Article

Development and Molecular Characterization of 55 Novel Polymorphic cDNA-SSR Markers in Faba Bean (\textit{Vicia faba} L.) Using 454 Pyrosequencing

Sundan Suresh, Jong-Hyun Park, Gyu-Taek Cho, Ho-Sun Lee, Hyung-Jin Baek, Sok-Young Lee and Jong-Wook Chung *

National Agrobiodiversity Center, National Academy of Agricultural Science, Rural Development Administration, Suwon 441-853, Korea

* Author to whom correspondence should be addressed; E-Mail: jwchung73@korea.kr; Tel.: +82-31-299-1880; Fax: +82-31-294-6029.

Received: 4 January 2013; in revised form: 17 January 2013 / Accepted: 23 January 2013 / Published: 31 January 2013

Abstract: Faba bean (\textit{Vicia faba} L.) is a major food source and fodder legume, popularly known for its high content of seed-protein. Its role is critical in crop rotation, and for fixing nitrogen effectively. Polymorphic simple sequence repeat markers from transcript sequences (cDNA; simple sequence repeat [SSR]) were developed for faba bean (\textit{Vicia faba}). We found that 1,729 SSR loci from 81,333 individual sequence reads and 240 primer pairs were designed and synthesized. In total, 55 primer pairs were found to be polymorphic and scorable consistently when screened in 32 accessions. The number of alleles ranged from 2 to 15, frequency of major alleles per locus varied from 0.17 to 0.91, the genotypes number ranged from 2 to 17, observed and expected heterozygosity values ranged from 0.00 to 0.44 and 0.17 to 0.89 and overall PIC values ranged from 0.16 to 0.88 respectively. These markers will be a useful tool for assessing the genetic diversity, understanding the population structure, and breeding patterns of faba bean.

Keywords: cDNA-SSR; genetic diversity; 454 sequencing; \textit{Vicia faba}

1. Introduction

Faba bean (\textit{Vicia faba} L.) is currently the third most important winter season food legume globally. Faba bean represent an important source for dietary protein to human beings, edible oil and animal feed. Its
critical role in crop rotation, effective nitrogen fixation, soil improvement abilities, and contribution to reducing energy input costs have long been recognized. Faba bean is a diploid with \(2n = 2x = 12\) chromosomes, is partially cross-pollinated (ranging from 4 to 84%) and possesses one of the largest genomes among crop legumes (~13,000 Mb) [1,2]. Despite being an alternative source of protein for human and effective in nitrogen fixation, the number of molecular markers available for faba bean is still scarce, with only 100 microsatellite (simple sequence repeat; SSR) markers [3,4] and only 32 EST–SSR markers [5,6] having been reported. The development of more reliable and informative molecular markers needs to be improved to enhance our understanding about the faba bean.

Next-generation transcriptome sequencing is an efficient means to generate superior resources for the development of cDNA–simple sequence repeat (SSR) markers. cDNA-SSR markers present some intrinsic advantages over genomic SSRs in their direct association with transcribed genes, low expense for development cost, and higher level of transferability to related species [7] and cDNA-SSRs, are highly polymorphic, rather than EST-derived SSR markers [8]. In a recent study, the authors sequenced faba bean transcriptomes using 454 pyrosequencing technologies and found that 1,729 SSR loci from 81,333 individual sequence reads and limited number of sequence 240 were submitted to GenBank, which paved the way for microsatellite marker development. In our study, we developed and characterized polymorphic cDNA-SSR markers based on these sequences for \(V.\ faba\) to facilitate the studies on molecular diversity of this species.

2. Results and Discussion

The \(V.\ faba\) transcriptome sequencing yielded 29.61 Mb and GS De Novo yielded 81,333 raw sequencing reads, based on the GS-FLX sequencer. SSR is one of the most popular marker systems, consisting of varying numbers of tandem repeated di-, tri-, or tetra-nucleotide DNA motifs. To identify SSR markers, we used the ARGOS program with default settings for the \(V.\ faba\) unigene collections. In total, 1,729 potential SSR motifs were identified, and the majority belonged to trinucleotide (67.61%) and dinucleotide (19.08%) repeats. All other types of SSRs such as tetra-, penta-, and hexa- nucleotide motifs were relatively low (13.3%), and the majority of trinucleotide SSRs had the GAA/AAG/AGA motif, followed by those with the TGG/CGT/GGT motif, and others with the CTT/TTC/TCT motif. The GA/AG, AT/TA, and GT/TG motifs were identified among the dinucleotide cDNA-SSRs. The relative proportion of SSR motif types in faba bean [9] were observed as similar to that of other plant species [10–12].

