S. L. J. Tea Sci. 74(1), 09 - 18, 2009, Printed in Sri Lanka

Construction of a Genomic and EST Simple Sequence Repeats (SSRs) based Genetic Linkage Map of Tea (*Camellia sinensis* L.)

K M Mewan[1,4], M C Saha[2], Y Pang[3], I S B Abeysinghe[1] and R A Dixon[3]

([1]Biochemistry Division, Tea Research Institute of Sri Lanka, Talawakelle, Sri Lanka
[2] Forage Improvement Division, The Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, OK 73401, USA
[3] Plant Biology Division, The Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, OK 73401, USA

ABSTRACT

Construction of a genetic linkage map for tea will be an important tool which provides great potential for increasing precision and speeding up the tea breeding programs while reducing the labor and land dedicated for the field assessments. However, to date, information on detail genetic maps is not available. Therefore, the goal of this study was to construct a molecular map of tea using simple sequence repeat (SSR) markers derived from expressed sequence tags (ESTs) and enriched genomic libraries.

The mapping population used for this study included 148 F_1 plants derived from a cross between two genetically diverse, non-inbred diploid tea cultivars, TRI 2043 ( ) and TRI 2023 ( ). Two separate sets of 192 EST and 192 genomic SSRs primers were first screened with two parents and six F_1 progenies, and 216 (114 EST and 102 genomic) polymorphic primers were identified to use for genotyping the mapping population. Out of that, 190 (104 of the EST and 86 of the genomic SSR) primers were used for construction of the genetic linkage maps. Tea EST-SSR primers were found to be more polymorphic than the genomic SSRs. The male parent (TRI 2023) map was constructed with 83 EST-SSR and 97 genomic SSR loci. A total of 180 markers were grouped in 16 linkage groups (LGs) covering a total length of 1,227.2 cM with an average marker density of 6.8cM per marker. In the female parent (TRI 2043) map, 146 markers (70 EST and 76 genomic SSR loci) were clustered in 15 LGs spanning 1,018.1cM with an average distance of 7.0 cM between adjacent markers. Fifteen LGs are consistent with the basic chromosome numbers of tea.

Key words: *Camellia sinensis*, genetic map, linkage mapping, microsatellites, SSRS, tea

INTRODUCTION

For tea (*Camellia sinensis* L.) which is a predominantly out breeding woody perennial (2n=2x=30) (Sealy, 1958), use of conventional methods for genetic improvement, is comparatively less effective, slow progressing and costly procedure. Construction of saturated genetic linkage maps is a necessary and fundamental step towards increasing

precision and speeding up the breeding process while reducing the labor and land dedicated for the field assessments especially in a woody perennial crop such as tea (Brondani *et al.*, 1998). In addition, genetic linkage maps can also be developed for economically important traits governed by polygenes (Quantitative Trait Loci; QTLs) (Jones *et al.*, 1997; Kashi *et al.*, 1997) and associated markers can be used effectively for cultivar improvement through marker-assisted selection (MAS) (Lande and Thompson, 1990; Dudley, 1993).

Genetic linkage maps can be constructed with one or more types of DNA markers, such as Randomly Amplified Polymorphic DNA (RAPD), Amplified Fragment Polymorphic DNA (AFLP), Restriction Fragment Length Polymorphisms (RFLP) *etc*, where each of which have their advantages and disadvantages. However, the relatively more informative simple sequence repeats (SSRs) or microsatellites have advantages in terms of their power, reliability and ease of use in plant genome analysis, and hence have gained wide acceptance among breeders. Their relatively high frequency of abundance in most genomes, high rate of mutations, simple Mendelian inheritance, co-dominant nature, locus specificity and interspecies transferability make them the choice for DNA marker-based studies in many crop species (Morgante and Olivieri, 1993; Gupta *et. al.*, 1996; Jarne and Lagoda, 1996; Powell *et al.*, 1996).

Among the two types of SSRs, genic or EST-SSRs, are developed from the more conserved transcribed regions of the genome with known or putative functions, whereas genomic SSRs are isolated from genomic libraries covering whole genome of the organism. Therefore, EST-SSR markers are considered as gene target markers that may lead to direct allele selection and are of importance for comparative mapping where genomic SSRs are considered as a good source for genome analysis due to their genome wide coverage and high polymorphic nature (Zane *et al.*, 2002; Varshney *et al.*, 2005). Therefore use of both types of SSRs in a mapping study will definitely be more informative and effective.

