Development, Characterization and Cross Species Amplification of Polymorphic Microsatellite Markers from Expressed Sequence Tags of Turmeric (Curcuma longa L.)

Article *in* Molecular Biotechnology · November 2009

DOI: 10.1007/s12033-009-9222-4 · Source: PubMed

Mol Biotechnol
DOI 10.1007/s12033-009-9222-4

RESEARCH

Development, Characterization and Cross Species Amplification of Polymorphic Microsatellite Markers from Expressed Sequence Tags of Turmeric (*Curcuma longa* L.)

S. Siju · K. Dhanya · S. Syamkumar · B. Sasikumar · T. E. Sheeja · A. I. Bhat · V. A. Parthasarathy

Abstract Expressed sequence tags (ESTs) from turmeric (*Curcuma longa* L.) were used for the screening of type and frequency of Class I (hypervariable) simple sequence repeats (SSRs). A total of 231 microsatellite repeats were detected from 12,593 EST sequences of turmeric after redundancy elimination. The average density of Class I SSRs accounts to one SSR per 17.96 kb of EST. Mononucleotides were the most abundant class of microsatellite repeat in turmeric ESTs followed by trinucleotides. A robust set of 17 polymorphic EST–SSRs were developed and used for evaluating 20 turmeric accessions. The number of alleles detected ranged from 3 to 8 per loci. The developed markers were also evaluated in 13 related species of *C. longa* confirming high rate (100%) of cross species transferability. The polymorphic microsatellite markers generated from this study could be used for genetic diversity analysis and resolving the taxonomic confusion prevailing in the genus.

Keywords Cross species amplification · *Curcuma longa* L. · Expressed sequence tags · Microsatellite markers · Turmeric

S. Siju · K. Dhanya · S. Syamkumar · B. Sasikumar (✉) · T. E. Sheeja · V. A. Parthasarathy
Division of Crop Improvement and Biotechnology, Indian Institute of Spices Research, Calicut, Kerala 673 012, India
e-mail: bhaskaransasikumar@yahoo.com

A. I. Bhat
Division of Crop Protection, Indian Institute of Spices Research, Calicut, Kerala 673 012, India

Introduction

The genus *Curcuma* belonging to the family Zingiberaceae has a widespread occurrence with a global representation of more than 80 species [1]. Turmeric (*Curcuma longa* L.), a perennial rhizomatous herbaceous plant of the genus is cultivated extensively in the tropics for its rhizomes, which are widely used for flavoring food preparations as a spice since time immemorial. By virtue of its medicinal properties, it is also widely used in the Indian (Ayurvedic, Unani, Siddha) and Chinese systems of medicine. Modern biomedical research also attests the medicinal value of turmeric in a variety of ailments [2]. Use of molecular markers in characterizing *C. longa* has been very limited and confined to isozyme studies [3], RAPD/ISSR analysis of cultivars [4, 5] and genetic stability analysis of micropropagated plants by RAPD [6–8]. As taxonomic confusion is reported to be prevailing in the genus which is often difficult to be discriminated based on conventional taxonomic tools [2], DNA marker studies offer a viable adjunct to the conventional classificatory studies of the genus in addition to effective bioprospecting [9]. The development of a highly reliable and reproducible molecular marker for assessing the genetic diversity within the germplasm accessions could also help to facilitate the crop improvement programme through molecular breeding.

Microsatellites or Simple Sequence Repeats (SSRs) are one of the most widely used molecular markers in plant breeding, agricultural genetics, mapping, marker assisted selection, and genetic diversity studies [10, 11]. They are stretches of DNA consisting of tandemly arranged units of 1–6 bp in length [12, 13], characterized by the relative abundance, hypervariable, locus specific, codominant, and multiallelic nature [10, 12]. SSRs are ubiquitous in the coding and non-coding regions [14, 15] of prokaryotes and

Published online: 19 November 2009

eukaryotes. Based on the length of SSR tracts and their potential as genetic markers, microsatellites are categorized into two groups—Class I or hypervariable markers (perfect SSRs ≥20 bp) and Class II or potentially variable markers (SSRs ≥12 and ≤20 bp) [16]. The standard method of microsatellite development involves the creation of a small insert genomic library, subsequent hybridization with tandemly repeated oligonucleotides and its capture followed by sequencing of candidate clones, thus making the entire process a time consuming and laborious one [13]. With the rapid increase in the deposition of nucleotide sequences in the public databases and with the advent of bioinformatics tools, it has become a cost effective and fast approach to scan for microsatellite repeats and exploit the possibility of converting it into potential genetic markers.

