

RESEARCH ARTICLE

Open Access

# Gene-based SSR markers for common bean (*Phaseolus vulgaris* L.) derived from root and leaf tissue ESTs: an integration of the BMc series

Matthew W Blair<sup>1\*</sup>, Natalia Hurtado<sup>1</sup>, Carolina M Chavarro<sup>1</sup>, Monica C Muñoz-Torres<sup>1,2,3</sup>, Martha C Giraldo<sup>1,4</sup>, Fabio Pedraza<sup>1,5</sup>, Jeff Tomkins<sup>2</sup>, Rod Wing<sup>2,6</sup>

## Abstract

**Background:** Sequencing of cDNA libraries for the development of expressed sequence tags (ESTs) as well as for the discovery of simple sequence repeats (SSRs) has been a common method of developing microsatellites or SSR-based markers. In this research, our objective was to further sequence and develop common bean microsatellites from leaf and root cDNA libraries derived from the Andean gene pool accession G19833 and the Mesoamerican gene pool accession DOR364, mapping parents of a commonly used reference map. The root libraries were made from high and low phosphorus treated plants.

**Results:** A total of 3,123 EST sequences from leaf and root cDNA libraries were screened and used for direct simple sequence repeat discovery. From these EST sequences we found 184 microsatellites; the majority containing tri-nucleotide motifs, many of which were GC rich (ACC, AGC and AGG in particular). Di-nucleotide motif microsatellites were about half as common as the tri-nucleotide motif microsatellites but most of these were AG<sub>n</sub> microsatellites with a moderate number of AT<sub>n</sub> microsatellites in root ESTs followed by few AC<sub>n</sub> and no GC<sub>n</sub> microsatellites. Out of the 184 new SSR loci, 120 new microsatellite markers were developed in the BMc (Bean Microsatellites from cDNAs) series and these were evaluated for their capacity to distinguish bean diversity in a germplasm panel of 18 genotypes. We developed a database with images of the microsatellites and their polymorphism information content (PIC), which averaged 0.310 for polymorphic markers.

**Conclusions:** The present study produced information about microsatellite frequency in root and leaf tissues of two important genotypes for common bean genomics: namely G19833, the Andean genotype selected for whole genome shotgun sequencing from race Peru, and DOR364 a race Mesoamerica subgroup 2 genotype that is a small-red seeded, released variety in Central America. Both race Peru and Mesoamerica subgroup 2 (small red beans) have been understudied in comparison to race Nueva Granada and Mesoamerica subgroup 1 (black beans) both with regards to gene expression and as sources of markers. However, we found few differences between SSR type and frequency between the G19833 leaf and DOR364 root tissue-derived ESTs. Overall, our work adds to the analysis of microsatellite frequency evaluation for common bean and provides a new set of 120 BMc markers which combined with the 248 previously developed BMc markers brings the total in this series to 368 markers. Once we include BMd markers, which are derived from GenBank sequences, the current total of gene-based markers from our laboratory surpasses 500 markers. These markers are basic for studies of the transcriptome of common bean and can form anchor points for genetic mapping studies in the future.

\* Correspondence: [mwblaircgar@gmail.com](mailto:mwblaircgar@gmail.com)

<sup>1</sup>CIAT - International Center for Tropical Agriculture, Biotechnology Unit and Bean Project, AA6713, Cali, Valle, Colombia

Full list of author information is available at the end of the article

## Background

Genic microsatellites are those microsatellites based on simple sequence repeats (SSRs) found within, or closely associated with, gene sequences from a given genome [1]. These SSRs tend to be more conserved and of different motifs than SSRs located in other non-gene containing regions of the genome, which are often referred to as genomic microsatellites simply to distinguish them from genic microsatellites [2]; although both gene and non-gene derived microsatellites are obviously part of the overall genome. Simple sequence repeats are defined as small stretches of repeated DNA, usually of two to six nucleotides, tandemly repeated and located in a given pattern between segments of non-repeated DNA [3]. In practice, remnant repeats can be found on either side of a stretch of SSR and in some occasions different motifs are combined together or either motif is interrupted [4]. This differentiates microsatellites into compound or simple microsatellites in the first case, and perfect and imperfect microsatellites in the latter case [5].

Common bean, *Phaseolus vulgaris* L., is an important food legume, basic to the diet of the poor in tropical regions of the world, and a major source of income for small farmers there. Genic microsatellites have been limited in number for this crop. This is perhaps due to two main reasons: 1) a lack of funding has precluded large scale expressed sequence tag (EST) sequencing or even the sufficient construction of many cDNA libraries for the crop and 2) those ESTs and cDNA libraries that exist have not been extensively screened for gene-based SSRs with the exception of the work of Blair et al. [6] and Hanai et al. [7,8]. Yet, common bean is essential for micronutrient nutrition and is adaptable to marginal areas for small-scale farm agriculture despite problems of low phosphorus soils or other abiotic constraints [9,10] and a range of diseases and pests [11]. Therefore a more complete toolbox of molecular tools for this crop is needed especially in the case of gene-based markers which can be based on SSRs polymorphisms as will be discussed here.

In our efforts to accumulate a larger set of genic SSRs, we previously constructed a leaf based cDNA library from Andean genotype G19833 [12] and used a hybridization approach to discover SSRs of various di-nucleotide and tri-nucleotide motifs and develop microsatellites from this library in the BMc (Bean microsatellites from cDNAs) series [6]. We have also recently developed two additional root based cDNA libraries under high and low phosphorus conditions from the Mesoamerican genotype DOR364, the other parent of the mapping population of Blair et al. [2] and sequenced ESTs from the libraries to discover new SSRs.

The EST sequencing of these libraries is used in this research as the basis for determining the frequency of

SSR sequences in root expressed genes as opposed to leaf expressed genes and for adding to the BMC series of microsatellites through an *in silico* approach to microsatellite discovery as described by Varshney et al. [13] for some species of cereals. EST-SSRs are more common in cereals than they are in legumes [14-16].

