sciendo

10.2478/botlit-2020-0019

BOTANICA
ISSN 2538-8657

2020, 26(2): 184–196

ASSESSMENT OF THE SUITABILITY OF MOLECULAR SCOT MARKERS FOR GENETIC ANALYSIS OF COFFEE SPECIES

Manoj Kumar Mishra*, Arun Kumar C. Huded, Pavankumar Jingade

Coffee Board, Central Coffee Research Institute, Division of Plant Biotechnology, Manasagangothri, 570006 Mysore, India
*Corresponding author. E-mail: manojmishra.m@gmail.com

Abstract

Mishra M.K., C. Huded A.K., Jingade P., 2020: Assessment of the suitability of molecular SCoT markers for genetic analysis of coffee species. – Botanica, 26(2): 184–196.

Start codon targeted polymorphism (SCoT), a novel and gene-targeted marker, has recently become the marker of choice in genetic diversity studies. In the present study, 31 SCoT primers were tested for their suitability in the genetic analysis of 21 coffee genotypes representing 18 species. A total of 647 distinct PCR amplified fragments were produced with a mean of 20.9 fragments per primer and 80.80% of which were polymorphic. The polymorphic information content of SCoT primers ranged from 0.16 to 0.86, with a mean value of 0.63. Resolving power ranged from 6.19 to 28.29, with a mean value of 20.2. Species-specific unique PCR amplified fragments were identified for 16 species, which could be used as genetic fingerprints. The genetic similarity among various coffee species calculated using the Dice similarity coefficient ranged between 0.60 and 0.89. The dendrogram constructed using the unweighted pair group of arithmetic means (UPGMA) clustered the 21 coffee genotypes into two major groups. The study revealed that *Coffea jenkinsii*, an indigenous species from India, showed the highest similarity with *C. arabica*, which is of Ethiopian origin. The results proved the suitability of SCoT markers in genetic analysis of coffee genotypes.

Keywords: *Coffea* species, DNA-based markers, genetic diversity, species-specific fragments, start codon targeted polymorphism (SCoT).

INTRODUCTION

The genus *Coffea* L. belongs to the family Rubiaceae that consists of more than 124 species (Razafinarivo et al., 2013). Most coffee species naturally occur in tropical forests of Africa, Madagascar and Mascarene islands, while some of the species are found in the Indian subcontinent, Southeast Asia and tropical Australasia (Davis et al., 2006, 2010, 2011; Davis, 2011). However, all the coffee species are not commercially cultivated except two main species *Coffea arabica* L. and *C. canephora* Pierre ex A. Froehner, which provides the global commodity for consumption (Mishra, 2019). Coffee is grown in about 10.2 million hectares of land spanning over 80 countries and is the second most important commodity in the world trade market after petroleum. In 2018, the total global production of green coffee was about 173.09 million bags (60 kg capacity) with an export earning of over US$ 30.1 billion during 2020 (ICO report 2020). More than 125 million people in the world derive their income directly or indirectly from the produce of this crop (Mishra, 2019).

Due to the prodigious amount of economic importance, intensive research has been carried out on evolutionary and genetic aspects of various coffee species by using conventional and molecular analysis of both nuclear and plastid genomes (Davis et al., 2007; Maurin et al., 2007; Tesfaye et al., 2007; Anthony et al., 2010; Davis et al., 2011; Nowak et al.,

2012). However, most of the previous studies have not included the indigenous coffee species from India while delineating the genetic relationship among the species. To date, only a few papers have been published on DNA based molecular analysis of Indian wild coffee species (Mishra et al., 2011a, b). Understanding the genetic diversity and relationship among different species would be useful for the genetic improvement of coffee because wild species are the genetic reservoirs of various traits not usually observed in cultivated species (Brar, 2005; Hajjar & Hodgkin, 2007). Previous studies have indicated the introgression of resistance genes from wild coffee species to commercially grown lines. For example, the genes imparting resistance for coffee leaf rust (*Hemileia vastatrix* Berk and Broome) and coffee leaf minor have been introgressed in commercially cultivated lines from *C. liberica* and *C. racemosa*, respectively (Guerreiro et al., 1999; Mishra & Slater, 2012). Other spontaneous natural interspecific hybrids such as Hybrido-de Timor, Devamachy (*C. arabica* × *C. canephora*) and S.26 (*C. arabica* × *C. liberica*) have been extensively used as the source for pest and disease resistance in the *C. arabica* breeding programme. Recently, Mishra et al. (2018) have studied the physiological and molecular characters of a natural interspecific hybrid involving *C. excelsa* and *C. arabica* and observed that the hybrid is more tolerant of biotic and abiotic stress. This gives credence to the characterisation and exploitation of diploid coffee species in the genetic improvement of coffee. The information on genetic variability among various coffee species would be useful for designing suitable strategies for their ex-situ conservation of rare plants with valuable new alleles. Since different molecular markers unmask the hidden genetic variability by targeting other parts of the genome, it is necessary to use various markers for comparison.

In recent years, a novel marker system called Start codon targeted polymorphism (SCoT) (Collard & Mackill, 2009) has been extensively used for genetic analysis of many plant species. SCoT is a simple and novel marker system that is based on the short conserved region in plant genes surrounding the ATG translation start (or initiation) codon. Since the marker system focuses on genic regions, which usually have low recombination levels between marker alleles and the gene or trait as compared to RAPDs, ISSRs and SSRs, the SCoT marker is considered more advantageous and reliable (Hajibarat et al., 2015). SCoT markers have been successfully used in genetic diversity analysis, cultivars identification, quantitative trait loci (QTL), differential gene expression and screening of stress-related genes, and assessment of genetic fidelity of tissue culture plants in different plant species including grape (Guo et al., 2012), mango (Luo et al., 2011, 2014), chickpea (Amirmoradi et al., 2012), sugarcane (Que et al., 2014), wheat (Hamidi et al., 2014), *Albizia* spp. (Rahmani et al., 2015) and bottle gourd (Bhawna et al., 2017). However, to date, there is not much data available to establish the genetic relationships and functional diversity between different species of coffee, using SCoT molecular markers. The objective of the present research was to test the suitability of SCoT markers for genetic analysis of coffee species.

MATERIALS AND METHODS

Plant material

All the 21 genotypes representing 18 coffee species were grown in the nursery of Tissue Culture and Plant Biotechnology Division, Mysore. The details of coffee genotypes, along with their codes and country of origin, are provided in Table 1.

