Sri Lanka J. Tea Science Volume 77, Part (1/2); 22-44

Development of Genomic and EST Microsatellite Markers and Their Usefulness in Genetic Conservation and Improvement of Tea (*Camellia sinensis* L.)

K M Mewan[1], I S B Abeysinghe[1], M C Saha[2], Yongzhen-Pang[2], E H Karunanayake[3], T L S Tirimanne[4], J M D T Everard[5] and R A Dixon[2]

([1]*Tea Research Institute of Sri Lanka, Talawakelle, Sri Lanka*
[2]*The Samuel Roberts Noble Foundation, Ardmore, OK 73401, USA*
[3]*Institute of Biochemistry Molecular Biology and Biotechnology, University of Colombo, Sri Lanka*
[4]*Faculty of Science, University of Colombo, Colombo, Sri Lanka*
[5]*Coconut Research Institute, Lunuwila, Sri Lanka*)

ABSTRACT

During past couple of decades, due to its' high abundance, power and informative nature, microsatellites or simple sequence repeats (SSRs) has become one of the most popular sources of genetic markers. Currently, a limited number of markers are available for tea (*Camellia sinensis* L.), which in-turn possess a major constraint on applying in genetic analysis, mapping, and marker assisted breeding in tea. Therefore the objective of the study was to generate a set of EST and genomic microsatellite markers of tea and to evaluate their potential in molecular marker based studies in tea.

For isolation of EST-SSRs, expressed sequence tag (EST) sequences were characterized from cDNA libraries of tea cultivars TRI2043 and DT1 whereas for genomic SSR, sequences characterized from two genomic libraries contracted from cultivar TRI2023 enriched for two repeat motifs $(CA/GT)_n$ and $(CTT/GAA)_n$. Both types of sequences were mined for SSRs using the software phpSSRMiner and primer3 program of the same software was used to construct primers.

Of the raw sequences characterized from the two tea genomic libraries, a total of 1,120 contained inserts with good quality clean sequences. Out of them, 945 (84%) were mined for SSRs, where 339 (36%) were identified as containing SSR-positive sequences with repeat motifs $\geq$ 20. Of the 2718 ESTs characterized (1322 from DT1 and 1396 from TRI2043) 571 (21%) contained SSRs with repeat motifs length $\geq$ 18 bases. For both types of SSRs, the majority

of the isolated perfect repeats were di-nucleotides (43% and 67% respectively) followed by hexa- (for EST-SSRs) and tri-nucleotides (for genomic SSRs) (19% and 27% respectively). GA/CT was the major repeat type present in the EST-SSRs and CA/GT and CTT/GAA were predominant in the genomic libraries. A total 305 (157 perfect and 148 imperfect) genomic-SSR and 777 (385 perfect and 392 imperfect) EST-SSR primer pairs were constructed.

Effectiveness of the markers was assessed for genetic analysis and mapping in tea, using the information generated for a set of genomic and EST-SSR primers. Both microsatellite marker sources had similar mean polymorphic information content (PIC) values (0.64). However, a higher percentage of genomic SSRs were found more informative than the EST-SSRs. The results demonstrate the usefulness of above SSRs for molecular marker based genetic studies in tea.

Key words: ESTs, genomic sequences, microsatellites, SSRs

INTRODUCTION

Use of molecular markers in genetics and plant breeding has become a major strategy regularly applied by researches in broad spectrum of areas and in wide range of plant species. Of the various marker systems now available, microsatellites (Litt and Luty, 1989) or simple sequence repeats- SSRs (Tautz, 1989) have proven its power and efficiency in a range of such applications. Microsatellites are generally described as short stretches of DNA, containing tandem repeating blocks of di-, tri-, tetra- or penta- nucleotides, ubiquitously distributed throughout in both eukaryotic and prokaryotic genomes. They exhibit range of unique characteristics such as high reproducibility, multi-allelic nature, co-dominant inheritance and wide genome coverage and hence have gained a wide acceptance among plant breeders (Gupta and Varshney, 2000). In addition they are thought to be one of the major sources of genetic variation in quantitative traits. These variations are originate from the unequal crossing-over or replication errors resulting in formation of unusual DNA secondary structures such as hairpins or slipped strands mis-paring (Pearson and Sinden, 1998).

One major drawback of microsatellites is their species-specific nature, which necessitates initial isolation and characterization of the markers from the species of interest. Once developed, practically, it is a rapid, robust, simple and highly sensitive PCR-based system, with a low demand of technical inputs. The traditional method for isolation of SSR loci is development of genomic libraries with small genomic fragments and screening several thousand of the genomic clones with appropriate repeat-containing probes (Rassmann *et al.,* 1991). This method is relatively inefficient, and even more so for plant or bird species, which usually exhibit a low frequency of microsatellites (Zane *et al.,* 2002). A

number of protocols have been developed that to address and overcome various limitations in isolation and characterization of genomic SSRs (Wu *et al.,* 1994; Zietkiewicz *et al.,* 1994; Provan *et al.,* 1999; Witsenboer *et al.,* 1997; Padegimas, 1998; Hayden *et al.,* 2001; Hamilton *et al.,* 1999; Kandpal *et al.,* 1994). Among them, use of microsatellite enriched libraries is a reliable and widely employed method, especially for taxa with low microsatellite frequency (Zane *et al.,* 2002). Development of EST (Expressed Sequence Tag) SSRs is an alternative option to genomic SSRs, but is limited to species for which sequence databases already exist.

Application of molecular marker techniques for tea is comparatively new, with only a few instances reporting use of RAPD, AFLP, ISSR and isozymes mainly for assessment of genetic diversity (Wachira *et al.,* 1995; Wright *et al.,* 1996; Tanaka and Yamaguchi 1996; Chen *et al.,* 1998; Mewan *et al.,* 2001; Mewan *et al.,* 2005; Mewan, 2011). The value of dominant markers such as RAPD and AFLP, or co-dominant markers such as isozymes, in genetic diversity studies, parentage/ hybrid analysis, construction of genetic frame-work/ linkage/ QTL maps, gene tagging and Marker Assisted Selection (MAS), genetic characterization and fingerprinting is limited especially in closely related individuals like tea. In contrast, microsatellites/ SSRs have proven useful as an alternative and highly reproducible marker system, especially for closely related individuals (Bowcock *et al.,* 1994). Although the potential is enormous, the major constrain in application of this marker system to tea is limited availability of tea-specific SSR markers. Transferability of SSR loci for cross-species amplification has also been reported (Scribner and Pearce, 2000; Saha *et al.,* 2004; Peakall *et al.,* 1998; Gaitàn-Solịs *et al.,* 2002; Dirlewanger *et al.,* 2002; Eujayl *et al.,* 2004), but this is not applicable for tea as there are no such reports on closely related species with developed SSR primers.

