Plant Omics Journal POJ

POJ 15(01):17-24 (2022) ISSN:1836-3644
doi: 10.21475/POJ.15.01.22.p3580

Development and characterization of EST-SSR derived functional domain marker (FDM) in *Phaseolus vulgaris* (common bean)

Shatabdi Basu and Saurav Bhattacharya*

Department of Biotechnology, Techno India University West Bengal-700091

***Corresponding author: https://orcid.org/0000-0002-3715-9803**

Abstract

EST-SSR markers have been increasingly useful in crop improvement studies owing to their rapid and less expensive development, better chances of association with protein-coding regions and cross-taxon transferability. In the present study, SSR sequences were mined from 117,596 EST sequences of *Phaseolus vulgaris* (common bean) (from NCBI) representing expressed cDNA data from different plant tissues (leaves, roots) grown under variable growth conditions. A total 1481, perfect non redundant SSR containing contig sequences were assembled that showed highest percentages of mononucleotide (44.8%) followed by trinucleotide (30.4%) di-nucleotide (14.0%), tetra-nucleotide (7.0%) and hexanucleotide (3.6%) repeat sequences. Among the distribution pattern, trinucleotide SSR sequences (65.4%) were found to be relatively frequent in the coding regions, the mononucleotide (25.6%) and dinucleotide (38.4%) SSRs were found to be highly frequent in the UTR region. Assembled SSR sequences were assigned to 2700 unique functional domain markers using the InterproScan program, the most significant were heat shock protein DnaJ, a cysteine-rich domain, RNA binding domain, S-adenosyl methionine decarboxylase, Brassinosteroid (BES1/BZR1) plant transcription factor, B3 DNA binding domain. Functional annotation results of selected 949 blasted EST-SSR sequences using BLAST2GO revealed its association with several important molecular functions (MF), biological processes (BP) and cellular components (CC) in *P. vulgaris*. Our results also showed a relatively higher density of SSR (1SSR/0.2kb) that were used to design 146 SSR primer pairs from 280 annotated sequences (52.1%) using Primer3 and successfully amplified in silico using FastPCR. The present findings report a new array of SSR markers with significant association with diverse functional domains that could be used in future studies on genetic diversity of common bean cultivars as well as in marker-assisted breeding.

Keywords: EST-SSR marker, Functional domain marker, common bean.
Abbreviations: Expressed Sequence Tag (EST), Simple Sequence Repeat (SSR), Functional Domain Marker (FDM), Untranslated Region (UTR), Gene ontology (GO)

Introduction

Phaseolus vulgaris (common bean), is an important low-cost edible grain legume and an important source of staple food in several countries, particularly in Asia, Africa and Latin America. The plant is one of the major cultivated species with production estimated to be 28.9 million Metric tons globally as a dry grain in 2019 (http://www.fao.org/faostat/). The plant is an important source of protein, complex carbohydrates, dietary fibre, water-soluble vitamins and minerals (Celmeli et al., 2018). Significant advancements in traditional as well as molecular breeding of common beans have been made for both simple and complex traits leading to the marker-assisted selection of elite genotypes (Kelly et al., 2003; Miklas and Singh 2007; Xu and Crouch 2008).

Simple sequence repeat markers (SSRs, or microsatellite markers) are considered to be a versatile tool for mapping, breeding and genetic diversity and structure analysis because of their reproducible, highly polymorphic with high mutation rate, codominance and multi-allelic nature (Gupta and Varshney, 2000; Varshney et al., 2005). These markers are present in both protein-coding and non-coding regions of the genome with an extensive genome coverage (Tautz and Renz, 1984; Vieira et al., 2016). Among the legumes, microsatellite markers have been reported in Soybean (Bisen et al., 2015; JunTae-Hwan et al., 2011), Vigna (Somta et al., 2008; Tangphatsornruang et al., 2009), Lentil (Andeden et al., 2015), Medicago (Mun et al., 2006), Chickpea and common bean with reports of both inter-specific as well as intraspecific transferability. However, traditional SSR marker development faces several constraints such as the need for sequence information, time and labour-intensive method of development with sensitive methods of detection and analysis that has made de novo development of SSR marker a challenging task (Iniguez-Luy et al., 2008; Squirrell et al., 2003). Moreover, in several reports amplified locus were found to be species-specific and less useful in inter-taxon or larger groups (Ellis and Burke, 2007). One important low-cost alternative is

