TÜBİTAK

Turkish Journal of Botany

http://journals.tubitak.gov.tr/botany/

Research Article

Turk J Bot
(2022) 46: 109-122
© TÜBİTAK
doi:10.3906/bot-2109-27

DNA barcoding analysis and phylogenetic relationships of Indian wild coffee species

Manoj Kumar MISHRA[1]*, Pavankumar JINGADE[2], Arun Kumar C. HUDED[3]
Plant Biotechnology Division, Unit of Central Coffee Research Institute,
Coffee Board, Manasagangothri, Mysore, Karnataka, India

Received: 14.06.2021 • **Accepted/Published Online:** 19.01.2022 • **Final Version:** 29.03.2022

Abstract: Wild coffee species are the reservoirs of genetic diversity and could play a critical role in the genetic improvement of coffee. However, the conservation and genetic assessment of wild coffee species has been largely neglected. In the present study, DNA barcoding approaches were employed to assess the phylogenetic relationships between five indigenous wild coffee species from India and compared with two cultivated, and four wild coffee species of African origin. The efficacy of three barcoding loci namely *matK*, *rbcL*, and *trnL-trnF* was investigated using PCR amplification and sequence characterization. The intergenic spacer trnL-trnF is the highly polymorphic loci followed by matK and rbcL chloroplast gene. Among the three barcoding loci, the matK locus has the maximum number of parsimony informative sites, whereas the trnL-trnF locus contains maximum singleton variable sites. Although all the three loci contain a few unique fixed nucleotides (UFNs), no individual barcode locus has the critical nucleotide sequence tags for all the five Indian wild coffee species that help in species discrimination. However, the multilocus combinations are efficient in discriminating the species due to the presence of SNPs and specific sequence tags. The phylogenetic tree constructed using the maximum likelihood analysis of the combined barcoding loci separated all the Indian wild coffee species from African wild coffee species compared to phylogeny inferred using individual barcoding loci. Our study supports the utility of DNA barcoding as a useful tool for coffee species identification, which can be used for conservation purposes.

Key words: Indian wild coffee species, DNA barcoding, matK, rbcL, *trnl*-trnF, phylogenetics

1. Introduction

Coffee is one of the most popular beverages consumed by people all over the world. It is cultivated in 10.2 million ha of land, spanning over 80 countries in tropical and subtropical regions, mostly in Africa, Asia, and Latin America (Mishra and Slater, 2012). Although the genus *Coffea* contains more than 125 species (Davis, 2011), commercial coffee cultivation centres around only two species, *Coffea arabica* L. (known as arabica coffee) and C. *canephora* Pierre ex A.Froehner (known as robusta coffee) for beverage production (Mishra, 2019). Another coffee species, *C. liberica* Hiern (Liberian coffee) is also cultivated on a small scale, but its commercial cultivation is insignificant compared to the other two. The annual global production of coffee during 2020 exceeded 1,05,21,000 tons 175.35 million bags (60 kg capacity) comprising 63,15,000 tons 105.25 million bags (60.02%) of arabica and 42,06,000 tons 70.1 million bags (39.97%) of robusta coffee (ICO, 2021). In 2019, the total world coffee export value was estimated to be USD 30.1 billion/year in the international market. Despite its prodigious economic importance, the pace of genetic improvement of coffee is abysmally slow. The genetic improvement and sustainability of any crop plant species is directly associated with the collection, characterization, and rational exploitation of its available germplasm (Langridge and Waugh, 2019; Nguyen and Norton, 2020). Therefore, enriching the diversity from both indigenous and exotic sources assumes considerable significance (Wang et al., 2017; Migicovsky et al., 2019) for crop improvement programs. In the case of coffee, the primary gene pool comprising numerous wild diploid species could serve as the reservoirs of genetic diversity.

During the last few years, many wild coffee species were discovered from the rainforest of tropical Africa, Indian Ocean islands, and Australasia (Davis et al., 2019). However, to date, only a few species such as Coffea liberica, Coffea congensis A.Froehner, Coffea racemosa Lour., and Coffea eugenioides S.Moore have been used, and sparingly so, in the genetic improvement of either C. arabica or C. canephora (Mishra et al., 2020). The two diploid and tetraploid hybrids that have been successfully generated using the conventional breeding programs so far include the C×R hybrid involving C. canephora and C. congensis in India and C. arabica × C. racemosa

* Correspondence: manojmishra.m@gmail.com

from Brazil, of which C×R is extremely popular and commercially exploited. Despite the importance of diploid coffee species as the potential source of genetic variation to counter specific pests and diseases, alleviate abiotic stress, breed low caffeine coffee, and organoleptic quality improvement, there is very little information available on the diverse characteristic features of various coffee species. Considering the accelerated climatic change, accumulating the comprehensive information and characterization of diploid coffee species is of paramount importance since approximately 60% of wild coffee species are on the verge of extinction (Davis et al., 2019).

In India, several indigenous wild coffee species were reported by various researchers (Narasimhaswamy and Vishweswara, 1963; Sivarajan et al., 1992). Among these, three species Coffea bengalensis Roxb. ex Schult., Coffea travancorensis Wight & Arn. and Coffea wightiana Wall. ex Wight & Arn. have either traces or nil caffeine content and could therefore be important in the breeding programs aimed at developing commercial coffee varieties with lower caffeine content. Although minor differences in leaf morphology exist between these three species, (Jingade et al., 2021) they are not sufficient to distinguish the species easily and authentically. Furthermore, detailed information on the ecology, distribution, and population dynamics of indigenous coffee species are not available as some of the species are either vulnerable or endangered due to the destruction of their natural habitats. Until now, these species have been characterized using different molecular techniques (including RAPD, ISSR, and SRAP), their genome size, ploidy level, and stomatal analyses (Mishra et al., 2011; Jingade et al., 2021). In recent years, DNA barcoding techniques have become the standard for rapid identification and discrimination of plants at the genus and species levels (Girma et al., 2016; Ismail et al., 2020; Tanaka and Ito, 2020). The principle of DNA barcoding is based on the analysis of the distinct sequence variations in short genomic regions that are universally present in target lineages and that vary in the level of variation they present. The matK and rbcL loci are proposed to be the standard plant DNA barcoding markers by the CBOL Plant Working Group of the Consortium for the Barcoding of Life due to their low cost, universality, better sequence quality, and high discriminatory power among angiosperm especially at the family level (CBOL Plant Working Group, 2009; Hollingsworth et al., 2009). Compared to matK, rbcL has a higher PCR amplification success, but lower discriminatory power due to its higher conservation among plant lineages. Along with coding chloroplast genes, noncoding regions have also been successfully used as barcode markers to develop a barcode database for phylogenetic assessment and conservation strategies (Taberlet et al., 1991). A range of loci such as matK, rbcL, and trnL-trnF have been extensively used as DNA barcodes in different plant species, including some species of Coffea (Maurin et al.,2007; Anthony et al., 2010; Kaya et al.,2018; Unsal et al., 2019). Guyeux et al. (2019) carried out the chloroplast genome annotation in 16 species of the Coffea and Psilanthus genera including C. bengalensis to reconstruct the ancestral chloroplast genomes and to evaluate their phylogenetic relationships. However, comprehensive DNA barcoding data for Indian indigenous coffee species are not yet available.

Keeping the above facts in view, the present study was undertaken with the following objectives: (1) investigate whether the matK, rbcL, and trnL-trnF regions of the chloroplast DNA can be used as barcodes to distinguish Indian indigenous coffee species; (2) assess the phylogenetic relationships between the Indian species of Coffea and other cultivated and wild coffee species of different geographical regions of the world. The results obtained may also contribute to the conservation status of the species.

