RESEARCH ARTICLE

Mining and Validation of Polymorphic EST-SSR Markers for Analysing Genetic Diversity among Interspecific Hybrids of Tea

J.A. Maangi[1,2], R. K. Korir[2], J. L. Bargul[1,3], S.M. Kamunya[2], R. C. Muoki[2*]

Abstract

Molecular markers are useful tools for studying gene flow and genetic diversity within a population. With the establishment of online databases, sequence can be now be downloaded for identification of Expressed Sequence Tags- Simple Sequence Repeat (SSR) markers designed to amplify the microsatellite loci present in the genes. Of the 170 microsatellites detected from the downloaded 789 ESTs belonging to the nine Camellia spp, only fourteen showed functional EST-SSR markers. Among them, five polymorphic markers specific to tri-nucleotide SSRs were randomly selected and screened alongside five established primers on four tea genotypes (TRFK 570/2, TRFK 688/1, TRFK 83/1, and TRFK 6/8,) The polymorphism information content (PIC) of the SSR markers had mean of 0.33 that ranged between 0.16 and 0.53, whereas discriminating power (D) had mean value of 0.23 ranging between 0.00 and 0.75 with. A total of 64 alleles were detected with an average of 6.4 alleles per SSR locus. Based on a PIC ≥ 0.40, D value ≥ 0.20 and number of polymorphic bands ≥ 1, five functional SSR markers comprising two novel EST-SSRs (Camjap A1, Camjap A4) and three adapted microsatellite markers (TM 134, A37, and A47) were recommended for use in discriminating interspecific hybrids of tea.

Keywords: Interspecific hybrids, microsatellite marker, polymorphism information content, Camellia *spp.,* Discriminating power

International Journal of Tea Science (2019); DOI: 10.20425/ijts1515

Introduction

Tea [*Camellia sinensis* (L.) O. Kuntze] is a non-alcoholic, caffeine-rich beverage widely consumed for its attractive aroma, medicinal value, and mildly stimulating effect.[1] The genus *Camellia* belongs to Theaceae family that is indigenous to Central Asia and with over 200 reported species that naturally hybridize.[2,3] From here, tea was spread to other parts of the world and is currently cultivated in diverse environments, ranging from 49°N to 30°S and altitudes from sea level to 2,700 m above mean sea level (amsl).[4] The performance of the tea industry is vital to the global and specifically Kenyan economy, for instance in 2019 tea was one of the leading foreign exchange earner in Kenya contributing over USD 1.098 billion.[5] It accounts for approximately 26% of export earnings and contributes about 4% to the GDP.[6] More than 750,000 farmers directly earn a living from tea and more than 6 million Kenyans directly or indirectly depend on tea.[6]

Tea improvement depends on the extent of genetic diversity within the available population. Cultivated tea naturally hybridize freely with some closely related 'wild' species in the genus Camellia.[7,8] As such, gene flow involving wild species and their domesticated counterparts is valuable in the enrichment of effective breeding population. The initial dispersal of tea led to varietal speciation and the evolution of three distinct cultivated taxa namely: var. *sinensis* ('China tea'), var. *assamica* (Masters) Kitamura ('Assam tea'), and var. *assamica* spp. *lasiocalyx* (Planchon ex Watt) ('Cambod tea').[3] In an effort to access diversity from the secondary and tertiary gene pools of tea, several *Camellia* species namely *C. japonica, C. brevistyla, C. sasanqua, C. irrawadiensis, C. assimilis, C. oleifera, C. kissi, C. chrysantha, C. furfuraceae,* and *C. brevistyla* were introduced in Kenya.[5] The prospect of using such interspecific hybridization in improving some traits in tea e.g. cold hardiness, drought tolerance, specific characters in chemical components, disease and pest resistance among others exists. However, the contribution of these species to the cultivated gene pool is presently unknown. Evolutionary effects that arise from spontaneous mating of domesticated plants with their wild relatives may lead to gene flow that may significantly impart evolutionary change in the recipient populations.[9] In tea, the breeding strategies are determined by its highly outbreeding nature, the long generation time of seed to flower and its amenability to vegetative propagation.

