and sonication produced fragments that were too difficult to clone, resulting in libraries with higher numbers of missing inserts or low transformation efficiency. Furthermore, unlike the other restriction digestions, which were carried out at 37 °C, the *Taq*I digestion was carried out at the high temperature of 65 °C, and therefore this enzyme was no longer active when the ligation step occurred, which could explain the low efficiency of cloning.

After comparing library quality, each of the selected libraries was plated and picked into 384-well plates which were used to create high-density filter arrays. In this step, we adapted a high-throughput protocol for the screening of plasmid-based libraries based on robotic colony picking and filter preparation, which allowed us to produce large libraries of 18 342 clones each to screen for microsatellites. Given that we were working with non-enriched small-insert libraries, it was important to screen large numbers of random genomic clones to obtain an accurate picture of microsatellite frequency and to increase the chances of recovering SSR-containing fragments. Our analysis of the non-enriched libraries through sequencing showed that there was no clone redundancy, which is a significant advantage over enriched libraries, where redundancy is problematic (Squirrell et al. 2003). In common bean, enriched libraries for GA-motif microsatellites have also had a large number of redundant clones (Gaitán et al. 2002), while in our current non-enriched microsatellite libraries, all positive clones identified new microsatellite loci.

Another advantage of the non-enriched libraries was the ability to screen for various motifs through probe hybridization, which would have been less possible with enrichment methods, which by nature are restricted to the motifs used for SSR capture (Gaitán et al. 2002). In small-insert libraries such as the ones represented on our high-density filters, no such limitation is present because the filter sets can be

probed multiple times with different motifs or combinations of motifs. Meanwhile, enrichment procedures that use PCR steps in their protocols can cause artifacts such as overrepresentation of certain clones that are difficult to detect until sequencing or marker testing steps (Zane et al. 2002). These artifacts do not occur in non-enriched libraries. In addition, enriched libraries often have the disadvantage of diminishing returns to sequencing investment due to redundancy, whereas genomic libraries such as the ones we developed can easily be scaled up even further to represent up to 1 or more genome equivalents while still fitting into a few high-density filters. To obtain higher numbers of microsatellites, additional libraries with larger numbers of clones could be made, as the robotic equipment for clone picking, filter preparation, and even re-arraying is widely available and inexpensive to operate. Finally, the libraries we produced are equivalent to small-insert versions of genomic shotgun libraries which could potentially be screened for further microsatellites in the future by next generation sequencing (NGS) technologies such as 454 Titanium sequencing. Given that the cost of these new sequencing methods is coming down, this may be a very useful method for identifying SSR-containing sequences at relatively low cost. Furthermore, various genotypes could be analyzed with NGS technologies to discover allelic differences at given microsatellite loci.

The disadvantages to our small-insert library screening methods were the time and radioactive label needed for the filter hybridization step, the need to align positive clones, and the low frequency of positive clones overall. Some of these problems could be solved by use of a phosphoimager for positive clone detection and automated ID calls for positive hits, as is done with overgos on large-insert clone libraries (International Human Genome Mapping Consortium 2001). Computerized calling of positive hits to immediately identify the clone from the library address would be especially recommended for screening filter sets with even larger numbers of clones. In any case, the double duplicate pattern we used for filter set printing allowed us to positively identify clones and avoid false positives more easily than with traditional unreplicated filter patterns or with the older time-consuming method of small-insert phage library screening with plating, plaque lifts, filter processing, and other steps (Panaud et al. 1995; Zane et al. 2002). The low frequency of positive clones with non-enriched libraries is more difficult to address but can provide an estimate of the efficiency of SSR discovery through the NGS techniques discussed above. Next generation sequencing would be able to generate massive amounts of sequence data which could then be screened bioinformatically for SSR motifs such as the ones identified here. In summary, screening of non-enriched, small-insert libraries as shown here is a useful alternative to some of the traditional screening or enrichment methods of SSR isolation and was informative as a prelude to new sequencing techniques which could enhance the efficiency of SSR discovery.

Our screening technique gave us estimates of the frequencies of di- and trinucleotide motifs across the 3 libraries, and in general terms the dinucleotide motifs we screened were more frequent than the trinucleotide motifs we screened, although this may have been a result of the motifs used in the hybridization probes. Microsatellite frequency overall appears to be less than 1% of clones, with 1 microsatellite every 60 to 70 kb of sequence. This frequency is in agreement with the evaluation of random common bean genomic sequences found in Schlueter et al. (2008), where simple sequence repeats were found in 0.32 of BAC end sequences and represented 0.64% of total repetitive sequences. Given the average distance between microsatellites of approximately 70 kb, it should be possible to identify up to 9 000 SSRs in the bean genome.