2. Materials and methods

2.1. Plant material

The plant material is comprised of eleven coffee species which include two cultivated (C. canephora Pierre ex A.Froehner, cv. S. 274 and C. arabica L. 'Kents'), four exotic coffee species of African origin (C. eugenioides S.Moore, C. liberica Hiern, C. racemosa Lour., C. stenophylla G.Don), and five Indian indigenous wild coffee species (C. bengalensis Roxb. ex Schult., C. travancorensis Wight &Arn. C. wightiana Wall. ex Wight &Arn. C. jenkinsii Hook. f. C. khasiana (Korth.) Hook.f.) (Figure 1 and Table 1). The Indian wild coffee species were collected from a single population and maintained in germplasm blocks at Central Coffee Research Institute, India (Table1). The exotic coffee species of African origin were collected through FAO and USDA expeditions and were maintained in germplasm blocks. Both the Indian cultivars C. canephora cv. S. 274 and C. arabica L. 'Kents' are commercially cultivated in India. Three representative samples from each species were analysed.

2.2. DNA extraction, PCR amplification, and sequencing

The fresh young and tender leaves from three representative samples of all the eleven coffee species were collected and the total genomic DNA was isolated, implementing a modified CTAB protocol as described earlier (Mishra et al., 2020). The quantity and quality of extracted DNA were assessed using nanodrop (Thermo Fischer Scientific) and by running 0.8% agarose gel electrophoresis. The DNA was diluted to a working concentration of 10 ng/µL and stored at a –20 °C freezer for further use.

PCR amplification of three plastid DNA barcode loci such as matK, rbcL, and trnL-trnF were carried out using

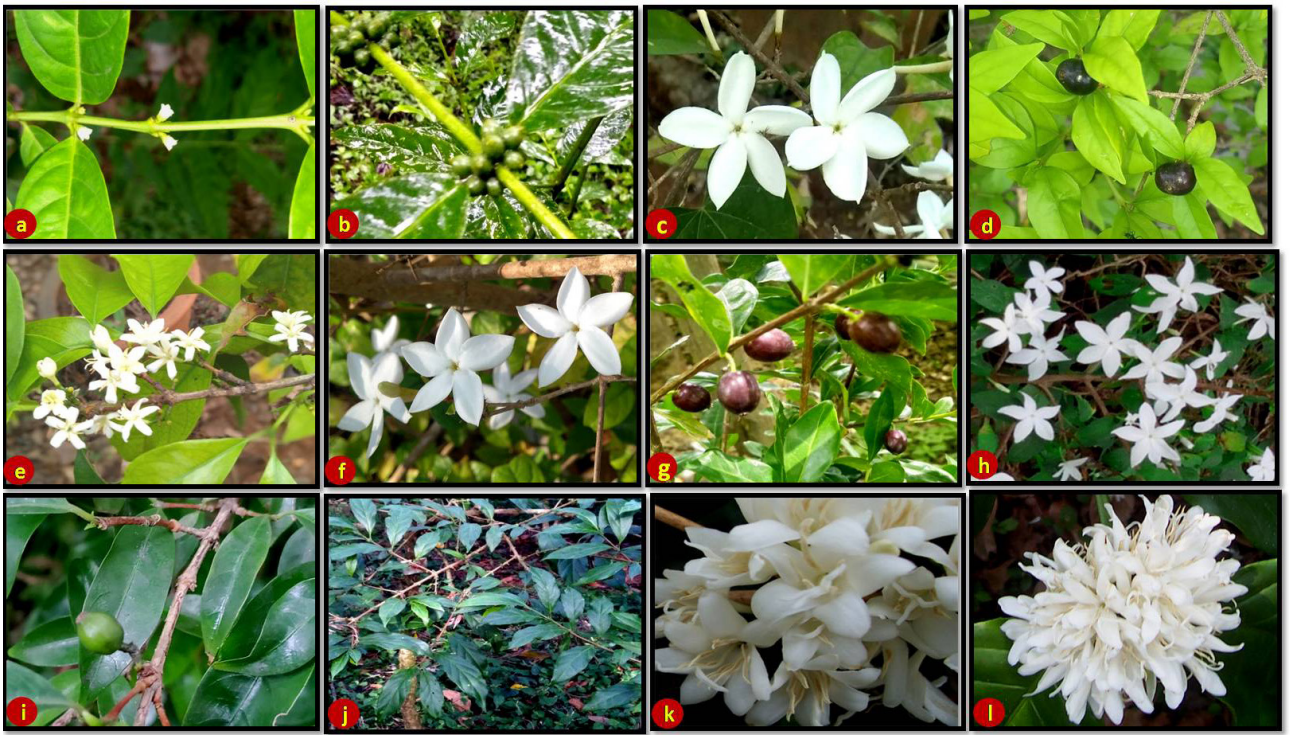

Figure 1. Picture of plant material used in the study a: C. jenkinsii flower b: C. jenkinsii fruits c: C. travancorensis flower d: C. travancorensis fruits e: C. stenophylla flower f: C. wightiana flower g: C. racemosa fruits h: C. bengalensis flower i: C. eugenioides fruit j: C. khasiana k: C. arabica flower k: C. canephora flower

respective universal primers (Table 2). PCR reactions were performed in a total 20 µL reaction volume consisting of 2.0 µL of 10x PCR buffer, 2.0 µL of 2 mM dNTP, 2.0 µL of 25 mM MgCl2, 2.0 µL each of forward and reverse primer (3 µM) specific to each gene (Table 2), 3.0 µL of 10 ng/µL high-quality template DNA, 1.0 µL of 1 U/µL Taq DNA polymerase and 6.0 µL of nuclease-free water. All the PCR components were obtained from Thermo Fisher Scientific. In all the eleven coffee species, two chloroplast regions such as rbcL and trnL-trnF were successfully amplified and produced a single band, which resulted in high-quality sequences (Figure S1). However, amplification of the matK region using the universal primer (Cuenoud et al., 2002) was only partially successful as few indigenous and exotic wild coffee species were not amplified (Figure 2). This problem was resolved by designing the newly developed matK gene-specific primers (Table 2).

The optimized PCR conditions employed for amplification of matK, rbcL, and trnL-trnF comprised of an initial denaturation of 94 °C for 5 min followed by 35 cycles consisting of denaturation at 94 °C for 45 sec, annealing for rbcL and matK at 52 °C and trnL-trnF at 50 °C for 45 s and extension at 72 °C for 1 min. After 35 cycles, all the samples were kept for a final extension at 72 °C for 10 min. The PCR products were resolved on 1.3% (w/v) agarose gel containing 0.05 µg/mL ethidium bromide using 1X TBE running buffer. A 1-kb bp plus DNA ladder was used as a molecular size standard to determine the size of the amplified products. The amplified products were visualized under UV and documented using the Gel Doc system (BioRad) (Figure S1).

The PCR amplified bands were extracted and cleaned using a QIAquick PCR purification kit procured from QIAGEN (Hilden, Germany), following the manufacturer's protocol. The PCR amplified and cleaned products were directly sequenced by the Sanger sequencing method using respective forward and reverse primer pairs in a 96-capillary 3730xl sequencer (Applied Biosystems) at Eurofins Scientific India Pvt., Ltd. Bangalore, India.

2.3. Sequence alignment and analysis

The sequence quality was assessed using Sequence Scanner Software 2 (Applied Biosystems). The quality sequences of three barcoding loci obtained from three samples belonging to each species were manually curated and aligned using the contig assembly program (CAP) of BioEdit 7.0 software. The respective barcoding sequences of matK, rbcL and trnL-trnF obtained from 11 species were assembled and aligned using the default parameters of the ClustalW algorithm of MEGA X software version 10.1.8. The assembled sequences were submitted using the

Table 1. Details of the plant material and their place of origin and distribution with NCBI accession Id's of barcoding loci.