Although initial isolation and characterization is laborious, genomic SSRs provide the most comprehensive and accurate evaluation of the distribution and abundance of microsatellites in a particular genome and are, therefore, considered the best source of markers for many SSR-based applications. In contrast, since EST-SSRs are related to coding regions, reported frequency and degree of polymorphism are usually less than for genomic SSRs (Wang *et al.,* 1994; Cho *et al.,* 2000; Morgante *et al.,* 2002). Therefore, large scale genomic and EST sequencing enable to assess the relative abundance and distribution of microsatellites in non transcribed and transcribed regions in the genome.

Therefore considering the importance and relevance to tea, in the present study, we report the isolation and characterization of genomic and EST-SSRs, as a first step towards construction a comprehensive set of genomic and genic (EST) SSR

primers for tea and exploitation of their potential and usefulness in future applications in tea.

MATERIALS AND METHODS

1. Isolation and characterization of genomic and EST SSRs of tea

a. Isolation and characterization of genomic SSRs

Construction, enrichment and sequencing of genomic DNA Libraries: Tea cultivar, TRI2023 was used for construction of genomic library and genomic DNA was extracted using the DNeasy® Plant Mini DNA Extraction Kit (QIAGEN Inc. Valencia, CA) and quality and the quantity were checked by spectrophotometer (NanoDrop ND 1000 spectrophotometer equipped with the software Nano Drop 1000 ver. 3.1.2) and 1.2% agarose/TAE gel electrophoresis. The microsatellite enrichment procedure was carried out following the steps described in the protocol by Hamilton *et al.,* (2002) (http:// www.bioserver. georgetown.edu/ faculty/hamilton). Two synthetic biotinylated oligonucleotides, $(AC)_n$ and $(CTT)_n$, were used for the capture and enrichment. Transformed cells were plated onto IPTG/X-gal supplemented LB-agar plates containing 100 µg/ ml ampicillin. SSR positive clones were directly selected using X-gal/ IPTG by blue/ white colony selection followed by determination of insert size through colony PCR using randomly selected white & blue colonies. To check for the presence of SSR motifs within the inserts, some of the colony PCR carried out using randomly selected SSR positive clones were subjected to direct sequencing using T_3 and T_7 universal primers. Plasmid DNA was extracted from the white colonies using Biomek 2000 robotic work station using the protocol Biomek 2000 96-well plasmid double stranded DNA isolation for DNA sequencing templates (http://www.noble.org/ PlantBio/ Genomics/ ProtocolBiomek.htm) and sequencing was performed using T_3 and T_7 primers.

b. Isolation and characterization of EST-SSR

Construction and sequencing of TRI2043 and DT1derived cDNA Libraries: Total RNAs was isolated separately from leaves of the tea cultivars TRI2043 and DT1, using the CTAB method (Jaakola *et al.,* 2001) and poly (A)+ RNA isolated using an Oligotex mRNA Mini Kit (Qiagene, Valencia, CA). The Creator Smart cDNA Library Construction Kit (Clontech, Mountain View, CA) was used for the cDNA library construction according to the manufacturer's instructions (Yongzhen Pang *et al.,* unpublished data). Colonies were randomly picked for overnight culture in Terrific Broth liquid media supplemented with 30 mg/ l chloramphenicol. Biomek 2000 robots were used to isolate sequencing templates, and sequencing was performed on an ABI 3730 sequencer. An automated in-house NEST-PIPE (Noble internal EST-PIPline) system was used for the analysis of tea EST sequences. Features of the NEST-PIPE include sequencing error detection, sequence cleaning, vector removal and sequence assembly.

c. Detection and construction of genomic and EST-SSR primers

After deleting the DNA sequences corresponding to vector arms and SNX linkers, all the insert sequences were then subjected to search for SSRs using the software phpSSRMiner: Simple Sequence Repeat Marker Miner (http://bioinfo.noble.org/ phpSSRMiner/design index?lid=6). Among all the SSR positive sequences, EST sequences having di-, tri-, penta- and hexa-nucleotide repeat motifs with more than 9, 6 and 4 repeat units respectively were used for primer designing where as for genomic inserts, both perfect and imperfect repeats with di-, tri-, tetra- and penta- nucleotide repeat sequences (with a minimum of 10, 7, 6 and 4 repeat motifs, respectively) were used. Sequences with good flanking regions on either sides of the repeat were used for designing of primers using primer3 program of phpSSRMiner software. Primer design criteria were based on the length (min18- max 24 bp), GC content (min 25- max 60,) melting temperature (min 52 ^{0}C- max 68 ^{0}C), primer product size (100-500 bp) and lack of secondary structures. SSR forward primers were modified by 5' concatenation of the 18mer 5'tgt aaa acg acg gcc agt3', which permitted concurrent fluorescence labeling of PCR products by a third primer with an incorporated fluorophore (Schuelke, 2000).

2. Evaluation of primers in genetic conservation and improvement of tea

a. Genetic conservation studies

The aim of following studies was to evaluate the usefulness of SSR primers to genetic studies on tea. For this purpose we used two diverse set of accessions *i. e.* improved tea cultivars, with known pedigrees and a set of old seedlings, which suppose to have 'China' origin.

Assessment of genetic diversity of improved tea cultivars: To check the readiness of the primers constructed for genetic diversity studies, 20 randomly selected SSR primers from above, comprising 10 genomic (designated as GMST) and 10 EST-SSR (designated as EMST) were assayed using 27 tea accessions which included twenty five TRI recommended improved cultivars, one estate selection and one phenotipically distinct tea accession. Genomic DNA was extracted from approximately 50 mg of freeze dried tender leaves from each genotype using Wizard Genomic DNA Purification kit (cat #A1120, Promega, USA) following manufacturer's instructions and quality and the quantity was checked on 1.0% agarose (1×TAE) electrophoresis and spectrophotometer.