One such source of nucleotide sequences are expressed sequence tag (EST) databases. ESTs are short and single pass sequence reads from mRNA (cDNA) [17], representing a snapshot of genes expressed in a given tissue and or at a given developmental stage. EST databases have been proven to be a valuable source of polymorphic SSRs (EST–SSRs or genic SSRs) in a number of plant species including barley [13], grape [18], and coffee [19]. EST–SSR markers are credited with the following advantages: easy identification by electronic sorting, presence in the gene rich regions of the genome, relative abundance, and easy transferability to related species [20]. So far no SSR markers have been reported in *Curcuma*, though a related work on the development of eight SSR markers is reported in ginger, a member of the same Zingiberaceae family [21]. The present study aims at scanning ESTs deposited in the public domain for the presence of hypervariable repeats and generating a robust set of polymorphic markers for turmeric (*C. longa* L.) and related *Curcuma* species.

Materials and Methods

EST Database and Sequence Retrieval

A total of 12,593 *C. longa* L. EST sequences were downloaded from the dbEST database hosted in GenBank (National Centre for Biotechnology Information, http://www.ncbi.nlm.nih.gov/dbEST/) using the keyword "*Curcuma longa*".

Sequence Analysis and Detection of SSRs

In a preliminary step, poly A and poly T-stretches of the ESTs corresponding to the poly A tails of eukaryotic mRNA were removed using TRIMEST program of EMBOSS suite using the parameters described by Kumpatla and Mukhopadhyay [22]. The sequences were assembled into contigs for creating a non-redundant dataset using EGassembler [23]. The identification of Class I (hypervariable) microsatellite repeats [16] in the generated non-redundant EST dataset were calculated using MSAT-COMMANDER [24]. When two SSRs were present close to each other in one EST, they were counted as individual SSRs rather than compound SSRs [25].

Designing of Primers

Primers were designed only for Class I SSR containing EST sequences by using Primer 3 [26] and also manually. The quality of the designed primers was validated using NetPrimer (http://www.premierbiosoft.com/netprimer/index.html). Primers having a score of more than 75 as evidenced by the absence of self dimer and or primer dimer and or cross dimers were selected for synthesis. A selected set of 45 primers were synthesized at Sigma Technologies, Bangalore, India. These primers were tested for functionality and polymorphisms against a panel of 20 *C. longa* accessions. Functional primers were validated for its effective transferability in 13 *Curcuma* species. The putative functions of sequences containing polymorphic EST–SSRs were detected using BLAST X [27] by comparing against the non-redundant protein database.

Plant Materials and DNA Isolation

Twenty diverse turmeric accessions collected from different geographical locations (states) of India and maintained at the Indian Institute of Spices Research (IISR) Experimental farm, Peruvannamuzhi, Calicut, Kerala were used in the present study to validate the SSR markers. High quality genomic DNA was isolated from the fresh rhizomes following the protocol of Syamkumar et al. [28]. The concentration and quality of the genomic DNA was assessed using spectrophotometer and agarose gel electrophoresis.

For cross species amplification studies, DNA was isolated from 13 *Curcuma* species (*C. amada* Roxb., *C. aromatica* Salisb., *C. aeruginosa* Roxb., *C. caesia* Roxb., *C. comosa* Roxb., *C. ecalcarata* Sivarajan & Indu., *C. haritha* Mangalay & Sabu, *C. montana* Wall., *C. malabarica* Vela. et al., *C. pseudomontana* Grah., *C. raktakanta* Mangalay & Sabu, *C. sylvatica* Val., and *C. zedoaria* Rosc).

PCR Amplification and Electrophoresis

Amplification of microsatellite repeats were carried out in a 25.0 µl reaction mixture containing 1× buffer with 1.5 mM $MgCl_2$ (Sigma, USA), 0.2 mM dNTP's (Bangalore Genei, Bangalore, India), 5.0 pmols each of primers, 50 ng genomic DNA, and 1 U *Taq* DNA polymerase (Sigma, USA). The amplification was performed in a

thermal cycler (Eppendorf, Mastercylcer Gradient S) programmed at an initial denaturation at 94°C (3 min) followed by a 35 cycle reaction profile programmed at 94°C (30 s), T_a°C (45 s, T_a varied for individual primers), 72°C (1 min), and a final extension of 72°C (10 min).