Similar microsatellite frequencies are found in other plant species, especially small-genome species like common bean, according to various authors (Cardle et al. 2000; Tóth et al. 2000; Zane et al. 2002). In the future it will be interesting to determine the relative frequencies of different SSR motifs in different parts of the bean genome, given that microsatellites are reported to vary in frequency at telomeres or centromeres, within certain parts of genes, and near retrotransposons (Areshchenkova and Ganal 1999; Hancock 1999; Ramsay et al. 1999; Fujimori et al. 2003). In terms of the frequency of specific motifs, our sequence analysis found that among dinucleotide markers, CA-motif microsatellites were more common than GA-motif microsatellites and that among the trinucleotide motifs the most common were CAC, CTG, GCA, and ATA. We also found some non-targeted motifs such as AT and GC in compound microsatellites along with the targeted dinucleotide and trinucleotide motifs. These results agree with the analysis of an enriched library from Hanai et al. (2007), where of 40 genomic microsatellites developed, 70% had CA-related motifs and only 30% had GA-related motifs, and compound repeats with non-target sequences were common. Blair et al. (2008) showed that ATA microsatellites are easily obtained from common bean genomic sequences. These results contrast with data from gene-based SSR analysis of EST sequences by Hanai et al. (2007), where GA, GAA, CAT, and TGG motifs were the most common. From computational studies in arabidopsis and rice, it appears that both GA and CA dinucleotide motifs are common and that the most frequent trinucleotide motif is variable (Panaud et al. 1995; Cardle et al. 2000; Temnykh et al. 2001).

It is likely that differences between various researchers' results are due to the source sequences used (whether from coding or non-coding regions of the genome), the method of generating the SSR library or identifying the microsatellites, or the specific species studied. To evaluate various parts of the genome, we selected enzymes with different recognition sites for the construction of the small-insert libraries, with *Hae*III (GG^CC) expected to target more GC-rich fragments of the genome perhaps associated with the gene-coding fraction, and *Alu*I (AG^CT) and *Rsa*I (GT^AC) targeting more AT-rich fragments of the genome associated with the non-coding fraction. Future work could use methylation-sensitive enzymes for library construction to isolate sequences even more likely to be associated with the transcriptome.

Other comments on our work are first that the markers developed in this study are complementary to previously developed microsatellites and for the most part may represent independent loci that might be less likely to be genetically linked in clusters than other SSR markers from enriched

libraries, which was a problem identified by Blair et al. (2003) for GA-motif microsatellites from the enriched libraries of Gaitán et al. (2002). Secondly, the markers based on GA or various trinucleotide repeats were moderately to highly polymorphic, but CA-based markers were not. The higher polymorphism of GA- versus CA-based microsatellites agrees with results from Blair et al. (2006), Benchimol et al. (2007), and Hanai et al. (2007). Sequence comparisons showed that 29 of the sequences from this study were related to common bean SSR loci from previous studies (Buso et al. 2006; Benchimol et al. 2007; Hanai et al. 2007) although they contained different motifs. This shows that our hybridization technique probably targeted some of the same repeats as these studies and that microsatellites in common bean may sometimes be embedded in repetitive elements and have similar flanking sequences associated with transposon families, as has been found in other crops (Ramsay et al. 1999).

Finally, the new markers also complement other common bean SSR markers in terms of the source genotype they represent. For example, our use of DOR364, a red-seeded Mesoamerican bean from Central America, to create the non-enriched SSRs complements previous microsatellite development with enriched libraries based on the Nueva Granada race genotype G4494, a dry bean variety in Colombia named Diacol-Calima, used by Gaitán et al. (2002); Fin de Bagnols, a snap bean variety from France used by Métais et al. (2002); or the carioca and black-seeded Mesoamerican genotypes Perola and IAC-UNA used by Buso et al. (2006) and Benchimol et al. (2007) and Hanai et al. (2007), respectively. We are currently also creating microsatellites for sequences from G19833, a Peru race genotype, and expect them to be novel given that race Peru is very distinct from other races in terms of microsatellite alleles and other character states (Blair et al. 2006). In the future, these various types of microsatellites should provide more complete coverage of the genetic map of common bean than in previous SSR mapping studies (Blair et al. 2003; Grisi et al. 2007). In this regard, it would be useful to have a genetic map for common bean consisting entirely of microsatellites, since these second-generation markers would be easy to assay and would enable a large number of segregating individuals to be analyzed in gene and QTL tagging studies. For this reason the markers developed here from non-enriched libraries will be complementary to other markers from BAC, cDNA, or SSR-enriched libraries.

