Genes Genom (2013) 35:1–12
DOI 10.1007/s13258-013-0064-x

RESEARCH ARTICLE

Comparative analysis of EST-derived markers for allelic variation in *Jatropha curcas* L. and cross transferability among economically important species of Euphorbiaceae

Wachira SaiSug · Kittipat Ukoskit

Received: 5 March 2012 / Accepted: 11 July 2012 / Published online: 29 January 2013

Abstract The large number of expressed sequence tags (ESTs) now available has facilitated the development of molecular markers in many plant species. In the present study, primer pairs amplifying 133 ESTs containing simple sequence repeats (EST-SSR) and 112 intron-length polymorphism (ILP) markers were designed from *Jatropha* unigenes and unique transcripts from cassava, respectively. Of the primers designed, 71 and 57 were polymorphic for EST-SSR and ILP markers, respectively, in 59 *Jatropha* accessions. A similar average polymorphism information content (PIC) was observed for each marker system (0.37 for EST-SSR and 0.35 for ILP), suggesting that the two marker systems were similarly effective in determining polymorphisms in *Jatropha*. A comparison of PCR amplification ability between the two types of markers and cross species transferability indicated that ILP markers, for which a large proportion (68.6 %) of both primers was found in the coding sequence, were significantly more transferable in *J. gossypifolia*, cassava, rubber tree, and castor bean than EST-SSR markers, for which most (78.9 %) of one or both primers were located in the untranslated regions. This may be attributed to sequence conservation at the primer binding sites in the protein coding region. A comparison between two dendrograms constructed on the basis of EST-SSRs and ILPs revealed that discrimination of same-species alleles by EST-SSRs is clearer than that by ILPs, while ILPs seem to be more reliable for identification between species. The EST-derived markers developed in the present study should be useful for genetic diversity and comparative mapping across species in Euphorbiaceae.

Keywords *Jatropha* · Euphorbiaceae ILP · EST · SSR

Introduction

Jatropha curcas, a member of the euphorbia family, is a promising biofuel crop, which may emerge as a major alternative to Diesel. Breeding of high yielding *J. curcas* has therefore become a focus of attention within the research community. The success of these programs lies in the identification of genetically divergent material and development of genetically superior stocks. Therefore, information on the number of introductions and the genetic diversity of *J. curcas* populations is critical to increasing the efficiency of *J. curcas* improvement programs.

The recent rapid increase in the number of expressed sequence tag (EST) sequences in the public database means that the development of EST-containing simple sequence repeats (EST-SSRs) is an attractive option for the development of SSR markers. EST-SSRs have some advantages over genomic SSRs. Since these markers are derived from expressed genes, they are more conserved and have a better potential for applications used for identifying conserved genomic regions among species and genera, comparative genomics, and evolutionary studies (Thiel et al. 2003; Eujayl et al. 2004; Ukoskit et al. 2012). Because of the above advantages, SSR markers have been developed from

Electronic supplementary material The online version of this article (doi:10.1007/s13258-013-0064-x) contains supplementary material, which is available to authorized users.

W. SaiSug · K. Ukoskit (✉)
Department of Biotechnology, Faculty of Science and Technology, Thammasat University, (Rangsit Campus) Klong Luang, Pathumtani 12121, Thailand
e-mail: ku@tu.ac.th

W. SaiSug
e-mail: miniature_dachshund_11@hotmail.com

ESTs in various crops such as sugarcane (Ukoskit et al. 2012), *Medicago truncatula* (Eujayl et al. 2004), barley (Thiel et al. 2003), rubber (Li et al. 2012), cotton (Han et al. 2006), and *Capsicum* (Ince et al. 2010), and *J. curcas* (Kumar et al. 2011; Wen et al. 2010).

Introns are noncoding sequences that are widespread and abundant in eukaryotic genomes. Compared with exons, introns are more variable because selective pressure in intronic regions is much less strong than that in exonic regions (Choi et al. 2004). Various types of polymorphisms are seen in introns. For example, among eight varieties of rice, the average number of single nucleotide polymorphisms (SNPs) per 1,000 bp in introns is over three times higher than that in exons (Feltus et al. 2006). Among the different types of polymorphism, length polymorphism is the most sensitive type of variation in introns. Intron length polymorphism (ILP) is a codominant marker, which is conveniently detected by polymerase chain reaction (PCR), and is relatively easy to interpret and record. ILPs are comparable between different species (Feltus et al. 2006; Wang et al. 2006) because the exon–intron structures of genes are highly conserved across species, and often across higher taxonomic ranks (Rogozin et al. 2003; Roy and Gilbert 2005). A general method for identifying introns is to compare cDNA/EST sequences with the genomic sequences encoding them. Primers can be designed on adjacent exon regions to prime amplification across the intron by so-called exon-primed intron-crossing (EPIC) markers (Bierne et al. 2000). ILP has been exploited as a molecular marker in various plant species including rice (Wang et al. 2006; Ming et al. 2010), conifers (Chen et al. 2010), ryegrass (Tamura et al. 2009), *Rhododendron* (Wei et al. 2005), and rubber (Li et al. 2012).