After PCR amplification, 3.0 µl of PCR products were mixed with an equal volume of loading buffer (98% formamide, 10 mM EDTA [pH 8.0], 0.05% xylene cyanol, 0.05% bromophenol blue), denatured at 94°C for 5 min, and snap-cooled using ice. Samples were resolved on 10% denaturing polyacrylamide gel (7 M Urea; 1× TBE buffer) at a constant power of 220 V for 4–6 h using the Mini-Protean system (Biorad, USA). The gels were silver stained [29]. The fragment size of the amplified products was compared against 50 bp DNA ladder (New England Bio-Labs) or 10 bp DNA ladder (Invitrogen, USA). The gel was documented using the Kodak MI Logic 200 system and allele sizes were scored using the inbuilt KODAK MI software. The variability at each locus was characterized in terms of number of alleles.

Results

Frequency and Distribution of Class I EST–SSRs

The 12,593 redundant EST sequences retrieved from NCBI represented approximately 8.3 Mb of *C. longa* transcriptome. A total of 389 Class I SSRs were detected in this redundant dataset corresponding to 1.0 SSR per 21.54 kb. Trimming of poly A and poly T tails resulted in the removal of 0.40% of this dataset. These sequences were clustered and assembled into a non-redundant dataset of 5,041 unique gene sequences (3,037 contigs and 2,004 singletons). Scanning of Class I microsatellites in this non-redundant dataset revealed a total of 231 unique SSRs. This accounts to a frequency of one SSR per 17.968 kb of *C. longa* transcriptome (Table 1).

Among the Class I microsatellite repeats detected, mononucleotides were the most abundant (39.8% of 231) followed by trinucleotides (23.81% of 231). The most abundant repeat motif was (A/T) in mononucleotides, (AG/CT) in dinucleotides, (AAG/CTT) in trinucleotides, (AAAC/GTTT) in tetranucleotides, (AAGAG/CTCTT) in pentanucleotides, and (AGCAGG/CCTGCT) in the hexanucleotides. The mononucleotide (A/T; 34.6% of 231) and the trinucleotide (AAG; 8.2% of 231) repeat motif constituted the major individual repeat motifs in turmeric ESTs. The dinucleotide (CG/GC) and trinucleotide (ACT/AGT) repeat motifs were not detected within the dataset. The distribution of individual Class I microsatellite repeats in the non-redundant dataset based on repeat motifs and number of repeat units are summarized in Table 2.

Table 1 Summary of EST–SSRs developed from the EST database of *Curcuma longa* L

Parameter	Value
Total number of ESTs searched	12593 (8366842 bp)
Total number of SSRs identified including poly A & poly T	389
Length of ESTs after poly A/T tail removal	8332776 bp
Total number of unigenes	5041 (4150834 bp)
Total number of contigs	3037 (2750907 bp)
Total number of singletons	2004 (1399927 bp)
Total number of unique Class I SSRs	231
Frequency of Class I SSRs in turmeric EST	1 per 17.968 kb
Number of sequences suitable for primer designing	163
Number of primers designed and tested	45
Functional polymorphic markers developed	17

Development of EST–SSRs

Primers could be designed only for 163 repeats (70.56% of 231). Designing of primers failed in the remaining sequences due to lack of sufficient flanking sequences or non-compatibility with the standard primer designing criteria. Forty-five of these sequences representing repeat motifs (dinucleotides to hexanucleotides) were short listed for primer designing and conversion to EST–SSRs. Out of the 45 primers (27.6% of 163) shortlisted for amplification studies, nine did not generate any amplicons. Six primers produced smearing pattern and thus were eliminated from further studies. Successful amplification indicated by the presence of prominent bands was observed for 30 loci, of which 17 revealed considerable level of polymorphism in the 20 turmeric accessions tested. The generated polymorphic markers were assigned the prefix CLEST SSR (*Curcuma longa* EST derived Simple Sequence Repeat). The number and range of the amplified alleles were determined across the tested individuals. All the primers together could detect a total of 79 alleles across 20 turmeric accessions ranging from 3 to 8 alleles per locus. Only clear prominent bands were considered for the determination of alleles. One marker (CLEST SSR-11) generated amplicons larger than the expected size deduced from the EST sequence. The details of the polymorphic EST–SSR markers and other related information are listed in Table 3.