Among the identified SSR loci, we selected 240 sequences that were deposited in GenBank (GenBank accession number: KC218573–KC218812). Of the 240 primer pairs, only 55 primer pairs produced consistently amplified (Table 1). These 55 cDNA-SSR loci were screened in 32 accessions. The number of alleles \((N_a)\) per locus varied widely among the markers (Table 2), ranged from 2 to 15, with an average of 6.0 alleles. The frequency of major alleles \((M_{AF})\) per locus varied from 0.17 to 0.91 with an average of 0.563, the genotypes number \((N_G)\) ranged from 2 to 17, with an average of 6.3. The \(H_o\) values were ranged from 0.00 to 0.44 with an average of 0.074, the \(H_e\) values were ranged from 0.17 to 0.89 with an average of 0.587 and overall \(PIC\) values ranged from 0.16 to 0.88, with an average of 0.550 respectively. Similar observation was also found in the \(Vicia\ faba\) [6]. These cDNA-SSR markers were developed in our study are found to be a useful tools for further studies on molecular diversity and population structure of faba bean.
| Locus       | Primer sequences (5’–3’)                                                                 | Product Size | Modif | GenBank Accession No. | Ta (°C) | BLAST top hit accession no. | Description                                                                 | E-value |
|------------|-------------------------------------------------------------------------------------------|--------------|-------|-----------------------|---------|-----------------------------|------------------------------------------------------------------------------|---------|
| GBSSR-VF-8 | F:TAAAGCAGCTCCGGATGAR:TCGGTGAGGAGTTTGG                                                  | 242          | (TTG)5 | KC218580              | 58      | None                        | None                                                                         | None    |
| GBSSR-VF-19 | F:TCCATCAACCTCAAATCAR:CGTGACTCTGTCACGGG                                                | 201          | (CAA)5 | KC218591              | 58      | XP_003637109.1              | MTR_072s1002 [Medicago truncatula]                                             | 1.00E-41|
| GBSSR-VF-20 | F:TCCACCAAGTCCACCTGAR:AATAGGGGCAGAG                                                | 198          | (GTG)6 | KC218592              | 58      | None                        | None                                                                         | None    |
| GBSSR-VF-21 | F:CGAAGCTCTTCTCTTCTC R:GGTGTTGTATTCGGG                                                | 199          | (CCT)4 | KC218593              | 58      | ABF66654.1                 | EBPI [Ammopiptanthus mongolicus]                                              | 3.00E-37|
| GBSSR-VF-22 | F:CGAAGCTCTTCTCTTCTC R:CAAGTGCGCTTTTTCA                                                | 271          | (CCT)4 | KC218594              | 58      | None                        | None                                                                         | None    |
| GBSSR-VF-32 | F:CGAAGCTCTTCTCTTCTC R:GGTGTTGTATTCGGG                                                | 197          | (CCT)4 | KC218604              | 58      | XP_003614315.1             | None                                                                         | None    |
| GBSSR-VF-34 | F:CGGGAACCAATCAACAAR:CCACCAACGCAACTATG                                              | 186          | (CGA)4 | KC218606              | 58      | None                        | None                                                                         | None    |
| GBSSR-VF-38 | F:ACCATTGGGCTTCTCTCTC R:CGCTACCAAAATGCGA                                              | 226          | (GTG)6 | KC218610              | 58      | CAA61492.1                 | phloem specific protein [Vicia faba]                                          | 7.00E-18|
| GBSSR-VF-52 | F:GGTTCTTCTGCTCAAAATAGCACG R:TCCGATTCTGGAAATTG                                      | 261          | (CAG)5 | KC218624              | 58      | AFW76468.1                 | putative protein phosphatase 2A family protein [Zea mays]                      | 5.00E-27|
| GBSSR-VF-84 | F:CGCGGTCTAGACGTGTG R:AACTAGCGCAGCTCATCG                                              | 243          | (GC)5  | KC218656              | 58      | XP_003555733.1             | PHRED: ammonium transporter 1 member 1-like isoform 1 [Glycine max]          | 1.00E-40|
| GBSSR-VF-113 | F:TGGTGCTTCTTCTCA R:TGGTGAGCCTTGGAACTGC                                             | 213          | (CTT)5 | KC218685              | 59      | XP_002524296.1             | Abscisic stress ripening protein, putative [Ricinus communis]                | 4.00E-05|
| Locus             | Primer sequences (5′–3′)                                                                 | Product Size | Modif | GenBank Accession No. | Ta (°C) | BLAST top hit accession no.                                      | Description                                                                                   | E-value |
|-------------------|------------------------------------------------------------------------------------------|--------------|-------|-----------------------|---------|------------------------------------------------------------------|---------------------------------------------------------------------------------------------|---------|
| GBSSR-VF-115      | F: TGCTGCTTTTCCAAACCAT R: GTGCAATGCCCATAAACAAAA                                         | 177          | (AT)7 | KC218687              | 57      | XP_003600578.1                                                   | NAD(P)H-quinone oxidoreductase subunit N [Medicago truncatula]                               | 0.018   |
| GBSSR-VF-119      | F: GTGGCCTGTACTGGTGGA R: ACTCGTGGGGCTAGGA                                               | 225          | (GAA)5| KC218691              | 58      | AFK37381.1                                                       | unknown [Medicago truncatula]                                                               | 1.00E-80|
| GBSSR-VF-131      | F: CCGTACTAAATGAAGCCTTT R: GGCAATCAAGTCCGGTAA                                         | 238          | (TA)6 | KC218703              | 57      | XP_003546500.1                                                  | RNA polymerase-associated protein CTR9 homolog [Glycine max]                               | 1.00E-05|