With the recent progress made in large-scale plant functional genome sequencing projects, a substantial amount of EST and genomic sequence data are becoming available. Among the Euphorbiaceae, *Manihot esculenta* (cassava) has the largest database of EST and genomic sequences, and this information has shown a considerable degree of transferability to *J. curcas* (Wen et al. 2010). This transferability should facilitate the development of ILP markers in *J. curcas*, with primers designed in exons flanking the target intron, based on *M. esculenta* sequence data. In the present study, EST-SSRs were developed by mining SSR data from *J. curcas* EST sequences, and ILPs were developed by comparing EST sequences with the genomic sequence of *M. esculenta*. EST-SSR and ILP markers were compared for amplification and allelic variation among *J. curcas* accessions collected from different geographic regions in Thailand and from those introduced from other countries. Cross-species/genera transferability of the two marker types was examined among four species of the Euphorbiaceae family: *J. gossypifolia*, *M. esculenta*, *Hevea brasiliensis* (rubber tree), and *Ricinus communis* (castor bean). Cross-species/genera transferable markers will be valuable for comparative genome studies among these five species. These two sets of markers will not only increase the number of available markers in *J. curcas,* but also in *M. esculenta*, *H. brasiliensis,* and *R. communis,* serving to reduce the cost and time needed for marker development in at least four genera of this family.

Materials and methods

Plant materials

J. curcas germplasm was used for comparison of amplification and allelic variation determined by EST-SSRs and ILPs comprising 59 accessions, which included 42 accessions collected from four different geographic regions of Thailand (15, 6, 17, and 4 accessions from Northern, Northeastern, Central, and Southern Thailand, respectively), and 17 accessions from eight countries representing a wide geographic distribution (one accession each from Vietnam, Laos, Sri Lanka, and Senegal, two accessions each from USA and Mexico, three accessions from India, and six accessions from Myanmar). The transferability of EST-SSR and ILP markers for cross-species analysis within Euphorbiaceae was evaluated using one accession of *J. gossypifolia* and two accessions each of *M. esculenta*, *H. brasiliensis,* and *R. communis*. To assess phylogenetic relationships, five accessions of *J. curcas* representing three accessions from Asia (Thailand, Laos, and India) and one accession each from Mexico and Senegal were included in the analysis, in addition to the four species of Euphorbiaceae used in the transferability study. Leaf material from one individual per accession was collected and DNA was extracted following the cetyltrimethylammonium bromide (CTAB) method described by Gawel and Jarret (1991).

Development of EST-SSR markers

A total of 6,996 PlantGDB-assembled Unique Transcript (PUT) sequences from *J. curcas* were collected from *Jatropha_curcas/175a* assembly results (assembled by PlantGDB; http://www.plantgdb.org/), and screened for mono- to hexamicrosatellite repeats containing more than 20 nucleotides using Websat (Martins et al. 2009). Orf-Predictor (Min et al. 2005) was used to predict the most probable untranslated region (5′-UTR), the protein-coding region (open reading frame; ORF), and the 3′-UTR. Primer3 software (Rozen and Skaletsky 2000) was used to design the primers to amplify the selected EST-SSRs based on the following parameters: primer length from 18 to

25 bp with 21 bp as the optimum; PCR product size from 100 to 400 bp; optimum annealing temperature of 59 °C; and GC content between 45 and 65 %, with 50 % as the optimum. PCR amplifications were carried out in 20-μL reaction volumes containing 1× Buffer, 3.0 mM $MgCl_2$, 400 μM dNTPs, 0.4 μM of each primer, 5 U *Taq* DNA polymerase, and 50 ng DNA template, and were performed in an MJ Thermal Cycler at 94 °C for 5 min, followed by 35 cycles of 30 s at 94 °C, 30 s at the appropriate annealing temperature (50–57 °C) for each primer pair, and 1 min at 72 °C, with a final extension of 7 min at 72 °C. PCR products were run on 6 % denaturing polyacrylamide gels and silver-stained as described by Benbouza et al. (2006).