Isolation of genomic DNA

About 100 mg of tissue from the young leaves was ground to a fine powder using liquid nitrogen and transferred to 1.5 ml microcentrifuge tubes containing 750 μL of preheated extraction buffer (2% CTAB (w / v), 100 mM Tris-HCl (pH 8.0), 25 mM EDTA, 2M NaCl and 0.1% beta-mercaptoethanol). The tubes containing the lysates were incubated in a hot water bath at 65ºC for 45 min and then centrifuged at 6000 rpm for 15 min. The supernatant was transferred to a fresh tube, and equal volumes of chloroform: isoamyl alcohol (24:1) was added. The samples were gently mixed by inverting the tubes and centrifuged at 6000 rpm for 15 min. The clear upper phase was transferred into a fresh 1.5 ml microcentrifuge tubes and DNA was precipitated by adding to 0.7 volume of ice-cold isopropanol and incubated at -20ºC for 15 min. The samples were centrifuged at 8000 rpm for 15 min, and the DNA pellet was washed

with 75% (v/v) ethanol for 10 min following which the tubes were centrifuged at 8000 rpm for 5 min, and the ethanol was discarded. The DNA pellet was dried at room temperature until traces of ethanol was removed and dissolved in 100 µl of TE buffer (10 mM–Tris and 1 mM–EDTA). RNase treatment was carried out by adding 5 µL of RNase enzyme solution and incubating at 37ºC. The RNase was deactivated by heating the samples to 70ºC. The quality of DNA was accessed by separating it on 0.8% agarose gel stained with ethidium bromide (0.5 µg / ml) and quantified using a UV spectrophotometer at 260 nm and 280 nm. The DNA samples were diluted to a working concentration of 10 ng/µl using sterilised distilled water and stored at -20ºC for future use.

SCoT marker analysis

Thirty-one SCoT primers developed by Collard & Mackill (2009) (Table 2) were used for assessing their suitability in the genetic analysis of coffee species. The integrity and reproducibility of the bands amplified by SCoT primers were validated by conducting the PCR amplification at least two times. PCR amplification was carried out in a total reaction volume of 20 µL using Bio-Rad Thermal cycler S1000. PCR reaction mixtures contained 5.0 µL of template DNA (10 ng/µL), 3.0 µL of 3 µM primer, 2.0 µL of 2 mM dNTPs, 2.0 µL of 10X Taq buffer, 2.0 µL of 25 mM $MgCl_2$ (all from Thermo Fisher Scientific, Waltham, USA), and 0.5 µL of 3 units/µL Taq DNA polymerase enzyme (GeNei). Standard PCR cycling parameters were used, which includes an initial denaturation step of 5 min at 94°C, followed by 30 cycles of 94°C for 30 s, primer annealing at 48°C for 2 min, primer extension at 72°C for 2 min, and final extension of 15 min at 72°C. The PCR amplified products were mixed with 5 µL Bromophenol blue dye (99.5% deionised formamide, 10 mM EDTA pH 8, 0.05% Bromophenol blue, xylene-cyanol dye solution, 1µL pure, sterile water) and separated on 1.5% agarose gel (SeaKem, Rockland USA) containing 0.5 µg ethidium bromide/ml in 1 × TBE (Tris-HCl, Boric acid, EDTA) buffer. After electrophoresis, the gels were visualised and documented using Gel Doc System (BioRad) with a Multi Analyst software programme.

Table 1. List of coffee species along with their code, place of origin and conservation status used for SCoT marker analysis

Species	Species code	Place of origin/distribution	Conservation status
C. arabica L. 'Kents'	S-1	Ethiopia/India	Vulnerable
C. canephora Pierre ex A. Froehner, cv. S. 274	S-2	West Tropical Africa/India	Least concern
C. canephora var. *ugandae* (Cramer) A.Chev.	S-3	West Tropical Africa	Least concern
C. canephora var. *quillon* Philipp.	S-4	West Tropical Africa	Least concern
C. congensis var. *froehneri* Pierre ex De Wild.	S-5	West Central Africa Congo	Least concern
C. eugenioides S.Moore	S-6	West Tropical Africa	Least concern
C. zanguebariae Lour.	S-7	East Tropical Africa	Vulnerable
C. racemosa Lour.	S-8	Southern Tropical Africa	Near threatened
C. kapakata (A. Chev.) Bridson	S-9	West Angola	Vulnerable
C. stenophylla G.Don	S-10	West Tropical Africa	Least concern
C. salvatrix Swynn.& Philipson	S-11	East Tropical Africa	Near threatened
C. abeokutae P. J. S. Cramer	S-12	West Tropical Africa	Least concern
C. liberica Hiern	S-13	West Tropical Africa	Least concern
C. dewevrei De Wild. & T.Durand	S-14	Democratic Republic of Congo	Vulnerable
C. arnoldiana De Wild.	S-15	West Central Africa	Least concern
C. dewevrei var. *excelsa* (A.Chev.) A.Chev.	S-16	West Central Africa	Least concern
C. bengalensis Roxb.ex Schult.	S-17	India	Endangered
C. travancorensis Wight & Arn.	S-18	India	Endangered
C. khasiana (Korth.) Hook.f.	S-19	India	Endangered
C. wightiana Wall. ex Wight & Arn.	S-20	India	Endangered
C. jenkinsii Hook.f.	S-21	India	Endangered

Table 2. Sequences of SCoT primers used in the analysis of coffee species

SCoT primer	Sequence (5' to 3')
SCoT 1	CAACAATGGCTACCACCA
SCoT 2	CAACAATGGCTACCACCC
SCoT 3	CAACAATGGCTACCACCG
SCoT 4	CAACAATGGCTACCACCT
SCoT 5	CAACAATGGCTACCACGA
SCoT 6	CAACAATGGCTACCACGC
SCoT 7	CAACAATGGCTACCACGG
SCoT 8	CAACAATGGCTACCACGT
SCoT 9	CAACAATGGCTACCAGCA
SCoT 10	CAACAATGGCTACCAGCC
SCoT 11	AAGCAATGGCTACCACCA
SCoT 13	ACGACATGGCGACCATCG
SCoT 14	ACGACATGGCGACCACGC
SCoT 15	ACGACATGGCGACCGCGA
SCoT 16	ACCATGGCTACCACCGAC
SCoT 17	ACCATGGCTACCACCGAG
SCoT 18	ACCATGGCTACCACCGCC
SCoT 19	ACCATGGCTACCACCGGC
SCoT 20	ACCATGGCTACCACCGCG
SCoT 21	ACGACATGGCGACCCACA
SCoT 22	AACCATGGCTACCACCAC
SCoT 28	CCATGGCTACCACCGCCA
SCoT 30	CCATGGCTACCACCGGCG
SCoT 31	CCATGGCTACCACCGCCT
SCoT 34	ACCATGGCTACCACCGCA
SCoT 36	GCAACAATGGCTACCACC
SCoT 39	CAATGGCTACCACTAGCG
SCoT 48	ACAATGGCTACCACTGGC
SCoT 61	CAACAATGGCTACCACCG
SCoT 62	ACCATGGCTACCACGGAG
SCoT 63	ACCATGGCTACCACGGGC