Apart from our efforts, currently there are approximately 70,000 other EST sequences from common bean including collections from Ramirez et al. [17], Melotto et al. [18] and Thibivilliers et al. [19] along with small groups of GenBank entries and a wish-list of further EST efforts [9]. However most of these libraries have not been screened nor compared for SSR markers. The Melotto et al. [18] libraries from anthracnose infected common bean leaves which contain together approximately 4,000 unigenes has been screened for microsatellites, yielding a set of 140 EST-based SSRs for Hanai et al. [7,8], although many of these have been used for genetic mapping rather than for germplasm characterization.

The objective of this research was to evaluate the frequency of microsatellites in sequences from different leaf and root EST libraries made in our laboratory, comparing the types of microsatellites from each source tissue. From there we developed the most promising microsatellite loci as gene-based SSR markers that we added to the BMC series of markers [6]. To validate these BMC markers we compared their ability to detect polymorphism in a standard germplasm panel of 18 mapping parent genotypes, which included Mesoamerican, Andean, wild and cultivated accessions that were useful for determining polymorphism information content of the different groups of markers. A final objective was to determine whether any difference in the ability to uncover polymorphism existed between the newly developed BMC markers found in the random EST sequencing versus BMC markers developed by our previous hybridization-based approach.

## Methods

### cDNA library and EST sequencing

Three cDNA libraries were searched for microsatellite containing sequences. These libraries were based on 1) mRNA from leaf and stem tissues as described in Blair et al. [12] and Ramirez et al. [17] for the genotype G19833; 2) a library that was made in the pBS-SKII vector from mRNAs of hydroponically grown DOR364 roots which were produced under low phosphorus (LP) conditions and 3) a final library also made in the pBS-SKII vector from mRNAs of hydroponically grown DOR364 roots but which were from a high phosphorus (HP) treatment.

In total, 1308 ESTs were sequenced from the G19833 leaf and stem tissue library: 540 sequences (GenBank entry BQ481427-BQ481965) from Blair et al. [12] and 768 sequences (HS089176-HS089943) sequenced for

this study. Meanwhile, a total of 1815 ESTs were sequenced from the DOR364 root tissue libraries: these being 862 from the HP library (GenBank entries, HS103978-HS104836) and 953 from the LP library (GenBank entries, HS103028-HS103977). Clones from all cDNA libraries were sequenced from the 5' end using BigDye chemistry (Applied Biosystems by Life Technologies; Carlsbad, CA) and di-deoxy-based Sanger sequencing reactions at the Clemson University Genomics Institute (CUGI). All EST sequences were screened for microsatellites to be assigned to the BMc series as described in Blair et al. [6] and with the methods given below.

#### **SSR identification, primer design and microsatellite amplification**

SSRs were identified by screening the EST collections with SSR Locator [20] with the default option of 1 to 6 nucleotide repeats. Primers were designed using Primer3 [21] with the following conditions: optimum primer length of 20 nucleotides (nt, minimum 18 nt - maximum 26 nt), optimum melting temperature of 50°C (min. 45°C - max. 55°C), an optimum product size of 125 base-pairs (bp, min. 100 bp - max. 350 bp) and an optimum G/C content of 50% (min. 45% - max. 55%). New markers were submitted as STS entries to GenBank and are listed in the Additional file 1 (Table S1).

PCR reaction conditions for all newly designed BMC markers and for the 248 BMc markers from Blair et al. [6] are as follows: 30 ng of genomic DNA, 0.16 µM of mixed forward and reverse primers, 1X Buffer (10 mM de Tris-HCl pH 8.2, 50 mM KCl, Triton 0.1%, BSA 1mg/ml), 1.5 mM MgCl<sub>2</sub>, 0.2 mM dNTPs and 1 U *Taq* polymerase in 12 µL reaction volumes. Amplification conditions were based on those described in Blair et al. [6,22] with 35 cycles and 47°C annealing temperature. PCR reaction products were run on PTC-200 thermal cyclers (MJR, Bio-Rad Laboratories; Hercules, CA) and then denatured at 94°C and run on 4% polyacrylamide gels (5M urea, 0.5X TBE) in metal backed Owl T-Rex vertical S3S gel units (Thermo Fisher Scientific Inc; Waltham, MA) at constant 120 W. Silver staining was performed as described in Blair et al. [22,23].

#### **Germplasm survey**

The set of genotypes used for the polymorphism survey in this study was based on a germplasm panel of 18 genotypes described in Blair et al. [22] as panel I. Both the DOR364 genotype, a Mesoamerican gene pool advanced line from the International Center for Tropical Agriculture (CIAT), and the G19833 genotype, an Andean gene pool Peruvian landrace in the FAO collection at CIAT were obtained from the gene bank in the Genetic Resources Unit (GRU), and used in a polymorphism survey since

these were the sources of the EST libraries we screened for microsatellite loci. Along with these two genotypes the germplasm survey included nine more domesticated Mesoamerican accessions and varieties (G3513, G4825, G11360, G11350, G14519, G21212, BAT477, BAT881 and DOR390), four other domesticated Andean accessions or varieties (G21078, G21657, G21242, Radical Cerinza) and three wild accessions (G19892, G24390 and G24404) representing Andean, Mesoamerican and Colombian wild sub-populations) which were also provided by the GRU.

DNA extraction consisted in a CTAB based mini-prep procedure as described in Afanador et al. [24] using bulk leaf tissue from four greenhouse grown plants per genotype or line. Since the accessions were from lines separated by seed color and maintained at the gene bank, or from advanced lines from the CIAT collection, we assumed homozygosity for all the germplasm but noted any double banding that could indicate a heterozygote or heterogeneous mixture from the four plants. Although beans are a highly inbreeding species (95 to 99%) some outcrossing occurs occasionally so there can be some within accession or intra-population variation and this would be observable in any lanes containing more than one band, representing more than one allele in seeds of the accession.

#### **Data analysis**

Allele sizes were estimated for the survey panel and mapping gels based on comparison with 10 and 25 bp molecular weight ladders that were distributed twice on each silver stained gel. A neighbor-joining (NJ) dendrogram was constructed with the proportion of shared alleles coefficient and matrix of alleles and genotypes for the survey panel with the software programs Darwin [25]. Polymorphism information content (PIC) was calculated for each marker with Powermarker [26].