Development of ILP markers

The genome sequence data for cassava were downloaded from Phytozome V5.0 (http://www.phytozome.net/). A total of 30,401 PUT sequences from cassava were downloaded from PlantGDB (http://www.plantgdb.org). Cassava was chosen for the study because it is phylogenetically closely related to *J. curcas*, and at the time of this study had the largest number of assembled EST sequences and was the most intensively studied species within Euphorbiaceae. The prediction of intron lengths and the positions and the design of ILP primers were performed following the method of Yang et al. (2007). The procedure was as follows: the coding sequences of cassava were aligned with their corresponding genome sequences using GeneSeqer (Brendel et al. 2004) to acquire information on intron positions and lengths in this species. A pair of specific primers flanking each of the predicted intron positions was designed using the Primer3 program. Since large intron fragments can be scored inaccurately for small differences in fragment size, only introns smaller than 1,000 bp were selected for primer synthesis and evaluation of length polymorphism. Amplification by PCR was performed in a 20-μl reaction mixture containing 50 ng template DNA, 1 × PCR buffer (20 mM Tris pH 9.0, 100 mM KCl, 3.0 mM MgCl2), 400 μM of each of the four dNTPs, 0.4 μM of each of the forward and reverse primers, and one unit of Taq DNA polymerase. The following PCR conditions were used: 94 °C for 1 min, followed by 35 cycles of 94 °C for 30 s, the appropriate annealing temperature (45–57 °C) for each primer pair for 30 s, 72 °C for 1 min, and 5 min at 72 °C for the final extension. PCR products were separated on a 2 % agarose gel, or 6 % on denaturing polyacrylamide gels and silver-stained as described by Benbouza et al. (2006). The PCR products, amplified from a highly polymorphic marker, were purified using QIAquick spin columns (QIAGEN USA, CA, USA) and directly sequenced from both directions using the dideoxy method to confirm that they originated from the target genes. The obtained genomic sequences were aligned with the corresponding cassava EST sequences using Clustal W (Larkin et al. 2007) to assess the PCR amplification specificity and determine the exon/intron structure.

Data analysis

The genotypic profiles generated by each set of markers were used to determine the genetic diversity parameters using PowerMarker (Liu and Muse 2004). Two matrices, one for each type of marker data set, were generated. Various parameters were calculated, including the number of genotypes (NG), the number of alleles (NA), gene diversity (GD), heterozygosity (H), and polymorphism information content (PIC). Pairwise genetic distances between accessions were calculated based on similarity coefficients following Nei and Li (1979). The resulting similarity matrices for each marker system were used to construct dendrograms using the Unweighted Pair Group Method with Arithmetic Averages (UPGMA). Comparison between EST-SSR and ILP data was performed by calculating the correlation between the two data sets using the mantel test with 1,000 permutations. All procedures were performed using the computer package, NTSYS-PC 1.8 (Rohlf 1998).

Results

Type and frequency of EST-SSRs

In the present study, 6,996 non-redundant sequences were searched for ESTs containing SSR motifs. A search using the default parameters of the Websat program for ESTs found 203 ESTs containing repeat motifs accounting for approximately 2.9 % of the total sequences searched. As a result, 251 SSR motifs were obtained. Analysis of SSR motifs revealed that mononucleotide motifs (79) were the most abundant type of SSRs, followed by tri- (66), and an equal number of hexa- and dinucleotide motifs (47 each) (Fig. 1). The tetra- and pentanucleotide motif sequences were much rarer (4 and 8, respectively). For the dinucleotide motif sequences, the TC motif was the most common (13) followed by CT (12) and AT (8) motifs, whereas the AC motif was the least common (1). Among the trinucleotide motif sequences, the CAG motif was the most common (7) followed by ATT (6). Of all of the EST-SSRs identified, 133 loci, accounting for 175 SSR motifs, had long enough flanking sequences for primer design and were selected for further study. Primer pairs for this complete set of EST-SSRs were compared for redundancy with those from other EST-SSR data mining research (Wen et al.

2010) and are shown in Supplementary Table 1. The SSRs comprised 57, 40, 49, 4, 7, and 18 mono-, di-, tri-, tetra-, penta-, and hexanucleotide motifs, respectively.

ExPASy was used to compare the distribution of the 133 EST-SSR markers in putative genic regions to determine whether the priming sites and repeat motifs were located in protein-coding sequences or in UTRs. Of the 133 markers, 28 had both of the primer sites (and thus the associated repeat motifs) located within coding sequence, whereas both primers (along with the associated repeat motifs) were located in a presumed UTR in nine of the markers (Table 1). The remaining 96 markers had one primer in a coding sequence and one primer in a UTR sequence. The SSRs themselves were located in a coding sequence in 97 markers and in an UTR in 36 markers; more frequently in the 5′ UTRs than in the 3′ UTRs (Larkin et al. 2007). Of the total SSR motifs, trinucleotide motifs (36.4 %) represented the most common type of motif in exons.

