Development of Genic and Genomic SSR Markers of Robusta Coffee (Coffea canephora Pierre Ex A. Froehner)

Prasad S. Hendre*, Ramesh K. Aggarwal*
Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, Telangana, India
*rameshka@ccmb.res.in

Abstract

Coffee breeding and improvement efforts can be greatly facilitated by availability of a large repository of simple sequence repeats (SSRs) based microsatellite markers, which provides efficiency and high-resolution in genetic analyses. This study was aimed to improve SSR availability in coffee by developing new genic—/genomic-SSR markers using in-silico bioinformatics and streptavidin-biotin based enrichment approach, respectively. The expressed sequence tag (EST) based genic microsatellite markers (EST-SSRs) were developed using the publicly available dataset of 13,175 unigene ESTs, which showed a distribution of 1 SSR/3.4 kb of coffee transcriptome. Genomic SSRs, on the other hand, were developed from an SSR-enriched small-insert partial genomic library of robusta coffee. In total, 69 new SSRs (44 EST-SSRs and 25 genomic SSRs) were developed and validated as suitable genetic markers. Diversity analysis of selected coffee genotypes revealed these to be highly informative in terms of allelic diversity and PIC values, and eighteen of these markers (~27%) could be mapped on a robusta linkage map. Notably, the markers described here also revealed a very high cross-species transferability. In addition to the validated markers, we have also designed primer pairs for 270 putative EST-SSRs, which are expected to provide another ca. 200 useful genetic markers considering the high success rate (88%) of marker conversion of similar pairs tested/validated in this study.

Introduction

Coffee tree belongs to the genus Coffea, comprising two main cultivated species C. arabica L. (2n=4x=44) and C. canephora Pierre ex A. Froehner (diploid,
2n=2x=22), yielding arabica and robusta type of coffees, respectively. Arabica coffee is known for excellent cup quality but suffers from a narrow genetic base due to its domestication history and susceptibility to diseases and pests. In contrast, robusta coffee though poor in quality has better adaptability to various stresses. To keep pace with the environment and also of the sensibilities of market, there is a continuous need for genetic improvement of coffee, which unfortunately is severely constrained owing to inherently slow pace of tree breeding using conventional methods, and variety of other reasons [1, 2]. The situation demands development of new, easy, practical technologies that can provide acceleration, reliability and directionality to the breeding efforts, as well as characterization of cultivated/secondary gene pool for proper utilization of the available germplasm in coffee genetic improvement programs. In this context, DNA polymorphism based genetic markers becomes important that have proven to be of immense value in characterization and genetic improvement of plant germplasm resources.

Among different types of DNA markers, microsatellites or SSR markers are the most ideal for studying genetic diversity, population structure, phylogenetic relationships, construction of frame-work linkage maps, QTL interval mapping, marker-assisted selection (MAS), etc., thereby aiding in genetic improvement of crop plants [1]. In the last few years a number of efforts have lead to development of a few hundred SSR markers in coffee [2–12], but these are insufficient to realize the full potential of markers for mapping/linkage studies in coffee, more so in arabicas which have an extremely narrow genetic base. Moreover, most of the described markers are poorly validated, especially for their utility in cultivated genepool comprising arabicas and robusta coffee. The situation thus calls for newer efforts to generate additional validated markers for them being of any gainful utility in marker-based genetic studies/coffee breeding.

With advancements in genomic studies, there has been an huge burst in the EST sequences in the public domain that provide an easy and economic/cost-effective opportunity to identify and develop EST based SSR markers, which have the additional advantage of assessing the functionally effective genetic diversity [13, 14], and also have very high cross-species transferability [8, 11]. In this study, we have used the coffee EST database containing 13,175 unigene [15] to identify SSRs in the expressed part of coffee genome, and use the same to develop novel coffee-specific EST-SSRs for use as efficient genetic markers. Thus we describe here 44 new validated genic-SSRs, and another set of 270 putative similar markers that need further validation. In addition, we also describe 25 new genomic SSR markers that were developed using an affinity capture approach based SSR-enriched partial, small-fragment genomic library.

Materials and Methods

Plant material and DNA extraction

The plant material used for the validation of SSR markers comprised a set of 16 elite coffee genotypes belonging to C. arabica (tetraploid arabicas) and
C. canephora (diploid robustas) and 14 related wild species belonging to Coffea and Psilanthus [2] that were available in the Coffee Germplasm Bank maintained at Central Coffee Research Institute, Balehonnur, Karnataka, India. The fresh leaf samples collected from each genotype were used for DNA isolation as described by Aggarwal et al. [16]. The DNA isolated from robusta variety CxR was used for constructing SSR enriched small-insert genomic library.

**Microsatellite screening of coffee transcriptome, identification of SSRs and marker development**

An EST database of robusta coffee comprising 13,175 unigene ESTs [15] was downloaded from ftp site (ftp://ftp.sgn.cornell.edu/coffee/) maintained by Sol Genomics Network (SGN, http://www.sgn.cornell.edu/coffee.pl). The database was used for: (i) identification and localization of SSRs using microsatellite search module MISA (MIcroSAtellite, http://www.pgrc.ipk-gatersleben.de/misa), and the criteria being- a minimum repeat core of 12 bp, considering the base complementarities and a minimum distance of 50 bp between two SSRs; (ii) selecting the ‘usable/candidate SSRs’ for marker development, being those that carried a minimum of 18-bp long repeat core (nine repeat units of DNRs, six of TNRs, five of TtNRs, four of PNRs, or three of HNRs) (iii) designing of primer pairs for the selected usable SSR sequences using PRIMER 3 tool embedded in MISA and/or GENETOOL Lite version 1.0 (http://www.biotools.com/downloads/brochures/GeneTool2.pdf); and (iv) standardizing PCR conditions followed by validation of working primer pairs for genetic studies as described earlier [2].

**Construction of an SSR-enriched small-insert genomic library/ development of genomic SSRs**

A partial genomic DNA library enriched for microsatellite repeats was constructed using the methods described earlier [17]. Briefly, the method involved: one-step restriction digestion of genomic DNA with Hae III enzyme (NEB) and ligation of resulting fragments with ds Mlu-I adaptor (Mlu-F: CTC TTG CTT ACG CGT GGA CTA and Mlu-R: pTAG TCC ACG CGT AAG CAA GAG CAC) [18]; amplification of the restricted-ligated DNA pool using Mlu-F primer; SSR enrichment of the amplified DNA pool using liquid phase hybridisation (in 6X SSC) with streptavidin coated paramagnetic beads (Dynal) attached with biotinylated equimolar pool of four oligos (CA)_{15}, (GA)_{15}, (GAA)_{15} and (CAA)_{15}. This was followed by amplification of the hybridized/trapped genomic DNA fragments by PCR and construction of partial genomic library in TA vector (Invitrogen) as per the manufacturer’s instructions. A number of positive (white) recombinant clones were randomly picked up from the library, amplified and sequenced for both the strands using M13 universal primers on ABI 3730 DNA Analyzer (Applied Biosystems, USA). The sequences were aligned and edited using Autoassembler (Applied Biosystems, USA). The SSR-positive sequences were identified and used for development of new genomic SSRs as described earlier [2].
The amplified PCR products generated using all the new SSRs were resolved using capillary-based ABI 3730 DNA Analyzer and were precisely sized for major, comparable and conspicuous peaks using GeneMapper 3.7 (Applied Biosystems), using default parameters.

**Statistical, genetic and diversity analysis**

The data for EST-SSRs and genomic SSRs were analyzed separately for various genetic parameters, viz., mean, standard deviation, expected heterozygosity ($H_e$), Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium, (LD), using Arlequin ver 3.1 [19], polymorphism information content (PIC) and private alleles (Pas) using Convert ver. 1.3.1 [20]. Cross-taxa transferability ($T_{mark}$) was calculated over 15 species (except C. canephora) as proportion of primers showing successful amplification vis-à-vis all the tested primers whereas primer conservance ($C_{taxa}$) was calculated as proportion of the species displaying successful amplification vis-à-vis all the tested markers.

Genetic diversity analysis to infer generic relatedness/affinities was performed over informative Pms (polymorphic markers) for cultivated genotypes/related species using MicroSatellite Analyzer [21] with Nei’s genetic distance [22]. The genetic distance matrices were used to construct Neighbour Joining (NJ) consensus tree using Phylip ver 3.6 [23], which was viewed using Treeview ver 1.6.6 [24].

We also attempted mapping of the new markers on a robusta linkage map using JoinMap ver 4.0 [25] as described earlier [2, 26] using group LOD score of 5.0.

**Results**

In this study, we undertook in-silico analysis of a robusta coffee transcriptome to identify and develop coffee-specific EST-SSR markers. Simultaneously, we also attempted development of genomic-SSRs by constructing a small-insert SSR enriched partial genomic DNA library. The new markers (44 EST-SSRs and 25 genomic-SSRs) were validated for their utility in genetic studies using panels of elite coffee genotypes, and related taxa of coffee for cross-species transferability. The details of these new markers viz., locus designation, primer sequences, repeat motifs, amplification temperature, amplicon size, and SGN ID or Genbank accession numbers are given in Tables 1 and 2. Details of additional primer pairs for 270 putative EST-SSRs, which were designed but need to be validated have been provided in Table 3.