#### **Results**

##### **Comparisons of EST-SSR repeat types and marker development**

Among the SSR motifs identified (Table 1), trinucleotides were the most common with 99 out of 184 found (53.8%) while di-nucleotide repeats were the second most common with 57 out of 184 found (30.9%). Meanwhile, only a few tetra-nucleotide (23) and penta- or hexa-nucleotide (5) SSRs were observed. Across all the EST sequencing sets the percentage of ESTs containing SSRs varied from 3.5 to 11.9% with the highest number found in the first sequencing of the leaf library and the least in the second sequencing of the leaf library which may have been due to sampling differences. The numbers of SSRs per ESTs in the two root libraries were similar, with 5.4% for the HP library and 4.8% for the LP library. When comparing the leaf versus root tissues we found

**Table 1 Microsatellites, simples sequence repeat (SSR) class and motif type found with in EST collections positive for SSR loci**

Tissue/Library type	Genotype/Gene pool	EST collection/author	EST No.	EST-SSRs found	2-nt	3-nt	4-nt	5/6nt	% EST-SSRs	GenBank entries for ESTs
Leaf cDNA	G19833	Blair (2002)	540	64	9	34	21	0	11.9	BQ481427-BQ481965,
Leaf cDNA	G19833	Blair (this study)	768	27	10	16	0	1	3.5	HS089176-HS089943
<b>subtotal</b>	<b>Andean</b>	<b>NA</b>	<b>1308</b>	<b>91</b>	<b>19</b>	<b>50</b>	<b>21</b>	<b>1</b>	<b>7.0</b>	<b>NA</b>
HP root cDNA	DOR364	Blair (this study)	862	47	20	23	2	2	5.5	HS103978-HS104836
LP root cDNA	DOR364	Blair (this study)	953	46	10	26	0	2	4.8	HS103028-HS103977
<b>subtotal</b>	<b>Mesoamerican</b>	<b>NA</b>	<b>1815</b>	<b>93</b>	<b>30</b>	<b>49</b>	<b>2</b>	<b>4</b>	<b>5.1</b>	<b>NA</b>
<b>grand total</b>	<b>Andean/Mesoamerican</b>	<b>NA</b>	<b>3123</b>	<b>184</b>	<b>57</b>	<b>99</b>	<b>23</b>	<b>5</b>	<b>5.9</b>	<b>NA</b>

that 6.9% of the leaf ESTs had SSRs while 5.1% of the root ESTs had SSRs so the values were similar overall. More tetra-nucleotide SSRs were found in leaf ESTs than in root ESTs while the number of di-nucleotide SSRs in relationship to the number of ESTs sequenced was similar in the two EST collections. Similar numbers of tri-nucleotides were found in ESTs from each type of tissue.

When comparing the specific motifs for SSRs found in each set of ESTs (Table 2) we observed similar frequencies of specific types of motifs among the di-nucleotides but different frequencies of specific types of motifs among the tri-nucleotides. Overall among the di-nucleotides AG/CT/GA/TC microsatellites were much more common than other types of di-nucleotide motifs with 41 out of 57 of these SSRs (71.9%). The next most common was the AT/TA microsatellites with 12 out of 57 of these SSRs (21.1%) while no CG/GC microsatellites were found. Only four AC/GT/CA/TG microsatellites were found constituting only 7.0% of the total

di-nucleotide repeat motif SSRs identified. Among the tri-nucleotide SSRs, AAG/AGA/GAA/TTC/TCT/CTT was the most common motif with 23% of the total followed by AGG/GAG/GGA/TCC/CTC/CCT with 16%. The CGC and ATA-rich microsatellites were the least common with all others being intermediate.

In the effort to develop additional cDNA-derived microsatellites, we added 120 new BMc (bean microsatellites from cDNAs series) markers to the 248 previously developed BMc markers [6]. Among the microsatellites, the first seventeen (BMc1 to BMc17) were developed from leaf cDNAs in the library described in [6,12] and as shown in the Additional file 1 (Table S1). A second set of leaf cDNA derived microsatellites from our second EST sequencing effort in this library were designated as BMc18 to BMc27. Meanwhile, 47 microsatellite markers (BMc28 to BMc74 plus BMc77 to BMc109 except BMc55 and BMc59) were developed from the HP root library and 46 other microsatellite markers (BMc55,

**Table 2 Percentage of SSR types across four EST collections**

SSR Type/Genotype/Tissue source	G19833 set 1 leaf cDNAs	G19833 set 2 leaf cDNAs	DOR364 root HP	DOR364 root LP	Total SSR and Seq.
<i>Di-nucleotide motifs<sup>1</sup></i>					
ac/gt/ca/tg	11.1	10.0	0.0	11.1	7.0
ag/ct/ga/tc	88.9	50.0	85.0	61.1	71.9
at/ta	0.0	40.0	15.0	27.8	21.1
gc/cg	0.0	0.0	0.0	0.0	0.0
<i>Tri-nucleotide motifs</i>					
aag/aga/gaa/ttc/tct/ctt	11.8	25.0	30.4	30.8	23.2
aat/ata/taa/tta/tat/att	2.9	0.0	4.3	7.7	4.0
aac/aca/caa/ttg/tgt/gtt	8.8	6.3	13.0	3.8	8.1
acc/cac/cca/tgg/gtg/ggt	17.6	6.3	4.3	19.2	13.1
agc/cag/gca/tcg/gtc/cgt	17.6	12.5	13.0	3.8	12.1
agg/gag/gga/tcc/ctc/cct	20.6	6.3	17.4	15.4	16.2
atc/cat/tca/tag/gta/agt	2.9	37.5	8.7	11.5	12.1
ccg/gcc/cgc/ggc/cgg/cgc	2.9	0.0	4.3	0.0	2.0
gac/cga/acg/ctg/gct/tgc	14.7	6.3	4.3	7.7	9.1

Motifs are distinguished for di- and tri-nucleotide based simple sequence repeats (SSRs) used to create new BMc markers.