Introns predicted in *J. curcas*

The program GeneSeqer was used to align each PUT sequence with the genomic sequence of cassava to examine gene structure (number of introns and positions of splice sites) and putative ILPs. By aligning 17,021,699 bp of 30,401 PUT sequences from cassava with 386,403 bp of cassava genomic sequence, 99 PUTs were mapped to the cassava genome. Of these, 153 introns and 262 exons were identified, averaging 2.5 exons per PUT. The length of candidate ILP markers on the genomic sequences ranged from 47 to 9,369 bp, with an average of 614 bp. The distribution of intron lengths is shown in Fig. 2. Of the total predicted introns, 133 had long enough flanking exon sequences to successfully design ILP primers. Since large intron fragments can be scored inaccurately on the basis of small differences in fragment size, only introns smaller than 1,000 bp were selected for primer synthesis and for evaluation of length polymorphism. Of 133 designed ILP primers, 112 were synthesized and used to perform PCR on the genomic DNA of 59 *J. curcas* accessions. The characteristics of all of the ILP primer pairs are shown in Supplementary Table 2. For 77 markers, the priming sites of both primers appeared to be located within coding sequence, while the remaining 35 markers had one primer each in coding and UTR sequence (Table 1).

Table 1 Distribution of the priming sites for 133 EST-SSR and 112 ILP markers in putative genic regions

Forverse primer	Reverse primer	Number	
		EST-SSR	ILP
5′ UTR	5′ UTR	8	0
5′ UTR	Coding sequence	18	0
5′ UTR	Coding sequence	44	23
Coding sequence	Coding sequence	28	77
Coding sequence	3′ UTR	25	12
Coding sequence	3′ UTR	9	0
3′ UTR	3′ UTR	1	0

In the present study, *J. curcas* ILP primers were designed according to the exon/intron junction site information obtained by aligning cassava ESTs with the corresponding genomic sequences. To test the specificity of the genomic PCR in this study, the PCR products generated by the highest polymorphic primer (IF7; PIC = 0.58), which amplified three variant alleles, were sequenced. Sequencing of the selected genomic PCR bands indicated precise sizes of 425, 543, and 581 bp (Fig. 3). The gel-based size estimation of the PCR bands was comparable with that derived from sequencing analysis (data not shown). Sequence alignments showed that priming sites were conserved, and revealed the presence of variable-size indels

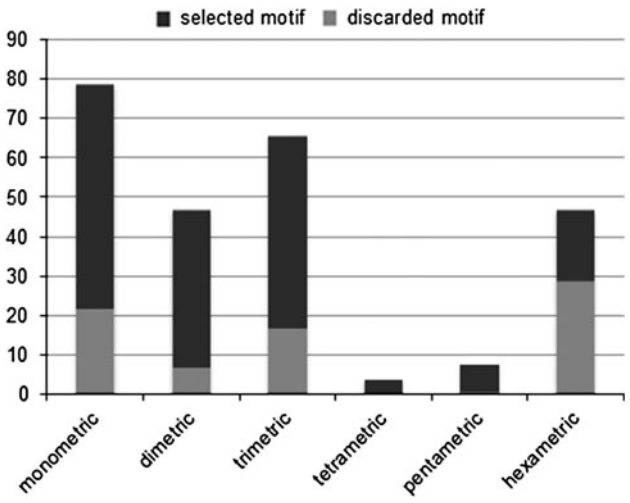

Fig. 1 Frequency of repeat motif types, number of selected SSR motifs with long enough flanking sequences for primer design, and number of discarded SSR motifs

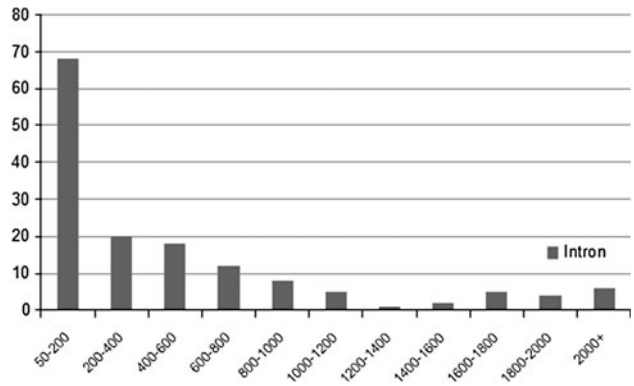

Fig. 2 Distribution of lengths (bp) for the 153 introns identified

Table 2 Percentage rates of PCR amplification for 133 EST-SSR and 112 ILP primers as determined from 59 *J. curcas* accessions

	EST-SSR	ILP
PCR success rate	97.7	67.0
PCR failure rate	2.3	33.0
Scorable PCR product	60.9	63.4
Multiple band PCR product	36.8	3.6
Polymorphism rate	53.4	50.9

and several point mutations in the intron. This result confirmed that the ESTs were indeed derived from the amplified genomic sequences, and that the PCR specificity was satisfactory.

