Scientia Horticulturae 134 (2012) 167–175

Contents lists available at SciVerse ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Microsatellite markers from tea plant expressed sequence tags (ESTs) and their applicability for cross-species/genera amplification and genetic mapping

Jian-Qiang Ma, Chun-Lei Ma, Ming-Zhe Yao, Ji-Qiang Jin, Zhi-Lan Wang, Xin-Chao Wang, Liang Chen*

Tea Research Institute of the Chinese Academy of Agricultural Sciences, National Center for Tea Improvement, 9 South Meiling Road, Hangzhou, Zhejiang 310008, China

ARTICLE INFO

Article history:
Received 1 August 2011
Received in revised form 26 October 2011
Accepted 28 October 2011

Keywords:
Tea plant
EST-SSR
Transferability
Genetic mapping

ABSTRACT

Bioinformatic mining of EST datasets is an efficient and economical method for developing functional microsatellite markers. The aim of this work is to generate new tea plant EST-SSR markers and to evaluate their potentials for cross-species/genera amplification and in particular genetic linkage mapping. A total collection of 8008 untapped tea plant ESTs were retrieved from GenBank and then clustered into 2982 non-redundant unique sequences, in which 561 microsatellites were identified. Of these SSRs, the di-nucleotide repeats were the most common repeat types (255, 45.5%), then followed by tri- (18.9%), hexa- (14.3%), penta- (13.5%), tetra- (5.3%), hepta- (2.1%), octa- (0.2%), and deca-nucleotide repeats (0.2%). AG/CT and AGA/TCT predominated in dimeric and trimeric repeat motifs, respectively. BLASTX homology searches found that 319 SSR-containing ESTs had significant *Arabidopsis* protein hits (*E*-value < 1.00*E*−05). In total, a set of 97 primer pairs was successfully designed for marker development and validation. Finally, fifty-four primers yielded repeatable and distinct amplification products among 21 tea plant genotypes, however, only 30 loci were polymorphic with an average allele number of 3.4/locus (range: 2–6 alleles/locus). Cross-amplification showed that 87% (26/30) primers could yield scorable products in at least one of nineteen Theaceae species. The transferability ranged from 7 to 77% with a mean of 30%. Furthermore, twenty-five SSR loci were polymorphic for the parents of a pseudo-testcross mapping population of tea plant, indicating they could be utilized for genetic linkage mapping. These novel informative EST-SSR markers will be a useful addition to genetic analysis and molecular breeding programs in tea plant and its related species.

1. Introduction

Microsatellites or simple sequence repeats (SSRs) are a ubiquitous class of repetitive DNA sequences consisting of tandemly repeated short (1–6 nucleotides) motifs (Morgante et al., 2002). They can exhibit a high level of length variability and allelic diversity which are thought to be generated by mutational changes associated with insertion or deletion of motif repeats during replication slippage, mismatch/double strand break repair, recombination and retrotransposition events (Ellegren, 2004; Seyfert et al., 2008). Features such as locus-specificity, co-dominant inheritance, high reproducibility and PCR-based detection make microsatellites ideal for assaying genetic variation (Varshney et al., 2005). They have been extensively employed in genotyping, genetic diversity evaluation, linkage and association mapping, and molecular marker-assisted selection (MAS) studies (Kalia et al., 2011). However, before microsatellites can be validated by PCR amplification, sequences information on SSR flanking regions is required to design specific primers. Thereby the first step for microsatellites mining is to obtain SSR-containing sequences. But conventional methods, such as generation of SSR-enriched genomic library, can be time-consuming and costly due to complicated process and technical requirements (Gupta and Varshney, 2000). Hence, only a few microsatellite markers had been developed for many organisms owing to the lack of available sequences information. However, with the increase of cDNA sequences information in public database, microsatellites exploitation from expressed sequence tags (ESTs) has already become an attractive alternative strategy (Gupta and Varshney, 2000). Additionally, EST-based microsatellites or EST-SSRs are derived from transcripts, wherein flanking regions are expected to be more conserved and thus tend to show higher level of transferability in cross-amplification and possibility of tagging and mapping of genes and quantitative trait loci (Varshney et al., 2005).

