Fingerprinting of Indian Coffee Selections and Development of Reference DNA Polymorphism Panels for creating Molecular IDs for Variety Identification

RAMESH K. AGGARWAL[*], R. RAJKUMAR, P. RAJENDRAKUMAR, PRASAD S. HENDRE, A. BARUAH, R. PHANINDRANATH, V. ANNAPURNA, N. S. PRAKASH[1], A. SANTARAM[1], C. S. SRINIVASAN[1], LALJI SINGH

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

Abstract

There are two main cultivated species of coffee namely, tetraploid *C. arabica* and diploid *C. canephora* (also called robusta coffee). A number of breeding programs were initiated in India at Central Coffee Research Institute, Chikamagalur, with main thrust being building resistance coupled with productivity without affecting quality in arabica and improvement of bean quality and yield in robusta. Such efforts spread over many decades resulted in the development of 16 superior selections of coffee (14 arabica and 2 robusta selections). The present study was undertaken to fingerprint these station-bred selections using high-resolution DNA marker approaches for their individualization and to ascertain their genetic base. DNA typing was performed using fluorescence-AFLPs, RAPDs and a large number of in-house developed coffee-specific microsatellite markers. The data revealed very limited variability in arabica genotypes, which could be visualized for their discrimination only by employing a very large number of DNA markers. The fingerprinting data were analyzed to create reference 'DNA polymorphism panels' (Microsatellite, AFLP, RAPD-based) which can readily be used for: a) generating 'Molecular IDs' needed for genotype individualization/registration/IPR protection; b) selecting genetically diverse genotypes and suitable/efficient marker approaches for breeding programs; and c) other genetic studies. In our opinion, this study involving DNA fingerprinting of Indian coffee selections, in persepctive strongly suggests the need for creation of molecular data banks (enumerating DNA marker efficiencies and DNA polymorphism status) for globally available elite coffee germplasm for use as reference resources. Availability of such resources would increase the practical utility and efficiency of the DNA typing approaches that are exorbitantly costly but are highly desired in coffee genetic improvement programs, especially on arabica coffee that inherently suffers with low genetic base.

Introduction

Coffee, a commodity of great importance in global trade, is cultivated in over 50 countries and revenue from this beverage crop contributes substantially to the national exchequer of the producing countries. The commercial coffee production relies on only two species *C. arabica*

(Arabica coffee) and *C. canephora* (Robusta coffee), which accounts for 70% and 30%, respectively, of world production. *C. arabica* is the only self-compatible tetraploid species of the genus characterized by low genetic diversity which has been attributed to its allotetraploid origin, reproductive biology and evolution process (Lashermes *et al* 1996). In contrast, considerable variability was reported among diploid species that form valuable gene reservoir for different breeding purposes (Berthaud and Charrier 1988). Indian coffee improvement programs to develop elite cultivars of both arabica and robusta, initiated in early 20th century at the Central Coffee Research Institute (CCRI), Balehonnur, have resulted in many superior coffee selections (14 of arabica and three cultivars of robusta) that are widely grown in the country.

High cost of development and need for germplasm protection, have increased the need for reliable methods of cultivar/variety identification, which in turn has served as the driving force for development of more reliable and cost effective approaches for genotype individualization. Traditionally, it is the morphological markers that have been for cultivars identification and purity. However, in recent past, molecular characterization (DNA fingerprinting) using DNA markers has become popular as it provides quicker and accurate assessment of genetic structure of the individual genotype without the confounding effect of environment, and high-resolution of discrimination and reliability. In coffee, potential of DNA marker technology is now being increasingly realized viz., germplasm characterization and genetic diversity analysis (Lashermes *et al* 1996, Anthony *et al* 2002, Steiger *et al* 2002), analysis of alien genome introgression (Prakash *et al* 2002, Herrera *et al* 2002) and identification of markers linked to the resistance genes (Noir *et al* 2003, Prakash *et al* 2004). In this paper we report the molecular characterization of Indian coffee selections using RAPD (Randomly Amplified Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism) and in-house developed microsatellite markers for generating molecular IDs for their individualization.

Material and Methods

The plant material used for analysis comprised of 16 station-bred selections (14 arabica and 2 robusta) developed at the Central Coffee Research Institute, India. Genomic DNA was isolated from fresh leaves as described by Aggarwal *et al* (2002), and used for molecular

characterization using three DNA typing approaches i.e., RAPD, f-AFLP and microsatellites. A total of 37 random decamer primers, 7 f-AFLP primer-pairs (each primer having 3-selective bases) and 150 in-house developed microsatellite markers were used for analysis. The procedures followed and PCR conditions used for AFLP and Microsatellite markers were as described by Aggarwal *et al* (2002) and Baruah *et al* (2003) respectively.