Molecular markers are useful tools for studying gene flow and genetic diversity within a population.[6] DNA based marker, microsatellites or simple sequence repeats (SSRs) consisting of tandem repeats of 2 – 6 bp with a motif length of 20 - 100 bp, are ideal for estimating variation between closely related genotypes.[7] SSRs are reproducible, multi-allelic, informative, polymorphic, relatively abundant in the genome, and co-dominantly inherited.[8,9] With the establishment of expressed sequence tag (EST) sequencing projects for gene discovery programs in several plant species, a wealth of DNA sequence information has been generated and deposited in online databases.[10] By using computer programs, the

[1]Department of Biochemistry, Jomo Kenyatta University of Agriculture & Technology, P.O. Box 62000-00200, Nairobi, Kenya

[2]Crop Improvement and Management Programme, Kenya Agricultural and Livestock Research Organization - Tea Research Institute, P.O. Box 820-20200, Kericho, Kenya

[3]International Centre for Insect Physiology and Ecology, P.O. Box 30772-00100, Nairobi, Kenya

Corresponding Author: R.C. Muoki, Crop Improvement and Management Programme, Kenya Agricultural and Livestock Research Organization - Tea Research Institute, P.O. Box 820-20200, Kericho, Kenya, e-mail: Richard.Chalo@kalro.org, rmchalo@gmail.com

How to cite this article: Maangi, J.A., Korir, R.K., Bargul, J.L., Kamunya, S.M., Muoki, R.C. Mining and Validation of Polymorphic EST-SSR Markers for Analysing Genetic Diversity among Interspecific Hybrids of Tea. International Journal of Tea Science 2020, 15(1):40-45.

Source of support: Nil

Conflict of interest: None

Received: 04/06/2020; **Revised:** 10/08/2020; **Accepted:** 19/09/2020

sequence data for ESTs, genes and cDNA clones can be downloaded from GenBank, scanned for identification of SSRs (typically referred to as EST-SSRs or genic microsatellites), locus-specific primers flanking EST-SSRs designed to amplify the microsatellite loci present in the genes, and then tested for their ability to give reproducible discrete and polymorphic fragments.[10] Here, ten SSR markers (5 novel EST-SSRs and 5 adapted from existing publications) were assessed for their applicability in characterizing genetic diversity in interspecies hybrids of tea.

Materials And Methods

Plant materials

Fresh tender shoots (two leaves and a bud) were harvested from 4 tea cultivars growing in experimental tea gardens of KALRO-TRI Centre in Timbilil (0°22'S, 35°21'E, 2200 m.a.s.l.), Kericho County, Kenya. The cultivars consisted of TRFK 570/2 (progeny of cross TRFK 301/3 ♀ and *C. japonica* ♂), TRFK 688/1 (progeny of cross *C. irrawadiensis* ♀ and TRFK 303/577 ♂), TRFK 83/1 (clonal bush obtained from Kapchomo Estate, EPK Nandi in 1966), and TRFK 6/8 (commercial standard cultivar in processing high quality black tea). The samples were collected using khaki bags and transported in a cool box at 4°C to the laboratory, washed in running water and stored at -20°C for subsequent DNA extraction.

Genomic DNA extraction

Genomic DNA was extracted from the leaf samples using modified cetyltrimethylammonium bromide (CTAB) protocol.[17] The DNA quality and quantity were assessed using Nanodrop spectrophotometry (Thermo Scientific NanoDrop 2000 UV-Vis Spectrophotometer), while integrity was checked using 1.8% agarose gel electrophoresis.

Data mining and processing of EST-SSRs

A total of 789 Expressed Sequence Tags (ESTs) belonging to *C. japonica* (519 ESTs), *C. taliensis* (67 ESTs), *C. brevistyla* (59 ESTs), *C. chrysantha* (45 ESTs), *C. furfuracea* (37 ESTs), *C. sasanqua* (28 ESTs), *C. kissi* (19 ESTs), *C. irrawadiensis* (9 ESTs), and *C. assimilis* (6 ESTs) were downloaded from the NCBI GenBank (http://www.ncbi.nlm.nih.gov/) in their FASTA format on 20th September 2019. After removal of redundancy in the sequences using CAP 3 program with default parameter values (i.e. base quality cutoff for clipping = 12, overlap length cutoff = 30, overlap percentage identity cutoff = 75, overlap similarity score cutoff = 500 and minimum number of good reads at clip position = 2)[18], 440 non-redundant unigenes (80 contigs and 360 singletons) were generated. Further, contaminating sequences such as adapters, linkers, PCR primers, and vector sequences were removed and poly A/T tails trimmed.[19]

Identification of SSR Motifs and Primer Design

Non-redundant EST datasets of the nine *Camellia* spp. were separately processed using the Sequence Repeat Identification Tool, SSRIT.[20] The criteria used were as follows: maximum motif length = decameter (10 identical and repetitive nucleotides); minimum number of repeats allowed = 3. The microsatellites were classified into Class I (≥ 20 nucleotides) and Class II (12 to ≤ 20 nucleotides) and used in primer design.

