Characterization of a highly repetitive DNA sequence in *Camellia sinensis* (L.) O. Kuntze genome

Mahipal Singh[1, 2, *], Bandana Dhiman[2], Chandan Sharma [2, 3]

[1]Animal Science Division, Fort Valley State University, Fort Valley, GA 31030, USA; [2]Division of Biotechnology, Institute of Himalayan Bioresource Technology, Palampur (HP) 176061, India; [3]Department of Cancer Immunology and AIDS, Dana Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA.

Tea is a widely consumed non-alcoholic natural beverage grown mostly in tropical asia and Africa. It is a vegetatively propagated crop and thus has low genetic variability. Repetitive DNA sequences in the genome, although, are commonly recognised very useful to differentiate individuals, they have not been explored extensively in tea genome. Here we report characterization of novel repetitive motiffs in a genomic DNA clone, designated as pMST11. It was isolated and sequenced from a genomic library of *Camellia sinensis* (L.) O. Kuntze by reverse genomic hybridisation approach. Sequence analysis of the insert displayed a highly A+T rich (66.8%) fragment of 894 base pairs (bp) containing a 168 bp single open reading frame in minus strand. The insert exhibited twelve 15-45 bp long direct repeats, of which, six were novel, non-overlapping and more than 30 bp long. The insert sequence of the clone displayed two unique microsatellite loci, having dinucleotide repeats (AG)15 and (AG)17. Interestingly, both microsatellites contained a 15 bp-consensus sequence attached to their 3' ends. Hybridization of the pMST11 clone with the tea genomic DNA, digested individually with 17 restriction endonucleases, exhibited a continuous smear suggesting that one or more repeats are highly dispersed through-out the tea genome. Dot blot analysis of the 894 bp insert sequence with genomes of 10 out of 12 different plant species exhibited positive signals of various intensities. Human and animal genomes exhibited negligible signals suggesting plant-specific nature of the repeats. The role of these repeat sequences in context of evaluating genetic variation in vegetatively propagated tea is discussed.

*Corresponding author: Animal Science Division, Agricultural Research Station, 1005 State University Drive, Fort Valley State University, Fort Valley, GA 31030, USA. Tel: 478 825 6810. Fax: 478 825 6376. Email: singhm@fvsu.edu

Financial Support: This research was supported by in-house funds of the institute and partially by a grant (BT/R&D/19/05/94) from the Department of Biotechnology, Govt. of India.

Abbreviations: Randomly Amplified Polymorphic DNA (RAPD); Amplified Fragment Length Polymorphism (AFLP); Cleaved Amplified Polymorphic Sequences (CAPS); Tris-ethylenediaminetetraacetic acid (TE); Nitro blue tetrazolium (NBT); 5-Bromo-4-chloro-3-indolyl phosphate (X-phosphate); Saline-Sodium Citrate (SSC)

Introduction

The genome of plants is complex with abundance of reiterated sequences, which are highly variable in repeat frequency and complexity [1]. They diverged rapidly during evolution and are constantly homogenized, giving rise to sequences that are species or genome-specific. Therefore, the repetitive DNA sequences are source of molecular markers that are useful in plant genetic and phylogenetic studies [2, 3]. The arrangement of these repeat sequences could be a random dispersal in genome or in the form of short repeat sequences arranged in tandemly repeated arrays, also known as satellite DNA [4]. Satellite DNA is said to have a role in concerted evolution [5], as enhancers [6] and in gene

silencing [7]. They have been utilized extensively in differentiating individuals in human, animals and plants [8, 9, 10]. Most common repetitive motifs are 2-6 bp tandemly repeated simple sequences, also known as microsatellites [11. 12]. They are scattered through out the genome of eukaryotes [13]. Most frequently occuring microsatellites in mammals are (AC/TG)n dinucleotide repeats, whereas in plants (AT/TA)n repeats are the most common [14]. They have proved to be very useful markers to study genetic variations [15, 16, 17].

