

Identification and Mapping of Simple Sequence Repeat Markers from Common Bean (*Phaseolus vulgaris* L.) Bacterial Artificial Chromosome End Sequences for Genome Characterization and Genetic–Physical Map Integration

Juana M. Córdoba, Carolina Chavarro, Fernando Rojas, Claritza Muñoz, and Matthew W. Blair*

Abstract

Microsatellite markers or simple sequence repeat (SSR) loci are useful for diversity characterization and genetic–physical mapping. Different *in silico* microsatellite search methods have been developed for mining bacterial artificial chromosome (BAC) end sequences for SSRs. The overall goal of this study was genome characterization based on SSRs in 89,017 BAC end sequences (BESs) from the G19833 common bean (*Phaseolus vulgaris* L.) library. Another objective was to identify new SSR taking into account three tandem motif identification programs (Automated Microsatellite Marker Development [AMMD], Tandem Repeats Finder [TRF], and SSRLocator [SSRL]). Among the microsatellite search engines, SSRL identified the highest number of SSRs; however, when primer design was attempted, the number dropped due to poor primer design regions. Automated Microsatellite Marker Development software identified many SSRs with valuable AT/TA or AG/TC motifs, while TRF found fewer SSRs and produced no primers. A subgroup of 323 AT-rich, di-, and trinucleotide SSRs were selected from the AMMD results and used in a parental survey with DOR364 and G19833, of which 75 could be mapped in the corresponding population; these represented 4052 BAC clones. Together with 92 previously mapped BES- and 114 non-BES-derived markers, a total of 280 SSRs were included in the polymerase chain reaction (PCR)-based map, integrating a total of 8232 BAC clones in 162 contigs from the physical map.

COMMON BEAN (*Phaseolus vulgaris* L.) is the most popular grain legume for the human diet and represents half of the food pulses consumed worldwide (see Broughton et al., 2003). Characterization of the common bean genome is necessary for a better understanding of the genes and molecular processes involved in agronomic adaptation, nutritional traits, disease or insect resistance, drought tolerance, and even yield of this important crop (Blair and Ishitani, 2009). There are several methods to study the genome; however, one of the most commonly used and the cheapest is the *in silico* design of molecular markers. The most targeted loci are either simple sequence repeats (also known as microsatellites) or single nucleotide polymorphisms (SNPs), due to the broad applicability of both marker types in genetic studies (Hamblin et al., 2007).

Simple sequence repeats (SSRs) are tandem repeats made up of repeated nucleotide (nt) motifs up to 6 bp long. As loci, microsatellites are widely distributed throughout the genome and highly variable (Schlötterer, 2000). Furthermore they can be easily characterized in almost any molecular lab with inexpensive equipment, and primer pairs that can be electronically distributed and readily ordered from many sources. Meanwhile, SNP markers are usually biallelic and require expensive equipment and somewhat more work to confirm as useful markers, thus SSRs remain important

J.M. Córdoba, C. Chavarro, C. Muñoz, and M.W. Blair, International Center for Tropical Agriculture (CIAT) Bean Project, A.A. 6713, Cali, Colombia; F. Rojas, International Center for Improvement of Maize and Wheat (CIMMYT), C.P. 56130, El Batán, Texcoco, México. Received 21 June 2010. *Corresponding author (m.blair@cgiar.org).

Published in The Plant Genome 3:154–165. Published 18 Dec. 2010.
doi: 10.3835/plantgenome2010.06.0013

© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

Abbreviations: AMMD, Automated Microsatellite Marker Development; BAC, bacterial artificial chromosome; BES, bacterial artificial chromosome end sequence; nt, nucleotide; BMa, bean microsatellite from AT-rich sequences; BMb, bean microsatellite derived from bacterial artificial chromosome end sequences; GC, guanine and cytosine; PCR, polymerase chain reaction; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; SSRL, SSRLocator; T_m , melting temperature; TRF, Tandem Repeats Finder; TROLL, Tandem Repeat Occurrence Locator.

for various applications (Hamblin et al., 2007). Most SSRs are neutral markers, although some can be related to promoters and other regulatory regions of DNA, and these can represent functional polymorphisms that, with gene-based SNPs, can be useful for association mapping (Gore et al., 2009). Simple sequence repeats have been used for high throughput fingerprinting and fine mapping in species such as rice (*Oryza sativa* L.) (Coburn et al., 2002), rye (*Secale cereale* L.) (Kofler et al., 2008), or soybean [*Glycine max* (L.) Merr.] (Kim et al., 2010) and recently, Blair et al. (2008, 2009a, b) have worked on developing a large number of new microsatellites for common bean.

In silico identification of SSRs can involve various different computer programs such as Automated Microsatellite Marker Development (AMMD) (Martins et al., 2006), Batchprimer3 (You et al., 2008), MISA (Garnica et al., 2006; Portis et al., 2007), Tandem Repeats Finder (TRF) (Benson, 1999), SSRSEARCH (Nicot et al., 2004), and SSRLocator (SSRL) (da Maia et al., 2008). This large number of computer programs for SSR analysis have been created for in silico identification of SSRs and their number reflects the challenges of identifying specific motifs in genomic sequences and the need for analyzing extensive databases from expressed sequence tag (EST) projects or high-throughput sequencing. Given that the distribution, length, and composition of SSR motifs are highly variable, SSRs are sometimes difficult to find with a single search engine (Benson, 1999). Therefore, no unique optimal solution to in silico microsatellite identification exists and each algorithm has its own strengths to find SSRs (Le et al., 2010).

Bacterial artificial chromosome (BAC) clones have been the most widely used vector for constructing physical maps and are often used for full genome sequencing projects using minimum tiling paths (Kelley et al., 1999). Bacterial artificial chromosomes clones usually range from 100 to 200 Kbp in insert length and represent most of the genome when the library is properly constructed. A byproduct of BAC libraries has been BAC end sequences (BESs), produced by sequencing each clone in a BAC library from each end of the genomic insert. Bacterial artificial chromosome end sequence projects generate large datasets that are proportional with the size and genome coverage of the BAC library and have often been used to identify molecular markers for genetic or physical mapping (Mun et al., 2006; Shoemaker et al., 2008; Wondji et al., 2005). When dense enough, these sequences can also be used as means for selecting minimally overlapping BAC clones for sequencing large DNA regions (Kelley et al., 1999), as a primary scaffold for whole-genome sequencing, or as a way to verify the accurateness of physical map assembly (Meyers et al., 2004). Additionally, BES projects often provide insights into genome structure (Hong et al., 2006) and relative abundance of single copy sequences, retrotransposons (RTs), transposable elements (TEs), and other repeat types.

In common bean, BAC end sequencing and fingerprinting of over 40,000 BAC clones has been used to create an initial physical map and to evaluate the frequency of

various repetitive sequences such as centromeric satellite repeats and housekeeping genes as well as genes from the families involved in transcription and other functions from the BAC library for genotype G19833 (Schlueter et al., 2008). At present, our lab has created and tested an initial set of 230 BES-derived SSRs from this library using BatchPrimer3 software (Córdoba et al., 2010) but more are needed since coverage of the initial set was limited and overall the number of mapped microsatellites is low (Blair et al., 2003, 2008; Grisi et al., 2007). Therefore, the objective of this study was to identify new BES-SSRs useful for producing a more saturated integrated physical-genetic map using additional SSR search engines to analyze the common bean BES database. We were especially interested in AT-rich microsatellites since these were located in previously underrepresented regions of the genome (Blair et al., 2008). Therefore, the software producing more AT-rich, di-, and trinucleotide motif BES-SSRs was used to create 323 new primer pairs and the polymorphic markers were mapped onto the genetic map for the DOR364 × G19833 cross. The importance of the integrated physical and genetic map in its ability to cross-link molecular markers and a set of contigged large-insert clones with a high degree of accuracy is highlighted.