Amplification of EST-SSRs and ILPs

All 133 EST-SSR primer pairs were determined using genomic DNA from 59 *J. curcas* accessions to assess their amplification and polymorphism in different *J. curcas* genotypes. Amplification of expected-size, scorable PCR products (one or two clear bands) was observed with 60.9 % (81/133) of these primer pairs (Table 2). Of these primer pairs, 2.3 % (3/133) did not amplify any bands, and 36.8 % (49/133) yielded multiple-band PCR products; these were discarded and not considered in further analyses. Of the primers pairs tested, 53.4 % (71/133) identified polymorphism among all of the *J. curcas* accessions examined.

Of the 112 ILP primer pairs tested, 63.4 % (71/112) showed scorable PCR products, with clear electrophoretic bands (Table 2). A higher PCR failure rate for ILP, (15.2 %; 17/112) compared with EST-SSR (2.3 %), was observed. This was because the ILP primers were designed indirectly from *M. esculenta* sequences. Of the total primers tested, 3.6 % (4/112) generated multiple bands. Since multiple-copy sequences are not desirable for use as molecular markers, these primer pairs were discarded. Of the primer pairs tested, 50.9 % (57/112) showed polymorphism within 59 *J. curcas* accessions. The estimated intron length determined from the 71 primer pairs varied from 153 to 481 bp, with an average of 284 bp. Most (88.73 %) of the introns were shorter than 400 bp.

Allelic variation of EST-SSRs and ILPs

To make comparisons between the levels of polymorphism detected by each marker system, both sets of data were analyzed as genotypes. A total of 238 alleles corresponding to 243 genotypes were identified by 71 polymorphic EST-SSR markers (Table 3) generated from 59 *J. curcas* accessions. The number of alleles and genotypes ranged from 2 to 6 and 2 to 5, with an average of 3.01 ± 1.01 and 3.08 ± 0.94, respectively. PIC values revealed low to medium allelic variation with values ranging from 0.04 to 0.54, averaging 0.37 ± 0.08. The gene diversity ranged from 0.04 to 0.61 with an average of 0.47 ± 0.01. For the ILP markers, a total of 152 alleles corresponding to 164 genotypes were generated by 57 polymorphic ILP markers (Table 3) among 59 *J. curcas* accessions. The number of alleles and genotypes varied from 2 to 5 and 2 to 7, with an average of 2.67 ± 0.81 and 2.88 ± 1.05, respectively. The PIC values varied from 0.06 to 0.58, with an average of 0.35 ± 0.10. The gene diversity varied from 0.07 to 0.65, with an average of 0.44 ± 0.14. Although more EST-SSR primers generated alleles than ILP primers, a similar level of allelic variation of ILP and EST-SSR markers was observed.

Cross-species/genera transferability of EST-SSRs and ILPs

The 81 EST-SSR and 71 ILP primers amplifying *J. curcas* genomic DNA with scorable PCR products were assessed for cross-species/genera transferability. The amplification frequency across *J. gossypifolia*, *M. esculenta*, *H. brasiliensis,* and *R. communis* revealed that 76.5 (62/81), 54.3 (44/81), 49.4 (40/81), and 25.9 % (21/81) of EST-SSR primers could amplify DNA with scorable PCR product in each of the four species, respectively, with an average of 51.5 % (Table 4). While 63.4 % (45/71) of ILP primers tested gave scorable amplifications in *J. gossypifolia*, they yielded 95.8 % (68/71) in *M. esculenta*, 93.0 % (66/71) in *H. brasiliensis*, and 69.0 % (49/71) in *R. communis*, with an average of 80.3 %. Of all the EST-SSR markers tested, 12.3 (10/81), 12.3 (10/81), 6.2 (5/81), and 46.9 % (38/81) failed to amplify genomic DNA from *J. gossypifolia*, *M. esculenta*, *H. brasiliensis,* and *R. communis*, respectively, with an average of 19.4 %. Of the ILP primers, 31.0 % (22/71) failed to yield any PCR products in *J. gossypifolia*, 2.8 % (2/71) in *M. esculenta*, 4.2 % (3/71) in *H. brasiliensis*, and 14.1 % (10/71) in *R. communis*, with an average of 13.0 %. Approximately 11.1 (9/81), 33.3 (27/81), 44.4 (36/81), and 27.2 % (22/81) of EST-SSR primers produced multiple-band PCR products in *J. gossypifolia*, *M. esculenta*, *H. brasiliensis,* and *R. communis*, respectively, with an average of 29.0 %. For the ILP markers, multiple-band PCR products were observed in 5.6 % (4/71) of *J. gossypifolia* sequences, 1.4 % (1/71) of *M. esculenta* sequences, 2.8 % (2/71) of *H. brasiliensis* sequences, and 16.9 % (12/71) of *R. communis* sequences, with an average of 6.7 %. Compared with EST-SSR primers, a higher average percentage of ILP primers successfully amplified products, while a lower average percentage of those primers failed to yield amplification products, in cross-species/genera within Euphorbiaceae. This reflects the

