Genes Genom (2013) 35:661–670
DOI 10.1007/s13258-013-0118-0

RESEARCH ARTICLE

Development of EST-SSR markers through data mining and their use for genetic diversity study in Indian accessions of *Jatropha curcas* L.: a potential energy crop

Maya Kumari · Atul Grover · Patade Vikas Yadav · Mohammad Arif · Zakwan Ahmed

Received: 6 February 2013 / Accepted: 22 May 2013 / Published online: 11 June 2013

Abstract *Jatropha curcas* L. is gaining importance as a potential energy crop. However, lack of sufficient numbers of molecular markers hinder current research on crop improvement in *Jatropha*. The expressed sequences tags (EST) sequences deposited in public databases, offers an excellent opportunity to identify simple sequence repeats (SSRs) through data mining, for further research on molecular breeding. In the present study 42,477 ESTs of *J. curcas* were screened, out of which 5,673 SSRs were identified with 48.8 % simple (excluding mononucleotide repeats) and 52.2 % compound repeat motifs. Amongst these repeat motifs, dinucleotide repeats were abundant (26.5 %), followed by trinucleotide (23.1 %) and tetranucleotide repeats (0.8 %). From these microsatellites, 32 EST-SSR (genic microsatellite) primer pairs were designed. These primers were used to analyze the genetic diversity among 42 accessions collected from different parts of India. Out of the 32 EST-SSR primers, 24 primer pairs exhibited polymorphism among the genotypes, with amplicons varying from one to eight, giving an average of 2.33 alleles per polymorphic marker. Polymorphic information content value ranged from 0.02 to 0.5 with an average of 0.402 indicating moderate level of informativeness within these EST-SSRs markers. The EST-SSR markers developed here will serve as a valuable resource for genetic studies, like linkage mapping, diversity analysis, quantitative trait locus/association mapping, and molecular breeding. The current study also revealed low diversity in the screened Indian *Jatropha* germplasm. Therefore, the future efforts must be made to broaden the gene pool of *Jatropha* for the creation of genetic diversity that can be further used for crop improvement through breeding.

Keywords Energy crop · EST-SSR · Polymorphism · Genetic diversity · *Jatropha* germplasm

Introduction

The genus *Jatropha* belonging to the family Euphorbiaceae, contains about 175 species in the world. *Jatropha curcas* (physic nut) is a semi woody perennial tree native to the America but found throughout the tropical and sub tropical regions of the world (Heller 1996). *Jatropha* is well adapted to arid, semi arid, low rain fall to high rainfall conditions and can be cultivated in soft, rocky, gravelly, sandy, calcareous, and saline soil (Francis et al. 2005). It is propagated both by seeds and cuttings. In India, 18 species are found scattered in different states of which *J. curcas* is gaining prominence as biofuel crop because of its adaptability to different habitats, larger fruits and seeds, soil conservation capabilities, use as live fence, and mainly for high oil yielding capacity leading to biodiesel production (Divakara et al. 2010). The plant is quite hardy, requires low soil fertility and moisture and can live for about 40 years in fruit bearing stage. *Jatropha* can be rightly called a wonder crop since no part of this plant goes waste. Its bark is used as raw material for dye, seed cake after oil extraction is used as a bio-fertilizer, leaf extract is used as insecticides/pesticides, and seed is used for extracting oil for biodiesel production (King et al. 2009). Oil of this plant is being used in traditional medicine and veterinary purposes since long time as it

Electronic supplementary material The online version of this article (doi:10.1007/s13258-013-0118-0) contains supplementary material, which is available to authorized users.

M. Kumari (✉) · A. Grover · P. V. Yadav · M. Arif · Z. Ahmed
Defence Institute of Bio-Energy Research (DIBER), Goraparao, Haldwani, Nainital 263 139, Uttarakhand, India
e-mail: madhulogin@gmail.com

prevents many infections, good for skin and can also be used as an antiseptic, anti inflammatory, molluscide, insecticide and fungicide (Lakshmanan et al. 1990; Rug and Ruppel 2000). The seeds of *J. curcas* contain 40–45 % oil (Openshaw 2000) with a high percentage of monounsaturated oleic and polyunsaturated linoleic acid. Thus, *Jatropha* oil has been recognized for the production of biodiesel (Fairless 2007). It is crucial to develop high seed yield and high oil content *Jatropha* varieties, well adapted to diverse conditions. For the improvement of *J. curcas* conventional, breeding programs might not be very effective due to its heterozygous genome, so use of genomic assisted breeding strategies becomes empirical.

Before the advent of molecular markers, morphological, qualitative and quantitative trait data were used to screen diversity in the germplasm. However, these markers are limited in number and offer very low polymorphism. DNA and PCR based molecular marker technology provides an efficient tool to assess genetic diversity, phylogenetic relationships, hybrid parent selections, population structure and distribution, mapping and tagging of genes or quantitative trait loci linked to important traits (Yadav et al. 2011). Currently there is very limited information available with respect to molecular markers in *J. crucas*. Genetic variability information is the prime requirement for any breeding programme. With the advent of molecular markers, different PCR based markers are being used for genetic studies, fingerprinting, gene mapping, etc., with their own advantages and limitations (Wang et al. 2011). Randomly amplified polymorphic DNA is cheap but non reproducible; restriction fragment length polymorphism is costly and laborious; amplified fragment length polymorphisms is costly, laborious though informative; simple sequence repeats (SSRs) are informative and reproducible but its generation is costly and require sequence information (Gao et al. 2003), while EST-SSRs or genic microsatellites or transcriptomic microsatellites are the markers of choice since they represent coding regions of the genome, easy to generate, involves less cost, co-dominant, informative and are highly repeatable and reproducible (Varshney et al. 2005).