PCR primers were designed using Primer3Plus software based on the regions flanking each SSR motif.[21,22] The design parameters were set as follows: primer length 18-27bp, optimum 20bp; annealing temperature (Tm) minimum 57°C, maximum 63°C and optimum 60°C; %GC content min 40, max 60, and optimum 50;

Table 1: Characteristics of the EST-SSR and adapted microsatellite markers used to study genetic diversity among interspecific hybrids of tea

Code	*Sequences (5' to 3')*	*Target motif*	*Product size range (bp)*	*Reference*
Camjap A1	F_ AACAGCAGCAACAGCAACAA	(GCA)5	280	Novel EST
	R_TCCATCCAATACTGCAAGTCC			
Camtal A1	F_CCTTCGCTCACCATTCTTTC	(TTC)6	168	Novel EST
	R_TGTAGCCCATTCCCTTTGTC			
Camjap A2	F_CCTTGTCTGTAATGCCTCTCAA	(CAG)4	259	Novel EST
	R_TGTTGTTGTTGCCTGTTGGT			
Camjap A3	F_AGCCAAGAAGATGTCCTCCA	(AAC)4	176	Novel EST
	R_CATCACCACCAACTCCATCA			
Camjap A4	F_CACGATTCCTCTCAGCAACA	(AAC)5	184	Novel EST
	R_GACTTCCATCGGAATCCTCA			
TM 134	F-TTCCGTGACTGATTTATGTG	(CAT)8	221-251	Wambulwa et al. (2016)
	R-TTGAGACTCGGGGTTTT			
A37	F-TCTGCCCTTCCCTAAATC	(AAG)9	170-182	Wambulwa et al. (2016)
	R-ATGTTTGGTCTCGGTTGTT			
A 47	F-TCCCTACAAACCCTAACCG	(GCC) 5	171-201	Wambulwa et al. (2016)
	R-GAGCAGCATCAGAGTCACGT			
Camsin M2	F_CCT CTG GGT GTC CTA CAC CT	(GT)17	240–260	Freeman et al. (2004)
	R_AAA GCC TTG ATG CCT TTC G			
Camsin M3	F_GGT GTG GTG TTT TGA AGA AA	(CA)18	190–210	Freeman et al. (2004)
	R_TGT TAA GCC GCT TCA ATG C			

maximum Tm difference between sense and antisense primer 2°C; and amplicon size range from 125 – 300 bp.[22] Five EST-SSR primers (Table 1) matching the criteria described by Zhang et al.[23] were synthesized by Inqaba Biotec, South Africa.

Screening and validation of SSR markers

A total of ten SSR primers comprising five novel and five adapted from published work (three from Wambulwa et al.[30] and two from Freeman et al.[31]) (Table 1) were screened for polymorphism. PCR amplifications were performed in 10 µL reaction volume using a thermal cycler (TC-5000, Techne Inc., Thermo Scientific) each consisting 40 ng genomic DNA, 0.2mM dNTPs, 0.5U *Taq* polymerase (recombinant), 2 mM $MgCl_2$, 0.5 µM of each primer (Forward and Reverse), and 1× PCR buffer (100 mM Tris-HCl, 500 mM KCl; pH 8.3). PCR were run with initial denaturation for 4 min at 94°C, followed by 35 cycles of 94°C for 30s, 55°C for 1 minute, 72°C for 30s, and a final extension of 7 minute at 72°C. The PCR products were resolved on 2.0% agarose gel stained and visualized under UV light (UVP PhotoDoc-it™ imaging system + Benchtop Variable Transilluminator Upland, CA, USA).

Data analysis

SSR products (genotypic) were scored manually as binary data using '1' (band present) and '0' (null allele) based on the SSR pattern of amplification of each primer-pair. The Online Marker Efficiency Calculator (iMEC) was used to compute key indices of polymorphism: the polymorphic information content (*PIC*) and discriminating power (*D*).[32] The *PIC* of each primer-pair was estimated using the following formula implemented in the iMEC program;

$$PIC = 1 - (\sum_{i=1}^{n} {p_i}^2) - \sum_{i=1}^{n-1} \sum_{j=1+1}^{n} (2{p_i}^2 {p_j}^2)$$

Where *pi* and *pj* are the distribution frequencies of the *i*-th and *j*-th alleles in the population, whereas *n* denotes the number of alleles identified by a marker.[33] *PIC* indicates the discriminating power of a marker based on allele distribution frequency and the number per locus in the genotypes being studied.[15] Co-dominant markers such as microsatellites SSR markers with a *PIC* value of ≥0.3 can detect moderate to high genetic diversity in a population.[33,34] The discriminating power (D) of each marker was computed using the following formula;

$$D = 1 - C_i = 1 - \sum_{i=1}^{I} P_i \frac{(NP_i - 1)}{N - 1}$$

Where *I* is the total number of genotypes (banding patterns) produced by a marker, *Pi* denotes the frequency of *ith* genotype of the *jth* primer, *N* represents the number of individuals tested, and *Ci* the confusion probability of *jth* SSR, which is the likelihood that any two individuals selected randomly from a sample possess a similar banding pattern.[35]