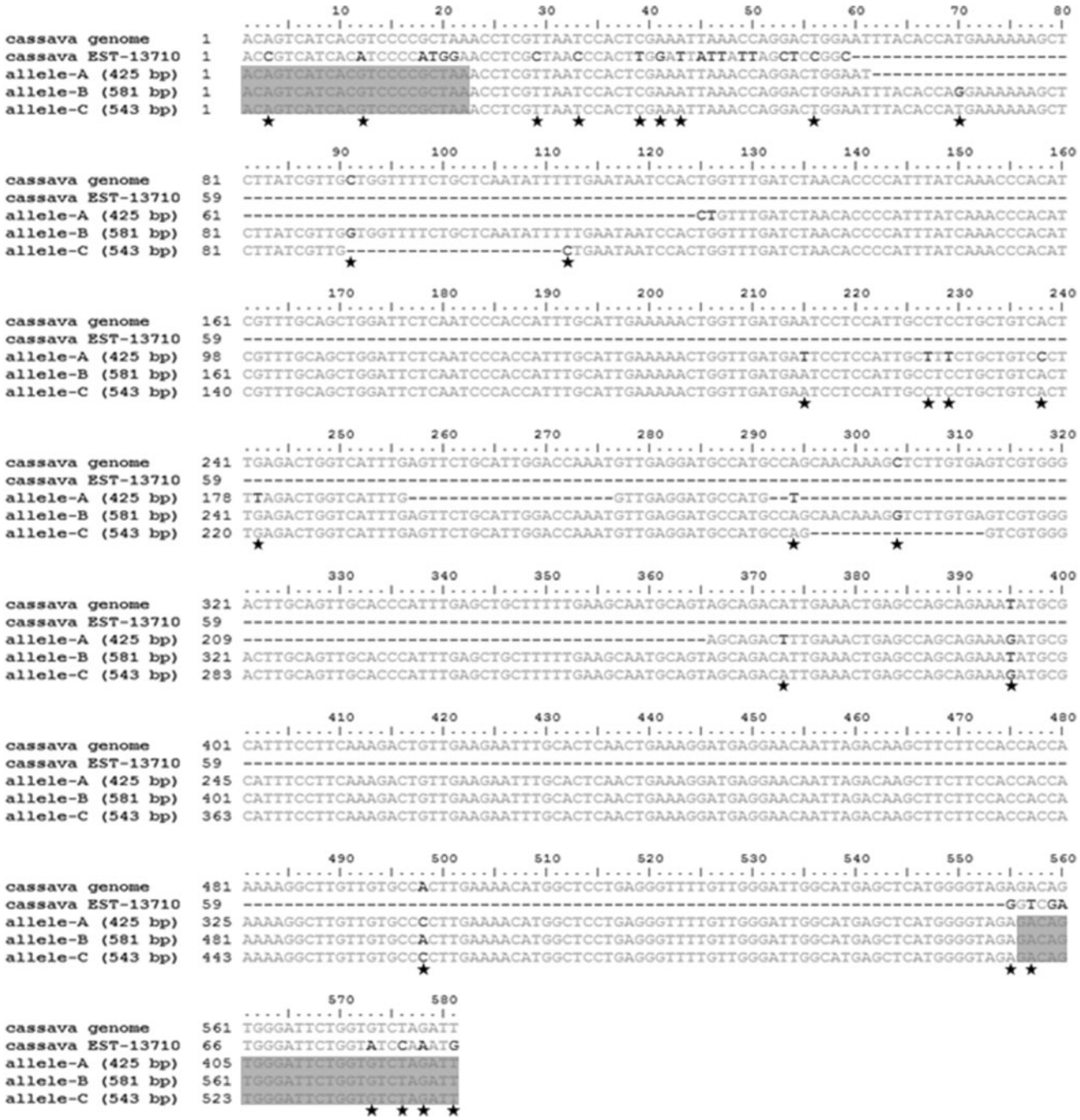

Fig. 3 Clustal W sequence alignment of the ILP alleles amplified using IF7 marker (PIC = 0.576) compared with the reference genomic and EST sequences from cassava. Primer sequences are *shaded*, single nucleotide polymorphisms (SNP) are in *bold* and indicated with *stars*, and *dashes* represent alignment gaps (indels)

more conserved nature of the ILP primer sequences compared with those of the EST-SSR primers.