For genetic diversity studies different marker types have been used but there are very few reports on the use of EST-SSR markers in *Jatropha* (Yadav et al. 2011; Chun and Zhong 2011; Wen et al. 2011). In recent years, EST-SSRs have received greater attention due to the increasing amounts of EST sequences being deposited in databases for various plant species (Wen et al. 2011). EST-SSR can be easily developed from EST databases through data mining. Due to their existence in transcribed regions of the genome, they can lead to the development of gene-based maps, which may help to identify candidate function genes and increase the efficiency of marker-assisted selection (Gupta and Rustgi 2004). EST-SSRs also show a higher level of transferability to closely related species and also genera (Grover et al. 2009; Wen et al. 2011).

To date, the number of available markers is grossly inadequate for genetic studies in *Jatropha*. No genome information is available in *Jatropha*; hence, there is a need for development of novel markers. Here, we screened large number of ESTs from *Jatropha* ET sequence databases with the objective to develop microsatellite markers in the transcribed regions of *Jatropha* genome and to assess the genetic diversity in *Jatropha* germplasm, collected from different regions of India.

Materials and methods

Germplasm material and DNA isolation

To analyze the genetic diversity of *J. curcas*, 64 accessions were collected through written request to different institutes and also through local collection from three different locations-semi-arid Secunderabad (17.45°N, 78.5°E; 543 m asl), foothills at Haldwani (29.22°N, 79.52°E; 424 m asl) and from hills at Pithoragarh (29.35°N, 80.13°E; 1,514 m asl) (Supplementary Table 1). Out of the 64 genotypes, seeds of only 42 germinated (Table 1). The accessions have been transplanted in field for further growth and development. At this stage data on 100 seed weight has been recorded (not produced here). Morphological and other quantitative data will be recorded as the plants grow. *Ricinus communis* (Castor), belonging to the same family Euporbiaceae was selected as an out-group. Total genomic DNA was isolated from young leaves using CTAB protocol standardized in our laboratory for high yield and quality (Supplementary File 2). The genomic DNA was quantified and diluted to a concentration of 25 ng/μl for PCR amplification.

Sequence resources, microsatellite mining and primer designing

Jatropha EST sequences were downloaded from EST db of NCBI and scanned locally on a PC terminal using MISA (Thiel et al. 2003). We systematically targeted simple microsatellite repeats (avoiding mononucleotide repeats) having lengths >30 bp for primer design. Primers were designed at least 30 bp upstream or downstream of the microsatellite span, using Primer3 (frodo.wi.mit.edu/cgi_bin/primer3/primer3_www.cgi). Complex repeats with no interruption and the interrupted repeats with interruption creating imperfection (An imperfect repeat is the one, in which 10 bp of flanking sequence on either side shows presence of the same motif or repeatability of a derived motif) were considered potentially polymorphic and were targeted for primer design. For design of primers adherence to the product %GC ≥40 and ≤80 was also considered. Primers were custom synthesized

Table 1 Functional annotation of EST sequences from *J. curcas*, using BLAST2GO and their repeat motifs, which were used to design the primers

Primer ID	GenBank ID	Seq. description	Seq. length	Min. eValue	Mean similarity (%)	GOs	SSR motif
DBR*Jc*001	FM891783.1	F-box protein phloem protein 2-b1 [*O. sativa* Japonica]	373	1.52E−21	80.15	Protein binding	(CT)17(AT)17
DBR*Jc*002a and 2b	FM896533.1	Protein	555	3.08E−26	73.20	Structural constituent of ribosome; translation	(TCT)13 (CT)10
DBR*Jc*004	FM895395.1	GA18412 [*Drosophila pseudoobscura*]	575	2.95E−04	54.00	Nucleic acid binding; nucleotide binding	(AT)19
DBR*Jc*003	FM889176.1	NA	563			–	(ATT)12
DBR*Jc*005	FM889738.1	Unknown protein [*G. max*]	417	2.06E−05	59.90	–	(TTCTT)7
DBR*Jc*006	FM895253.1	Conserved hypothetical protein [*Ricinus communis*]	463	8.34E−47	62.90	FMN binding	(TCT)11
DBR*Jc*007	GT228727.1	NA	326			–	(TATTT)6
DBR*Jc*008	FM889501.1	Sjchgc01971 protein	345	9.92E−08	70.27	–	(TA)28
DBR*Jc*009	GT229302.1	Hydroxyproline-rich glycoprotein precursor	526	1.27E−04	45.25	Protein binding	(TC)21
DBR*Jc*010	GT228457.1	Unknown protein [*L. japonicus*]	567	6.15E−21	69.15	Cytoplasmic membrane-bounded vesicle	(AAC)12
DBR*Jc*011	FM889075.1	Zinc finger a20 and an1 domain-containing stress-associated protein 8-like isoform 2	611	1.67E−84	80.45	Zinc ion binding; DNA binding	(CT)17
DBR*Jc*012	FM887106.1	NAC domain ipr003441	485	1.48E−19	82.50	DNA binding	(CAG)11
DBR*Jc*013	FM894501.1	NA	501			–	(AG)16
DBR*Jc*014	FM896040.1	Conserved hypothetical protein [*R. communis*]	485	1.78E−32	58.43	FMN binding	(TCT)10
DBR*Jc*015	FM891340.1	Mitochondrial import inner membrane translocase subunit tim17 tim22 tim23 family protein	465	7.39E−21	92.75	P–P-bond-hydrolysis-driven protein transmembrane transporter activity; mitochondrial inner membrane presequence translocase complex; integral to membrane; protein transport	(AT)27
DBR*Jc*016	FM891770.1	Late embryogenesis abundant protein	549	1.59E−39	91.80	Response to stress	(TA)22
DBR*Jc*017	FM896491.1	NA	583			–	(CT)18
DBR*Jc*018	FM896869.1	Protein binding	591	2.38E−61	77.65	Cytoplasmic membrane-bounded vesicle; zinc ion binding	(TC)16
DBR*Jc*019 and DBR*Jc*020	FM889451.1	Amidohydrolase family expressed	540	6.61E–16	80.50	Atrazine chlorohydrolase activity; hydrolase activity, acting on carbon–nitrogen (but not peptide) bonds	(AT)24
DBR*Jc*021	FM890349.1	NA	468			–	(TA)17
DBR*Jc*022	FM892249.1	Transcriptional corepressor	436	1.17E−14	86.40	Transcription cofactor activity; nucleus	(TC)16
DBR*Jc*023	FM896553.1	Hypothetical protein [*Plasmodium berghei* strain ANKA]	389	1.52E−04	65.00	–	(AT)24
DBR*Jc*024	FM887200.1	Tran membrane protein	333	1.44E−06	56.00	Nucleosome assembly; DNA binding	(TTA)11