Transferability of EST–SSR Markers

Transferability of the 17 EST–SSR markers was examined by testing for its amplification in 13 *Curcuma* species. All the 17 markers could generate prominent bands in the

Table 2 Occurrence of unique Class I SSRs in 12593 *C. longa* L. ESTs

SSR motif	Number of repeats																		Total number of repeats
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	>20	
A/T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8	72	80
G/C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7	5	12
AC/GT	-	-	-	-	-	-			1		1							0	2
AG/CT	-	-	-	-	-	-	3	5				1	1	2	1	1		2	16
AT/AT	-	-	-	-	-	-			2	2		3	1	2		1	1	2	14
AAC/GTT	-	-	-	1															1
AAG/CTT	-	-	-	9	2	1		3	3					1					19
AAT/ATT	-	-	-	4	3			1	1										9
ACC/GGT	-	-	-				1												1
ACG/CGT	-	-	-	2															2
AGC/GCT	-	-	-	5	1														6
AGG/CCT	-	-	-	7	1		1	3											12
ATC/GAT	-	-	-		1														1
CCG/CGG	-	-	-	2	1		1												4
AAAC/GTTT	-	3	2																5
AAAG/CTTT	-	3	1																4
AAAT/ATTT	-	1																	1
AAGG/CCTT	-	1		1															2
AATT/AATT	-		1																1
AGAT/ATCT	-	2			1														3
AGCC/GGCT	-	1																	1
AGGG/CCCT	-		1																1
ATCG/CGAT	-	4																	4
AAAAC/GTTTT		1																	1
AAAAT/ATTTT		1																	1
AAAGG/CCTTT	1																		1
AACAG/CTGTT	1																		1
AAGAG/CTCTT	2																		2
AATGG/CCATT		1																	1
ACCGG/CCGGT	1																		1
AGAGG/CCTCT	1																		1
AAAATG/CATTTT	1																		1
AAATCT/AGATTT	1																		1
AACGCC/GGCGTT	1																		1
AAGAGG/CCTCTT	1	2																	3
AATGGG/CCCATT	2																		2
ACCATC/GATGGT	3																		3
ACCCCC/GGGGGT	1																		1
ACCTCC/GGAGGT	1																		1
AGATCG/CGATCT	1																		1
AGCAGG/CCTGCT	2	1			1														4
AGGCGG/CCGCCT		1																	1
AGGGCG/CGCCCT	1																		1
ATCGGC/GCCGAT	1																		1
N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	77	92
NN	-	-	-	-	-	-	3	5	3	2	1	4	2	4	1	2	1	4	32

Table 2 continued

SSR motif	Number of repeats																		Total number of repeats
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	>20	
NNN	-	-	-	30	9	1	3	7	4										55
NNNN	-	15	5	1	1														22
NNNNN	6	3																	9
NNNNNN	16	4			1														21
Total repeat number	22	22	5	31	11	1	6	12	7	2	1	4	2	5	1	2	16	81	231

(-) repeat frequencies not included in the analysis

tested species indicating high transferability efficiency (100%) of the developed EST–SSR markers.

Discussion

Frequency and Distribution of Class I Microsatellites

Random sequencing of clones within the cDNA library often enables the representation of a stretch of nucleotides in one or more reads during sequencing, thereby leading to the deposition of redundant sequences in the database. Refining the original dataset reduces the data size and redundancy for frequency analysis and primer designing. Estimation of the frequency of SSRs in non-redundant ESTs accurately reflects the density of SSRs in the transcribed portions of the genome [20]. Assembled ESTs also facilitated the designing of primers in certain cases when SSR motifs were located close to the end of an individual EST.

Though frequency analysis of microsatellite repeats have been reported in different plant species [13, 19, 22, 25, 30–39, 41], it was observed that the parameters/limits for SSR detection for frequency calculations varied in all the studies. However, the limits used for SSR identification in the present study is similar to that of Gao et al. [33] with exceptions in limits (TNR = ≥7, HNR = ≥4) for retrieving SSRs ≥20 bp. Thus, the average density of non-redundant EST–SSRs in turmeric (1 SSR/17.96 kb) is almost on par with that of wheat (1 SSR/17.42 kb), higher than soybean and maize (23.80 and 28.32 kb, respectively), but lower than rice (11.81 kb).