Results

Development of EST-SSR primers

A total of 1331 potential SSR repeats (221 in 80 contigs and 1,110 in 360 singletons) were identified by the SSRIT tool from the 789 ESTs belonging to the nine Camellia spp. downloaded from the database. This represents 18.6% of the unigenes with microsatellite motifs. Di-nucleotides were the most abundant repeat motif with 836 (62.90%) loci followed by tri-nucleotides with 456 (34.31%) loci (Figure 1). The remaining loci consisted tetra-, penta-, hexa-, and hepta-nucleotides accounted for 37 (2.78%). Mononucleotides were omitted since they could have resulted from sequencing errors.[36] On average, the maximum number of repeats (CT) found in one unigene was 14. Di-nucleotide repeats of the (TA)n, (AT)n, and (AG)n type were the most abundant microsatellite at 30.8% followed by (TG)n, (GA)n, and (TC)n at 20.8% (Figure 2). Among tri-nucleotides, (CCA)n, (ACC)n, (GAA)n and (CAC)n repeats were the most prevalent, cumulatively occurring in 28 sequences. The number of SSR sequences containing tetra-, penta-, hexa-, and hepta-nucleotide were lowest. Higher DNA polymerase-mediated slippage events in shorter units can explain this variation in microsatellite density.[24] In total, 170 microsatellites comprising 39 Class I and 131 Class II types were detected but only fourteen returned functional EST-SSR markers based on Primer3Plus design parameters. Finally, 5 polymorphic markers specific to tri-nucleotide SSRs were randomly selected on the prefix that they were likely to be maintained in related species due to triplet codon, with fairly similar Tm and %GC between forward and reverse primers, minimal or no secondary structure (primer-dimers), annealing temperature ~60^{0}C, and GC content of less than 50%[23]. The primers were synthesized and tested alongside five adapted primers whose polymorphism had been established.

Polymorphisms and discriminating power of SSR Markers

To validate the polymorphisms of these 10 SSR markers, PCR-based genotyping was performed using three interspecific hybrids and one intraspecific (control) cultivar. All the ten markers produced PCR amplicons separable by size in all three hybrid samples screened and nine in the commercial intraspecific cultivar (Figure 3). A total of 64 SSR alleles (bands) over 10 loci were identified (Table 2) with the number of alleles per locus ranging from 2 (Camsin M3) to 22 (Camjap A1) on an average of 6.4 (SD = 5.5). Primers with high allele frequency were Camjap A1, TM 134, and Camjap A4 that showed 22, 10, and 8 alleles, respectively (Table 2). The size of the amplified

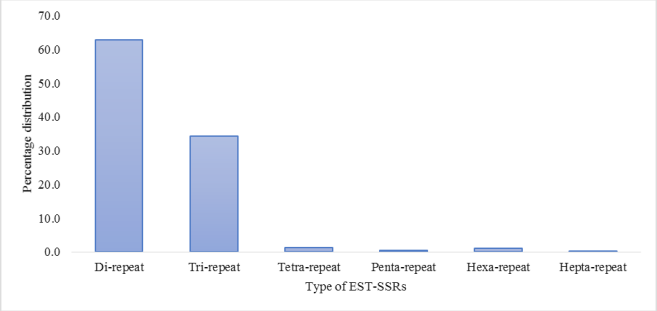

Figure 1: EST-SSRs repeat motif percentage distribution in unigenes belonging to Camellia spp.

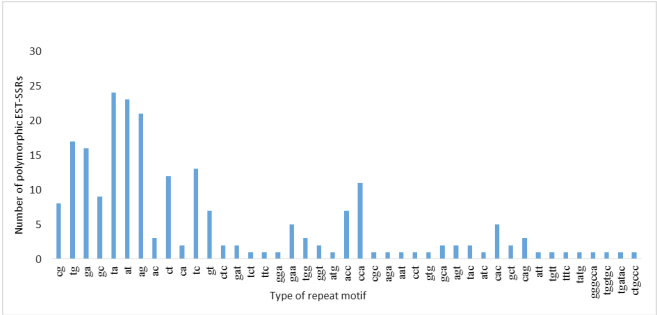

Figure 2: Repeat motif type distribution of polymorphic EST-SSRs belonging to Camellia spp.