Table 1 continued

Primer ID	GenBank ID	Seq. description	Seq. length	Min. eValue	Mean similarity (%)	GOs	SSR motif
DBRJc025	FM889664.1	Predicted protein [*Populus trichocarpa*]	593	7.39E−34	88.15	Biological process	(AC)15
DBRJc026	FM890563.1	Cytosolic class I small heat shock protein partial	420	2.10E−44	96.75	Response to stress	(TA)19
DBRJc027	FM890538.1	Cytosolic class I small heat shock protein partial	403	1.43E−39	99.20	Response to stress	(TA)25
DBRJc028	FM889625.1	At4g01090 f2n1_13	453	1.16E−22	88.90	Plasma membrane	(AT)16
DBRJc029	GO247057.1	Early nodulin 20 precursor	610	8.64E−77	62.90	Cell part	(TCT)11
DBRJc030	FM895034.1	Protein trichome birefringence-like 41	548	1.53E−44	75.75	DNA binding; zinc ion binding; DNA integration	(CT)15
DBRJc031	FM888881.1	Major facilitator super family	181	1.07E−04	65.14	Hydrolase activity, hydrolyzing O-glycosyl compounds; carbohydrate metabolic process; cation binding; catalytic activity	(AT)19
DBRJc032	FM889042.1	NA	553			–	(AG)19

from a commercial source (Imperial Life Sciences, India). The ESTs containing microsatellites used in this study were annotated for their function using Blast2GO web based programme (Conesa et al. 2005).

PCR amplification

PCR amplification in 20 µl reaction volume was optimized with 1 µl of DNA template (25–50 ng), 10 mM Tris–HCl (pH 8.3), 1.5 mM $MgCl_2$, 0.5 µM primers, 0.2 mM dNTPs, and 1U *Taq* DNA polymerase (Banglore Genei, India). A Mastercycler gradient (Bio-Rad, USA) was used for thermal cycling as: 2 min at 94 °C, followed by 30 cycles of 30 s at 94 °C, 45 s at Primer Tm/Ta annealing temperature, 45 s at 72 °C, and final extension for 7 min at 72 °C. Gradient PCR was done to standardize the annealing temperature of the primer pairs. PCR products were separated on a 4 % agarose gel through electrophoresis, stained with ethidium bromide (50 µg/ml) and visualized with UV light. To check the reproducibility of the results, the PCR amplification with each primer pair was carried out at least three times.

Data collection and analysis

Clearly resolved, unambiguous bands generated by EST-SSR primer pairs were scored visually for their presence (1) or absence (0) with each primer pair. As the measure of usefulness of each marker in distinguishing one individual from another, polymorphic information content (PIC) was calculated using formula:

$$\text{PIC} = 1 - \Sigma p_i^2$$

where p_i is the frequency of the *i*th allele for individual *p*.

Cluster analysis

The binary data was used to construct a dendrogram. The genetic associations among the germplasm lines were evaluated by calculating the dice's similarity coefficient for pairwise comparisons based on the proportions of shared bands produced by primers. Similarity matrix was generated using the NTSYS-pc 2.0 (Rohlf 1998). The similarity coefficients were used for cluster analysis and dendrogram was constructed by the unweighted pair-group method using arithmetic averages (UPGMA) (Sneath and Sokal 1973).

Results

Identification, characterization and development of SSR markers from *J. curcas* EST sequences

Analysis of the 42,477 ESTs of *Jatropha* from NCBI GenBank database, revealed 5,673 SSRs in *J. curcas*

(Additional file in excel sheet) with 48.8 % simple (excluding mononucleotide repeats) and 52.2 % compound repeat motifs. Amongst these repeat motifs, dinucleotide repeats were abundant (26.5 %), followed by trinucleotide (23.1 %) and tetranucleotide repeats (0.8 %). Pentanucleotide and hexanucleotide motifs were also found with very low frequency (<1 %) (Fig. 1). The EST sequences spanned 5.6 % of the estimated genome (410 Mb) of *Jatropha*. Out of the 5,673 SSR containing sequences, 689 sequences contained more than one SSR. Among the single nucleotide repeats "A/T" was the most common (46.34 %), "AG/CT" was the most common dinucleotide repeats (19.7 %) and "AAG/CTT" was the most common trinucleotide repeat (7.3 %). Rest of the repeats were found in very small numbers (Supplementary Table 2). Six hundred and forty microsatellites fitted in the criteria described earlier for their exploitation as molecular markers; however, to find primers in the desired range of % GC and Tm, 32 ESTs were selected for primer designing. The species distribution of the EST sequences from which primers were designed showed maximum synteny with *Glycine max*, followed by *R. communis* and *Arabidopsis* (Supplementary Fig 1). The functions of the ESTs were annotated using Blast2Go software. Out of the 32 ESTs, 25 were annotated and functionally classified into different categories mainly biological processes, cellular processes, metabolic processes, stress responses, Nitrogen compound metabolic process and biosynthesis process (Supplementary Table 3).