**Types, Frequency and Distribution of SSRs in the coffee transcriptome**

The coffee EST unigene database analyzed here, comprised 13,175 unigenes having a total length of 8923 kb and an average length of ca. 677 bases/unigene [15]. These ESTs were found to contain a total of 2,589 SSRs (having a minimum...
| Sl. No. | Marker Id   | Primer sequence          | Repeat motif | Allele size (bp)* | Source EST ID** | Linkage group |
|--------|-------------|--------------------------|--------------|-------------------|-----------------|---------------|
| 1      | CCESSR01    | F: TGGTAGACACTGTCCGGAAAGCATAT  | (AGC)7       | 239               | 119516          | –             |
|        |             | R: GACCCATCTAACTTGTGCCCGATTTC |             |                   |                 |               |
| 2      | CCESSR02    | F: AAGATATGTGGTTTTGACGCAAGTGTGAC | (AT)11      | 168               | 119534          | –             |
|        |             | R: ATGGTTGACTTGTGGTACCTGTCAT |             |                   |                 |               |
| 3      | CCESSR03    | F: CAGCCGCTACGACCAGCAT     | (TC)10       | 217               | 119559          | CLG02         |
|        |             | R: TTCCCAACCGCTCAAAATGCCT  |             |                   |                 |               |
| 4      | CCESSR04    | F: GTGACAAACAGCTCGATGAGTC   | (TCT)6       | 259               | 119613          | –             |
|        |             | R: CAAAAAGACTGGAAGAGGTTTAG |             |                   |                 |               |
| 5      | CCESSR05    | F: AGGGGCCTGTTATTTTTTGGG   | (GCA)6       | 141               | 119723          | –             |
|        |             | R: GGGGTGAAATACGGGAAAGCAGA |             |                   |                 |               |
| 6      | CCESSR06    | F: AACCTCCCTCTCTCCACCTTTTC | (AC)23       | 224               | 119781          | CLG09         |
|        |             | R: GGAGGTGTTGTTGTTGTTAAAAAG |             |                   |                 |               |
| 7      | CCESSR07    | F: CCCCCCTCCATTTTTTCTCCT | (AGC)6       | 194               | 119864          | –             |
|        |             | R: CCGGTTGAGAGGTGGTGTGCTGCA |             |                   |                 |               |
| 8      | CCESSR08    | F: AAACCCACATCCCCAAGAAGTT | (GCA)8       | 149               | 120008          | –             |
|        |             | R: GCCAGTAGAAATGGAGGCTGAA |             |                   |                 |               |
| 9      | CCESSR09    | F: CCCCCCACCCTTCTTCGGT     | (TT)C8       | 184               | 120113          | –             |
|        |             | R: ACAAAACAAGGACGCTCTGATAAA |             |                   |                 |               |
| 10     | CCESSR10    | F: GCAGAAGAGCAGACTGAGCAGAA-GAG | (GCT)7 | 207               | 120120          | –             |
|        |             | R: TGCCCTCTACTTC CTCTCTTC |             |                   |                 |               |
| 11     | CCESSR11    | F: AGGGGAAAAAGAAGTGAGAAGAAGATATT | (GAA)6 | 140               | 120217          | –             |
|        |             | R: GGCACACTCACTATATCTGCAAC- TTAG |             |                   |                 |               |
| 12     | CCESSR12    | F: CCGCCATCCCTTGGCCCTTTC | (CCA)6       | 200               | 120227          | CLG02         |
|        |             | R: ACGCCACGCAAGTGGAGGTGTTT |             |                   |                 |               |
| 13     | CCESSR13    | F: GCGGGGTAGTTTTGGGATATGG | (TTA)6-It-(TTC)6 | 124               | 120252          | CLG02         |
|        |             | R: TTTGGGCTCTTCCCCCCATCCACAT |             |                   |                 |               |
| 14     | CCESSR14    | F: CTTGCCCCCCCTTCCCTCCTACTC | (CT)10       | 232               | 120329          | –             |
|        |             | R: TTGCGCTCTTGGTTGGTGGA |             |                   |                 |               |
| 15     | CCESSR16    | F: AGAGCATAATGAGAAGAAACAAGAAGACT | (CAT)6 | 173               | 120439          | –             |
|        |             | R: AGTGTCCTAAAATCTCCAGAAT |             |                   |                 |               |
| 16     | CCESSR17    | F: CTCGCCACAAACAAAAATCCCAGTTT | (CAC)6 | 169               | 120475          | –             |
|        |             | R: CCACATCTGAGTCTGTGACTAA |             |                   |                 |               |
| 17     | CCESSR18    | F: GGGGAGGATGCTTATGATGAGGTTAG | (CAC)6 | 152               | 120498          | –             |
|        |             | R: TCGGTTACCTCTGGTTTTCTT |             |                   |                 |               |
| 18     | CCESSR19    | F: CATCTGTATCTGCCTTCTCCTCCTC | (CA)13       | 219               | 120514          | –             |
|        |             | R: CCACAAACAGTACACACACACAGG |             |                   |                 |               |
| 19     | CCESSR20    | F: TCTTGGCCGGATTTGCTGATTAG | (AAG)7       | 143               | 120538          | –             |
|        |             | R: GCACAGACGCAGTGAACACTTACAC |             |                   |                 |               |
| 20     | CCESSR21    | F: CGAGCTAGTGCAAAGACATGAGAGATG | (AG)17 | 164               | 120628          | CLG13         |
|        |             | R: GTCCCTGGCGAAATCCCTCAG |             |                   |                 |               |
| 21     | CCESSR22    | F: CCCTAATCTCGTCCCTCCTCT | (TC)9        | 153               | 120823          | –             |
|        |             | R: CCCTCCATAATCTCTCTTTCAGTACTC |             |                   |                 |               |
| Sl. No. | Marker Id | Primer sequence | Repeat motif | Allele size (bp)* | Source EST ID** | Linkage group |
|-------|-----------|-----------------|--------------|------------------|----------------|---------------|
| 22    | CCESSR23  | F: GGCCTCTCTTTAATTTTCTTTTTTTTC | (TTC)8 | 160             | 120860        | –             |
|       |           | R: ATGGAGGGTAGGGTTTGCAGAGTGAG   |             |                  |                |               |
| 23    | CCESSR26  | F: AACCAGCCTCTTTGATGATCTTCTTTCTCTCT | (CAT)8 | 206             | 121392        | CLG07         |
|       |           | R: TTGGCTAACCTCCTACTCTCTCCCTACTA |             |                  |                |               |
| 24    | CCESSR27  | F: GCCCACTCACTGCTACTGCTTTCCC   | (CT)12 | 120             | 121464        | –             |
|       |           | R: GCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT |             |                  |                |               |
| 25    | CCESSR28  | F: AAATGAGGGATGGGCTTAGGGAGAGAG | (CCA)7 | 193             | 121482        | –             |
|       |           | R: AAGGGAAGGCAGATCAGCTGAGGG |             |                  |                |               |
| 26    | CCESSR29  | F: GCCGCTAGGTGTGTTGTTGTTGTTG | (CTTCT)6 | 95              | 121548        | –             |
|       |           | R: CGCCATTGGAAACCCGGACGAAC |             |                  |                |               |
| 27    | CCESSR31  | F: AGAAGAGAGTGAGGCTAGGCTGGAAGA | (GT)6 | 220             | 121671        | –             |
|       |           | R: ACGCATCGACCTCTCCATAGC |             |                  |                |               |
| 28    | CCESSR32  | F: CTTGGCCTTTAGCTTGCTCTCATTT | (TA)9 | 163             | 121811        | –             |
|       |           | R: GTCTCAACCAAATCGATATTTCTCTCTTCT |             |                  |                |               |
| 29    | CCESSR33  | F: GCCCCGATGGGACGACTTGGGA | (AGG)7 | 227             | 121876        | CLG13         |
|       |           | R: CCGTTGAGATTTGCTTAGTTGCT |             |                  |                |               |
| 30    | CCESSR34  | F: GCATTTCTGCTCTGCTCAT | (CTC)6 | 168             | 121905        | –             |
|       |           | R: GACGCTGAGGGAAGGAGGAAGAGGA |             |                  |                |               |
| 31    | CCESSR35  | F: CTGCTAATGGTGCTGAAAGAAATGATACC | (AG)10 | 107             | 121994        | CLG13         |
|       |           | R: GGCTGTGAATTCTTGTGACTTGTGACT |             |                  |                |               |
| 32    | CCESSR36  | F: AGCCTTTTGCAATCCCTCTGTACA | (AG)9 | 100             | 122089        | –             |
|       |           | R: GCCGCTGCTAGGCGATGAGACGA |             |                  |                |               |
| 33    | CCESSR38  | F: GCCGCAAGGTGATGAAGTGATAGTACA | (AG)12 | 162             | 122114        | –             |
|       |           | R: CTGTCTTCTGTTGATTTTTGTCTTCTA |             |                  |                |               |
| 34    | CCESSR39  | F: GCGACCCGAGCACAAGAAGAATATTTTAT | (GAA)7 | 131             | 122147        | –             |
|       |           | R: GCACCCTCGCTGACGGCTAATTAAATCA |             |                  |                |               |
| 35    | CCESSR40  | F: CGTGGGGGTTTTGTCTTCTCG | (AGG)7 | 205             | 122194        | –             |
|       |           | R: GTCGCCCTCAGCCGCTTTTTTT |             |                  |                |               |
| 36    | CCESSR41  | F: GGGCTGACGGCTTGTACACGACCGA | (GA)10 | 205             | 122322        | –             |
|       |           | R: ATTCGTTTTTAGTTTGTCTTCTCCTAAC |             |                  |                |               |
| 37    | CCESSR42  | F: CGGGCCGAAACCGCTAGATC | (GCA)7 | 117             | 122295        | –             |
|       |           | R: TGGCCCTGTTTTGCTGCT |             |                  |                |               |
| 38    | CCESSR43  | F: CCCAGCAAGACTCAACCCACCTAC | (TTC)8 | 172             | 122653        | CLG02         |
|       |           | R: TGCCCTAAATGAAATGACGGCTTGCTGATG |             |                  |                |               |
| 39    | CCESSR44  | F: AGGAATAATGGAGGAGACGTGTTG | (CTT)7 | 236             | 122680        | CLG02         |
|       |           | R: GCAACCAATCCCAGTACCTGCTCATAGA |             |                  |                |               |
| 40    | CCESSR45  | F: AAATGAGGCCAGATAGAGGAAAGGA | (AGG)7 | 135             | 122764        | –             |
|       |           | R: CAAAATCCCTCCGCGTGCTTGATC |             |                  |                |               |
| 41    | CCESSR47  | F: GCAGCACAATCATCCACACAGC | (GCA)6 | 198             | 122922        | –             |
|       |           | R: TGCTGTGTACACTGGAGGTTTGTT |             |                  |                |               |
| 42    | CCESSR48  | F: GCAGGCGATCTCAGTTTGCTCTCAACTT | (AAATCA)5 | 188             | 122975        | –             |
|       |           | R: CAGGGAGAATATGGCCAGCTCATA |             |                  |                |               |
| 43    | CCESSR49  | F: CGGGCCATCCTTGGCTCTTCG | (ATC)8 | 186             | 122978        | CLG02         |
numbers of repeats as: six for DNRs, four for TNRs and three for all other HO-NRs) located in 2,028 unigenes. The identified 2,589 SSRs comprised- 502 DNRs, 1285 TNRs, 144 PNRs and 155 HNRs, which differed significantly in their relative distribution and abundance across the unigene ESTs (Tables S1 & S2). The mean length of repeat iterations (RI) for all the SSRs was 5.2, whereas average length of DNRs was maximum (9.6 RI) followed by TNRs (4.6 RI). Among the individual SSRs, AC had the maximum average repeat length of 10.5 RI, followed by AG (8.6 RI), AT (8.3 RI), CG (6.3 RI); all the TNRs had RI in the range of 4.3 to 4.8, whereas all other larger SSRs had an average RI of less than three. The identified SSRs having a repeat core of 18 bp or more were selected as candidate ‘usable’ SSRs for further primer designing/marker conversion. Overall, in-silico analysis of the EST unigenes revealed one EST-SSR (having a minimum repeat core of 12 bp) per 3.4 kb and one usable SSR (having a minimum repeat core of 18 bp) per 15.9 kb of robusta transcriptome (Table S2). Among the individual SSRs, the most abundant EST-SSR motif was AG, followed by AAG.

Development of microsatellite markers from usable EST-SSRs
Only 483 (18.7%) of the total 2589 identified SSRs had a repeat core>18 bp, which were used for marker conversion. Primer pairs could be designed for 320 of these SSRs, of which randomly chosen 50 pairs were further tested for validation studies. These included SSRs with DNRs (30%), TNRs (64%), PNRs, HNRs (2% each) and complex SSRs (see Table 1 for marker ID, primer sequences, repeat motifs, amplicon size, sequence ID and functional identity). Of the selected 50 primer pairs, 44 could be successfully amplified as single locus SSR marker, indicating 88% primer to marker conversion ratio. Considering this high conversion ratio, another ca. 200 useful genetic markers are expected from the remaining 270 putative EST-SSRs (Primer IDs: CCESSR51 to CCESSR320) that are listed in Table 3.

| Sl. No. | Marker Id | Primer sequence | Repeat motif | Allele size (bp)* | Source EST ID** | Linkage group |
|--------|-----------|-----------------|--------------|------------------|----------------|--------------|
| 44     | CCESSR50  | F: GGGATGATGTTGATCATGTTCTACTA (CAG)7 | 108           | 123181 CLG10     |                |              |
|        |           | R: ATGCCATTTACACTTCTCCTCA       |              |                  |                |              |

CCES: CCMB Coffee EST SSR marker;
F: forward primer;
R: reverse primer;
–: Unmapped;
CLG: Combined Linkage Group [1];
*: Predicted ampiclon size based on source EST sequence;
**: Source EST ID as per the downloaded SGN database (ftp://ftp.sgn.cornell.edu/coffee/).

doi:10.1371/journal.pone.0113661.t001
Table 2. Details of the new genomic SSR markers developed using streptavidin-avidin affinity capture SSR-enriched library in the present study.

| Sl. No. | Primer Id  | Primer sequence                      | Repeat motif | T<sub>a</sub> (°C) | Allele size (bp)* | GeneBank acc. No. | Linkage group |
|---------|------------|--------------------------------------|--------------|---------------------|-------------------|------------------|---------------|
| 1       | CCRM02     | F: AATGGTGCGACTCTGGAAAGATC           | (GA)12       | 57                  | 268               | KM874369         |               |
|         |            | R: AGATCAACTTCTTGTGOCGTC             |              |                     |                   |                  |               |
| 2       | CCRM06     | F: TTCTTATACCTTGGGCTACCTTCTTC       | (AG)8        | 57                  | 146               | KM874370         |               |
|         |            | R: AACATCAACTTCTTGTGOCGTC             |              |                     |                   |                  |               |
| 3       | CCRM07     | F: TAAAGGATGGTATATGTGGCTGGAGTA      | (AT)8        | 57                  | 126               | KM874371         | CLG01         |
|         |            | R: CGGACCTCCGTTGGATTATCATATATAT     |              |                     |                   |                  |               |
| 4       | CCRM10     | F: AAAAGAGAAGATTACCATCGTCAACTGAGT  | (GT)9        | 57                  | 104               | KM874372         |               |
|         |            | R: TTTCCACCCCCCAAAAAAAA               |              |                     |                   |                  |               |
| 5       | CCRM14     | F: ATTTGATTCTTCTCTCTGTGGTC          | (CT)22       | 55                  | 130               | KM874373         |               |
|         |            | R: AAAAAAGCCCTGAAATAATAGATCTA       |              |                     |                   |                  |               |
| 6       | CCRM15     | F: CGGAGATGCAAGGCTCCTTGTT             | (CTT)6       | 57                  | 243               | KM874374         |               |
|         |            | R: TGGCTAGTCTCAGAAATCGTGTAAGGAC     |              |                     |                   |                  |               |
| 7       | CCRM16     | F: TCCATAGGAAACAAACAAATGCAGACAG   | (TC)26       | 55                  | 223               | KM874375         |               |
|         |            | R: GGGTTTTGGGCTCTTTAAGATCATAA       |              |                     |                   |                  |               |
| 8       | CCRM17     | F: TACGGGGATGATCCCTAATATCTCT       | (CT)17       | 57                  | 228               | KM874376         | CLG01         |
|         |            | R: ACAGCTAAGAAAACAAATGAACCAGT       |              |                     |                   |                  |               |
| 9       | CCRM19     | F: GGGTTTTTTTTTTTTTTTTTTTTTTTTT  | (GA)26       | 57                  | 252               | KM874377         | CLG06         |
|         |            | R: AAGGCAATGAGTGTCAGCGGTG          |              |                     |                   |                  |               |
| 10      | CCRM21     | F: CACCCCCCTCCCATCGGTGAAACAT     | (GA)16       | 57                  | 258               | KM874378         | CLG06         |
|         |            | R: AATGATGCTCCAGTGTTGATGGA          |              |                     |                   |                  |               |
| 11      | CCRM22     | F: CTTGAGTTTAATCTCCTCTCTTGGTTG  | (GA)29       | 57                  | 241               | KM874379         |               |
|         |            | R: TTTCTCTCTATATGTGGAGTTCTCTC      |              |                     |                   |                  |               |
| 12      | CCRM23     | F: CGGAGATGCTGCTGCCCCCTTTGGAT    | (GTT)6       | 57                  | 141               | KM874380         |               |
|         |            | R: AAAAAACTCACACTCTATCAAACAAACTAAGG |              |                     |                   |                  |               |
| 13      | CCRM24     | F: GAGTGTGATGTTTTTTTTTTTGGCTCACTTAA | (GA)9,(GA)9 | 57                  | 213               | KM874381         |               |
|         |            | R: ACCCCCCCATCCCTCCATCATCCCATCC    |              |                     |                   |                  |               |
| 14      | CCRM28     | F: GGGCGAACAGTGGTAGGATGGAAGAC    | (GA)10       | 57                  | 209               | KM874382         |               |
|         |            | R: CGCTTCTCTATGTTTTTTGCTCTTAA     |              |                     |                   |                  |               |
| 15      | CCRM31     | F: CTTTTATGCTTCTATGCTCTCTGCTTCTCT | (CA)10       | 57                  | 114               | KM874383         |               |
|         |            | R: CCTGAGTAGTTTCACTCTTATCTCC        |              |                     |                   |                  |               |
| 16      | CCRM33     | F: ACAACGCTGTTAATCTGGAAGATTACA      | (CT)12       | 57                  | 118               | KM874384         |               |
|         |            | R: ACAAAAGGATGAGGAGAGGATGGAATAAC   |              |                     |                   |                  |               |
| 17      | CCRM34     | F: CCCCGAGAACGAAAGGGGAATCAT       | (GA)10       | 57                  | 165               | KM874385         | CLG05         |
|         |            | R: TTGGGACTTATTTACTTGGGGAAGAA      |              |                     |                   |                  |               |
| 18      | CCRM35     | F: GGGTTAAAATCAGGGGAAAAAGTGG      | (TG)12       | 57                  | 144               | KM874386         |               |
|         |            | R: AAGGCGAGGAGAGACAGACGATC        |              |                     |                   |                  |               |
| 19      | CCRM36     | F: CCAAAGGGCAGACGACAATTCTAT     | (CT)18       | 57                  | 171               | KM874387         |               |
|         |            | R: TCCAGGCAGGGCCTCCAGGAGTATAA    |              |                     |                   |                  |               |
| 20      | CCRM37     | F: TGCTTCTCTTCTACCTCGATCTTCTT     | (GT)10       | 57                  | 146               | KM874388         |               |
|         |            | R: AATCCATCAAAACTTTGGCATACCA     |              |                     |                   |                  |               |
| 21      | CCRM38     | F: TGAGAATTAAGGGACAGGGGTATG      | (CA)10       | 57                  | 204               | KM874389         | CLG05         |
|         |            | R: GCAGAAAAAGGCAAAAAAGCATTACAC    |              |                     |                   |                  |               |
| 22      | CCRM40     | F: ATTCAGGTTTTTCACTTTCTCT       | (CT)29       | 57                  | 176               | KM874390         | CLG05         |
Identification and development of genomic SSRs using affinity capture

Sequencing of randomly chosen 66 recombinant clones from the small-insert SSR-enriched robusta genomic library (prepared in this study), revealed 81 potential SSRs distributed in 62 sequences. A redundancy analysis of these sequences indicated presence of a total of 60 non-redundant sequences of which 56 were SSR+ive (93.3% of non-redundant sequences) containing 72 non-redundant SSRs. The non-redundant SSRs contained 56.9% AG, 33.3% AC, 2.8% AT, 2.8% AAC, 1.4% AAG and 2.7% A/T repeat motifs (Table S3).

