Development and use of cleaved amplified polymorphic sequence (CAPS) markers in tea.

Shiv Shankhar KAUNDUN and Satoru MATSUMOTO

National Institute of Vegetable and Tea Science
2769 Kanaya, Kanaya-cho, 428-8501 Shizuoka, Japan

Abstract

In order to breed for new tea types and tastes, to fingerprint commercial teas as a means of controlling tea identity and to collect and conserve abandoned tea, a perfect knowledge of the genetic background of the existing teas is a prerequisite. To these ends, we have developed informative CAPS markers based on three key genes, PAL, CHS and DFR, involved in the phenyl propanoid pathway and playing an important role in the organoleptic properties of tea. Fifty tea samples, including both Assam and China tea genotypes were initially analyzed to detect for polymorphic primers/restriction enzymes, and the markers were subsequently used to screen Japanese tea cultivars as well as abandoned tea in Japan, China and Korea. A multivariate analysis: markers X genotypes separated *assamica* and *sinensis* teas into two distinct groups, thus reflecting the large differences in their chemical composition. On the other hand, all the forty-six Japanese tea cultivars studied could easily be distinguished from each other, with Yabukita, the predominant tea cultivar in Japan, displaying a unique restriction profile. China and Korea teas were found to be closely related to each other and relatively distinct from Japanese tea, even though a higher proportion of diversity resided within rather than between groups. In the light of these results, collection and conservation strategies of tea for future breeding programs are discussed.

Key words: Tea, CAPS, DNA markers, Genetic Diversity

Introduction

Tea has been introduced in many countries where it contributes significantly to the local economy. Due to intensive selection and use of few successful clones in most countries, tea diversity has been declining over time. Several studies have shown that diversity within different national germplasm collections is low and represents a tiny fraction of the existing variability initially introduced. In order to limit further reduction in tea gene pool as well as breeding for new tea types, a detailed knowledge of the existing diversity is a prerequisite.

Several types of DNA markers are nowadays available to quantify genetic diversity effectively. Restriction Fragment Length Polymorphism (RFLP) was initially used and has provided a wealth of information in many crop species (Matsumoto et al 1994). This technique is however time consuming, costly as well as being technically demanding and requiring large amounts of relatively pure DNA. Amplified Fragment Length Polymorphism (AFLP), though showing the highest marker index ratio, also requires appreciable amounts of pure DNA and is labour intensive (Shim and Jorgensen 2000). Random Amplified Polymorphic DNA (RAPD) offers several advantages over the other markers by generating a considerable number of markers, from low amounts of template DNA and without prior knowledge of genome information. This technique is however poorly reproducible and cannot be transferred from laboratory to laboratory (Prenner et al. 1993). Simple sequence repeat markers are highly polymorphic, but are yet to be developed in tea, and in any case require sophisticated separating systems for precise allele scoring (Diwan and Cregan 1997).

Cleaved Amplified Polymorphic Sequence (CAPS or PCR-RFLP) methodology generates another type of molecular markers combining both PCR and RFLP techniques. It requires minute amounts of genomic DNA and simple electrophoretic systems to reveal polymorphism. We have employed this technique to detect polymorphism in three genes involved in the phenyl-propanoid pathway, namely phenylalanine ammonia-lyase (PAL), chalcone synthase (CHS) and

dihydroflavonol 4-reductase (DFR), all three playing and important role in tea taste and quality.

Material and methods

Materials
Fifty tea samples including 24 and 24 plants from *assamica* and *sinensis* (comprising the leading Japanese tea cultivar Yabukita) varieties as well as two hybrid teas between a *sinensis* X (*sinensis* X *assamica*) genotype were used to develop markers. Forty-eight samples resulting from a cross between two local cultivars were analyzed to test the inheritance of the markers. Forty-six other common Japanese tea cultivars were subsequently screened and compared with 180 abandoned tea samples from China, Japan and Korea.

DNA isolation
Total DNA was extracted from fresh or frozen tissue using a slightly modified version of the CTAB method as described in Kaundun et al (2000).

Primer design
PCR primers were designed with the Genetix software (Tokyo) and based on cDNA nucleotide sequences of PAL, CHS2 and DFR in tea (cultivar: Yabukita). Three primer pairs were used to tag both the coding and non-coding regions of PAL, while one and three other primer pairs were sufficient to amplify DNA from CHS2 and DFR.