Marker informativeness and the assessment of genetic diversity in *J. curcas*

Thirty two EST sequences showing GC content >40 % were selected for SSR marker development. Out of the 32 ESTs, 22 ESTs showed dinucletotide repeat motifs, eight ESTs showed trinucleotide repeat motifs and two ESTs showed tetranucleotide repeat motifs, as most common (Table 1). Out of the 32 primer pairs synthesized, one (DBR*Jc*003) did not show any amplification while seven primer pairs amplified but gave no polymorphism (Table 2). Polymorphic patterns were produced only by 24 primer pairs with more than two amplicons per primer pair. The number of amplicons for EST-SSRs ranged from one to eight with a mean of 2.19. DBR*Jc*011 primer pair gave maximum of eight amplicons followed by DBR*Jc*005 (four amplicons) (Table 2). The sizes of the amplicons for the EST-SSRs frequently violated the expected size ranges and spanned 100 to 600 bp, when analyzed on 4 % agarose gel. A total of 68 alleles were detected across 42 genotypes of which 56 alleles (82.35 %) were polymorphic with an average of 2.33 alleles per polymorphic marker (Table 2). Polymorphic information content was high for 13 markers (~0.50), moderate for 5 markers and low for 4 markers. The average PIC for polymorphic markers was 0.402, showing moderate level of informativeness for the markers. Five primer pairs, i.e., DBR*Jc*002b, DBR*Jc*004, DBR*Jc*007, DBR*Jc*009 and DBR*Jc*010 could also be cross-amplified in castor coming from the same family.

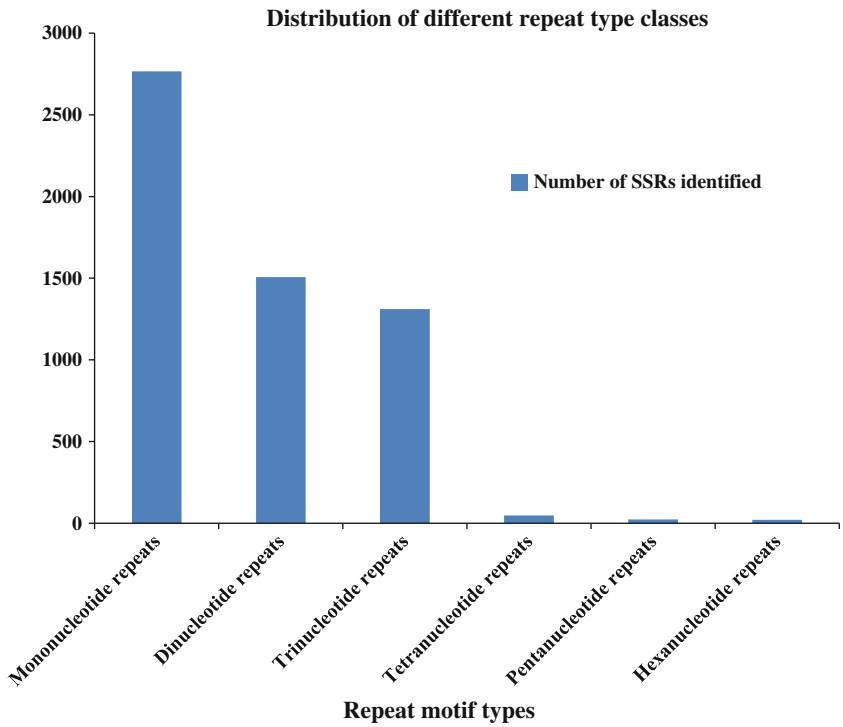

Fig. 1 Distribution of different types of repeat motifs for the SSRs identified from the EST sequences of *Jatropha curcas*. Single nucleotide repeats were the most abundant followed by di nucleotide repeats. Hexa and pentanucleotide repeats were found in nonsignificant amounts

Table 2 Primer sequence and informativeness of the EST-SSRs developed from *Jatropha curcas* (primer sequence) EST sequences