From the 56 SSR+ive sequences, a total of 41 primer pairs could be designed successfully (with five pairs containing two SSRs each). Of these, 25 pairs encompassing 28 SSRs resulted in robust PCR amplifications (Table 2), and all of them could further be validated as single locus markers indicating ~61% primer to marker conversion ratio.

Validation of EST-SSRs for use in genetic studies

All the new 44 EST-SSRs resulted in good amplicons exhibiting low to medium allelic diversity when tested on a panel of 16 elite robusta and arabica genotypes (Figure 1). Overall, a maximum of six and seven alleles ($N_A$) with an average of 2.1 and 3 alleles/SSR were obtained for the tested markers of which 65.9% and 81.8% were polymorphic/informative for tetraploids and robusta genotypes, respectively (Table 4). Fifteen markers in the case of tetraploids and eight for robustas were found to be monomorphic. Moreover, 14 markers resulted in double alleles (i.e. consistent presence of two allelic amplicons across the tested samples) indicative

---

**Table 2. Cont.**

| Sl. No. | Primer Id | Primer sequence | Repeat motif | $T_a$ (°C) | Allele size (bp)* | GeneBank acc. No. | Linkage group |
|---------|-----------|-----------------|--------------|------------|-------------------|-------------------|--------------|
| 23      | CCRM41    | F: AGCAGAAAAACACAAATGACACAGAGCA R: AATGGTCCAAGGAAATGAAAAATGTT | (CT)24 | 57 | 161 | KM874391 | – |
| 24      | CCRM42    | F: CGGAGAAGAGCAATATAACAGCAAGG R: GCCACCCAGAACAATTTT6CAA | (GA)13 | 55 | 143 | KM874392 | CLG11 |
| 25      | CCRM45    | F: CTTCAAGCAAATTTTCAACAGCAGACG R: GGCCCTTTTTTAGTCTCAACCACATT | (GA)10 | 57 | 187 | KM874393 | – |

CCRM: CCMB CXR Microsatellite marker; $T_a$: annealing temperature; F: forward primer; R: reverse primer; –: Unmapped; CLG: Combined Linkage Group [1]; *: Expected amplicon size in the robusta variety CxR.

doi:10.1371/journal.pone.0113661.t002
| Sl No. | Marker ID | Unigene ID* | SSR repeat unit | Forward Primer | Predicted Tm (°C) forward primer | Reverse Primer | Predicted Tm (°C) reverse primer | Predicted amplicon size (bp) |
|--------|-----------|-------------|-----------------|----------------|-------------------------------|----------------|-------------------------------|-----------------------------|
| 1      | CCESSR51  | 119463 (AGAGCA)5 | AGCAGAGATCGAGACAGAG-AGA | 59 | ACATATCTGAAACC-CTCGGC | 59 | 477 |
| 2      | CCESSR52  | 119606 (CT)9 | AATCGGAGGATTTGTGCTTC | 59 | GGCCTAAATCAC-CATATT | 59 | 390 |
| 3      | CCESSR53  | 119736 (GA)11 | CCAGGGCTAGCTTACTACA | 58 | ACACAAATCAAC-CCATCC | 58 | 266 |
| 4      | CCESSR54  | 119787 (GGA)6 | CGCAGTCTGAATGAGGAGGAG | 59 | TGGGAGGTTGATCA-TCTGAA | 59 | 341 |
| 5      | CCESSR55  | 119820 (CT)11 | AGGACAGGAGTGTGATCCCT | 59 | GCCATGCTCTCTT-TGTAT | 59 | 230 |
| 6      | CCESSR56  | 119897 (TC)11 | AGGTTTTTGTGCTCCTTTTG | 58 | CATCGATGAAAAAGCA-GCAAT | 58 | 184 |
| 7      | CCESSR57  | 119919 (AT)15 | GAATTTGTCAGCCAGAGCA | 59 | GACGGAAAGATTCTG-GCTTT | 59 | 296 |
| 8      | CCESSR58  | 119987 (TCC)7 | AGCTACGCTAGGAATTGGT | 59 | GACCAAAACAAACCAG-CAACA | 59 | 418 |
| 9      | CCESSR59  | 120008 (TC)12 | GGAACAGACTCTCTTGGCA | 58 | TCACTACACAAAGAG-GCAAT | 58 | 198 |
| 10     | CCESSR60  | 120008 (TG)9 | ATTGCCTCCTTGTGATGTA | 59 | CCGGTGCTAACAAC-TCTTA | 59 | 232 |
| 11     | CCESSR61  | 120045 (TCT)6 | GTTAAAGGTTCTTTACTGTA | 59 | GAGATGCTACTTGG-CTACTGC | 59 | 320 |
| 12     | CCESSR62  | 120064 (GATT)5 | CTCTGTGTTTCAAACCAACC | 59 | CGGACACTTGAGAGGAGA-GAGAA | 59 | 432 |
| 13     | CCESSR63  | 120107 (TCT)6 | AGTCCAGTCCAGTCTCTGCC | 59 | CCGATATGATTGTG-GCTG | 59 | 241 |
| 14     | CCESSR64  | 120121 (GCT)7 | CGAAGTTGTCAGAGATGGAAC | 59 | GGAGCTGCTTGTCT-TCTTT | 59 | 203 |
| 15     | CCESSR65  | 120179 (AT)10 | GTGATGCTCGGTGTATCTGG | 59 | ACTAGAGGCGAGAA-TTGGA | 59 | 447 |
| 16     | CCESSR66  | 120206 (AGA)10 | GTAGAGCTTCAAAACAAAGCA | 59 | AAAACTTCCAGGCTC-CAGA | 59 | 273 |
| 17     | CCESSR67  | 120260 (CTCC)5 | GTGTCGTGTTTCTCTCTGAC | 60 | CCACACTACCGAGCA-AAAAAT | 59 | 469 |
| 18     | CCESSR68  | 120291 (AAG)8 | CCACGCGAATAATCATAAC | 59 | AAGCAGCTTATCCCCA-ACAG | 59 | 408 |
| 19     | CCESSR69  | 120316 (TCA)6 | TTGAGAAAACCATAAGAGGCGG | 59 | ACCCCTCATCAACTCTCT-TGCC | 59 | 343 |
| 20     | CCESSR70  | 120320 (GCCACC)4 | TTAATACGCCCCTCAACGG | 59 | CTTGAAATCTTGCGGCA-CTGT | 59 | 342 |
| 21     | CCESSR71  | 120322 (TTT)6 | TTGAGAAAGCTGGAGGAGCAGA | 59 | TTGACCACTTCCCCT-TC-TCT | 59 | 382 |
| 22     | CCESSR72  | 120543 (AT)9 | GATTGCTGCTTCTTCTGGG | 59 | TCCGCNGCAGTTAGATT-TTAT | 59 | 214 |
| 23     | CCESSR73  | 120545 (AT)9 | GTTTCCGCGCTAGCTTGGTC | 59 | GTGACATGAGGTGAAGA-TGGA | 59 | 485 |
| 24     | CCESSR74  | 120579 (CTT)7 | TGTAATATGCTTCTGGCCTG | 59 | AAGGGATGATGCCTAG-TTGG | 59 | 408 |
| Sl No. | Marker ID | Unigene ID* | SSR repeat unit | Forward Primer | Predicted Tm (˚C) forward primer | Reverse Primer | Predicted Tm (˚C) reverse primer | Predicted amplicon size (bp) |
|-------|-----------|-------------|-----------------|----------------|----------------------------------|----------------|-----------------------------------|----------------------------|
| 25    | CCESSR75  | 120596      | (TCC)7          | GGCTCAATCGATAAGCAACA | 59                    | CATGAACATATCCTGG-AGCG | 59                    | 429                        |
| 26    | CCESSR76  | 120656      | (AAC)6          | AAGCGAAGATGGGAGAGCAAT | 59                    | CGCCTTTTGTAGAAG- GAACC | 59                    | 316                        |
| 27    | CCESSR77  | 120656      | (TCA)6          | TTTTAGGCACTCTTCACC  | 59                    | GTATGTGAGGGCGTA-ACTGTG | 59                    | 454                        |
| 28    | CCESSR78  | 120720      | (GAA)6          | GAGCCCTTTTCTCCTCTT  | 59                    | CGTAAATGATAACAAC- CCCC | 59                    | 326                        |
| 29    | CCESSR79  | 120764      | (CATATA)4–13 bp- (AT)9 | AGCAGCACCAGACTAGCTCA | 59                    | GTCCCGAGAAACGAAG- ATGGT | 59                    | 399                        |
| 30    | CCESSR80  | 120821      | (ACAGG)4        | CAGATCCCATATTCCCAAC  | 59                    | TTGCGTGTCTGATTTCC- AAGC | 59                    | 324                        |
| 31    | CCESSR81  | 120883      | (TCAAT)4        | CAGATTCCGGTGTCAAACAG | 59                    | TTTGCTGCAGTGTCA-CCTGT | 59                    | 107                        |
| 32    | CCESSR82  | 120965      | (TCGAA)4        | GACTATGGATGGCTTTGCT | 59                    | GGGCGGATGGTTGAT-AAGAT | 59                    | 397                        |
| 33    | CCESSR83  | 121010      | (CCTGCA)4       | TGAATCAACTCTCTGCTCTG | 59                    | GGTAGTTCCAGCCAC- CCAGAT | 59                    | 115                        |
| 34    | CCESSR84  | 121086      | (TC)11          | GGATCAAACGTGCTAAGGT | 59                    | AAGAAAAAGGGCTAC- AACGGA | 59                    | 203                        |
| 35    | CCESSR85  | 121131      | (AAG)6          | CCCCGGGCTGCAAAAAACA- GT | 60                    | AGCTCCGGAATGACC- TCAATA | 59                    | 389                        |
| 36    | CCESSR86  | 121240      | (GCA)6          | AACCTCACCCTCTTTTCA  | 59                    | ACTATTTTGCTGCTG- CTGG | 59                    | 347                        |
| 37    | CCESSR87  | 121289      | (TTT)6          | GAAGAACCCTCTCTAAGAT  | 59                    | AGCACCTCTGCTCCAT- AATCT | 59                    | 345                        |
| 38    | CCESSR88  | 121377      | (GT)9           | CACGGGGAATACACCGG  | 59                    | AGTTGCGTCTGCTT- CTCAT | 59                    | 191                        |
| 39    | CCESSR89  | 121392      | (GCT)6          | GCACGTTTTTGAATGGTGG  | 59                    | TCCGCACTACAATGC- ACCGTC | 59                    | 159                        |
| 40    | CCESSR90  | 121439      | (GATTA)4        | TACTTGGACGCAGTCAACCG | 59                    | TAATCTGCGTGCTA- TTTG  | 59                    | 455                        |
| 41    | CCESSR91  | 121548      | (CTTCT)5        | AGGGCGCTAGGTGTTGGT  | 59                    | CAAAGTAACGCAAC- GGCATT | 59                    | 108                        |
| 42    | CCESSR92  | 121580      | (TC)11          | CCTACATCCACGACCTTCT  | 59                    | GGTAGTGCTTTTGGG- GTGGTG | 59                    | 361                        |
| 43    | CCESSR93  | 121610      | (AT)9           | CTGCCCTATGATGAATGGTG  | 59                    | TCAATCTCAAGCATCAG- TCTCC | 59                    | 157                        |
| 44    | CCESSR94  | 121752      | (TGCTCC)4       | CTTCTCATGGACGCAACTAA | 59                    | AAGGCTAGTGGGAACTG- GAAAG | 59                    | 258                        |
| 45    | CCESSR95  | 121841      | (CT)10          | TGAGCGAGAGATTGCAAGG  | 59                    | CCAGCTTAACCTGC- CCTG | 59                    | 349                        |
| 46    | CCESSR96  | 122004      | (AT)19          | ACATTCGAAAACCTTGGGGAG  | 59                    | CAAGCTCTTGAGTTTGGT- TCTCG | 59                    | 232                        |
| 47    | CCESSR97  | 122006      | (AG)14          | CTCGTGCTGTAACCCCTTCA  | 59                    | AGTGTGATGAAAC- GGGATA | 59                    | 439                        |
| 48    | CCESSR98  | 122012      | (AGC)6          | GACGCCAGTCTTTCAAGTA  | 59                    | GCTCTCAAGTCTGCTT- TCCAA | 59                    | 373                        |
| Sl No. | Marker ID   | Unigene ID* | SSR repeat unit | Forward Primer                  | Predicted Tm (˚C) forward primer | Reverse Primer                  | Predicted Tm (˚C) reverse primer | Predicted amplicon size (bp) |
|-------|-------------|-------------|-----------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|-----------------------------|
| 49    | CCESSR99    | 122077      | (CGC)6          | TCGGACTGTACTAGTGACG              | 59                               | TCCTCCAGAGCTTA-GCCTT             | 59                               | 215                         |
| 50    | CCESSR100   | 122089      | (AG)9           | TTCTGCAATTCTCTGTA               | 59                               | AGGGGTCAAGAAGA-GGGATT           | 59                               | 214                         |
| 51    | CCESSR101   | 122184      | (AT)10          | AGGGGTGATTCAGTAGACG             | 59                               | ATGCAATTGACCCGAG-CTCAG          | 59                               | 466                         |
| 52    | CCESSR102   | 122195      | (AGGCTC)6       | AGTCAGCTTCTTGCTGCTA             | 59                               | CCCTCCAGAGTACC-ATTTG            | 59                               | 336                         |
| 53    | CCESSR103   | 122254      | (TCGT)5         | GGAATGAGCTAAAGCCAG              | 59                               | TTACCGGTAGTTGT-GTGA             | 59                               | 318                         |
| 54    | CCESSR104   | 122257      | (TTCA)5         | GTCAGTGAGGTCTGAAAACA            | 59                               | ACCGTGAGTTTGAG-GGAA             | 59                               | 310                         |
| 55    | CCESSR105   | 122324      | (AGA)6          | AGGGGCTAGTGGTGAAACAA            | 59                               | ACAAGCAGACGAAC-ACCA             | 59                               | 319                         |
| 56    | CCESSR106   | 122340      | (TC)10          | AGCTGATGAGGATGGCT                | 59                               | CATTGCTTTCCTGAC-ATCT            | 59                               | 484                         |
| 57    | CCESSR107   | 122383      | (CTT)8          | CTTTCCTCCCTTCTCCTT              | 59                               | TGAGTGACAGCCTG-AACAAA           | 59                               | 459                         |
| 58    | CCESSR108   | 122387      | (CTT)7          | GTGGGAAATCTCAAATGTGCT          | 59                               | CCTTCTTCTGGTGG-GTGTGA           | 59                               | 143                         |
| 59    | CCESSR109   | 122514      | (AGG)6          | GGAAGACCCAACAAACCAAGA           | 59                               | GGAGTAGCATTCAACA-TGGGAC         | 59                               | 340                         |
| 60    | CCESSR110   | 122533      | (TA)10          | GTGCCGTGTTGTTTTCTCAT            | 59                               | GTGCACTGAGTGGAT-TCAGG            | 59                               | 324                         |
| 61    | CCESSR111   | 122550      | (TCA)6          | TTCATGGACCATTCTCAG              | 59                               | TAGCCCAAGACAGAG-CATTG            | 59                               | 290                         |
| 62    | CCESSR112   | 122619      | (GAA)6          | TCCCCGAAATGAGGCGTTCTCT          | 59                               | ACAAGGCAACAGGT-TTACC            | 59                               | 452                         |
| 63    | CCESSR113   | 122646      | (AGC)6          | GCATCCTGTATTGCAGGAGGT            | 59                               | ACCACTCGCTTCAAC-CTCTT            | 59                               | 115                         |
| 64    | CCESSR114   | 122681      | (AAAG)5         | CCAAGCTAAGTGCTCATC              | 59                               | GGCAAGCAGTAA-GTGCA              | 59                               | 209                         |
| 65    | CCESSR115   | 122758      | (TC)13          | TTCCATTTTCTCTTCTTCTT            | 59                               | TAGGGTTTTGAAAG-GCAATC           | 59                               | 274                         |
| 66    | CCESSR116   | 122793      | (TC)18          | GAGAAGTTTGGCAGAAACCT            | 59                               | ATCAACCTTCAAG-GGCCAC            | 59                               | 453                         |
| 67    | CCESSR117   | 122797      | (CT)12          | CTCGTCCTCTCCCTTGACT             | 59                               | ACCACAAACAGCGCAGGACATCAT        | 59                               | 160                         |
| 68    | CCESSR118   | 122811      | (AAG)7          | CTTTCTCAGCTTCCACCACA            | 59                               | CCGCCTGTGTCTGTGT-CCTGTGTT       | 59                               | 371                         |
| 69    | CCESSR119   | 122842      | (ATA)6          | TGCAAGTTGTGATGAATGTTG            | 59                               | CTAAGGCCAGAAAA-CAAGGG           | 59                               | 431                         |
| 70    | CCESSR120   | 122880      | (GAAGCA)4       | TAGGGGGAACAAATAGCAAAG            | 59                               | AGACGCTTGCGAC-AAAAGTC            | 59                               | 212                         |
| 71    | CCESSR121   | 122996      | (TTA)6          | CCGGACAGGAACAGAAGAAT            | 59                               | CATCGAATTGCGCTG-ACCATC          | 59                               | 443                         |
| 72    | CCESSR122   | 123010      | (TCC)6–30 bp-(TCA)7 | GTGTCATCTGCTTCTGCTACT          | 59                               | AAAAAAAAAAAAAAACATCC-AGCCAG       | 59                               | 438                         |
Table 3. Cont.