MATERIALS AND METHODS

Identification of Simple Sequence Repeats in the Bacterial Artificial Chromosome-Ends

A total of 89,017 BES produced as part of the physical mapping project described in Schlueter et al. (2008) and originally from a BAC library of the Andean common bean genotype G19833 constructed by CIAT and the Clemson University Genomic Institute (CUGI) were searched for SSR repeats with three software programs: AMMD from Martins et al. (2006), SSRL from da Maia et al. (2008), and TRF from Benson (1999). Automated Microsatellite Marker Development is a package consisting of TROLL (Tandem Repeat Occurrence Locator) used for microsatellite detection and the rest of the Staden Package used for sequence assembly and analysis. Pre-GAP4 and GAP4 are used by this software and the program has a filter option in which it is possible to select against vector contamination and poor sequence quality.

The search parameters for AMMD and SSRL corresponded to motif lengths of di-, tri-, tetra-, penta-, and hexanucleotide repeats with minimum repeat numbers of 5, 4, 3, 3, and 3, respectively. Primers were designed around the SSR motif using Primer 3 (Rozen and Skaltsky, 2000) such that the polymerase chain reaction (PCR) product sizes would be between 100 and 300 bp. The other parameters were: primer size from 18 to 22 bp, optimal primer melting temperature (T_m) of 50°C, and guanine and cytosine (GC) content between 45 and 55%. The TRF algorithm was run with a minimum repeat length stretch of 12 nt for the same five types of SSR motifs (di- to hexanucleotide repeats) and additional search parameters of match, mismatch, indel, and minimum alignment scores of 2, 7, 7, and 50, respectively.

Simple Sequence Repeat Amplification and Detection

For selected primer pairs from the AMMD software analysis, PCR reactions were performed in a final volume reaction of 15 μ L containing 20 ng of total genomic DNA, 0.15 μ M each of the forward and reverse primers, 2.0 mM of $MgCl_2$, 200 μ M of total dNTP, and 1 unit of Taq polymerase. The PCR program involved a hot start of 93°C for 3 min, followed by denaturation for 30 s at 92°C, annealing for 30 s at the T_m (+4°C) of the lower annealing primer temperature, extension for 45 s at 72°C, followed by a touchdown thermocycling profile with a 1°C drop per cycle in extension temperatures for 8 cycles, followed by 27 cycles of denaturation for 30 s at 92°C, annealing for 30 s at the T_m (-4°C), and extension for 45 s at 72°C to ensure strong PCR products. Post-amplification, there was a 5 min extension period at the same 72°C temperature. Reactions were performed on PTC-200 (MJ Research Inc, Watertown, MA) thermocyclers.

After the PCR reaction, 5 μ L of formamide containing 0.4% w/v bromophenol blue and 0.25% w/v xylene cyanol was added to each PCR reaction and the mixture was denatured at 96°C for 6 min. Subsequently, the mixtures were loaded with a eight-syringe, multipipette into alternate position of a 100-well, shark-tooth comb set into a 4% denaturing polyacrylamide (29:1 acrylamide:bis-acrylamide) gel that contained 5 M urea and 0.5x TBE buffer (44.57 mM trizma base, 44.46 mM boric acid, 1mM EDTA, and pH 8.0). The gels were run in Owl Sequencing Units (Thermo Fisher Scientific Inc, Waltham MA) at a constant 50°C and 100 W for approximately 1 h. Detection of PCR amplification products was via silver staining according to Blair et al. (2003) and the allele sizes were estimated based on 10, 25, and 50 bp molecular weight (MW) ladders.

Genetic Mapping and Integration with Physical Map

The first step in genetic mapping was a polymorphism survey using DOR364 and G19833, in which DOR364 is a Mesoamerican advanced breeding line from CIAT and G19833 is an Andean germplasm accession from Peru that was used to construct the BAC library from which BES information was developed. After the parental genotypes were scored for their alleles, any polymorphic microsatellites were then mapped using the full set of 89 $F_{9:11}$ recombinant inbred lines from the cross DOR364 \times G19833 as described by Blair et al. (2003). Segregation data and the software program MapDisto v.1.7 with a LOD > 3.0 were used to place the new markers in the genetic map for the population from Blair et al. (2008). Genetic distances were calculated from the recombination fraction based on the Kosambi function.

To integrate the physical and genetic maps, the mapped BES-SSR marker loci were compared with the positions of the corresponding BACs contained in the fingerprint-assembled contigs at the Phaseolus WebFPC database (<http://phaseolus.genomics.purdue.edu> [verified 11 Nov. 2010]). The integrated map was drawn to scale using an

in-house, MS Excel 2007 macro with relative marker distances in cM and contig sizes in Kbp. The physical map was therefore shown as discontinuous line but did indicate the size and orientation of each contig relative to the markers on the genetic map. Finally the integrated map was compared for each linkage group or chromosome with the most recent cytogenetic map obtained for common bean by Fonsêca et al. (2010) to see if there was a correlation between our genetic distances and their physical distances.

RESULTS

Software Comparisons in the Identification of Common Bean BES-Simple Sequence Repeats

A large number of specific software programs have been created to search for SSRs in various types of DNA sequences and in this study of common bean BES-SSRs we compared a total of three new programs (AMMD, SSRL, and TRF) to one previously used program (BatchPrimer3). Each one produced differing results based on their unique algorithms and various features and in this study we found the software programs AMMD and SSRL to be the most complete in these characteristics and the best for displaying information required about SSRs and microsatellite flanking primers as well as in specifying the motif length for each type of repeat. In contrast, TRF did not produce information regarding primer design and the SSR search engine was not very efficient and therefore was not used further in this study.

In the case of AMMD software that was finally chosen over SSRL, we used the same parameters for motif length as in our previous work with BatchPrimer3 as described in Córdoba et al. (2010) so as to facilitate comparisons with that software. The main advantage we found was the higher number of microsatellites found by AMMD (4727 loci) in comparison to the number of SSRs identified with BatchPrimer3 by Córdoba et al. (2010). Given the low number of loci identified by BatchPrimer3, further marker development was a priority in the current research and we found AMMD to be the ideal software in the identification of better quality SSRs in comparison to the other two programs tested (SSRL and TRF).

Differences in Simple Sequence Repeat Motifs Found by the Search Engines and Initial Genome Characterization

The types and number of SSRs identified in the 89,017 BAC end sequences from the physical mapping project differed among the microsatellite search engines (Table 1). With AMMD software, a moderate number of SSRs were found, totaling 4727, of which 3095 were found in BES associated with BAC contigs, 639 were from BAC singleton BES included in the physical map through overgo hybridization, and 993 were found in BAC singleton BES not included in the physical map. The SSRs located in contigs covered 794 BAC groupings with an average of four SSRs per BAC contig. One advantage of AMMD was that it identified perfect and compound

Table 1. Comparison of the results of different microsatellite (simple sequence repeat [SSR]) search software programs showing motif type and class for the perfect microsatellites and number of compound SSRs found with Automated Microsatellite Marker Development (AMMD), SSRLocator (SSRL), and Tandem Repeats Finder (TRF).

Motif	AMMD			SSRL [§]			TRF	
	Perfect [†]		C [‡]	Perfect		C	Perfect	
	Class I [‡]	Class II [‡]		Class I	Class II		Class I	Class II
Di-nt [#]	105 (91.3%)	1270 (29.7%)	—	771 (93.3%)	1674 (13.8%)	—	805 (71.8%)	—
Tri-nt	10 (8.7%)	1212 (28.3%)	—	55 (6.7%)	4012 (33.2%)	—	236 (21.1%)	42 (9.0%)
Tetra-nt	—	1219 (28.5%)	—	—	4337 (35.9%)	—	9 (0.8%)	23 (4.9%)
Penta-nt	—	262 (6.1%)	—	—	1009 (8.3%)	—	15 (1.3%)	31 (6.6%)
Hexa-nt	—	313 (7.3%)	—	—	1055 (8.7%)	—	56 (5.0%)	372 (79.5%)
Subtotals	115	4,276	336	826	12,086	1,844	1,121	468
Total		4,727			14,756			1,589

[†]Percentages indicate amount of microsatellite loci in that category of motif and class for perfect microsatellites compared to the total number of SSRs identified with the software program in a given class.