Camellia sinensis (L.) O. kuntze (tea) is a widely consumed non-alcoholic natural beverage and is grown mostly in asia (India, Sri lanka, China, Indonesia, Japan) and Africa (Kenya, Uganda, Malawi) and to some extent in Argentina. It is a vegetatively propagated crop and thus has low genetic variability. It is important to find marker system which will give enough differentiating power (polymorphism) in order to develop molecular maps useful in mapping genes, future molecular breeding as well as diversity studies in vegetatively propagated tea. Repetitive DNA sequences in the genome, although, are commonly recognised very useful to differentiate individuals, they have not been explored extensively in tea genome. In an earlier study, genomic library of *C. sinensis* (L.) O. Kuntze was generated and several independent clones containing repetitive sequences were isolated [18]. Here, we report characterization of novel repetitive motiffs in a genomic DNA clone, designated as pMST11, isolated from the genomic library.

Materials and methods

Plant material and genomic DNA extraction

UPASI-9 clone of the commercial tea cultivars maintained at the Banuri Experimental Farm of the IHBT (Palampur, India) was used as a source of genomic DNA. The total genomic DNA was extracted from fresh young leaves using modified protocol of Saghai-Maroof et al 1984 as described earlier [19]. Human and animal DNA was extracted from 1 ml of fresh blood using a commercial DNA extraction kit (Sigma-Aldrich, St. Louis, MO, USA).

Restriction digestion, electrophoresis and Southern transfer

Total genomic DNA (3µg) was digested overnight at 37°C with 17 different restriction endonucleases, i.e. *Hinf* I, *Hind* III, *Hinc* II, *Pst* I, *Eco* RV, *Xba* I, *Xho* I, *Bgl* II, *Nhe* I, *Sac* I, *Hae* III, *Kpn* I, *Sau* 3AI, *Sca* I, *Taq* I, *Dra* I, and *Nde* I (New England Biolabs, Ipswich, MA, USA). The digested DNA was precipitated, washed once with 70% ethanol and dissolved in minimal volume (50 µl) of tris-ethylenediaminetetraacetic acid buffer (TE). The DNA was electrophoretically seperated on a 1% agarose gel and transferred on Hybond N+ nylon membrane using a 20X SSC transfer buffer (3M NaCl, 300mM Sodium citrate, pH 7.0) as recommended by the manufacturer (Amersham, Piscataway, NJ, USA). The DNA was fixed on membrane by baking for 30 min at 120°C under vaccum.

Probe-labelling with digoxigenin, Southern hybridization, and detection

Tea genomic DNA insert (~0.9 kb) of the clone pMST11 was released by double digestion with *Pst* I and *Not* I enzymes. The digested DNA was separated on 1% agarose gel, subsequently 0.9 kb band was excised from the gel and DNA was purified using QIAquick Gel Extraction kit (Qiagen Inc., Valencia, CA, USA). The purified insert was labelled with digoxigenin (DIG) random primer labelling kit (Boehringer Mannheim, Indianapolis, IN, USA) as recommended by manufacturer. The pre-hybridization (5X SSC, 1% blocking reagent, 0.1% N-lauraylsarcosine, 0.2% SDS), hybridization and washing were performed as described in the DIG system user's guide for filter hybridization (Boehringer Mannheim, Indianapolis, IN, USA). Hybridizations were carried out at 68°C in 20X SSC buffer. Filters were washed twice in 2X SSC buffer containing 0.1% SDS at RT for 15 min, twice in 0.5X SSC