| Sl. No. | Marker ID   | Unigene ID* | SSR repeat unit | Forward Primer | Predicted Tm (˚C) forward primer | Reverse Primer | Predicted Tm (˚C) reverse primer | Predicted amplicon size (bp) |
|--------|-------------|-------------|-----------------|----------------|----------------------------------|----------------|----------------------------------|-----------------------------|
| 73     | CCESSR123   | 123023      | (CTG CCT)4      | GACAGAGGAATCTTGGCAT | 59                      | TCCCAAGAAAATCACCTTTCC | 59                      | 494                          |
| 74     | CCESSR124   | 123110      | (A ACTA)4       | GAGAAGGCCAACACACTTGA | 59                      | ACTTGAGCAGCAACCCCTTT   | 59                      | 183                          |
| 75     | CCESSR125   | 123132      | (CT ATGA)4      | GAGCGCTATGGATGAGCATA | 59                      | GCCTGACAGGATTGGAAGCT   | 59                      | 478                          |
| 76     | CCESSR126   | 123134      | (CT G)6         | ACCACCAAGCAACACAAAT | 59                      | TGACCCCCAAACAAACAAAG    | 61                      | 457                          |
| 77     | CCESSR127   | 123160      | (T A)11         | CTGATGGCGTCATCAATAC | 59                      | GGGAAAGGGAGTCTCAAAACA  | 59                      | 394                          |
| 78     | CCESSR128   | 123185      | (ACG)6          | GAGGAAATCCGGAGGATGGA | 59                      | CATCAGTCATCATCATGTCGA  | 59                      | 280                          |
| 79     | CCESSR129   | 123240      | (TT C)6         | CCAGAAACAACTAGGTTT   | 59                      | TCATTTCTGGCAGAAAGGC   | 59                      | 498                          |
| 80     | CCESSR130   | 123248      | (GCA)6          | TCTTCTACAGCTGCCACAC  | 59                      | ATTCGCGAAGAATGAAGGGTC  | 59                      | 137                          |
| 81     | CCESSR131   | 123291      | (CT)10          | CTTCTCCAACGGACAAAC   | 60                      | CTCAGGCTCTAATGCCACTAA  | 59                      | 172                          |
| 82     | CCESSR132   | 123309      | (AT)20          | CTGAAATTAGGGGGCTTGGT  | 60                      | TTGACCAAAACACTGATGSTA  | 60                      | 227                          |
| 83     | CCESSR133   | 123323      | (AC AT)5        | TCTCCCATGGTACTCCACA  | 59                      | CCGATGCATTTCAGACATA    | 59                      | 463                          |
| 84     | CCESSR134   | 123329      | (TC)13          | TCACCTGCGTGGGCTTTATG | 59                      | GAGCATGTCGCCGTGTAAGAT   | 59                      | 380                          |
| 85     | CCESSR135   | 123332      | (AG)10          | TTGTCTCTCCAGGGGAAGTC | 59                      | CGTGAACCAATCCAGACACA   | 59                      | 210                          |
| 86     | CCESSR136   | 123342      | (AC)9           | AATATGAGGGGAGGCAAGG  | 59                      | ACAAGTGCCGAACTGTTGA    | 59                      | 347                          |
| 87     | CCESSR137   | 123362      | (AA T)6         | GGTCCTGAAAGTGCCAAAAT | 59                      | GGAGGATACTGCAGAGGC TT  | 59                      | 496                          |
| 88     | CCESSR138   | 123381      | (CGG)6          | AAACCTAGTGTTGGAGGTTGG | 59                      | CAGCTTTGTTCCACGTCCTAA  | 59                      | 240                          |
| 89     | CCESSR139   | 123445      | (TC)9           | AATCTTGGGGAGAAATGCC  | 59                      | GCTTTTTGGTTGCGATTCCTC  | 59                      | 349                          |
| 90     | CCESSR140   | 123662      | (CT)9           | CGAGGCTCTATCCTCTCTCTC | 59                      | CTGACGCGATCTGACCTCACT  | 59                      | 182                          |
| 91     | CCESSR141   | 123682      | (CT C)7–10 bp-  | AGTACCTGCAACACCCCAAC | 59                      | AGGCTCTGCAAAATCAGGT    | 59                      | 479                          |
|        |             |             | (TT C)7         | AGTACCTGCAACACCCCAAC | 59                      | AGGCTCTGCAAAATCAGGT    | 59                      | 479                          |
| 92     | CCESSR142   | 123741      | (CAGCAA)4       | GAGATTGTTGAGCTTGAAGGC | 59                      | TTCCCGCTATCTGAACTTTG   | 59                      | 478                          |
| 93     | CCESSR143   | 123742      | (TGT A)4        | AACCAGTTTGGATTCCAGGC  | 59                      | CTCCACTCCACTGGTTTTTT   | 59                      | 294                          |
| 94     | CCESSR144   | 123743      | (TTGT A)4       | AACCAGTTTGGATTCCAGGC  | 59                      | AAAAGGTCGAAACGTCAGCC   | 59                      | 269                          |
| 95     | CCESSR145   | 123795      | (GCG)8          | AGGCTTGAAGGCGAAATGCT  | 59                      | GCCTTTTTGAGTCTGACCCTCC | 59                      | 309                          |
| 96     | CCESSR146   | 123812      | (AT CTTT)4      | ACACCTGCAGGCTTTGACC  | 59                      | AGTGATCATCCATCCGTTCCA  | 59                      | 482                          |
| SI No. | Marker ID  | Unigene ID* | SSR repeat unit | Forward Primer | Predicted Tm (˚C) forward primer | Reverse Primer | Predicted Tm (˚C) reverse primer | Predicted amplicon size (bp) |
|-------|------------|-------------|-----------------|----------------|---------------------------------|----------------|---------------------------------|-------------------------------|
| 97    | CCESSR147  | 123819      | (CTTC)5         | ACGAGGGCAAAACAAAGAACG | 59                      | TCAAGACGAAGAATG- CAGG | 59                      | 112                           |
| 98    | CCESSR148  | 123822      | (GCT)6          | ATCCCGCAAGACTATTTGTTT | 59                      | ACGAATTCTTGTCCA- CTCC | 59                      | 133                           |
| 99    | CCESSR149  | 123903      | (CCG)6          | GGCTCTGAGGCTTTGAATA | 59                      | GGCGAGCATGATTAG- ACAGA | 59                      | 256                           |
| 100   | CCESSR150  | 123909      | (CT)14          | CTTCAGGTTCTTGTCTCTC | 59                      | CCATCGAGCATTTCCAT- CATC | 59                      | 389                           |
| 101   | CCESSR151  | 123967      | (TTTC)5         | CTCTTCTCATTTGTCTGC | 59                      | AGGGTTTTCTGTGGCTG- TTCT | 59                      | 374                           |
| 102   | CCESSR152  | 123993      | (TG)10          | GCCTAGTACTCCCAAATCCA | 59                      | ATGATGGATGATTGACCT- TCCA | 59                      | 251                           |
| 103   | CCESSR153  | 124161      | (CT)12          | TCTGGGCTCCAAAATTCAT  | 59                      | ACATGCGGTTTATGCAG- TCTAT | 59                      | 453                           |
| 104   | CCESSR154  | 124171      | (CCA)6          | ACCACACCTCAGAGAATC  | 59                      | CGTAGAAATTGCTATCCG- GCA  | 59                      | 421                           |
| 105   | CCESSR155  | 124195      | (AGC)6          | ATCCCCATCAGAGACCTCA | 59                      | GAAGCTCGACAAAACAGGACA | 59                      | 296                           |
| 106   | CCESSR156  | 124198      | (GAAC)4         | CCACAGAGAAATCCCCAGTG | 59                      | GACACATCGAGGCC- AATAAC | 59                      | 240                           |
| 107   | CCESSR157  | 124268      | (TGC)6          | TTCTTCTCACAACGCCAAGA | 59                      | GCTGATGATGCGAT- CAAAGT | 59                      | 406                           |
| 108   | CCESSR158  | 124269      | (CCT)7          | AGTCGCTTCTTTGGTGCTT  | 59                      | GAGGAGAGGTGG- TGGTG | 59                      | 195                           |
| 109   | CCESSR159  | 124309      | (AAAGG)4        | TCTGAGTCTCTGTGGCTTG | 59                      | TGTTGTATGGAACACAGGA | 59                      | 345                           |
| 110   | CCESSR160  | 124355      | (AAAA)4         | TCACGGTCCTTCCTCTACCT | 59                      | GCTCGTCTCCTCTTGAG- CTCC | 59                      | 359                           |
| 111   | CCESSR161  | 124358      | (ACA)6          | ATGAGGATGAGGAGGATTG  | 59                      | TCCCTCAAGCTGCTTT- TCCT | 59                      | 407                           |
| 112   | CCESSR162  | 124358      | (AAG)6          | AGCTTCTCTGCGAAACCTCA | 59                      | ATCATCATCAGGGCCT- CCT | 59                      | 147                           |
| 113   | CCESSR163  | 124365      | (AAC)6          | TCTGCTCTCAGCTTATCCCAGCG | 59                      | CTTCACGAAGATGCTCT- AGAA | 59                      | 171                           |
| 114   | CCESSR164  | 124426      | (TTTTTA)5      | GACGTGGAGGGAGATGGTTT | 59                      | CAATTCGAGCTTT- ATCAG | 59                      | 140                           |
| 115   | CCESSR165  | 124428      | (TTTTTA)4      | GACGTGGAGGGAGATGGTTT | 59                      | TGAATATGACCCGCA- GAGT | 59                      | 239                           |
| 116   | CCESSR166  | 124436      | (GAG)6          | CGAGTACCATTTGAGTAGAT | 59                      | AGTGTACCATTTGAGTGC- TCCG | 59                      | 304                           |
| 117   | CCESSR167  | 124457      | (AC)9          | CAAAGAAAAACCGCTTCTCG | 59                      | AGTGGAGGATTGG- AAACG | 59                      | 367                           |
| 118   | CCESSR168  | 124661      | (ATA)6         | CTCTGCCCCTCTTTCTTCTC | 59                      | CACTTGGGCATCTGA- GAAGA | 59                      | 354                           |
| 119   | CCESSR169  | 124465      | (TTCA)6        | TCCCCCATTTGATTTGAGTAGT | 59                      | GCAAGGAGACCTCAGA- GAAGG | 59                      | 296                           |
| 120   | CCESSR170  | 124468      | (TTCTCTC)4–24 bp- (ATC)6 | TCTCCACAGCAATGCAGTTGG | 59                      | GTAAGTGGCTGAATGAT- GAGG | 59                      | 342                           |
| Sl No. | Marker ID | Unigene ID* | SSR repeat unit | Forward Primer | Predicted Tm (°C) forward primer | Reverse Primer | Predicted Tm (°C) reverse primer | Predicted amplicon size (bp) |
|-------|-----------|-------------|----------------|----------------|---------------------------------|----------------|---------------------------------|-------------------------------|
| 121   | CCESSR171 | 124565      | (ATT)6         | GCTCTGGTCAGGAAGGGAAG | 59 | GTATCAGAATCCCCAC-CCCA | 60 | 406 |
| 122   | CCESSR172 | 124577      | (AAG)6         | TGGCTTTTTCTCGGTATCCT | 59 | TTCTTTCTGGTGCTTG-ATCG | 59 | 293 |
| 123   | CCESSR173 | 124612      | (CGG)6         | TCTTTTCTAGCCCTCTGGA | 59 | AGAGAAAAAGAGCGG-AGAG  | 59 | 331 |
| 124   | CCESSR174 | 124697      | (GAC)6         | GAAAATCGAGAAAGGGTAAGAT | 59 | GAGTGTTCTCCAAGT-CCCAAT | 59 | 242 |
| 125   | CCESSR175 | 124700      | (GCT)6         | ACACCAACGACGGGTAAGAT | 59 | CATCTCAATCAAATGCGCCAC | 59 | 346 |
| 126   | CCESSR176 | 124759      | (AG)11         | GGGATTTGCTGATAGGAAT | 59 | ACCTTCAGTCTCCG-TCAC | 59 | 444 |
| 127   | CCESSR177 | 124765      | (CCA)7(AA-C)6 | CTGCAGTCTCAGACATCCACT | 59 | ACCTGTGCTCTCCG-CCAC | 59 | 369 |
| 128   | CCESSR178 | 124767      | (TCAT)5        | GCCGCTAGTCCTATACGGAG | 59 | TCTCTCAAAAGCGGAT-TCCT | 59 | 354 |
| 129   | CCESSR179 | 124817      | (TC)9          | TCATTGCTGCTTCAGTCTGC | 59 | TCTCTCAAAAGCGGAT-TCCT | 59 | 354 |
| 130   | CCESSR180 | 124851      | (TTC)11        | GACTTCTCCTCCCTCTCCCTC | 59 | TTGGACGCTAATGCTG-AGG | 59 | 422 |
| 131   | CCESSR181 | 124871      | (AGC)6         | ACCAATTTCAAAACCTCCAGC | 59 | CGAACACCATGCACTG-ACA | 59 | 475 |
| 132   | CCESSR182 | 124949      | (CCT)7         | CCCCTTAAATCGATCCCTCCT | 59 | CGGAGAGGAAATTTCG-AGCAT | 59 | 328 |
| 133   | CCESSR183 | 124969      | (AATTG)4       | TCTGGCTTGAAGGGGATGTT | 59 | TAAGCTCTAAGCAAAG-AGCA | 59 | 243 |
| 134   | CCESSR184 | 124995      | (GAGAAG)4      | CACCAAGCCATGAAAGCTA | 59 | TCCCAAATTTCCTCCTC-TCATC | 59 | 224 |
| 135   | CCESSR185 | 125025      | (TC)11         | CCTCTTTAAAACCTTAAACC | 59 | TCTAGATTTTTGGTGGGGGC | 59 | 302 |
| 136   | CCESSR186 | 125054      | (TA)9          | TCTTGCTTGCACATTTCCC | 59 | GGTCTCCGTTGGTAGGATT | 59 | 283 |
| 137   | CCESSR187 | 125099      | (ACC)6         | TTCCCGCTTTACAGATACC | 59 | TTCTCTCAACAGAC-ACAC | 59 | 388 |
| 138   | CCESSR188 | 125102      | (ACC)6         | TTCCCGCTTTACAGATACC | 59 | AAACATAAGGTGGG-AGTAGG | 59 | 281 |
| 139   | CCESSR189 | 125104      | (AT)11         | CTGGCTTGCATGCAATT | 59 | TTCAGAAGGCGGTAG-TCGGT | 59 | 387 |
| 140   | CCESSR190 | 125128      | (TC)12         | TCACCTTTCTTCTCTCCTCT | 59 | CTGGAAAGGTTGAAAAGGGA | 59 | 418 |
| 141   | CCESSR191 | 125139      | (CAGGG)4       | AGCACATCCCAACACACAG | 59 | AACCCACGTTTAGGAGGACAC | 59 | 204 |
| 142   | CCESSR192 | 125151      | (ACA)8         | ATGCCGAATCGGTAGTACAGAG | 59 | AATGTACATCCCTCC-CCAC | 59 | 400 |
| 143   | CCESSR193 | 125186      | (AAG)6         | GCTCCAACACTTGGCTATT | 59 | CGTGACGATGAGCCC-CCAC | 59 | 372 |
| 144   | CCESSR194 | 125232      | (AT)13         | ATTCAGTGGACGATGGGAG | 59 | ATGATTGGAGAGAGGACAG | 59 | 314 |
| Sl No. | Marker ID   | Unigene ID* | SSR repeat unit | Forward Primer | Predicted Tm (˚C) forward primer | Reverse Primer | Predicted Tm (˚C) reverse primer | Predicted amplicon size (bp) |
|-------|-------------|-------------|----------------|----------------|---------------------------------|----------------|---------------------------------|------------------------------|
| 145   | CCESSR195   | 125265      | (GT)14         | TGATGATAGCTCCGCTTTTG | 59 | AGCCAATCACCCAGCA-AGATT | 60 | 254 |
| 146   | CCESSR196   | 125286      | (CTC)8         | GGCTCCTATGTCGCTTTTTT | 60 | GAGACCAAGCTTGAA-GCAA | 59 | 304 |
| 147   | CCESSR197   | 125289      | (ATT)5         | GAGGATCATTCTGACCCTCCT | 59 | GAAAAGGTCAAGCACATCCCC | 56 | 470 |
| 148   | CCESSR198   | 125382      | (AC)16         | CTCTCCCATCCCTCAACT | 59 | GTGAAGGGAGGTTGAGGTTTCT | 59 | 365 |
| 149   | CCESSR199   | 125416      | (AT)16         | ATCATCCCTGCATAGCAAC | 59 | TCCCCGAGGAAAAGAA-TGAG | 59 | 183 |
| 150   | CCESSR200   | 125444      | (TG)11         | TGTTGTAGGTTGGTGTCTCAT | 59 | GCTCGGGGATATCGATCTTA | 59 | 495 |
| 151   | CCESSR201   | 125484      | (CTC)G4        | AGGGTTTCGCGAGTAGGAC | 60 | CCGGGAAACTAGAG-GCAAAG | 59 | 424 |
| 152   | CCESSR202   | 125529      | (TC)9          | GTGCTGTGCAATTGAAACT | 59 | ATGCCCGAGTGGATGTCATG | 59 | 490 |
| 153   | CCESSR203   | 125676      | (AAACAA)4      | TGATCTTGATCCCCTCATCA | 59 | GATGCATTCCAAAAC-GGTC | 59 | 393 |
| 154   | CCESSR204   | 125734      | (GAG)6         | TTCTGGTCTGAGATCTTTT | 59 | TTTGTGCTCCTTCTC-GGAC | 59 | 434 |
| 155   | CCESSR205   | 125776      | (CT)14         | AGGATTGCCTCCGCTTTTGTA | 59 | GATTTAATCGTGAGGAA-CCTT | 59 | 225 |
| 156   | CCESSR206   | 125893      | (CAT)7         | CCCCTTTTCTCTTTTTGTG | 60 | TGCACTATCCCTTGTATCGT | 59 | 271 |
| 157   | CCESSR207   | 126139      | (GGC)6         | TCCGTTCCCTGCTTGTTT | 59 | CAGAAGACCATGGGA-TAACA | 60 | 411 |
| 158   | CCESSR208   | 126219      | (GAG)6         | TGGTCTTCAATCACCAGGA | 59 | GACACATTCGACTGAGTCCC | 59 | 134 |
| 159   | CCESSR209   | 126243      | (CAC)8         | GGTGAGCAGGAGCTTTTATGT | 59 | CGGAGCAAAACTTG-AACGCC | 59 | 180 |
| 160   | CCESSR210   | 126250      | (CAC)9         | GCCAAAAATCCCTTCTCCTCAT | 59 | TGGTGGATGACCAGA-GTCC | 59 | 245 |
| 161   | CCESSR211   | 126255      | (GAA)6         | ACGAAGGGGTTGAAGGTGAC | 59 | GACAGAGACGACGAA-GACCA | 59 | 412 |
| 162   | CCESSR212   | 126427      | (ATC)7         | TCAATGTTTTACAGGAGCA | 59 | TCTCCGGTGAAACT-GCTG | 59 | 341 |
| 163   | CCESSR213   | 126427      | (CAGGAG)6      | GGGCGGCTTCCGCTATTATTA | 59 | GCACATAGTTCTTGGAGTGCC | 59 | 184 |
| 164   | CCESSR214   | 126506      | (TTTC)6        | TTCTTGCTTTTCCGAGCCAG | 59 | CCCCATCCTACTTCTCC | 59 | 339 |
| 165   | CCESSR215   | 126540      | (AAG)9         | AAATCGAATGGCTTGGGAGAC | 59 | GTGCAAGTGAAAATCGC-GACA | 59 | 309 |
| 166   | CCESSR216   | 126661      | (AGT)6         | TTTGTCATTTCTTCTGCTG | 59 | AGTTGGGGGAGATTG-CCTG | 59 | 440 |
| 167   | CCESSR217   | 126701      | (TTCA)5        | AAAGGATACGGGCTCACAAGG | 59 | GAATGAATGAACGAGG-GGTT | 59 | 124 |
| 168   | CCESSR218   | 126701      | (TTCA)6        | AAGGAGGCAATCGTTCATT | 59 | TACCCCGTGATTTGCAGG | 59 | 175 |
Table 3. Cont.