S. N.	EST-SSR ID	Primer (5′–3′) forward sequence	Primer (5′–3′) reverse sequence	Observed product size range (bp)	Expected product size (bp)	Alleles observed	Polymorphic alleles	PIC
1	DBR*Jc*001	TGAGGAGATGGAAGGTGGTC	GCAGGTGACAATCTTATTGCTTC	75–150	238	2	2	0.5
2	DBR*Jc*002a	CACATGGGGATACACAGCAG	GTGGCTAACCTCCTCATTTCAG	50–120	236	2	2	0.5
3	DBR*Jc*002b	ACCCATCATCAACGTTTTCC	TGCATTTGCAATTGAAGGAG	50–120	225	2	2	0.5
4	DBR*Jc*003	AAATCAAACTAATCGCATTTTGAA	TCAAATCCCTGATCATCTCTACC	No amplification	248	–	–	–
5	DBR*Jc*004	CGTGCGTTACAACCAAAATG	CCTTCGGCTTTGATAACTCG	100–150	236	2	1	0.172
6	DBR*Jc*005	ACTGTACACGCCTTCCTCAC	GGACCAAGATCACCTTCTTTTG	150–200	229	5	4	0.172
7	DBR*Jc*006	TCAACCTTGGCAGTGAGTTC	GAAGCAAAGAGCCTGAGCTG	100–150	214	1	1	0
8	DBR*Jc*007	GCCTGAGGCTATTTACTTGAG	TTATGCCTGAAAGCCATTTTG	100–150	244	2	2	0.482
9	DBR*Jc*008	AAGTGGGTCGTTGGAGAGAAC	TTGTAAAGGATGAGTAACACAGAC	100–200	223	2	2	0.456
10	DBR*Jc*009	GTTCTCATCTCCGGTGAACG	AAGCTCAACTCGCCACACTT	100–150	154	2	2	0.444
11	DBR*Jc*010	AACTCCACCATCCTATCACC	R-CCCAATTGATGGCAGTAACC	100–150	244	2	2	0.472
12	DBR*Jc*011	AACGCCTCTCTCTCCGTTTG	TTGTTGCTGCACTTCCAAAG	150–600	238	8	8	0.5
13	DBR*Jc*012	TGGGAGCTAAATTGGACAAA	GTCGAAGTGCAGCCCATAAT	100–150	152	1	0	0
14	DBR*Jc*013	CCATAAACATGCGTTTCACG	GTCCCAACTACAAGCCAAGC	100–150	204	2	2	0.464
15	DBR*Jc*014	TTCCCTGAAACAGTCCCAAT	TTCCCCACCTGATAGCTCAC	75–120	186	2	2	0.314
16	DBR*Jc*015	TGGCAAGTTCTTTTCCACAAC	TCCTTCCCAACTGTGATTGTTTG	75–120	218	2	1	0.024
17	DBR*Jc*016	GCATGTTATTTTGGGGTTGC	GATACCCTTCATGCCCAATC	100–250	229	3	2	0.497
18	DBR*Jc*017	GCGACCTGAAAGCAACTCTC	GGCCGTCTAGGGAATGTGTG	100–150	171	1	0	0
19	DBR*Jc*018	TCGAGGTTTCAATACAAAGCAA	TGAAAAAGGGTAGCTTTTGAGG	75–100	227	1	0	0
20	DBR*Jc*020	TGGCCCTTACATGAAATTACG	GATCTGGACACTGGATGGTG	100–150	243	2	1	0.024
21	DBR*Jc*021	TGATGTTCCAGGAATCGAGA	GGAATTTACCAACCCATGTCC	100–120	199	1	0	0
22	DBR*Jc*022	TGGCCTAGTCATTCCCTGTC	GCAAGACTTGACAGAAACTCTCAG	100–150	250	2	2	0.495
23	DBR*Jc*023	AACATCACGCTGAGAACTCC	TCATCTCCAGGTGAAAGTTGC	50–100	228	2	2	0.5
24	DBR*Jc*024	GTTTTTGGAAGATGGCATGAC	GTTTCACAGAAAGTACAAACCTTTTG	100–120	174	1	0	0
25	DBR*Jc*025	TCATACGGCAGTTCAAGCTC	TTGATTTCATCCCCAACACC	120–200	235	2	2	0.408
26	DBR*Jc*026	CTTTCAGTACATACTTTCTCACACAC	CCGGAGAATGCTAAGATGGA	100–150	228	3	2	0.5
27	DBR*Jc*027	TTTCAGTACATACTTTCTCACACAC	CGGAGAATGCAAAGATGGATC	100–150	229	2	2	0.5
28	DBR*Jc*028	ACATCCACAAACGCAAAGTT	TGGAATTCGAAGCACCTTCT	75–100	221	2	2	0.495
29	DBR*Jc*029	GGGGAGACTCAAACTGAAAC	TGCTGATTTTGGTGGTGGAG	100–150	204	3	2	0.308
30	DBR*Jc*030	CAGACTGCTTAGTTAGACTTGTAAAAA	AGTGTCCGAATGGTGGAGAG	50–100	155	1	1	0
31	DBR*Jc*031	TTGAAATCAAAGCATAGGGAGAA	TTTAGAAAGGAAATTAATATTTTGAGG	50–100	108	2	2	0.5
32	DBR*Jc*032	CTTGAATTCTCAGACGGTCTTT	CATGTCTGCCTTTGTTGGAA	100–120	246	3	3	0.423
					Total	68	56	
					Mean	2.13	2.33	0.402

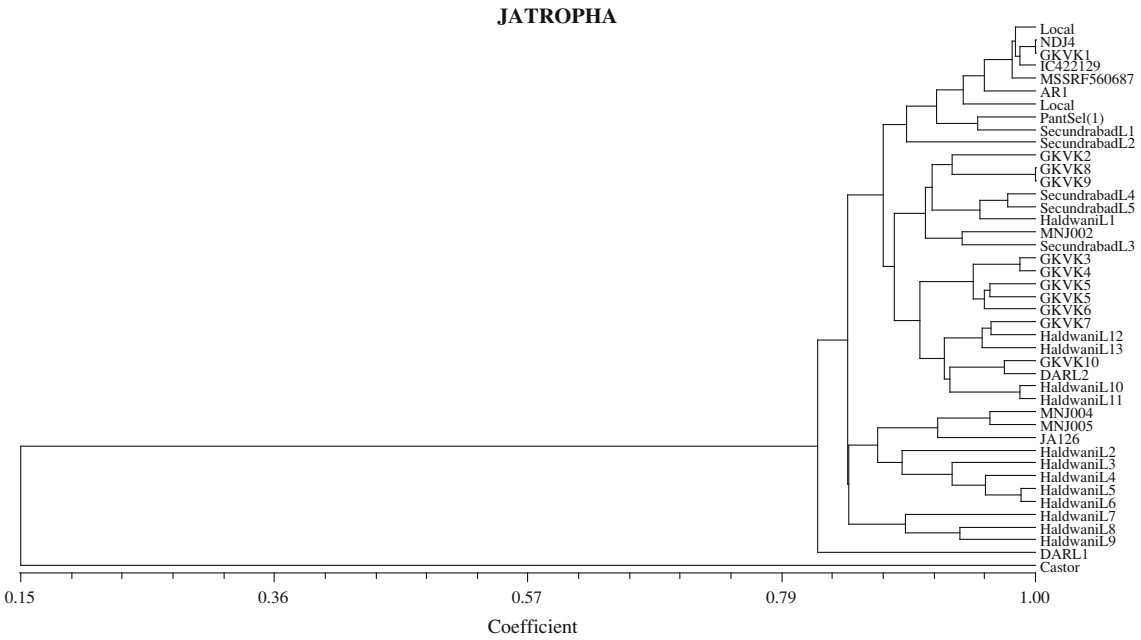

Fig. 2 An UPGMA tree constructed based on Dice's similarity coefficient using NTsys-PC, showing Dice's similarity coefficients between *Jatropha* accessions. Clearly, a high level of genetic similarity was displayed, with Dice's similarity coefficients distributed in a narrow range of 0.82 (divergence of accession DARL-1)–1.0 (Acc. NDJ4, GKVK1, GKVK8 and GKVK9). Two consecutive minor units on the coefficient scale are separated by a value of ~0.04

