Development and Characterization of Coffee Specific Microsatellite Markers for use as Potential Genetic Markers

RAMESH K. AGGARWAL, A. BARUAH, V. NAIK, PRASAD S. HENDRE, A. ASHRAF, P. RAJENDRAKUMAR, R. RAJKUMAR, V. ANNAPURNA, R. PHANINDRANATH, N. S. PRAKASH*, C. S. SRINIVASAN*, LALJI SINGH

Centre for Cellular and Molecular Biology, Uppal Road, Tarnaka, Hyderabad-500007, India.
1Central Coffee Research Institute, Chikamagalur-577117, India
(*Correspondence: rameshka@ccmb.res.in)

Abstract

Availability of informative genetic markers is prerequisite for effective genetic-linkage studies. Recent advances in molecular biology have led to the development of a plethora of DNA variation based efficient marker systems that promise- impetus, dependability and directionality to the genetic improvement efforts. Among various types of DNA markers the short-sequence repeat (SSR) based microsatellite markers have proven to be the most desirable for genetic studies due to their codominant nature, stability, abundance, sensitivity, ease and speed of analysis, minimal sample requirements and suitability for automation. Despite their potential and desirability, very few such markers have been developed and described in the literature for coffee tilldate. Under the first Indian initiative on Coffee Genomics, we have been able to develop 150+ new coffee microsats that provide potential genetic markers for germplasm characterization and molecular linkage studies in coffee.

For development of microsatellite markers, small-insert, partial, genomic libraries (comprising ~ 75,000 plasmid clones) were constructed from total DNA of *Coffea arabica* var.HdeT (Hibrido-de-Timor) and *C. canephora*. These libraries spotted on high-density nylon filters were screened for SSR positive clones through Southern hybridization with different synthetic oligonucleotide repeat probes. Based on hybridization signals, a large number of putative repeat-positive clones were selected and sequenced. The sequence data revealed relatively low abundance of SSR motifs in coffee genome, with AG di-repeat being most frequent. The clone sequences having 18+ bp repeat-motifs were used for designing PCR primers, which were subsequently validated using panels of elite coffee genotypes for their suitability as genetic markers. In total >170 informative markers could be developed that also showed broad cross-species transferability. It is hoped that this new set of markers would serve as an important resource for coffee genomics.

Introduction

Coffee is an important plantation crop belonging to the genus *Coffea* of the family Rubiaceae. Although, >100 species of coffee are known, its commercial cultivation relies only on two species, amphidiploid *Coffea arabica* L. (2n = 4x = 44) and diploid *C. canephora* Pierre (2n = 22). Considerations like sustainability over ever-changing demands of agro-climatic

conditions and the commercial markets calls for continuous efforts to develop better coffee genotypes. The conventional breeding efforts to this end, especially in arabica coffee have been seriously constrained mainly due to the narrow genetic base of the arabica genepool, lack of efficient breeding tools and linkage maps, long pre-bearing periods and difference in ploidy level. Recent advances in DNA marker technology offer possibilities to overcome some of these limitations by providing highly informative genetic markers and marker assisted breeding.

Out of various kinds of DNA markers, PCR-based microsatellites are most desirable for germplasm characterization and crop improvement and are widely used for genetic analysis, development of linkage maps and population improvement by employing marker aided selection techniques (Gupta & Varshney 2000). Despite these advantages, only ~150 microsatellite markers have been reported till to date for coffee (Baruah *et al*. 2003; Bhat *et al*. 2004), signifying the need for expanding the repertoire of these genetically highly informative markers for efficient management and improvement of coffee germplasm resources. Under the first Indian initiative on Coffee genomics, at our center we have initiated efforts for development of microsatellite markers to complement genetic studies on coffee and this paper deals with the leads achieved in this regard.

Material and Methodology

We have used the conventional approach involving construction and screening of genomic library for development of SSR-markers (Sambrook *et al*. 1989). For the purpose, small-insert partial genomic libraries were constructed from genomic DNA of two coffee cultivars, *C. arabica* var.HdeT and *C. canephora* var.sln-274. Genomic DNA used for library construction was isolated from leaves as detailed by Aggarwal (2002). Ten micrograms of total genomic DNA was restriction digested with *Sau*3aI (NEB) in case of *C. arabica* and double digested with *Rsa* I and *Hae* III (NEB) restriction enzymes in case of *C. canephora*. Digested DNA samples were size-fractionated on agarose gel and fragments in the range of 500 to 1500 bp were eluted using GFX column (Amersham Pharmacia). The eluted DNA fragments were cloned in plasmid vectors pBKS(+) and/or pMOS (Amersham Pharmacia) and transformed into electro-competent *E. coli* (DH10B) cells by electroporation. The recombinant clones were selected by blue-white selection using X-gal (@ 40µg/mL) and

IPTG (@ 20µg/mL). The individual white recombinant colonies were then curetted into 96 or 384 well microtitre plates in 10% glycerol LB-ampicillin medium, grown overnight and stored at –70 °C till further use. The curetted clones were spotted on nylon membrane using 96-/384- pin replicators (Nunc Technologies) to prepare high-density library filters for screening and identification of SSR-positive clones by Southern hybridization with labeled-oligo probes.