The family Theaceae includes 7–40 genera depending on the classification system employed, comprising diverse economically important species, such as beverage, oil, ornamental, medicine (Prince and Parks, 2001). Among these, tea plant is one of the most popular beverage crops grown in tropical and subtropical regions, mostly in developing countries. Cultivated tea plant is usually a

* Corresponding author. Tel.: +86 571 86652835; fax: +86 571 86652835.
E-mail address: liangchen@mail.tricaas.com (L. Chen).

0304-4238/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.scienta.2011.10.029

diploid ($2n = 2x = 30$) and belongs to *Camellia* L. Sect. *Thea* (L.) Dyer, consisting of five species and two varieties (Chen et al., 2000). Of these, *C. sinensis* (L.) O. Kuntzevar. *sinensis* and *C. sinensis* var. *assamica* (Masters) Kitamura are world widely planted commercially. Tea is made from leaves, leaf buds and internodes of tea plant. Its popularity has been increasing and spreading due to its potential medical consequences such as tumorigenesis inhibition (Yang et al., 2009a). In 2010, the world produced 4.16 billion kg dry tea and consumed 3.98 billion kg tea (ITC, 2011); tea industry contributed a lot to the local economy. The yield and quality of tea can be greatly affected by tea cultivars, and heroic efforts have been therefore made to develop the best cultivar. But progress with conventional breeding strategies has been very slow and mightily constrained by various factors, such as long juvenile period, self-incompatibility, and large plant size. As with tea plant, many of other valuable Theaceae species such as *C. nitidissima* Chi, *Stewartia sinensis* Rehd. Te Wils, are also confronted with the scarcity of genome information.

Over the past decades, molecular markers have reinvigorated the art of plant breeding and improved the efficiency of breeding programs through their use in MAS (Xu and Crouch, 2008). So far, EST-SSRs have been extensively generated and utilized in some woody plants such as cassava (Sraphet et al., 2011), cocoa (Lima et al., 2010), and coffee (Aggarwal et al., 2007; Hendre et al., 2008). However, the total available tea plant EST-SSRs, approximate 270 (Zhao et al., 2008; Sharma et al., 2009, 2011; Ma et al., 2010a) are rather insufficient. Their intergeneric transferability and mappability are not clear yet. Here, we report the generation and characterization of thirty novel EST-SSR markers and their applicability for cross-species/genera amplification and in particular genome mapping of tea plant.

2. Material and methods

2.1. Plant materials

For the present study a panel of 21 representative tea genotypes from *Camellia* L. Sect. *Thea* (L.) Dyer, including eight of *C. sinensis* (L.) O. Kuntze var. *sinensis*, eight of *C. sinensis* var. *assamica* (Masters) Kitamura, and each one of *C. sinensis* var. *pubilimba* Chang, *C. gymnogyna* Chang, *C. tachangensis* F. C. Zhang, *C. taliensis* (W. W. Smith) Melchior, *C. crassicolumna* Chang were selected and used along with nineteen related species belonging to nine genera of family Theaceae for SSR validation and cross-amplification. The parents of a pseudo-testcross mapping population, *C. sinensis* var. *sinensis* cv. *Yingshuang* and *C. sinensis* var. *pubilimba* cv. *Beiyue Danzhu*, were used to test the potentials of novel EST-SSR markers for linkage mapping (Table 1). Healthy 'two and a bud' for each of them were sampled from China National Germplasm Hangzhou Tea Repository at the Tea Research Institute of the Chinese Academy of Agricultural Sciences. Total genomic DNA was isolated using a modified CTAB method (Dellaporta et al., 1983).