Assessment of genetic relationships based on EST-SSR and ILP markers

Cluster analysis was performed based on Nei and Li (1979) similarity coefficient matrices, which were calculated based on scorable PCR products from 81 EST-SSR and 71 ILP markers to generate two separate phylogenetic trees of five species in Euphorbiaceae. Five accessions of *J. curcas,* representing a wide geographic distribution, were included in the analysis to compare the classifications based on EST-SSR and ILP markers both within and between *J. curcas* species. The similarity coefficient values based on EST-SSR markers ranged from 0.04 to 0.67, averaging 0.12.

Table 3 Allelic variation of 71 polymorphic EST-SSRs and 57 polymorphic ILPs evaluated from 59 *J. curcas* accessions

	EST-SSR				ILP			
	NG	NA	GD	PIC	NG	NA	GD	PIC
Mean	3.08	3.01	0.47	0.37	2.88	2.67	0.44	0.35
Max	5	6	0.61	0.54	7	5	0.65	0.58
Min	2	2	0.04	0.04	2	2	0.07	0.06
Total	243	238			164	152		

NG number of genotypes; NA number of allele; GD gene diversity; PIC Polymorphism information content

While the similarity coefficient values based on ILP markers differed in that they varied from 0 to 0.79, they had the same average of 0.12. In addition, Mantel's test showed a high correlation ($r = 0.90$, $P < 0.00$) between the two genetic similarity matrices revealed by the two marker systems.

The phylogenetic tree generated from EST-SSR (Fig. 4a) data separated the five species into two clusters. The first cluster included two species within the *Jatropha* genus. Within this genus, the five accessions of *J. curcas* obtained from distant sources of collection or origin were clustered into two subgroups that corresponded well to geographic distribution, while *J. gossypifolia* clearly fell outside the cluster of *J. curcas* accessions, with a genetic similarity coefficient of 0.15. Within *J. curcas*, the genetic similarity coefficient ranked from 0.07 to 0.38, averaging 0.25. A group of three accessions (Thai, Lao, and Indian) from Asia was separated from a group of two accessions from North America (Mexico) and Africa (Senegal), with similarity coefficients of 0.20. The second cluster, *R. communis*, *M. esculenta*, and *H. brasiliensis*, formed a monophyletic clade in which *M. esculenta* and *H. brasiliensis* were close to each other, while *R. communis* was placed more distantly, with a similarity coefficient of 0.06.

Based on the ILP marker data, three separate groups were revealed. The five accessions of *J. curcas* were clustered together, with genetic similarity coefficients ranging from 0.25 to 0.75 (averaging 0.45), and separated from *J. gossypifolia,* with a similarity coefficient of 0.20. The group of three *J. curcas* accessions from Asia was separated from the group of two accessions from North America (Mexico) and Africa (Senegal), with a similarity coefficient of 0.33. The ILP marker data showed a monophyletic clade composed of *M. esculenta* and *H. brasiliensis,* with a similarity coefficient of 0.03, and *R. communis* was placed in its own clade, with no shared genetic similarity with *M. esculenta* and *H. brasiliensis*. This result suggested that, compared with EST-SSRs, ILP markers showed stronger species specificity for the inference of genetic distance between species within the Euphorbiaceae.

Discussion

The discovery of SSR markers within ESTs provides the opportunity to develop microsatellites in a simple and direct way, by data mining of EST databases. Large collections of ESTs have become available for many plant species, including *J. curcas*. This report found the frequency of EST-SSRs to be at least one SSR motif per 16.3 kb in the approximately 4.3 Mb of ESTs that were searched. This EST-SSR frequency was lower than the genomic SSR frequency analyzed by Sato et al. (Sato et al. 2010). Data mining of EST-SSRs in cotton and wheat showed similar values, with one SSR for every 20.0 (Cardle et al. 2000) and 15.6 kb (Kantety et al. 2002), respectively. This number is lower than that for other species within the Euphorbiaceae, such as *H. brasiliensis*, *M. esculenta*, and *R. communis*, which have values of one SSR for every 2.25 kb (Li et al. 2012), 10.3 kb (Raji et al. 2009), and 680 bp (Sharma and Chauhan 2011),

Table 4 Percentage of PCR amplifications and cross species/genera transferability of 81 EST-SSR and 71 ILP markers for four species of Euphorbiaceae