PCR amplifications were carried out on Gene Amp 9700 thermo cycler (Applied Biosystems) in a final volume of 20 µl, containing 20-30 ng of genomic DNA, 0.2 µl of reverse primer (10 µM), 0.4 µl of forward primer (5 µM), 10 mM $MgCl_2$ (Promega), 2 mM each dNTP (Promega), 0.2 U of *Taq*-polymerase (Promega) with 1× PCR buffer (Promega) using following thermal

profile: initial denaturation at 94 ^{0}C for 1min. followed by 35 cycles each consisting 45 sec at 94 ^{0}C, primer annealing 1min at T_m ^{0}C (annealing temperature for each primer pair) followed by 45 sec at 72 ^{0}C for elongation with a final extension of 7 min at 72 ^{0}C. Amplified PCR products were separated on polyacrylamide (1×TBE) gels (Bio-Rad, protean^{TM} 11 Slab cell) stained with silver nitrate (Bassm *et al.,* 1991). Only intensely stained, unambiguous bands were visually scored for presence (1) and absence (0) and assembled in a data matrix. Nei and Li's (1979) similarity coefficient was employed to generate a genetic distance matrix. Cluster analysis (Unweighted Pair Group Method with Arithmetic Means - UPGMA) based on Nei and Li's (1979) genetic similarity coefficient was used to construct a dendrogram using DRAWGRAM function PHYLIP 3.64 software package.

Assessment of genetic diversity of old seedling teas: A total of 30 accessions which include 24 old seedlings (OSDs) were assessed using 20 randomly selected SSR primer pairs *i.e.* 10 Genomic (designated as GMST) and 10 EST (designated as EMST). Genomic DNA was extracted from approximately 50 mg of freeze dried tender leaves from each genotype using DNeasy Plant Mini Kit (QIAGEN Inc, Valencia, CA) according to the manufacturer's instructions. Quality and the quantity of extracted DNA were tested by 1% agarose gel containing 0.5 µg/ ml EtBr in 1X TAE buffer on Biorad Gel Documentation System, USA

PCR reactions were carried out on Gene Amp 9700 thermal cycler (Applied Biosystems, USA) in a final volume of 25 µl containing, 20-30 ng genomic DNA, 5X Go taq flexi buffer (Promega), 5 µl, 1.8 µl of 25 mM $MgCl_2$ (Promega), 2 mM each dNTPs (Promega), 10 pmol forward and reverse primers, and one unit of Go Taq Flexi DNA Polymerase (Promega) in a thermal profile of initial denaturation (1min at 94 °C) followed by 35 cycles of denaturation (at 94 °C for 45 sec), annealing (at respective Tm for 1 min), primer extension (at 72 °C for 60 sec) and final extension at 72 °C for 7 min. PCR products were separated on 6% polyacrylamide gel electrophoresis (1X TBE) and visualized by silver staining (Bassm *et al.,* 1991). Only intensely stained, unambiguous bands were visually scored for presence (1) and absence (0) and assembled in a data matrix. Nei and Li's (1979) similarity coefficient was employed to generate a genetic distance matrix. Cluster analysis (Unweighted Pair Group Method with Arithmetic Means - UPGMA) based on genetic distances was used to construct a dendrogram using DRAWGRAM function of PHYLIP 3.64 software package.

b. Genetic improvement studies

The aim of following study was to evaluate the informative and effectiveness of above developed genomic and EST-SSR loci in genetic mapping of tea.

Testing SSR primers for genetic mapping: 384 primer combinations consisted of 192 genomic and 192 EST-SSR primers were screened with a panel of genomic DNA from 8 individual tea plants (including two parents and a subset of six randomly selected individuals from the progeny lines) to assess the level of polymorphisms of SSR primers. The whole idea was to evaluate the readiness of primers for genotyping of mapping population towards construction of genetic map.

Genomic DNA was extracted from above individuals using DNeasy® Plant Mini DNA Extraction Kit (QIAGEN Inc. Valencia, CA). PCR reactions were performed in a final volume of 20 µl, containing 20 ng of template DNA, 0.2 µM of reverse primer, 0.2 µM forward primer, 2.5 mM each dNTPs (Promega), 1x buffer (GeneScript Corp, Piscataway WI), 0.09 units *Taq* DNA polymerase (GeneScript corp, Piscataway WI) and 1 µM M13 Dye (Applied Biosystems, Foster City, CA). PCR reactions were carried out in an Applied Biosystems Geneamp 9700 thermocycler with the following thermal profile: 5 min at 95 ^{0}C, followed by 30 cycles of 30 sec at 95 ^{0}C, 45sec at optimum annealing temperature for each reverse primer, 45 sec at 72 ^{0}C followed by another 10 cycles of 30 sec at 95 ^{0}C, 45 sec at 53 ^{0}C, 45 sec at 72 ^{0}C with final extension of 10 min at 72 ^{0}C. The amplified products were separated by capillary electrophoresis on an ABI 377 DNA sequencer (Applied Biosystems) using five color dye chemistry. Sizes of amplified alleles were determined using the software GeneMapper ver 3.0 with the size standard LIZ®™ 500 (Applied Biosystems). Amplification data for each primer were analyzed using ABI Prism Gene Scan Analyses (ver.2.1, Applied Biosystems) and the resulting GENESCAN trace files were imported into GeneMapper Ver 3.0 which were finally transformed to binary mode using scores 1/0 for presence/ absence of the allele, respectively.

RESULTS AND DISCUSSION

1. Isolation and characterization of genomic and EST SSR of tea

a. Isolation and characterization of genomic SSRs

Out of 1,120 genomic sequences generated, 945 (84.3%) good quality clean sequences were used for mining of SSRs, 68% of that contained 1739 SSR motifs with a motif length ≥ 10 bases, with a abundance of 2.7 SSR/ sequence. Similar results were reported for tall fescue (70%- Saha *et al.*, 2006) and for white clover (71%- Kölliker *et al.*, 2001). However, SSR-positive sequences with repeat motif length ≥ 20 bases (339 sequences out of 945- 35.9%) were used for SSR characterization. Of the above, 275 (81%) sequences contained

perfect SSRs where as 252 (74%) were with imperfects. Accordingly, 331 perfect and 294 imperfect repeats were detected and a total of 305 genomic SSR primers, which consisted of 157 (47.4%) perfect and 148 (50.3%) imperfect microsatellite were constructed using the software (Figure 1). Three hundred and five of the 625 candidate SSRs (≥ 20 base pairs) that contain in 339 sequences were converted into SSR primers (30% of the total quality genomic sequences), representing a substantial improvement over previous reports on both enriched (9% for tall fescue -Saha *et al.*, 2006; 10% white clover-Kölliker *et al.*, 2001) and non-enriched genomic libraries (0.4% - Liu *et al.*, 1995; Kubik *et al.*, 1999).