Table 2: Characteristics of the 10 SSR primers used in the study

Primer #	*Primer's Code*	*Allele No.*	*Size range (bp)*		*No. of polymorphic bands*	*PIC value*	*Discriminating power (D)*
			Min.	*Max.*			
1	Camjap A1	22	50	1500	12	0.38	0.75
2	Camtal A1	3	50	200	0	0.16	0.00
3	Camjap A2	4	50	450	1	0.24	0.08
4	Camjap A3	3	50	200	0	0.16	0.00
5	Camjap A4	8	50	400	3	0.45	0.15
6	TM 134	10	125	1350	1	0.37	0.70
7	A37	4	50	750	1	0.45	0.20
8	A47	5	50	750	2	0.53	0.22
9	Camsin M2	3	200	250	1	0.34	0.13
10	Camsin M3	2	150	200	1	0.19	0.06
Total	-	64	-	-	22	-	-
Average	-	6.4	-	-	2.2	0.33	0.23

Table 3: Characteristics of SSR primers that provided reliable informativeness on four Camellia genotypes

Primer #	*Primer's Code*	*Allele No.*	*No. of polymorphic bands*	*PIC value*	*Discriminating power (D)*
1	Camjap A1	22	12	0.38	0.75
2	Camjap A4	8	3	0.45	0.15
3	TM 134	10	1	0.37	0.70
4	A37	4	1	0.45	0.20
5	A47	5	2	0.53	0.22
Total	-	49	19	-	-
Average	-	9.8	3.8	0.44	0.40

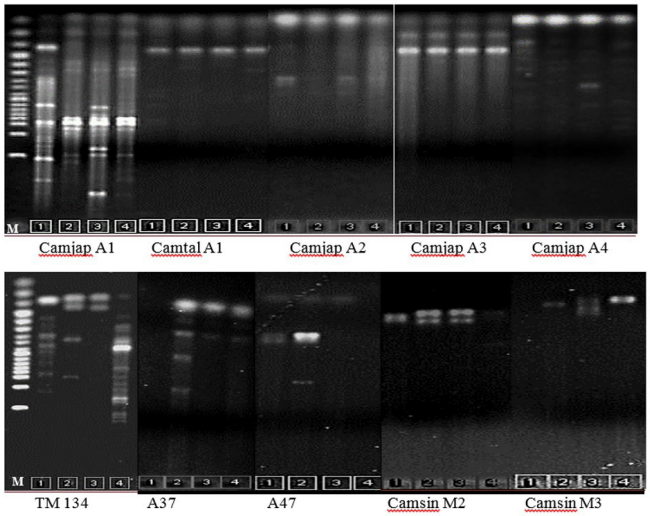

Figure 3: SSR markers Camjap A1, Camtal A1, Camjap A2, Camjap A3, Camjap A4, TM 134, A37, A47, Camsin M2, and Camsin M3 profiles of intraspecific hybrid 1 (6/8) and interspecific hybrids 2 (570/2), 3 (688/1), and 4 (83/1) on Etbr-stained 2% agarose gel using 50 bp DNA size marker (M) (Inqaba Biotech, South Africa).

alleles at all loci varied between 50 bp and 1500 bp, with mean of 491.8 bp (SD = 439.3).

A total of 22 polymorphic bands were produced by eight SSR primers (TM 134, A37, A47, Camsin 2, Camsin 3, Camjap A1, Camjap A2, and Camjap A4), accounting for 34.4% polymorphism in the hybrids (Table 2). The *PIC* value ranged from 0.16 (for Camtal A1 and Camjap A3) to 0.53 (for A47), with a mean of 0.33 per marker (SD = 0.13) (Table 2). Mean *PIC* value for genomic microsatellites was 0.37 compared to 0.26 for EST-SSR markers (Table 2). On the basis of *PIC* values, three markers – A37, A47, and Camjap A4 were highly informative (PIC ≥ 0.6), whereas TM134 and Camjap A1 were relatively informative (PIC ≥ 0.4). However, *PIC* values were not significantly correlated with the number of alleles detected (r = 0.35, p <0.33). Two primers showed a higher discriminating power D ≥ 0.5 (M = 0.23, SD = 0.26) (Table 3).