| GBSSR-VF-149      | F: ACGACATGGAGATGAATCCT R: ACCTGAGGAGATGAGC                                            | 211          | (CAA)8| KC218721              | 58      | XP_003610732.1                                                  | NTR_5g006400 [Medicago truncatula]                                                          | 1.00E-58|
| GBSSR-VF-153      | F: TCCCGACGCTACTTTCTCA R: CCGAGATCTGCAAACACGC                                         | 225          | (CCA)5| KC218725              | 58      | AFK44330.1                                                      | unknown [Medicago truncatula]                                                               | 3.00E-44|
| GBSSR-VF-154      | F: ACACCAATGGTTTTTGCGG R: TCCTGACTTTTGCGG                                              | 247          | (GAA)5| KC218726              | 58      | XP_003522646.1                                                  | uncharacterized protein LOC100777431 [Glycine max]                                      | 1.00E-22|
| GBSSR-VF-159      | F: GTGCCATCATCCTCGAAA R: CAGCTGCTAGGTCCTG                                              | 235          | (TCT)4| KC218731              | 58      | ABD32307.2                                                      | Nuclear factor related to kappa-B-binding protein, related [Medicago truncatula]           | 3.00E-61|
| GBSSR-VF-164      | F: ACCATTGGCCCTCTTCT R: CAAGGAGGAGTTTTTACGA                                             | 199          | (GTG)6| KC218736              | 57      | CAA61492.1                                                      | phloem specific protein [Vicia faba]                                                        | 3.00E-11|
| GBSSR-VF-168      | F: TCTCCAACCCCTCTCTG T: CAGCCACAAAATCAGCA                                               | 250          | (TGT)6| KC218740              | 57      | XP_003605197.1                                                  | GRAS family transcription factor [Medicago truncatula]                                   | 9.00E-05|
| GBSSR-VF-172      | F: CGGTTCCTAAATCTGCCT G: GTCTCCATGAAACCAAATCTCTTCT                                      | 232          | (TTG)5| KC218744              | 57      | XP_003552039.1                                                  | uncharacterized protein LOC100790537 [Glycine max]                                       | 1.00E-15|
| Locus        | Primer sequences (5′–3′)                                                                 | Product Size | Modif | GenBank Accession No. | Ta (°C) | BLAST top hit accession no. | Description                                      | E-value  |
|-------------|-----------------------------------------------------------------------------------------|--------------|-------|-----------------------|---------|----------------------------|--------------------------------------------------|----------|
| GBSSR-VF-173 | F: CACAGACAGGGTTTCGGGA R: TTGGGTGCAACATCATCA                                           | 247          | (CAC) 6 | KC218745              | 58      | XP_003607781.1             | Low-temperature inducible                         | 7.00E-10 |
| GBSSR-VF-175 | F: TTGGGTGCAACATCATCA R: CCAGGCAATGGAATCTGA                                           | 256          | (TCC) 7 | KC218747              | 58      | ABN08705.1                 | DDT; Homeodomain-related                         | 1.00E-16 |
| GBSSR-VF-180 | F: CCATCAACGGAGGACTCA R: TAGGGGAACAGGGGCTA                                              | 183          | (TGC) 4 | KC218756              | 58      | XP_003594716.1             | Low-temperature inducible                         | 0.002    |
| GBSSR-VF-184 | F: TTTATGGAAGCGAAGCCG R: TTAGGGGAACAGGGCTA                                              | 156          | (TCT) 5 | KC218762              | 58      | XP_003608427.1             | Myocyte-specific enhancer factor 2B                | 3.00E-17 |
| GBSSR-VF-203 | F: TTCCCTGACCTTCAGCA R: TGGTGAGGACCTCAAGTT                                              | 231          | (CTC) 4 | KC218775              | 58      | XP_003555847.1             | CAX-interacting protein 4-like [Glycine max]      | 2.00E-14 |
| GBSSR-VF-209 | F: CCATCAACGGAGGACTCA R: GGGAACAGGGGCTAGGAGA                                             | 181          | (TGC) 4 | KC218781              | 58      | XP_002519249.1             | Prefoldin subunit, putative [Ricinus communis]    | 3.00E-28 |
| GBSSR-VF-213 | F: CCAAGTGTTCTTCTCCAGGA R: TTAAATTGGGGCGGAGAT                                            | 271          | (TC) 7 | KC218785              | 58      | None                       | None                                             | None     |
| GBSSR-VF-215 | F: ACAACTGAGTCCAAACCATGT R: TGCCACTGCTTCTTATGG                                         | 279          | (CAA) 4 | KC218787              | 58      | XP_003638360.1             | Cytochrome b reductase [Medicago truncatula]      | 1.00E-79 |
| GBSSR-VF-216 | F: CCATCCAGACGTGGAA R: GGCTGCAATCTACCACCA                                              | 194          | (ACA) 4 | KC218788              | 58      | ACM50914.1                 | Aquaporin [Medicago falcata]                      | 9.00E-18 |
| GBSSR-VF-218 | F: GTCTCGCATATACAAAGATG R: ATGGCGGTGCTACATATG                                           | 211          | (CCA) 5 | KC218790              | 57      | None                       | None                                             | None     |
| Locus         | Primer sequences (5′–3′)                                                                 | Product Size | Modif  | GenBank Accession No. | Ta (°C) | BLAST top hit accession no. | Description                                                                 | E-value |
|--------------|------------------------------------------------------------------------------------------|--------------|--------|-----------------------|---------|----------------------------|----------------------------------------------------------------------------|---------|
| GBSSR-VF-220 | F: ACCATTTGGCCTGTTCCT R: CAAATGCTGAAAATGGCCT                                             | 217          | (GTG)6 | KC218791              | 58      | CAA61492.1                 | phloem specific protein  [Vicia faba]                                      | 2.00E-13|