Majority of the SSRs identified were of di-nucleotide in nature (54.4%) followed by tri- nucleotides (29%). However a considerably higher number of di- repeats were observed in perfect (66.9%) SSRs than to imperfect (38.6%) SSRs. Tri- nucleotides were represented almost similarly in both perfect and imperfect SSRs (27% and 28% respectively) where as a comparatively higher percentage of longer repeats, namely hexa- (8.23%), hepta- 6.33%) and octa- (8.23%) were observed especially among imperfect SSRs (Table 1). The frequencies of different genomic microsatellites vary significantly among different organisms (Morgante and Oliveri, 1993; Wang *et al.*, 1994; Gupta *et al.*, 1996). The results obtained from the above study is in accordance with the reported fact that, of the majority of SSRs (48-67%) found in many species are of di-nucleotides (Wang *et al.*, 1994; Schug *et al.*, 1998). Furthermore, it has also observed that tri-nucleotide and tetra-nucleotide repeats are also found abundantly in plant genomes as well (Gupta *et al.*, 1996).

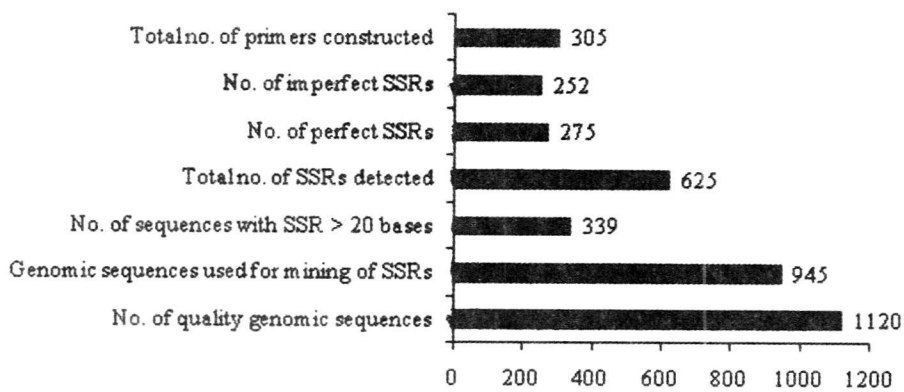

Figure 1. Statistics on genomic sequences and types of SSRs detected from the enriched genomic library developed using tea cultivar TRI 2023

Table 1. Distribution of the different types of perfect and imperfect SSR characterized from the genomic library of tea cultivar TRI 2023

Motif type	SSR type				Total	
	Perfect		Imperfect			
	No. of sequences	%	No. of sequences	%	No. of sequences	%
Di	105	66.88	61	38.61	166	54.4
Tri	43	27.39	45	28.48	88	29.0
Tetra	1	0.64	-		1	0.33
Penta	2	1.27	2	1.27	4	1.3
Hexa	2	1.27	13	8.23	15	5
Hepta	1	0.64	10	6.33	11	3.6
Octa	1	0.64	14	8.86	15	5
Nano	-		3	1.9	3	1
Deca	2	1.27	-		2	0.7
Total	157	100	148	100	305	100

b. Isolation and characterization of EST-SSRs

Of 6267 ESTs characterized, 4490 were generated from the cDNA library constructed from tea cultivar TRI2043, and the rest from the library of the cultivar DT1. Out of above, 2718 sequences (1322 from DT1 and 1396 from TRI2043) were subjected for SSR mining, 571 (21%) were identified as containing both perfect and imperfect SSRs with repeat motifs length ≥ 18 bases. The percentage contribution to above 571 from the cDNA library developed from DT1 was comparatively significant (64.9%) than to the cDNA library of TRI2043 (35.1%). A total of 777 SSRs were identified which consisted of approximately equal amounts of perfect (385-14.2%) and imperfect SSRs (392-14.4%) (Figure 2). For the cDNA library of DT1, majority of the SSR detected was of imperfect types (390 imperfect 156 perfect) where as for TRI2043 the situation was *vise-versa* (229 perfect and 2 imperfect).

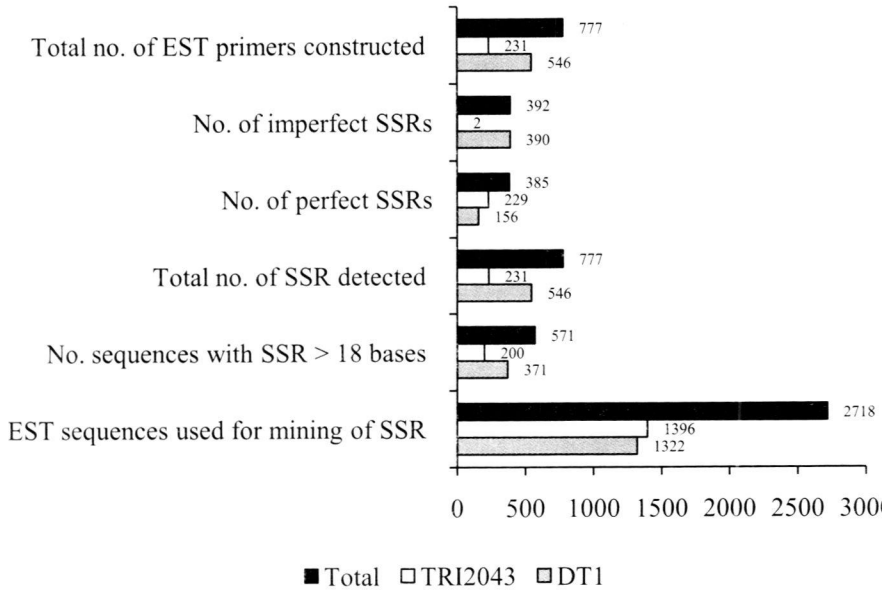

Figure 2. Graphical presentation of EST sequence data and EST-SSRs isolated from tea cDNA libraries developed from cultivar TRI2043 and DT 1

Within all the EST-SSR detected (both perfect and imperfect), the majority consisted of hexa- (35% of the total) nucleotide motifs followed by di- (26.5%) nucleotides and a comparatively low abundance of tri- nucleotides (11%) were seen (Table 2). However, a considerably higher number of nano- (7.6%) and deca- (5.5%) nucleotide repeats was observed for the EST libraries than the genomic library (1% and 0.7% respectively).