The first genetic linkage map of tea was constructed using RAPD markers (Tanaka, 1996) which was further expanded with additional RAPD and AFLP markers (Hackett *et al.*, 2000). In this study the female parent map covered a total length of 1349.7 cM with an average distance of 11.7 cM between loci. Huang *et al.* (2005) recently developed AFLP linkage maps for female and male parents of a tea population covering a total length of 2457.7 cM and 2545.3 cM respectively. In addition, mapping populations used for the above studies were too small for precise mapping. Although a precise genetic map is not available for tea, the potential of maker assisted selection is clear and strong. Thus, the objective of this study is to construct a genetic linkage map for tea using EST and genomic SSRs.

MATERIALS AND METHODS

The mapping population

The mapping population used for this study was derived from the cross between two genetically diverse, heterozygous, and diploid genotypes of tea cultivars, TRI 2043 (♀) and TRI 2023 (♂). One hundred and forty eight F_1 plants were germinated from the seeds obtained from TRI 2043 and constituted the 'pseudo test cross' mapping population (Grattapaglia and Sederoff, 1994).

Extraction of DNA

Genomic DNA was extracted from approximately 200 mg of freeze dried tender leaves from each genotype using a DNeasy® Plant Mini DNA Extraction Kit (QIAGEN Inc. Valencia, CA) following the manufacturer's instructions. The quality and the quantity of the extracted DNA were checked using a NanoDrop ND 1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) equipped with the software Nano Drop 1000 ver. 3.1.2.

SSR Primers

EST-SSRs were developed from two cDNA libraries of the tea cultivars TRI 2043 and DT 1 (Yongzhen *et. al.*, unpublished data), and genomic SSRs were developed from the cultivar TRI 2023 (Mewan *et al.*, unpublished data) using an enriched protocol (Hamilton *et al.*, 1999). 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 (Schnelke, 2000).

Pre-screening of SSR primers

Two separate sets of 192 EST (designated as TESR) and 192 genomic (designated as TGSR) SSR primers were prescreened with two parents and a subset of six randomly selected individuals from the progeny lines to select polymorphic SSR primers (PSPs) with strong or good PCR amplification products to use in the mapping study.

Simple sequence repeat polymorphisms and SSR marker data: Out of 384 primers pre-screened, 216 (114 TESR and 102 TGSR) primers, that were polymorphic between the parental cultivars and segregated in the initial primer screening genotypes, were used for mapping study.

PCR conditions

PCR reactions were performed for all selected PSPs in a final volume of 20 µl, containing 20 ng of template DNA, 1µM of reverse primer, 0.25 µM forward primer, 3.0 mM $MgCl_2$, 2.5 mM each dNTPs, 1 µM M13 fluorescent Dye (Applied Biosystems, Foster City, CA) and 0.09 units *Taq* DNA polymerase with 1x PCR Buffer (GeneScript Corp, Piscataway WI). PCR reactions were carried out in Applied Biosystems Geneamp 9700 thermocycler

(Applied Biosystems, Foster City, CA) using the following thermal profile: 5 min at 95 °C, followed by 30 cycles of 30 s at 95 °C, 45 s at optimum annealing temperature for each reverse primer, 45 s at 72 °C followed by another 10 cycles of 30 s at 95 °C, 45 s at 53 °C, 45 s at 72 °C with final extension of 10 minutes at 72 °C.

Detection and analysis of amplification products

Amplification products were resolved on an ABI 3730 capillary electrophoresis array (Applied Biosystems) with the size standard LIZ®™ 500 (Applied Biosystems) and detected using four color dye chemistry. Amplified alleles 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.

Genetic linkage analysis

Initially SSR markers within the size range of 100-500 bp in length, segregated from one parent or both, were extracted from the mapping population. The names of the individual SSR marker loci amplified were described using the SSR primer code (either as TESR 1 to 192 or as TGSR 1 to 192) followed by the relative allele size (bp). JOINMAP ver. 4.0 (http://www.kyazma.nl) was used for defining the parental linkage groups (LGs). Kosambi mapping function was applied for calculating the map distances (Kosambi, 1944). Following construction of the LGs, the graphics were structured using MAPCHART 2.1 (Voorrips, 2002).

RESULTS

SSR markers

Out of 384 primers tested, 216 (114 TESR and 102 TGSR) were polymorphic between the parental clones and segregated in the initial primer screening genotypes. Out of 216 used for genotyping of two parents and 148 individuals, a total of 190 PSPs consisting of 104 TESR primers (54% of the total pre-screened) and 86 TGSR primers (45% of the total pre-screened) were finally used to construct genetic linkage maps (Table 1).