Scoring and construction of the dendrogram

Gel images of PCR amplicons were scored as absent (0) or present (1) based on size compared to a marker (1Kb plus Ladder, Thermo Fisher Scientific, Waltham, USA) to form a binary matrix. The total number of fragments, the distribution of fragments across the species, polymorphic fragments, species-specific fragments, and the mean number of fragments per primer were calculated. The value of each SCoT primer was assessed using two indices; PIC (Polymorphic Information Content), which is the same as Diversity Index (DI) (Botstein et al., 1980; Milbourne et al., 1997) and resolving power (Rp) (Prevost & Wilkinson, 1999). The PIC or DI was estimated as PIC = $(1-p_i^2)/n$, where n is the number of band positions analysed in all the species, p_i is the frequency of the banding pattern. The resolving power of a primer is Rp = I_b, where I_b (band informativeness) takes the value of 1 – [2x (0.5–p)] and p is the ratio of samples sharing the band. A pairwise similarity matrix was constructed using the Dice similarity coefficient (Sneath & Sokal, 1973). The relationship between the species was displayed as a dendrogram constructed using NTsys 2.10e software (Rohlf, 1995) based on the Unweighted Pair Group Method using Arithmetic averages (UPGMA) with a bootstrap value of 1000.

The multivariate analysis by employing GenAIEX 6.5 software (Peakall & Smouse 2012) was performed to define relationships between 21 coffee genotypes employing Principal Coordinate Analysis (PCoA) using the Dice similarity coefficient.

RESULTS

In the present study, highly polymorphic fingerprinting patterns were obtained by using 31 SCoT primers in 21 genotypes belonging to 18 different coffee species (Figs 1–2).

The 31 primers produced 647 distinct scoreable fragments with the number of amplified fragments ranging from 7 (SCoT-10) to 30 (SCoT-4 and SCoT-17) with a mean of 20.87 fragments per primer (Table 3). Out of the total 647 amplified fragments, 544 were polymorphic, with a mean of 18 polymorphic fragments per primer. The amplified product size generated using different primers ranged from 300bp to 4kb. The per cent polymorphism ranged from 15% to a maximum of 100% with a mean of 80.80%. Among 31 primers screened, six primers showed 100% polymorphism. To determine the informative and discriminatory capacity of each primer, polymorphic information content (PIC) and resolving power (RP) were calculated (Table 3). The resolving power ranged from 6.19 (SCoT-10) to 28.29 (SCoT-4) with a mean of 20.2. Similarly, the polymorphism information content (PIC) ranged from 0.16 (SCoT-14) to 0.86 (SCoT-8) with a mean value of 0.63 (Table 3). The fragments generated by SCoT primers across the 21 genotypes were computed to understand the frequency occurrence of each amplified fragment, and it was observed that 44% of the amplified fragments

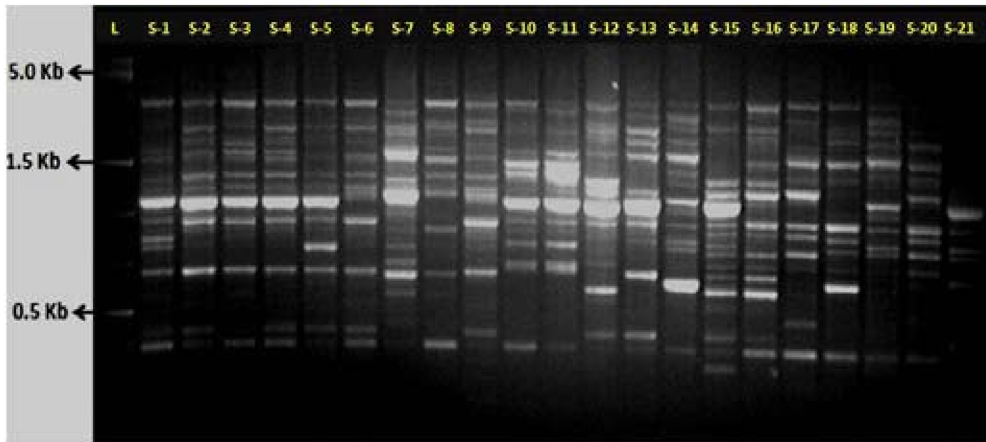

Fig. 1. Gel picture depicting the amplification pattern of coffee species obtained using SCoT 17 primer. Lane L – molecular ruler (1 Kb plus ladder), lanes S-1 to S-21 – coffee species listed in Table 1

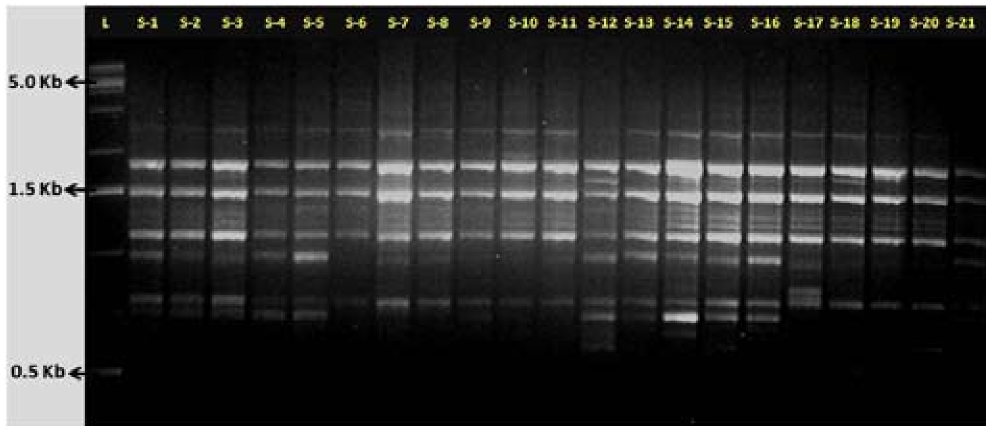

Fig. 2. Gel picture depicting the amplification pattern of coffee species obtained using SCoT 20 primer. Lane L – molecular ruler (1 Kb plus ladder), lanes S-1 to S-21 – coffee species listed in Table 1

occurred in 90–100% frequency, whereas 0.1 to 4.99% of fragments occurred in 1% frequency across the genotypes (Table 4).