to exploit publicly available genomic resources and sequence data for searching, identification and characterization of deposited SSR sequences from the transcribed genomic regions (Hu et al., 2011). EST markers have better transferability across taxa than genomic SSR markers, better association and physical linkage with expressed genes as they are designed from coding regions and their ease of analysis and represent functional domain markers that are key regions for efficient marker-assisted selection (Carson and Botha, 2000; Saha et al., 2004; Varshney et al., 2005). Such expressed sequences (ESTs) derived from the entire expressed cDNA pool represent robust functionally annotated marker sequences with predicted protein domain signatures. Hence these DNA markers derived from functionally defined and validated sequences, have better chances of association with polymorphic traits so that can be successfully employed in molecular breeding approaches (Thiel et al., 2003; Yu et al., 2010). A limited number of studies have explored the potential of genic microsatellite data (EST) based on marker analysis in the *Phaseolus vulgaris* (Blair and Hurtado, 2013; Blair et al., 2011; Blair et al., 2009; Garcia et al., 2011; Hanai et al., 2010). With the availability of an increasing number of sequence data in the public domain, and considering the high transferability of SSR markers across species including common bean (Yu et al., 1999) we aimed to develop functionally utilizable polymorphic SSR markers in the present study. Hence the present work attempted to explore the pooled EST sequences of common bean to achieve the following two broad objectives: (1) Development of novel simple sequence repeat-functional domain markers (FDM) using pooled EST sequences of *Phaseolus vulgaris* and (2) Characterization of microsatellite (SSR) sequence in the expressed sequences of common bean.

Results and Discussions

Processing and assembly and SSR detection

In the present study, a total of 117,596 ESTs (retrieved in Oct 2020) from *P. vulgaris* representing different plant parts (leaf, stem, and roots) were retrieved and assembled. The proportion of assembled EST contigs was 68.5%, suggesting that the majority of ESTs shared sequences in common with other ESTs (Table 1). Table 1 summarizes details on the quantity of SSR discovered from assembled EST sequences and associated metadata. Following assembly, a non-redundant group of ESTs were generated consisting of contigs and singletons, hereafter referred to as "assembled EST sequences." SSR detection by MISA and Krait tool revealed a total of 1,481 numbers of perfect SSRs of which 10 were in compound formation and 820 were VNTRs. In-depth analysis revealed that the combined SSR frequency was 8.63% across all assembled EST sequences. We detected that mononucleotide repeats (664, 44.83%) were the most abundant followed by trinucleotide (450, 30.38%), di-nucleotide (208, 14.04%), tetra-nucleotide (105, 7.09%) and hexanucleotide (54, 3.65%) (Table 2). In addition, the majority of SSR repeats were detected in the contig sequence (83.3%) compared to singleton (16.7%). Similarly, among the length distribution of SSR repeat types, mononucleotide repeats (9527, 41.9%) were found to be the highest followed by trinucleotide (6996, 30.7%) and dinucleotide (3188. 14%) (Fig 1). The most abundant repeat motif was (AG)n followed by (AT)n and (AC)n in di-nucleotide, and (AAG)n in trinucleotide followed by (ACC)n, (AGG)n, (ATG)n, (CCG)n, and (ATC)n (Fig 2).

Functional domain marker (FDM) analysis of SSR-ESTs

A total of 1381 SSR-containing sequences were selected for FDM analysis excluding less abundant mononucleotide repeats. The Interpro scan result assigned 2700 unique SSR-FDM to 1381 contigs. The domain was analyzed from Interpro member databases including SignalPHMM, TMHMM, HMMPanther, Pfam, SMART, Panther and Gene3D. Among the important functional domains identified were heat shock protein DnaJ, a cysteine-rich domain, RNA binding domain, S-adenosyl methionine decarboxylase, Brassinosteroid (BES1/BZR1) plant transcription factor, B3 DNA binding domain.

Gene ontology analysis

GO terms were assigned to EST-SSRS with significant matches. The function of 949 SSR-containing sequences was annotated against the non-redundant (nr) protein database performed using the Blast2GO module. Annotations were recorded for a total of 280 (29.5%) sequences. The molecular function (MF) refers to the product function of the gene at the molecular level and includes catalytic and binding activities of a gene. The main important proteins with molecular functions include DNA binding transcription factor activity (124 EST SSR, 13%), ATP binding (10%), metal ion binding (9%), zinc ion binding (8), mRNA binding (7%), methyl transferase (3%), and GTP binding (1%) (Fig 3).