The EST PSPs generated 264 polymorphic loci, with an average of 2.5 loci per PSP while the genomic PSPs generated 295 polymorphic loci with an average of 3.4 loci per PSP (Table 1). Overall the level of polymorphism was higher for EST-SSRs than for the genomic SSRs. However, genomic SSRs generated more loci per PSP than did EST-SSRs. In this study, 180 loci were incorporated to construct the TRI 2023 map and 146 loci were incorporated to the TRI 2043 map (Table 2).

TRI 2023 (♂) parental map

The TRI 2023 parental map was constructed with 83 EST-SSR loci and 97 genomic SSR loci. The 180 SSR marker loci were distributed among 16 LGs with genomic SSRs constituting the major part of the LGs. Markers in the 16 LGs covered a total length of

Table 1. Summary of the genomic and EST SSR primers and SSR loci used for the tea pseudo-testcross mapping population

	Genomic SSRs (TGSR)	EST – SSRs (TESR)	Total
No. of primers			
Pre-screened	192	192	384
Selected for mapping	102 (53%)	114 (59%)	216 (56%)
Used to construct linkage maps	86 (45%)	104 (54%)	190 (49%)
SSR loci			
Total no. alleles amplified	295	264	559
Average/primer	3.4	2.5	2.9
No. of alleles incorporated to parental maps	173	153	326

* Figures in the parenthesis indicate the percentage value from the total primers pre-screened.

1227.2 cM with an average of 11.25 loci per LG. The average size of a LG was 76.7 cM. The marker density ranged between 0.7 (LG 7) to 13.6 (LG 16) with an average density of 6.8 cM/ marker. The highest number of marker loci was observed in LG 3 (29 loci) with a marker density of 2.8 cM per marker whereas the lowest were LG 10 (5 loci) with marker densities of 3.2 cM/marker (Table 2). The highest number of genomic and EST SSR loci were distributed in LG 1 (12 and 11 respectively) and in LG 3 (14 and 15 respectively). Thirteen of the LGs were covered by both types of SSR loci whereas LG 7 only comprised EST-SSR derived loci (7 loci). Likewise, LG 9 and 14 comprised only 7 and 10 genomic SSR loci, respectively (Table 3).

TRI 2043 (♀) parental map

TRI 2043 map was constructed with a total of 146 SSR marker loci generated from EST-SSRs (70 loci) and genomic SSRs (76 loci), and all the markers were distributed on 15 LGs. The total length covered by all LGs was 1018.1cM with an average length of 67.9 cM/LG. The average marker density was 7.0 loci/ LG (Table 2). The marker density ranged between 0.8 (LG 9 with a map length of 5.6 cM and 7 marker loci) to 17.1 (LG 12 with a map length of 68.4 cM and 4 marker loci) with an average of 7.0 cM/ loci. The highest and lowest number of marker loci was observed in LG 2 (22 loci) and LGs 12, 14 and 15 with 4 loci each (Table 2). EST-SSR marker loci were predominant in LGs 6 (9 out of 10) and 8 (7 out of 11), while missing in LGs 9 and 15. LGs 1 (10 out of 16), 2 (16 out of 22) and 4 (10 out of 17) were enriched with genomic SSR derived marker loci, while LGs 10 and 12 did not have any genomic SSR loci. Eleven out of 15 LGs consisted of both kinds of SSR loci (Table 3).

Table 2. Distribution of EST-SSR and genomic SSR marker loci in male (TRI 2023) and female (TRI 2043) parental maps

	Parental map	
	TRI 2023 (♂)	TRI 2043 (♀)
No. of LGs	16	15
Total length	1227.2 (cM)	1018.1 (cM)
Distribution of SSR marker loci		
EST	83	70
Genomic	97	76
Total	180	146
Marker density		
Highest	7 markers in 4.9 cM (LG7)	7 markers in 5.6 cM (LG9)
Lowest	6 markers in 81.6 cM (LG 16)	4 markers in 68.4 cM (LG 12)
Range (cM)	0.5 – 13.6	0.8 – 17.1
Average	6.8 (cM/marker)	7.0 (cM/marker)
No. of loci/ LG		
Range	5 (LG10) – 29 (LG3)	4 (LGs12,14,15) – 22 (LG2)
Average	11.25	7.0
Size of LGs (cM)		
Range	4.9 (LG7) – 112.7 (LG1)	5.6 (LG9) – 105.0 (LG5)
Average	76.7	67.9