Mononucleotides formed the major category of microsatellite repeats identified from turmeric ESTs. The relative abundance of mononucleotide repeats even after trimming of terminal poly A/T tracts clearly indicates their occurrence in the turmeric transcriptome. The abundance of mononucleotide repeats are also reported in the ESTs of *Citrus* [35, 36], *Medicago truncatula* [37], and *Hevea brasiliensis* [38]. The second most abundant repeat in turmeric ESTs were the trinucleotides, which is reported to be the most frequent repeats in a majority of different plant ESTs [13, 30–34, 39], due to the negative selection against frameshift mutations in the coding regions [40]. Dinucleotides constituted the third most abundant microsatellite repeat in turmeric ESTs, which is also reported to be the major repeats in ESTs of certain plants [19, 22, 41].

When compared with the earlier reports [13, 19, 22, 30–34, 39, 41], the relatively low level of trinucleotide and dinucleotide representations observed in the present study could be due to the application of stringent limits used for SSR identification and variation in the quantity of sequence data analyzed. The frequency analysis described in the present study offer insights into the distribution of different classes/types of microsatellite motifs in the turmeric transcriptome. This could also facilitate the designing of suitable SSR probes for the effective targeting and isolation of genomic microsatellite markers from turmeric.

Development of EST–SSRs

The present study could generate 17 polymorphic EST–SSR markers from a total of 45 primers tested. The non-amplification of the nine primers might be due to the usage of questionable or chimeric sequence information for primer designing [13]. The increase in the product size of CLEST SSR-11 than expected from the EST sequence might be due to the amplification of a small intron. During validation of the markers, the strongest bands were assigned to be alleles because of the possibility of the associated lighter bands being stutters arising due to the slippage of *Taq* polymerase during PCR [42, 43] and prevalence of staining artifacts.

Though intraspecific variability in chromosome number and assignment of a new ploidy status have been recently reported in *C. longa* [44], the study does not negate the earlier reports on the triploid nature of turmeric [45, 46]. The polymorphic SSR markers generated in the present study also revealed three or less alleles in most of the turmeric accessions analyzed, which is probably an indication of the triploid status of the plant. Hence, considering the allelic banding pattern observed in the present study