[‡]Class I microsatellite loci defined as those SSRs with 10 or more repeats and class II microsatellite loci are defined as those SSRs with less than 10 repeats according to Shultz et al. (2007).

[§]Total number of identified SSRs, including SSRs with and without primer design.

[‡]C = compound SSRs were not detected by TRF.

[#]nt, nucleotide.

microsatellites, meaning those loci that had single or multiple motifs, respectively. However, imperfect microsatellites with a single motif interrupted by one or more nucleotides were not discovered unless the first or second part of the SSR was long enough to be detected.

With this in mind, it was notable that even with the limitation of searching for mostly perfect repeats, dinucleotide-, trinucleotide-, and tetranucleotide-based SSRs were almost equally common (29.1, 26.6, and 25.5%, respectively). Meanwhile, the compound, hexanucleotide-, and pentanucleotide-based loci represented together less than 20% of the total SSRs identified by AMMD. The specific motifs AG/TC (37.5%) and AT/TA (33.7%) were the most frequent among the total of dinucleotide repeats; while among the trinucleotide repeats, AGA/TCT (44.6%) and ATA/TAT motifs (18.4%) were relatively common. The microsatellites found by AMMD were also classified according to their length. Here, we found that only 2.6% of SSRs were longer than 10 repeats and belonged to class I type microsatellites, 7.4% were compound, and the rest of the microsatellites had less than 10 repeats and were classified as class II type SSRs according to the classification system of Shultz et al. (2007). The motifs found among class I microsatellites were mainly AT/TA (75%) and AG/TC (6%). While among the class II microsatellites, dinucleotide, trinucleotide, and tetranucleotide repeat motifs were about equally abundant.

In the second software analyzed (SSRL), we were able to identify a very large number of SSRs. In total, 14,756 SSRs were found in the 89,017 BES; however, it was only possible to design primers for 1842 SSRs. The other 12,914 SSRs that could not be used to develop primers pairs had their repeats in inadequate positions within the BES fragments or had low T_m or poor GC content in part of their sequence. Of the 1842 SSRs with primer pairs designed for them, the majority consisting of 1317 SSRs were located in BES from BAC contigs while 525 SSRs were found associated with BAC singletons. The SSRL software was useful in identifying perfect, compound,

and imperfect microsatellites; however, the predominant microsatellites among these were perfect, dinucleotide-based SSRs (44.8% of the total of both class I and class II loci) followed by compound SSRs (16.1%) and trinucleotide based SSRs (15.8%). Simple sequence repeats based on tetranucleotide, pentanucleotide, and hexanucleotide motifs corresponded to a total of 23.3% of loci with primer pairs. The most frequent motifs over the 1842 dinucleotide and trinucleotide repeat microsatellites were those rich in AT-based sequences such as AT/TA based SSRs (24.9%) and ATA/TAT based SSRs (6.2%).

When SSRL-derived microsatellites were classified according to their length, class I type SSRs were predominant (44.8%) over class II type SSRs (39.1%) with the remainder compound loci (16.1%). The remaining 12,914 SSRs identified by SSRL, but for which primers were not identified, provided information about the common bean genome. For example, the analysis of all SSRs identified placed tetra- and trinucleotides and the motifs AT/TA, ATA/TAT, and CTTGGG/GAACCC in the top positions within their motif groups as the most common SSR types. Furthermore, class II SSRs were more frequent than class I type SSRs among those SSRs without flanking primers. Interestingly, the longest motifs were of the AT/TA, AG/TC, ATA/TAT, and AGA/TCT repeat types.

Tandem Repeats Finder, the third software program used, identified the lowest number of SSRs in comparison to the previously mentioned programs. Among the 89,017 BES fragments searched, we found a total of only 1589 SSRs. For comparison sake, the dinucleotide motif SSRs (805 loci out of 1589 in total) as well as the hexanucleotide motif SSRs (428 loci) were the predominant repeat types, along with AT/TA (715 loci), CTTGGG/GAACCC (196 loci), and ATA/TAT (191 loci) repeats. Tandem Repeats Finder identified only perfect SSR motifs and most of them were classified as class I (70.5%) compared to a smaller number of class II (29.4%) with 12 nt of repetitive bases as a minimum threshold for

each SSR. As previously observed, the class I SSRs were mainly based on the motif AT/TA (88.8%), with very few long GA/CT, CA/GT, or GC/CG motif markers (11.2%).

BES-Simple Sequence Repeat Marker Polymorphism Screening

As described above, the evaluation of the three different SSR search software programs AMMD, SSRL, and TRF, produced very different results. The comparison of the performance and output of each of the programs showed that AMMD was the best computer program for SSR identification in terms of types of repeat motifs found and primer design. As a consequence, a subset of SSRs from the AMMD output (Supplementary Table 1) were selected for a parental survey of DOR364 and G19833. These microsatellites were named using a combination of the series name BMb (bean microsatellite derived from BAC end sequences) with a sequential number starting at 1000 to distinguish them from BES microsatellites developed with BatchPrimer3 in our previous study (Córdoba et al., 2010).

In total, we selected 323 primer pairs for the new BMb microsatellites based on dinucleotide AT/TA motifs and trinucleotide ATA/TAT motifs located in BES fragments from contigged BACs. In addition, our selection of SSR markers tried to anchor contigs not previously included in the integrated map made using SSRs identify with BatchPrimer3 (Córdoba et al., 2010). In terms of microsatellite class, 32 BES-SSRs belonged to class I and 291 to class II. The molecular characterization of the parental genotypes showed an amplification success rate of 70.0% (226 out of 323) and a polymorphism rate of 37.6% (85 out of 226). We found that a higher percentage of class I microsatellites (43.7%) showed amplification problems such as multiple banding or stuttering while, in contrast, class II microsatellites (24.7%) were less likely to have these problems. Any nonamplifying or poor-banding microsatellites were not used for analysis.

The comparison between the DOR364 versus G19833 (Mesoamerican versus Andean) polymorphism rate of dinucleotide and trinucleotide motif-based SSRs showed in both cases more than half of the microsatellites were monomorphic. In total among both classes of SSRs, 81 out of 146 successfully evaluated dinucleotide SSRs were monomorphic (55.4%) while 61 out of the 80 trinucleotide SSRs amplified were monomorphic (76.2%). As a result polymorphism was higher for the dinucleotide SSRs (44.6%) than for the trinucleotide SSRs (23.8%). A final interesting point was that in spite of the amplification problems encountered for class I SSRs, these SSRs were predominantly polymorphic (55.5%) in comparison with class II SSRs (35.6%). More specifically, a total of 10 out of the 18 markers developed for class I loci were polymorphic (55.5%) while out of 208 class II loci a total of 75 markers were polymorphic (36.0%).

Genetic Mapping of BES-Simple Sequence Repeats

The 85 polymorphic BMb markers identified in the parental screening corresponded to 81 contigs and were

amplified in DNA of the recombinant inbred line mapping population based on the cross DOR364 × G19833 as described in Blair et al. (2008). Map integration with a high LOD score (>3.0) was successful for a total of 75 new BMb markers placed on linkage groups as shown in Fig. 1 and described in Table 2. Four BMb markers could not be mapped since they were assigned to more than one linkage group (BMb1138, BMb1180, and BMb1273) or because their LOD scores were lower than 3.0 (BMb1098). In six cases, markers presented distances from neighboring markers longer than 20 cM (BMb1023, BMb1090, BMb1099, BMb1145, BMb1181, and BMb1285) and were not used in the final genetic map.

