Development and characterization of microsatellite markers in coffee (*Coffea arabica*) for their use as potential genetic markers

Baruah A, Naik V, Rajkumar R, Rajendrakumar P, Hendre P S, Prakash NS*, Santaram A*, Sreenivasan CS*, Aggarwal RK

Abstract

Genetic improvement of coffee, an important plantation crop, through conventional breeding approach is very difficult and time consuming. Lack of genetic markers, suitable screening tools, information on the genetic make-up of the available genepool and long generation advancement are among the major constraints. Molecular Breeding, the newly emerging discipline, based on polymorphic DNA markers/techniques offers possibilities to overcome these constraints but require development and mapping of coffee specific DNA markers and molecular linkage maps. In this context, it is the small tandem repeat based microsatellite markers that have proven to be most desirable in various genetic studies on germplasm characterization and linkage analysis due to their codominant nature, stability, abundance, sensitivity, ease and speed of analysis, minimal sample requirements and suitability for automation. To our knowledge, only about 10 such markers have been developed and described in the literature for coffee till to date. We describe here the development of 27 new coffee specific microsatellite markers that can potentially be used for DNA typing of germplasm collections and also as genetic landmarks on molecular linkage map of coffee.

For the purpose, we have constructed a small insert, partial, genomic library comprising about 50,000 clones having inserts of 0.8-1.0 kb, from the total DNA of an elite genotype of *Coffea arabica* var. HDT (Hibrido De Timor). The latter is a putative natural interspecific hybrid of *C. arabica* and *C. canephora* and is used as a universal donor of rust resistance. High density filters representing around 12,000 of these clones were screened through Southern hybridization with 9 different synthetic oligonucleotide repeat probes viz., $(CA)_{15}$, $(GA)_{15}$, $(CAA)_{10}$, $(GGT)_{10}$, $(ATT)_{10}$, $(AGG)_{10}$, $(AGA)_{10}$, $(ACT)_{10}$ and $(CATA)_8$. Based on the hybridization signals, a large number of putative repeat-positive clones were identified of which 288 were fully sequenced for both the strands. Of these 143 clones were found to carry the repeats suggesting an estimated occurrence of ~$1.4x10^5$ repeats per haploid genome of *C. arabica*. Among various repeats, $(AT)_n$ repeats were found to be the most abundant followed by $(GA)_n$, $(CA)_n$ and $(ATT)_n$ that were estimated to be ~$9.6x10^4$, $1.1x10^4$, $9.1x10^3$ and $7.1x10^3$, respectively, in the arabica genome. These estimates corroborates well with our dot-blot based comparative hybridization results on relative abundance and distribution of different repeat motifs in the genome of different coffee species.

Of the repeat positive clones, 27 were selected for development of microsatellite markers. In each case, primers were designed from the flanking regions of the repeat and PCR conditions were standardized for successful amplification. Subsequently, each of the putative microsatellite markers was validated for polymorphism information content (PIC) and cross-species specificity using panels of elite arabica genotypes (23 improved selections and 33 rust differentials) and 17 *Coffea* species. The analysis revealed few alleles coupled with low heterozygosity in arabica genotypes but very large number of alleles (2 to 24, average of 13.06 alleles per locus) and high heterozygosity between the species. The study thus provide the largest number of coffee specific microsatellite markers described till to date, that can be used to investigate the level of genetic variation of coffee germplasm, for evolutionary/taxonomic studies and development of molecular linkage map of coffee.

Centre for Cellular and Molecular Biology, Hyderabad-500007, India.
* Central Coffee Research Institute, Chikamagalur-577117, India