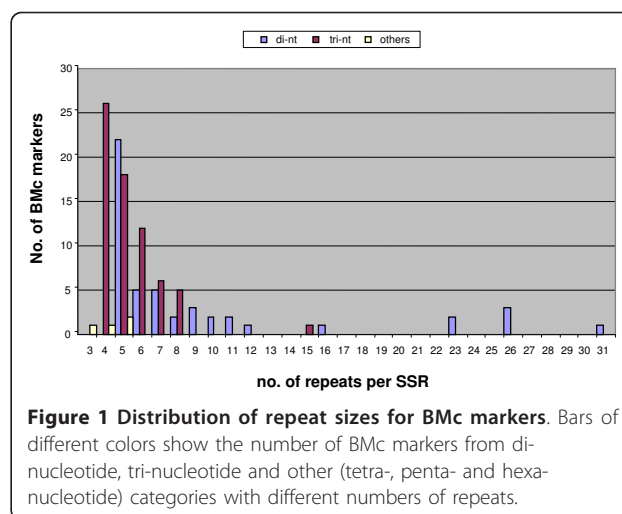
<sup>1</sup>Complementary sequences for a given motif are given and were the basis for grouping of di-nucleotide and tri-nucleotide motif SSR.

BMc59, BMc75, BMc76 and BMc78 to BMc108 as well as BMc110 to BMc120) were developed from the LP root cDNA libraries. In summary the largest number of new cDNA derived microsatellites were found in the root libraries (93 out of the 120) compared to the leaf library (27 out of the 120).

Among the newly developed markers 50 were based on di-nucleotide repeats, 66 on tri-nucleotide repeats and 4 on tetra-, penta- or hexa-nucleotide repeats which we generally avoided for primer design (Table 3). The new markers produced expected product sizes from as small as 80 to as large as 298, although the majority were designed to be small PCR amplicons to avoid the possibility of including exons. The average number of repeats in the BMc markers (including both compound and simple SSRs) was 6.8 repeats per microsatellite but this varied from an average of 9.1 for di-nucleotide motifs to 5.3 for tri-nucleotide motifs and 4.3 repeats for other tetra, penta or hexa-repeat based motifs.

The highest repeat numbers were found for BMc70 (31 repeats) and BMc58 (26 repeats) as well as BMc30 and BMc33 (23 repeats, each); all of which were based on di-nucleotide motifs; the first and last two based on GA<sub>n</sub> with the second based on CA<sub>n</sub>. Surprisingly, there were few long AT<sub>n</sub> microsatellites, with the exception of BMc3 (26 repeats), but this may be due to the genic nature of the microsatellites developed. The distribution of repeat sizes among the BMc markers was skewed generally to the smaller number of repeats; the reader is reminded that the minimum number of repeats for di-nucleotides was five and for tri-nucleotides was four while for all other types it was three (Figure 1). Interestingly, a small group of di-nucleotide microsatellites with large numbers of repeats were found to the right of the graph and greater skewing of di-nucleotide compared to tri-nucleotide microsatellites was found towards the left of the graph.

When comparing the source tissue for the BMc markers, the ratio of di-nucleotide and tri-nucleotide markers was similar for root and leaf derived microsatellites



**Figure 1** Distribution of repeat sizes for BMc markers. Bars of different colors show the number of BMc markers from di-nucleotide, tri-nucleotide and other (tetra-, penta- and hexa-nucleotide) categories with different numbers of repeats.

(Table 3). These ratios held true for the proportion of markers that had problems of non-amplification (16 out of 120) or that were multi-copy (6 out of 120). The markers showing multiple monomorphic banding were BMc30, BMc58, BMc60, BMc70, BMc92, and BMc96. The ratio of simple to compound SSRs was 102 to 18 among the new BMc markers, 85% and 15% of the total number of markers, respectively. Among the compound repeats many were just due to an interruption of the same repeat (7 out of 18). Therefore the percentage of truly compound repeats was even lower (11 out of 120) corresponding to 9.2% and the vast majority were simple, perfect motif SSRs. Amplification strength was similar for SSRs of different motifs and repeat lengths (Figure 2).

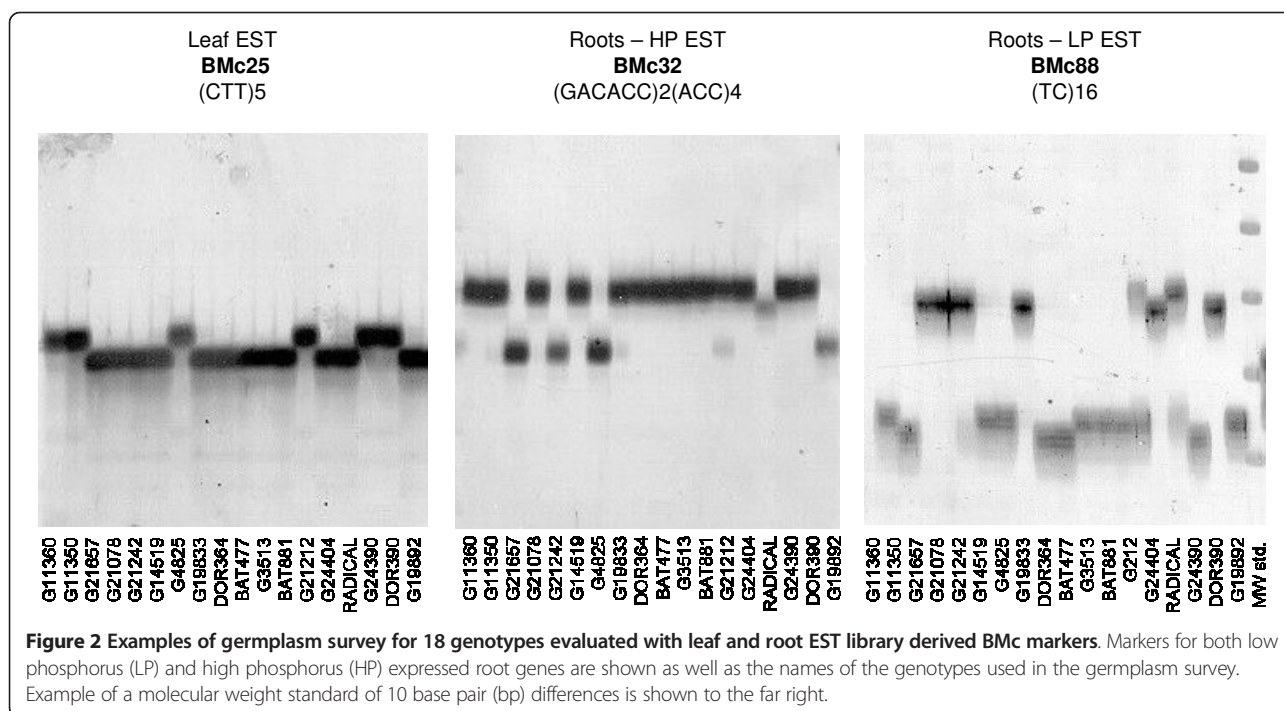
#### Genetic diversity detected

As described above, out of the 120 new BMc markers a total of 98 microsatellites amplified well in the survey panel and these were used for polymorphism survey for the germplasm panel and diversity analysis. In this final set of 98 functional markers, 59 (60.2%) were