Genomic DNA isolated from young leaves of 42 *Jatropha* accessions were used for genetic diversity study. Out of the 64 germplasm accessions collected, 42 accessions germinated. The rest 22 genotypes did not germinate, maybe the seeds were very old and so lost viability. Loss of seed viability upon storage is one of the major problems in *Jatropha*. Genotyping data were used to estimate pairwise similarity comparisons among the accessions of *Jatropha*. The Dice's similarity coefficient values of the phenogram ranged from 0.82 to 1.00 (Fig. 2) indicating very low diversity in the screened germplasm. There was high degree of similarity in accessions coming from Bangalore (GKVK2–GKVK10). DARL-1 from Uttarakhand, identified by DIBER, appeared to be the most distinct; therefore it can be used in the breeding programme. The results indicated narrow genetic base (Fig. 2; Table 3) in the screened accessions of *Jatropha*.

Discussion

Genetic variability is one of the essential requirements for crop improvement through plant breeding. The prime step in this process involves germplasm screening for establishment of genetic diversity. With the increasing knowledge on genome, molecular markers have been assisting the other marker systems like morphological and quantitative trait data for genetic characterization studies. There are very few reports available on the development of genome-derived SSR markers in *J. curcas*. Hence, there is need to develop more number of efficient genetic markers to enrich the marker pool of *J. curcas,* to be utilized in genetic improvement program through molecular breeding. Recent trends in marker technology favour the development of functional markers instead of anonymous markers (Andersen and Lübberstedt 2003), since they can be used to assess the functional diversity in the populations and may prove more powerful for marker-assisted breeding. In the present investigation, screening of 42,477 ESTs sequences from NCBI database for SSR search, revealed 5,673 SSRs containing sequences. In *J. curcas*, the highest proportion of repeat motifs comprised of single nucleotides (48.8 %) followed by dinucleotides (26.5 %) and trinucleotides (23.1 %). The present finding is contradictory to the earlier report in *Jatropha*, (Yadav et al. 2011) where they found abundant of dinucleotide repeats (66 %). These differences in the frequency of different nucleotide motifs to some extent may be attributed to SSR search criteria used for data mining and the characteristics of the EST database analyzed (Yadav et al. 2011).

Out of the 32 selected primer pairs, 31 primer pairs gave successful amplification. One primer pair (DBR*Jc*003) failed to produce any amplification. The failure might have been due to unfavorable primer location or sequence problem during synthesis (Nicot et al. 2004). Seven primer pairs gave monomorphic amplification which maybe due to

Table 3 List of *Jatropha* accessions, along with their source, used for genetic diversity study using EST-SSR markers

S N	Germplasm/accessions	Code	Region	S N	Germplasm/accessions	Cbode	Region
1	Local collection	A	Gujarat	22	DARL-1	N1	Uttarakhand
2	NDJ-4	B	UP	23	Secundrabad Local Collection1	O1	Hyderabad
3	IC-422129	C	Rajasthan	24	Secundrabad Local Collection2	O2	Hyderabad
4	Local collection	D	Bangalore	25	Secundrabad Local Collection3	O3	Hyderabad
5	GKVK-1	E1	Bangalore	26	Secundrabad Local Collection4	O4	Hyderabad
6	GKVK-2	E2	Bangalore	27	Secundrabad Local Collection5	O5	Hyderabad
7	GKVK-3	E3	Bangalore	28	Lama chaur Local Collection6	P1	Uttarakhand
8	GKVK-4	E4	Bangalore	29	Transport Nagar Local Collection	P2	Uttarakhand
9	GKVK-5	E5	Bangalore	30	Halduchaur—Local collection1	P3	Uttarakhand
10	GKVK-6	E6	Bangalore	31	Halduchaur—Local Collection2	P4	Uttarakhand
11	GKVK-7	E7	Bangalore	32	Halduchaur— Local Collection3	P5	Uttarakhand
12	GKVK-8	E8	Bangalore	33	Halduchaur—Local Collection4	P6	Uttarakhand
13	GKVK-9	E9	Bangalore	34	Sangam Vihar—Local Collection1	P7	Uttarakhand
14	GKVK-10	E10	Bangalore	35	Sangam Vihar—Local Collection2	P8	Uttarakhand
15	MNJ-002	F1	Manipur	36	Sangam Vihar—Local Collection3	P9	Uttarakhand
16	MNJ-004	F2	Manipur	37	Beriparao—Local Collection1	P10	Uttarakhand
17	MNJ-005	F3	Manipur	38	Beriparao—Local Collection2	P11	Uttarakhand
18	JA-126	H	DBT (Rajasthan)	39	Kaladhungi Local Collection1	P12	Uttarakhand
19	AR-1	J	DBT	40	Kaladhungi Local Collection2	P13	Uttarakhand
20	MSSRF-560687	K	DBT (Tamil Nadu)	41	DARL-2	N2	Uttarakhand
21	Pant–Sel (1)	M	Uttarakhand	42	Castor (*Ricinus communis*)	C	