A biological process (BP) is a sequence of actions carried out by one or more organized assemblies of molecular function. Annotated EST-SSRs with most abundant biological processes include transcription-regulation (19%), followed by oxidation-reduction homeostasis (13%), phosphorylation (6%), photosynthesis (5%), and stress response (5%). Information on other biological corresponding to EST-SSRs was shown in Fig 3.

The remaining GO category is the cellular component (CC) that describes subcellular structures and macromolecular complexes. GO-CC terms may thus be used to annotate cellular locations of gene products. The most abundant cellular process from annotated SSR-ESTs in our results was shown to be associated with plant Cell walls (15%). Plasma membrane (13%) followed by Nucleolus (10%), Nucleus (9%) and Chloroplast (8%). The remaining EST-SSRs with their cellular components have been shown in Fig 3.

Results from the ORF finder showed a relatively higher distribution of SSR sequences within the UTR region (859, 58%) than the coding region (ORF) (622. 42%) indicating its possible roles as binding sites of proteins and regulatory elements such as transcription factors. While the trinucleotide SSRS (65.4%) were found to be relatively frequent in the coding regions, the mononucleotide (25.6%) and dinucleotide (38.4%) SSRs were found to be highly frequent in the UTR region. Similar observations were reported in other plants such as Prunus (Sorkheh et al., 2016) and Ocimum (Gupta et al., 2010). The distribution of SSR sequences was found to vary in several plant genomes including model plants like rice and Arabidopsis and highly dense particularly in the UTR regions of the genome (Lawson and Zhang, 2006).

Table 1. Summary of perfect microsatellite repeats in the assembled contig EST sequence of *P. vulgaris.*

Item	Description	Number
Total number of perfect SSRs	Counts	1481
Total length of perfect SSRs	Bp	22759
The average length of SSRs	total ssr length/total ssr counts (bp)	15.37
The percetage of sequence covered by SSRs	total ssr length/total sequence length (%)	8.64
Relative abundace	total SSRs/total valid length (loci/Mb)	5618.66
Relative density	total SSR length/total valid length (bp/Mb)	86343.74

Fig 1. Distribution of different SSR repeat types in terms of their length and count in *P. vulgaris* assembled EST contig sequence.

Table2. Table showing the counts, length and abundance of microsatellite repeats types in the assembled contig EST sequence of *P. vulgaris.*

Repeat Type	Counts	Length (bp)	Percent (%)	Average Length (bp)	Relative Abundance (loci/Mb)	Relative Density (bp/Mb)
Mono	664	9527	44.83	14.35	2519.1	36143.8
Di	208	3188	14.04	15.33	789.12	12094.72
Tri	450	6996	30.38	15.55	1707.22	26541.62
Tetra	105	1680	7.09	16	398.35	6373.63
Hexa	54	1368	3.65	25.33	204.87	5189.96

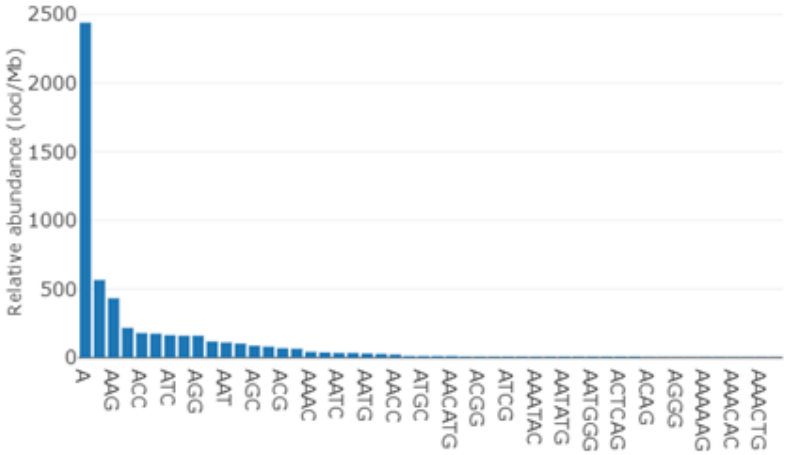

Fig 2. The most abundant SSR motif categories found in assembled EST-contigsequences of *P. vulgaris.*

Table 3. List of primer sequences, melting temperature (Tm) and amplicon sizes used for successful amplification of most abundant SSR motif in *P.vulgaris* EST-contig sequence in the FastPCR programme.