DISCUSSION

The pseudo test cross strategy employed here has been widely used for the construction of genetic linkage maps in out-crossing plant species such as apple (Hemmat *et al.*, 1994), eucalyptus (Grattapaglia and Sederoff, 1994), alfalfa (Echt *et al.*, 1994) and tea (Hackett *et al.*, 2000). Microsatellite or SSR markers have become the marker class of choice due to their manifold advantages over other marker classes and have been widely used for the molecular mapping of many crop species (Doligez *et al.*, 2006; Yi *et al.*, 2006; Dida *et al.*, 2007; Bindler *et al.*, 2007).

EST-derived markers derived from the transcribed region of the genome offer the opportunity for gene discovery and increase the value of genetic markers by surveying variation in transcribed gene-rich regions of the genome (Decroocq *et al.*, 2003; Varshney *et al.*, 2005). Genomic SSRs tend to be widely distributed throughout the genome resulting in better map coverage (Taramino *et al.*, 1997; Warnke *et al.*, 2004; La Rota *et al.*, 2005). Thus, a combination of EST- and genomic-SSR markers is an important strategy for genetic linkage mapping.

Table 3. Distribution of EST-SSR and genomic SSR marker loci in the linkage groups of the two parental maps of the tea pseudo-testcross mapping population

Linkage group	Parental map							
	TRI 2023				TRI 2043			
	EST SSR	Genomic SSR	Total	Marker density (cM/marker)	EST SSR	Genomic SSR	Total	Marker density (cM/marker)
1	11	12	23	4.9	6	10	16	6.3
2	5	10	15	6.3	6	16	22	3.9
3	15	14	29	2.8	2	6	8	8.9
4	7	6	13	5.9	7	10	17	6.1
5	8	3	11	9.9	7	8	15	7.0
6	4	8	12	8.1	9	1	10	9.6
7	7	0	7	0.7	3	4	7	7.4
8	3	3	6	9.8	7	4	11	7.1
9	0	7	7	10.4	0	7	7	0.8
10	2	3	5	3.2	5	0	5	1.6
11	3	3	6	10.1	6	3	9	7.7
12	5	1	6	12.2	4	0	4	17.1
13	6	9	15	6.0	6	1	7	9.9
14	0	10	10	9.5	2	2	4	14.2
15	3	6	9	11.4	0	4	4	11.7
16	4	2	6	13.6				
Total/Avg.	83	97	180	6.8	70	76	146	7.0

The length of the female parent map was estimated as 1,018 cM and that of the male parent was 1,227 cM. Similar variations in parental maps coverage were reported by Huang *et al.* (2005) in tea AFLP maps. Warnke *et al.*, (2004) and Saha *et al.*, (2005) also observed similar variation between the parental clones in ryegrass and tall fescue, respectively. Differences in map length can result from variation in the number of recombination events in the two parents as well as variations in the numbers and locations of mapped loci.

In summary, this is the first genomic and EST-SSR based framework of tea. The TRI 2023 map covered 1,227 cM of the genome with a marker density of 6.8 cM per marker. The TRI 2043 parental map covered 1,018 cM of the genome with a marker density of 7.0 cM per marker, which is a substantial improvement over the existing RAPD-based map (Tanaka, 1996), RAPD and AFLP map (Hackett *et al.*, 2000) and AFLP map (Huang *et al.*, 2005), especially with regards to marker density. Mapping of SSR loci, especially the EST-SSRs,

will be of great value for comparative mapping with related species. Mapped microsatellite loci will also be of value for their practical application in marker-assisted selection. Tea is a clonally propagated plant. Parents and all progenies of the mapping population are maintained vegetatively and planted in Sri Lanka for field evaluation. This will facilitate further improvement in the mapping effort and enable QTL mapping studies over multiple environments.

ACKNOWLEDGEMENTS

The authors thank K. Chekhovskiy, Patrick Zhao, Ann Harris and Jarrod Steele of Noble Foundation, USA for their technical assistance. This work was supported by a grant from the National Science Foundation Plant Genome Program (DBI-0109732-004) and constitutes a part of a PhD training program. Any opinion, finding and conclusion expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Bindler G, Van der Hoeven R, Gunduz I, Plieske J, Ganal M, Rossi L, Gadani F and Donini P 2007 A microsatellite marker based linkage map of tobacco. Theoritical and Applied Genetics. 114, 341 – 349.