the highly conserved region of the sequences from which the primers were designed. The remaining 24 primer pairs amplified to give more than one amplicons. Most of the primer pairs gave the amplicon size smaller than the expected size while very few (DBR*Jc*011) gave amplicon size larger than the expected, which maybe due to the presence of introns (Yadav et al. 2011). Out of the 32 primer pairs, 5 could cross amplify in castor (*R. communis*) which was used as an outlier, indicating that these sequences were conserved across genera of the same family. This shows there is some degree of synteny between the genomes of *Jatropha* and castor, coming from the same family, Euphorbiaceae. Hence, they can also be used to check for cross-transferability of markers in related species, genera and even in the same family. Sato et al. (2011) reported high degree of microsynteny with the genome of castor and to a lower extent with soybean and *Arabidopsis*. However in the present study maximum hit was obtained from soybean (Supplementary Fig 1). This discrepancy might be due to small number of sequences used to find the species distribution. The PIC value of each EST-SSR locus was determined to assess the informativeness of the makers and accordingly defined into high (PIC $>$ 0.5), moderate (0.5 $>$ PIC $>$ 0.25), and low (PIC $<$ 0.25) categories (Bostein et al. 1980). The average PIC value of 0.402 was obtained with highest PIC value of 0.50, which is in congruence with recently reported PIC values of 0.04–0.61 of EST-SSRs developed by Yadav et al. (2011) in *Jatropha*. Nine out of thirty-two primers gave high PIC value indicating its good informativeness as a marker. Some of the markers gave low polymorphisms. This maybe because genic regions, where these EST-SSRs lie are relatively less susceptible to mutations. The microsatellites occurring in these regions might not show higher mutation rates because of the vitality of the function of the gene and that maybe a basis of detection of low polymorphism. Further if the level of polymorphism detected at a microsatellite locus is low, chances maybe high that selective forces are acting to conserve the locus (Eckert and Hile 2009). It is generally expected that a microsatellite locus with no or lesser polymorphism might be of recent origin (Schlotterer 2003) or associated to an important region of the genome (Grover and Sharma 2011). However, narrow genetic base of the *Jatropha* germplasm in India might be another and most practical reason for observing low polymorphism. *Jatropha* is an introduced species in India and the entire current germplasm has descended from the few originally introduced species.

In the present study the utility of the developed EST-SSR markers were tested for genetic diversity studies in the collected *Jatropha* germplasm. Here, we further add up 32 EST-SSRs in public domain for genetic studies in *J. curcas*, 24 of which revealed polymorphic patterns in the 42 accessions collected from different germplasm repositories

in India (Add. File 1). There are several reports in literature upon germplasm collection and genetic diversity assay using different marker systems throughout the *Jatropha* growing regions like China (Sun et al. 2008), Brazil (Rosado et al. 2010; Grativol et al. 2011), Mexico (Quintero et al. 2011), Thailand (Phumichai et al. 2011; Tanya et al. 2011) and India (Basha et al. 2009), but the germplasm screened and the number of polymorphic markers identified are still not sufficient (Supplementary Table 4). Genetic diversity levels in *Jatropha* around the world except at centre of its origin have been low to moderate (QiBao et al. 2008; Wen et al. 2010; Cai et al. 2010; Grativol et al. 2011; Quintero et al. 2011; Ricci et al. 2012). Very low diversity in Indian *Jatropha* germplasm has been reported by previous workers (Tatikonda et al. 2009; Basha et al. 2009; Sudheer et al. 2010) hence, need was felt to look for new and diverse sources of genetic variation in *J. curcas* that could be used in breeding programs. The germplasm collected is not a complete representation of the available germplasm in India. We collected *Jatropha* accessions from various germplasm centers and also through local collection. Many of the germplasm seeds failed to germinate maybe due to loss of viability upon storage. The age of the seeds obtained were not known. The Dice's similarity coefficient values (0.82–1.00) indicated very low diversity in the screened germplasm. The highest degree of similarity was displayed in accessions collected from Bangalore (GKVK2–GKVK10). Accessions from Uttarakhand were more diverse, followed by Hyderabad and Manipur. This is not conclusive of the diversity present in a particular state, because of small number of accessions received from some of the states, as we could not get germplasm from all the *Jatropha* research centers. Also the pedigrees for these accessions were unknown. DARL-1 from Uttarakhand, identified by DIBER, appeared to be the most distinct. It maybe used in the breeding programme. The results indicated narrow genetic base (Fig. 2, Table 3) in the screened accessions of *Jatropha*.

Based on the results, no obvious correlation of genotypes of the *J. curcas* to their geographical locations in India could be established. Genotypes from Uttarakhand indicated some diversity (Fig. 2). The genotypes received from Bangalore showed high degree of similarity. Local germplasm collected from in and around Haldwani area showed some diversity. The reason for low level of genetic diversity in the accessions maybe because of migration of the species due to human activities (reported or unreported) or because of outcrossing between these genotypes or because of clonal propagation of same genetic material in different geographical locations. Sudheer et al. (2010) also reported low genetic diversity in *J. curcas* and the clustering pattern suggested that the distribution of species might have happened through anthropogenic activity. Therefore, enrichment of the Indian gene pool of *J. curcas* is required to broaden its genetic base. We further recommend screening of more number of germplasm from different countries along with Indian accessions, with appropriate representation from different geographical regions and agroclimatic zones and also use all the polymorphic markers reported so far to screen all the available germplasm for genetic diversity.

Acknowledgments The authors express their gratitude to Dr. M.K. Rajesh, Senior Scientist, CPCRI, Kerala, India for helping in data analysis. Authors are also thankful to the donors of *Jatropha* accessions.