SL. No.	Contig number	Primer sequence		Tm (^{o}C)	Amplicon size (b.p)	SSR motif
1.	SSR2 Contig 2121-35	F	GCCCTCCGCAAAAGCATC	59	161	TCC
		R	GATCGCCGGGGAGTTTGC	61		
2.	SSR15 Contig152150-164	F	AAGTATGGAAACTGGGGCTGT	59	80	AAC
		R	TGGAAGTTGCTCTGGCTTCA	59		
3.	SSR17 Contig224240-257	F	CGGTGAAACCGGGATTGAAG	59	169	GGA
		R	ACCAACAGATAGCCTTGGCG	60		
4.	SSR20 Contig2961133-1150	F	GTGGGTCTTCAAGCTCGTCTT	60	193	TC
		R	AGACTCTTACTGACGCGGAA	58		
5.	SSR27 Contig34117-30	F	GAGGCGCAAAAATGGTGAAG	50	94	AG
		R	GTTCGCACTATCACCGCCT	57		
6.	SSR31 Contig43024-41	F	TTTCGCGGGGAACACTATCC	60	127	TTC
		R	CTCCAACAGTGTAGAGGCCC	59		
7.	SSR36 Contig499451-468	F	CCCACTGAAAACGGTCCTCT	59	197	AT
		R	GGCTGCAACCCTACTAGCTC	60		
8.	SSR38 Contig5961627-1644	F	GTAGCCTTGTGAGTGGCGAT	60	182	TG
		R	TAGTTCAAGGCCGCACCGA	61		
9.	SSR42 Contig6291219-1234	F	TTACAGCAATGGTGGTGTGGT	60	178	GTGA
		R	ACTCACATGACAATCTCTGGCT	59		
10.	SSR45 Contig69291-108	F	TTGGGCGGGGAGAAAGC	59	212	TGG
		R	ACTGGGTTTGGCTCAGTCAC	60		
11.	SSR54 Contig776288-302	F	GCTTGGCAGAGAAACCGTCA	60	146	TCT
		R	CTTGCAACCAGGGAGACCAA	60		
12.	SSR65 Contig1039243-257motifGCA		CAACACCAAGCAATCTGCGG	60	106	GCA
			CAAATTGGCGCTGATACGCA	59		
13.	SSR66 Contig1059332-349	F	CCGGTGGACCAATTGGCAT	60	198	CAC
		R	TGCCCCACTTCCTTTACTGA	58		
14.	SSR73 Contig1157323-340	F	GTTCACTCCGCCAATCCCTT	60	153	ACC
		R	GCTACGAGAGTGTTTCCCGT	59		
15.	SSR85 Contig1360856-873	F	AAGTGAGGAAGTTGAACGCAGA	60	192	CCG
		R	AGGTGGCAAAACAGTGGCTC	61		