Sl. No.	Plant Material	Details	Place of origin/distribution	Accession numbers of submitted sequences		
				matK	rbcL	trnL-trnF
1	C. canephora Pierre ex A.Froehner, cv. S. 274	Cultivar	West central Africa Congo (Indian cultivar)	MZ668308	MZ668319	MZ668330
2	C. arabica L. 'Kents'	Cultivar	Ethiopia (Indian cultivar)	MZ668309	MZ668320	MZ668331
3	C. eugenioides S.Moore	Wild species	East Africa	MZ668317	MZ668328	MZ668339
4	C. liberica Hiern	Wild species	West tropical Africa	MZ668314	MZ668325	MZ668336
5	C. racemosa Lour.	Wild species	Southern tropical Africa	MZ668313	MZ668324	MZ668335
6	C. stenophylla G.Don	Wild species	West tropical Africa	MZ668312	MZ668323	MZ668334
7	C. bengalensis Roxb.exSchult.	Wild species	Cooch Behar (West Bengal), India	MZ668318	MZ668329	MZ668340
8	C. travancorensis Wight &Arn.	Wild species	Nilgiri (Tamil Nadu), India	MZ668311	MZ668322	MZ668333
9	C. wightiana Wall. ex Wight &Arn.	Wild species	Travancore (Kerala), India	MZ668310	MZ668321	MZ668332
10	C. jenkinsii Hook.f.	Wild species	Khasi Hill (Meghalaya), India	MZ668316	MZ668327	MZ668338
11	C. khasiana (Korth.) Hook.f.	Wild species	Jaintia Hill (Meghalaya), India	MZ668315	MZ668326	MZ668337

Table 2. Details of the primer used for barcoding coffee species.

Sl. no.	Primer name	Primer sequence (5' - 3')	Forward/ reverse	Target gene/region	Reference
1	390f	CGATCTATTCATTCAATATTTC	Forward	Segment of Maturase K (matK)	Cuenoud et al. 2002 (Old matK primer)
2	1326r	TCTAGCACACGAAAGTCGAAGT	Reverse		
3	CBL MatK-F	CGATCAATTCATTCACCATTTC	Forward	Segment of Maturase K (matK)	New primers designed in this study (New matK primer)
4	CBL MatK-R	TTTAGCACAAGAAAGTCGAAGT	Reverse		
5	rbcLa-F	ATGTCACCACAAACAGAGACTAAAGC	Forward	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene	Fazekas et al., 2012
6	rbcLajf634-R	GAAACGGTCTCTCCAACGCAT	Reverse		
7	B49317	CGAAATCGGTAGACGCTACG	Forward	Non coding region of chloroplast DNA	Taberlet et al., 1991
8	A50272	ATTTGAACTGGTGACACGAG	Reverse		

BankIt sequence submission portal of the National Centre for Biotechnology Information (NCBI).

In the present study, the complete chloroplast genome sequence of Coffea arabica (NC_008535.1) available in NCBI was used as reference sequence to identify the DNA sequence polymorphism in all three barcoding loci of 11 coffee species. For each sequence, the percentage of polymorphic sites was calculated as (number of variable nucleotides/length of the entire region) × 100. The types of mutations that determined the polymorphisms were also identified and are presented in Table 3. The number of insertions events in the sequence is quantified as the number of variable sites whose polymorphism is signalized by the addition of one or more nucleotides and similarly, the number of deletions represented the number of variable sites whose polymorphism was created due to a deletion of one or more nucleotides. The number of transitions in the sequences is identified by the number of variable sites whose polymorphism was caused by an interchange between two purines (A and G) or two pyrimidines (C and T) and finally, the number of transversions comprised the number of variable sites whose polymorphism was because of the substitution of a purine into a pyrimidine (Table 4). The number of mutation events in the sequence is quantified by combining the number of both different and multiple polymorphisms detected in all samples (that

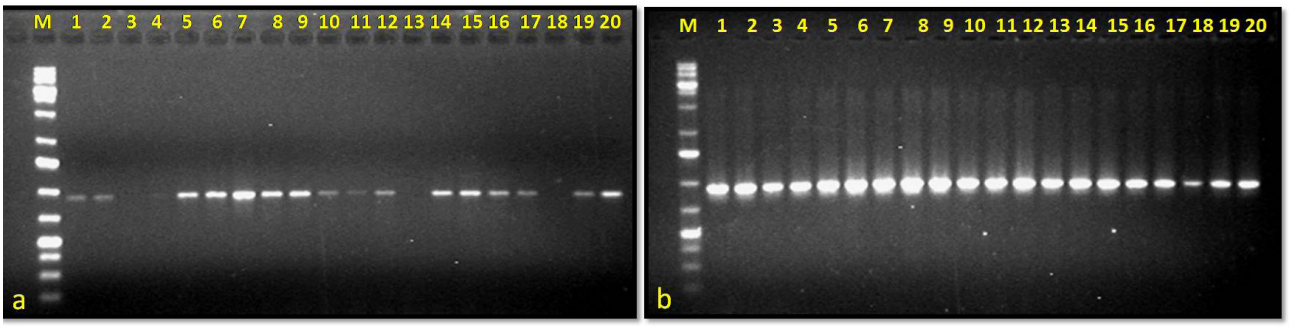

Figure 2. Gel pictures showing the amplification of matK barcoding primer from 20 coffee species. a: Amplification pattern of old matK primer b: Amplification pattern of new matK primer. Lane L: 1kb plus GeneRuler from Thermo Fischer Scientific, Lane 1: C. abekutae, Lane 2: C. arnoldiana, Lane 3: C. bengalensis, Lane 4: C. eugenioides, Lane 5: C. excelsa, Lane 6: C. jenkinsii, Lane 7: C. kappakatta, Lane 8: C. khasiana, Lane 9: C. liberica, Lane 10: C. racemosa, Lane 11: C. stenophylla, Lane 12: C. travancorosis, Lane 13: C. wightiana, Lane 14: C. zangebariae, Lane 15: C. salvatrix, Lane 16: C. congensis, Lane 17: C. dewevrei, Lane 18: C. canephora cv. Peridenia Lane 19: C. canephora cv. Uganda Lane 20: C. canephora cv. Quillon

Table 3. Details of singleton variable sites, parsimony informative sites and nucleotide diversity (Pi) of three barcoding gene sequence of different coffee species.

	matK	rbcL	trnL-trnF
Total aligned length (bp)	875	594	888
Number of monomorphic sites	790	548	769
Number of polymorphic sites	77	46	79
Total number of InDels sites	8	0	31
Overlapping Indels sites	0	0	9
Number of singleton variable sites	48	29	57
Total number of mutations, Eta	78	49	85
Parsimony informative sites, PIC	29	17	22
Nucleotide diversity, Pi	0.022	0.021	0.021
Mean nucleotide difference (k)	12.74	12.60	17.82
Number of Haplotypes, h	8	8	8
Haplotype diversity, Hd	0.925	0.926	0.925
Variance of Haplotype diversity	0.0011	0.0013	0.0009
Sequence conservation, C	0.910	0.923	0.906
Conservation threshold (CT)	1.0	1.0	1.0
Divergence time (T)	**6.73**	**6.73**	**6.73**

Table 4. Maximum likelihood estimate of substitution matrix of three bar-coding regions estimated using Tamura–Nei model.

rbcL gene					**matK gene**					**trnL- trnF gene**				
	A	**T/U**	**C**	**G**		**A**	**T/U**	**C**	**G**		**A**	**T/U**	**C**	**G**
A	-	7.71	5.88	12.11	A	-	9.59	4.76	5.85	A	-	7.80	4.70	6.70
T/U	7.68	-	7.80	6.45	T/U	7.62	-	10.29	3.93	T/U	9.28	-	9.96	4.17
C	7.68	10.23	-	6.45	C	7.62	20.72	-	3.93	C	9.28	16.52	-	4.17
G	14.40	7.71	5.88	-	G	11.35	9.59	4.76	-	G	14.91	7.80	4.70	-

is the sum of variable sites and the number of different mutations which took place on the same nucleotide site in different samples).

The aligned sequences of three chloroplast regions were characterized for DNA polymorphism using DnaSP software (Librado and Rozas, 2009; Rozas, 2009). For each sequence region, several discriminative parameters were estimated, including the number of monomorphic or polymorphic sites, parsimony informative sites, nucleotide diversity (π), haplotype diversity (Hd), number of mutations, sequence conservation, and conservation (DeSalle et al., 2005; Pettengill and Neel, 2010; Hosein et al., 2017) (Table 3).