The genomic libraries were screened using different combinations of di-, tri- and tetra-repeat SSR oligos of 30-mer lengths that were synthesized in-house (on DNA/RNA Oligosynthesizer394, Applied Biosystems) and 5`-end labeled using γ–P^{32-}dATP radionucleotide using Polynucleotide Kinase (NEB). The oligos used for library screening were selected based on their relative abundance in the coffee genome (our unpublished data). All library filters were screened using 9 oligoprobes, namely, $(CA)_{15}$, $(GA)_{15}$, $(CAA)_{10}$, $(GGT)_{10}$, $(ATT)_{10}$, $(AGG)_{10}$, $(AGA)_{10}$, $(ACT)_{10}$ and $(CATA)_{8}$, but *C. canephora* library was also screened using 5 additional oligoprobes i.e., $(GAC)_{10}$, $(CAT)_{10}$, $(CTG)_{10}$, $(CGG)_{10}$, and $(GATA)_{8}$. Hybridizations were performed in 6 x SSC at 55 °C for 12-14 hrs and filters were washed at relatively high increasing stringency till the counts have reduced significantly. The hybridized filters were exposed to phosphor imager for different lengths of time (usually 4-8 h) and scanned to identify probe-hybridized, SSR-positive clones. Plasmids were isolated from the positive clones using alkaline lysis method, and the cloned inserts were amplified and sequenced using M13 universal primers. Each clone was sequenced for both strands using BigDye terminator kit (ABI) on automated DNA sequencer ABI3700 (Applied Biosystems).

The sequences were searched using microsatellite search tool MISA (Theil *et al* 2003) for ascertaining frequency and distribution of SSRs having a minimum number of 12 repeat for mono, 6 for di, 5 for tri, tetra, penta and hexamer motifs. The clones found positive for SSR repeat length of at least 18-mer were used for designing primers and marker development. The primers were designed using Primer 3 (http://www-genome.wi.mit.edu/genome_software/other/primer3.html) or GeneTool Lite version 1.0. All primer pairs were used to standardize the PCR-amplification conditions and then the working pairs were tested for their utility as potential genetic markers using panels of elite genotypes/accessions of *C. arabica*, *C. canephora* (robusta) and 17 species of *Coffea* and

related genera *Psilanthus*. In general, PCR amplifications were carried out using ~10 ng genomic DNA at 2 mM $MgCl_2$ and annealing temperature ranging between 55 - 60 °C. Microsatellite alleles were resolved by GeneScan analysis on ABI-377 DNA sequencer and characterized using the software Genotyper 2.1 (Applied Biosystems).

Results and Discussion

More than 1300 genomic clones putatively positive for SSR-motifs were identified in Southern hybridization based screening of the two libraries, of which >1000 clones were sequenced for both the strands. The summary of the marker development efforts is summarized in Table-1, which clearly shows that the genomic-library approach is resource intensive but reasonably good success can be achieved if the hybridization based screening is done appropriately.

In silco analysis of the sequenced clones for ascertaining frequency and distribution of the SSR repeats revealed that the di-repeats were the most abundant in the coffee genome followed by the mono-, tri- and tetra-/higher-order repeats in the decreasing proportion. Significantly, the spread of repeats were broadly comparable in the two species despite their ploidy differences; the only exception being a relatively higher proportion of tri-repeats in the

Table 1: Summary statistics of the work done for development of microsatellite markers using genomic libraries of coffee

	Parameters	*C. arabica*	*C. canephora*
1	Total number of clones picked-up from HDT library	50,000	18,000
2	Total no.of clones screened	50,000	15,744
3	Total no.of clones selected from primary screening	5,400	1200
4	Total no.of clones selected from secondary screening	850	446
5	No. of Clones sequenced	711	310
	Size of genome sequenced	580686_{bp} ($\sim 0.58_{Mb}$)	266600_{bp} ($\sim 0.27_{Mb}$)
	Range of the clone size	300_{bp} to 3.0_{kb}	280_{bp} to 2.6_{kb}
	Average size of the clones	830 bp	860 bp
6	No. of clones with repetitive motif >18bp	320	148
7	No. of clones used for primer design flanking the repeats	220	59
8	No.of primer pairs synthesized (till date)	160	59
9	No. of primer pairs validated as SSR markers	122	51