The new genetic map, which included 114 previously mapped non-BES genomic or gene-based SSR markers from Blair et al. (2003, 2008) and 91 BES-derived markers from Córdoba et al. (2010) plus the 75 new BMb markers, was found to cover 1575 cM and had a total of 280 SSRs with an average distance between neighboring loci of 5.7 cM and linkage group lengths ranging from 84.1 cM (b06g) to 212.2 cM (b02d) and averaging 143.2 cM. A greater number of new BMb markers were assigned to b01h (13) and b02d (10) with less than ten on the remaining linkage groups. The greatest combined number of new and previous BES-based markers were found on b08f (23) along with b01h and b02d (both 21) and the average across all linkage groups was 15.1 markers. Although relatively few BES-based markers were found on linkage group b09k and b05e (nine each) and to a certain extent on b06g and b11j (11 each), intermediate numbers were found on other linkage groups (b03c, b04b, b07a, and b10i). Despite these differences, a chi square test (p -value > 0.05) of the new and previous BES-based markers across each linkage group confirmed the uniform distribution overall of BACs from which these markers were derived. In contrast, within each linkage group the distribution of the new BMb was not always even and some of the markers tended to cluster together with the previously mapped AT-rich markers from the BMa (bean microsatellite from AT-rich sequences) series (Blair et al., 2008).

Integrated Genetic-Physical Map for Common Bean

The integration of the physical-genetic maps of common bean was done using the mapping information for the BES-SSRs and the fingerprinting information from Phaseolus WebFPC. This integrated map had two main components: (i) a genetic map represented by a gray continuous line in Fig. 1 with the SSR markers and their genetic distance given in cM and (ii) the physical map displayed to the right as blue and black lines indicating newly mapped contigs among those placed by Córdoba et al. (2010), with each of the contigs drawn in proportion to its Kb length. In addition, the orientation of the gray box at the top or bottom of the line was used to indicate whether the BES-SSR was anchored, respectively to the forward/5' or reverse/3' end of the corresponding BAC clone within the contig. Integration points were numbered sequentially from the top to the bottom of the linkage group including both new and previous map

linkages from both AMMD and previous Batchprimer3 results as described in Supplementary Table 2.

Layout of the linkage groups followed Blair et al. (2008) although certain linkage groups were reoriented based on new cytogenetic mapping of the short and long arms of the chromosomes by Fonsêca et al. (2010). The total of 166 mapped BMb markers linked 8232 BAC clones assembled into 162 contigs representing 78.2 Mb of the common bean genome with an average of 14.7 contigs and 748.4 BAC clones linked to each linkage group. In addition, the physical to genetic ratio Kbp:cM was calculated using the estimated chromosome length reported by Fonsêca et al. (2010) and the genetic distances of the present integrated genetic map. Significant differences were seen among the linkage groups and the values ranged from 269.6 Kbp:cM on b01h to 549 Kbp:cM on b06 g. Overall the average was 425.1 Kbp:cM, which is similar to previous estimates for the common bean genome. These results suggest no relation between physical and genetic length of each linkage groups as seen in the physically longest chromosome (b10i) that differed from the genetically longest linkage group (b02d). Interestingly, b06 g had an exact correspondence to the average even though this is the only chromosome that was acrocentric (Fonsêca et al., 2010).

The metacentric and submetacentric chromosomes varied in Kbp:cM ratio as compared to linkage group physical or genetic length with no significant correlations with the number of BAC contigs in either case ($r = 0.01$ and $r = 0.33$, respectively; $p > 0.05$). In contrast, there was a significant positive correlation of $r = 0.80$ ($p = 0.000$) between the number of SSR markers and the genetic length.

DISCUSSION

In Silico Characterization of the Common Bean Genome

The identification of SSRs using different computer programs allowed us not only to evaluate the performance and effectiveness of the software but also for in silico characterization of the genome and microsatellite types. This was especially useful in terms of motif frequency and determining the prevalence of each class of microsatellites. The combined analysis of the SSRs identified by the different software programs pointed out that microsatellite loci in the common bean genome are mainly composed of dinucleotide repeats with the AT/TA and AG/TC motifs being most common among these. Trinucleotide motifs were the next most common with many being AT rich such as the ATA/TAT and AGA/TCT. These microsatellite loci along with the AT/TA and AG/TC motifs tended to be the longest SSRs found. However, the hexanucleotide motif CTTGGG/GAACCC and CTTGGG were also prevalent in the SSRs found by TRF, which may indicate an association with retrotransposons or heterochromatin (Ramsay et al., 1999; Fonsêca et al., 2010). In rice, those hexanucleotide motif SSRs were also found to be related with transcriptional start sites (Fujimori et al., 2003).

The high prevalence of dinucleotide motif SSRs found by all the search engines used in this study agrees with

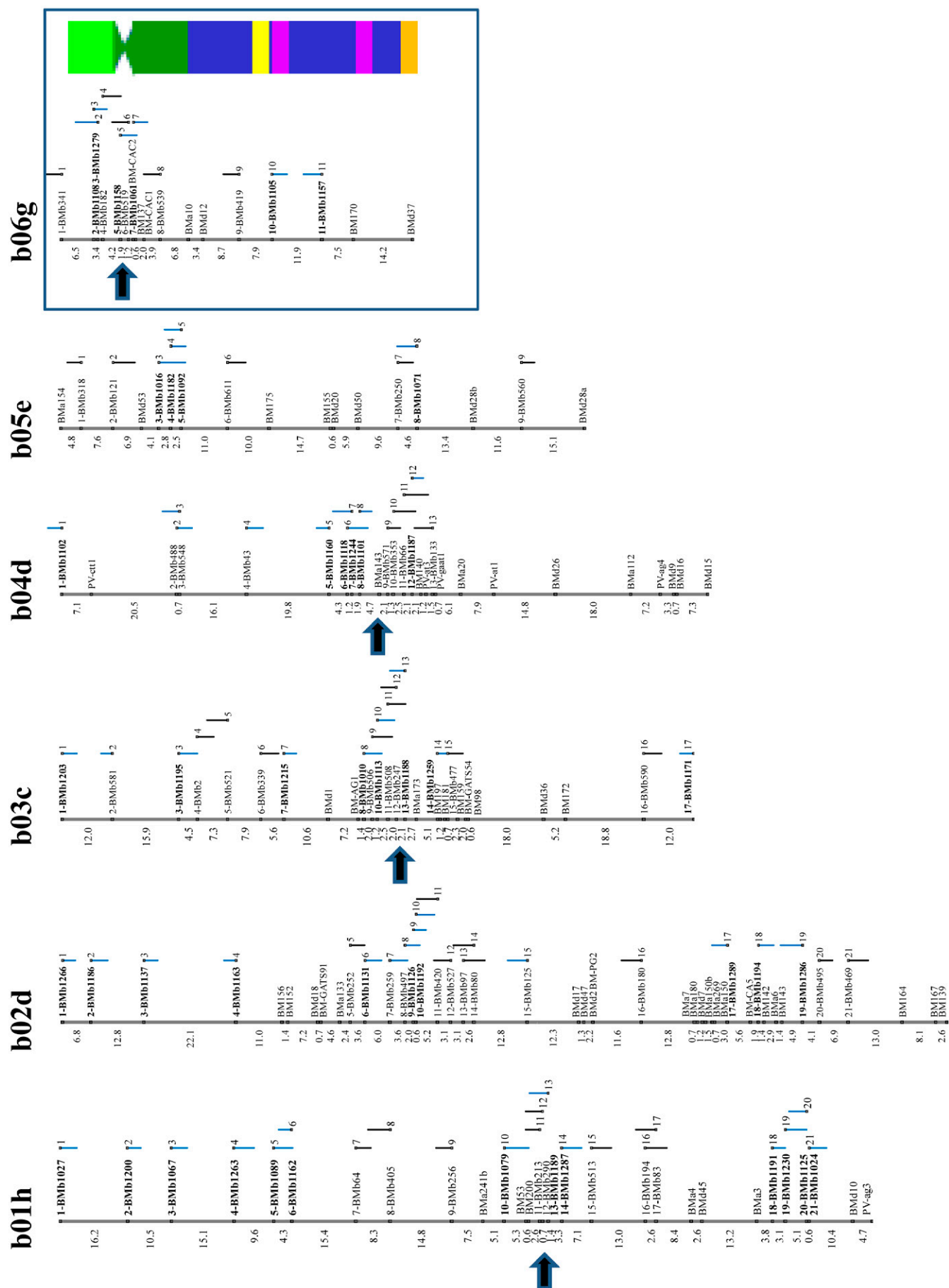
results from previous analysis of common bean microsatellites using enriched libraries. Since BES are generally from noncoding DNA, it can be expected that dinucleotide repeats would be more common than trinucleotide repeats and this would seem to be the case based on the categorization of the common bean BES by Schlueter et al. (2008). Regarding the motifs found for dinucleotide repeats in this study, the most common was the AT/TA repeat type for SSRL and TRF agreeing with previous results for BatchPrimer3 from Córdoba et al. (2010), while AG/TC was the most common motif among AMMD microsatellites. Both motifs have been reported as frequent in plant species generally (Mun et al., 2006) and in common bean in particular (Blair et al., 2003; Yu et al., 2000). If some of these microsatellites were to encode parts of proteins, their high levels of adenine and guanine would encode hydrophilic amino acids such as glutamic acid, glutamine, lysine, and serine, which are found at higher frequency in most proteins (Garnica et al., 2006; Portis et al., 2007).