Thirty-one SCoT primers generated 62 unique fragments in 18 coffee species out of which five Indian species (*C. bengalensis*, *C. travancorensis*, *C. wightiana*, *C. khasiana* and *C. jenkinsii*) generated 18 unique fragments (Table 5). Among the Indian species, the maximum number of unique fragments was generated by *C. bengalensis* (7) followed by *C. travancorensis*, *C. wightiana* (4 each), and the least number of unique fragments were observed in *C. khasiana* (3). However, *C. arabica*, *C. canephora* var. *ugandae*, *C. eugenioides*, *C. stenophylla* and *C. jenkinsii* failed to produce any unique fragments (Table 5). The genetic similarity calculated between different species ranged from 0.60 (*C. abeokutae* and *C. khasiana*) to 0.89 (*C. canephora* cv. S.274 and *C. canephora* var. *ugandae*, *C. canephora* var. *quillon* and *C. canephora* var. *ugandae*). Among the indigenous species, *C. bengalensis* showed the highest similarity of 0.78 with *C. travancorensis*, and *C. wightiana*, whereas the least genetic similarity of 0.66 was observed between *C. bengalensis* and *C. jenkinsii.*

Based on the SCoT marker data analysis, a strict consensus tree was constructed using the UPGMA clustering algorithm, and all the 21 coffee genotypes were grouped into two major clusters (Fig. 3). The first major cluster consisted of only two species *C. dewevrei* and *C. khasiana*. The second major cluster was divided into four minor clusters. The first minor cluster brought together *C. bengalensis*, *C. wightiana* and *C. travancorensis*. In the second minor cluster, *C. abeokutae*, *C. arnoldiana* and *C. dewevrei*, var. *excelsa* were grouped together. The third minor clus-

Table 3. Evaluation of 31 SCoT primers in genetic analysis of coffee species

Primers	Total bands	Size range (bp)	Number of bands in each species (mean)	No. of polymorphic bands	Percentage of polymorphism	Resolution power	Polymorphic information content
SCoT 1	18	450–3000	2–10 (6.43)	17	94.44	12.86	0.81
SCoT 2	29	450–3000	9–16 (12.76)	28	96.55	25.52	0.72
SCoT 3	21	500–3000	5–15 (9.38)	19	90.47	18.76	0.71
SCoT 4	30	450–2500	10–19 (14.14)	28	93.33	28.29	0.69
SCoT 5	28	500–3000	4–17 (11.95)	28	100.00	23.90	0.74
SCoT 6	20	500–4000	8–16 (11.95)	17	85.00	23.90	0.53
SCoT 7	19	700–2500	4–11 (7.00)	17	89.47	14.00	0.76
SCoT 8	27	700–3000	6–12 (08.43)	27	100.00	16.86	0.86
SCoT 9	19	800–3000	4–12 (7.67)	15	78.94	15.33	0.73
SCoT 10	7	1400–3000	2–06 (3.10)	5	71.42	6.19	0.66
SCoT 11	16	450–3500	8–13 (11.76)	8	50.00	23.62	0.30
SCoT 13	20	700–3000	8–16 (12.67)	16	80.00	25.33	0.49
SCoT 14	13	500–3000	11–12 (11.33)	2	15.38	22.67	0.16
SCoT 15	11	500–2500	7–10 (8.14)	4	36.36	16.29	0.29
SCoT 16	22	900–2500	7–17 (11.81)	18	81.81	23.62	0.59
SCoT 17	30	380–2800	9–17 (13.33)	28	93.33	26.67	0.71
SCoT 18	19	500–3000	10–16 (12.62)	11	57.89	25.24	0.42
SCoT 19	28	550–2200	9–19 (12.71)	24	85.71	25.43	0.67
SCoT 20	21	400–2500	5–14 (9.86)	16	76.19	19.71	0.63
SCoT 21	26	100–2500	4–17 (9.38)	26	100.00	18.76	0.80
SCoT 22	19	500–2500	3–12 (8.09)	18	94.73	16.19	0.73
SCoT 28	27	300–3500	9–19 (13.62)	21	77.77	27.24	0.61
SCoT 30	15	550–2800	10–14 (11.90)	6	40.00	23.81	0.27
SCoT 31	22	300–2000	7–2 (9.43)	21	95.45	18.86	0.68
SCoT 34	17	550–2500	5–09 (6.86)	14	82.35	13.71	0.71
SCoT 36	17	520–2500	4–11 (8.67)	17	100.00	17.33	0.66
SCoT 39	21	700–2800	3–11 (6.57)	21	100.00	13.14	0.86
SCoT 48	20	550–2500	9–13 (11.43)	15	75.00	22.86	0.56
SCoT 61	21	500–3000	5–13 (9.91)	19	90.47	19.81	0.68
SCoT 62	22	550–2800	3–10 (6.86)	22	100.00	13.71	0.84
SCoT 63	22	550–3000	11–18 (13.33)	16	72.72	26.67	0.53
Total	647		201–427 (313.09)	544	2504.78	626.3	19.40
Mean	20.87		6.5–14.0 (10.10)	18	80.80	20.20	0.63

ter comprised of *C. arabica* 'Kents', *C. canephora* var. *ugandae*, *C. canephora* var. *quillon*, *C. eugenioides*, *C. congensis*, *C. canephora* cv. S.274 and *C. jenkinsii*, which is an Indian wild coffee species. The fourth minor cluster included *C. zanguebariae*, *C. salvatrix*, *C. racemosa* and *C. stenophylla*.

Using SCoT markers, 21 genotypes were clustered into four main groups on the PCoA scatter plot (Fig. 4). The PCoA mapped three Indian coffee species *C. bengalensis*, *C. travancorensis* and *C. wightiana* along with *C. racemosa* and *C. stenophylla* in the first group. The second and third groups included most wild coffee species of African origin. The fourth group compiled of *C. arabica*, *C. canephora*, *C. congensis*, *C. eugenioides* and *C. kapakata*. The dendrogram and the PCoA analysis revealed that species clustered together in the dendrogram were placed together using PCoA analysis (Fig. 4).