PCR and restriction product analysis
The polymerase chain reaction was carried out in a 25 μl volume containing 50-100 ng of genomic DNA on a Perkin-Elmer thermocycler model 2400. The PCR products were separated on 2% agarose gel and visualized under UV to reveal Amplicon Length Polymorphism (ALP) and Cleaved Amplified Polymorphic Sequence (CAPS) in which case 2 μl of PCR products were digested with 32 different restriction endonucleases.

Cloning and Sequencing of amplification products
To confirm the identity of the PCR bands generated by each of the seven primer pairs, the corresponding amplification products from Yabukita were ligated into pGEM-T easy vector (Promega) and the nucleotide sequences of the inserts determined on an Perkin-Elmer ABI 373 automated sequencer using the dye primer or dye terminator kits (Perkin Elmer).

Data analysis
The presence of an allele was coded as 1 and its absence as 0 for both ALP and PCR/RFLP markers. Popgene v3.1; http://www.ualberta.ca/~fyeh/index.htm) was used to estimate heterozygosity values and apportion diversity into within and among groups, while factorial correspondence analysis were performed by Statitcf v5 (ITCF, Paris) to study the samples at an individual level.

Results and Discussion

Seven primer pairs were sufficient to tag PAL, CHS2 and DFR genes in a manner comprising, PAL exon 1, PAL intron and PAL exon 2, CHS exon2, DFR intron 1+2, DFR intron 3 and DFR intron 4+5. The primers amplifying the exons generated bands corresponding to the expected sizes as deduced from cDNA sequences, while those flanking the introns were greater in length. Sequencing reactions carried out on the seven amplification products for cultivar Yabukita confirmed that they corresponded to the targeted genes. Five of the seven primer pairs produced a single band while those flanking PAL intron and DFR intron 1+2 gave one or two bands with a total of three and five amplicon length polymorphisms respectively due to the presence of mutation deletion.

In the absence of ALP, the PCR products were digested with a set of 32 restriction

enzymes. Four base pair cutters were the most successful at revealing polymorphism among the samples due to their higher number of recognition sites. The higher proportions of AT in the introns were responsible for multiple restriction sites for enzymes with AT in their recognition sites, while most of those with GC sequences were more suitable for exons. Characteristic DNA patterns obtained with CHS exon2 digested by *HaeIII* are shown on figure 1. Three types of profiles could be observed: homozygote for a restriction site, homozygote for the absence of the site and a heterozygote state due to the co-dominant nature of the polymorphism. In all, thirteen loci were polymorpic including both ALPs and PCR/RFLP polymorphisms.

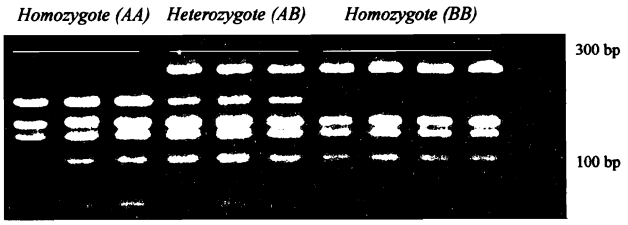

Figure 1. DNA profiles obtained with CHS2 exon2 digested by *HaeIII*: homozygote for a polymorphic restriction site (AA), homozygote for the absence of the site (BB) and a heterozygote state (AB) with only one of the two alleles being cut.

In order to test the inheritance of the markers, we examined the progenies of a cross between two local tea cultivars in Japan. The progenies segregated in a mendelian fashion at all informative loci and were consistent with the parental genotypes. Therefore the markers represented different alleles of a unique locus rather than different loci. They could subsequently be used in total confidence at estimating the different diversity parameters within and between *sinensis* and *assamica* teas.

The basic statistics given by Popgene revealed that *sinensis* tea was more diverse than *assamica* tea. Overall, however, 83 % of the total diversity resided within as compared to between groups (17 %). This is a characteristic of highly outcrossing species, maintaining high within groups as compared between groups diversity (Hamrick 1990). Also, *assam* and *sinensis* types can cross-hybridize very easily and it is thought that pure tea archetypes no longer exist (Visser 1969). Classification of many hybrids and introgressants are determined based on their morphological proximity with either varieties, and it would not be surprising that some of the samples considered as pure *assamica* genotypes in this study contain a percentage of the *sinensis* genome and vice versa.