The phylogenetic tree was constructed using the sequences of individual barcoding loci (rbcL matK and trnL-trnF), as well as the concatenating the sequence data of all three barcoding loci. In addition to the 11 coffee species used in the present study, coffee species for which sequence information for all the three barcoding loci was available in the NCBI database along with five outgroup species such as Solanum tuberosum, Solanum melongena, Lycopersicon esculentum, and Nicotiana tabacum belong to Solanaceae and Camellia sinensis belonging to Theaceae were downloaded. The concatenated data was utilised to construct the dendrogram by using the maximum likelihood (ML) method based on the Tamura–Nei model (Tamura and Nei, 1993) with 1000 bootstrap replications in MEGA X Version 10.1.8. Software developed by Pennsylvania State University, USA. The evolutionary relationship of Indian wild coffee species was ascertained by constructing a phylogenetic tree using 138 plastid DNA sequences of various other coffee species comprising 76 sequences of trnL-trnF, 44 sequences of matK, and 18 sequences of rbcL retrieved from the NCBI GenBank.

3. Results

3.1.Amplification and nucleotide sequencing of barcoding loci

The three barcoding loci specific primers resulted in robust amplification in the 11 species. The PCR amplified bands were sequenced and the sequence homology was determined using the Basic Local Alignment Search Tool (BLAST; https: //blast.ncbi.nlm.nih.gov/Blast.cgi) of NCBI. No sequence variation in the barcoding loci was observed among the three individuals belonging to the same species. All the 33 nucleotide sequences of three chloroplast loci were deposited to a public depository (NCBI GenBank) and individual accession numbers were obtained (Table 1).

3.2. Variation in aligned sequences

Except for the rbcL locus, the sequence length of matK and trnL-trnF DNA barcode regions varied considerably among different coffee species. For example, the length of trnL-trnF ranged from 859 bp to 871 bp and matK ranged from 870 bp to 873 bp in different species (data not shown). The nucleotide base composition of three barcode loci was computed in different coffee species (Table 5). The average G+C content was 44.5% for rbcL, 33.6% for matK, and 34.2% for trnL-trnF loci (Table 5). The trnL-trnF region has the most variable sequence with 8.9% recorded variability followed by matK (8.8%) and the least sequence variability (7.7%) was recorded in the rbcL region (Table 3). Among the three chloroplast loci examined, all of

Table 5. Base composition matrix of three bar-coding genes in coffee species.

Species	rcbL						MatK						trnL-trnF					
	T(U)	C	A	G	C+G	Total	T(U)	C	A	G	C+G	Total	T(U)	C	A	G	C+G	Total
C. canephora cv. S.274	27.8	21.0	27.9	23.2	44.3	594	36.7	18.7	29.0	15.7	34.4	873.0	29.7	18.1	36.1	16.1	34.2	866.0
C. arabica cv. S795	27.8	21.0	27.9	23.2	44.3	594	37.2	18.1	29.6	15.1	33.2	873.0	30.2	17.9	35.9	16.0	33.9	864.0
C. eugenioides	27.9	21.0	27.8	23.2	44.3	594	36.9	18.5	29.3	15.3	33.8	870.0	29.8	17.3	35.6	17.3	34.6	866.0
C. liberica	27.9	21.2	27.6	23.2	44.4	594	37.4	18.3	29.3	15.1	33.3	870.0	29.9	18.3	35.8	16.0	34.2	865.0
C. racemosa	27.9	21.2	27.8	23.1	44.3	594	37.7	18.4	29.0	14.9	33.3	873.0	30.1	17.9	35.8	16.2	34.1	864.0
C. stenophylla	27.9	21.4	27.8	22.9	44.3	594	37.1	18.5	29.4	14.9	33.4	870.0	30.0	18.0	35.9	16.1	34.1	863.0
C. bengalensis	27.9	21.0	27.6	23.4	44.4	594	36.9	18.5	29.5	15.1	33.6	870.0	30.4	18.3	35.6	15.7	34.0	871.0
C. travancorensis	27.8	21.5	26.8	23.9	45.5	594	36.9	18.4	29.5	15.2	33.6	870.0	30.1	18.2	36.0	15.7	33.9	870.0
C. wightiana	27.9	21.0	27.6	23.4	44.4	594	37.0	18.2	29.8	15.0	33.2	873.0	30.1	18.2	35.8	15.9	34.1	867.0
C. jenkinsii	27.8	21.4	28.1	22.7	44.1	594	36.9	18.4	29.6	15.1	33.6	873.0	30.6	18.6	35.3	15.5	34.1	864.0
C. khasiana	27.3	21.4	27.6	23.7	45.1	594	36.8	18.2	29.6	15.5	33.7	873.0	29.8	18.5	35.5	16.2	34.7	859.0
Average	27.8	21.2	27.7	23.3	44.5	594	37.0	18.4	29.4	15.2	33.6	871.6	30.1	18.1	35.7	16.1	34.2	865.4

Table 6. Parsimony informative sites (PIC) and singleton variable sites.

Gene	Parameter	Sites	Positions	Total	Grand total
rbcL	PIC	Two variants	15, 20, 23, 44, 65, 82, 88, 110, 202, 257, 283, 399, 418, 423 and 555	15	**17**
		Three variants	66 and 256	2	
	Singleton variable sites	Two variants	43, 49, 57, 67, 69, 81, 91, 101, 141, 149, 150, 152, 156, 185, 188, 189, 192, 227, 246, 265, 276, 281, 393, 424, 426, 441, 468 and 492	28	**29**
		Three variants	425	1	
matK	PIC	Two variants	89, 196, 215, 382, 581, 626, 629, 636, 653, 657, 658, 667, 670,707, 712, 713, 719, 724, 725, 727, 737, 788, 802, 803, 818, 835, 858, 859 and 869	29	**29**
	Singleton variable sites	Two variants	22, 36, 37, 39, 59, 116, 164, 187, 211, 235, 268, 275, 305, 344, 431, 474, 536, 537, 549, 553, 559, 560, 575, 593, 611, 623, 632, 637, 642, 649, 669, 674, 688, 691, 710, 716, 744, 752, 753, 754, 757, 764, 805, 809, 824, 847 and 866	47	**48**
		Three variants	620	1	
trnL-trnF	PIC	Two variants	46, 52, 116, 254, 274, 302, 422, 436, 533, 638, 744, 808, 855, 857, 870, 873, 875, 877	18	**22**
		Three variants	854, 871, 872 and 878	4	
	Singleton variable sites	Two variants	2, 9, 28, 31, 33, 38, 57, 58, 65, 73, 127, 143, 151, 235, 275, 297, 322, 347, 567, 587, 615, 630, 659, 718, 771, 777, 785, 794, 803, 815, 817, 818, 823, 824, 826, 827, 828, 830, 836, 840, 844, 845, 849, 850, 851, 852, 853, 856, 858, 861, 862, 863, 864, 868 and 879	55	**57**
		Three variants	859 and 876	2	