**Table 3** Summary of the motif and polymorphism characteristics of microsatellites found in BMc markers

BMc marker types	Leaf EST source	Root EST source	Number of SSRs	Percentage of total
Di-nucleotide based	10	40	50	41.7
Tri-nucleotide based	16	50	66	55.0
Tetra, penta or hexa-nt based	1	3	4	3.3
Multi-copy	0	6	6	5.0
Non-Amplifying	6	10	16	13.3
Monomorphic in survey	12	47	59	49.2
Polymorphic in survey	9	30	39	32.5
Polymorphic in cultivated	7	27	34	28.3
Polymorphic in wild only	1	4	5	4.2

BMc markers developed from leaf and root expressed sequence tags (EST) are separately and jointly considered for number of simple sequences repeats (SSRs) and percentage of total.



monomorphic and 39 (39.8%) were polymorphic. The average PIC value of the new polymorphic BMC markers was 0.310 and ranged from 0.099 for the least polymorphic markers to 0.657 for the most polymorphic marker (BMc70).

Polymorphism comparison of the di-nucleotide and tri-nucleotide markers showed that they had similar average PIC values (0.131 and 0.125, respectively) when considering both monomorphic and polymorphic microsatellites together. A similar situation was observed when considering only polymorphic microsatellites, where di- and tri-nucleotide based markers again had similar PIC values (0.322 and 0.301, respectively). None of the tetra-, penta- or hexa-nucleotide repeat-based markers was polymorphic.

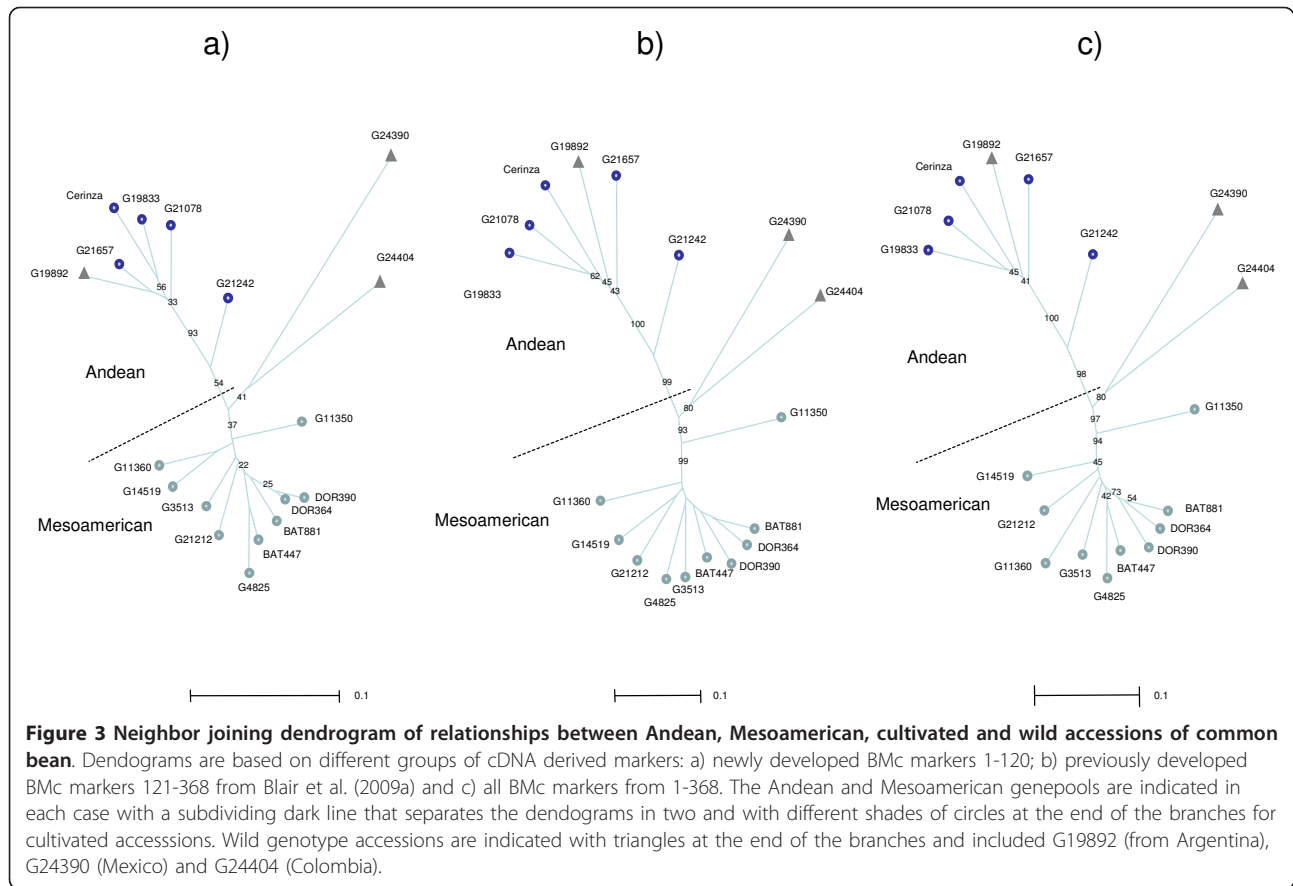
Polymorphic markers were in similar proportion (38% in each case) for the BMC markers from leaf ESTs (8 out of 21 functioning markers) and for the BMC markers from root ESTs (30 out of 77 functioning markers). Interestingly some polymorphic root-derived BMC markers (BMc30, BMC40, BMC58, BMC60 and BMC70) showed monomorphic background bands suggesting they were members of gene families with different degrees of diversity in different homologs.