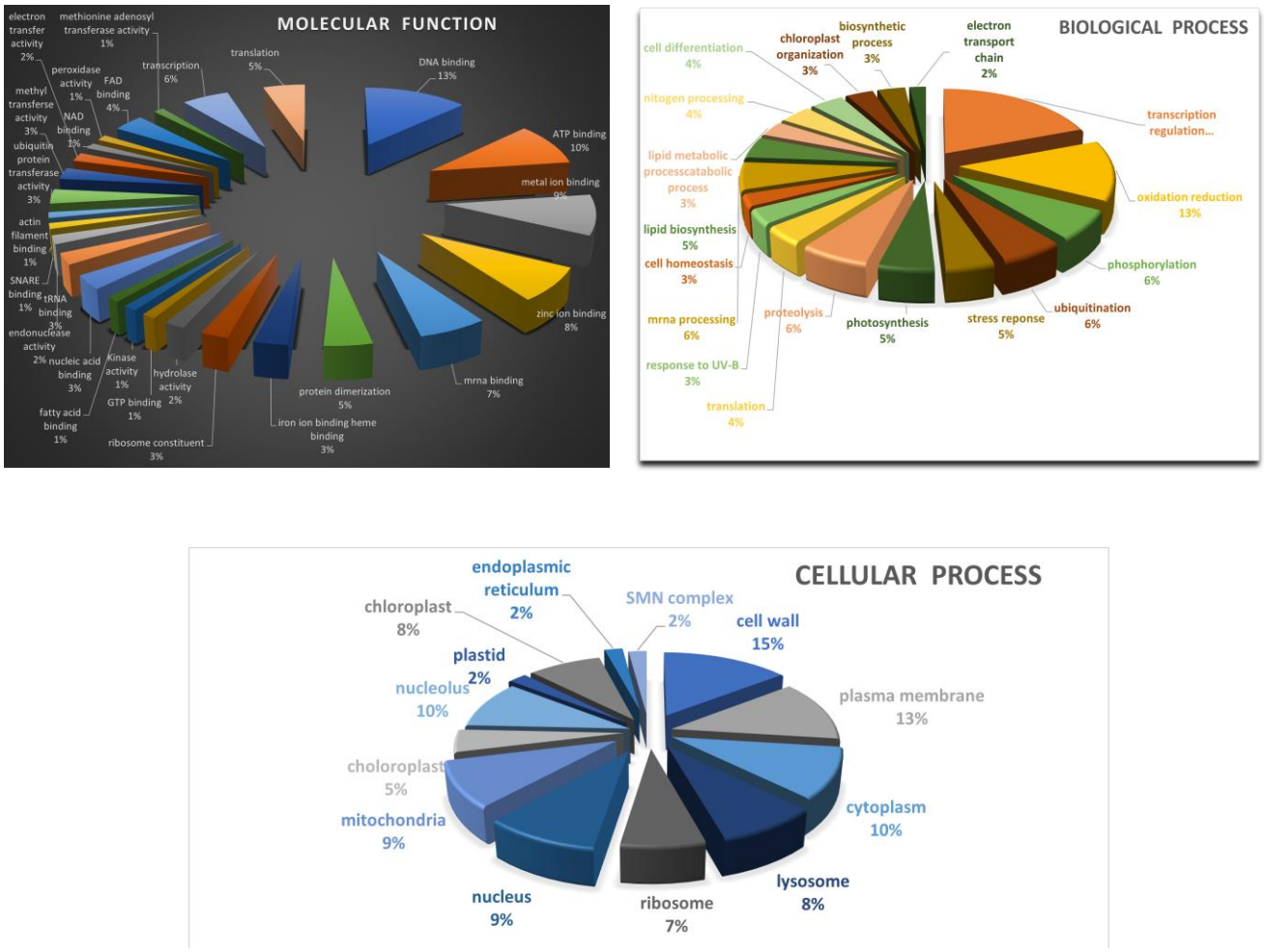

Fig 3. GO annotation results of SSR-contigs showing assigned annotated sequences assigned to various attributes of Molecular functions (MF), Biological Processes (BP) and Cellular Processes (CC) obtained using the Blast2Go program.

SSR primer design and in silico PCR

Out of a total of 280 annotated sequences, it was possible to design primers for 146 SSR-containing contig sequences (52.14%), whereas proper reproducible primer pairs could not be generated for the remaining 134 SSR (47.8%) sequences. Out of a total of 460 primer pairs designed for all the sequences, 52 primer pairs were successfully validated using *insilico* PCR using SSR containing contig sequences. The average amplicon SSR allele size ranged between 50– 500 bp, with an average T_m of 60ºC. The average length of forward and reverse primer was kept at 20bp and 21 bp respectively. The selected list of the primers with fragment sizes generated after successful amplification was given in Table 3. Using the default settings and standard genetic code, the open reading frames (ORFs) for all the SSR-containing sequences with FDM markers were predicted using the ORF Finder tool available at NCBI. As the principal encoding segment, only sequences that reach the maximum length without a stop codon were chosen (ORF). The relative location of SSRs, i.e., whether the SSR was present within the ORF, in the 5′ or 3′ untranslated regions, was identified in all predicted ORFs (UTR).

Common bean (*P. vulgaris* L.) is considered to be one of the most important grain legumes with high nutritional value and huge germplasm diversity, especially in the North Western parts of India (Choudhary et al., 2018; Mir et al., 2021).