In terms of the predominant microsatellite length, TRF was the only software program that identified a large number of class I SSRs while the other programs identified mainly class II SSRs. The lack of consensus among programs highlights the differences between the algorithms used by each software. In the case of AMMD, the subcomponent used for SSR search is TROLL (Castelo et al., 2002) based on an Aho-Corasick algorithm that consists in a motif dictionary-matching approach that locates elements in a finite set of strings within an input text. The algorithm moves along the input and conserves only the longest matches. However, the algorithm stops searching when faced with a discontinuity in the motif and is unable to detect imperfect microsatellites where one motif is interrupted by one or more nucleotides. This event could restrict the number of SSR repetitions found and produce more output matches corresponding to class II microsatellites.

The differences found in the motif characteristics and types of SSRs identified can also be related with the information used for the search algorithms or with the algorithm itself. Specifically, TRF algorithm used a probabilistic method that found mainly class I or perfect SSRs. In contrast, AMMD and SSRL employed motif dictionaries that contained all the possible motif combinations for their SSR searches and identified many class II SSRs but also recognized imperfect SSRs in the case of SSRL. These differences may explain why AMMD and SSRL found the highest number of SSRs since their motif libraries are very efficient at microsatellite identification. For now we can conclude that the main reason for the differences between programs was based on the heuristic nature of each algorithm, meaning each algorithm found optimal instead of best solutions within the limits of a reasonably quick run time.

Simple Sequence Repeat Polymorphism of the New BMb Markers

From the comparison of the search engines described above, we decided to concentrate our marker development based on the AMMD software with a total of 323 primer pairs selected



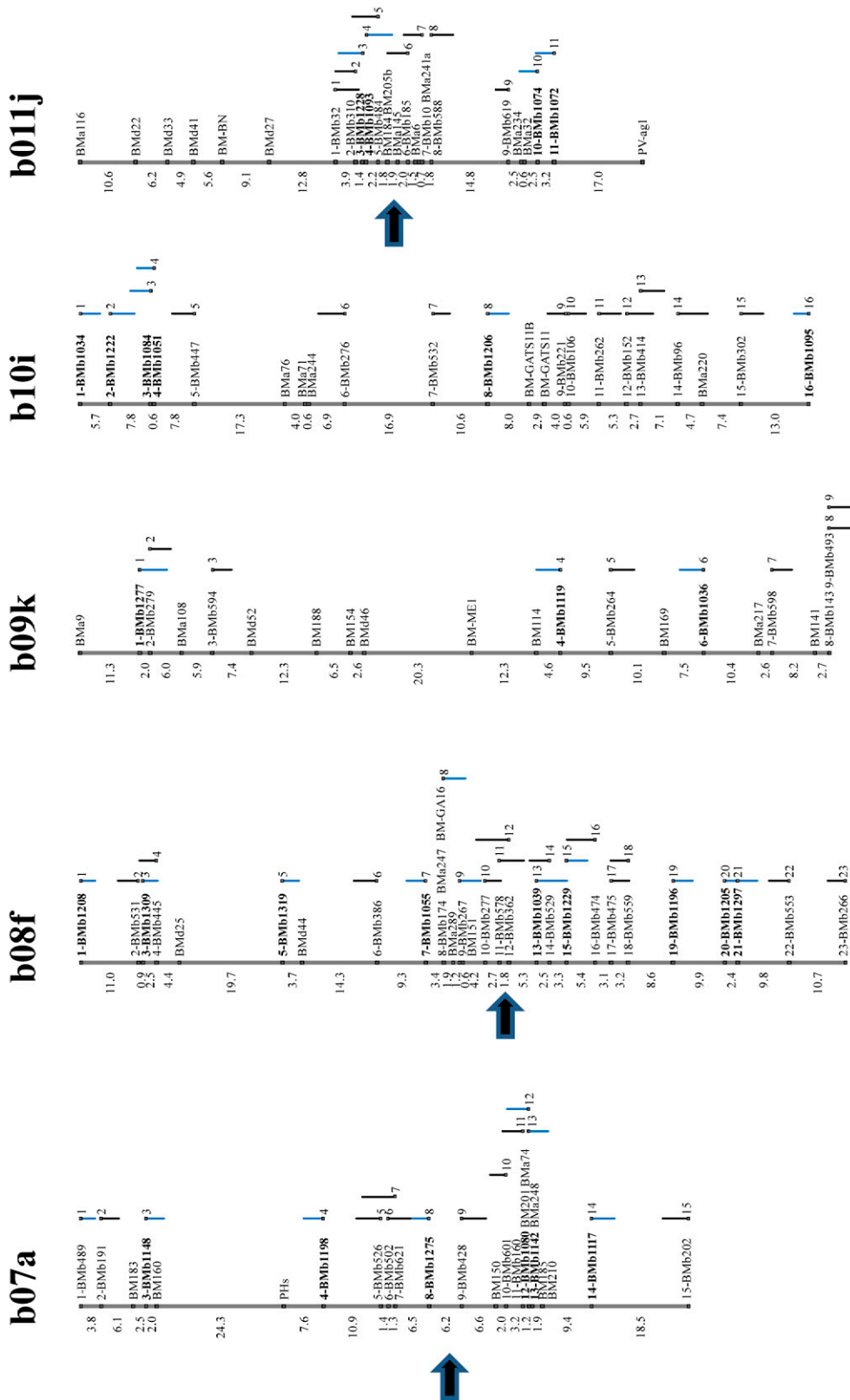


Figure 1. (continued from previous page) Integrated common bean genetic-physical map showing positions of bacterial artificial chromosome end sequence (BES)-simple sequence repeat (SSR) and non-BES-SSR markers. Abbreviations for the markers are: BM, genomic microsatellites made from enriched libraries by Gaitán et al. (2002); BMa, ATA-based genomic markers from enriched libraries described in Blair et al. (2008) or Métais et al. (2002); BMb, BES-derived markers from this study or from Córdoba et al. (2010); and BMd and PV, database-derived markers from Blair et al. (2003) and Yu et al. (2000), respectively. The new BMb markers (BMb1000 to 1322) designed with Automated Microsatellite Marker Development software (AMMD) are shown in bold compared to those from Córdoba et al. (2010) designed with BatchPrimer3 and shown unbolded (BMb001 to BMb625). All linkage groups or chromosomes are oriented according to cytogenetic chromosome length estimates of Fonsêca et al. (2010) with short arm shown above long arm based as shown for the example of linkage group b06 g (see insert to the far left of the first set of integrated maps). Large arrows to the left of the linkage groups point to clusters of BMa or BMb microsatellites, which may represent pericentromeric regions of reduced recombination.