A set of five microsatellites (BMc17, BMC36, BMC44, BMC61 and BMC68) was only polymorphic in the wild accessions but not in the cultivated accessions or varieties. These markers had relatively low PIC values of 0.099 to 0.157. From the 368 current BMC markers,

including the 248 from the previous study of Blair et al. [6] and the 120 described here, a total of 209 (56.8%) of the BMC markers yielded monomorphic results while 159 (43.2%) produced polymorphic results in the germplasm survey.

The average PIC value of the full set 368 BMC microsatellites was calculated to be 0.291, while for all those that were polymorphic the PIC value was 0.424. When the diversity analysis with the newly-developed, cDNA-derived markers BMC1 to BMC120 was undertaken (Figure 3a) we found that the Andean and Mesoamerican gene pools represented the main axis of the neighbor joining tree upon which two of the wild accessions then showed a divergence from the cultivated genotypes. The Argentinian accession G19892 grouped within the Andean gene-pool, while the highly diverse Colombian (G24404) and Mexican (G24390) accessions were near the division of the two gene pools. These results agree with the neighbor joining analysis of Blair et al. [6] who evaluated the markers BMC121 to BMC368 (Figure 3b).

When the results of the phylogeny analysis of the newly developed makers were combined with the previous markers from Blair et al. [6] an even clearer picture of the associations emerged (Figure 3c). Although all dendrograms showed very highly supported nodes for the separation of the two main gene pools and the two wild accessions; in the combined analysis, we found very high bootstrap values (ranging from 90 to 100%) based on the strength of the total set of markers evaluated.



## Discussion

The major achievements of this research were 1) to evaluate microsatellite frequency in three cDNA libraries from root and leaf tissues with one of the root libraries developed for the abiotic stress of low phosphorus and 2) to create additional genic microsatellite markers based on low-level sequencing of these EST libraries to use in a polymorphism survey both to understand common bean genetic diversity and to understand the differences in various microsatellite types from different sources and their ability to uncover bean diversity. The creation of new genic microsatellites is especially pressing as only about 230 [2,7,8,27,28] had been reported before we started our work on the design of BMC microsatellite markers.

In total we have now designed 368 genic microsatellites in the BMC series between the efforts of this study and the previous work of Blair et al. [6]; all BMC markers were designed from cDNA libraries made from different tissues of the mapping population parents used by Blair et al. [12,23]. In addition, with this study we have created BMC markers from two different genotypes including G19833 and DOR364 and from leaf tissue and root tissues subjected to low or high phosphorus conditions. The advantage of having markers developed from

sequences of both genotypes resides is the fact that the Andean G19833 is being used for whole-genome shotgun sequencing and the Mesoamerican DOR364 provides a commercially useful tropical, small red seeded counterpart to the Andean genotype and to black beans which have been better studied in terms of agronomy as well as EST development [17].

In addition, both marker types from both genotype sources are useful for evaluation in the reference map based on DOR364 × G19833 studied by Blair et al. [2,23] which is linked both the UC-Davis [29] and Univ. of Florida [30] genetic maps. In terms of the practical use of the microsatellites, the PCR amplification strength was similar for SSRs of different motifs and repeat lengths, which may be typical of gene-derived microsatellites and distinct from genomic microsatellites as first suggested by Blair et al. [22].

In our previous study of cDNA derived microsatellites [6] we found that uniformly strong PCR products were obtained with the specific primer sets around the SSR loci in cDNA sequences. In comparison, amplification with non-gene based microsatellites is prone to some pitfalls as discussed by Blair et al. [23] for AT-rich microsatellites and Blair et al. [31] for hybridization

derived genomic microsatellites. Differences between genic and different kinds of genomic microsatellites have been observed for other marker sets as well [7,32,33]. Although the SSR and EST sequencing effort from most of these projects has been small it is useful to have added their sequences to GenBank to compare in the future to larger EST collections from Ramirez et al. [17], Melotto et al. [18] and Thibivilliers et al. [19] as well as future genomic sequences for common bean or related species. Furthermore the possible role of microsatellites as promoters or gene expression enhancers especially in root genes where many AG<sub>n</sub> microsatellites were found could be studied.

In terms of other di-nucleotide motifs, the lack of GC microsatellites has been observed before within the bean genome [6,31], while AT-rich microsatellites were not expected to be found in genic sequences neither as di-nucleotides nor tri-nucleotides such as those studied by Blair et al. [23]. There were only a few AC<sub>n</sub> based microsatellites which was surprising given that enrichment for this motif has yielded about the same number of markers as enrichment with AG<sub>n</sub> or GA<sub>n</sub> based probes [7,34]. Among the tri-nucleotide motifs it appears that AAG (23), ACC (12), AGC (12), AGG (16) and ATC (12) microsatellites are the most common and this may have to do with their frequency in triplet codon use for amino acid incorporation into polypeptides. Additionally, open reading frames are known to have a higher GC percentage than non-translated regions [35] which might favor tri-nucleotide motifs such as ACC, AGC and AGG. Compared to the results of Blair et al. [6,31] the ratio of tri-nucleotide to di-nucleotide motifs was fairly high (99 *versus* 57 in total). Perhaps this was due to a majority being located in the open reading frame rather than in untranslated regions of the original mRNA transcripts represented by the cDNA sequences.

In the second step of this study, we analyzed the potential of two different groups of BMC markers, one from cDNA clone sequencing (120 BMC markers) and one from cDNA hybridization with SSR motifs (248 BMC markers developed from 497 positive cDNA clones) to be used in phylogeny analysis. The full group of markers, therefore, included a total of 368 BMC microsatellites all evaluated against the same germplasm survey from Blair et al. [6]. In that evaluation, genetic diversity was reliably predicted by both types of cDNA based BMC microsatellites. Both sets of markers were useful in separating the Andean and Mesoamerican gene pool and accurately placing the wild accessions within each gene pool. Two wild accessions (Colombian and Mexican) were separated from the cultivated accessions. Similar results were found with the same diversity panel in Blair et al. [6].