DISCUSSION

Characterisation and evaluation of genetic diversity among coffee species are essential to understand the value of germplasm collections and their usefulness in breeding programmes. Previously, the genetic diversity of coffee species has been studied by using RAPD, ISSR and SRAP markers (Mishra et al., 2011a, b). However, with the rapid development

Table 4. Details of species-specific fragments amplified by SCoT markers in different coffee species

Coffee species	SCoT marker
C. arabica L. 'Kents'	–
C. canephora Pierre ex A.Froehner, cv. S. 274	SCoT-28-3500
C. canephora var. *ugandae* (Cramer) A.Chev.	–
C. canephora var. *quillon* Philipp.	SCoT-18-720, SCoT-19-1400, SCoT-19-600
C. congensis var. *froehneri* Pierre ex De Wild.	SCoT-9-2900,SCoT-34-1000
C. eugenioides S.Moore	–
C. zanguebariae Lour.	SCoT-1-900, SCoT-2-680, SCoT-11-1400, SCoT-14-1200, SCoT-17-1300, SCoT-19-550, SCoT-28-480, SCoT-34-600, SCoT-39-2200, SCoT-62-710
C. racemosa Lour.	SCoT-6-720,SCoT-22-1500, SCoT-36-2500
C. kapakata (A. Chev.) Bridson	SCoT-7-1000, SCoT-9-1900
C. stenophylla G.Don	–
C. salvatrix Swynn.& Philipson	SCoT-5-750, SCoT-11-900, SCoT-31-1200
C. abeokutae P. J. S. Cramer	SCoT-20-1600, SCoT-61-1000
C. liberica Hiern	SCoT-2-600, SCoT-4-300, SCoT-6-800, SCoT-31-2500, SCoT-31-600, SCoT-62-1000, SCoT-63-2800
C. dewevrei De Wild. & T.Durand	SCoT-6-750, SCoT-7-900, SCoT-7-800, SCoT-18-1400, SCoT-20-800
C. arnoldiana De Wild.	SCoT-3-600, SCoT-17-390, SCoT-34-1200, SCoT-48-650
Coffea dewevrei var. *excelsa* (A.Chev.) A.Chev.	SCoT-20-650, SCoT-21-1400
C. bengalensis Roxb.ex Schult.	SCoT-2-400, SCoT-7-1200, SCoT-8-3100, SCoT-17-500, SCoT-20-750, SCoT-20-710, SCoT-21-1150
C. travancorensis Wight & Arn.	SCoT-19-2400, SCoT-21-1200, SCoT-22-1300, SCoT-62-2500
C. khasiana (Korth.) Hook.f.	SCoT-3-680, SCoT-15-700, SCoT-62-550
C. wightiana Wall. ex Wight & Arn.	SCoT-2-680, SCoT-5-1000, SCoT-5-1100, SCoT-61-1400
C. jenkinsii Hook.f.	–

in the field of genomics, new functional markers targeting the genes and promoters are available and used in the genetic analysis of many plant species with better resolution and reproducibility (Poczai et al., 2013). SCoT marker assay preferentially targets the coding sequences in the genome, because the primers are based on the short conserved region surrounding the ATG translation start codon (Collard & Mackill, 2009; Xiong et al., 2009). Therefore, the polymorphism revealed by the SCoT marker is directly related to the diversity at the gene level, which could be possibly involved in phenotypic trait variation (Andersen & Lubberstedt, 2003). Although a large amount of data was obtained with SCoT markers for various plant species, the genetic diversity and breeding value of the coffee gene pool using the SCoT marker is yet to be ascertained. The present study is the first report on the use of SCoT markers to estimate the breeding value of the coffee gene pool.

The data generated in the present study by screening 21 coffee genotypes with 31 SCoT revealed a mean of 80.80% polymorphic fragments. The per cent polymorphism obtained using SCoT markers was low as compared to SRAP and ISSR marker analysis, wherein 96.12 and 93.06% polymorphism was obtained, respectively (Mishra, 2011b). The lower percentage of mean polymorphism could be due to the differential working principle of the SCoT marker system, wherein only the functional domains of genes were targeted. The functional domains of genes comprise a smaller portion of the total genome. They are more conserved, unlike non-functional regions, resulting in a lower polymorphic percentage of the marker system. Nevertheless, the polymorphism exhibited by the SCoT marker system is more critical as it is directly linked with gene sequence, which in-turn governs phenotypic traits. However, the mean resolving power (Rp) detected by SCoT (20.2) markers was much higher than in previous studies employing SRAP (9.74) and ISSR (8.64) (MISHRA et al., 2011b). The higher Rp of SCoT can be related to the amplification of both dominant and co-dominant fragments in the marker system (Poczai et al., 2013). Similarly, the PIC values, which are used as allele

Table 5. Frequency class of PCR amplified bands generated by analysing 21 coffee genotypes using SCoT assay

Primers	Frequency class (%)							Total
	0.1–4.99	5.0–9.99	10.0–29.99	30.0–49.99	50.0–69.99	70.0–89.99	90.0–100	
SCoT 1	1	2	43	28	22	18	21	135
SCoT 2	4	6	12	56	63	86	41	268
SCoT 3	2	6	21	28	60	18	62	197
SCoT 4	1	6	44	35	64	67	80	297
SCoT 5	3	2	27	31	114	34	40	251
SCoT 6	3	2	6	15	25	117	83	251
SCoT 7	4	0	27	14	23	18	61	147
SCoT 8	1	6	61	51	25	33	0	177
SCoT 9	2	4	34	25	12	0	84	161
SCoT 10	0	2	10	0	11	0	42	65
SCoT 11	2	4	0	0	0	17	225	248
SCoT 13	0	6	11	20	36	51	142	266
SCoT 14	1	0	0	10	0	17	210	238
SCoT 15	1	2	3	0	0	18	147	171
SCoT 16	0	8	11	33	37	18	141	248
SCoT 17	3	8	31	35	75	48	80	280
SCoT 18	2	0	17	10	13	15	208	265
SCoT 19	4	6	28	9	23	93	104	267
SCoT 20	5	0	19	16	14	48	105	207
SCoT 21	3	4	45	34	40	51	20	197
SCoT 22	2	0	31	32	12	52	41	170
SCoT 28	2	4	41	14	39	0	186	286
SCoT 30	0	4	0	9	13	35	189	250
SCoT 31	3	8	21	7	13	66	80	198
SCoT 34	3	4	28	0	12	34	63	144
SCoT 36	1	2	7	39	28	66	39	182
SCoT 39	1	8	22	33	58	16	0	138
SCoT 48	1	0	21	33	24	18	143	240
SCoT 61	2	4	22	19	64	17	80	208
SCoT 62	4	10	29	26	39	16	20	144
SCoT 63	1	0	36	0	50	48	145	280
Total	62	118	708	662	1009	1135	2882	6576

diversity and frequency among different coffee species, varied from 0.163 to 0.864 with a mean of 0.625 and was lower than the mean PIC value obtained using ISSR (0.81) and SRAP (0.81) marker analysis (Mishra et al., 2011b). The PIC value obtained in our study was much higher than the PIC value (0.33) obtained in *Cocos nucifera* L. using SCoT marker analysis (Rajesh et al., 2015).