Molecular markers from both genic as well as non-genic regions have been employed in breeding programs of *P. vulgaris* including microsatellite and EST sequences (Blair et al., 2011; Šajgalík et al., 2019). Genic EST sequences are considered a valuable resource to screen DNA markers that are largely linked with protein-coding regions of the genome and serve as a potential source to identify candidate genes. The frequency of distributed SSR in our assembled contig sequence was found to be 8.64%, which indicated a relatively high abundance of SSRs in the available *P. vulgaris* EST sequence. This is in congruence with other reports where high SSR distribution across the EST sequence has been reported in some other plant species such as in Euphorbia (11.77%) (Sen et al., 2017; Sen et al., 2018), Prunus (8.32%) (Sorkheh et al., 2016) and Mentha (8.4%) (Kumar et al., 2015). However, the frequency of SSR distribution in our findings was lower compared to tea (15.5%) (Ma et al., 2010; Sahu et al., 2012) and castor bean (28.4%) (Qiu et al., 2010). Nevertheless, several other reports have shown a relatively low abundance of SSR motifs, especially in important crop plants including rice (4.7%), barley (3.4%), maize (1.4%), sorghum (3.6%), and

wheat (3.6%) (Kantety et al., 2002). Additionally, considering the small genome size of the plant,
the density of SSR in our study was found to be 1 SSR per 0.2 kb which was relatively higher compared to soybean (1/7.4kb), maize (1/8.1kb), cotton (1/20kb), tomato (1/11.1 kb).

Materials and Methods

Retrieval, processing and assembly of EST sequence

All deposited raw EST sequences of *P.vulgaris* were retrieved from National Centre for Biotechnology Information (NCBI) and downloaded in Fasta format. A total of 128,868 EST sequences were downloaded (October 2020) representing expressed cDNA data from different plant tissues (leaves, roots) grown under variable growth conditions. All steps of EST pre-processing, clustering and assembling were done using the online web server EGassembler (https://www.genome.jp/tools/egassembler/) (Masoudi-Nejad et al., 2006). The available ESTs were trimmed at 5' end or 3' end for any poly A or poly T stretches and further cleaned for vector and adaptor sequence contamination, low complexity filtering, and other contamination. The cleaned quality sequences were subjected to contig assembling with default parameters with standalone processing using 6 CPUs for data analysis. The final contig output file was downloaded and saved in FASTA format and further utilized for SSR mining.

SSR mining and identification of functional domain marker

For SSR identification, assembled contig sequences were only used and mined for SSR-containing regions using KRAIT Program (Du et al., 2017) and cross-checked using the MISA tool (MIcro SAtellite identification tool; http://pgrc.ipkgatersleben.de/misa/misa.html) (Beier et al., 2017) with default search settings. SSR containing contig sequences were further analyzed for functional domain markers (FDMs) using the InterProScan tool in Blast2GO program (Götz et al., 2008). InterProScan provides the platform to analyze functional domains with the help of member databases, such as BlastProDom, FPrintScan, HMMPIR, HMMPfam, HMMSmart, HMMTigr, ProfileScan, HAMAP, PatternScan, SuperFamily, SignalPHMM, TMHMM, HMMPanther, and Gene3D. Additionally, EST-SSR sequences were searched for significant matches against the non-redundant protein database using BLASTx (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and further mapped to obtain associated GO hits. GO hits from Interpro scan and BLASTx were further pooled to obtain functional GO annotations and GO terms depicting three descriptors i.e., biological process (BP), cellular component (CC) and molecular function (MF) were assigned to them. All the analyses were performed using the Blast2GO program. SSR contigs were annotated and mapped against the KEGG database to retrieve the enzyme commission (EC) IDs.

SSR primer designing, ORF prediction and in silico amplification

Primers were designed for SSR containing contig sequences using Primer3 software with default parameters: optimum primer size = 20.0 (range of 18–27), optimum annealing temperature = 60.0 (range of 57.0–63.0), GC content of 20–80 %. Open reading frames (ORFs) were predicted for all SSR-containing sequences using the ORF Finder available at NCBI using standard genetic code. The relative position of the SSR motif i.e., whether the SSR was present within the ORF, in the 5' or 3'-untranslated region (UTR) was also recorded. The screened primers were then crosschecked for *in silico* amplification by FastPCR (Kalendar et al., 2017) version 6.5.

Conclusion

The development and characterization of molecular markers especially from the coding regions of the genome have immense potential for genetic improvement and diversity studies. The new EST-SSRs identified in the present study revealed several unique marker sequences associated with functional domains that could enrich the EST resources of the common bean. The derived EST-SSR markers in the present study can be used in intraspecific and interspecific genetic studies as well as in related species where limited genetic information is available. These findings can also be utilized in the development of polymorphic SSR markers linked to agronomically important genes/QTL and their further studies through molecular breeding.

Acknowledgement

The authors acknowledge Techno India University, West Bengal for its support and motivation during the research work.