Discussion

Microsatellite/SSR markers are powerful molecular markers for genetic diversity assessment, gene flow rate and molecular breeding in crops compared to RFLP, RAPD, or AFLP markers due to their multiallelic nature codominant inheritance, reproducibility, high variability and genome coverage.[14,37] A wide increase in sequencing projects has provided a wealth of DNA sequence information that is useful for mining EST-SSR markers for genetic improvements. The development and use of EST-SSR markers in diversity studies has been previously reported. Yao et al. developed and utilized 96 polymorphic EST-SSR markers for population structure analysis in 450 Chinese tea accessions.[25] Ma et al. also reports the development and polymorphism validation of 74 EST-SSR markers in 45 tea cultivars belonging to 7 different varieties.[26] In a recent study, 82 SSRs were developed from sequences available in public databases such as ESTs, GSS and RNA-seq and validated using 36 tea genotypes.[27]

In the present study, frequency analysis revealed that di-nucleotide repeats were the most frequent motif type (62.9%) in the wild Camellia genomes followed by tri-nucleotides (Figure 1). This confirm earlier studies where high di-nucleotide (over 51%) repeat density relative to the other repeat units was been reported in *C. sinensis* ESTs.[28,29]. Different parameter values used for motif detection could explain the slight variation in the number of reported SSR classes between studies.[27] The most abundant di-nucleotide repeats were (TA)n, (AT)n and (AG)n, making up 43.9% of all di-repeats, which is consistent with Tan et al.[30] who reported that AG/CT motif was most frequent repeat unit followed by AT/ TA in tea. Among tri-nucleotides, ACC/CCA motif was the most abundant, accounting for 30.1% of all tri-repeats detected (Figure 2).

Of the total of 170 microsatellites (39 Class I and 131 Class II SSRs) detected, 14 returned novel polymorphic EST-SSR markers based on the Primer 3 Plus design parameters.[16] Out of these, five markers specific to tri-nucleotide SSRs were randomly selected and validated in 3 inter- and 1 intraspecific cultivars. An additional five established markers were adapted from earlier publications for verification.[18,19] Most of the markers (70%) revealed moderate to high polymorphism (PIC ≥ 0.33) in the four cultivars tested, demonstrating efficiency in distinguishing intra- and inter-specific hybrids of tea (Table 2). The five novel markers detected 64 alleles with an average of eight alleles per marker compared to 4.8 for the adapted markers. Generally, PIC values that estimate the informativeness of a marker based on the allelic frequency and total alleles detected[15,16], were not significantly correlated with allele frequency data ($r = 0.35$, $p<0.33$). This suggests that the usefulness of a marker was not dependent on detecting a higher number of alleles as opposed to when the value exceeds 0.5 which indicates informativeness.[34] Three SSR markers comprising one novel EST-SSR marker (Camjap A4) and two adapted markers (A37 and A47) gave an average informative value of 0.50 (Table 2). Moreover, the average PIC value for genomic microsatellites (core markers of *C. sinensis*) was not significant ($p \leq 0.05$) (0.38) from that of novel EST-SSR markers (0.28) (Table 2). Occurrence of EST-SSR markers in a more conserved region of the genome may account for the mean differences in PIC recorded.[39]

Discriminating power (D) is also a useful estimator of the informativeness of a marker.[33] SSR markers with higher discriminating power ($D \geq 0.7$) give an optimal primer combination for discriminating cultivars.[36] In this study, the D values of two primers, namely, Camjap A1 and TM 134 were 0.70 and 0.75, respectively (Table 2). These markers are thus efficient tools for definitive identification of inter- and intra-specific hybrids. Both *PIC* and *D* are dependent on allele frequency. However, the discriminating efficiency of a primer is not exclusively dependent on the number of polymorphic bands it produces implying that SSRs with a similar frequency of unique bands can have different *D*, e.g., TM134 and A37 (Table 2). On the other hand, two markers producing significantly different numbers of polymorphic bands may have fairly similar *D*, e.g., Camjap A4 and Camsin M2 (Table 2). Differences between the frequencies of alleles produced by primers could help explain this result.[36]

The number of polymorphic bands can be useful for evaluating genetic divergence.[40] Mutations occurring in SSR regions produce fragments of different sizes detectable on amplification. Thus, polymorphic alleles can indicate unique genetic variants in the population. Here, 64 bands were detected with an average of 6.4 bands per marker (Table 2). Eight of the markers were polymorphic with the five EST-SSR markers accounting for 72.7% of the polymorphism. These SSR markers are potential candidates for selecting inter- and intra-specific hybrids within the genus *Camellia*. Multiple alleles were revealed in the genotypes by all the markers tested (Table 2), an indication of high heterozygosity among the hybrids.[41] Moreover, tea plants are known to outcross freely, with hybrids exhibiting high genetic diversity.[3] The genotypes tested in this study are half-sib progenies between cultivated tea and related wild *Camellia* species. Therefore, these multi-allelic SSR markers can be useful for distinguishing a specific genotype.