| SI No. | Marker ID    | Unigene ID* | SSR repeat unit | Forward Primer | Predicted Tm (˚C) | Reverse Primer | Predicted Tm (˚C) | Predicted amplicon size (bp) |
|--------|--------------|-------------|-----------------|----------------|------------------|----------------|------------------|-----------------------------|
| 169    | CCESSR219    | 126730      | (CT)10          | GAGTTGGCTTGAAAGCATGA | 59               | CCAATCCCCCAAGGAAAGTGTATA | 60               | 371             |
| 170    | CCESSR220    | 126767      | (GA)9           | GCACAAAGACCTCCCTTTGT | 59               | CTCTTGACGTGCTTTTGT | 59               | 428             |
| 171    | CCESSR221    | 126881      | (GCA)6          | CGGTGCTGCAAGTCCTACAC | 60               | CAAGGCACAGAATCGGA | 59               | 351             |
| 172    | CCESSR222    | 126910      | (AAT)8          | CCAACAGAAAAAGTGGGGAAC | 59               | TTATCCGAAATCCTTTTCCC | 59               | 163             |
| 173    | CCESSR223    | 127055      | (TTC)6          | GAAATCCACTCGTGAAAGT | 59               | CCAATCCCCCAAGGAAAGTGTATA | 60               | 401             |
| 174    | CCESSR224    | 127171      | (AAGG)5         | ATTCAGCTCTGGGTTAGCAAG | 60               | TCTCTTGCTGCTTTTGGTCTG | 59               | 470             |
| 175    | CCESSR225    | 127376      | (GT)10          | TCCACAAACAAACGAGACA | 59               | CTATCCCTACGCGCTTCACA | 59               | 270             |
| 176    | CCESSR226    | 127478      | (CT)23          | CAGCCCTCCGTTAGCAAG | 60               | TGCCTCAATAACCTCCTGGGCT | 59               | 179             |
| 177    | CCESSR227    | 127479      | (ACC)10         | GTACCCCCACAACACTCACA | 59               | GATGCTTCTCATCGTCACCACAA | 59               | 404             |
| 178    | CCESSR228    | 127712      | (TTA)5          | GATTGTTGCAAAAGGGAAG | 60               | GCCCTGACGGAGAGATGTCAT | 59               | 115             |
| 179    | CCESSR229    | 127774      | (CT)6           | CCGCCGTTAATCCAAACCTCT | 59               | AAATCCAAACAGCACTCCTCTGC | 59               | 120             |
| 180    | CCESSR230    | 127785      | (AT)13          | GCTGTCGTGTTGCTTTTCTC | 60               | AGCTTCGAAATGTCATGTCAT | 60               | 333             |
| 181    | CCESSR231    | 127800      | (GAA)9          | GCTACAGCGTGCAGAAAGAAG | 59               | TACTTCTCAGGGCTTCATTTT | 59               | 318             |
| 182    | CCESSR232    | 127828      | (TCA)6          | ACCTGCGAGGACTGATAAGC | 59               | CCAATCCGATCCATTTTGCCTG | 59               | 177             |
| 183    | CCESSR233    | 127910      | (AAAAGG)4       | GAAGACTCTGTAAGAGGACGAGACC | 59               | GTGTATGGCTCAACACCATGC | 59               | 347             |
| 184    | CCESSR234    | 127938      | (CTG)8          | GCCGTCGTAGTTCTATCTCTCAG | 59               | ATGGCTGACAGCCTCCCTTCTCTT | 59               | 442             |
| 185    | CCESSR235    | 127992      | (TAA)6          | AGTGGGAAAAGTGAAGCTGG | 59               | ATTAATCCCTCCTCCCTAACGAT | 59               | 225             |
| 186    | CCESSR236    | 128005      | (CTC)8          | AGTCAGGTATGCGTCCATTTG | 59               | GGAATCTTTGGAAGAGATGTCAA | 59               | 322             |
| 187    | CCESSR237    | 128130      | (CCT)6          | GTCTCAATAATGTTGATTGC | 59               | GGGAGAATGGGAGAGTGAAGA | 59               | 142             |
| 188    | CCESSR238    | 128268      | (TA)11          | CCTGCTGCTATGGCAAAGAAGA | 59               | ACAGAGGGGTTTCCACACAGG | 59               | 254             |
| 189    | CCESSR239    | 128301      | (AAGAAC)4       | TCCCAAGTCCTCTATTGTCGC | 59               | CAGCTGAGGTGTTTACGAGCA | 59               | 123             |
| 190    | CCESSR240    | 128306      | (TCT)7          | TCTGCTGCTCCTTTCAGCAG | 59               | GAGATTGTCATACGAGTATGAG | 59               | 221             |
| 191    | CCESSR241    | 128309      | (CGG)8          | ATATCATAGGATGTGGCTGCC | 61               | TCAGGATTTGCTGCTTTCCT | 59               | 355             |
| 192    | CCESSR242    | 128315      | (TTC)6          | GCACGAGGCTTCTCTCTTCT | 59               | CACGAGGGAAACTGGAGAGTCG | 59               | 260             |
Table 3. Cont.