Declaration

Authors declare that no part of the submitted work has been published in any other journal or is under consideration for publication or accepted for publication in any other journal. The authors also declare no conflict of interest in the present submission.

References

Andeden EE, Baloch FS, Çakır E, Toklu F, Özkan H (2015) Development, characterization and mapping of microsatellite markers for lentil (Lens culinaris Medik.). Plant Breed. 134(5):589-598.

Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics. 33(16):2583-2585.

Bisen A, Khare D, Nair P, Tripathi N (2015) SSR analysis of 38 genotypes of soybean (Glycine Max (L.) Merr.) genetic diversity in India. Physiol Mol Biol Plants 21(1):109-115.

Blair MW, Hurtado N (2013) EST-SSR markers from five sequenced cDNA libraries of common bean (Phaseolus vulgaris L.) comparing three bioinformatic algorithms. Mol Ecol Resour. 13(4):688-695.

Blair MW, Hurtado N, Chavarro CM, Muñoz-Torres MC, Giraldo MC, Pedraza F, Tomkins J, Wing R (2011) 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 Biol. 11:50-50.

Blair MW, Torres MM, Giraldo MC, Pedraza F (2009) Development and diversity of Andean-derived, gene-based microsatellites for common bean (Phaseolus vulgaris L.). BMC Plant Biol. 9:100.

Carson DL, Botha FC (2000) Preliminary Analysis of Expressed Sequence Tags for Sugarcane. Crop Sci. 40(6):1769-1779.

Celmeli T, Sari H, Canci H, Sari D, Adak A, Eker T, Toker C (2018) The Nutritional Content of Common Bean (Phaseolus vulgaris L.) Landraces in Comparison to Modern Varieties. Agronomy. 8(9):166

Choudhary N, Hamid A, Singh B, Khandy I, Sofi PA, Bhat MA, Mir RR (2018) Insight into the origin of common bean (Phaseolus vulgaris L.) grown in the state of Jammu and Kashmir of north-western Himalayas. Genet Res Crop Evol. 65(3):963-977.

Du L, Zhang C, Liu Q, Zhang X, Yue B (2017) Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. Bioinformatics. 34(4):681-683.

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99(2):125-132.

Garcia RAV, Rangel PN, Brondani C, Martins WS, Melo LC, Carneiro MS, Borba TCO, Brondani RPV (2011) The characterization of a new set of EST-derived simple sequence repeat (SSR) markers as a resource for the genetic analysis of Phaseolus vulgaris. BMC Genet. 12(1):41.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36(10):3420-3435.

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica. 113(3):163-185.

Gupta S, Shukla R, Roy S, Sen N, Sharma A (2010) In silico SSR and FDM analysis through EST sequences in Ocimum basilicum. Plant Omics J. 3(4):121-128

Hanai LR, Santini L, Camargo LE, Fungaro MH, Gepts P, Tsai SM, Vieira ML (2010) Extension of the core map of common bean with EST-SSR, RGA, AFLP, and putative functional markers. Mol Breeding. 25(1):25-45.

Hu J, Wang L, Li J (2011) Comparison of genomic SSR and EST-SSR markers for estimating genetic diversity in cucumber. Biol Plant. 55(3):577-580.

Iniguez-Luy FL, Voort AV, Osborn TC (2008) Development of a set of public SSR markers derived from genomic sequence of a rapid cycling Brassica oleracea L. genotype. Theoret Appl Genet. 117(6):977-85.

JunTae-Hwan, P. M, Rouf MA, GustafsonPerry (2011) Development of soybean aphid genomic SSR markers using next generation sequencing. Genome. 54(5):360-367.

Kalendar R, Khassenov B, Ramankulov Y, Samuilova O, Ivanov KI (2017) FastPCR: An in silico tool for fast primer and probe design and advanced sequence analysis. Genomics. 109(3):312-319.

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol. 48(5):501-510.

Kelly JD, Gepts P, Miklas PN, Coyne DP (2003) Tagging and mapping of genes and QTL and molecular marker-assisted selection for traits of economic importance in bean and cowpea. Field Crops Res. 82(2):135-154.

Kumar B, Kumar U, Yadav HK (2015) Identification of EST–SSRs and molecular diversity analysis in Mentha piperita. Crop J. 3(4):335-342.

Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the Arabidopsis thaliana and rice genomes. Genome Biol. 7(2):R14.