The clustering pattern of 21 coffee genotypes in UPGMA dendrogram is largely in congruence with the results of PCoA analysis. In both dendrogram and PCoA plot, the Indian wild coffee species formed a separate group (I). African wild coffee species were arranged in II and III groups. In contrast, the cultivated species *C. arabica* and *C. canephora* were placed along with *C. eugenioides* and *C. congensis* in group IV (Fig. 4). Both *C. canephora* and *C. eugenioides* were considered as progenitors of *C. arabica*, and similarly, *C. congensis* is close to *C. canephora* and used in the breeding programme (Mishra, 2019). Hence it is quite apparent that all the three species are grouped together. The results indicated that SCoT markers could be efficiently used to identify polymorphism between coffee species.

In the present study, 31 SCoT primers amplified 62 unique fragments in 15 species, with a maximum number of 10 unique fragments generated by *C. zanguebariae* and followed by *C. liberica* and *C. bengalensis*. However, no unique fragments were generated in *C. arabica*, *C. congensis*, C. *steno-*

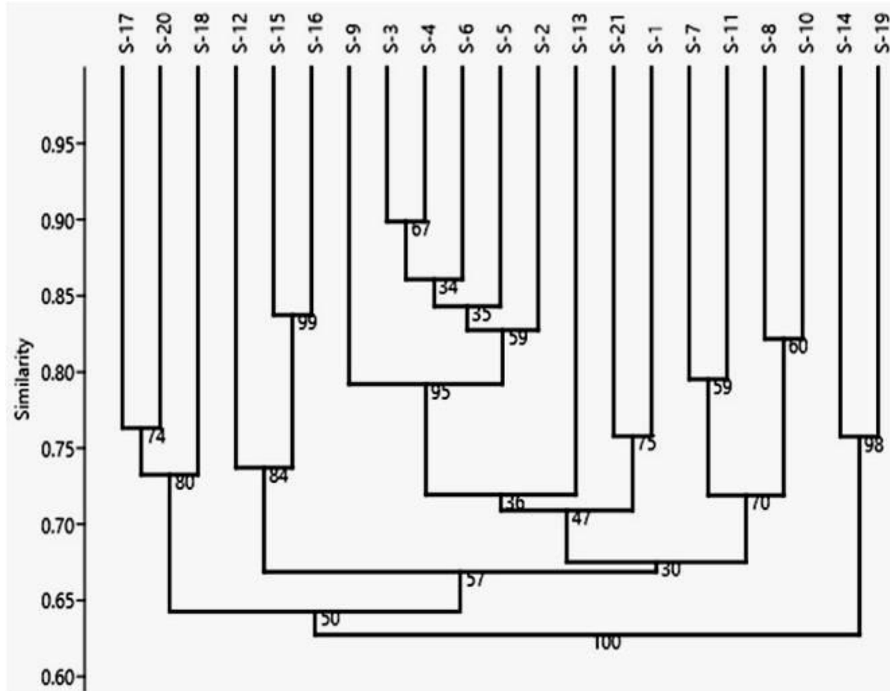

Fig. 3. Dendrogram generated using the unweighted pair group method with arithmetic average analysis (UPGMA) showing relationships among different coffee species using SCoT data. The numbers at the nodes indicate the bootstrap value. Coffee species listed in Table 1

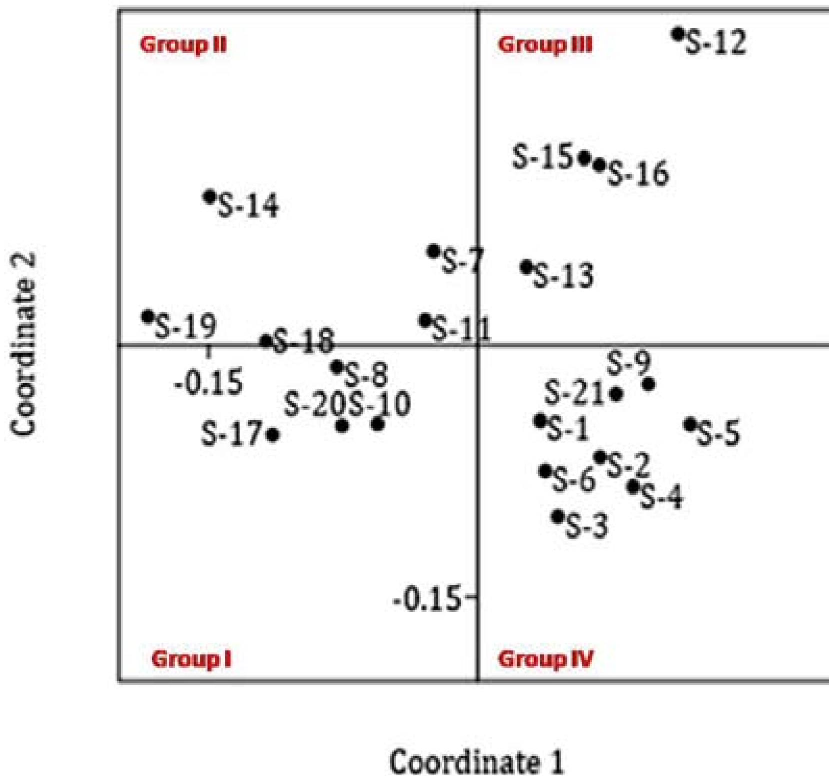

Fig. 4. Principal Coordinate Analysis (PCoA) of SCoT marker data obtained by evaluation of coffee species. Coffee species listed in Table 1

phylla and *C. jenkinsii*, which could be attributed to the insufficient number of SCoT primers used. These unique fragments could be used as diagnostic fingerprinting tools to discriminate species. A close look at the fingerprinting pattern generated by different SCoT primers revealed that SCoT 20, SCoT 19, SCoT 7, SCoT 62 and SCoT 2 are more informative markers as they amplified the maximum number of unique fragments across the species (Table 5). Since single accession per species was used in the present study, the study could facilitate in selecting the highly polymorphic SCoT primers suitable for genetic analysis of coffee species. In future research, more number of plants belonging to each species could be included for assessing the total genetic variability in the coffee gene pool.