For the ESTs developed from the library DT1, contrastingly higher percentage of imperfect repeats (71.4%) was observed than perfect repeats. Furthermore, hexa nucleotide repeats were the most abundant (41.9%) among imperfect repeats, followed by hepta-, di- and tri- repeats (Table 2) whereas di- (38.5%), hexa- (19.2%) and nano (16.6%) were the most abundant for the perfect. Similarly, for the TRI 2043 library, we observed significantly higher percentage of di- repeats (46.7%) under perfect category, where tri- (18.9%) and hexa- (18.6%) remained as the 2nd and 3rd abundant respectively.

Table 2. Distribution of the different types of perfect and imperfect EST-SSR detected in two cDNA libraries constructed from tea cultivars DT1 and TRI2043

Motif Type	No. of EST-SSRs								
	from cDNA library of cultivar DT1			from cDNA library of cultivar TRI2043			Both libraries		
	Perf	Imperf	Total	Perf	Imperf	Total	Perf	Imperf	Total
Di	60(38.5)†	39(9.4)	99(18)	107(46.7)	-	107(46.7)	167(43.72)	39(9.4)	206(26.5)
Tri	16(10.4)	32(8.3)	48(8.8)	43(18.9)	-	43(18.9)	59(15.34)	32(8.3)	91(11.7)
Tetra	0	6(1.6)	6(1.1)	4(1.76)	-	4(1.76)	4(1.0)	6(1.6)	10(1.3)
Penta	8(5.12)	20(5.2)	28(5.1)	10(4.4)	-	10(4.4)	18(4.7)	20(5.2)	38(4.9)
Hexa	30(19.2)	199(51.7)	229(41.9)	41(18.0)	2	43(18.6)	71(18.5)	201(51.7)	272(35)
Hepta	1(0.6)	41(10.7)	42(7.7)	1(0.4)	-	1(0.4)	2(0.5)	41(10.7)	43(5.5)
Octa	0	15(3.9)	15(2.7)	0	-	0	0	15(3.9)	15(1.9)
Nano	26(16.6)	15(3.9)	41(7.5)	18(7.9)	-	18(7.9)	44(11.4)	15(3.9)	59(7.6)
Deca	15(9.6)	23	38(7)	5(2.2)	-	5(2.2)	20(5.2)	23(6.0)	43(5.5)
Total	156	390	546	229	2	231	385	392	777

†Values in parenthesis give the percentage from the total

It has been reported that all types of SSRs except triplets and hexa- nucleotides are significantly less frequent in the protein coding sequences compared with the non-coding fraction in six plant species (Morgante *et al.,* 2002). Therefore relatively high abundance of di nucleotide repeats in ESTs developed from TRI2043 is interesting.

When consider genomic and EST libraries (Table 3), comparatively high relative abundance as well as frequency (*i.e.* no. of SSRs per Sequence) were observed for genomic SSRs (35.9% and 1.84 respectively). As the EST-derived microsatellites are found within transcribed regions of the genome these markers may be less polymorphic than those from un-transcribed regions, but possible conservation of the primer sites could make them more transferable across species (Varshney *et al.,* 2005). Majority of the repeats detected in genomic SSRs represented by di- and tri- repeats (83.3% of the total) and hexa- and di- repeats (61.5%) for EST-SSRs. Among the EST derived SSRs, availability of tri- nucleotides was comparatively low (11.7%) as compared to the genomic library (29%).

Table 3. Comparison of statistics on efficiency, relative abundance and major types of SSRs observed in genomic- (with motifs ≥ 10 repeat units) and EST–SSRs (with repeat motifs ≥ 9 repeat units) characterized from tea

		Genomic SSR	EST-SSR
a.	No. of sequences used for SSR mining	945	2718
b.	Relative abundance of SSRs		
	- SSR positive sequences used for SSR mining	339 (35.9) #	571 (21.0) #
	- Frequency (SSR/sequence)	1.84	1.36
	- Perfect	331 (53%)#	385 (50%)#
	- Imperfect	294 (47%)#	392(50%)#
c.	Most abundant SSRs		
	- Motif types and its' percentage	Di (54.4%)# Tri (29%)#	Hexa (35%)# Di (26.5%)# Tri (11.7%)#

Value in parenthesis give the percentage from the total

For genomic libraries, the promonent SSR motif was the expected types i. e. di- AC/TG and tri- CTT/GAA. In addition to those, considerable amount of TC/AG repeats were seen. Interestingly, among all the other types of di- repeat motifs, above two types were the (*i. e* AG/TC and TC/Ag respectively) most abundant in EST-SSRs as well (Table 4). Approximately 62% of the characterized SSRs were of enriched types. Thus, this enrichment procedure was an efficient system to generate a large number of SSR loci. Although AT/TA is the most abundant di-nucleotide motif recorded in plants (Cardle *et al.,* 2000), this motif is not usually use in SSR enrichment procedures since to its self-complementary nature, hence was not included in this study. The next most common in the plant was GA/TC and CA/TG (Cai *et al.,* 2003 & 2005; Cardle *et al.,* 2000) and this could be the reason for presence of higher amount of TC/AG in genomic library of present study.

Table 4. Distinct SSRs motifs observed in genomic- (with motifs ≥ 10 repeat units) and EST–SSR (with repeat motifs ≥ 9 repeat units) characterized from tea

	Motif type	Motif	Remarks
Genomic	Di	AC/TG	Type used for enrichment
		TC/AG	-
	Tri	CTT/GAA	Type used for enrichment
EST	Di	AG/TC	
		TC/AG	

c. Construction of genomic and EST-SSR primers

Out of 625 genomic SSRs isolated from 339 SSR positive sequences, a total of 305 (~49% from the total SSR isolated) were used for automated primer designing following the criteria described in Materials and Methods, where as all the SSRs containing EST sequences for generation of EST- SSR primers. Lack of adequate flanking sequences limited potential primer design to the rest of the genomic sequences.