Results and Discussion

The DNA typing of the Indian-bred selections revealed that, irrespective of the marker approach used (RAPD, SSR and AFLP), the DNA profiles were very distinctive between diploids and tetraploids and the polymorphism was always more for diploid robustas. The RAPD analysis with 37 random decamer primers generated a total of 690 markers of which 80% markers were polymorphic among the 16 genotypes tested. The AFLP assays with 7 primer pairs generated 320 markers out of which 69% were polymorphic. Majority of the polymorphism revealed with these two multi-locus DNA typing approaches was accounted by diploid robusta selections (61% for RAPDs and 51% in case of AFLPs) as against only 18-19% in tetraploid arabica selections (Figure 1). Almost similar level of polymorphism was seen with microsatellite markers, wherein 23% markers were found to be polymorphic when only the tetraploid selections were considered compared to 57% being informative in case of the 2 diploid robustas (Figure-1).

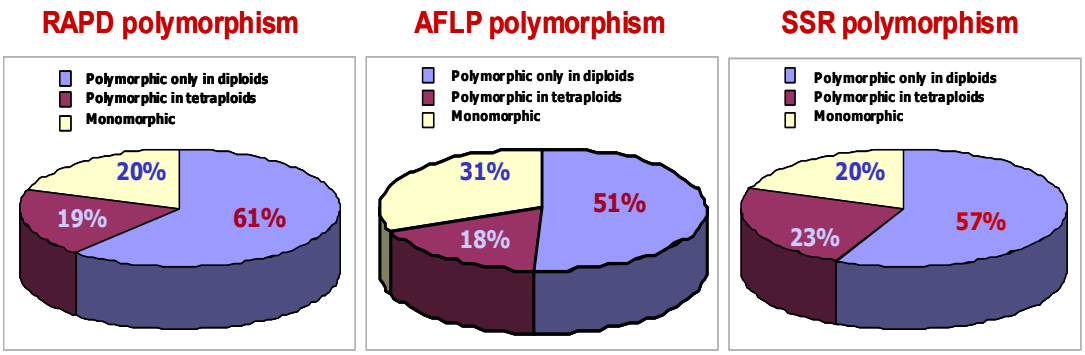

Figure 1. Distribution of DNA polymorphism seen in station-bred arabica (tetraploid) and robusta (diploid) selections using 3 different DNA typing approaches.

Significant differences were observed between different RAPD primers, AFLP-combinations and microsatellite markers with respect to the informative polymorphism revealed by them. A number of these marker/primer combinations were found to be completely monomorphic (and thus uninformative) for the materials analyzed (especially arabica selections) and thus of low to negligible utility for the purpose of germplasm characterization.

Although genetic variation among arabica cultivars was limited, the DNA polymorphism obtained was found to be sufficient to distinguish and individualize the arabica genotypes studied. The fingerprinting data were also analyzed to identify the most informative markers (primer combinations/approaches) that could distinguish the coffee selections especially the arabicas. Of the many informative RAPD primers, a sub-set of 6 primers could be identified that discriminated all the coffee selections. Similarly, 4 of the 7 AFLP primer-combinations were identified that could resolve all the genotypes. Further, compared to RAPDs and AFLPs, although relatively a lower proportions of the ~150 tested microsatellite markers were informative for arabicas, but these were relatively more efficient to discriminate all the 16 selections (e.g., a sub-set of 6 of the ~30 microsatellites that showed some polymorphism among the 14 arabica selections, could reliably distinguish all of them).

	CM 42					CM 54					CM 136			CM 115					CM 72					CM 27					CM 158			cm 83		CM 92		CM 82	
1		132		140				228			110			150				177			198			230			170	175		197			199	221	269	281	193	201
2		132		140				228			110			150				177			198			230		168		175		197			199	221	269	281	193	201
3		132		140				228			110		146	150			175				198		220			168	170			197			199		269		193	205
4		132		140				228			110		146	150				177			198		220			168	170			197			199	231	269		193	201
5		132		140				228			110		146	150				177			198		220		156	168	170			197			199	221	269		193	201
6			134		146				233		110		146	150				177			198			230		168	170			197	204	220	199	221	269		193	201
7	130			140				228	233		110		146	150		160	175				198			230		168	170			197	204	220	199	221	257		193	201
8			134	140				228	233		110		146	150		160	175					200		230		168	170			197	204	220	199	221	257		193	201
9			134	140					233		110		146		153			177				200	220			168	170			197			199	221	269		193	201
10		132		140				228			110			150					179		198			230		168			177	197			199	231	269	281	193	201
11		132		140				228			110		146	150				177			198		220		156	168				197			199	221	269		193	201
12		132		140				228			110			150				177			198					168	170			197	204		199		269		193	201
13					146		224				110	123		150				177		186	198					168	170			197	204	212	203	219	257	269	201	
14					146	214	224				110	123		150				177				200					170			197	204		203	219	269		201	
15			134		146			228			110		146	150				177				200	220			168	170			197			199	229	269		193	201
16				140		214		228			110	123		150		160	175					200		230		168	170			197			199	219	269		193	201