In summary, cDNA derived markers seem to be very useful for diversity analysis due to the fact that they are derived from genic sequences that are conserved and are highly transferable between different accessions of beans. They were critical in recent studies of diversity in both dry and snap bean cultivars of *Phaseolus vulgaris* [36,37]. Therefore, in the future we plan to analyze the frequency of gene-based microsatellites in larger collections of ESTs such as those of Ramirez et al. [17] or Thibivilliers et al. [19] which surpass the numbers of ESTs evaluated in the libraries we used here. It will be interesting to see if SSR frequency is similar or different for the multiple libraries used by the first of these authors or the larger set of ESTs from a single rust-infected leaf library evaluated by the second research group. One lesson from this microsatellite evaluation is that it is important to test new markers for consistent patterns of genetic diversity detection. We also plan to test the gene-derived markers in related *Phaseolus* species.

## Conclusions

In terms of the evaluation of genetic diversity we found that genic microsatellites from both EST sequencing and hybridization based approaches performed equally well in distinguishing Andean and Mesoamerican gene pools and the Argentinean, Colombian and Mexican wild beans as separate accessions. Therefore, these markers can be used for diversity analysis and for breeding especially in crosses between wild and cultivated beans or between gene pools. We expect that next generation sequencing will make the discovery of new transcriptome-based SSRs even easier than the two approaches used so far. Nonetheless, the utility of cDNA derived microsatellites for diversity analysis is well established and is perhaps best explained due to their conservation and slower rate of evolution than genomic microsatellites. In summary, gene-based or 'genic' microsatellites appear to be especially useful for genetic analysis of common bean and it would be ideal to have a larger set of these markers for functional diversity analysis and perhaps association mapping once they are genetically mapped which will be the subject of a separate manuscript to define the regions of the genome that are part of the transcriptome. Finally, these gene-based markers may be the keys to selection of specific traits as they represent expressed genes some of which are likely to have multiple functional alleles with diverse phenotypes as a result. Simple sequence repeats in promoter regions have sometimes been found to be important in controlling gene expression and this may be the case for some of the genic markers discovered in this study as well.



## Additional material

### Additional file 1: Supplementary Table S1. Primer sequences and simple sequence repeat motif for new set of cDNA-derived BMC (Bean microsatellite derived from cDNA sequence) series markers.

GenBank entry, predicted product size based on EST sequence and polymorphism information content (PIC) given for each marker.

## Acknowledgements

We are grateful to Agobardo Hoyos for germplasm curation and development. We also wish to thank the staff of CUGI that made the sequencing possible including Christopher Saski, Diane Cohen, Michael Atkins and Michael Palmer. Joe Tohme in CIAT and Dorrie Main in CUGI are acknowledged for advice. The funding from USAID-SLO linkage grants is gratefully recognized.

## Author details

<sup>1</sup>CIAT - International Center for Tropical Agriculture, Biotechnology Unit and Bean Project, AA6713, Cali, Valle, Colombia. <sup>2</sup>Clemson University Genomics Institute, Clemson, South Carolina, USA. <sup>3</sup>Department of Biology, Georgetown University, Washington DC, USA. <sup>4</sup>Department of Plant Pathology, Kansas State University, Manhattan, Kansas, USA. <sup>5</sup>Sun Seeds, Fargo ND, USA. <sup>6</sup>Arizona Genomics Institute, Tuscon, Arizona, USA.

## Authors' contributions

MWB conceived and organized the study and wrote the manuscript. NH and MCC and MCG performed the laboratory work for BMC marker evaluation. NH and MCC helped in writing the manuscript and preparing tables and figures. MCMT contributed to writing and designed the primers. FP and MCMT constructed, arrayed and screened all the libraries at CIAT and CUGI. JT and RW assisted with library preparations at CUGI. All authors read and approved of the manuscript.