The discovery of SSR markers in ESTs provided the opportunity to develop microsatellites in a relatively simple, fast and direct way. This has gain wide attraction among the researchers all over the world as an inexpensive source for isolation of microsatellites, but is inherently restricted to those species for which there are a sufficiently large number of EST databases available. EST-SSRs are useful as they represent transcribed regions of genes where putative functions can be often deduced by homology search to generate additional information on the loci. In addition, as EST-SSRs directly sample variation in transcribed

regions of the genome, they may provide an estimate of functional diversity. Future plans of our research involves mapping of chromosomal areas that confer resistance towards the most important fungal diseases of tea, blister blight, The sources we have used to develop EST-SSRs were related to above resistant trait and therefore, the EST-SSR primers developed from the present study may be of immense value towards potential tagging and mapping of resistant trait.

2. Evaluation of primers in genetic conservation and improvement of tea

a. Genetic conservation studies

Assessment of genetic diversity of improved tea cultivars: For 20 SSR primers, a total of 128 alleles were scored (57 from EST and 71 from genomic primers) and all alleles were 100% polymorphic among all 27 accessions (Table 5). Average number of alleles developed per primer was 6.4 and higher number of alleles was observed with genomic SSRs primers (average 7.1/primer) compared to EST primers (5.7/ primer).

Dendrogram generated using 128 SSR marker loci separated studied tea accession into two groups in which Group A comprised of open pollinated progeny of ASM 4/10 and group B mainly consisted of crosses derived from control hybridization. These results further strengthen the previous findings generated for the same accessions on their origins (Mewan *et al.,* 2005; Goonetilleke *et al.,* 2009; Mewan, 2011). Therefore above results clearly prove the fact that the primers developed in this study could be use very effectively to genetic diversity studies in tea. In addition, according to the results (not shown), it was evident that those primers could be used very accurately for other studies such as assessment of genetic relationships, exploitation of and effective management of germplasm, parentage and hybrid analysis of tea.

Table 5 . Level of polymorphism observed with SSR primers against improved tea accessions

	EST primers	Genomic primers	Total
Total no. of primers	10	10	20
Total no. of bands	57	71	128
Polymorphism %	100	100	100
Average no. of alleles/ primer	5.7	7.1	6.4

Assessment of genetic diversity of old seedling teas: A total of resulted 139 alleles were scored for 20 SSR primer-pairs (PP) with an average of 7.0 alleles/ primer, out of which 87.5% exhibited polymorphism (Table 6). Both the highest number of alleles (11 alleles) and monomorphic alleles (3) were resulted with

PP EMST 73 whereas maximum number of polymorphic alleles was resulted with the PP GMST 118 (8 alleles). The average number of polymorphic alleles per primer was 6.1. The average number of total and polymorphic alleles generated by EMST primers (6.5 and 5.1 respectively) was higher than to that of the GMST (7.4 and 7.0 respectively) primers whereas the percentage of polymorphism was higher in GMST primers.

Table 6. Level of polymorphism observed with SSR primers against old seedling tea accessions

	EST-SSR primers	Genomic - SSR primers	Total
Total no. of primers	10	10	20
Total no. of bands	65	74	139
Polymorphism %	79%	95%	87.5%
Average no. of alleles/ primer	6.5	7.4	7.0
Average no. of polymorphic alleles/ primer	5.1	7.0	6.1

The dendrogram separated accessions into two groups where group I consisted mainly with OSD accessions from Great Western, and Chelsy/ Maha Eliya estates. The accessions 'China' and DT1 were also clustered into the same group, reflecting the origin of OSDs close to 'China' type. Group II consisted mainly with 'TRI cultivars' together with OSDs collected from St Coombs (which having characters closer to 'Cambod' type) and Logie estates reflecting their close affinity to 'Cambod' type. Although it involved a small number of SSR loci, the results generated through this study clearly reflected the power and the effectiveness of above developed primers to be used in SSR marker based studies related to tea.

b. Genetic improvement studies

Genetic mapping:

Primer polymorphisms: Among 384 primers tested (192 genomic and 192 EST SSR primers), 330 (173 EST-SSR and 157 genomic- SSR) produced amplification products across a panel of eight individuals including two parents (TRI 2043 X TRI 2023) and six individuals randomly selected from the progeny. Out of that 158 EST and 141 genomic- SSR primers (a total of 299 SSR primers) resulted with strong amplification products. Among the 299 primers, 80 (39 genomic and 41 EST-SSR primers) showed mono-morphic amplifications across test individuals and thus were not selected for further studies towards map construction (Table 7).

The number of markers obtained from each SSR primer pair varied from one to ten. On an average, each of the genomic SSR primers produced 3.4 markers and the EST-SSR primers generated 2.5 markers. In general, more EST-SSR primers produced 1-4 markers than did genomic-SSRs, and the reverse scenario was observed for 5-10 markers per primer (Figure 3).

Table 7. Summary of the prescreening of EST and genomic SSR primers across two parents (TRI2043 x TRI2023) and six individuals from the progeny. Values in the parenthesis indicate the percentage against total screened.

No. of primers	Genomic-SSRs	EST-SSRs	Both types
Pre-screened	192	192	384
Without any amplification	35 (18.2)	19 (9.9)	54 (14.1)
With amplification products	157 (81.8)	173 (90.1)	330 (85.9)
With poor amplifications	16 (8.3)	15 (7.8)	31 (8.1)
Total no. with good products	141(73.4)	158 (82.3)	299 (77.9)
With mono-morphic products	39 (20.3)	41 (21.4)	80 (20.8)
With poly-morphic products	102 (53.1)	114 (59.4)	219 (57.0)

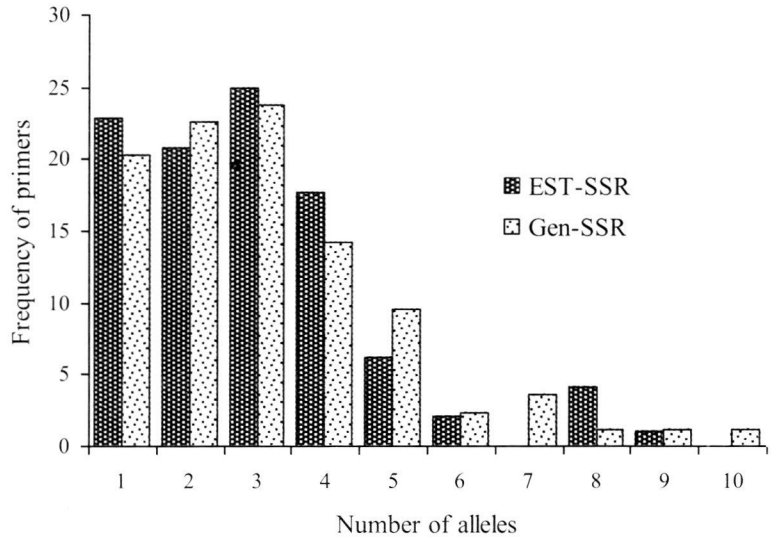

Figure 3. Number of alleles obtained from different EST- and genomic (Gen)-SSRs primers

Polymorphic information content of the primers: Polymorphic information content values for the EST-SSRs and genomic-SSRs ranged from 0.40-1.00 and from 0.18-1.00 for all primer pairs tested, respectively (Table 8).