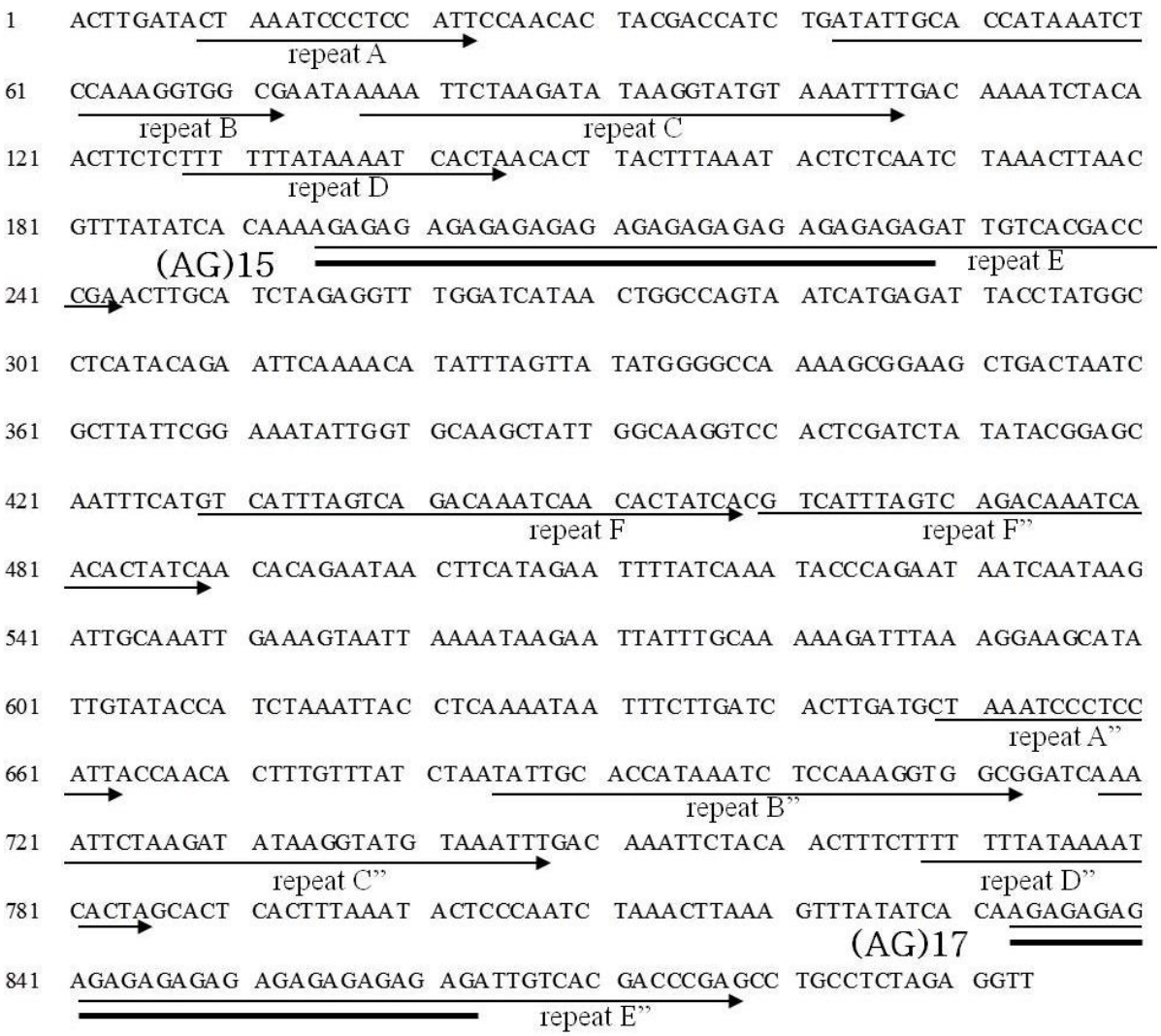

Figure 1. Complete nucleotide sequence of tea genomic DNA insert in plasmid DNA clone pMST11 (genbank Acc.No: AF546881). Underlined portions represent direct repeats. Arrows indicate 5'-3' orientation. Repeat A, A"; B, B"; C, C"; D, D"; E, E"; and F, F" represent DNA sequence repeat pairs. Dinucleotide repeat stretch (AG)15 and (AG)17 is marked by double (thick) underline. Nucleotide 1-168 represents an ORF in this whole sequence.

buffer containing 0.1% SDS at 68°C for 15 min and finally in 0.1X SSC buffer containing 0.1% SDS at RT for 15 min. The detection of bands was carried out with X-Phosphate and NBT color substrates according to manufacturer's instructions (Boehringer Mannheim, Indianapolis, IN, USA). Dot blots were also performed using the same DIG labelling kit following manufacturer's recommendations for dot blots (Boehringer Mannheim, Indianapolis, IN, USA).

DNA sequencing and sequence analysis

DNA Sequencing of the clone pMST11 was performed on an automatic DNA analyser (Applied Biosystems, Inc.) at Delhi University, New Delhi (Central DNA Sequencing Facility of DBT, Govt. of India) using T7 and T3 flanking primers of pBluescript SK+ phasmid as sequencing primers. The REPEAT program of PC Gene package (Intelligenetics, Inc., UK) was used to identify direct repeats. BLAST program was used to perform homology searches [20].