On the basis of PIC (≥ 0.40), D (≥ 0.2), and number of polymorphic bands (≥ 1), a set of five polymorphic SSR primers (Camjap A1, Camjap A4, TM 134, A37, and A47) were selected as ideal for studying genetic diversity in interspecific tea hybrids. Their mean PIC and D values of 0.44 and 0.40 (Table 3), respectively, indicate their efficient ability to discriminate hybrid clones. This is consistent with Wambulwa et al.[31] who identified 23 polymorphic SSR loci that separated East African teas into groups based on geographical origin. Further, identified SSR markers are efficient tools for studying diversity in closely related breeding lines.[42,43] In all these studies, *PIC* and related indices of polymorphism were used as a benchmark for assessing the effectiveness of SSR markers. Our data indicate that the SSR markers identified can effectively be used to discriminate interspecific hybrids of tea as well as identify germplasm for future inclusion in tea improvement programs.

Conclusions

In conclusion, five candidate EST-SSR markers from 789 ESTs were designed and tested alongside other five published microsatellite markers on three interspecific hybrids and a commercial intraspecific cultivar. On average, PIC and discriminating power of five primers (Camjap A1, Camjap A4, TM 134, A37, and A47) were 0.44 and 0.40 (Table 3), respectively, implicating the markers as ideal for studying genetic diversity and population structure of tea. The high cross-species transferability of these SSRs indicated their value as useful markers for tea cultivar differentiation.

Acknowledgement

This study was supported by the National Research Fund and the Tea Research Institute, Kericho, Kenya.

REFERENCES

1. Karak T & Bhagat RM. 2010. Trace E elements in tea leaves, made tea and tea infusion: A review. *Food Res Int* 43(9): 2234–2252.
2. Macfarlane A & Macfarlane I. 2004. *The Empire of Tea*. The Overlook Press: New York, 32 p.
3. Preedy, VR. 2012. *Tea in Health and Disease Prevention*. Academic Press: New York, 67 p.
4. Zhen Y, Chen Z, Cheng S & Chen M. 2005. *Tea: Bioactivity and Therapeutic Potential*. Taylor & Francis: New York, 109 p.
5. International Tea Committee (ITC). 2020. *Annual Bulletin of Statistics*. London: Author.
6. Tea Board of Kenya. 2010. *Annual Report and Accounts 2009/2010*. Nairobi, Kenya: Author.
7. Wachira FN, Powell W & Waugh R. 1997. An assessment of genetic diversity among Camellia sinensis L. (cultivated tea) and its wild relatives based on randomly amplified polymorphic DNA and organelle-specific STS. *Heredity* 78: 603–611.
8. Kerio L, Wachira FN, Kanyiri WJ & Rotich MK. 2012. Characterization of anthocyanins in Kenyan teas: Extraction and identification. *Food Chem* 131(1): 31–38.
9. Ellstrand NC, Prentice, HC & Hancock JF. 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annu Rev Ecol Syst* 30: 539–563.