References

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Basha SD, Francis G, Makkar HPS, Becker K, Sujatha M (2009) A comparative study of biochemical traits and molecular markers for assessment of genetic relationships between *Jatropha curcas* L. germplasm from different countries. Plant Sci 176:812–823

Bostein D, White RL, Sholnick M, David RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet 32:314–331

Cai Y, Sun D, Wu G, Peng J (2010) ISSR-based genetic diversity of *Jatropha curcas* germplasm in China. Biomass Bioenerg 34:1739–1750

Chun Y, Zhong LA (2011) Development of EST-SSR markers from *Jatropha curcas* and their application in genetic diversity analysis among germplasms. Plant Diversity Resour 33:529–534

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Divakara BN, Upadhyaya HD, Wani SP, Gowda CL (2010) Biology and genetic improvement of *Jatropha curcas* L.: a review. Appl Energ 87:732–742

Eckert KA, Hile SE (2009) Every microsatellite is different: intrinsic DNA features dictate mutagenesis of common microsatellites present in human genome. Mol Carcinog 48:379–388

Fairless D (2007) Biofuel: the little shrub that could maybe. Nature 449:652–655

Francis G, Edinger R, Becker K (2005) A concept for simultaneous wasteland reclamation, fuel production, and socio-economic development in degraded areas in India: need, potential and perspectives of *Jatropha* plantations. Nat Resour Forum 29:12–24

Gao LF, Tang J, Li H, Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breeding 12:245–261

Grativol C, Catarina da Fonseca LM, Hemerly AS, Ferreira PCG (2011) High efficiency and reliability of inter-simple sequence repeats (ISSR) markers for evaluation of genetic diversity in Brazilian cultivated *Jatropha curcas* L. accessions. Mol Biol Rep 38:4245–4256

Grover A, Sharma PC (2011) Is spatial occurrence of microsatellites in genome, a determinant of their function and dynamics contributing to genome evolution? Curr Sci 100:859–869

Grover A, Ramesh B, Sharma PC (2009) Development of microsatellite markers in potato and their transferability in some members of Solanaceae. Physiol Mol Biol Plants 15:343–358

Gupta PK, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomics 4:139–162

Heller J (1996) Physic nut, *Jatropha* curcas. Promoting the conservation and use of underutilized and neglected crops 1. IPGRI, Rome, p 1–66. http://www.bioversityinternational.org/fileadmin/bioversity/publications/pdfs/161_Physic_nut__Jatropha_curcas_L.pdf

King AJ, He W, Cuevas JA, Freudenberger M, Ramiaramanana D, Graham IA (2009) Potential of *Jatropha curcas* as a source of renewable oil and animal feed. J Exp Bot 13:1–9

Lakshmanan P, Mohan S, Jeyarajan R (1990) Antifungal properties of some plant extracts against *Thanatephorus cucumeris*, the causal agent of collar rot disease of *Phaseolus aureus*. Madras Agric J 77:1–4

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109:800–805

Openshaw K (2000) A review of *Jatropha Curcas* L: an oil plant of unfulfilled promise. Biomass Bioenerg 19:1–15

Phumichai C, Phumichai T, Kongsiri N, Wongkaew A, Sripichit P, Kaveeta R (2011) Isolation of 55 microsatellite markers for *Jatropha curcas* and its closely related species. Biol Plantarum 55:387–390

QiBao S, Lin FL, Yong L, Guo JW, Xue JG (2008) SSR and AFLP markers reveal low genetic diversity in the biofuel plant *Jatropha curcas* in China. Crop Sci 48:1865–1871

Quintero VP, López JLA, Colmenero AZ, García NM, Colín CAN, Bonilla JLS, Rangel MRA, Langarica HRG, Bustamante DJM (2011) Molecular characterization of *Jatropha curcas* L. genetic resources from Chiapas. México through AFLP markers. Biomass Bioenerg 35:1897–1905

Ricci A, Chekhovskiy K, Azhaguvel P, Albertini E, Falcinelli M, Saha M (2012) Molecular characterization of *Jatropha curcas* resources and identification of population-specific markers. Bioenerg Res 5:215–224

Rohlf FJ (1998) On applications of geometric morphometrics to studies of ontogeny and phylogeny. Systems Biol 47:147–158

Rosado TB, Laviola BG, Faria DA, Pappas MR, Bhering LL, Quirino B, Grattapaglia D (2010) Molecular markers reveal limited genetic diversity in a large germplasm collection of the biofuel crop *Jatropha curcas* L. in Brazil. Crop Sci 50:2372–2382

Rug M, Ruppel A (2000) Toxic activities of the plant *Jatropha curcas* against intermediate snail hosts and larvae of schistosomes. Trop Med Int Health 5:423–430

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N et al (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Res 18:65–76

Schlotterer C (2003) Hitchhicking mapping: functional genomics from the population genetics perspectives. Trends Genet 19:32–38

Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman Press, San Francisco

Sudheer PDVN, Mastan SG, Rahman H, Reddy MP (2010) Molecular characterization and genetic diversity analysis of *Jatropha curcas* L. in India using RAPD and AFLP analysis. Mol Biol Rep 37:2249–2257

Tanya P, Taeprayoon P, Hadkam Y, Srinives P (2011) Genetic diversity among *Jatropha* and *Jatropha*-related species based on ISSR markers. Plant Mol Biol Rep 29:252–264

Tatikonda L, Wani SP, Kannan S, Beerelli N, Sreedevi TK, Hoisington DA, Devi P, Varshney RK (2009) AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L., a biofuel plant. Plant Sci 176:505–513

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Varshney RK, Graner A, Sorrells ME (2005) Geneic microsatellites in plants: features and applications. Trends Biotechnol 23:48–55

Wang CM, Liu P, Yi C, Gu K, Sun F, Li L, Lo LC, Liu X, Feng F, Lin G et al (2011) A first generation microsatellite and SNP based linkage map of *Jatropha*. PLoS One 6:e23632. doi: 10.1371/journal.pone.0023632

Wen MF, Wang HY, Xia Z, Zou M, Lu C, Wang WQ (2010) Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha Curcas* L. BMC Res Notes 3:42

Wen MF, Chen X, Wang HY, Lu C, Wang WQ (2011) Transferability analysis of cassava EST-SSR and genomic-SSR markers in *Jatropha* and rubber tree. Acta Agrono Sinica 37:74–78

Yadav HK, Ranjan A, Asif MH, Mantri S, Sawant SV, Tuli R (2011) EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species/genera. Tree Genet Genomes 7:207–219