| Sl No. | Marker ID   | Unigene ID* | SSR repeat unit | Forward Primer | Predicted Tm (˚C) forward primer | Reverse Primer | Predicted Tm (˚C) reverse primer | Predicted amplicon size (bp) |
|--------|-------------|-------------|-----------------|----------------|----------------------------------|----------------|----------------------------------|-----------------------------|
| 193    | CCESSR243   | 128377      | (AG)12          | GGTCTTCTGTGTGGGTGTG | 59 | GTGAAGCCAGCTAAT-CCGGCC | 59 | 107 |
| 194    | CCESSR244   | 128432      | (TG)15          | CACACACCAACACACAGAAG | 59 | GCGGAGGGAATTTTGAGCCAGAG | 60 | 482 |
| 195    | CCESSR245   | 128504      | (GAG)6          | TCTATCCCATATCCCATCTCTT | 59 | GTGCGGGTCTCTTCCTTCTTGGC | 59 | 454 |
| 196    | CCESSR246   | 128527      | (GA)10          | CACACACAAATCTTCTATCCC | 59 | TCCATCTCCAACACTACCA | 59 | 427 |
| 197    | CCESSR247   | 128527      | (CTG)6          | CACACACAAATCTTCTATCCC | 59 | TCCATCTCCAACACTACCA | 59 | 427 |
| 198    | CCESSR248   | 128527      | (CTG)6          | TGGGTAGGTGGAGATGGA | 59 | CATTGAACCGCTGTATCTG | 59 | 260 |
| 199    | CCESSR249   | 128550      | (ATC)6          | CATAGTCATGGTCACCCA | 59 | TTCTTCGAGGAGCTCTGTA | 59 | 416 |
| 200    | CCESSR250   | 128587      | (TC)11          | GAGGGCAACACCAATGAGAT | 59 | ATGGAAGGCGCTGCTGACC | 59 | 358 |
| 201    | CCESSR251   | 128673      | (CT)9           | TTTCCTGTGTGCTGTCTT | 59 | GGCACACATTATCACTCG | 59 | 245 |
| 202    | CCESSR252   | 128894      | (AAAAG)4        | GCCCAAGGCTTGTAGAGACC | 59 | TCTCTGTATGACTTCTTTGCTC | 59 | 344 |
| 203    | CCESSR253   | 129060      | (TCT)7          | CAATTCAGCTAGCTCCGT | 59 | ACATCTTGCCACTGCTTGGAGC | 59 | 216 |
| 204    | CCESSR254   | 129129      | (GAG)7          | ACAGATGCTTTTATGGAGG | 59 | GGGAAGGTATTGTGGCTGGTA | 59 | 272 |
| 205    | CCESSR255   | 129269      | (GCA)9          | TGAAGGTCCATGATTGGA | 59 | TTGTGGTGTGTAACTGCTGCTG | 59 | 477 |
| 206    | CCESSR256   | 129431      | (AGG)9          | CGGCACACATGACAGTAAG | 59 | TCCGACATATCGAAACATTGAGCTAT | 59 | 496 |
| 207    | CCESSR257   | 129439      | (CA)10          | AGTCACCACACAGATGCTGC | 59 | GCCACTTCGAAGGTTCTTTTGT | 59 | 296 |
| 208    | CCESSR258   | 129524      | (GAA)8          | CTAATCTTCAACAGCCTCCTCC | 59 | TTTGTTACAGCGAAAGGTGTC | 59 | 412 |
| 209    | CCESSR259   | 129581      | (AAAAG)4        | CCAACCTCAGGTCTACACTA | 59 | GCCAGTGACAGACAGACACTTTT | 59 | 102 |
| 210    | CCESSR260   | 129646      | (TC)11          | CTTCAAGCAACACAGATGCT | 59 | AAAAGAGCAGCAAGCAGTAGG | 59 | 179 |
| 211    | CCESSR261   | 129668      | (CT)12          | CATGATTAGCTGCTTGGCTTC | 59 | GTGTTGATGCTGTTGAATGAGG | 56 | 163 |
| 212    | CCESSR262   | 129745      | (CAC)10         | GCCTACAGCTCAAAACAAA | 59 | TGTGTTCTTCCTTCTCTTGTC | 59 | 252 |
| 213    | CCESSR263   | 129793      | (CA)10          | TAGCAGGGAAGATTGATAGG | 59 | AGGGGTTTTCTCCACTTCTT | 59 | 487 |
| 214    | CCESSR264   | 129797      | (TCC)6          | TCAGATGAGGTCTACAGCTTCC | 59 | TGGCTCTGAGAAACACTGCTGG | 59 | 312 |
| 215    | CCESSR265   | 129906      | (TA)11          | GACCGAGAGAGAGACACCTACTTTT | 57 | CTCTTCTAGCAGCAGACAGA | 59 | 186 |
| 216    | CCESSR266   | 129913      | (CAG)6          | TCTCTTTAAAACCGTCAGGG | 59 | CTGCTAGCTCTTGTCTCT | 59 | 264 |
| SI No. | Marker ID   | Unigene ID* | SSR repeat unit | Forward Primer                      | Predicted Tm (°C) forward primer | Reverse Primer                      | Predicted Tm (°C) reverse primer | Predicted amplicon size (bp) |
|-------|-------------|-------------|-----------------|------------------------------------|----------------------------------|-----------------------------------|----------------------------------|-----------------------------|
| 217   | CCESSR267   | 129917      | (AAGAGC)5       | GAGATGGAGAGACCATCATTCC             | 58                               | AACTAGACTTTTGCCTGCGT             | 59                               | 267                         |
| 218   | CCESSR268   | 129950      | (TCCACT)4       | TTCCCATCTAAGGAACAAATGGT          | 59                               | CTATATCAGGGCTCTCTTATT             | 59                               | 403                         |
| 219   | CCESSR269   | 129957      | (GAGCAG)4       | AACCCACCTTCTCCTCTT              | 59                               | CATGCATATCGGACACAAAC             | 59                               | 294                         |
| 220   | CCESSR270   | 129964      | (GAT)7          | AAGGCTGACATGTTTTTCC            | 59                               | CCTACGCAACATCATCACC             | 59                               | 221                         |
| 221   | CCESSR271   | 129972      | (AG)9           | CTTTCGCGCTTTCTCGTACTC           | 59                               | AACACAGCACAACACAGA               | 59                               | 392                         |
| 222   | CCESSR272   | 130030      | (AG)19          | GGTGCAAGAAACATAGAGCA             | 59                               | GAAATCTCTGTTGGCTGGT             | 59                               | 369                         |
| 223   | CCESSR273   | 130066      | (GAG)6          | AAAAGGGACGCTCTCCACAA             | 59                               | GCTCCAAAAACCCACCAATTACC         | 59                               | 212                         |
| 224   | CCESSR274   | 130235      | (TGC)6          | GTTATGAGCGTTGGAGAC             | 59                               | GGCGCTATGAGGGCTGAGAAT            | 59                               | 198                         |
| 225   | CCESSR275   | 130276      | (GCC)7          | GGGCGCTACTCAGAGACTTCTC          | 59                               | TGGAACCTTCAATTCCTTCC            | 59                               | 300                         |
| 226   | CCESSR276   | 130329      | (AT)13          | AGGCTGCGTTCCATATAAAGA            | 59                               | CCAAAGAATCCCACATTGTGAC          | 59                               | 410                         |
| 227   | CCESSR277   | 130342      | (ATC)6          | CTCTTGGAGAAACACCTTCA            | 59                               | CCAAACGCTAAAGATGCCAA           | 59                               | 296                         |
| 228   | CCESSR278   | 130353      | (CA)11          | TCTCTCTCTCTTCATTGCC             | 59                               | GGTACAGAGGTGGATGCTGGG            | 59                               | 154                         |
| 229   | CCESSR279   | 130353      | (CAC)6          | ATGGCTCCATCTATGCTCAC            | 59                               | ATGGTCGCTGGCTCTCTCAATTCC         | 59                               | 215                         |
| 230   | CCESSR280   | 130368      | (AG)17          | GGCTGCTTTAAGGCGCTAGTT            | 59                               | ATCAAGTGATGCTGTCGGGC            | 59                               | 436                         |
| 231   | CCESSR281   | 130475      | (CTC)7          | CGCACAACACCCCTTATCTAT           | 59                               | CATAGGAAATGGAGACGTG             | 59                               | 485                         |
| 232   | CCESSR282   | 130516      | (CCT)6          | ACCCTCTCTCCACCTTTACAC           | 59                               | GAGCAAAACCTTCATCCTGCCTG         | 59                               | 357                         |
| 233   | CCESSR283   | 130575      | (AGA)7          | GTGGTCAAGATGGTTGGTGGG           | 59                               | ACGCACAACACCATCTCCTG             | 59                               | 456                         |
| 234   | CCESSR284   | 130635      | (AG)13          | AGACACCGTGCACTCTCCATC           | 59                               | TAAGCACAACGCTGCTCGTAG           | 59                               | 283                         |
| 235   | CCESSR285   | 130647      | (CAC)6          | GAGGCCGAGTTACAGAGAGTC           | 59                               | AGCATGCCAACCTTATTCC             | 59                               | 366                         |
| 236   | CCESSR286   | 130785      | (TTAAA)4        | TGCTCGGACCTCTCTAGTTG           | 59                               | TGAGTCTGTTAAAAAGGGG             | 59                               | 334                         |
| 237   | CCESSR287   | 130804      | (AG)10          | CGTTACAGAATTTGCCGAGT            | 59                               | CAACTCTAACGCGTCTCAGTA            | 59                               | 360                         |
| 238   | CCESSR288   | 130947      | (GCAGAG)4       | TTTTGAAACACGCTAAGGCTC           | 59                               | CGAGAAGGGGACTGAGAAAG             | 59                               | 273                         |
| 239   | CCESSR289   | 130972      | (TCA)9          | GCACAAACCATATGACGAAG            | 59                               | GGAGGAACAGTTGGAGAATGA           | 59                               | 107                         |
| 240   | CCESSR290   | 130978      | (AG)16          | TTATATGGACAGCCCCGCC            | 60                               | AACAATCCATCTTTGCTCCC           | 59                               | 388                         |
| Sl No. | Marker ID     | Unigene ID* | SSR repeat unit | Forward Primer       | Predicted Tm (˚C) forward primer | Reverse Primer       | Predicted Tm (˚C) reverse primer | Predicted amplicon size (bp) |
|--------|---------------|-------------|-----------------|----------------------|----------------------------------|----------------------|----------------------------------|-----------------------------|
| 241    | CCESSR291     | 130993      | (TC)14          | CGGTGGGGGTCTGATAATT  | 59                               | CGGTTCAGCTAAGG-     | 59                               | 313                         |
| 242    | CCESSR292     | 131127      | (TC)11          | ATGCTGTGTTAGCATGACCC | 59                               | GCAGATATCATACCTCGT  | 59                               | 143                         |
| 243    | CCESSR293     | 131130      | (AAG)6          | CACTACTCAACACTGAGCA  | 59                               | CCATCTCCACCATTTCT   | 59                               | 426                         |
| 244    | CCESSR294     | 131306      | (TCGA)4         | CTTTGCGGGCTAGAGTGA   | 59                               | GCCTCATATTGCAAGGTC  | 59                               | 432                         |
| 245    | CCESSR295     | 131334      | (CCA)7a(AA-C)6  | TAATAACATCCACTCGCAAA | 59                               | TACATCAACAACCTCGCC  | 59                               | 320                         |
| 246    | CCESSR296     | 131479      | (TG)9           | GCAATTGTGATGCTGAG    | 59                               | AAAACACTGGGAAAGAGCA | 59                               | 424                         |
| 247    | CCESSR297     | 131560      | (ACC)6          | CAATCTGTCCCCAGTATGC  | 59                               | ACAAGTGCTCCAAACACC  | 59                               | 328                         |
| 248    | CCESSR298     | 131583      | (AT)12          | CCACACCGTGCAGATTGGT  | 60                               | AAGATCTGTGGGCGATCAAC | 59                               | 453                         |
| 249    | CCESSR299     | 131633      | (CTT)6          | AGGTTCGCGCCAGAAGAATA | 59                               | TGTGCCTTTGGATGACTCTC | 59                               | 141                         |
| 250    | CCESSR300     | 131669      | (AG)12          | TCCGGAAAAGAGTAGTGGAG  | 59                               | TGGCTGAAACAGTCTCCCTC| 59                               | 203                         |
| 251    | CCESSR301     | 131714      | (GA)22          | ATTAATGCGGGAATCCCTTG  | 59                               | TGGTAATCTCGAGTGAGGC | 59                               | 491                         |
| 252    | CCESSR302     | 131742      | (TA)18          | TAATGCACGGCAACAACAC  | 59                               | ATCATGGAAATCTCTCCTCC | 59                               | 112                         |
| 253    | CCESSR303     | 131761      | (CA)10          | CATGGATTGGAATTCTGCT  | 59                               | CAGATATCCAGCACATCAA | 59                               | 454                         |
| 254    | CCESSR304     | 131806      | (ATG)8          | GCACCTAAGAGGAGTGACC  | 59                               | ATGAATTTGCGGTGAAGTGA | 59                               | 401                         |
| 255    | CCESSR305     | 131831      | (GCCTGC)4       | CGATTTCCGCTGTAATGTA  | 59                               | TATCAATACGGAGCTCTGGG | 59                               | 462                         |
| 256    | CCESSR306     | 131899      | (GAT)9          | TCCGGCTGTGACCAAAGAAT | 59                               | ATCAAACCTCGACACCCCTC | 59                               | 184                         |
| 257    | CCESSR307     | 131901      | (AAC)6          | AAACCTTTGATACCGGCTG  | 59                               | GCCGGCTTTGTGAGAGAAC | 59                               | 175                         |
| 258    | CCESSR308     | 131958      | (TCA)9          | GATTCAGATTTCCGCTCC   | 59                               | TGGTCTAGATAGCGTGCTC | 59                               | 211                         |
| 259    | CCESSR309     | 131966      | (AC)10          | TTGGAAACTTTGCTTTTCCC | 59                               | TGGTGAAACTTTCTGGTAT | 59                               | 392                         |
| 260    | CCESSR310     | 132036      | (AAG)6          | CTCTCAACAATCTCGAGCA  | 59                               | CTCCTCTGAGCTCTATGC  | 59                               | 194                         |
| 261    | CCESSR311     | 132205      | (GT)10          | CTTTGATTGTGAGTGTATGC | 59                               | AAGATATGGCAGAGTCATGC | 60                               | 424                         |
| 262    | CCESSR312     | 132207      | (TCGT)7         | AAAGGATTACCGGTCAACAAG| 59                               | ACCCTGTATGGTTGAGGAGA | 59                               | 198                         |
| 263    | CCESSR313     | 132325      | (AAG)8          | TGAGGACTCAAGGGAAATTG | 59                               | TGGTCCAGGGAGATAAGCTTT | 59                               | 462                         |
| 264    | CCESSR314     | 132326      | (AAG)6          | GCTTTGGTCGAGGATGAT   | 59                               | GGAGGCTCCAGGAGTGAAGA | 59                               | 336                         |
of duplicated loci in case of all the tested tetraploid arabicas. In general, no private alleles were evident except in one robusta genotype (Sln274) for marker CCESSR14.

The PIC values were comparable (0.19–0.67 and 0.11–0.77), and no significant differences were seen in the observed/expected heterozygosity ($H_o/H_e$; $t$-value=0.70; $P=0.49$; and $t$-value=0.68; $P=0.40$) for the new markers across the tested tetraploids and diploid robustas, respectively. However, significant differences were observed in the total number of amplified alleles ($N_A$; $t=3.74$, $P<0.005$), as well as, the behaviour of the polymorphic markers ($P_m$s) when tested for HWE and LD in the tested tetraploids and the robusta genotypes (Table 4). In general, more markers were in HWE and only a relatively small proportion of markers exhibited LD and heterozygote excess and/or deficiency in case of robustas, in comparison to tetraploid arabicas.

Validation of genomic SSRs for use in genetic studies

A total of 25 putative genomic SSRs were also validated as genetic markers (Table 5). When tested on the panel of 16 elite robusta and arabica (tetraploid) genotypes, five of these markers in arabicas and one in robustas were found to be monomorphic. Twelve of the polymorphic markers in arabicas resulted in double alleles (putative duplicated loci). In total, a maximum of seven and eight alleles ($N_A$) with an average of 2.7 and 4.3 alleles/marker were obtained for the tested polymorphic markers of which 32% and 96% were informative in arabicas and robustas, respectively (Figure 1). The PIC values varied considerably, with mean PIC value being 0.47 (range 0.12–0.78) for tetraploids, which was significantly less than 0.60 (0.12–0.85) observed for robusta (Table 5). Further, the Student’s $t$-test revealed significant differences in $N_A$ ($t=4.09$, $P=0.00$) but non-significant
differences in PIC estimates ($t=1.26$, $P=0.13$), as well as, for the observed/expected heterozygosity estimates ($H_e/H_o$) for the comparable markers of arabica and robusta genotypes.