Table 8. Polymorphic information content of the microsatellite markers generated from EST- and genomic-libraries

PIC*	EST-SSR	Genomic-SSR
0.21 - 0.40	1 (1)†	4 (5)
0.41 - 0.60	23 (24)	18 (21)
0.61 - 0.80	44 (46)	36 (43)
0.81 - 0.99	6 (6)	9 (11)
1.00	22 (23)	17 (20)
Range	0.40-1.00	0.18-1.00
Mean	0.64	0.64

* PIC = Polymorphic information content
† figures in parenthesis indicate percent values

Both marker sources had the same mean PIC values (0.64). The lowest PIC content for EST-SSRs was 0.40 (TEST004) where as that of genomic SSRs was 0.18 (TGSR138). In general, PIC values for the genomic SSR primers varied widely with the highest frequency for both less (5%) and highly (11%) informative markers. Among the EST-SSRs; TESR002, TESR008, TESR115, TEST179, TESR183 and TESR191 were found highly informative with PIC values above 0.80. TGSR024, TGSR105, TGSR138 and TGSR156 were the least (PIC <0.40) and TGSR005, TGSR011, TGSR022, TGSR036, TGSR053, and TGSR134 were the most (PIC >0.85) informative genomic-SSR primers (data not shown). Although, the mean PIC score for both marker systems was same, higher percentage of genomic SSRs was found highly polymorphic than the EST-SSRs.

The PIC is an estimate of the ability of a marker to differentiate genotypes based on both the number of alleles expressed and their relative frequencies (Weir 1990). Both microsatellite marker sources had similar mean PIC values (0.64). The similarity of mean PIC scores for both marker systems implies that the two marker sources had similar power to differentiate the accessions tested. However, both the least (TGSR138, PIC = 0.18) and highly informative (TGSR011, PIC = 0.90) primers were observed in genomic SSRs. Higher

percentage of genomic SSRs was found more informative than the EST-SSRs. The PIC values depend on the number and frequency of alleles per primer pairs. The SSR loci are multi allelic and co-dominant and therefore expected to have high PIC values. Previous studies indicated that EST SSR markers have lower PIC values relative to genomic SSRs and thus less informative (Cho *et al.,* 2000; Eujayl *et al.,* 2002). However, the tea EST-SSR primers developed in this study were also found highly informative. Saha *et al.,* (2004) and Varshney *et al.,* (2005) also found their EST-SSRs highly informative in tall fescue, and barley, respectively.

Therefore, based on above information, genomic and EST primers developed from above study would be of a great source, which has the potential to use in genetic mapping and marker assisted breeding of tea.

CONCLUSIONS

Use of enriched genomic libraries and EST sequencing is efficient and appropriate approach for isolation and characterization of microsatellites, especially for the species where there is no such information on markers/ primers available or sequence information. The microsatellite primers constructed from two SSR sources are informative enough and with good genome coverage for genetic studies well as for genetic mapping and marker assisted selection and hence, could be used very effectively to enhance tea improvement program through marker assisted breeding.

ACKNOWLEDGEMENT

This work was supported by a grant from the National Science Foundation, Plant Genome Program (DBI-0109732-004) to RAD.

REFERENCES

Bassm B J, Anolles G C and Gresshoff P M 1991 Fast and Sensitive Silver Staining of DNA in Polyacrylamide gels. Analytical Chemistry. 196, 80-83.

Bowcock A M, Ruiz-Linares A, Tomfohrde A, Minch E, Kidd J R and Cavalli-Sforza L L 1994 High-resolution human evolutionary trees from polymorphic microsatellites. Nature. 368, 455- 457.

Cai H W, Yuyama N, Tamaki H and Yoshizawa A 2003 Development and characterization of simple sequence repeat (SSR) markers in hexaploid forage grass timothy (*Phleum pretense* L.). Theor. Appl. Genet. 107, 1337-1349.

Cai H W, Inoue M, Yuyama N, Takahashi W, Hirata M and Sasaki T 2005 Isolation, characterization and mapping of simple sequence repeat markers in Zoysia (*Zoysia* spp.). Theor. Appl. Genet. 112 (1), 158-166.

Cardle L, Ramsay L, Milbourne D, Macualay M, Marshall D and Waugh R 2002 Computational and experimental characterization of physically clustered simple sequence repeats in plant. Genetic. 156, 847-854.

Chen L Y, Yang J, Yu F L, Gao Q K and Chen D M 1998 Genetic diversity of 15 tea (*Camellia sinensis* L. O. Kuntze) cultivars using RAPD markers. Journal of Tea Science. 18, 21-27.

Cho Y G, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch S R, Park W D, Ayres N and Cartinhour S 2000 Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet. 100, 713-722.

Dirlewanger E, Cosson P, Tavaud M, Aranzana M, Poizat C, Zanetto A *et al.*, 2002 Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). Theor. Appl. Genet. 105(1), 127-138.

Doyle J J and Doyle J L 1990 Isolation of plant DNA from fresh tissue. Focus. 12, 13-15.

Eujayl I, Sledge M, Wang L, May G D, Chekhovskiy K, Zwonitzer J C and Mian M A 2004 *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for Isolation of EST- derived microsatellite markers for *Medicago* spp. Theor. Appl. Genet. 108, 414- 422.

Eujayl I, Sorrells M E, Baum M, Wolters P and Powell W 2002 Isolation of EST- derived microsatellite markers for genotyping the A and B genomes of wheat. Theor. Appl. Genet. 104:399-407.

Gaitán-Solís E, Duque M C, Edwards K J and Tohme J 2002 Microsatellite repeats in common bean (*Phaseolus vulgaris*): Isolation, characterization, and cross-species amplification in *Phaseolus* ssp. Crop Science. 42(6), 2128-2136.