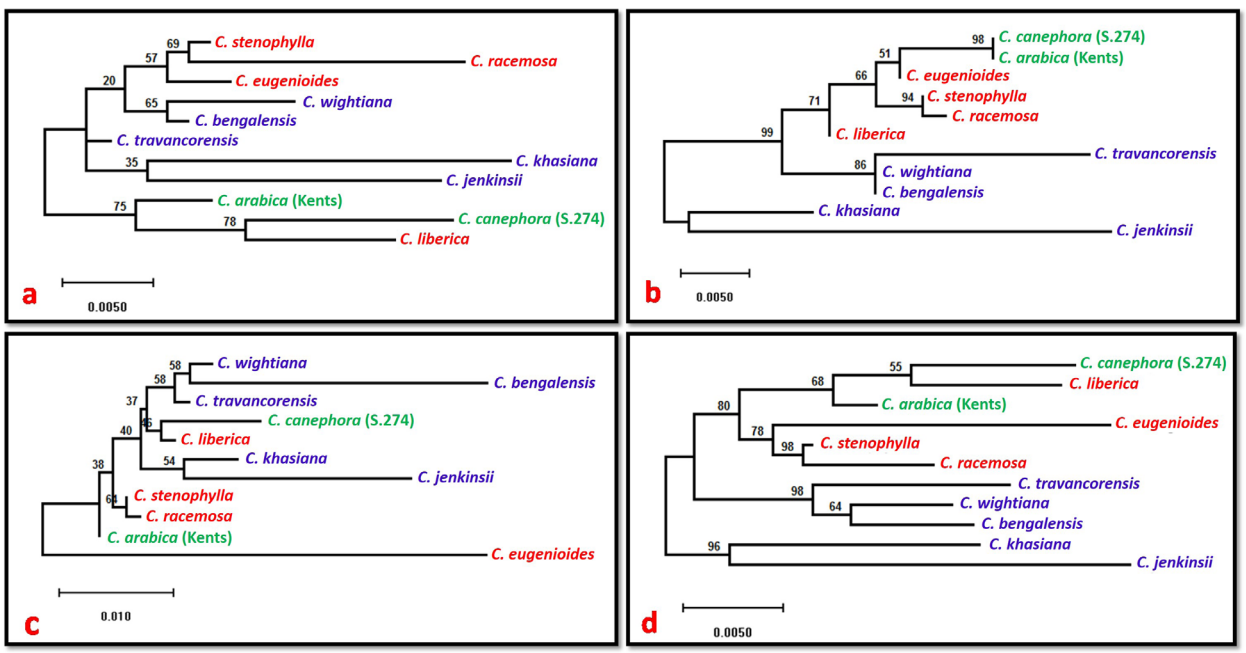

Figure 3. Phylogenetic tree constructed using sequence information obtained by sequencing barcode regions from 11 coffee species by Maximum Likelihood approach using Tamura-Nei model. The Indian wild species mentioned in purple font, cultivated species in green font and Excotic species from African origin mentioned in red font. a: Phylogenetic tree constructed using matK sequences, b: Phylogenetic tree constructed by using rbcL sequences. c: Phylogenetic tree constructed using trnL-trnF sequences. d: Combined phylogenetic tree constructed using matK, rbcL and trnL-trnlF sequences.

Table 7. Details obtained by analysing matK, rbcL and trnL- trnFbar-coding regions from coffee species.

Gene	Parameters	Coffea species										
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI
matK	SNP (bp)	27	8	3	21	16	4	5	5	10	15	24
	SNP (%)	3.10	0.91	0.34	2.40	1.83	0.46	0.57	0.57	1.14	1.71	2.74
	Transition sites	11	6	2	11	5	1	2	3	3	7	13
	Transition (%)	1.26	0.69	0.23	1.26	0.57	0.11	0.23	0.34	0.34	0.80	1.49
	Transversion sites	16	2	1	10	11	3	3	2	7	8	11
	Transversion (%)	1.83	0.23	0.11	1.14	1.26	0.34	0.34	0.23	0.80	0.91	1.26
	Insertion (bp)	1	-	-	-	-	-	-	-	-	-	-
	Deletion (bp)	1	-	3	3	-	3	3	3	-	-	-
	InDel (%)	0.23	-	0.34	0.34	-	0.34	0.34	0.34	-	-	-
rbcL	SNP (bp)	4	4	-	3	4	3	9	19	9	28	16
	SNP (%)	0.67	0.67	-	0.50	0.67	0.50	1.51	3.20	1.52	4.71	2.70
	Transition sites	1	1	-	2	1	1	5	9	5	15	7
	Transition (%)	0.17	0.17	-	0.33	0.17	0.17	0.84	1.52	0.84	2.53	1.18
	Transversion sites	3	3	-	1	3	2	4	10	4	13	9
	Transversion (%)	0.50	0.50	-	0.17	0.50	0.33	0.67	1.68	0.68	2.18	1.52
	Insertion (bp)	-	-	-	-	-	-	-	-	-	-	-
	Deletion (bp)	-	-	-	-	-	-	-	-	-	-	-
	InDel (%)	-	-	-	-	-	-	-	-	-	-	-
trnL-trnF	SNP (bp)	12	-	34	5	3	2	26	11	17	21	9
	SNP (%)	1.35	-	3.23	0.56	0.34	0.22	2.93	1.24	1.91	2.36	1.01
	Transition sites	4	-	12	4	2	1	12	2	8	13	7
	Transition (%)	0.45	-	1.35	0.45	0.22	0.11	1.35	0.22	0.90	1.46	0.79
	Transversion sites	8	-	22	1	1	1	14	9	9	8	2
	Transversion (%)	0.90	-	2.48	0.11	0.11	0.11	1.58	1.01	1.01	0.90	0.22
	Insertion (bp)	2	-	6	2	-	-	4	-	3	5	5
	Deletion (bp)	6	6	8	7	6	6	3	-	6	11	12
	InDel (%)	0.90	0.67	1.58	1.01	0.67	0.67	0.79	-	1.01	1.80	1.91

Note: I: C. canephora; II: C. arabica; III: C. eugenioides; IV: C. liberica; V: C. racemosa; VI: C. stenophylla; VII: C. bengalensis; VIII: C. travancorensis; IX: C. wightiana; X: C. jenkinsii; XI: C. khasiana.

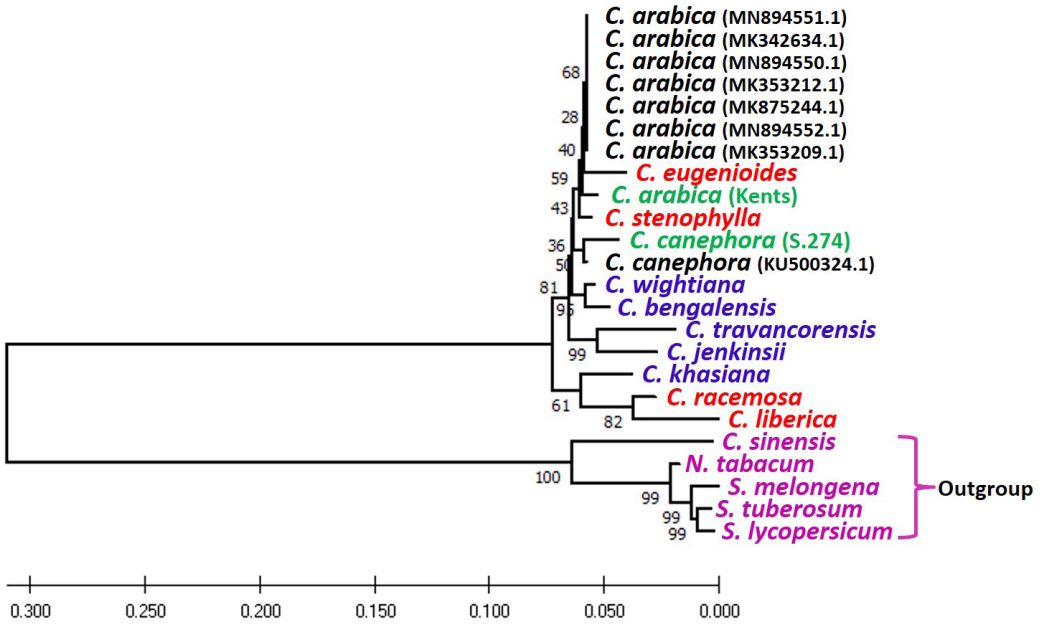

Figure 4. Maximum Likelihood phylogenitic tree constructed using concatenated sequence of matK, rbcL and trnL-trnF barcoding loci of different coffee species from current study and the available sequences in NCBI GenBank along with out-group species using Tamura-Nei model. The numbers in the nodal region indicates bootstrap values. The Indian wild species mentioned in purple font, cultivated species in green font and Excotic species from African origin mentioned in red font.

them have similar conservation thresholds (Table 3). The parsimony informative sites for rbcL, trnL-trnF, and matK regions were recorded as 17, 22 and 29 respectively (Table 6). Similarly, the singleton variable sites identified in rbcL, matK, and trnL-trnF loci were 29, 48, and 57, respectively (Table 6). All the three loci, trnL-trnF, rbcL, and matK, have 8 haplotypes and some of the haplotypes are shared among the species.

Among all the eleven species examined, C. eugenioides showed maximum single nucleotide substitution in trnL-trnF (34), C. jenkinsii in rbcL (28), and C. canephora in matK (27) loci (Table 7). In three barcoding loci, transversion events cause more sequence substitutions than transition events (Table 4). It was observed that among the substitution types, A to G and T to C are more frequent compared to G to A and C to T types in all the three loci (Table 4). In the present study, the comparative sequence analysis of individual species revealed a variable number of indels at both matK and trnL-trnF chloroplast loci (Table 7). In matK, a maximum number of indels in the form of deletion were recorded in C. eugenioides, liberica, C. stenophylla, C. bengalensis, and C. travancorensis whereas in the trnL-trnF region maximum numbers of indels are observed in C. khasiana (Korth.) Hook.f. followed by C. jenkinsii. No indel is observed in rbcL loci.