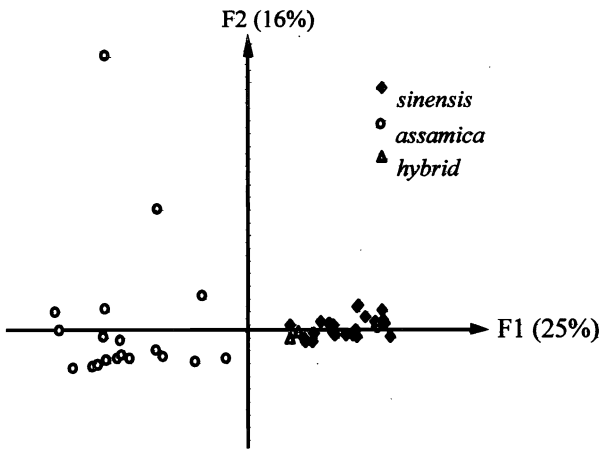

Figure 2. Principal plane (F1 X F2) of a factorial correspondence analysis based on 50 tea genotypes and 31 polymorphic alleles.

When the teas were analyzed at an individual level through a factorial correspondence analysis (genotypes X alleles) the *assam* and *sinensis* teas were completely separated from each other (figure 2). This result is in agreement with former studies based on anonymous RAPD and AFLP markers (Wachira et al 1995; Paul et al 1997). It also highlights the overall physiological and chemical differences between these two tea varieties. The two hybrid samples grouped closer with the *sinensis* subgroup, due to the fact that three quarter of their genomes were of *sinensis* type as they resulted from a cross involving a *sinensis* with a *sinensis* X *assamica* genotype. These hybrids contain larger amounts of catechins and tannins and are used to produce semi-fermented and

fermented teas in Japan.

When the same markers were used in fingerprinting studies, all the 46 most common Japanese teas cultivars were easily distinguishable, with Yabukita, the leading tea cultivar showing a unique profile. These markers also show a greater diversity within as compared to between countries. The diversity was the highest in China, the center of Origin of tea. Korea and Japan teas appeared close to each other and relatively distant from Japanese tea. This suggests that present Japanese tea originated from Chinese tea populations that have disappeared by now, or simply that these Chinese tea populations have not been included in this study. Japanese cultivars highly overlapped with local and abandoned tea (Yamacha) implying that enhancement of Japanese tea germplasm collection would rationally be achieved through exchange of genetic material between China, Japan and Korea.

The development of STS-based markers may be relatively time consuming in the first instance because sequence information is required for primer design, sequencing reactions needed to confirm that the right DNA fragment have been amplified and a number of restriction enzymes have to be tested to reveal polymorphism. Nevertheless, once markers are established they can find wide application in mapping, population genetics and fingerprinting studies.

Acknowledgements

The authors are very grateful to A. Nesumi for providing plant materials for inheritance analysis and to Dr. N. Mizuno for fruitful discussions about statistical calculations. The first author was supported by a fellowship and grant from the Science and Technology Agency (STA, Japan) for the period 1999-2001.

References

Diwan N, Cregan PB (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. **95**: 723-733

Hamrick JL (1990) Isozymes and the analysis of genetic structure in plant populations. In: Soltis ED,Soltis PS (eds) Isozymes in plant biology. Chapman and Hall, London., pp 87-90

Kaundun SS, Zhyvoloup A, Park YG (2000) Evaluation of the genetic diversity among elite tea (*Camellia sinensis var. sinensis*) genotypes using RAPD markers. Euphytica **115**: 7-16

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 tea Pal cDNA probe. Theor. Appl. Genet. **89**: 671-675

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

Prenner GA, Bush A, Wise R, Kim W, Dommier L, Kasha K, Laroche A, Scoles G, Molnar SJ, Fedak G (1993) Reproducibility of Random Amplified Polymorphic DNA (RAPD) analyis among laboratories. PCR methods Applic. **2**: 341-345

Shim SI, Jorgensen RB (2000) Genetic structure in cultivated and wild carrots (Daucus carota L.) reveled by AFLP analysis. Theor. Appl. Genet. **101**: 227-233

Visser (1969). Tea, *Camellia sinensis* (L.) O. Kuntze. Outline of perrenial crop breeding in the tropics. F. P. Ferwarda and F. Wit. Wageningen, The Netherland, Veenman, H. Zonen, N.V.: 459-493.

Wachira FN, Waugh R, Hackett CA, Powell W (1995) Detection of genetic diversity in tea (*Camellia sinensis*) using RAPD markers. Genome **38**: 201-210