Results

Highly repetitive DNA sequence motifs in the pMST11 clone

We had isolated pMST11, a plasmid DNA clone, containing a 0.9 kb tea genomic DNA fragment from a genomic library of *C. sinensis* (L.) O. Kuntze [18]. This 0.9 kb DNA insert in plasmid pMST11 was sequenced (Genbank Accession Number: AF546881). Sequence analysis revealed that the insert was 894 bp long, and showed to be AT rich with only 33.8% of GC content. The sequence of the pMST11 had only single open reading frame (ORF) on minus strand. When analysed for direct repeats, longer than 15 bases in length, we observed 12 repeats of various sizes ranging from 15-45 bp (Figure 1). Six of these repeat units were unique and non-overlapping, four of which were 30 bases or longer. We also observed five inverted repeats, all of which were 10 bases in length. All inverted repeats, except one, were palindromic. Two more repeat units of 17 and 36 bases with only a slight difference of 3 and 4 bases, respectively, were observed. Two long stretches of AG dinucleotide repeats, (AG)15 and (AG)17, which were separated by 600 bp were also observed (Figure 1, double underlined sequence). The uniqueness of these (AG)n repeat loci was a consensus sequence of 15 bps (ATTGTCACGACCCGA) attached to their 3' end.

Dispersal of repetitive DNA in tea genome

Using the 0.9 kb insert as a probe and the UPASI-9 tea cultivar as a source of genomic DNA, we analyzed patterns of Southern hybridizations. Total genomic DNA was individually digested with 17 different restriction endonucleases, recognizing 4 or 6 base sequences. (it should be noted that thirteen of these restriction enzymes had no recognition site in the sequence of the probe). The digested DNA, when transferred onto nylon membrane and probed with labelled insert of pMST11 clone, exhibited a continuous smearing (in all cases) even with the harsh stringent washing conditions. An example of the observed smearing with eight different restriction enzymes is shown in figure 2.

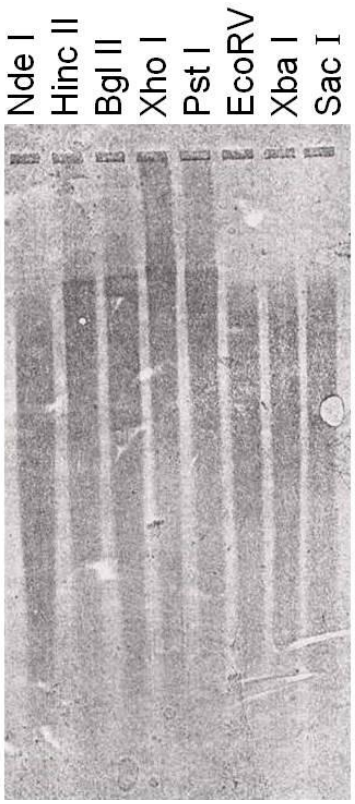

Figure 2. Southern blot of a genomic DNA of a tea cultivar (UPASI-9) independently digested with eight different restriction enzymes and probed with DIG labelled 894 bp insert from plasmid pMST11.

Conservation of repetitive DNA in various plant species

In order to discover if the repeat sequences are specific to tea or if they are ubiquitously distributed across the plant species, we used the 0.9 kb tea DNA insert as a hybridization probe and performed a dot blot analysis. For this experiment genomic DNA from 12 different plant species, which were selected randomly, was used. All of the plant species, except *Ocimum sanctum* and *Olracia sp.* (Figure 3, lane 6 & 7) exhibited positive signals of varying intensities suggesting that one or more repeat sequences in clone pMST11 are highly conserved in plants. Human as well as animal DNA failed to exhibit any signals (Figure 3, lane 13 & 14).