| GBSSR-VF-221 | F: CCGAAATGAGATGATGATGA R: TGAAAGGAAAACGTGAAAGTCA                                        | 243          | (CAT)5 | KC218792              | 58      | None                       | None                                                                        | None    |
| GBSSR-VF-237 | F: TTGGTGCAACATCATCA R: CGGGAAGAATCAGAGGC                                               | 234          | (GTG)5 | KC218793              | 58      | AKF45200.1                 | unknown  [Medicago truncatula]                                             | 2.00E-05|
| GBSSR-VF-245 | F: TTGTAACAGCATTTAATTTTC R: CATTACACCGTGAGGCA                                            | 279          | (TCA)5 | KC218794              | 57      | None                       | None                                                                        | None    |
| GBSSR-VF-255 | F: GGAGATGCTTTTGGCCT R: TGGTCTGCAGTTCCAT                                                | 262          | (GTT)5 | KC218795              | 58      | XP_003544671.1             | uncharacterized protein  [Glycine max]                                      | 2.00E-24|
| GBSSR-VF-258 | F: AACCCCTCATCTTCGGA R: CAGGAGATGATCTTGATAAGGC                                          | 165          | (GTG)4 | KC218796              | 58      | CAB16318.1                 | cysteine proteinase precursor  [Vicia narbonensis]                          | 5.00E-30|
| GBSSR-VF-262 | F: TCTGGCGAGTGGCATACT R: GCCTCTCACAACCGGCTTT                                              | 224          | (GAA)5 | KC218797              | 58      | CAB07811.1                 | sucrose transport protein  [Vicia faba]                                    | 9.00E-14|
| GBSSR-VF-263 | F: ATGCCACCTCTACTTCC R: TCCTCCTCAAATTCAGATCC                                             | 157          | (ACA)6 | KC218798              | 58      | XP_003592083.1             | Polypyrinidine tract-binding protein-like protein  [Medicago truncatula]   | 2.7     |
| GBSSR-VF-266 | F: CAAATGCAATGTGCTGAAA R: AACAGGTGGTGCTGTGATG                                             | 292          | (TCA)5 | KC218799              | 58      | None                       | None                                                                        | None    |
| GBSSR-VF-270 | F: CAGGGATGGACACACAACA R: TGAAAGGAAATGGGAAGGGG                                            | 252          | (TCA)4 | KC218800              | 58      | AAD50628.1                 | alpha-tubulin  [Gossypium hirsutum]                                       | 0.017   |
| Locus        | Primer sequences (5’–3’)                        | Product Size | Modif | GenBank Accession No. | Ta (°C) | BLAST top hit accession no. | Description                                                                                     | E-value |
|--------------|-------------------------------------------------|--------------|-------|-----------------------|---------|----------------------------|-----------------------------------------------------------------------------------------------|---------|
| GBSSR-VF-271 | F:TTTATGAAAGCGAACCGR:TTTGAACAAAGGGAAGGCAACA    | 158          | (TCT)5 | KC218801              | 56      | ACJ84526.1                 | unknown [Medicago truncatula]                                                                  | 1.00E-05|
| GBSSR-VF-276 | F:GCCCTTCAACCTGCCTTCAAGGCAACGCTTGGAGGCCAACA    | 150          | (CAA)4 | KC218802              | 58      | AAU11490.1                 | monodehydroascorbate reductase I [Pisum sativum]                                                | 9.00E-55|
| GBSSR-VF-285 | F:AAGAAGTTGTCGCGGAAGGCCGCGGCAAATATACACTTCCTTCT| 176          | (GAT)5 | KC218803              | 58      | None                       | None                                                                                           | None    |
| GBSSR-VF-288 | F:GCCTGTGGCTAGAAGCAAGCCGCGGCAAATATACACTTCCTTCT| 237          | (TCA)5 | KC218804              | 57      | None                       | None                                                                                           | None    |
| GBSSR-VF-293 | F:CCCCATTCAGAACGCTGAGAAGCGACGTGGAAAATCAGCAGAGCTTGTGGA | 160          | (ACA)4 | KC218805              | 58      | ACU17538.1                 | unknown [Glycine max]                                                                          | 0.018   |
| GBSSR-VF-295 | F:CAGGGTTACGATGTTGCTGAAATGTCAGGAGATACGTTGCAAGCAGACA | 175          | (TGC)4 | KC218806              | 58      | XP_003594716.1            | Prefoldin subunit [Medicago truncatula]                                                        | 3.00E-35|
| GBSSR-VF-297 | F:AGACCGAATCAAGGTCACACAAGCGACGTTCACTCTTCCAGGAGGAAGCAGACCTT | 150          | (AAT)6 | KC218807              | 58      | CAB39664.1                 | putative protein [Arabidopsis thaliana]                                                         | 4.00E-05|
| GBSSR-VF-298 | F:AATCATCAGCGAAGACCATCCAGCGACGTTCACTCTTCCAGGAGGAAGCAGACCTT | 249          | (TTC)4 | KC218808              | 58      | XP_003624491.1            | hypothetical protein MTR_7g083930 [Medicago truncatula]                                        | 3.00E-25|
| GBSSR-VF-311 | F:GGCCCTTCAACAGACGAGGGACTTGGCTGTTCCTTCTATGTTGAGGAAGG | 207          | (CAC)6 | KC218809              | 58      | CAA61492.1                 | phloem specific protein [Vicia faba]                                                            | 2.00E-16|
| GBSSR-VF-312 | F:GGCTATGCTGGTGACTGGACAGATCGCTGCTTCACTTGAAGGAGG | 151          | (TGG)5 | KC218810              | 58      | CAH40798.1                 | putative glycine rich protein precursor [Pisum sativum]                                       | 6.00E-08|
| GBSSR-VF-319 | F:CATGCAATTGGCTGCTCAGAGGAAATCGCAGGCGCACTACATGGGCACTGCAGG | 158          | (TTG)5 | KC218811              | 58      | NP_001237269.1            | VHS and GAT domain protein [Glycine max]                                                       | 5.00E-28|
| GBSSR-VF-323 | F:TCTGGCTTCCATCTTCCATCGAAGTTGCTCCTCTGTA         | 180          | (ACA)4 | KC218812              | 58      | None                       | None                                                                                           | None    |