Further, it was notable that while $>83\%$ of the $Pms$ were in HWE and only few markers showed significant heterozygote deficiency to varying extent in both arabicas and robustas, the number of marker-pairs that exhibited LD was significantly more in arabicas ($28.0\%$; 8 of 28 pairs) that that seen in robustas ($14.2\%$; 36 of 254 marker pairs).
| Species      | C. arabica (n=8) | C. canephora (n=8) | Coffea spp. (n=12) | Psilanthus spp. (n=2) |
|--------------|------------------|-------------------|-------------------|----------------------|

| Primer ID    | Allele range (bp) | Allele range (bp) | Allele range (bp) | Allele range (bp) |
|--------------|-------------------|-------------------|-------------------|-------------------|
| CCESSR01     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR02     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR03     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR04     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR05     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR06     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR07     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR08     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR09     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR10     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR11     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR12     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR13     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR14     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR15     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR16     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR17     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR18     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR19     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR20     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR21     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR22     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR23     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR24     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR25     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR26     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR27     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR28     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR29     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR30     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR31     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR32     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR33     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR34     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR35     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR36     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR37     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR38     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR39     | 251–254           | 251–254           | 251–254           | 251–254           |
| CCESSR40     | 251–254           | 251–254           | 251–254           | 251–254           |
| Species: C. arabica (n=8) | C. canephora (n=8) | Coffea spp. (n=12) | Psilanthus spp. (n=2) |
|--------------------------|---------------------|--------------------|----------------------|
| **Primer ID** | **N_A** | **PA** | **Allele range (in bp)** | **H_o** | **H_e** | **PIC** | **N_A** | **PA** | **Allele range (in bp)** | **N_A** | **PA** | **Allele range (in bp)** |
| CCESSR41 | 3 | 0 | 212–219 | DL | 4 | 0 | 214–223 | 0.71 | 0.76 | 0.65 | 9 | 0 | 212–226 | 1 | 0 | 139 |
| CCESSR42 | 1 | 0 | 129 | MM | 1 | 0 | 129 | MM | 5 | 1 | 120–141 | 1 | 0 | 132 |
| CCESSR43 | 2 | 0 | 184–190 | DL | 3 | 0 | 187–199 | 0.33 | 0.32 | 0.27 | 5 | 0 | 178–199 | 3 | 0 | 174–187 |
| CCESSR44 | 3 | 0 | 250–259 | 0.88 | 0.58 | 0.45 | 3 | 0 | 238–259 | 1.00 | 0.59* | 0.46 | 8 | 1 | 236–259 | 3 | 0 | 247–159 |
| CCESSR45 | 2 | 0 | 144–151 | DL | 1 | 0 | 151 | MM | 8 | 3 | 137–166 | 2 | 0 | 125–151 |
| CCESSR47 | 1 | 0 | 211 | MM | 1 | 0 | 211 | MM | 2 | 0 | 211–213 | 1 | 0 | 207 |
| CCESSR48 | 2 | 0 | 199–211 | DL | 2 | 0 | 192–205 | 0.13 | 0.13 | 0.11 | 4 | 0 | 192–211 | 1 | 0 | 182 |
| CCESSR49 | 2 | 0 | 192–203 | DL | 3 | 0 | 192–203 | 0.50 | 0.43 | 0.35 | 7 | 1 | 188–203 | 2 | 0 | 185–197 |
| CCESSR50 | 1 | 0 | 122 | MM | 2 | 0 | 116–122 | 0.38 | 0.33 | 0.26 | 4 | 0 | 116–128 | 2 | 0 | 122–128 |
| **Range** | 1–6 | 0 | **0.00–0.88** | **0.23–0.78** | **0.19–0.67** | 1–7 | 0–1 | **0.00–1.00** | **0.13–0.85** | **0.11–0.77** | 1–17 | 0–5 | **0.00–7.52** | **0.13–1.57** | **0.34–0.89** |
| **Average** | 2.14 | 0.00 | 0.47 | 0.54 | 0.43 | 3.00 | 0.02 | 0.54 | 0.53 | 0.44 | 7.52 | 1.57 | 2.11 | 0.34 |
| **SD (±)** | 1.21 | 0.00 | 0.36 | 0.13 | 0.16 | 1.57 | 0.15 | 0.26 | 0.22 | 0.19 | 3.62 | 1.59 | 0.89 | 0.81 |
| **SE (±)** | 0.18 | 0.00 | 0.10 | 0.03 | 0.03 | 0.24 | 0.02 | 0.04 | 0.04 | 0.03 | 0.55 | 0.24 | 0.14 | 0.12 |

**Note:** N_A: Number of amplified alleles; PA: Number of Private Alleles; H_o: Observed heterozygosity; H_e: Expected heterozygosity; PIC: Polymorphism Information Content; PI: Probability of Identity; NA: Not amplified; *: Significant HW dis-equilibrium at P<0.05; **: Highly significant HW dis-equilibrium at P<0.01; The putative DL (duplicated loci) markers were not considered for calculation of various estimates as these appear to be fixed exhibiting no segregation.

doi:10.1371/journal.pone.0113661.t004
Table 5. Allelic diversity attributes of the newly developed 25 genomic SSRs when tested over cultivated and wild related coffee genera.

| Species: | C. arabica (n=8) | C. canephora (n=8) | Coffea spp. (n=12) | Psilanthus spp. (n=2) |
|-----------|------------------|--------------------|--------------------|----------------------|
| Primer ID | N_A  | PA  | Allele range (in bp) | N_A  | PA  | Allele range (in bp) | N_A  | PA  | Allele range (in bp) | N_A  | PA  | Allele range (in bp) |
| CCRM02   | 3    | 0   | 252–262              | DL   | 7    | 2 | 252–278              | 2    | 1   | 162                   |
| CCRM06   | 2    | 1   | 143–156              | 0.14 | 0.14 | 0.12 | 143 MM | 1    | 0   | 143                   |
| CCRM07   | 5    | 0   | 124–140              | 0.71 | 0.82 | 0.78 | 124–146 | 1    | 0   | 115–136               |
| CCRM10   | 3    | 0   | 96–106               | 0.29 | 0.38 | 0.32 | 96–106 | 0.00 | 0.23 | 98–100                |
| CCRM14   | 3    | 0   | 109–188              | DL   | 7    | 2 | 109–132              | 0.63 | 0.68** | 109–155             |
| CCRM15   | 1    | 0   | 298 MM               | 0.17 | 0.62** | 0.61 | 298 MM | 1    | 0   | 123                   |
| CCRM16   | 3    | 0   | 198–202              | 0.00 | 0.70** | 0.67 | 198–202 | 0.88 | 0.89 | 180–226               |
| CCRM17   | 1    | 0   | 230 MM               | 0.71 | 0.82 | 0.78 | 230 MM | 1.00 | 0.86 | 216–250               |
| CCRM19   | 4    | 0   | 217–234              | 0.43 | 0.69 | 0.62 | 217–234 | 0.83 | 0.92 | 188–246               |
| CCRM21   | 7    | 1   | 256–290              | DL   | 7    | 0 | 256–288              | 0.63 | 0.90 | 234–320               |
| CCRM22   | 3    | 0   | 194–203              | DL   | 5    | 1 | 194–254              | 0.25 | 0.67** | 194–254             |
| CCRM23   | 2    | 0   | 140–162              | DL   | 3    | 0 | 140–151              | 0.38 | 0.49 | 140–162               |
| CCRM24   | 3    | 0   | 202–213              | DL   | 3    | 0 | 202–213              | 0.63 | 0.57 | 202–213               |
| CCRM28   | 2    | 0   | 203–206              | 0.00 | 0.26 | 0.23 | 203–206 | 0.13 | 0.13 | 202–210               |
| CCRM31   | 1    | 0   | 111 MM               | 0.38 | 0.44 | 0.42 | 111 MM | 0.38 | 0.44 | 101–115               |
| CCRM33   | 2    | 0   | 110–116              | 0.14 | 0.14 | 0.33 | 110–116 | 0.88 | 0.64 | 106–122               |
| CCRM34   | 2    | 0   | 130–146              | DL   | 4    | 0 | 130–146              | 0.75 | 0.75 | 130–146               |
| CCRM35   | 1    | 0   | 156 MM               | 0.00 | 0.43** | 0.42 | 156 MM | 1.00 | 0.90 | 139–183               |
| CCRM36   | 2    | 0   | 139–146              | DL   | 7    | 1 | 139–175              | 1.00 | 0.90 | 139–183               |
| CCRM37   | 2    | 0   | 140–150              | DL   | 3    | 0 | 140–150              | 1.00 | 0.63 | 140–150               |
Table 5. Cont.

| Species: | C. arabica (n=8) | C. canephora (n=8) | Coffea spp. (n=12) | Psilanthus spp. (n=2) |
|----------|------------------|-------------------|-------------------|-----------------------|
| Primer ID | N_A | PA | Allele range (in bp) | Ho | He | PIC | N_A | PA | Allele range (in bp) | Ho | He | PIC | N_A | PA | Allele range (in bp) | N_A | PA |
| CCRM38   | 4   | 1  | 175–217 | DL | 3   | 1  | 175–206 | 0.38 | 0.34 | 0.33 | 12 | 5  | 163–214 | 2   | 0  | 175–200 |
| CCRM40   | 2   | 0  | 144–151 | DL | 5   | 0  | 159–177 | 0.63 | 0.72 | 0.68 | 8  | 0  | 151–177 | 2   | 0  | 151–153 |
| CCRM41   | 1   | 0  | 101    | MM | 2   | 0  | 95–101 | 0.14 | 0.14 | 0.14 | 2  | 0  | 95–110 | 1   | 0  | 101    |
| CCRM42   | 5   | 0  | 130–161 | DL | 5   | 1  | 138–157 | 0.75 | 0.75 | 0.72 | 13 | 2  | 121–165 | 4   | 0  | 134–155 |
| CCRM45   | 3   | 0  | 183–195 | DL | 5   | 0  | 187–212 | 0.50 | 0.76 | 0.72 | 7  | 1  | 183–201 | 2   | 0  | 187–189 |
| Range    | 1–7 | 0–1 | – | 0 | 0.14 | 0.12 | 1–8 | 0–2 | – | 0.00 | 0.13 | 0.12 | 1–16 | 0–6 | – | 0–4 | 0–2 | – |
| Mean     | 2.68 | 0.12 | – | 0.32 | 0.45 | 0.47 | 4.28 | 0.48 | – | 0.57 | 0.63 | 0.60 | 7.92 | 2.08 | – | 2.04 | 0.30 | – |
| SD (±)   | 1.46 | 0.33 | – | 0.32 | 0.28 | 0.24 | 1.97 | 0.71 | – | 0.33 | 0.25 | 0.22 | 3.67 | 1.78 | – | 0.77 | 0.56 | – |
| SE (±)   | 0.30 | 0.07 | – | 0.07 | 0.06 | 0.05 | 0.40 | 0.15 | – | 0.07 | 0.05 | 0.05 | 0.75 | 0.36 | – | 0.16 | 0.11 | – |

Note: N_A: Number of amplified alleles; PA: Number of Private Alleles; Ho: Observed heterozygosity; He: Expected heterozygosity; PIC: Polymorphism Information Content; PI: Probability of Identity; NA: Not amplified; **: Highly significant HW dis-equilibrium at P<0.01; The putative DL (duplicated loci) markers were not considered for calculation of various estimates as these appear to be fixed exhibiting no segregation.

doi:10.1371/journal.pone.0113661.t005
Figure 2. Map positions of 18 new SSR markers developed in this study (11 CCESSR and 7 CCRM markers) on robusta linkage map; mapping population was derived from a cross between CxR (a commercial robusta hybrid) and a local selection Kagganahalla [26]. The SSR markers of the existing map used as anchor markers are shown in italic and bold face.

doi:10.1371/journal.pone.0113661.g002

Figure 3. Unrooted phenetic trees based on the allelic diversity across the tested 44 EST-SSRs showing generic affinities between the: a) *C. arabica* and *C. canephora* genotypes, and b) 14 *Coffeea* and two *Psilanthus* taxa; (only >50 bootstrap values are shown).

doi:10.1371/journal.pone.0113661.g003
Mapping of new EST- and genomic SSRs

The 69 new SSR markers were also tested for their suitability in linkage mapping. In total, 11 of the 44 EST-SSRs (~39%) and seven of the 25 genomic SSRs (28%) could be mapped onto an existing first-generation framework linkage map of robusta coffee [2, 26]. This map comprised a total of 374 mapped markers (71 SSRs, 185 RAPDs and 118 AFLPs) on 11 major and 5 minor linkage groups. The new markers developed in the present study were mapped using the existing SSRs on the map as anchor markers. The 18 new markers that could be mapped, occupied positions on eight distinct linkage groups, with eight markers on CLG03; two markers each on CLG06, CLG11, CLG15; one marker on CLG02, CLG04, CLG05, CLG08 (Tables 1 & 2). The position of these 18 markers on robusta linkage groups along with positions of SSRs used as anchors (CM62, CM115, CM12, CM100, Cof_EST01_150, CaM46, CaM44, CM39_302, CM39_273) is shown in figure 2.

Cross-species/-genera transferability and marker conservation

New SSR-markers when tested on 13 related Coffea and two Psilanthus species, exhibited robust cross-species amplifications with alleles of comparable sizes in the tested taxa (Figure 1, Tables 4 & 5, Table S4). The EST-SSRs showed 100% transferability across the tested Coffea and Psilanthus spp., whereas the genomic-SSRs indicated 96% amplification and transferability for Coffea spp. and 92% for the related Psilanthus spp. The analysis also indicated some private alleles (PAs), which possibly could be species-specific (Tables 4 & 5).

Generic affinities within/between cultivated and wild coffee germplasm by new SSRs

The SSR allelic data were examined for their utility in ascertaining the genetic diversity and generic inter-relationships between the cultivated, as well as, the wild coffee gene pool. The average genetic distance values calculated using the EST-SSR allelic data were in general, significantly less but comparable to that obtained using the genomic SSRs for the tested arabicas, robustas, as well as, for different Coffea and Psilanthus species.

The NJ phenetic tree generated using the genetic distance estimates of EST-SSRs allelic data clearly resolved the tested germplasm in two distinct clusters, one representing all the tetraploid arabicas, while the other comprised all the diploid robusta genotypes (Figure 3a) with significant branch support. The selections formed a single cluster within the tetraploids cluster, while pure arabicas and hybrid-selections appeared in distinct sub-clusters. Similarly, in clustering analysis of 14 related species (12 Coffea and two Psilanthus spp.; Figure 3b) along with two genotypes each from C. arabica and C. canephora, tetraploid Erythrocoffeas (C. arabica) and diploid Erythrocoffeea (C. canephora) formed coherent clusters. Moreover, the grouping of the related taxa, in general, was as per their botanical type with few changes. Though all the entries from Erythrocoffeeas fell into one
cluster, it contained two entries from Pachycoffeas (C. dewevrei with C. canephora and C. liberica with C. congensis). The remaining four of the Pachycoffeas (C. excelsa, C. arnoldiana, C. aruwemiensis, C. abeokutae) grouped with each other with good bootstrap support. The C. salvatrix a Mozambicoffea was also grouped with these Pachycoffeas, while the other three Mozambicoffeas (C. racemosa, C. eugenioides, C. kapakata) and two Paracoffeas (Psilanthus spp.) appeared as independent strong groups. Single Melanocoffea species (tested in this study), C. stenophylla was not grouped with any of the above species cluster but was found close to the Coffea species than the Psilanthus spp.