Table 3 Details of 17 EST–SSR markers developed in *Curcuma longa* L

Locus	Accession Number (Genbank)	Repeat motif#	Primer sequence (5′-3′)	T_a (°C)	Expected product size (bp)	N_a	Allele size (bp)	BLAST X top hit description [species] (Accession no:)
CLEST SSR-01	DY394887*	(CT)^15	FP-TTTGAGATGGCGAGTAGAAC RP-ATGAGGGAAGAGAGGAGAAG	60	180	5	189–157	Predicted protein [*Populus trichocarpa*] (XP_002311289)
CLEST SSR-02	DY389303*	(AAG)^12	FP-ACCGTAGCAAAGAAATAGGAC RP-AAGGTGGAAGGAAACTCG	62	183	6	204–152	UDP-glucose 6-dehydrogenase [*Zea mays*] (ACG34621)
CLEST SSR-03	DY394652	(ATT)^8	FP-AGGGAAAATAGAGTAGGCAAG RP-TGAAGGATTACAGTCAGCAAA	62	172	6	173–133	Alpha-1,4-glucan-protein synthase [*Ricinus communis*] (EEF36430)
CLEST SSR-04	DY393469*	(AAAC)^6	FP-ACACAACATTCAGTTTAGCAC RP-TCCCTATTCTTTCCTCTCG	62	184	5	188–172	Hypothetical protein [*Zea mays*](NP_001145684)
CLEST SSR-05	DY394828*	(CCTCTT)^5	FP-TATCCTCCCTGGTCGTTT RP-GATTCCCTTTCCTTTCTTTTG	58	184	4	185–175	Hypothetical protein [*Zea mays*] (NP_001132525)
CLEST SSR-06	DY393861*	(ATT)^8..(T)^10	FP-TCATCGTCTGCTTTAGTTTTC RP-ACGCTCTGCTCCTTCAAC	65	202	3	199–191	S-adenosyl-L-methionine synthetase [*Dendrobium crumenatum*] (AAL16064)
CLEST SSR-07	DY393567	(AT)^8(AC)^14	FP-AGACAGAAGAAGAGGCAGAAG RP-AAATGATGACCACGGACTAC	62	152	8	181–104	Hypothetical protein [*Oryza sativa* Japonica Group] (EAZ27998)
CLEST SSR-08	DY394311*	(AGAT)^8	FP-GATGCACACATTGCCCGTG RP-GGGTGCAATTCTTGGTCCG	60	171	4	181–154	Hypothetical protein [*Sorghum bicolor*] (EER92298)
CLEST SSR-09	DY394891*	(CTT)^7	FP-TCGGTTCTACTGAATCTTTACTCG RP-AGACTGTTTTCCCATTGTTGC	65	188	5	218–184	Hypothetical protein [*Oryza sativa* Japonica group] (EEE50593)
CLEST SSR-10	DY393238*	(ATC)^8	FP-GTGGTGGAGGAGGAAGAGAAG RP-TTGAGGGAACAAAAGGAAGAC	65	196	3	200–188	Photosystem-1 F subunit precursor [*Oryza sativa* Japonica group] (AAM19016)
CLEST SSR-11	DY390357	(AAAG)^5	FP-FP-TTCATTCGACGCAAACAGC RP-CGACGCAATAGTCGAAGGC	65	209	3	305–292	Putative NADH-ubiquinone oxidoreductase 1[*Ricinus communis*] (EEF50873)
CLEST SSR-12	DY391910*	(AGG)^6	FP-GGGATTGAGGTGGAGGTAGG RP-GCTGGCGAAGTAGAAGAAGAAG	65	150	4	162–140	Predicted protein [*Populus trichocarpa*] (EEE88547)
CLEST SSR-13	DY393462	(AAT)^7	FP-TGTACAAGCTCCAAATAAGTCAAG RP-CAGGAGTGTTCTAATGTTGCCC	65	154	4	158–136	Hypothetical protein[*Oryza sativa* Indica group] (EAY89333)
CLEST SSR-14	DY391880*	(GCT)^7	FP-CACCTCTCCTTCCCCAACC RP-GCCGTCCTCGCTTCTTCTTA	65	176	5	186–166	Hypothetical protein [*Vitis vinifera*] (XP_002275328)
CLEST SSR-15	DY391143*	(CCT)^7	FP-GCCAAAGAAAGAACTGACATCC RP-TTACAACCCTCCTCCCATTAGA	65	172	6	198–162	Hypothetical protein [*Vitis vinifera*] (CAN69661)
CLEST SSR-16	DY389892	(AGG)^5(AAG)^7	FP-AAGCAGTCCGTGGGAGAAG RP-CTTCCTCAATCGAATGGCCG	65	173	4	174–164	Conserved hypothetical protein [*Ricinus communis*] (EEF32318)
CLEST SSR-17	DY388605	(ATT)^7	FP-GTGCCTGTGGACCTATCCG RP-GAAGCATGCGAATTCATCTAAAC	65	174	4	192–174	AP2/ERF domain-containing transcription factor [*Populus trichocarpa*] (EEE79445)

T_a (°C)—annealing temperature of primer pairs; N_a—number of alleles; (N/NN/NNN)^ shows non-Class I repeat motifs included in the amplicons; # repeat motif detected within the assembled contig/singleton; * represents an EST sequence within the contig; *FP* forward primer; *RP* reverse primer

and triploid status of the plant [45, 46], the traditional measures for analyzing genetic variability (Ho, He, tests of Hardy–Weinberg, and linkage disequilibrium) were not carried out. More extensive study of these loci will be needed for estimating these parameters. The polymorphism generated by the markers by revealing variable number of alleles (3–8) offers possibility for their future utilization in revealing the genetic diversity of turmeric.

Cross Species Transferability

The high rate (100%) of cross species transferability of the generated SSR markers might be due to the sequence conservation in the ESTs of related *Curcuma* species. Thus, the EST–SSR markers generated in the present study offer potentiality for its future use in characterization and genetic diversity analysis of related *Curcuma* species. Moreover, the transferability efficiency indicates that the approach of utilizing EST sequences distributed in the public databases would help to generate reliable markers for comparative mapping in crops whose genetic/molecular information is little known.

Conclusion

In the present work, a robust set of 17 polymorphic markers were developed, showing that EST databases would be a useful source for the generation of genic SSR markers in turmeric. This study gives an insight into the frequency, type and distribution of Class I SSRs in turmeric transcriptome. The developed EST–SSR loci were able to generate sufficient polymorphism to ensure their use as molecular markers for genetic investigations in turmeric. The high transferability rate (100%) obtained by the amplification of markers in 13 *Curcuma* species confirm the scope of exploiting these markers for the effective fingerprinting of related *Curcuma* species and species delimitation studies. This study is the first to utilize the turmeric ESTs distributed in public databases for the identification of SSRs and generation of a robust set of microsatellite markers in turmeric.