Note: Ta, annealing temperature.
Table 2. Diversity statistics from initial primer screening in 32 accessions of faba bean (*Vicia faba* L.).

| Marker          | \(M_{AF}\) | \(N_G\) | \(N_A\) | \(H_O\) | \(H_E\) | PIC |
|-----------------|------------|--------|--------|--------|--------|-----|
| GBSSR-VF-8      | 0.17       | 17     | 15     | 0.28   | 0.89   | 0.88|
| GBSSR-VF-19     | 0.50       | 6      | 6      | 0.00   | 0.66   | 0.61|
| GBSSR-VF-20     | 0.42       | 8      | 9      | 0.09   | 0.74   | 0.71|
| GBSSR-VF-21     | 0.38       | 6      | 6      | 0.06   | 0.74   | 0.70|
| GBSSR-VF-22     | 0.47       | 7      | 8      | 0.03   | 0.71   | 0.67|
| GBSSR-VF-32     | 0.66       | 5      | 5      | 0.00   | 0.54   | 0.51|
| GBSSR-VF-34     | 0.65       | 6      | 6      | 0.00   | 0.56   | 0.53|
| GBSSR-VF-38     | 0.69       | 7      | 7      | 0.10   | 0.49   | 0.46|
| GBSSR-VF-52     | 0.47       | 7      | 7      | 0.00   | 0.69   | 0.65|
| GBSSR-VF-84     | 0.58       | 4      | 4      | 0.00   | 0.60   | 0.56|
| GBSSR-VF-113    | 0.38       | 7      | 7      | 0.00   | 0.76   | 0.73|
| GBSSR-VF-115    | 0.34       | 10     | 9      | 0.03   | 0.79   | 0.76|
| GBSSR-VF-119    | 0.28       | 6      | 6      | 0.00   | 0.80   | 0.77|
| GBSSR-VF-131    | 0.30       | 8      | 8      | 0.06   | 0.80   | 0.78|
| GBSSR-VF-149    | 0.47       | 9      | 9      | 0.00   | 0.73   | 0.70|
| GBSSR-VF-153    | 0.50       | 4      | 4      | 0.00   | 0.67   | 0.62|
| GBSSR-VF-154    | 0.47       | 7      | 7      | 0.06   | 0.71   | 0.68|
| GBSSR-VF-159    | 0.41       | 5      | 5      | 0.00   | 0.70   | 0.65|
| GBSSR-VF-164    | 0.52       | 8      | 7      | 0.06   | 0.67   | 0.64|
| GBSSR-VF-168    | 0.25       | 10     | 11     | 0.13   | 0.81   | 0.78|
| GBSSR-VF-172    | 0.30       | 18     | 14     | 0.44   | 0.85   | 0.84|
| GBSSR-VF-173    | 0.52       | 7      | 7      | 0.09   | 0.66   | 0.63|
| GBSSR-VF-175    | 0.23       | 13     | 12     | 0.25   | 0.82   | 0.80|
| GBSSR-VF-180    | 0.50       | 5      | 5      | 0.09   | 0.66   | 0.61|
| GBSSR-VF-184    | 0.61       | 5      | 5      | 0.03   | 0.56   | 0.51|
| GBSSR-VF-185    | 0.91       | 3      | 3      | 0.06   | 0.17   | 0.17|
| GBSSR-VF-190    | 0.56       | 9      | 9      | 0.28   | 0.64   | 0.61|
| GBSSR-VF-203    | 0.91       | 2      | 2      | 0.00   | 0.17   | 0.16|
| GBSSR-VF-209    | 0.77       | 4      | 4      | 0.03   | 0.39   | 0.35|
| GBSSR-VF-213    | 0.47       | 5      | 5      | 0.00   | 0.69   | 0.64|
| GBSSR-VF-215    | 0.72       | 3      | 3      | 0.00   | 0.44   | 0.39|
| GBSSR-VF-216    | 0.75       | 2      | 2      | 0.00   | 0.38   | 0.30|
| GBSSR-VF-218    | 0.81       | 2      | 2      | 0.00   | 0.30   | 0.26|
| GBSSR-VF-220    | 0.68       | 4      | 4      | 0.06   | 0.50   | 0.46|
| GBSSR-VF-221    | 0.45       | 7      | 6      | 0.13   | 0.68   | 0.63|
| GBSSR-VF-237    | 0.63       | 6      | 5      | 0.09   | 0.56   | 0.53|
| GBSSR-VF-245    | 0.72       | 5      | 5      | 0.00   | 0.45   | 0.42|
| GBSSR-VF-255    | 0.75       | 4      | 4      | 0.00   | 0.41   | 0.37|
| GBSSR-VF-258    | 0.56       | 6      | 6      | 0.13   | 0.63   | 0.60|
| GBSSR-VF-262    | 0.44       | 11     | 10     | 0.25   | 0.74   | 0.71|
| GBSSR-VF-263    | 0.41       | 8      | 8      | 0.19   | 0.70   | 0.65|
| GBSSR-VF-266    | 0.59       | 4      | 4      | 0.00   | 0.58   | 0.53|
| GBSSR-VF-270    | 0.69       | 3      | 3      | 0.00   | 0.48   | 0.43|
| GBSSR-VF-271    | 0.38       | 10     | 8      | 0.22   | 0.74   | 0.70|
Table 2. Cont.