Similar results were obtained using the data from genomic SSRs (CCRMs, data not shown).

Discussion
SSR motifs in coffee transcriptome, and development of new EST-SSR markers
In the present study, 15.4% of the coffee ESTs were found to contain SSRs, which is comparable with our earlier study [11], but much higher than 2.7–10.8% that was reported for 18 representative dicotyledonous species [27], and 7 –10% reported for monocot species [28]. Notwithstanding this apparent enrichment/higher abundance, the SSRs in coffee transcriptome were very comparable to other plant species in observations like: 1. Abundance of TNRs than DNRs; 2. Abundance of AG among the DNRs followed by AT; 3. CG as the least abundant among the DNRs; 4. Abundance of AAG among TNRs (among the dicots); 5. Predominance of GC-rich TNRs (but not CCG/GGC) than the non-GC-rich TNRs.

A total of 18.7% of the detected EST-SSRs were found to be suitable candidates for primer design, a comparatively lower proportion (~50%) then we reported earlier [11]. The main attributes that rendered majority of the identified SSRs unsuitable for marker development were: a shorter repeat core (<18 bp) and/or flanking sequences of low complexity (AT/GC-rich and/or regions prone to secondary structure formation) or shorter lengths seriously constraining designing of optimal primer pairs. However, in this study, primer-to-marker conversion ratio (ca. 88%), was higher than many earlier similar studies in other crops. Such differences in marker conversion ratios are expected due to differences in the quality of primers designed), GC content of the genome, the genome complexity, and/or genome size [13].

SSR enrichment and development of genomic SSRs
The genomic DNA library constructed in this study, resulted in very high proportion of SSR+ive sequences, with very low degree of redundancy (9.1%; 6 out of 66 identified SSRs positive sequences). This was notable, as in earlier similar studies the apparent high success rates were generally confounded by high
degree of redundancy [29]. Similarly, the proportion of SSR positive sequences found suitable for primer designing was also higher (87%) in our study than the average of 54 ± 3% recorded in other species [30]. These observations suggest that the enrichment approach used in this study may be a desirable strategy for efficiently entrapping and targeting the SSRs even in genome(s) like coffee that are relatively poor in SSR motifs [2].

Utility of new EST- and genomic SSRs as genetic markers

The SSRs provide desirable markers for studying genetic diversity, germplasm characterization, constructing reference panels/bar codes, for individualization of genotypes, linkage mapping, population biology, and taxonomic relationships of related taxa [2]. Therefore, it becomes desirable to validate the new markers for their utility in genetic studies, which unfortunately has been lacking in majority of published studies describing development of coffee-specific SSR markers.

Various genetic parameters viz., allelic diversity, PIC, \( H_o \), \( H_e \), HWE, LD calculated for all the new EST and genomic SSRs and mapability on linkage map, amply suggested their possible utility as genetic markers (see Table 4 & 5). In general, the extent and pattern of allelic/genetic diversity revealed by the new markers conform to that reported earlier for the coffee genomic SSRs [5, 6, 31], and the EST-SSRs [8, 11].

Different genetic parameters/tests such as \( H_o \), \( H_e \), LD, HWE are important indicators of origin, evolution and distribution of diversity in the available genepool. The heterozygosity measures (\( H_o \), \( H_e \)) for the new SSR markers indicated heterozygote decay (deficiency) in the tested germplasm. The HWE and LD analysis of the polymorphic markers were in general agreement with our earlier observations with genomic as well as EST-SSRs [5, 8, 11]. Overall, these studies indicated that the tested robusta germplasm comprised allogamous, relatively unrelated genotypes, while autogamous tetraploids comprised mostly of hybrid varieties/selections with overlapping/shared pedigrees. The results thus suggest the suitability of the new markers for reliably ascertaining genetic diversity in the coffee gene pool.

Cross-species/−generic transferability

All the new EST- and genomic SSR markers revealed very high and robust cross species/−generic amplifications with alleles of comparable sizes when tested on 12 other Coffea and two Psilanthus taxa. The data revealed that the markers described here show much higher taxa transferability than earlier published genomic−/EST-SSR markers [2, 5, 7, 11]. This is significant as successful cross-species amplification is generally restricted to related species within a genus and reduces when tested for different genera [32]. Further, it was interesting to note that the new SSRs that were monomorphic/uninformative for the tested arabica/robusta germplasm, exhibited considerable polymorphism across the tested related taxa (the only exceptions were the marker CCESSR16 and 18 that showed a very low
conservation even across the *Coffea* spp.). Thus the new SSR markers described here strengthen the possibility of their use as Conserved Orthologous Sets (COS) for genetic characterization of different related wild coffee taxa, and also for coffee taxonomic/synteny studies.

**Diversity analysis and genetic relatedness within/between *Coffea* and *Psilanthus* species**

The EST−/genomic-SSRs described in this study were able to group all the 16 genotypes (representing the cultivated genepool) in phenetic clustering that was indicative of their species status and known pedigrees (Figure 1a). Similarly, the analysis 14 *Coffea* and two *Psilanthus* species, revealed generic affinities that were largely in agreement with their known taxonomic relationships (Figure 1b), based on their geographical distribution as well as Chevalier’s botanical classification [33]. Importantly, the analysis distinctly separated the two Paracoffea species (*P. bengalensis* and *P. wightiana*) from all the other *Coffea* spp. These results are similar to the earlier published studies undertaken to ascertain species relationships using SSRs [2, 7, 8, 11], as well as other marker approaches [34–36]. These results, thus, amply demonstrate that the new SSR markers developed in the present study can be considerably informative in exploring the taxonomic relationship of coffee species complex.

**Conclusions**

The present study describes a total of 69 new validated SSRs; 44 EST-SSRs developed from coffee transcriptome using *in-silico* methodology, and 25 genomic SSRs developed using SSR enrichment approach. In addition, it provides primer pairs for additional 270 putative EST-SSRs. Analysis of the identified SSR-positive ESTs also provided insights into the relative abundance and distribution pattern of different SSR motifs in the coffee transcriptome, which was found to be relatively rich in its SSR abundance. Among the identified EST-SSRs, TNRs followed by DNRs were more abundant than other SSRs, and among different types of SSR motifs, AG was the most abundant. All the 69 markers were found to be polymorphic in the tested coffee/related germplasm and their utility as efficient genetic markers could be demonstrated for diversity analysis, germplasm individualization, linkage mapping, cross-species transferability and taxonomic studies. As many of these SSRs showed a very high cross-species transferability, they can aid in conservation, management and resolving taxonomic relationships, as Conserved Orthologous Sets (COS) for *Coffea* and *Psilanthus* species and more importantly as efficient, and informative genetic landmarks on molecular linkage maps.
Supporting Information

Table S1. Summary statistics of screening of the coffee unigene ESTs for SSRs.
doi:10.1371/journal.pone.0113661.s001 (PDF)

Table S2. Summary statistics of distribution and abundance of detected SSRs in the unigene ESTs and SSR frequency estimates for coffee transcriptome.
doi:10.1371/journal.pone.0113661.s002 (PDF)

Table S3. Characteristics and distribution of the detected SSR motifs (without MNRs) across the non-redundant 56 SSR+ive sequences generated using SSR enrichment approach.
doi:10.1371/journal.pone.0113661.s003 (PDF)

Table S4. Inter-species and inter-generic transferability of the new EST-SSRs and genomic SSR markers.
doi:10.1371/journal.pone.0113661.s004 (PDF)

Acknowledgments

RKA thanks the Director, CCMB, Hyderabad for the facilities to undertake the study; Director Research, Coffee Board, Bangalore for coffee materials, and R. Phanindranath for providing lab help as and when required. PSH acknowledges CSIR for junior and senior research fellowship during his doctoral research and DBT for the Postdoctoral fellowship.

Author Contributions

Conceived and designed the experiments: RKA. Performed the experiments: PSH. Analyzed the data: PSH RKA. Contributed reagents/materials/analysis tools: RKA. Wrote the paper: PSH RKA.

References

1. Hendre PS, Aggarwal RK (2007) DNA markers: development and application for genetic improvement of coffee, In: Varshney RK, Tuberosa R (eds.), Genomic Assisted Crop Improvement: Genomics Applications in Crops Volume 2, Springer-Verlag, Germany, pp. 399–434.

2. Hendre PS, Phanindranath R, Annapurna V, Lairmeruata A, Aggarwal RK (2008) Development of new genomic microsatellite markers from robusta coffee (Coffea canephora Pierre ex A. Froehner) showing broad cross-species transferability and utility in genetic studies. BMC Pl Biol. 8:51, doi: 10.1186/1471-2229-8-51.

3. Combes M-C, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, et al. (2000) Characterization of microsatellite loci in Coffea arabica and related coffee species. Mol Ecol. 9:1178–1180.

4. Rovelli P, Mettulio R, Anthony F Anzueto F, Lashermier P, et al. (2000) Microsatellites in Coffea arabica L. in Coffee Biotechnology and Quality. Kluwer, Dordrecht, pp. 123–133.

5. Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P, et al. (2003) Isolation and characterization of nine microsatellite markers from Coffea arabica L., showing wide cross-species amplifications. Mol Ecol Notes. 3:647–650.

6. Moncada P, McCouch S. (2004) Simple sequence repeat diversity in diploid and tetraploid Coffea species. Genome. 47:501–509.
7. Poncet V, Hamon P, Minier J, Carasco C, Hamon S, et al. (2004) SSR cross-amplification and variation within coffee trees (Coffeea spp.). Genome. 47:1071–1081.
8. Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, Varshney RK, et al. (2005) Identification and characterization of expressed sequence tags-derived simple sequence repeats markers from robusta coffee variety ‘CXR’ (an interspecific hybrid of Coffeea canephora and Coffeea congensis). Mol Ecol Notes. 5:80–83.
9. Poncet V, Dufour M, Hamon P, Hamon S, de Kocho A, et al. (2007) Development of genomic microsatellite markers in Coffeea canephora and their transferability to other coffee species. Genome. 50:1156–1161.
10. Poncet V, Rondeam M, Tranchant C, Cayrel A, Hamon S, et al. (2006) SSR mining in coffee tree EST databases: potential use of EST-SSR as markers for the Coffeea genus. Mol Gen Genomics. 276:436–449.
11. Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, et al. (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor Appl Genet. 114:359–372.
12. Plechakova O, Tranchant-Dubreuil C, Benedet F, Couderc M, Tinaut A, et al. (2009) MoccaDB - an integrative database for functional, comparative and diversity studies in the Rubiaceae family. BMC PI Biol. 9:123, doi: 10.1186/1471-2229-9-123.
13. Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 23:48–55.
14. Riar DS, Rustgi S, Burke IC, Gill KS, Yenish JP (2010) EST-SSR development from 5 Lactuca species and their use in studying genetic diversity among L. serriola biotypes. J Heredity doi:10.1093/jhered/esq103.
15. Lin C, Mueller LA, McCarthy J, Crouzillat D, Pe’tiard V, et al. (2005) Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. Theor Appl Genet. 112:114–130.
16. Aggarwal RK, ShenoyVV, Ramadevi J, Rajkumar R, Singh L (2002) Molecular characterization of some Indian Basmati and other elite rice genotypes using fluorescent-AFLP. Theor Appl Genet. 105:680–690.
17. Hendre PS, Aggarwal RK (2004) High throughput isolation of microsatellites using affinity capture, In: Aggarwal RK (ed), Course Manual, Hands-on Training course: DNA markers: Development and Applications. CCMB, Hyderabad, India, pp. 2.1–2.24.
18. Edwards KJ, Barker JH, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques. 20:758–760.
19. Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online. 1:47–50.
20. Glaubitz JC (2004) Convert: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Mol Ecol Notes. 4:309–310.
21. Dieringer D, Schlötterer C (2003) MicroSatellite Analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Mol Ecol. 3:167–169.
22. Nei M (1972) Genetic distance between populations. Am Naturalist. 106:238–292.
23. Felsenstein J (2004) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
24. Page RDM (2001) TreeView Software. Available: http://taxonomy.zoology.gla.ac.uk/rod/rod.html.
25. van Ooijen JW (2006) JoinMap 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.
26. Hendre PS (2007) Development of coffee specific microsatellite markers and construction of a framework molecular linkage map of robusta coffee, Coffeea canephora. PhD thesis, School of Life Sciences, Jawaharlal Nehru University, New Delhi, India.
27. Kumpatla SP (2004) Computational mining and survey of simple sequence repeats (SSRs) in expressed sequence tags (ESTs) of dicotyledonous plants. M.S. thesis submitted to the School of Informatics, Indiana University-Purdue University, Indianapolis, USA.
28. Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett. 7:537–546.

29. Squirrel J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, et al. (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol. 12:1339–1348.

30. Moriguchi Y, Iwata H, Ujino-Ihara T, Yoshimura K, Taira H, Tsumura Y (2003) Development and characterization of microsatellite markers for Cryptomeria japonica D. Don. Theor Appl Genet. 106:751–758.

31. Aggarwal RK, Rajkumar R, Rajendrakumar P, Hendre PS, Baruah A, et al. (2004) Fingerprinting of Indian coffee selections and development of reference DNA polymorphism panels for creating molecular IDs for variety identification. In: Proceedings of 20th international conference on coffee science (ASIC) Bangalore, India, pp. 751–755.

32. Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross-species amplification of soybean (Glycine max) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Mol Biol Evol. 15:1275–1287.

33. Chevalier A (1947) Les cafeiers du globe. III. Systematique des cafeiers at Faux cafeiers. Maladieset insect nuisible, in P. Lechevalier (Ed.) Encyclopedie de biologique Paris, France, pp. 356.

34. Orozco-Castillo C, Chalmers KJ, Powell W, Waugh R (1996) RAPD and organelle specific PCR reaffirms taxonomic relationships within the genus Coffea. Pl Cell Rep. 15:337–341.

35. Lashermes P, Combes M-C, Trouslot P, Charrier A (1997) Phylogenetic relationship of coffee-tree species (Coffea L.) as inferred from ITS sequences of nuclear ribosomal DNA. Theor Appl Genet. 94:947–955.

36. Ruas PM, Ruas CF, Rampim L, Carvaljo VP, Ruas EA, et al. (2003) Genetic relationship in Coffea species and parentage determination of interspecific hybrids using ISSR (inter-simple sequence repeat) markers. Genet Mol Biol. 26:319–327.