Received: 26 November 2010 Accepted: 22 March 2011

Published: 22 March 2011

## References

1. Varshney RK, Graner A, Sorrells ME: **Genic microsatellite markers in plants: features and applications.** *Trends Biotech* 2005, **23**:48-55.
2. Blair MW, Pedraza F, Buendia HF, Gaitán-Solis E, Beebe SE, Gepts P, Tohme J: **Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.).** *Theor Appl Genet* 2003, **107**:1362-1374.
3. Hancock JM: **Microsatellites and other simple sequences: genomic context and mutation mechanisms.** Edited by: Goldstein DB and Schlotterer C. *Microsatellites: Evolution and Applications.* Oxford Univ. Press, New York; 1999:1-9.
4. Amos W: **A comparative approach to the study of microsatellite evolution.** Edited by: Goldstein DB and Schlotterer C. *Microsatellites: Evolution and Applications.* Oxford Univ. Press, New York; 1999:66-79.
5. Ellegren H: **Microsatellites: Simple sequences with complex evolution.** *Nature* 2004, **5**:435-445.
6. Blair MW, Muñoz-Torres M, Giraldo MC, Pedraza F: **Development and diversity assessment of Andean-derived, gene-based microsatellites for common bean (*Phaseolus vulgaris* L.).** *BMC Plant Bio* 2009, **9**:100.
7. Hanai LR, de Campos T, Camargo LEA, Benchimol LL, de Souza AP, Melotto M, Carbonell SAM, Chioratto AF, Consoli L, Formighieri EF, Siquiera MF, Tsai SM, Vieira MLC: **Development, characterization and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources.** *Genome* 2007, **50**:266-277.
8. Hanai LR, Santini L, Aranha LEC, Pelegrielli MHF, Gepts P, Tsai SM, Carneiro ML: **Extension of the core map of common bean with EST-SSR, RGA, AFLP, and putative functional markers.** *Mol Breeding* 2010, **25**:25-45.
9. Broughton WJ, Hernández G, Blair MW, Beebe SE, Gepts P, Vanderleyden J: **Beans (*Phaseolus* spp.) - Model Food Legumes.** *Plant Soil* 2003, **252**:55-128.
10. Rao IM: **Role of physiology in improving crop adaptation to abiotic stresses in the tropics: The case of common bean and tropical forages.** Edited by: Handbook of plant and crop physiology (Pessaraki M.). Marcel Dekker Inc, New York, USA; 2002:583-613.
11. Miklas PN, Kelly JD, Beebe SE, Blair MW: **Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding.** *Euphytica* 2006, **147**:105-131.
12. Blair MW, Munoz MC, Pedraza F, Gaitan E, Tohme J, Main D, Frisch D, Wing R: **Generation of expressed sequence tags (ESTs) from vegetative tissues of a common bean (*Phaseolus vulgaris*) mapping parent, G19833.** *GenBank* 2002, BQ481427-965.
13. Varshney RK, Thiel T, Stein N, Landridge P, Graner A: **In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species.** *Cell Mol Biol Lett* 2002, **7**:537-546.
14. Gao L, Tang J, Li H, Jia J: **Analysis of microsatellites in major crops assessed by computational and experimental approaches.** *Molecular Breeding* 2003, **12**:245-261.
15. Choumane W, Winter P, Baum M, Kahl G: **Conservation of microsatellite flanking sequences in different taxa of Leguminosae.** *Euphytica* 2004, **138**:239-245.
16. Kumpatla SP, Mukhopadhyay S: **Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species.** *Genome* 2005, **48**:985-998.
17. Ramírez M, Graham MA, Blanco-López L, Silvente S, Medrano-Soto A, Blair MW, Hernández G, Vance CP, Lara M: **Sequencing and analysis of common bean ESTs: Building a foundation for functional genomics.** *Plant Physiol* 2005, **137**:1211-1227.
18. Melotto M, Monteiro-Vitorello CB, Bruschi AG, Camargo LEA: **Comparative bioinformatic analysis of genes expressed in common bean (*Phaseolus vulgaris*) seedlings.** *Genome* 2005, **48**:562-570.
19. Thibivilliers S, Joshi T, Campbell KB, Scheffler B, Xu D, Cooper B, Nguyen HT, Stacey G: **Generation of *Phaseolus vulgaris* ESTs and investigation of their regulation upon *Uromyces appendiculatus* infection.** *BMC Plant Bio* 2009, **9**:46.
20. da Maia L, Palmieri D, Queiroz V, Marini M, Félix FA, Costa A: **SSRLocator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation.** *Int J Plant Genomics* 2008, 1-9.
21. Rozen S, Skaletsky HJ: **Primer3 on the WWW for general users and for biologist programmers.** Edited by: Krawetz S, & Misener S. *Bioinformatics Methods and Protocols: Methods in Molecular Biology,* New Jersey, U.S.A.: Humana Press, Ottawa, CA; 2000.
22. Blair MW, Giraldo MC, Buendia HF, Tovar E, Duque MC, Beebe S: **Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.).** *Theor Appl Genet* 2006, **113**:100-109.
23. Blair MW, Buendia HF, Giraldo MC, Metais I, Peltier D: **Characterization of AT-rich microsatellites in common bean (*Phaseolus vulgaris* L.)** *Theor Appl Genet* 2008, **118**:91-103.
24. Afanador L, Haley S, Kelly JD: **Adoption of a "mini-prep" DNA extraction method for RAPD's marker analysis in common bean.** *Phaseolus vulgaris Bean Imp Coop* 1993, **36**:10-11.
25. Perrier X, Flori A, Bonnot F: **Data analysis methods.** Edited by: Hamon P., Seguin M., Perrier X., Glaszmann J. C. *Genetic diversity of cultivated tropical plants.* Enfield, Science Publishers. Montpellier; 2003:43-76.
26. Liu K, Muse SV: **PowerMarker: an integrated analysis environment for genetic markers analysis.** *Bioinformatics* 2005, **21**:22128-2129.
27. Yu K, Park SJ, Poysa V: **Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*).** *Genome* 1999, **42**:27-34.
28. Yu K, Park SJ, Poysa V, Gepts P: **Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.).** *J Hered* 2000, **91**:429-434.
29. Freyre R, Skroch PW, Geffroy V, Adam-Blondon AF, Shirmohamadali A, Johnson WC, Llaca V, Nodari RO, Periera PA, Tsai SM, Tohme J, Dron M, Nienhuis J, Vallejos CE, Gepts P: **Towards an integrated linkage map of common bean. 4 Development of a core linkage map and alignment of RFLP maps.** *Theor Appl Genet* 1998, **97**:847-856.
30. Vallejos CE, Sakiyama NE, Chase CD: **A molecular marker based linkage map of *Phaseolus vulgaris* L.** *Genetics* 1992, **131**:733-740.
31. Blair MW, Muñoz M, Pedraza F, Giraldo MC, Buendia HF, Hurtado N: **Development of microsatellite markers for common bean (*Phaseolus vulgaris* L.) based on screening of non-enriched, small-insert genomic libraries.** *Genome* 2009, **52**:772-782.
32. Benchimol LL, de Campos T, Carbonell SAM, Colombo CA, Chioratto AF, Formighieri EF, Gouvêa LRL, de Souza AP: **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* 2007, **54**:1747-1762.
33. Campos T, Benchimol LL, Carbonell SAM, Chioratto AF, Formighieri EF, de Souza AP: **Microsatellites for genetic studies and breeding programs in common bean.** *Pes Agropec Bras* 2007, **42**:589-592.
  34. Gaitán-Solis E, Duque MC, Edwards KJ, Tohme J: **Microsatellite Repeats in Common Bean (*Phaseolus vulgaris*): Isolation, Characterization, and Cross-Species Amplification in *Phaseolus* ssp.** *Crop Sci* 2002, **42**:2128-2136.
  35. Li YC, Korol AB, Fahima T, Nevo E: **Microsatellites within genes: structure, function, and evolution.** *Mol Bio Evol* 2004, **21**:991-1007.
  36. Blair MW, Gonzales LF, Kimani P, Butare L: **Inter-genepool introgression, genetic diversity and nutritional quality of common bean (*Phaseolus vulgaris* L.) landraces from Central Africa.** *Theor Appl Genet* 2010, **121**:237-248.
  37. Blair MW, Chaves A, Tofiño A, Calderón JF, Palacio JD: **Extensive diversity and inter-genepool exchange of phaseolin alleles found in world-wide snap bean germplasm analyzed with AFLP and microsatellite markers.** *Theor Appl Genet* 2010, **120**:1381-1391.

doi:10.1186/1471-2229-11-50

**Cite this article as:** Blair et al.: Gene-based SSR markers for common bean (*Phaseolus vulgaris* L.) derived from root and leaf tissue ESTs: an integration of the BMC series. *BMC Plant Biology* 2011 **11**:50.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