Brondani R P V, Brondani C, Tarchini R and Grattapagila D 1998 Development, characterization and mapping of microsatellite markers in *Eucalyptis grandis* and *E. urophylla*. Theoritical and Applied Genetics. 97, 816 – 827.

Decroocq V, Favé M G, Hagen L, Bordenave L and Decroocq S 2003 Development and transferability of apricot and grape EST microsatellite markers across taxa. Theoritical and Applied Genetics. 106, 912 – 922.

Dida M M, Srinivasachary R S, Bennetzen J L, Gale M D and Devos K M 2007 The genetic map of finger millet, *Eleusine coracana*. Theoritical and Applied Genetics. 114, 321 – 332.

Doligez A, Adam-Blondon A F, Cipriani G, Di Gaspero G, Laucou V, Merdinoglu D, Meredith C P, Riaz S, Roux C and This P 2006 An integrated SSR map of grapevine based on five mapping populations. Theoritical and Applied Genetics. 113, 369 – 382.

Dudley J W 1993 Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Science. 33, 660 - 668.

Echt C S, Kidwell K K, Knapp S J, Osborn T C and McCoy T J 1994 Linkage mapping in diploid alfalfa (*Medicago sativa*). Genome. 37, 61 - 71.

Grattapaglia D and Sederoff R 1994 Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics. 137, 1121 - 1137.

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

Hackett C A, Wachira F N, Paul S, Powel W and Waugh R 2000 Construction of a genetic linkage map of *Camellia sinensis* (tea). Heredity. 85, 346 - 355.

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.

Hemmat M, Weeden N F, Manganaris A G and Lawson D M 1994 Molecular marker linkage map for apple. Journal of Heredity. 85 (1), 4 - 11.

Huang J A, Li J X, Huang Y H, Luo J W, Gong Z H and Liu Z H 2005 Construction of AFLP molecular markers linkage map in tea plant. Journal of Tea Science. 25 (11), 7 - 15.

Jarne P and Lagoda P J L 1996 Microsatellites, from molecules to populations and back. Trends in Ecology and Evolution. 11, 424 - 429.

Jones N, Ougham H and Thomas H 1997 Marker and mapping: we are all geneticists now. New Phytology. 137, 165 - 177.

Kashi Y, King D and Soller M 1997 Simple sequence repeats as a source of quantitative genetic variation. Trends in Genetics. 13, 74 - 78.

Kosambi D D 1944 The estimation of map distances from recombination values. Annals of Eugenics. 12, 172 - 175.

Lande R and Thompson R 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics. 124, 743 - 756.

La Rota M, Kantety R V, Yu J K and Sorrells M E 2005 Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice wheat and barley. Genomics. 6, 23 - 35.

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

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A 1996 The utility of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding. 2, 225 - 238.

Saha M C, Mian M A R, Zwonitzer J C, Chekhovskiy K and Hopkins A A 2005 An SSR- and AFLP-based genetic linkage map of tall fescue (*Festuca arundinacea* Schreb.). Theoritical and Applied Genetics. 110, 323 - 336.

Schuelke M 2000 An economic method for the fluorescent labeling of PCR fragments: A poor man's approach to genotyping for research and high-throughput diagnostics. Nature Biotechnology. 18, 223 - 234.

Sealy J R 1958 A revision of the Genus *Camellia*. Journal of Royal Horticultural Society of London. 62, 934 - 940.

Tanaka J 1996 RAPD linkage map of tea plant and the possibility of application in tea genetics and breeding. Tea Research Journal. 84 (S), 44 - 45.

Taramino G, Tarchini R, Ferrario S, Lee M and Pe M E 1997 Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theoritical and Applied Genetics. 95, 66 - 72.

Varshney R K, Graner A and Sorrells M 2005 Genic microsatellite markers in plants: features and applications. Trends in Biotechnology. 23 (1), 48 - 55.

Voorrips R E 2002 MAPCHART: software for the graphical presentation of linkage maps and QTLs. Journal of Heredity. 93, 77 – 78.

Warnke S E, Barker R E, Jung G, Sim S C, Mian M A R, Saha M C, Brilman L A, Dupal M P and Forster J W 2004 Genetic linkage mapping of an annual x perennial ryegrass population. Theroritical and Applied Genetics. 109, 294 - 304.

Yi G, Lee J M, Lee S, Choi D and Kim B D 2006 Exploitation of pepper EST–SSRs and an SSR-based linkage map. Theoritical and Applied Genetics. 114, 113 – 130.

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