Gupta P K, Balayan H S, Sharma P C and Ramesh B 1996 Microsatellites in plants: a new class of molecular markers. Curr. Sci. 70: 45-54.

Gupta P K and Varsheny R K 2000 The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica. 113, 163-185.

Hamilton M B, Pincus E L, DiFiore A and Fleischer R C 1999 A universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. BioTechniques. 27(3), 500- 507.

Hayden M J and Sharp P J 2001 Sequence-tagged microsatellite profiling (STMP): a rapid technique for developing SSR markers. Nucleic Acid Research. 2 (8), 43.

Jaakola L, Pirttila A M, Halonen M and Hohtola A 2001 Isolation of high quality RNA from bilberry (Vaccinium myrtillus L.) fruit. Mol. Biotechnol. 19, 201-203.

Kandpal R P, Kandpal G and Weissman S M 1994 Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. Proceedings of the National Academy of Sciences of the USA. 91, 88-92.

Kölliker R, Jones E S, Drayton M C, Dupal M P and Forster J W 2001 Development and characterisation of simple sequence repeat (SSR) markers for white clover (*Trifolium repens* L.). Theor. Appl. Genet. 102, (2-3), 416-424.

Kubik C, Meyer W A and Gaut, B S 1999 Assessing the abundance and polymorphism of simple sequence repeats in perennial ryegrass. Crop Science, 39, 1136–1141.

Litt M and Luty J A 1989 A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American Journal of Human Genetics. 44, 397- 401.

Liu Z W, Jarret R L, Kresovich S and Duncan R R 1995 Characterization and analysis of simple sequence repeat (SSR) loci in seashore paspalum (*Paspalum vaginatum* Swartz). Theor. Appl. Genet. 91, 47–52.

Mewan K M 2011 Ph. D. Dissertation, University of Colombo.

Mewan K M, Liyanage A C, Everard J M D T, Gunasekare M T K, Fernando U and Karunanayake E H 2001 Estimation of genetic relatedness of Tea *Camellia sinensis* L. clones in Sri Lanka using Isozyme polymorphisms and RAPD markers assessed by RAPD markers. Proceedings of 2001 International Conference on O-CHA (tea) culture and science, Sizuoka-Japan.

Mewan K M, Everard J M D T, Liyanage A C, Gunasekare M T K, Tirimanne T L S and Karunanayake E H 2005 Studying genetic relationships among tea (*Camellia sinensis* L.) cultivars using RAPD markers. S. L. J. Tea Sci. 70(1), 10-17.

Morgante M and Olivieri A M 1993 PCR-amplified microsatellites as markers in plant genetics. Plant J. 3, 175-182.

Morgante M, Hanafey M and Powell W 2002 Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genetics. 30:194-200.

Nei M and Li W H 1979 Mathematical model for study genetic variation in terms of restriction endonucleases. Proceedings of National Academy of Science., 74, 5267-5273.

Peakall R, Gilmore S, Keys W, Morgante M and Rafalski A 1998 Cross species amplification of soybean (*Glycine max*) simple sequence repeat (SSRs) within the genus and other legume genera: Implication for transferability of SSRs in plants. Molecular Biology and Evolution. 15(10), 1275-1287.

Pearson C E and Sinden R R 1998 Trinucleotide repeat DNA structures: Dynamic mutations from dynamic DNA. Curr Opin Struct Biol 8: 321–330.

Proven J, Russell J R, Booth A and Powell W 1999 Polymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. Molecular Ecology. 8, 505-511.

Rassmann K, Schlötterer C and Tautz D 1991 Isolation of simple sequence loci for use in polymerase chain reaction-based DNA fingerprinting. Electrophoresis. 12, 113-118.

Saha M C, Mian M A R, Eujayl I, Zwonitzer J C, Wang L and May G D 2004 Tall fescue EST-SSR markers with transferability across several grass species. Theor. Appl. Genet. 109, 783-791.

Saha M C, Cooper J D, Rouf M A, Chekhovskiy K and May G D 2006 Tall fescue genomic SSR markers: development and transferability across multiple grass species. Theor. Appl. Genet., 113(8), 1449-1458.

Schug M D, Wetterstrand K A, Gaudette M S, Lim R H, Hutter C M and Aquadro C F 1998 The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. Molecular Ecology. 7, 57-70.

Scribner K T and Pearce J M 2000 Microsatellites: evolutionary and methodological background and empirical applications at individual, population and phylogenetic levels. In A. Baker, (Ed.) *Molecular methods in Ecology* (pp. 235- 271). Backwell Science Limited, London.

Tanaka J I and Yamaguchi S 1996 Use of RAPD markers for the identification of parentage of tea cultivars. Bulletin of the National Research Institute of Vegetables, Ornamental Plants and Tea (Series B). 9, 31-36.

Tautz D 1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research. 17, 6463- 6471.

Varshney R K, Graner A and Sorrells M E 2005 Genic microsatellite markers in plants: features and applications. Trends in Biotechnology. 23:48-55.

Wachira F N, Waugh R, Hackett C A and Powell W 1995 Detection of genetic diversity in tea (*Camellia sinensis*) using RAPD markers. Canadian Journal of Genome. (38-2), 201-210.

Wang Z, Weber J L, Zhong G and Tanskley S D 1994 Survey of plant short tandem DNA repeats. Theor Appl Genet. 88: 1-6.

Weir B 1990 Genetic data analysis: Methods for discrete population genetic data. Sinauer Assoc. Sunderland, MA.

Witsenboer H, Vogel J and Michelmore R W 1997 Identification, genetic localization, and allelic diversity of selectively amplified microsatellite polymorphic loci in lettuce and wild relatives (*Lactuca spp.*). Genome. 40, 923-936.

Wright L P, Apostolides Z, Louw A I, Whittle A M and Khumalo F R B 1996 DNA fingerfrinting of clones. Proceedings of the 1st Regional Tea Research Seminar, Blantyre, Malawi. 44 -50.

Wu M 1994 Studies on genetic relationships among Korean native tea trees and physio-chemical properties of its green tea. Ph. D. Thesis, Korean University, Korea.

Zane L, Bargelloni L and Patarnello T 2002 Strategies for microsatellite isolation: a review. Molecular Ecology. 11, 11-16.

Zietkiewicz E, Rafalski A and Labuda D 1994 Genome fingerprinting by simple- sequence repeat (SSR) anchored polymerase chain reaction amplification. Genomes, 20. 176-183.