3.3. Tree-based analysis using single and multilocus combinations

The sequences obtained from rbcL, matK, and trnL-trnF loci were used individually as well as in combination to infer trees using the maximum likelihood (ML) method (Figure 3). It was observed that among the three loci, the tree constructed using rbcL (Figure 3b) distinctly separated the Indian wild coffee species from other species compared to trees constructed using matK and trnL-trnF (Figure 3). Among the five indigenous wild coffee species, three Indian coffee species, C. bengalensis, C. wightiana, and C. travancorensis are clustered together while C. khasiana and C. jenkinsii are closely placed together and form a separate group (Figure 3). The tree constructed using combined analysis of all the three loci could further enhance the resolution of Indian coffee species compared to the resolution obtained by individual loci (Figure 3d).

3.4. Phylogenetic analyses using database sequences

The phylogenetic tree constructed by the concatenated sequences of three barcoding loci separated all the coffee species from the outgroup represented by S. tuberosum, S. melongena, L. esculentum, N. tabacum, and C. sinensis. Further, the Indian wild coffee species are clearly distinguished from other wild and cultivated coffee

species. All the coffee spices revealed a common ancestral origin (Figure 4)

The phylogenetic tree constructed using the 138 plastid DNA sequences of three barcoding loci revealed close clustering of Indian wild coffee spices using both rbcL and trnL-trnF barcoding loci (Figure S2).

3.5. Potential barcode regions for species identification

The aligned sequences of three chloroplast loci (matK, rbcL, and trnL-trnF) were evaluated and compared with the NCBI database to find out the unique sequences that could be used as a potential barcode to discriminate Indian wild coffee species from other coffee species. All three barcode loci matK, trnL-trnF and rbcL have species-specific sequences in C. jenkinsii, C. khasiana, C. travancorensis, C. wightiana, and C. bengalensis that could be used as potential barcodes to identify species (Table S1).

4. Discussion

Crop wild relatives (CWR) constitute an enormous reservoir of genetic variation that is crucial to the success of crop improvement programs. These wild crop species occur in a wide range of habitats which are currently at significant risk because of habitat degradation and climate change. It is estimated that worldwide approximately 60% of wild coffee species are threatened and on the verge of extinction. In this context, there is an urgent need to conserve wild coffee species both in the wild (in situ) and in genebanks (ex situ) to ensure that genetic diversity remains available for future utilizations. In general, morphological descriptors are widely utilised for the characterization of coffee species. Molecular markers are more efficient, precise, and more reliable than morphological and biochemical markers in discriminating closely related species and cultivars (Mishra et al., 2012). Further, it has been demonstrated that certain regions of the chloroplast genome have the potential to be used as DNA barcodes across a broad range of plant species (CBOL, 2009; Attigala et al., 2014; Kress et al., 2015; Li et al., 2015; Wu et al., 2019). Hence, the present research was undertaken to evaluate five Indian wild coffee species along with six wild and commercial coffee species using three DNA barcoding markers to find out their evolutionary and phylogenetic relationships.

4.1. Sequence variability and species discrimination

In the present study, three DNA barcode loci viz, rbcL, matK and trnL-trnF of chloroplast region have been evaluated to find out their utility in wild coffee species identification and phylogenetic information. Among the three candidate regions tested in this study, clear and robust amplification was obtained for both rbcL and trnL-trnF using universal primers whereas amplification of matK region was partially successful using universal primers already available (Cuenoud et al., 2002). In C. bengalensis, C. wightiana, C. eugenioides, and C. canephora cv.S.274, amplification could not be obtained using the universal primers even after repeated attempts. The new matK primers designed in the present study resulted in robust PCR amplification in 20 coffee species analysed. In the past, many authors have reported the problem of PCR amplification and sequencing of matK locus using universal primers (Du et al., 2011; Hollingsworth et al., 2011; Yan et al., 2011) and the research carried out in the present study agrees with the earlier reports. PCR primers of matK designed in the current study showed robust amplification among different coffee species. Other studies have reported high PCR amplification success with rbcL and trnL-trnF loci using universal primers (Wu et al., 2019; Jiang et al., 2020) and the present study further supports this contention.

Even though a large amount of data is currently available on angiosperm DNA barcoding, no specific and universal barcodes are available to date that would always provide a precise species identification. Very often, a barcode suitable for one group of plants is found inadequate for another group of plants, especially the recently diverged species (von Crautlein et al., 2011). However, many studies have reported that among different barcoding loci, matK and rbcL either singly or in combinations are efficient barcodes across different plant species (CBOL, 2009; Hollingsworth et al., 2009; de Melo Moura et al., 2019). Li et al. (2015) suggested that as a single barcode locus matK is effective in discriminating more than 90% of species in the Orchidaceae but less than 49% species in the nutmeg family. Many authors also advocated the utility of rbcL as an efficient DNA barcode locus due to its shorter length of about 500 bp, robust universal primer, high PCR success, and high-quality sequencing (von Crautlein et al., 2011; Wu et al., 2019). Moreover, many authors have suggested that any valid plant barcode should be multilocus, preferably comprising of conserved coding regions like rbcL and matK in combination with more rapidly evolving noncoding regions, such as trnL intron and trnL-trnF intergenic spacer of the chloroplast DNA (Kress et al., 2009; Mahadani and Ghosh, 2014; Kang, 2021). The effectiveness of any barcoding loci depends on the extent of sequence variability among the species or terminals analysed. In the present study, among the three barcoding loci, trnL-trnF and the matK have more variable sequences than rbcL. Further, the number of parsimony informative sites and singleton variable sites also varies considerably among the three loci of which trnL-trnF presented the highest numbers, followed by matK. No insertion/deletion (indels) events were recorded in the rbcL locus of any coffee species examined in the present study, whereas the variable length of indels was documented at different positions of trnL-trnF and matK. Earlier studies have

documented higher sequence polymorphism in trnL-trnF and matK loci compared to rbcL (Hao et al., 2009; Skuza et al., 2018; Meena et al., 2020) and the present findings are in agreement with such reports. In the present study, the GC content of rbcL locus is higher than that of both matK and trnL-trnF, and this phenomenon has been reported in other tropical plant species (de Melo Moura et al., 2019).

In our study, all candidate barcode markers rbcL, matK, and trnL-trnF fit the DNA barcoding criteria described by Hollingsworth et al. (2011), in terms of sequence variability that would make them appropriate to discriminate Indian wild coffee species. Previously, Mahadani and Ghosh (2014) reported that multiple residues indels can be used as potential diagnostic markers in species discrimination and phylogenetic analyses. We have observed that variability in the form of SNP or multiple residue indels in all the three barcode regions (Table 7) creates sufficient polymorphism which could be exploited for Indian wild coffee species discrimination. The species C. jenkinsii and C. khasiana exhibited the maximum number of SNPs compared to other coffee species, across all three barcoding loci. The presence of unique fixed nucleotides (UFNs) was also documented in all five Indian wild coffee species and successfully validated by repeated amplifications in duplicated test samples. No individual barcoding loci have UNF for all the five Indian wild coffee species. For example, matK has as UNF for C. khasiana, C. jenkinsii, C. travancorensis and C. whightiana, rbcL has UNF for C. khasiana, C. jenkinsii, C. travancorensis, and trnL-trnF has UNF for C. khasiana, C. jenkinsii, C. bengalensis. This study highlights the significance of SNP and UFNs in the identification and discrimination of Indian wild coffee species.