Ma J-Q, Zhou Y-H, Ma C-L, Yao M-Z, Jin J-Q, Wang X-C, Chen L (2010) Identification and characterization of 74 novel polymorphic EST-SSR markers in the tea plant, Camellia sinensis (Theaceae). Am J Bot. 97(12):153-156.

Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T, Goto S (2006) EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Res. 34(Web Server issue):W459-62.

Miklas PN, Singh SP (2007) Common Bean. In: Kole C (ed) Pulses, Sugar and Tuber Crops. Springer Berlin Heidelberg, Berlin, Heidelberg, p 1-31.

Mir RR, Choudhary N, Bawa V, Jan S, Singh B, Bhat MA, Paliwal R, Kumar A, Chitikineni A, Thudi M, Varshney RK (2021) Allelic Diversity, Structural Analysis, and Genome-Wide Association Study (GWAS) for Yield and Related Traits Using Unexplored Common Bean (Phaseolus vulgaris L.) Germplasm From Western Himalayas. Front Genet. 11:609603.

Mun J-H, Kim D-J, Choi H-K, Gish J, Debellé F, Mudge J, Denny R, Endré G, Saurat O, Dudez A-M, Kiss GB, Roe B, Young ND, Cook DR (2006) Distribution of microsatellites in the genome of Medicago truncatula: a resource of genetic markers that integrate genetic and physical maps. Genetics. 172(4):2541-2555.

Qiu L, Yang C, Tian B, Yang J-B, Liu A (2010) Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (Ricinus communisL.). BMC Plant Biol. 10(1):278.

Saha MC, Mian MA, Eujayl I, Zwonitzer JC, Wang L, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet. 109(4):783-91.

Sahu J, Sarmah R, Dehury B, Sarma K, Sahoo S, Sahu M, Barooah M, Modi MK, Sen P (2012) Mining for SSRs and FDMs from expressed sequence tags of Camellia sinensis. Bioinformation. 8(6):260-266.

Šajgalík M, Ondreičková K, Hauptvogel P, Mihálik D, Glasa M, Kraic J (2019) Higher effectiveness of new common bean (Phaseolus vulgaris L.) germplasm acquisition by collecting expeditions associated with molecular analyses. Sustainability. 11(19):5270.

Sen S, Dehury B, Sahu J, Rathi S, Yadav RNS (2017) In silico mining and characterization of simple sequence repeats (SSRs) from Euphorbia esula expressed sequence tags (ESTs): A potential crop for biofuel. Plant Omics J. 10(2):53-63.

Sen S, Dehury B, Sahu J, Rathi S, Yadav RNS (2018) Mining and comparative survey of EST-SSR markers among members of Euphorbiaceae family. Mol Biol Rep. 45(4):453-468.

Somta P, Musch W, Kongsamai B, Chanprame S, Nakasathien S, Toojinda T, Sorajjapinun W, Seehalak W, Tragoonrung S, Srinives P (2008) New microsatellite markers isolated from mungbean (Vigna radiata (L.) Wilczek). Mol Ecol Resour. 8(5):1155-7.

Sorkheh K, Prudencio AS, Ghebinejad A, Dehkordi MK, Erogul D, Rubio M, Martínez-Gómez P (2016) In silico search, characterization and validation of new EST-SSR markers in the genus Prunus. BMC Res Notes. 9(1):336.

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol. 12(6):1339-48.

Tangphatsornruang S, Somta P, Uthaipaisanwong P, Chanprasert J, Sangsrakru D, Seehalak W, Sommanas W, Tragoonrung S, Srinives P (2009) Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). BMC Plant Biol. 9(1):137.

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. 12(10):4127-4138.

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet. 106(3):411-22.

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends in Biotechnol. 23(1):48-55.

Vieira MLC, Santini L, Diniz AL, Munhoz CdF (2016) Microsatellite markers: what they mean and why they are so useful. Genet Mol Biol. 39(3):312-328.

Xu Y, Crouch JH (2008) Marker-Assisted Selection in Plant Breeding: From Publications to Practice. Crop Sci. 48(2):391-407.

Yu J-K, Paik H, Choi JP, Han J-H, Choe J-K, Hur C-G (2010) Functional Domain Marker (FDM): an In Silico Demonstration in Solanaceae Using Simple Sequence Repeats (SSRs). Plant Mol Biol Rep. 28(2):352-356.

Yu K, Park SJ, Poysa V (1999) Abundance and variation of microsatellite DNA sequences in beans (Phaseolus andVigna). Genome. 42(1):27-34.