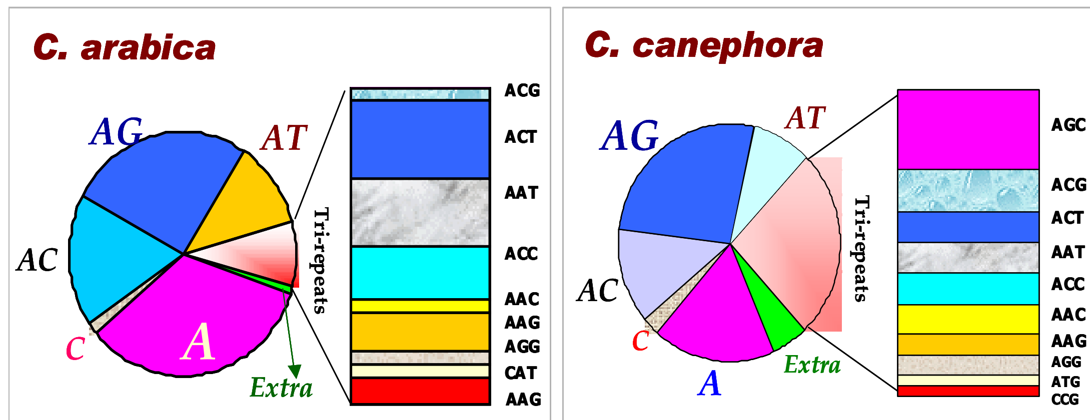

Figure 1. Distribution of various repeats among the sequenced clones of two coffee species.

diploid *C. canephora*. Overall, AG repeats were the most abundant among the di-repeats whereas poly-A characterized the mononucleotide stretches (Figure 1). Among the tri-repeat motifs for which both the libraries were screened, ACT, AAT and ACC were more abundant than the others, but in case of *C. canephora* library which was additionally screened for three other trirepeat motifs, the AGC repeats were the most abundant. The data also revealed many SSRs in the sequenced clones that were not used for screening the genomic libraries, e.g., AT repeats, stretches of mononucleotides etc. In general, these were seen in clones that also had parts of targeted repeats, thus indicating random overlapping distribution of SSRs in the coffee genome and the good quality of the genomic libraries generated in the study.

Hybridization based screening was quite efficient as 55% putatively selected clones were found positive for SSRs on sequencing. Of these, a large number had a repeat core length of >18 bp, an important attribute indicative of the suitability for possible conversion to an informative microsatellite marker. These short-listed clone sequences were used to design 219 primer pairs of which 173 pairs could be successfully converted to microsatellite markers. Detailed validation studies carried out using panels of arabica/robusta genotypes and related species representatives indicated their suitability for use in genetic studies, as well as broad cross-species transferability. In general, the newly developed microsatellites revealed an overall low heterozygosity among arabica genotypes but considerable allelic variability in the robustas genepool. These observations are strongly in agreement with the earlier studies on the genetic status of the coffee genepool that clearly show a much narrow genetic base of

the arabicas compared to robustas. Also, the genetic diversity seen using the new microsats is largely comparable with that seen using other DNA marker approaches (our unpublished data). Further, in comparison to the cultivated genepool, the new markers were found much more polymorphic in the related taxa (*Coffea* and *Psilanthus* spp.) with very broad cross-species specificities. The data thus suggest the potential of the new markers for genetic analysis of coffee especially its secondary genepool for conservation, management, resolving taxonomic relationships, and even more importantly for their use as efficient, informative genetic landmarks on interspecific molecular linkage maps being developed for genetic improvement of coffee.

Acknowledgement

We thank the Department of Biotechnology, New Delhi, India for the financial support, Dr Lalji Singh, Director, CCMB, Hyderabad for the facilities to undertake the study, Dr R Naidu, Director Research, Dr. N. SuryaPrakash, Dr. A. Shantaram, Central Coffee Research Institute, Chikamagalore for plant materials.

References

Aggarwal RK, Shenoy VV, Ramadevi J, Rajkumar R, Singh L (2002) Molecular characterization of some Indian Basmati and other elite rice genotypes using fluorescence-AFLP. *Theoretical and Applied genetics*, 105: 680-690.

Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P, Aggarwal RK (2003) Isolation and characterization of nine microsatellite markers from *Coffea arabica* (L.) showing wide cross-species amplifications. *Molecular Ecology Notes*, 3: 647-650.

Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, Varshney RK, Aggarwal RK (2004) Identification and characterization of gene-derived EST-SSR markers from robusta coffee variety ‘CxR’ (an interspecific hybrid of *Coffea canephora* x *C. congensis*). *Molecular Ecology Notes (in press).*

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica*, 113, 163–185.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, 106:411-422.