4.2. Phylogenetic relationships

The phylogenetic trees were constructed using both individual and combined barcode loci following the ML method to understand the relationships between different coffee species. Incidentally, the dendrogram constructed using rbcL placed all the five Indian wild coffee species together and separated them from African coffee species. However, the individual trees constructed using matK and trnL-trnF loci could not classify the Indian wild coffee species into a separate group. Among the Indian wild coffee species, C. jenkinsii and C. khasiana are closely placed together whereas C. bengalensis, C. travancorensis, and C. whightiana formed a separate group (Figure 3). Similarly, the three wild African coffee species like C. racemosa, C. stenophylla, and C. eugenioides are clustered together and similarly, C. canephora and C. liberica are placed closely in the dendrograms using individual and combined barcode loci analysis. The placement of species in the dendrogram has not altered significantly by combining three barcode loci (Figure 3). Interestingly, several authors have suggested using a combination of different barcoding loci or whole-chloroplast genome sequences for better resolution of plant groups (CBOL, 2009; Girma et al., 2016, Gogoi et al., 2020) whereas several other authors have also documented the utility of single barcoding loci in resolving diverse plant species such as wild cherry (Unsal et al., 2019), bamboo (Dev et al., 2020) Zingiber (Saha et al., 2020), and conifers (Pham et al., 2021). Consequently, a combination of three chloroplast marker loci was analysed in the present study, both individually and combined, and the resolution obtained indicated that the combination of loci led to a higher level of discrimination between Indian wild coffee species.

In the present study, a phylogenetic tree was constructed using four species (S. tuberosum, S. melongena, L. esculentum, N. tabacum) belonging to the family Solanaceae and a beverage crop (C. sinensis) as the outgroup to understand the evolutionary relationship. Previously several reports are available on comparative genomics and evolution of coffee and members of Solanaceae (Lefebvre- Pautigny et al., 2010; Barchi et al., 2019). The phylogenetic tree distinctly separated the coffee species from the outgroup indicating a common ancestral origin of all the coffee species (Figure 4). Further, the Indian wild coffee species are clustered together and separated from other wild and cultivated species, indicating their genetic uniqueness and independent evolution (Figure 4). Recently Hamon et al. (2017) studied the genetic relationship between 85 Coffea species including some Indian wild coffee species using GBS (genotype by sequencing) and observed that Indian wild coffee species are different and represented a separate cluster and the present study supports their contention. Further, based on the whole chloroplast genome annotation of 16 coffee samples including a few Psilanthus species, Guyeux et al. (2019) suggested that C. canephora is closely related to Psilanthus species. In the present study, the tree generated using the combined barcoding loci also placed C. canephora close to Indian wild coffee species.

The relationship of Indian wild coffee species was further evaluated by comparing the sequences of three barcoding loci from other coffee species (Figure S2). Based on the comparative analysis, it was observed that all the Indian wild coffee species are closely placed together using both rbcL and trnL-trnF barcoding loci whereas, the Indian wild coffee species were interspersed along with other species using matK (Figure S2).

5. Conclusion

In conclusion, we investigated the molecular relationships between five Indian wild coffee species and compared them with wild coffee species of African origin using a DNA barcoding approach. Among the three barcoding loci used, the trnL-trnF spacer region and matK gene

have shown higher sequence variability compared to the rbcL barcoding region. All three barcoding loci efficiently discriminate the Indian wild coffee species due to the presence of unique variable sites. The maximum likelihood tree-based phylogenetic analysis using both the individual and combined barcode loci DNA sequences yielded nearly identical dendrograms and separated the Indian and African wild coffee species effectively, though the combined analysis of barcoding loci provided better resolution. Our findings will be helpful in future research programs concerning wild coffee species identification and conservation efforts by using DNA barcoding markers.

Conflict of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Contribution of authors

The corresponding author was involved in project conceptualization, execution, sample collection, writing and editing of the manuscript. PJ and AKH conducted the laboratory work and analysed the data. All the authors read and approved the manuscript.

References

Anthony F, Diniz LEC, Combes MC, Lashermes P (2010). Adaptive radiation in Coffea subgenus Coffea L. (Rubiaceae) in Africa and Madagascar. Plant Systematics and Evolution 285:51-64. doi: 10.1007/s00606-009-0255-8

Attigala L, Triplett JK, Kathriarachchi HS, Clark LG (2014). A new genus and a major temperate bamboo lineage of the Arundinarieae (Poaceae: Bambusoideae) from Sri Lanka based on a multi-locus plastid phylogeny. Phytotaxa 174: 187-205. doi: 10.11646/phytotaxa.174.4.1

Barchi L, Pietrella M, Venturini L, Minio A, Toppino L et al. (2019). A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. Scientific Reports 9:11769. doi:10.1038/s41598-019-47985-w

CBOL Plant Wording Group (2009). DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America 106:12794-12797. doi: 10.1073/pnas.0905845106

Cuenoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ et al. (2002). Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. American Journal of Botany 89: 132-144. doi: 10.3732/ajb.89.1.132

Davis AP (2011). Psilanthus mannii, the type species of Psilanthus, transferred to Coffea. Nordic Journal of Botany 29: 471-472. doi:10.1111/j.1756-1051.2011.01113.x

Davis AP, Chadburn H, Moat J (2019). High extinction risk for wild coffee species and implications for coffee sector sustainability. Science Advances 5: eaav3473.doi: 10.1126/sciadv.aav3473

de Melo Moura CC, Brambach F, Jair Bado KJH, Krutovsky KV, Kreft H et al. (2019). Integrating DNA barcoding and traditional taxonomy for the identification of Dipterocarps in Remnant Lowland Forests of Sumatra. Plants 8 (11):461. doi:10.3390/plants8110461

DeSalle R, Egan MG, Mark S (2005). The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society B 360:1905-1916. doi:10.1098/rstb.2005.1722

Dev SA, Sijimol K, Prathibha PS, Sreekumar VB, Muralidharan EM (2020). DNA barcoding as a valuable molecular tool for the certification of planting materials in bamboo. 3 Biotech 10 (2):59. doi: 10.1007/s13205-019-2018-8

Du ZY, Qimike A, Yang CF, Chen JM, Wang QF (2011). Testing four barcoding markers for species identification of Potamogetonaceae. Journal of Systematics and Evolution 49:246-251.doi: 10.1111/j.1759-6831.2011.00131.x

Girma G, Spillane C, Gedil M (2016). DNA barcoding of the main cultivated yams and selected wild species in the genus Dioscorea. Journal of Systematics and Evolution 54: 228-237. doi: 10.1111/jse.12183

Gogoi B, Wann SB, Saikia SP (2020). DNA barcodes for delineating Clerodendrum species of North East India. Scientific Reports 10: 13490. doi:10.1038/s41598-020-70405-3

Guyeux C, Charr J-C, Tran HTM, Furtado A, Henry RJ (2019). Evaluation of chloroplast genome annotation tools and application to analysis of the evolution of coffee species. PLoS ONE 14 (6): e0216347.doi:10.1371/journal.pone.0216347

Hamon P, Grover CE, Davis AP, Rakotomalala JJ, Raharimalala NE et al. (2017). Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (Coffea) and insights into the evolution of caffeine content in its species: GBS coffee phylogeny and the evolution of caffeine content. Molecular Phylogenetics and Evolution 109:351-361. doi: 10.1016/j.ympev.2017.02.009

Hao DC, Huang BL, Chen SL, Mu J (2009). Evolution of the chloroplast trnL-trnF region in the gymnosperm Lineages Taxaceae and Cephalotaxaceae. Biochemical Genetics 47: 351-369. doi: 10.1007/s10528-009-9233-7

Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S et al. (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America 106:12794-12797. 10.1073/pnas.0905845106.

Hollingsworth PM, Graham SW, Little DP (2011). Choosing and using a plant DNA barcode. PLoS ONE 6. doi: 10.1371/journal.pone.0019254

Hosein FN, Austin N, Maharaj S, Johnson W, Rostant L et al. (2017). Utility of DNA barcoding to identify rare endemic vascular plant species in Trinidad. Ecology and Evolution 7: 7311-7333. doi:10.1002/ece3.3220

International Coffee Organization (ICO) Report (2021). Coffee market ends 2019/20 in surplus. https://icocoffeeorg.tumblr.com/post/178884744500/coffee-market-ends-201718-in-surplus. [accessed 11 August 2021].