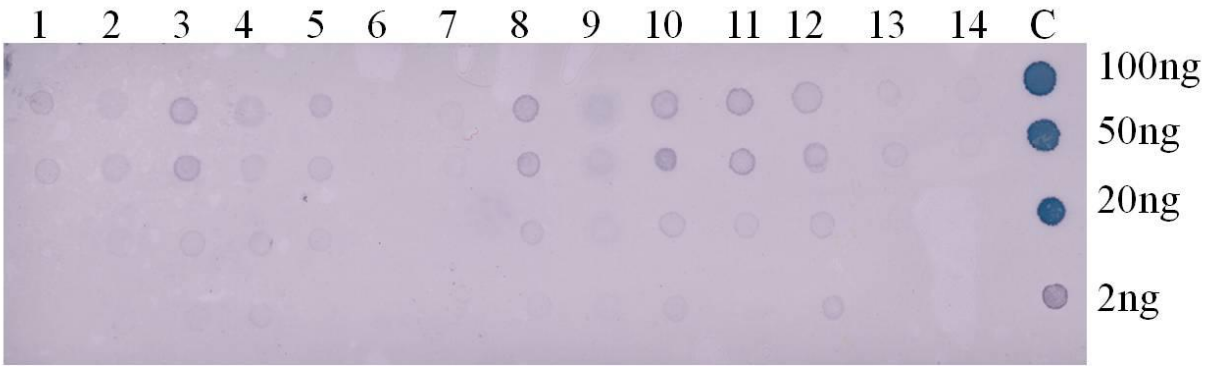

Figure 3. Dot blot of genomic DNA from 12 different species. Lanes 1-12 represent *Nicotiana tobaccum* (tobacco), *Artemisia vulgaris* (mugwort), *Camellia sinensis* (tea; total genomic DNA as a control), *Valeriana jatamansi* (Valeriana), *Rosa demascena* (Damask rose), *Ocimum sanctum* (holy basil), *Olracea sp*. (spinach), *Triticum aestivum* (bread wheat), *Oriza sativa* (rice), *Mentha piperita* (peppermint), *Phaseolus vulgaris* (common bean) and *Dendrocalamus hemiltonii* (bamboo). Lane 13-14 represent human and cow genomic DNA as controls. "C" represents a positive control containing the pMST11 sequence.

Discussion

Tea, a non-alcoholic beverage, is primarily obtained from *C. sinensis (L.) O. Kuntze* which is vegetatively propagated through stem cuttings and thus is genetically highly homogenous. Hence, it is difficult to detect genetic variability in tea. Various approaches including randomly amplified polymorphic DNA (RAPD) [21, 22], gene-specific probes [3], rDNA based probes [19], amplified fragment length polymorphism (AFLP) [24], cleaved amplified polymorphic sequences (CAPS) [25], and microsatellites [11] have been used to differentiate tea clones, or individuals of clones, with varying degree of success. Therefore, it is important to generate additional molecular markers to succeed in fine mapping of the tea genome, molecular breeding program, as well as assessing genetic diversity in commercial tea. In an effort to find out unique repetitive sequence loci in tea genome, we isolated several plasmid DNA clones from genomic library of a commercial tea cultivar [18]. Subsequently, one of these clones, designated as pMST11, was sequenced for the purpose of this study. The sequence analysis of this clone showed that it contained twelve different types of repetitive sequences within a very short span of sequence. Moreover, the length of the repeats varied. The DNA hybridization experiments demonstrated that the repeat sequences observed in the clone pMST11 are dispersed through-out the tea genome (Figure 2). However, which particular repeat sequence unit is abundantly dispersed in tea genome and which one is locus-specific remains to be explored. Since microsatellites are reported to be more abundantly dispersed in plants as compared to other type of repeats [12], it may be assumed that (AG)n repeats observed in this study may be the ones which are abundant in tea genome and the ones contributing significantly to the observed smear in our study as compared to other repeat units. Although, the insert of the clone pMST11 may not be useful as a hybridization probe in studies related to plant genetic variation, the sequence context and, in particular the observed (AG)n loci, should be useful in tea genome mapping program. The microsatellite loci with a high number of repeat units tend to vary in number, even among individuals within a species, which makes them a very useful marker system to study genetic variability of vegetatively propagated plants such as tea. A good linear relationship between number of alleles detected at a locus and the number of short sequence repeats (SSR) in microsatellite DNA repeats has been shown [26]. It is important because SSRs have highest allele number per locus compared to other marker systems providing more differentiation power

(polymorphism) to see differences between cultivars, clones and even between individuals derived from a single clone. Some of the repeat units, in particular the microsatellite loci, should contribute to the array of molecular markers used to differentiate commercial tea clones and individuals originating from a particular clone, especially the ones which produce high biomass and/or superior quality of tea. These biomarkers can ultimately be useful in preparing genetic maps and in tea breeding programs. Future studies should dissect individual repeat units which should shed more light on their value as molecular markers.

Acknowledgments

We would like to thank late Prof. A. K. Gupta and Dr. P. S. Ahuja, the former and current directors of IHBT, Palampur for providing facilities and support of this work. This work was supported by in-house funds of the institute and partly by a grant (BT/R&D/19/05/94) from the Department of Biotechnology, Govt of India. CS and BD are thankful to the CSIR, Govt. of India for Senior Research Fellowship awards.