In the present study, the pairwise similarity coefficient between different coffee species ranged from 0.60 to 0.89. The highest similarity of 0.89 was obtained between *C. canephora* cv. S.274 and *C. canephora* var. *ugandae*, and *C. canephora* var. *quillon* and *C. canephora* var. *ugandae* using SCoT marker. However, in a previous study, the pairwise similarity coefficient between different coffee species ranged from 0.11–0.90 using SRAP and 0.27–0.89 using ISSR markers in coffee (Mishra et al., 2011b). The highest similarity (0.90) was obtained between *C. canephora* and *C. canephora* var. *laurantii* using SRAP and (0.89) between *C. canephora* and *C. congensis* using ISSR assay. The genotypes with the highest coefficient of similarity using the SCoT, SRAP and ISSR assays had a common origin (West Tropical Africa). In the present study, the lowest genetic similarity was observed between *C. abeokutae* and *C. khasiana* based on the SCoT marker data analysis. In the previous study, the lowest genetic similarity was established between *C. bengalensis* and *C. liberica* using the SRAP marker and between C. *wightiana* and *C. congensis* using ISSR marker analysis (Mishra et al., 2011b). In this study, a narrow range of similarity coefficient was observed between five Indian species, which coincided with the previous research employing SRAP and ISSR markers (Mishra et al., 2011b). Further, Indian species, *C. bengalensis* and *C. travancorensis* were closely grouped in the UPGMA dendrogram, which has been confirmed in the previous study. However, C. *jenkinsii*, an indigenous Indian species, showed the highest similarity to *C. arabica*, originated from Ethiopia and widely cultivated in India. The maximum similarity observed between *C. jenkinsii* and *C. arabica* 'Kents' could be due to the close similarity at the functional loci revealed by SCoT marker analysis.

CONCLUSIONS

The extent of the genetic variability of coffee species at the functional level was successfully tested for the first time using SCoT markers. The markers with their high resolving power successfully evaluated the genotypes and generated a high level of polymorphism. The assay successfully differentiated coffee species based on functional diversity and geographical origin. The species-specific amplicons produced by the SCoT marker assay could be used as fingerprints of species identification. Since the SCoT assay targets functional domains, the sequence information obtained by sequencing differential fragments generated across the species can be used in selecting the right parental material in the coffee improvement and conservation programmes.

ACKNOWLEDGEMENTS

The authors thank the Director of Research, Central Coffee Research Institute, India for providing the laboratory facilities and encouragement. Funding support from Coffee Board, Govt. of India is gratefully acknowledged.

AUTHOR´S CONTRIBUTION

The corresponding author designed and collected the material for the experiment and wrote the manuscript. AKH and PJ carried out the experiments and participated in manuscript preparation. All authors read and approved the manuscript.

REFERENCES

Amirmoradi B., Talebi R., Karami E., 2012: Comparison of genetic variation and differentiation among annual *Cicer* species using start codon targeted (SCoT) polymorphisms, DAMD-PCR, and ISSR markers. – Plant Systematics and Evolution, 298: 1679–1688.

Andersen J., Lubberstedt T., 2003: Functional markers in plants. – Trends in Plant Science, 8: 554–560.

Anthony F.O., Diniz L.E.C., Combes M., Lashermes P., 2010: Adaptive radiation in *Coffea* subgenus *Coffea* L. (Rubiaceae) in Africa and Madagascar. – Plant Systematics and Evolution, 285: 51–64.

Bhawna Abdin M.Z., Arya L., Verma M., 2017: Use of SCoT markers to assess the gene flow and population structure among two different populations of bottle gourd. – Plant Gene, 9: 80–86.

Botstein D., White R.L., Skolnick M., Davis R.W., 1980: Construction of a genetic linkage map in man using restricted length polymorphism. – American Journal of Human Genetics, 32(3): 314–331.

Brar D.S., 2005: Broadening the gene pool and exploiting heterosis in cultivated rice. – In: Toriyama K., Heong K.L., Hardy B. (eds), Rice is life: scientific perspectives for the 21st century proceedings of the world rice research conference: 157–160 – Tokyo.

Collard B.C.Y., Mackill D.J., 2009: Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. – Plant Molecular Biology Reporter, 27: 86–93.

Davis A.P., 2011: *Psilanthus mannii*, the type species of *Psilanthus*, transferred to *Coffea*. – Nordic Journal of Botany, 29: 471–472.

Davis A.P., Chester M., Maurin O., Fay M.F., 2007: Searching for the relatives of *Coffea* (Rubiaceae, Ixoroideae): The circumscription and phylogeny of *Coffeae* based on plastid sequence data and morphology. – American Journal of Botany, 94: 313–329.

Davis A.P., Govaerts R.H.A., Bridson D.M., Stoffelen P., 2006: An annotated taxonomic conspectus of genus *Coffea* (Rubiaceae). – Botanical Journal of the Linnean Society, 152(4): 465–512.

Davis A.P., Rakotonasolo F., De Block P., 2010: *Coffeatoshii* sp. nov. (Rubiaceae) from Madagascar. – Nordic Journal of Botany, 28(2): 134–136.

Davis A.P., Tosh J., Ruch N., Fay M.F., 2011: Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. – Botanical Journal of the Linnean Society, 167: 357–377.

Guerreiro F.O., Silvarolla M.B., Eskes A.B., 1999: Expression and mode of inheritance of resistance in coffee to leaf miner *Perileucoptera coffeella*. – Euphytica, 105: 7–15.

Guo D.L., Zhang J.Y., Liu C., 2012: Genetic diversity in some grape varieties revealed by SCoT analysis. – Molecular Biology Reports, 39(5): 5307–5313.

Hajibarat Z., Saidi A., Hajibarat Z., Talebi R., 2015: Characterization of genetic diversity in chickpea using SSR markers, start codon targeted polymorphism (SCoT) and conserved DNA-derived polymorphism (CDDP). – Physiology and Molecular Biology of Plants, 21: 365–373. DOI 10.1007/s12298-015-0306-2

Hajjar R., Hodgkin T., 2007: The use of wild relatives in crop improvement: A survey of developments over the last 20 years. – Euphytica, 156(1): 1–13. doi: 10.1007/s10681-007-9363-0.

Hamidi H., Talebi R., Keshavarzi F., 2014: Comparative efficiency of functional gene-based markers, start codon targeted polymorphism (SCoT) and conserved DNA-derived polymorphism (CDDP) with ISSR markers for diagnostic fingerprinting in wheat (*Triticum aestivum* L.). – Cereal Research Communications, 42: 558–567.

Luo C., He X.H., Chen H., Ou S.J., Gao M.P., Brownj S., Tondo C.T., Schnell R.J., 2011: Genetic diversity of mango cultivars estimated using SCoT and ISSR markers. – Biochemical Systematics and Ecology, 39(4–6): 676–684.

Luo F., Fu Y.Y., Xiang Y., Yan S., Hu G., Huang X., Huang G., Sun C., Li X., Chen K., 2014: Identification and quantification of gallotannins in mango (*Mangifera indica* L.) kernel and peel and their antiproliferative activities. – Journal of Functional Foods, 8: 282–291.