Acknowledgment This work was supported by a research grant from Department of Biotechnology (DBT), Government of India, New Delhi. The authors thank DISC, IISR (Indian Institute of Spices Research) for providing the internet facility.

References

1. Velayudhan, K. C., Muralidharan, V. K., Amalraj, V. A., Gautam, P. L., Mandal, S., & Kumar, D. (1999). *Curcuma genetic resources. Scientific Monograph No. 4* (pp. 1–5). New Delhi: National Bureau of Plant Genetic Resources.
2. Sasikumar, B. (2005). Genetic resources of *Curcuma*: Diversity, characterization and utilization. *Plant Genetic Resources Characterization & Utilization, 3*, 230–251.
3. Shamina, A., Zachariah, T. J., Sasikumar, B., & George, J. K. (1998). Biochemical variation in turmeric (*Curcuma longa* L.) accessions based on isozyme polymorphism. *Journal of Horticultural Science & Biotechnology, 73*, 479–483.
4. Nayak, S., Naik, P. K., Acharya, L. K., & Pattnaik, A. K. (2006). Detection and evaluation of genetic variation in 17 promising cultivars of turmeric (*Curcuma longa* L.) using 4C nuclear DNA content and RAPD markers. *Cytologia, 71*, 49–55.
5. Syamkumar, S. (2008). *Molecular, biochemical and morphological characterization of selected Curcuma accessions*. Ph.D. Thesis, Calicut University, Calicut, India.
6. Salvi, N. D., George, L., & Eapen, S. (2001). Plant regeneration from leaf base callus of turmeric and random amplified polymorphic DNA analysis of regenerated plants. *Plant Cell, Tissue and Organ Culture, 66*, 113–119.
7. Panda, M. K., Mohanty, S., Subudhi, E., Acharya, L., & Nayak, S. (2007). Assessment of genetic stability of micropropagated plants of *Curcuma longa* L. by cytophotometry and RAPD analyses. *International Journal of Integrative Biology, 1*, 189–195.
8. Tyagi, R. K., Agrawal, A., Mahalakshmi, C., Hussain, Z., & Tyagi, H. (2007). Low-cost media for in vitro conservation of turmeric (*Curcuma longa* L.) and genetic stability assessment using RAPD markers. *In Vitro Cellular & Developmental Biology Plant, 43*, 51–58.
9. Parthasarathy, V. A., & Sasikumar, B. (2006). Biotechnology of *Curcuma*. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources, 1*(020), 9 pp.
10. Powell, W., Machray, G. C., & Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends in Plant Science, 1*, 215–222.
11. Gupta, P. K., & Varshney, R. K. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica, 113*, 163–185.
12. Gupta, P. K., Balyan, H. S., Sharma, P. C., & Ramesh, B. (1996). Microsatellites in plants: A new class of molecular markers. *Current Science, 70*, 45–54.
13. Thiel, T., Michalek, W., Varshney, R. K., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics, 106*, 411–422.
14. Tautz, D., & Renz, M. (1984). Simple sequence repeats are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research, 12*, 4127–4138.
15. Toth, G., Gaspari, Z., & Jurka, J. (2000). Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research, 10*, 967–981.
16. Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., & McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research, 11*, 1441–1452.
17. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., et al. (1991). Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science, 252*, 1651–1656.
18. Scott, K. D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E. M., Lee, L. S., et al. (2000). Analysis of SSRs derived from grape ESTs. *Theoretical and Applied Genetics, 100*, 723–726.
19. Aggarwal, R. K., Hendre, P. S., Varshney, R. K., Bhat, P. R., Krishnakumar, V., & Singh, L. (2007). Identification, characterization and utilization of EST-derived genic microsatellite