| Marker         | $M_{AF}$ | $N_G$ | $N_A$ | $H_O$  | $H_E$  | PIC  |
|----------------|----------|-------|-------|--------|--------|------|
| GBSSR-VF-276   | 0.72     | 3     | 3     | 0.00   | 0.44   | 0.40 |
| GBSSR-VF-285   | 0.73     | 5     | 5     | 0.26   | 0.45   | 0.43 |
| GBSSR-VF-288   | 0.64     | 5     | 5     | 0.03   | 0.53   | 0.49 |
| GBSSR-VF-293   | 0.78     | 3     | 3     | 0.00   | 0.35   | 0.31 |
| GBSSR-VF-295   | 0.70     | 5     | 5     | 0.03   | 0.47   | 0.43 |
| GBSSR-VF-297   | 0.59     | 8     | 6     | 0.16   | 0.59   | 0.55 |
| GBSSR-VF-298   | 0.72     | 3     | 3     | 0.00   | 0.44   | 0.40 |
| GBSSR-VF-311   | 0.72     | 7     | 6     | 0.09   | 0.45   | 0.42 |
| GBSSR-VF-312   | 0.69     | 2     | 2     | 0.00   | 0.43   | 0.34 |
| GBSSR-VF-319   | 0.80     | 4     | 4     | 0.03   | 0.35   | 0.32 |
| GBSSR-VF-323   | 0.66     | 7     | 6     | 0.23   | 0.52   | 0.49 |
| Mean           | 0.563    | 6.3   | 6.0   | 0.074  | 0.587  | 0.550|