10. Varshney RK, Graner A & Sorrells ME. 2005. Genic microsatellite markers in plants: Features and applications. *Trends Biotechnol* 23(1): 48–55.
11. Tea Research Foundation of Kenya. 2010. *Annual Technical Report 2010*. Kericho, Kenya: Author, 73 p.
12. Marić S, Bolarić S, Martinčić J, Pejić I & Kozumplik V. 2004. Genetic diversity of hexaploid wheat cultivars estimated by RAPD markers, morphological traits and coefficients of parentage. *Plant Breed* 123(4): 366–369.
13. Agrawal DC, Saha S, Jenkins J & Zipf, A. 1999. Use of cross-species simple sequence repeat (SSR) primers for developing polymorphic DNA markers. *J New Seeds* 1(2): 1–12.
14. Gupta P, Rustgi S & Kulwal P. 2005. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Mol Biol* 57(4)*:* 461–485.
15. Nagy S, Poczai P, Cernak I, Gorji AM, Hegedus G & Taller J. 2012. PICcalc: An online program to calculate polymorphic information content for molecular genetic studies. *Biochem Genet* 50(9-10): 570–672.
16. Reyes-Valdés MH. 2013. Informativeness of microsatellite markers. *Methods Mol Biol* 1006: 259–270.
17. Kamunya S. 2010. *DNA analysis*. Kericho, Kenya: Tea Research Foundation of Kenya, 2 p.
18. Huang X & Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Res* 9(9): 868–877.
19. Huang H, Zhang Z, Zhang Z, Mao S, Wang L & Zhang B. 2011. Analysis of SSRs information in *Capsicum* spp. from EST database. *Agric Sci China* 10(10): 1532–1536.
20. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S & McCouch S. 2001. Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* 11(8): 1441–1452.
21. Rozen S & Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132: 365–386.
22. SantaLucia J. 2007. Physical principles and visual-OMP software for optimal PCR design. *Methods Mol Biol* 402: 3–34.
23. Zhang K, Wu Z, Tang D, Lv C, Luo K, Zhao Y, Liu X, Huang Y & Wang J. 2016. Development and identification of SSR markers associated with starch properties and β-carotene content in the storage root of sweet potato (*Ipomoea batatas* L.). *Front Plant Sci* 7(223): 1–21.
24. Kruglyak S, Durrett, RT, Schug, MD & Aquadro, CF. 2000. Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. *Mol Biol Evol* 17: 1210–1219.
25. Yao MZ, Ma CL, Qiao TT, Jin J & Chen L. 2012. Diversity distribution and population structure of tea germplasms in China revealed by EST-SSR markers. *Tree Genet Genomes* 8(1): 205–220.
26. Ma JQ, Zhou YH, Ma CL, Yao MZ, Jin JQ, Wang XC & Chen L. 2010. Identification and characterization of 74 novel polymorphic EST-SSR markers in the tea plant, *Camellia sinensis* (Theaceae). *Am J Bot* 97(12): e153–e156.
27. Dubey H, Rawal HC, Rohilla M, Lama U, Kumar PM, Bandyopadhyay T, Gogoi M, Singh NK & Mondal TK. 2020. TeaMiD: A comprehensive database of simple sequence repeat markers of tea. *Database* 2020: 1–14.
28. Sharma RK, Bhardwaj P, Negi R, Mohapatra T & Ahuja PS. 2009. Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). *BMC Plant Biol* 9(53): 1–12.
29. Wu H, Chen D, Li J, Yu B, Qiao X, Huang H & He Y. 2013. De novo characterization of leaf transcriptome using 454 sequencing and development of EST-SSR markers in tea (*Camellia sinensis*). *Plant Mol Biol Rep* 31: 524–538.
30. Tan LQ, Wang LY, Wei K, Zhang CC, Wu LY, Qi GN, Cheng H, Zhang Q, Cui QM & Liang JB. 2013. Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). *PLoS ONE*, 8(e81611): 1–13.
31. Wambulwa MC, Meegahakumbura MK, Chalo R, Kamunya S, Muchugu A, Xu JC, Liu J, Li DZ & Gao L. 2016. Nuclear microsatellites reveal the genetic architecture and breeding history of tea germplasm of East Africa. *Tree Genet Genomes* 12(11): 11–21.
32. Freeman S, West J, James C, Lea V & Mayes S. 2004. Isolation and characterization of highly polymorphic microsatellites in tea (*Camellia sinensis*). *Mol Ecol Notes* 4(3): 324–326.
33. Amiryousefi A, Hyvönen J & Poczai P. 2018. IMEC: Online marker efficiency calculator. *Appl Plant Sci* 6(6): 1–4.
34. Botstein D, White RL, Skalnick MH & Davies RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am J Hum Genet* 32(3): 314–331.
35. Nei M & Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76(10): 5269–5273.
36. Tessier C, David J, This P, Boursiquot JM & Charrier A. 1999. Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theor Appl Genet* 98(1): 171–177.
37. Taheri S, Abdullah TL, Yusop MR, Hanafi MM, Sahebi M, Azizi P & Shamshiri RR. 2018. Mining and development of novel SSR markers using next generation (NGS) data in plants. *Molecules* 23(2): 399–406.
38. Fang W, Cheng H, Duan Y, Jiang X & Li X. 2012. Genetic diversity and relationship of clonal tea (Camellia sinensis) cultivars in China as revealed by SSR markers. *Plant Syst Evol* 298: 469–483.
39. Parthiban S, Govindaraj P & Senthilkumar S. 2018. Comparison of relative efficiency of genomic SSR and EST-SSR markers in estimating genetic diversity in sugarcane. *3 Biotech* 8(3): 144–149.
40. Rocha EA, Paiva LV, de Carvalho HH & Guimarães CT. 2010. Molecular characterization and genetic diversity of potato cultivars using SSR and RAPD markers. *Crop Breed Appl Biotechnol* 10(3): 1–8.
41. Tsykun T, Rellstab C, Dutech C, Sipos G & Prospero S. 2017. Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus Armillaria cepistipes. *Heredity* 119(5): 371–380.
42. Zhang Y, Zhang X, Chen X, Sun W & Li J. 2018. Genetic diversity and structure of tea plant in Qinba area in China by three types of molecular markers. *Hereditas* 155(22): 1–12.
43. Zhou Q, Li H, Hoang TX & Ruan X. 2019. Genetic diversity and relationship of dongting Bilouchun tea germplasm in Suzhou revealed by SSR markers. *Pak J Bot* 51(3): 1–8.