markers for genome analyses of coffee and related species. *Theoretical and Applied Genetics, 114*, 359–372.
20. Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genic microsatellite markers in plants: Features and applications. *Trends in Biotechnology, 23*, 48–55.
21. Lee, S. Y., Fai, W. K., Zakaria, M., Ibrahim, H., Othman, R. Y., Gwag, J. G., et al. (2007). Characterization of polymorphic microsatellite markers, isolated from ginger (*Zingiber officinale* Rosc.). *Molecular Ecology Notes, 7*, 1009–1011.
22. Kumpatla, S. P., & Mukhopadhyay, S. (2005). Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome, 48*, 985–998.
23. Masoudi-Nejad, A., Tonomura, K., Kawashima, S., Moriya, Y., Suzuki, M., Itoh, M., et al. (2006). EGassembler: Online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. *Nucleic Acids Research, 34*, W459–W462.
24. Faircloth, B. C. (2008). MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources, 8*, 92–94.
25. Gupta, P. K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., & Balyan, H. S. (2003). Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics, 270*, 315–323.
26. Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In S. Krawetz & S. Misener (Eds.), *Bioinformatics methods and protocols: Methods in molecular biology* (pp. 365–386). Totowa, NJ, USA: Humana Press.
27. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research, 25*, 3389–3402.
28. Syamkumar, S., Lowarence, B., & Sasikumar, B. (2003). Isolation and amplification of DNA from rhizomes of turmeric and ginger. *Plant Molecular Biology Reporter, 21*, 171a–171e.
29. Benbouza, H., Jacquemin, J. M., Baudoin, J. P., & Mergeai, G. (2006). Optimization of a reliable, fast, cheap and sensitive silver staining method to detect SSR markers in polyacrylamide gels. *Biotechnology, Agronomy, Society & Environment, 10*, 77–81.
30. Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D., & Waugh, R. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics, 156*, 847–854.
31. Kantety, R. V., La Rota, M., Matthews, D. E., & Sorrells, M. E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum, and wheat. *Plant Molecular Biology, 48*, 501–510.
32. Varshney, R. K., Thiel, T., Stein, N., Langridge, P., & Graner, A. (2002). *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cellular & Molecular Biology Letters, 7*, 537–546.
33. Gao, L., Tang, J., Li, H., & Jia, J. (2003). Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding, 12*, 245–261.
34. Poncet, V., Rondeau, M., Tranchant, C., Cayrel, A., Hamon, S., de Kochko, A., et al. (2006). SSR mining in coffee tree EST databases: Potential use of EST–SSRs as markers for the *Coffea* genus. *Molecular Genetics and Genomics, 276*, 436–449.
35. Dong, J., Guang-Yan, Z., & Qi-Bing, H. (2006). Analysis of microsatellites in citrus unigenes. *Acta Genetica Sinica, 33*, 345–353.
36. Palmieri, D. A., Novelli, V. M., Bastianel, M., Cristofani-Yaly, M., Astua-Monge, G., Carlos, E. F., et al. (2007). Frequency and distribution of microsatellites from ESTs of citrus. *Genetics & Molecular Biology, 30*, 1009–1018.
37. Gupta, S., & Prasad, M. (2009). Development and characterization of genic SSR markers in *Medicago truncatula* and their transferability in leguminous and non-leguminous species. *Genome, 52*, 761–771.
38. Feng, S. P., Li, W. G., Huang, H. S., Wang, J. Y., & Wu, Y. T. (2009). Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). *Molecular Breeding, 23*, 85–97.
39. Nicot, N., Chiquet, V., Gandon, B., Amilhat, L., Legeai, F., Leroy, P., et al. (2004). Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). *Theoretical and Applied Genetics, 109*, 800–805.
40. Metzgar, D., Bytof, J., & Wills, C. (2000). Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research, 10*, 72–80.
41. Rungis, D., Berube, Y., Zhang, J., Ralph, S., Ritland, C. E., Ellis, B. E., et al. (2004). Robust simple sequence repeat (SSR) markers for spruce (*Picea* spp.) from expressed sequence tags (ESTs). *Theoretical and Applied Genetics, 109*, 1283–1294.
42. Roa, A. C., Chavarriaga-Aguirre, P., Duque, M. C., Maya, M. M., Bonierbale, M. W., Iglesias, C., et al. (2000). Cross-species amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites: Allelic polymorphism and degree of relationship. *American Journal of Botany, 87*, 1647–1655.
43. Lagercrantz, U., Ellegren, H., & Andersson, L. (1993). The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Research, 21*, 1111–1115.
44. Skornickova, J. L., Sida, O., Jarolimova, V., Sabu, M., Fer, T., Travnicek, P., et al. (2007). Chromosome numbers and genome size variation in Indian species of *Curcuma* (Zingiberaceae). *Annals of Botany, 100*, 505–526.
45. Ramachandran, K. (1961). Chromosome numbers in the genus *Curcuma* Linn. *Current Science, 30*, 194–196.
46. Islam, M. A. (2004). *Genetic diversity of the genus Curcuma in Bangladesh and further biotechnological approaches for in vitro regeneration and long-term conservation of C. longa germplasm.* PhD thesis, University of Hannover, Germany.