Table 2. Characteristics of the integrated map showing number of simple sequence repeat (SSR) loci based on two microsatellite (SSR) search software programs as well as number of contigs and clones integrated per linkage group (LG) and bacterial artificial clone (BAC). Contig length and genetic versus physical distances and lengths are calculated for each LG.

LG or chromosome	No. BMB markers (BP31)	No. BMB markers (AMMD [†])	Total BMB markers	Total SSR markers	No. linked BAC contigs	No. linked BAC clones	BAC contig length (Mbp)	Genetic length (cM)	Genetic distance between markers (cM)	Physical length (Mbp) [§]	Kbp:cM ratio [¶]
b01h	8	13	21	29	21	1535	13.0	195.2	6.7	58.1	297.6
b02d	11	10	21	43	20	659	7.5	212.0	4.9	57.1	269.3
b03c	9	8	17	27	17	552	6.5	151.3	5.6	60.1	397.2
b04b	7	6	13	26	13	617	5.6	155.1	6.0	55.9	360.4
b05e	5	4	9	17	9	497	4.2	125.0	7.3	51.4	411.2
b06g	5	6	11	18	11	332	3.4	84.1	4.7	46.2	549.3
b07a	9	6	15	24	15	785	8.6	116.0	4.8	61.2	527.6
b08f	14	9	23	29	22	802	7.3	145.8	5.0	63.6	436.2
b09k	6	3	9	20	9	189	2.3	142.2	7.1	60.4	424.8
b10i	10	6	16	22	15	1467	2.3	139.9	6.3	65.0	464.6
b11j	7	4	11	25	10	797	6.9	108.3	4.3	58.3	538.3
\bar{x}	8.3	6.8	15.1	25.5	14.7	748.4	6.1	143.2	5.7	57.9	425.1
TOTAL	91	75	166	280	162	8732	78.2	1575	—	637.3	

[†]BP3, BatchPrimer3.

[‡]AMMD, Automated Microsatellite Marker Development software.

[§]Estimated physical length in Mbp for each linkage group (LG) or chromosome from Fonséca et al. (2010).

[¶]Kbp:cM ratio was calculated using the estimated chromosome length reported by Fonséca et al. (2010) and the genetic distances of the present integrated genetic map.

for testing in the parental survey mostly based on AT/TA repeats. As in previous studies, the AT-rich microsatellite were found to be highly polymorphic, something that seems to be common across all the legumes (Blair et al., 2008; Hüttel et al., 2000; Jayashree et al., 2006; Métais et al., 2002; Shoemaker et al., 2008). Higher polymorphism was found for the dinucleotide based microsatellites than for the trinucleotide based microsatellites; a trend that was observed in some previous evaluations of bean microsatellites (Blair et al., 2006, 2009b). This could be the result of the dinucleotides being mostly in noncoding regions where repeat length changes would be expected to have a lower mutational effects than in coding regions where trinucleotide repeat motifs are often found (Metzgar et al., 2000; Toth et al., 2000). The selection of a subset of AMMD microsatellites for the genetic characterization of parental genotypes would not preclude the possibility of using the SSRs identified with SSRL or TRF or use of additional software for microsatellite searches such as HIGEDA by Le et al. (2010), iMotifs by Piipari et al. (2010), and Tool for motif discovery (Tmod) by Sun et al. (2010).

With AMMD-derived microsatellites it was interesting that SSR length showed the expected relationship to the level of polymorphism. The longer class I microsatellite loci were more polymorphic than shorter class II microsatellites. The long AT-rich motifs among the class I types may have caused some of the amplification problems observed for this type of microsatellite marker, which generally used lower annealing temperatures than GA-based microsatellites (Gaitán et al., 2002; Blair et al., 2008). Higher polymorphism for class I SSRs contrasts with results observed with the SSRs identified with BatchPrimer3 (Córdoba et al., 2010). This suggests that longer microsatellite are not necessarily preferential targets for replication slippage or unequal crossing over compared to shorter microsatellites except in certain sequence contexts (Hüttel et al., 2000; Shoemaker et al., 2008).

Distribution of BES-Simple Sequence Repeat Markers on the Integrated Genetic-Physical Map

Upon genetic mapping of the polymorphic BMB markers, we found these to be evenly distributed across all linkage groups without a significant bias toward any particular chromosome. This was in agreement with the SSR mapping of AT-rich (BMA) microsatellites by Blair et al. (2008), who also found good coverage of this type of microsatellite. However, the distribution of both BMA and new BMB microsatellites inside of each linkage group was not random and both markers tended to cluster together toward the center and ends of all the linkage groups except b01h. Clustering of SSR loci is common for SSR markers identified from enriched libraries but not for those found in BES. However, in this case the clustering can be related to the AT motifs we selected and may reflect a bias of the AT-rich motifs

toward pericentromeric locations although this remains to be studied perhaps by fluorescent in situ hybridization (FISH). Despite the bias in location of the present set of BMb markers, the coverage together with previous BMb markers was very good overall as shown for the integrated genetic–physical map. The construction of a more saturated genetic map was one of the goals of this project and together with the genetic mapping of Blair et al. (2003, 2008) and Grisi et al. (2007) brings to over 400 the microsatellite loci mapped in common bean.

There are diverse methodologies for linking genetic markers to a physical map (Yuan et al., 2000). For example, BAC pooling and PCR screening (Klein et al., 2000) can be used as can hybridization using overgo probes (Chen et al., 2002; Cone et al., 2002; Yüksel et al., 2005). However the *in silico* identification and laboratory mapping of molecular markers from BAC end sequences (Kim et al., 2007; Mun et al., 2006; Shoemaker et al., 2008; Shultz et al., 2007; Troggio et al., 2007) is both a useful way to integrate physical–genetic maps and to develop new markers simultaneously. Thus, the justification for this approach in our study was that we continued to enrich the number of markers for the common bean genome, especially based on higher polymorphism SSR markers, which have greater utility than SNPs due to their multiallelism. Overall in this study, a total of 166 BMb and 114 non-BMb markers have been integrated together to form a highly saturated map of 280 SSR markers.

The low physical–genetic correlation found for all linkages groups other than b06 g is in agreement with previous results obtained for Pedrosa et al. (2003) and Fonsêca et al. (2010) and, as suggested by them, can represent the need to establish a more detailed physical map or the occurrence of a low physical–genetic correlation in common bean. The exceptional case of linkage group b06 g is related to the chromosome's acrocentric nature and the presence of a short arm consisting only of 45s rDNA sequences and no single copy sequences. The cluster of microsatellites at the top of our b06g linkage group may reflect the low recombination rate in the pericentromeric region and ultimately the location of the centromere. This is supported by the close linkage of many AT-rich BMb and BMa microsatellite sequences at that site, which may indicate more heterochromatin than euchromatin in that region.

This clustering of AT-rich markers agrees with the results of Blair et al. (2008) and is also seen at the center of some other linkage groups, which corresponded with the metacentric chromosomes represented by linkage groups b04d and b08f and submetacentric chromosomes represented by linkage groups b01h, b03c, b07a, and b11j (small arrows in Fig. 1). B02d was an exception in that it had two clusters, which may reflect low recombination in two regions of the corresponding chromosome. For linkages groups b05e, b09k, and b10i clustering of microsatellites markers was less evident suggesting less evidence for positioning of pericentromeric BACs and both b09k and b10i are known to have large stretches of rDNA in their short and long arms, respectively (Fonsêca et al., 2010), which may have influenced recombination rate in certain parts

of the genome. In summary, the genome architecture in terms of centromere position, rDNA loci and distribution of single-copy versus repetitive DNA may have influenced the overall distribution of SSR loci on our genetic map. In this context it was interesting that the three linkage groups with substantial rDNA (b06 g, b09k, and b10i) had among the highest Kbp:cM ratio. For the time being, this study has provided an initial view of the common bean genome from a genetic perspective with linkages to the only physical map for the species that is currently available (Schlueter et al., 2008). The genetic and physical map presented here will be complemented in the near future by high-throughput and whole-genome shotgun sequencing of the G19833 genotype (S. Jackson, personal communication, 2010), making further *in silico* marker development and even greater reality (Meyers et al., 2004).