$(M_{AF})$, Major allele frequency; $(N_G)$, number of genotype; $(N_A)$, number of allele; $(H_O)$, observed heterozygosity; $(H_E)$, expected heterozygosity; (PIC), polymorphic information content.

3. Experimental

3.1. Plant Material

Faba bean seeds were selected from the National Agrobiodiversity Center, Rural Development Administration, Suwon, Korea. Seedlings were germinated and grown in a glasshouse. The leaves of young seedlings were used to extract the mRNA required to synthesize the cDNA library and for 454 sequencing.

3.2. cDNA Preparation

Total RNA was extracted from $V. faba$ leaves that were frozen in liquid nitrogen, ground well into a powder, and then extracted using an RNeasy Plant Mini kit (Qiagen, Valencia, CA, USA) following the manufacturer’s instructions. The integrity of total RNA was determined using a BIOSPEC-NANO spectrophotometer (Shimadzu, Kyoto, Japan) and agarose gel electrophoresis. mRNA was purified using the PolyATract mRNA Isolation System IV (Promega, Madison, WI, USA), and the purified products were used to synthesize full-length cDNAs using a ZAP-cDNA Synthesis kit (Stratagene, Santa Clara, CA, USA). Finally, cDNA was fragmented by nebulization for library construction.

3.3. Library Preparation

Approximately 1 µg of cDNA was used to generate a DNA library to use with the rapid library preparation method manual (Roche Life Science Inc., Branford, CT, USA). The cDNA fragment ends were polished (blunted), and two short adapters were ligated to both ends according to standard procedures described previously. The adapters provided priming of the sequences for both amplification and sequencing of the sample library fragments, as well as the sequencing key, a short sequence of four nucleotides used by the system’s software for base calling. Following repair of any nicks in the double-stranded library, the unbound strand of each fragment was released (with
5-Adaptor A). Finally, the quality of this single-stranded template DNA library was assessed using a 2100 BioAnalyzer (Agilent, Waldbronn, Germany). The library was quantified to determine the optimal amount of the library needed as input for emulsion-based clonal amplification.

3.4. 454 Pyrosequencing

Single effective copies of template species from the DNA library to be sequenced were hybridized to DNA capture beads. The immobilized library was then resuspended in an amplification solution, and the mixture was emulsified, followed by polymerase chain reaction (PCR) amplification. The DNA carrying beads were recovered from the emulsion and enriched after amplification. The second strands of the amplified products were melted, leaving the amplified single-stranded DNA library bound to the beads. The sequencing primer was then annealed to the immobilized amplified DNA templates. After amplification, a single DNA carrying bead was placed into each well of a PicoTiterPlate (PTP) device. Simultaneous sequencing with multiple samples on a single PTP (four region gasket) was used. The PTP was then inserted into the FLX Genome Titanium sequencer for pyrosequencing [13,14], and sequencing reagent was sequentially flowed over the plate. Information from the PTP wells was captured simultaneously by a camera, and the images were processed in real-time by an onboard computer. Multiplex identifiers were used to specifically tag unique samples in a GS FLX Titanium sequencing run, which were recognized by the GS data analysis software after the sequencing run and provided high confidence for assigning individual sequencing reads to the correct sample. Sequence assembly was performed after sequencing using GS De Novo Assembler software (Roche) to produce contigs and singletons. All sequence data were conformed to references using GS Reference Mapper software (Roche).

3.5. Discovery of cDNA-SSR Markers

All contigs and singletons from both transcriptomes were then used to mine SSR motifs, and the SSR motifs were identified using the ARGOS pipeline program (version 1.46) at the default settings to survey the molecular markers present in the V. faba accessions [15]. Parameters were designed for identifying perfect di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of six repeats. Primer design parameters were set as follows: length range, 18–23 nucleotides with 21 as optimum; PCR product size range, 100–400 bp; optimum annealing temperature, 55 °C; and GC content 40–60%, with 50% as optimum. Faba bean genomic DNA was extracted from 18 diverse faba bean accession samples for EST-SSR marker validation using a DNeasy® Plant Mini kit (Qiagen, Valencia, CA, USA), according to the manufacturer’s instructions. Fresh leaf tissue from each accession was used for each extraction and ground well using liquid nitrogen. DNA was resuspended in 100 μL water, and dilutions were made to 10 ng/μL followed by storage at either −20°C or −80°C. Randomly selected EST-SSR primer pairs were validated experimentally, and forward primers were synthesized by adding the M13 sequence to enable fluorescent tail addition through the PCR amplification process [16]. PCR conditions included a hot-start at 95 °C for 10 min, followed by 10 cycles at 94 °C for 30 s, 60–50 °C for 30 s and 72 °C for 30 s, followed by 25 cycles at 94 °C for 30 s, 50 °C for 30 s, and 72 °C for 30 s, and a final elongation step of 72 °C for 10 min. PCR products were separated and visualized using the QIAxcel Gel Electrophoresis System (Qiagen).
3.6. Data Analysis

These 55 SSR loci were screened in 32 accessions (Table 3). The number of alleles ($N_a$), major allele frequency ($M_A$), observed heterozygosity ($H_O$), Expected heterozygosity ($H_E$), number of genotype ($N_G$), and polymorphism information content ($PIC$) were calculated using GenAlEx (version 6.5) [17].

Table 3. List of faba bean ($Vicia faba$ L.) accessions and details of collection sites.

| No. | Temp. ID | ARS No.   | Origin               |
|-----|----------|-----------|----------------------|
| 1   | K193517  | PI 221516 | Afghanistan          |
| 2   | K193518  | PI 223418 | Iran                 |
| 3   | K193519  | PI 234634 | Australia            |
| 4   | K193520  | PI 251331 | Israel               |
| 5   | K193521  | PI 253806 | Iraq                 |
| 6   | K193522  | PI 254920 | Spain                |
| 7   | K193523  | PI 275641 | Ethiopia             |
| 8   | K193524  | PI 284338 | Lebanon              |
| 9   | K193525  | PI 284345 | Italy                |
| 10  | K193526  | PI 319901 | Soviet Union         |
| 11  | K193527  | PI 358270 | Serbia and Montenegro|
| 12  | K193528  | PI 371806 | United Kingdom       |
| 13  | K193529  | PI 415027 | France               |
| 14  | K193530  | PI 415050 | Sudan                |
| 15  | K193531  | PI 430133 | Egypt                |
| 16  | K193532  | PI 442559 | BEL                  |
| 17  | K193533  | PI 458504 | Mexico               |
| 18  | K193534  | PI 469135 | Japan                |
| 19  | K193535  | PI 469144 | Cyprus               |
| 20  | K193536  | PI 469180 | Greece               |
| 21  | K193537  | PI 469182 | Jordan               |
| 22  | K193538  | PI 469189 | India                |
| 23  | K193539  | PI 478506 | Bolivia              |
| 24  | K193540  | PI 499959 | China                |
| 25  | K193541  | PI 577723 | Kyrgyzstan           |
| 26  | K193542  | PI 577735 | Chile                |
| 27  | K193543  | PI 577741 | Nepal                |
| 28  | K193544  | PI 655323 | Morocco              |
| 29  | K193545  | PI 655325 | Peru                 |
| 30  | K193546  | PI 655326 | Ecuador              |
| 31  | K193547  | PI 655330 | Kenya                |
| 32  | K193548  | PI 655344 | Pakistan             |