Figure 2: Part of 'reference polymorphism panel' showing allelic distribution (across 10 in-house developed coffee specific microsatellite markers) for 16 station-bred coffee selections developed in India. Note that each of the selection can be uniquely individualized in the panel. Sixteen coffee selections are: S.288, S.795, Sln-4 (Agaro), Sln-4 (Ciocee), Sln-5A, Sln-5B, Sln-7.3, Sln-8, Sln-9, Sln-10, Sln-11, Sln-12, S-274, CxR, Sln-4 (Tafarikela) and Sln-6.

To increase the long-term utility of the DNA fingerprinting analysis, all the DNA marker data were used to create 'Reference DNA polymorphism Panels', cataloguing allelic/marker distribution revealed by different DNA typing approaches for all the coffee genotypes. These panels can be used to generate 'Molecular IDs' for genotype individualization and for other advantages in programs of genetic improvement, such as, selection of suitable/diverse parental materials and also most informative DNA marker(s) for linkage analysis avoiding the need for 'parental survey'. A representative 'Reference panel' based on the allelic polymorphism data across 10 coffee specific microsatellite loci (developed at CCMB, Hyderabad) suitable for identifying/distinguishing the CCRI station-bred selections is shown in figure-2.

Conclusions and Perspectives

The present study demonstrated that DNA markers could be used reliably for identification of the coffee germplasm and that the microsatellites are the most efficient and desirable markers for the purpose. The data obtained can be used for ascertaining the genetic diversity in the available cultivated and wild exotic germplasm of coffee. The present study also highlights the overall low efficiency of these otherwise high-resolution approaches for coffee genepool characterization. The data clearly establish that despite their high-genetic resolution, DNA typing need to be carried out at large scale to resolve the low variation inherent in the coffee genepool especially of arabica, thus making the whole exercise resource intensive and practically non-viable. The later constrain warrants development of ways and means to increase the efficiency of DNA markers based expensive but unavoidable approaches for coffee genetic analysis. An easy way to achieve this can be creation of molecular data banks enumerating DNA marker (approach) efficiencies and more importantly, DNA polymorphism status (using defined markers/guidelines) for the analyzed germplasm, for use as reference resource. Based on our experience, we propose construction of 'Reference DNA polymorphism databases' of elite coffee germplasm available world-wide using standard repeatable markers (such as microsatellites) that can then be used by the coffee geneticists/breeders community for various advantages namely, for: a) better management, utilization, registration and IPR protection of elite coffee germplasm; b) selection of suitable material/genotypes and informative DNA markers for breeding and linkage analysis; c) exchange of germplasm, etc.

Acknowledgements

This study is part of the Coffee Genomics program at CCMB carried out under the First National initiative on Coffee Biotechnology financially supported by the Department of Biotechnology, Govt. of India. The authors thank DBT for financial support, Coffee Board the collaborating institute for the cooperation and plant material supply and Dr. Lalji Singh, Director CCMB, Hyderabad for the experimental facilities.

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.

Anthony F, Combes MC, Astorga C, Bertrand B, Graziosi G, Lashermes P (2002) The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics,* 104: 894-900.

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(4): 647-650.

Berthaud J and Charrier A (1988) Agronomy, Genetic Resources of *Coffea* In: *Coffee*, Vol 4 (eds, Clarke RJ, Macrae R), EA Science, London pp 1-42.

Herrera JC, Combes MC, Anthony F, Charrier A, Lashermes P (2002) Introgression into the allotetraploid coffee (*Coffea arabica* L.): Segregation and recombination of the *C. canephora* genome in the tetraploid interspecific hybrid (*C. arabica* x *C. canephora*). *Theoretical and Applied Genetics,* 104: 661-668.

Lashermes P, Trouslot P, Anthony F, Combes MC, Charrier A (1996) Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica* L. *Euphytica,* 87: 59-64.

Noir S, Anthony F, Bertrand B, Combes MC, Lashermes P (2003) Identification of a major gene (*Mex*-1) from *Coffea canephora* conferring resistance to *Meloidogyne exigua* in *Coffea arabica. Plant Pathology*, 52: 97-103.

Prakash NS, Combes MC, Naveen KS, Lashermes P (2002) AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid. *Euphytica*, 124: 265-271.

Prakash NS, Marqes DV, Varzea VMP, Silva MC, Combes MC, Lashermes P (2004) Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *C. arabica* L. *Theoretical and Applied genetics*, (*In press*); online: 10.1007/s00122-004-1748-z.

Steiger DL, Nagai C, Moore PH, Morden CW, Osgood RV, Ming R (2002) AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theoretical and Applied Genetics*, 105: 209-215.