Acknowledgments

We thank staff of the Clemson University Genome Institute for technical assistance and J. Tomkins and D. Frisch for helpful advice. This work was supported by the Generation Challenge Programme, USAID, and CIAT funds.

References

Afanador, L.K., Haley, S.D., and Kelly, J.D. 1993. Adoption of a "mini-prep" DNA extraction method for RAPD marker analysis in common bean (*Phaseolus vulgaris* L.). Annu. Rep. Bean Improv. Coop. **36**: 10–11.

Areshchenkova, T., and Ganal, M.W. 1999. Long tomato microsatellites are predominantly associated with centromeric regions. Genome, **42**(3): 536–544. doi:10.1139/gen-42-3-536. PMID: 10382301.

Benchimol, L.L., de Campos, T., Carbonell, S.A.M., Colombo, C.A., Chioratto, A.F., Formighieri, E.F., et al. 2007. Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. Genet. Resour. Crop Evol. **54**(8): 1747–1762. doi:10.1007/s10722-006-9184-3.

Blair, M.W., Pedraza, F., Buendia, H.F., Gaitán-Solís, E., Beebe, S.E., Gepts, P., and Tohme, J. 2003. Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. **107**(8): 1362–1374. doi:10.1007/s00122-003-1398-6. PMID:14504741.

Blair, M.W., Giraldo, M.C., Buendía, H.F., Tovar, E., Duque, M.C., and Beebe, S.E. 2006. Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. **113**(1): 100–109. doi:10.1007/s00122-006-0276-4. PMID: 16614831.

Blair, M.W., Fregene, M.A., Beebe, S.E., and Ceballos, H. 2007. Marker-assisted selection in common beans and cassava. *In* Marker-assisted selection. Current status and future perspectives in crops, livestock, forestry and fish. *Edited by* E. Guimarães, J. Ruane, B. Scherf, A. Sonnino, and J. Dargie. FAO, Rome. pp. 81–116.

Blair, M.W., Buendía, H.F., Giraldo, M.C., Métais, I., and Peltier, D. 2008. Characterization of AT-rich microsatellites in common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. **118**(1): 91–103. doi:10.1007/s00122-008-0879-z. PMID:18784914.

Broughton, W.J., Hernandez, G., Blair, M.W., Beebe, S.E., Gepts, P., and Vanderleyden, J. 2003. Beans (*Phaseolus* spp.) — model food legumes. Plant Soil, **252**(1): 55–128. doi:10.1023/A:1024146710611.

Buso, G.S.C., Amaral, Z.P.S., Brondani, R.P.V., and Ferreira, M.E. 2006. Microsatellite markers for the common bean *Phaseolus vulgaris*. Mol. Ecol. Notes, **6**(1): 252–254. doi:10.1111/j.1471-8286.2006.01210.x.

Caixeta, E.T., Borém, A., and Kelly, J.D. 2005. Development of microsatellite markers based on BAC common bean clones. Crop Breed. Appl. Biotechnol. **5**: 125–133.

Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D., and Waugh, R. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics, **156**(2): 847–854. PMID:11014830.

de Campos, T., Benchimol, L.L., Carbonell, S.A.M., Chioratto, A.F., Formighieri, E.F., and de Souza, A.P. 2007. Microsatellites for genetic studies and breeding programs in common bean. Pesqui. Agropecu. Bras. **42**: 589–592.

Dutech, C., Enjalbert, J., Fournier, E., Delmotte, F., Barrès, B., Carlier, J., et al. 2007. Challenges of microsatellite isolation in fungi. Fungal Genet. Biol. **44**(10): 933–949. doi:10.1016/j.fgb.2007.05.003. PMID:17659989.

Edwards, K.J., Baker, J.H.A., Daly, A., Jones, C., and Karp, A. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques, **20**: 759–760.

Fujimori, S., Washio, T., Higo, K., Ohtomo, Y., Murakami, K., Matsubara, K., et al. 2003. A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. FEBS Lett. **554**(1–2): 17–22. doi:10.1016/S0014-5793(03)01041-X. PMID:14596907.