Maurin O., Davis A.P., Chesterm M., Mvungi E., Jaufeerally-Fakim Y.J., Fay M.F., 2007: Towards a Phylogeny for *Coffea* (Rubiaceae): Identifying well-supported lineages based on nuclear and plastid DNA sequences. – Annals of Botany, 100(7): 1565–1583.

Milbourne D., Meyer R., Bradshaw J.E., Baird E., Bonar N., Provan J., Powell W., Waugh R., 1997: Comparison of PCR based marker systems

for the analysis of genetic relationships in cultivated potato. – Molecular Breeding, 3: 127–136.

Mishra M.K., 2019: Genetic resources and breeding of coffee (*Coffea* spp.). – In: Jameel M., Al-Khayri S., Jain S.M., Johnson D.V. (eds), Advances in Plant Breeding Strategies: Nut and Industrial Crops, 4: 475–515. – Switzerland.

Mishra M.K., Slater A., 2012: Recent advances in the genetic transformation of coffee. Biotechnology Research International, 1–17. https://doi.org/10.1155/2012/580857

Mishra M.K., Nishani S., Jayarama, 2011a: Genetic relationship among indigenous coffee species from India using RAPD, ISSR, and SRAP marker analysis. – Biharean Biologist, 5(1): 17–24.

Mishra M.K., Nishani S., Jayarama, 2011b: Molecular identification and genetic relationship among coffee species inferred from ISSR and SRAP marker analysis. – Archives of Biological Sciences, 63: 667–679.

Mishra M.K., Awati M., Anand C.G., Kumar A., 2018: Molecular and physiological characterization of a natural interspecific coffee hybrid. – Indian Journal of Plant Physiology, 23: 810–821. https://doi.org/10.1007/s40502-018-0410-8

Nowak M.D., Davis A.P., Yoder A.D., 2012: Sequence data from new plastid and nuclear COSII regions resolves early diverging lineages in *Coffea* (Rubiaceae). – Systematic Botany, 37(4): 995–1005.

Online Source: International Coffee Organization (ICO, 2020) http://www.ico.org/Market-Report-19-20-e.asp, Available at: http://www.ico.org/prices/po-production.pdf.

Peakall R., Smouse P.E., 2012: GenAlEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research-an update. Bioinformatics, 28(19): 2537–2539. doi:10.1093/bioinformatics/bts460.

Poczai P., Varga I., Laos M., Cseh A., Bell N., Valkonen J.P.T., Hyvonen J., 2013: Advances in plant gene-targeted and functional markers: A review. – Plant Methods, 9: 6. https://doi.org/10.1186/1746-4811-9-6

Prevost A., Wilkinson M.J., 1999: A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. – Theoretical and Applied Genetics, 98: 107–112.

Que Y., Pan Y., Lu Y., Yang C., Yang Y., Huang N., Xu L., 2014: Genetic analysis of diversity within a Chinese local sugarcane germplasm based on start codon targeted polymorphism. – BioMed Research International, 1–10. https://doi.org/10.1155/2014/468375.

Rahmani M., Pijut P.M., Shabanian N., Nasri M., 2015: Genetic fidelity assessment of in vitro-regenerated plants of *Albizia julibrissin* using SCoT and IRAP fingerprinting. – In Vitro Cellular & Developmental Biology – Plant, 51(4): 407–419.

Rajesh M.K., Sabana A.A., Rachana K.E., Rahman S., Jerard B.A., Karun A., 2015: Genetic relationship and diversity among coconut (*Cocos nucifera* L.) accessions revealed through SCoT analysis. – 3 Biotech, 5(6): 999–1006.

Razafinarivo N.J., Guyot R., Davis A.P., Couturon E., Hamon S., Crouzillat D., Rigoreau M., Tranchant C.D., Poncet V., De Kochko A.D.E., Rakotomalala J.J., Hamon P., 2013: Genetic structure and diversity of coffee (*Coffea*) across Africa and the Indian Ocean islands revealed using microsatellites. – Annals of Botany, 111(2): 229–248.

Rohlf F.J., 1995: NTSYS-pc: numerical taxonomy and multivariate analysis system version 2.10e. Exeter software, New York.

Tesfaye K., Borsch T., Govers K., Bekele E., 2007: Characterization of *Coffea* chloroplast microsatellites and evidence for the recent divergence of *C. arabica* and *C. eugenioides* chloroplast genomes. – Genome, 50: 1112–1129.

Xiong F.Q., Tang R.H., Chen Z.L., Pan L.H., Zhuang W.J., 2009: SCoT: a novel gene-targeted marker technique based on the translation start codon. – Molecular Plant Breeding, 7: 635–638.

MOLEKULINIŲ SCOT ŽYMENŲ TINKAMUMO *COFFEE* RŪŠIŲ GENETINEI ANALIZEI ĮVERTINIMAS

Manoj Kumar Mishra, Arun Kumar C. Huded, Pavankumar Jingade

Santrauka

Pastaruoju metu genetinės įvairovės tyrimuose vis plačiau naudojamas naujas startinio genų kodono (SCoT) gentinių žymenų metodas. Šiame darbe buvo išbandytas 31 SCoT genetinio pradmens efektyvumas, identifikuojant 18 kavos rūšių 21 genotipą. Buvo paruošti 647 genetiniai fragmentai, vidutiniškai 20,9 vienam pradmeniui, iš kurių 80,8 % buvo polimorfiški. Nustatytas SCoT pradmenų polimorfiškumas kito nuo 0,16 iki 0,86, o vidutinė vertė buvo 0,62. Skiriamoji pradmenų geba svyravo nuo 6,19 iki 28,29, vidutinė vertė – 20,2. Šešiolikoje kavos rūšių genotipų buvo nustatyti specifiniai DNR fragmentai, kurie gali būti naudojami kaip genetiniai „pirštų atspaudai“ rūšių identifikavimui. Kavos rūšių genetinis panašumas apskaičiuotas, naudojant Dice panašumo koeficientą, buvo 0,60–0,89. Dendrograma sudaryta remiantis porų grupavimo pagal aritmetinius vidurkius su vienodais svoriais metodu (UPGMA) parodė, kad visi kavos genotipai buvo suskirstyti į dvi pagrindines grupes. Statistinė tirtų genotipų vizualizacija parodė, kad Indijos vietinė rūšis *Coffea jenkensii* genetiškai artima etiopinės kilmės *C. arabica*. Gauti rezultatai patvirtino molekulinių SCoT žymenų metodo tinkamumą tolesnei kavos rūšių genetinei analizei.