Reference

1. Schmidt T, Heslop-Harrison JS. 1998. Genomes, genes and junk: the large scale organization of plant chromosomes. Trends Plant Sci. 3: 195-199.
2. Joshi SP, Ranjekar PK, Gupta VS. 1999. Molecular markers in plant genome analysis. Current Sci. (India). 77: 230-240.
3. Farrall M. 2004. Quantitative genetic variation: a post-modern view. Human Mol. Genet. 13: 1-7.
4. Metais I, Aubry C, Hamon B, Peltier D, Jalouzot R. 1998. Cloning, quantification and characterisation of a minisatellite DNA sequence from common bean *Phaseolus vulgaris* L. Theor. Appl. Genet. 97: 232-237.
5. Cronn RC, Zhao X, Peterson AH, Wendel JF. 1996. Polymorphism and concerted evolution in a tandemly repeated gene family: 5S rDNA in diploid and allopolyploid cottons. J Mol. Evol. 42: 685-705.
6. Ott RW, Hansen LK. 1996. Repeated sequences from the *Arabidopsis thaliana* genome function as enhancers in transgenic tobacco. Mol. Gen. Genet. 252: 563-571.
7. Meyer P. 1996. Repeat-induced gene silencing: common mechanisms in plants and fungi. Biol. Chem. Hoppe –Seyler. 377: 87-95.
8. Jeffreys AJ, Wilson V, Thein SL. 1985. Hypervariable, minisatellite, regions in human DNA. Nature. 314: 67-73.
9. Webster MS, Reichart L. 2005. Use of microsatellites for parentage and kinship analysis in animals. Methods in Enzymol. 395: 222-238.
10. Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB. 1995. The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. 90: 43-48.
11. Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ. 2011. Current trends in microsatellite genotyping. Mol Ecol Resour. 11(4): 591-611.
12. 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 Biology. 9: 53.
13. Gupta PK, Balyan IS, Sharma PC, Ramesh B. 1996. Microsatellites in plants: a new class of molecular markers. Current Sci. 70: 45-54.
14. Powell W, Machray GC, Provan J. 1996. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1: 215-222.
15. Becker J, Heun M. 1995. Barley microsatellites: allele variation and mapping. Plant Mol. Biol. 27: 835-845.
16. Alix K, Baurens F, Paulet F, Glaszmann J, D'Hont A. 1998. Isolation and characterization of a satellite DNA family in the Saccharum complex. Genome. 41: 854-864.
17. Varshney RK, Graner A, Sorrells ME. 2005. Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 23: 48-55.
18. Singh M, Sharma C, Dhiman B, Singh D, Raizada J. 2004. Cloning and characterization of repetitive DNA sequence elements from *Camellia sinensis* (L.) O. Kuntze. Physiol. Mol. Biol. Plants. 10: 209-215.
19. Singh D, Singh M. 2001. Organization of 5S ribosomal RNA genes in tea (*Camellia sinensis*). Genome. 44: 143-146.
20. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.
21. Wachira FN, Waugh R, Powell W, Hackett CA. 1995. Detection of genetic diversity in tea (*Camellia sinensis*) using RAPD markers. Genome. 38: 201-210.
22. Singh M, Saroop J, Dhiman B. 2004. Detection of intra-clonal genetic variability in vegetatively propagated tea using RAPD markers. Biol. Plantarum. 48: 113-115.
23. Matsumoto S, Takeuchi A, Hayatsu M, Kondo S. 1994. Molecular cloning of phenylalanine ammonia-lyase cDNA and classification of varieties and cultivars of tea plants (*Camellia sinensis*) using the tea PAL cDNA probe. Theor. Appl. Genet. 89: 671-675.
24. Paul S, Wachira FN, Powell W, Waugh R. 1997. Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O. Kuntze) revealed by AFLP markers. Theor. Appl. Genet. 94: 255-263.
25. Kaundun S, Matsumoto S. 2003. Development of CAPS markers based on three key genes of the phenylpropanoid pathway in tea, *Camellia sinensis* (L.) O. Kuntze, and differentiation between assamica and sinensis varieties. Theor Appl Genet. 106: 375-383.
26. Yang GP, Saghai-Maroof MA, Xu CG, Zhang Q, Biyashev RM. 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol. Gen. Gent. 245: 187-194.