Ismail M, Ahmad A, Nadeem M, Javed MA, Khan SH et al. (2020). Development of DNA barcodes for selected Acacia species by using rbcL and matK DNA markers, Saudi Journal of Biological Sciences 27 (12):3735-3742. doi: 10.1016/j.sjbs.2020.08.020

Jiang KW, Zhang R, Zhang ZF, Pan B, Tian B (2020). DNA barcoding and molecular phylogeny of *Dumasia* (Fabaceae: Phaseoleae) reveals a cryptic lineage. Plant Diversity 42(5):376-385. doi: 10.1016/j.pld.2020.07.007

Jingade P, Huded AKC, Mishra MK (2021). First report on genome size and ploidy determination of five indigenous coffee species using flow cytometry and stomatal analysis. Brazilian Journal of Botany 44:381-389. doi: 10.1007/s40415-021-00714-y

Kang Y (2021). Molecular identification of Aquilaria species with distribution records in China using DNA barcode technology, Mitochondrial DNA Part B 6 (4):1525-1535.doi: 10.1080/23802359.2021.1914210

Kaya E, Vatansever R, Filiz E (2018). Assessment of the genetic relationship of Turkish olives (Olea europaea subsp. europaea) cultivars based on cpDNA trnL-F regions. Acta Botanica Croatica 77 (1): 88-92.doi: 110.1515/botcro-2017-0019

Kress WJ, Erickson DL, Jones FA (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proceedings of the National Academy of Sciences of the United States of America 106 (44):18621-6. doi: 10.1073/pnas.0909820106

Kress WJ, García-Robledo C, Uriarte M, Erickson DL (2015). DNA barcodes for ecology, evolution, and conservation, Trends in Ecology & Evolution 30 (1):25-35. doi:10.1016/j.tree.2014.10.008

Langridge P, Waugh R (2019). Harnessing the potential of germplasm collections. Nature Genetics 51:200-201. doi: 10.1038/s41588-018-0340-4

Lefebvre-Pautigny F, Wu F, Philippot M, Rigoreau M, Priyono, et al. (2010). High resolution synteny maps allowing direct comparisons between the coffee and tomato genomes. Tree Genetics & Genomes. 6. 10.1007/s11295-010-0272-3.

Li X, Yang Y, Henry RJ, Rossetto M, Wang Y et al. (2015). Plant DNA barcoding: from gene to genome: Plant identification using DNA barcodes. Biological Reviews 90:157-166. doi: 10.1111/brv.12104

Librado P, Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25 (11):1451-1452. doi: 10.1093/bioinformatics/btp187

Mahadani P, Ghosh SK (2014). Utility of indels for species-level identification of a biologically complex plant group: a study with intergenic spacer in *Citrus*. Molecular Biology Reporter 41:7217-7222.doi:10.1007/s11033-014-3606-7

Maurin O, Davis AP, Chester M, Mvungi EF, Jaufeerally-Fakim Y et al. (2007). Towards a Phylogeny for Coffea (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences. Annals of Botany 100:1565-1583. doi: 10.1093/aob/mcm257

Meena RK, Negi N, Uniyal N, Shamoon A, Bhandari MS et al. (2020). Chloroplast-based DNA barcode analysis indicates high discriminatory potential of matK locus in Himalayan temperate bamboos. 3 Biotech 10:534. doi:10.1007/s13205-020-02508-7

Migicovsky Z, Warschefsky E, Klein LL, Miller AJ (2019). Using living germplasm collections to characterize, improve, and conserve woody perennials. Crop Science 59: 2365-2380. doi: 10.2135/cropsci2019.05.0353

Mishra MK, Slater A (2012). Recent advances in the genetic transformation of coffee. Biotechnology Research International. doi:10.1155/2012/580857

Mishra MK, Huded AKC, Jingade P (2020). Assessment of the suitability of molecular SCoT markers for genetic analysis of coffee species. Botanica 26 (2):184-196. doi: 10.2478/botlit-2020-0019

Mishra MK, Nishani S, Suresh N, Satheesh SK, Soumya PR et al. (2012). Genetic diversity among Indian coffee cultivars determined via molecular markers. Journal of Crop Improvement 26:1-24. doi: 10.1080/15427528.2012.696085

Mishra MK, Padmajyothi D, Surya Prakash N, Sreenivasan MS, Srinivasan CS (2011). Variability in stomatal features and leaf venation pattern in Indian coffee (Coffea arabica L.) cultivars and their functional significance. Botanica Serbica 35:111-119.

Mishra MK (2019). Genetic resources and breeding of coffee (Coffea spp.) In: Jameel M. Al-Khayri, S. Mohan Jain and Dennis V. Johnson (Editors) Advances in Plant Breeding Strategies, Vol. 4 Nut and Industrial Crops. Springer Publisher. pp. 475-515. doi: 10.1007/978-3-030-23112-5

Narasimhaswamy RL, Vishweshwara S (1963). A note on the occurrence of three species of Coffea indigenous to India. Turrialba 5:65-71.

Nguyen GN, Norton SL (2020). Genebank Phenomics: A Strategic Approach to Enhance Value and Utilization of Crop Germplasm. Plants 9 (7): 817. doi: 10.3390/plants9070817

Pettengill JB, Neel MC (2010). An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the genus Agalinis (Orobanchaceae). American Journal of Botany 97 (8):1391-406. doi: 10.3732/ajb.0900176

Pham MP, Tran VH, Vu DD, Nguyen QK, Shah SNM (2021). Phylogenetics of native conifer species in Vietnam based on two chloroplast gene regions rbcL and matK. Czech Journal of Genetics and Plant Breeding 57: 58-66.doi: 10.17221/88/2020-CJGPB

Rozas J (2009). DNA sequence polymorphism analysis using DnaSP. Methods in Molecular Biology 537:337-50. doi: 10.1007/978-1-59745-251-9_17

Saha K, Dholakia BB, Sinha RK (2020). DNA barcoding of selected Zingiberaceae species from North-East India. Journal of Plant Biochemistry and Biotechnology 29: 494-502. doi: 10.1007/s13562-020-00563-y

Sivarajan VV, Biju SD, Mathew P (1992). Revision of the genus Psilanthus Hook.f.(Rubiaceae, tribe Coffeeae), in India. Botanical Bulletin of Academia Sinica 33:209-224.

Skuza L, Szucko I, Filip E, Adamczyk A (2018). DNA barcoding in selected species and subspecies of Rye (secale) using three chloroplast loci (matK, rbcL, trnH-psbA). Notulae Botanicae Horti Agrobotanici Cluj-Napoca 47 (1):54-62. doi: 10.15835/nbha47111248

Taberlet P, Gielly L, Pautou G (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17:1105-1109. doi: 10.1007/BF00037152

Tamura K, Peterson D, Peterson N (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28:2731-2739. doi: 10.1093/molbev/msr121

Tanaka S, Ito M (2020). DNA barcoding for identification of agarwood source species using trnL-trnF and matK DNA sequences. Journal of Natural Medicine 74: 42-50. doi: 10.1007/s11418-019-01338-z

Unsal SG, Ciftci YO, Eken BU, Velioglu E, Marco GD et al. (2019). Intraspecific discrimination study of wild cherry populations from North-Western Turkey by DNA barcoding approach. Tree Genetics and Genomes 15:16. doi:10.1007/s11295-019-1323-z

von Crautlein M, Korpelainen H, Pietilainen M, Rikkinen J (2011). DNA barcoding: a tool for improved taxon identification and detection of species diversity. Biodiversity and Conservation 20: 373-380. doi: 10.1007/s10531-010-9964-0

Wang C, Hu S, Hu Z, Gardner C, Lubberstedt T (2017). Emerging Avenues for Utilization of Exotic Germplasm. Trends in Plant Science 22: 7. doi:10.1016/j.tplants.2017.04.002

Wu F, Li M, Liao B, Shi X, Xu Y (2019). DNA barcoding analysis and phylogenetic relation of mangroves in Guangdong Province, China. Forests 10 (1):56. doi:10.3390/f10010056

Yan HF, Hao G, Hu CM, Xue-Jun GE (2011). DNA Barcoding in closely related species: A case study of Primula L. sect. Proliferae Pax (Primulaceae) in China. Journal of Systematics and Evolution 49:225-236.: doi:10.1111/j.1759-6831.2011.00115.x