Gaitán, E., Duque, M.C., Edwards, K.J., and Tohme, J. 2002. Microsatellite repeats in common bean (*Phaseolus vulgaris*): isolation, characterization, and cross-species amplification in *Phaseolus* spp. Crop Sci. **42**: 2128–2136.

Grisi, M.C.M., Blair, M.W., Gepts, P., Brondani, C., Pereira, P.A.A., and Brondani, R.P.V. 2007. Genetic mapping of a new set of microsatellite markers in a reference common bean (*Pha-*

seolus vulgaris) population BAT93 × Jalo EEP558. Genet. Mol. Res. **6**(3): 691–706. PMID:18050090.

Guerra-Sanz, J.M. 2004. New SSR markers of *Phaseolus vulgaris* from sequence databases. Plant Breed. **123**(1): 87–89. doi:10.1046/j.0179-9541.2003.00929.x.

Gupta, P.K., and Varshney, R.K. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica, **113**(3): 163–185. doi:10.1023/A:1003910819967.

Hanai, L.R., de Campos, T., Camargo, L.E.A., Benchimol, L.L., de Souza, A.P., Melotto, M., et al. 2007. Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. Genome, **50**(3): 266–277. doi:10.1139/G07-007. PMID:17502900.

Hancock, J.M. 1999. Microsatellites and other simple sequences: genomic context and mutation mechanisms. *In* Microsatellites: evolution and applications. *Edited by* D.B. Goldstein and C. Schlötterer. Oxford University Press, New York. pp. 1–9.

International Human Genome Mapping Consortium. 2001. A physical map of the human genome. Nature, **409**(6822): 934–941. doi:10.1038/35057157.

Métais, I., Hamon, B., Jalouzet, R.,, and Peltier, D. 2002. Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. Theor. Appl. Genet. **104**: 1346–1352. doi:10.1007/s00122-002-0901-9. PMID:12582590.

Morgante, M., and Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J. **3**(1): 175–182. doi:10.1111/j.1365-313X.1993.tb00020.x. PMID:8401603.

Panaud, O., Chen, X., and McCouch, S.R. 1995. Frequency of microsatellite sequences in rice (*Oryza sativa* L.). Genome, **38**(6): 1170–1176. PMID:8654913.

Powell, W., Machray, G.C., and Provan, J. 1996. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. **1**: 215–222.

Ramsay, L., Macaulay, M., Cardle, L., Morgante, M., degli Ivanissevich, S., Maestri, E., et al. 1999. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J. **17**(4): 415–425. doi:10.1046/j.1365-313X.1999.00392.x. PMID:10205898.

Rozen, S., and Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist programmers. *In* Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, N.J. pp. 365–386.

Sambrook, J.M., and Russell, D.W. 2001. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. pp. 1.44–1.47.

Schlueter, J.A., Goicoechea, J.L., Collura, K., Gill, N., Lin, J.-Y., Yu, Y., et al. . BAC-end sequence analysis and a draft physical map of the common bean (*Phaseolus vulgaris* L.) genome. Trop. Plant Biol. **1**(1): 40–48. doi:10.1007/s12042-007-9003-9.

Squirrell, J., Hollingsworth, P.M., Woodhead, M., Russell, J., Lowe, A.J., Gibby, M., and Powell, W. 2003. How much effort is required to isolate nuclear microsatellites from plants? Mol. Ecol. **12**(6): 1339–1348. doi:10.1046/j.1365-294X.2003.01825.x. PMID:12755865.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S.R. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. **11**(8): 1441–1452. doi:10.1101/gr.184001. PMID:11483586.

Tóth, G., Gáspári, Z., and Jurka, J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. **10**(7): 967–981. doi:10.1101/gr.10.7.967. PMID:10899146.

Varshney, R.K., Graner, A., and Sorrells, M.E. 2005. Genic microsatellite markers in plants: features and applications. Trends Biotechnol. **23**(1): 48–55. doi:10.1016/j.tibtech.2004.11.005. PMID:15629858.

Yu, K., Park, S.J., and Poysa, V. 1999. Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*). Genome, **42**(1): 27–34. doi:10.1139/gen-42-1-27.

Yu, K., Park, S.J., Poysa, V., and Gepts, P. 2000. Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). J. Hered. **91**(6): 429–434. doi:10.1093/jhered/91.6.429. PMID:11218079.

Zane, L., Bargelloni, L., and Patarnello, T. 2002. Strategies for microsatellite isolation: a review. Mol. Ecol. **11**(1): 1–16. doi:10.1046/j.0962-1083.2001.01418.x. PMID:11903900.