2.2. SSR identification and functional annotation

All of the untapped tea plant ESTs available in the GenBank on 20 April 2011 were retrieved, and assembled using DNAStar (http://www.dnastar.com) with parameters (match size: 12, minimum match percentage: 80, minimum sequence length: 100). The non-redundant consistent ESTs were subsequently screened for microsatellites detection by SSRIT (http://www.gramene.org/db/markers/ssrtool; Temnykh et al., 2001) under criteria of a minimum of 8, 5, 4, 3 motif repeats for di-, tri-, tetra-, penta- and higher-order nucleotide microsatellites, respectively. The SSR-containing ESTs were compared against *Arabidopsis* protein sequences database by BLASTX program (http://blast.ncbi.nlm.nih.gov/Blast.cgi), and then sequences defined as *Arabidopsis* proteins with significant hit (*E*-value < 1.00*E*−05) were annotated and classified into functional categories using Gene Ontology (GO) Annotations from TAIR (http://www.arabidopsis.org; Berardini et al., 2004). A schematic drawing of EST-SSR mining and validating was provided in Fig. 1.

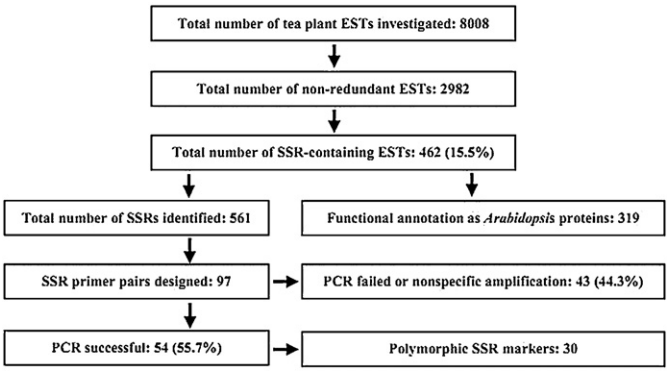

Fig. 1. Schematic drawing depicting the key outcome of each step of EST-derived SSR marker generation.

2.3. Primer designing, PCR amplification, amplicon visualization

PCR primer pairs were designed from SSR flanking regions using Primer Premier (http://www.premierbiosoft.com) with the following criteria: primer length from 18 to 24 bp, annealing temperature (Ta) from 40 to 60 °C, amplicon size from 100 to 300 bp, and GC content between 40 and 60%. PCR reactions were performed as described earlier (Zhao et al., 2008). Amplified products were electrophoretically separated on 10% polyacrylamide gels in 1 × TBE (Tris–Borate–EDTA) buffer, visualized by silver staining according to Charters and Wilkinson (2000), and amplicon sizes were estimated by molecular weight comparison with 50 bp DNA ladder mobility (Beijing Dingguo Biotech).

2.4. SSR scoring and data analysis

Each distinct allele with expected size, amplified with homologous SSRs in different tea plant genotypes and related species, was scored and coded into a binaryformat: 1 for presence and 0 for absence or smear, respectively. The allelic data for 21 tea genotypes were used to estimate different statistical and genetic parameters. Observed heterozygosity (H_O) and expected heterozygosity (H_E) were calculated using PopGene program (Yeh and Boyle, 1997); polymorphic information content (PIC) values were calculated by PowerMarker (Liu and Muse, 2005).

3. Results

3.1. Frequency and distribution of non-redundant EST-SSRs in tea plant

After sequence redundancy analysis, a total of 2982 consistent sequences (2085 singletons and 897 contigs) were predicted from 8008 tea plant ESTs, representing approximately 4238 kb of putative functional tea plant transcriptome. The searches detected 561 microsatellite loci among 462 (15.5%) non-redundant ESTs. Therefore, the average separation between two SSRs was ~7.55 kb, equivalent to one SSR in every 14.3 EST sequences. BLASTX against *Arabidopsis* protein sequences database found that 319 (69.0%) of the tea plant SSR-ESTs had significant hits (*E*-value < 1.00*E*−05). Most of them were annotated as hypothetical proteins of binding function, transferase activity, and transporter activity.