CONCLUSIONS

Genome characterization and map development were facilitated by using various microsatellite search engines to find BES-SSRs and to increase the genome coverage of the integrated map. Specifically in the case of AMMD-derived BMb markers there were a larger number of contig-based BES-SSR markers that were AT rich and that were used to genetically map additional microsatellites near those from Blair et al. (2008). The only limitation of AMMD was its discovery of lower polymorphism SSRs compared to those of Córdoba et al. (2010), although this was made up for by the finding of a larger number of SSRs. The search for more informative SSRs is not over, however, and should continue with complementary motifs or additional software programs. In summary, the *in silico* methods used for BES-SSR identification was a good option for obtaining widespread and well distributed markers of adequate polymorphism. More specifically, the BAC end sequences were a good sample of the common bean genome to use in microsatellite discovery and provided information on the genome itself. Finally, the molecular characterization of the BES-SSRs was a key step to advance physical–genetic map integration with 166 polymorphic BES-SSRs found so far that have shown utility in estimating the genetic to physical distance ratio across different linkage groups. The SSRs have also been useful in associating other genomic feature such as possible pericentromeric regions with types of SSR repeats prevalent in common bean. Therefore, the new BMb markers not only have contributed to the saturation of the common bean integrated map but the SSR search was a valuable exercise in genome analysis. Finally, the BMb markers presented here increase the number of points of comparison between the physical and genetic maps of common bean and therefore make genomic analysis of specific genes more likely and more accurate.

Authors' Contributions

JMC participated in planning of the study, constructed the integrated map, analyzed the results, and cowrote the paper. CC tested the AMMD software. FR downloaded, implemented, and tested the AMMD software. JMC and CM performed the SSR genotyping. MWB conceived

of and coordinated the study, obtained funding for the overall project, and cowrote the paper.

Acknowledgments

We acknowledge Fernando Niño of the National University of Colombia – Bogotá for advice on this work and on programming; Fredy Monserrate for computational assistance; Agobardo Hoyos for seed preparation; and Yarinka Rojas for laboratory help. The Generation Challenge Program and Tropical Legume I project of the Bill and Melinda Gates Foundation are gratefully acknowledged for funding this project.

References

- Benson, G. 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580.
- Blair, M., H. Buendia, M. Giraldo, I. Métais, and D. Peltier. 2008. Characterization of AT-rich microsatellite in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 118:91–103.
- Blair, M., M. Giraldo, H. Buendía, E. Tovar, M. Duque, and S. Beebe. 2006. Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 113:100–109.
- Blair, M., and M. Ishitani. 2009. Common bean: A model food legume for international agriculture. *Grain Legumes* 53:8–9.
- Blair, M., M. Muñoz, M. Giraldo, and F. Pedraza. 2009a. Development and diversity of Andean-derived, gene-based microsatellites for common bean (*Phaseolus vulgaris* L.). *BMC Plant Biol.* 9:100.
- Blair, M., M. Muñoz, F. Pedraza, M. Giraldo, H. Buendia, and N. Hurtado. 2009b. Development of microsatellite markers for common bean (*Phaseolus vulgaris* L.) based on screening of non-enriched, small-insert genomic libraries. *Genome* 52:1–11.
- Blair, M.W., F. Pedraza, H. Buendia, E. Gaitán-Solís, S. Beebe, and J. Tohme. 2003. Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 107:1362–1374.
- Broughton, W., G. Hernández, M. Blair, S. Beebe, P. Gepts, and J. Vanderleyden. 2003. Beans (*Phaseolus* spp.) model food legumes. *Plant Soil* 252:55–128.
- Castelo, A., W. Martins, and G. Gao. 2002. Tandem repeat occurrence locator. *Bioinformatics* 8:634–636.
- Chen, M., G. Presting, W. Barbazuk, J. Goicoechea, B. Blackmon, G. Fang, H. Kim, D. Frisch, Y. Yu, S. Sun, S. Hingebotton, J. Phimpilai, D. Phimpilai, S. Thurmond, B. Gaudette, P. Li, J. Liu, J. Hatfield, D. Main, K. Farrar, C. Henderson, L. Barnett, R. Costa, B. Williams, S. Walser, M. Atkins, C. Hall, M. Budiman, J. Tomkins, M. Luo, I. Bancroft, J. Salse, F. Regad, T. Mohapatra, N. Singh, A. Tyagi, C. Soderlund, R. Dean, and R. Wing. 2002. An integrated physical and genetic map of the rice genome. *Plant Cell* 14:537–545.
- Coburn, J., S. Temnykh, E. Paul, and S. McCouch. 2002. Design and application of microsatellite marker panels for semi-automated genotyping of rice (*Oryza sativa* L.). *Crop Sci.* 42:2092–2099.
- Cone, K., M. McMullen, I. Bi, G. Davis, Y. Yim, J. Gardiner, M. Polacco, H. Sánchez-Villeda, Z. Fang, S. Schroeder, S. Havermann, J. Bowers, A. Paterson, C. Soderlund, F. Engler, R. Wing, and E. Coe. 2002. Genetic, physical, and informatics resources for maize. On the road to an integrated map. *Plant Physiol.* 130:1598–1605.
- Córdoba, J.M., M.C. Chavarro, J.J. Schleuter, S.A. Jackson, and M.W. Blair. 2010. Integration of physical and genetic maps of the common bean genome through microsatellite markers. *BMC Genomics* 11:436.
- da Maia, L., D. Palmieri, V. Queiroz, M. Marini, F. Félix, and A. Costa. 2008. SSRLocator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int. J. Plant Genomics* 2008:1–9.
- Fonsêca, A., J. Ferreira, T. Barros dos Santos, M. Mosiolek, E. Bellucci, J. Kami, P. Gepts, V. Geffroy, D. Schweizer, K. dos Santos, and A. Pedrosa-Harand. 2010. Cytogenetic map of common bean (*Phaseolus vulgaris* L.). *Chromosome Res.* 18(4):487–502.
- Fujimori, S., T. Washio, K. Higo, Y. Ohtomo, K. Murakami, K. Matsubara, J. Kawai, P. Carninci, Y. Hayashizaki, S. Kikuchi, and M. Tomita. 2003. A novel feature of microsatellites in plants: A distribution gradient along the direction of transcription. *FEBS* 554:17–22.
- Gaitán, E., M. Duque, K. Edward, and J. Tohme. 2002. Microsatellite repeats in common bean (*Phaseolus vulgaris*): Isolation, characterization, and cross-species amplification in *Phaseolus* spp. *Crop Sci.* 42:2128–2136.
- Garnica, D., A. Pinzon, L. Quesada-Ocampo, A. Bernal, E. Barreto, N. Grünwald, and S. Restrepo. 2006. Survey and analysis of microsatellites from transcript sequences in *Phytophthora* species: Frequency, distribution, and potential as markers for the genus. *BMC Genomics* 7:245–255.
- Gore, M., M. Wright, E. Ersoz, P. Bouffard, E. Szekeres, T. Jarvie, B. Hurwitz, A. Narechania, T. Harkins, G. Grills, D. Ware, and E. Buckler. 2009. Large-scale discovery of gene-enriched SNPs. *Plant Gen.* 2:121–133.
- Grisi, M.C.M., M.W. Blair, P. Gepts, C. Brondani, P.A.A. Pereira, and R.P.V. Brondani. 2007. Genetic mapping of microsatellite markers in common bean (*Phaseolus vulgaris*) population BAT93 × Jalo EEP558. *Genet. Mol. Res.* 6:691–706.
- Hamblin, M., M. Warburton, and E. Buckler. 2007. Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PLoS One* 2:e1367.
- Hong, C., P. Plaha, D. Koo, T. Yang, S. Choi, Y. Lee, T. Uhm, J. Bang, D. Edwards, I. Bancroft, B. Park, J. Lee, and Y. Lim. 2006. A survey of the *Brassica rapa* genome by BAC end sequence analysis and comparison with *Arabidopsis thaliana*. *Mol. Cells* 22:300–307.
- Hüttel, B., P. Winter, K. Weising, W. Choumane, F. Weigand, and G. Kahl. 2000. Sequenced-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). *Genome* 12:210–217.
- Jayashree, B., R. Punna, P. Prasad, K. Bantte, T. Hash, S. Chandra, D. Hoisington, and R. Varshney. 2006. A database of simple sequence repeats from cereal and legume expressed sequenced tags mined *in silico*: Survey and evaluation. *In Silico Biol.* 6:607–620.
- Kelley, J., C. Field, M. Craven, D. Bocskai, U. Kim, S. Rounsley, and M. Adams. 1999. High throughput direct end sequencing of BAC clones. *Nucleic Acids Res.* 27:1539–1546.
- Kim, D., K. Kim, K. Van, M. Kim, and S. Lee. 2010. Fine mapping of a resistance gene to bacterial leaf pustule in soybean. *Theor. Appl. Genet.* 120:1443–1450.
- Kim, H., P. San Miguel, W. Nelson, K. Collura, M. Wissotski, J. Walling, J. Kim, S. Jackson, C. Soderlund, and R. Wing. 2007. Comparative physical mapping between *Oryza sativa* (AA genome type) and *O. punctata* (BB genome type). *Genetics* 176:379–390.
- Klein, P., R. Klein, S. Cartinhour, P. Ulanich, J. Dong, J. Obert, D. Morishige, S. Schlueter, K. Childs, M. Ale, and J. Mullet. 2000. A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. *Genome Res.* 10:789–807.
- Kofler, R., J. Bartos, L. Gong, G. Stif, P. Suchankova, H. Simkova, M. Berenyi, K. Burg, J. Dolezel, and T. Lele. 2008. Development of microsatellite markers specific for the short arm of rye (*Secale cereale* L.) chromosome 1. *Theor. Appl. Genet.* 117:915–926.
- Le, T., T. Altman, and K. Gardiner. 2010. HIGEDA: A hierarchical genetic based algorithm for finding subtle motifs in biological sequences. *Bioinformatics* 26:302–309.
- Martins, W., D. de Sousa, K. Proite, P. Guimarães, M. Moretzsohn, and D. Bertioli. 2006. New softwares for automated microsatellite marker development. *Nucleic Acids Res.* 34:e31.
- Métais, I., B. Hamon, R. Jalouzy, and D. Peltier. 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.
- Metzgar, D., J. Bytof, and C. Wills. 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome* 10:72–80.
- Meyers, B., S. Scalabrin, and M. Morgante. 2004. Mapping and sequencing complex genomes: Let's get physical. *Natl. Rev.* 5:578–589.
- Mun, J., D. Kim, H. Choi, J. Gish, F. DeBelle, J. Mudge, R. Denny, G. Endre, O. Saurat, A. Dudge, G. Kiss, B. Roe, N. Young, and D. Cook. 2006. Distribution of microsatellites in the genome of *Medicago*