Temp ID, Korean Gene Bank ID; ARS No, USDA-ARS Number.
4. Conclusions

In our study we have developed 55 cDNA-SSR markers, and they were successfully used to investigate the genetic diversity among 32 accessions of faba bean. However, there seems to be a relatively higher genetic diversity within *V. faba*, as only 55 of 240 cDNA-SSR loci exhibited polymorphism. The availability of co-dominant polymorphic cDNA-SSR markers provided a tool set for further study on molecular diversity, and will greatly facilitated the genetic structure studies of *V. faba* populations, the identification and conservation of faba bean.

Acknowledgments

This study was carried out with the support of “Research Program for Agricultural Science & Technology Development (Project No. PJ008623)”, National Academy of Agricultural Science, Rural Development Administration, Korea.

References

1. Bennett, M.D.; Smith, J.B. Nuclear DNA amounts in angiosperms. *Proc. Royal Soc. London. Ser. B. Biol. Sci.* 1982, 216, 179–199.
2. Johnston, J.S.; Bennett, M.D.; Rayburn, A.L.; Galbraith, D.W.; Price, H.J. Reference standards for determination of DNA content of plant nuclei. *Am. J. Bot.* 1999, 86, 609–613.
3. Pozarkova, D.; Koblizkova, A.; Roman, B.; Torres, A.M.; Lucretti, S.; Lysak, M.; Dolezel, J.; Macas, J. Development and characterization of microsatellite markers from chromosome 1-specific cDNA libraries of *Vicia faba*. *Biol. Plant* 2002, 45, 337–345.
4. Zeid, M.; Mitchell, S.; Link, W.; Carter, M.; Nawar, A.; Fulton, T.; Kresovich, S. Simple sequence repeats (SSRs) in faba bean: New loci from Orobanche-resistant cultivar ‘Giza 402’. *Plant Breed.* 2009, 128, 149–155.
5. Gong, Y.M.; Xu, S.C.; Mao, W.H.; Hu, Q.Z.; Zhang, G.W.; Ding, J.; Li, Z.Y. Generation and characterization of 11 novel EST derived microsatellites from *Vicia faba* (Fabaceae). *Am. J. Bot.* 2010, 97, e69–e71.
6. Ma, Y.; Yang, T.; Guan, J.; Wang, S.; Wang, H.; Sun, X.; Zong, X. Development and characterization of 21EST-derived microsatellite markers in *Vicia faba* (fava bean). *Am. J. Bot.* 2011, 98, e22–e24.
7. Wang, Z.; Fang, B.; Chen, J.; Zhang, X.; Luo, Z.; Huang, L.; Chen, X.; Li, Y. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genom.* 2010, 11, 726–739.
8. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.) *Theor. Appl. Genet.* 2003, 106, 411–422.
9. Kaur, S.; Pembleton, L.; Cogan, N.; Savin, K.; Leonforte, T.; Paull, J.; Materne, M.; Forster, J. Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genom.* 2012, 13, 104–115.
10. Kaur, S.K.; Cogan, N.O.I.; Pembleton, L.W.; Shinozuka, M.; Savin, K.W.; Materne, M.; Forster J.W. Transcriptome sequencing of lentil based on second generation technology permits large-scale unigene assembly and SSR marker discovery. *BMC Genom.* **2011**, *12*, 265–275.

11. Kumpatla, S.P.; Mukhopadhyay, S. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* **2005**, *48*, 985–998.

12. Luro, F.L.; Costantino, G.; Terol, J.; Argout, X.; Allario, T.; Wincker, P.; Talon, M.; Ollitrault, P.; Morillon, R. Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. *BMC Genom.* **2008**, *9*, 287–290.

13. Ronaghi, M. Pyrosequencing sheds light on DNA sequencing. *Genom. Res.* **2001**, *11*, 3–11.

14. Elahi, E.; Ronaghi, M. Pyrosequencing: A tool for DNA sequencing analysis. *Methods Mol. Biol.* **2004**, *255*, 211–219.

15. Kim, K.Y. *Developing one step program (SSR manager) for rapid identification of clones with SSRs and primer designing*. M.Sc. Thesis, the Graduate School, Seoul National University: Seoul, Korea, 2004.

16. Riley, M. Functions of the gene products of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **1993**, *57*, 862–952.

17. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **2012**, *28*, 2537–2539.

Sample Availability: Available from the authors’ institute.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).