- truncatula*: A resource of genetic markers that integrate genetic and physical maps. *Genetics* 172:2541–2555.
- Nicot, N., V. Chiquet, B. Gandon, L. Amilhat, F. Legeai, P. Leroy, M. Bernard, and P. Sourdille. 2004. Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). *Theor. Appl. Genet.* 109:800–805.
- Pedrosa, A., C. Vallejos, A. Bachmair, and D. Schweizer. 2003. Integration of common bean (*Phaseolus vulgaris* L.) linkage and chromosomal maps. *Theor. Appl. Genet.* 106:205–212.
- Piipari, M., T. Down, H. Saini, A. Enright, and T. Hubbard. 2010. iMotifs: An integrated sequence motif visualization and analysis environment. *Bioinformatics* 26:843–844.
- Portis, E., I. Nagy, Z. Sasvari, A. Stigel, L. Barchi, and S. Lanteri. 2007. The design of *Capsicum* spp. SSR assays via analysis of *in silico* DNA sequence and their potential utility for genetic mapping. *Plant Sci.* 172:640–648.
- Ramsay, L., M. Macaulay, L. Cardle, M. Morgante, S. degli Ivanissevich, E. Maestri, W. Powell, and R. Waugh. 1999. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J.* 17:415–425.
- Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In K. Misener (ed.) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. 2000. Humana Press, Totowa, NJ.
- Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109:365–371.
- Schlueter, J., J. Goicoechea, K. Collura, N. Gill, J. Lin, Y. Yu, D. Kudrna, A. Zuccolo, C. Vallejos, M. Muñoz-Torres, M.W. Blair, J. Tohme, J. Tomkins, P. McClean, R. Wing, and S. Jackson. 2008. BAC-end sequence analysis and a draft physical map of the common bean (*Phaseolus*). *Trop. Plant Biol.* 1:40–48.
- Shoemaker, R., D. Grant, T. Olson, W. Warren, R. Wing, Y. Yu, H. Kim, P. Cregan, B. Joseph, M. Futrell-Griggs, W. Nelson, J. Davito, J. Walker, J. Wallis, C. Kremitski, D. Scheer, S. Clifton, T. Graves, H. Nguyen, X. Wu, M. Luo, J. Dvorak, R. Nelson, S. Cannon, J. Tomkins, J. Schmutz, G. Stacey, and S. Jackson. 2008. Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. *Genome* 51:294–302.
- Shultz, J., S. Kazi, R. Bashir, J. Afzal, and D. Lightfoot. 2007. The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. *Theor. Appl. Genet.* 114:1081–1090.
- Sun, H., Y. Yuan, Y. Wu, H. Liu, J. Liu, and H. Xie. 2010. Tmod: Toolbox of motif discovery. *Bioinformatics* 26:405–407.
- Toth, G., Z. Gaspari, and J. Jurka. 2000. Microsatellites in different eukaryotic genomes survey and analysis. *Genome Res.* 10:967–981.
- Troggio, M., G. Malacarne, G. Coppola, C. Segala, D. Cartwright, M. Pindo, M. Stefanini, R. Mank, and M. Moroldo. 2007. A dense SNP-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring Pinot Noir BAC contigs. *Genetics* 176:2637–2650.
- Wondji, C., R. Hunt, P. Pignatelli, K. Steen, M. Coetzee, N. Besansky, N. Lobo, F. Collins, J. Hemingway, and H. Ranson. 2005. An integrated genetic and physical map for the malaria vector *Anopheles funestus*. *Genetics* 171:1779–1787.
- You, F., N. Huo, Y. Gu, M. Luo, Y. Ma, D. Hane, G. Lazo, J. Dvorak, and O. Anderson. 2008. BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinform.* 9:253–266.
- Yu, K., S. Park, V. Poysa, and P. Gepts. 2000. Integration of simple sequence repeats (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris*). *J. Hered.* 91:429–434.
- Yuan, Q., F. Liang, J. Hsiao, V. Zismann, M. Benito, J. Quakenbush, R. Wing, and R. Buell. 2000. Anchoring of BAC rice clones to the rice genetic map *in silico*. *Nucleic Acids Res.* 28:3635–3641.
- Yüksel, B., J. Bowers, J. Estill, L. Goff, C. Lemke, and A. Paterson. 2005. Exploratory integration of peanut genetic and physical maps and possible contributions from *Arabidopsis*. *Theor. Appl. Genet.* 111:87–94.