	J. gossypifolia	*M. esculenta*	*H. brasiliensis*	*R. communis*	Average
EST-SSR					
PCR success rate	97.5	87.7	93.8	53.1	83.0
PCR failure rate	2.5	12.3	6.2	46.9	17.0
Scorable PCR product	76.5	54.3	49.4	25.9	51.5
Multiple bands PCR product	21.0	33.4	44.4	27.2	31.5
Polymorphism rate	ND	38.3	38.3	25.9	34.2
ILP					
PCR success rate	69.0	97.2	95.8	85.9	87.0
PCR failure rate	31.0	2.8	4.2	14.1	13.0
Scorable PCR product	63.4	95.8	93.0	69.0	80.3
Multiple bands PCR product	5.6	1.4	2.8	16.9	6.7
Polymorphism rate	ND	18.3	31.0	16.9	22.1

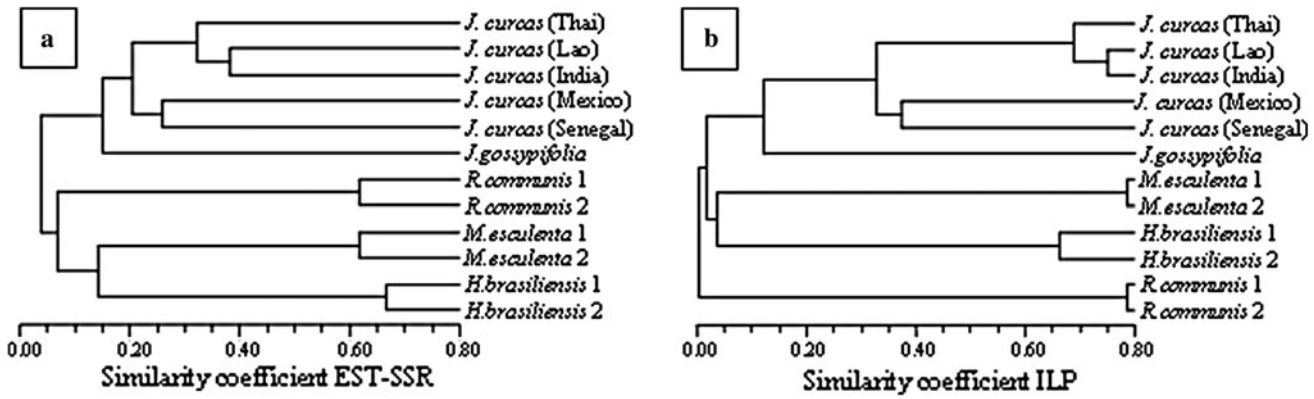

Fig. 4 UPGMA cluster analysis, based on Nei's genetic similarity derived from **a** EST-SSR markers and **b** ILP markers, showing the phylogenetic relationships between five accessions of *J. curcas* and five species of Euphorbiaceae

respectively. Since the criteria for SSR data mining depend on the parameters of the SSR search algorithm (for example, variation in sample sizes, search criteria, size of the database, and tools used for EST-SSR development (Varshney et al. 2005; Ukoskit et al. 2012), exact comparison of SSR characteristics between different plants reported in different studies is complicated.

Aside from mononucleotide repeats, the trinucleotide motif was the most common, followed by hexanucleotide motifs. This was also found in previous reports in *J. curcas* (Wen et al. 2010), citrus (Chen et al. 2006), barley (Thiel et al. 2003), and most dicotyledonous species (Raji et al. 2009). Additionally, trinucleotide motifs were found most frequently in coding regions. The dominance of tri- and hexanucleotide motifs in this study could be explained by the suppression of non-trinucleotide motifs in the coding regions because of the risk of frame shift mutations, which arise when motifs are not composed of multiples of three nucleotides (Thiel et al. 2003; Varshney et al. 2005). The distribution of EST-SSR markers in different putative genic regions of *J. curcas* revealed that a greater proportion of the EST-SSRs were located in the 5′ untranslated region rather than in the coding region, which was analogous to genomic distributions of SSRs in castor bean (Sharma and Chauhan 2011), *Arabidopsis* (Morgante 2002), turnip (Hong et al. 2007), and rice (Lawson and Zhang 2006). Since the *J. curcas* EST dataset used in this study was sequenced from the 5′ end, a greater proportion of the EST-SSRs were located in the 5′ UTR rather than in the coding region, as predicted by OrfPredictor. A computational (Fujimori et al. 2003) and an experimental (Ayres et al. 1997; Bao et al. 2002) study suggested that microsatellites located at high frequency in the 5′-flanking regions of plant genes can potentially function as factors in regulating gene expression.

ILP is another type of codominant molecular marker that can be conveniently detected by PCR with a pair of primers designed on flanking exons. The conservation of exon–intron structures allows the prediction of intron positions in other plants by referring to the genomes of model plants and, therefore, the development of ILP markers by designing primers on exons flanking the target introns. In this study, comparison of the EST sequences with the genome sequence of an evolutionarily close relative within the Euphorbiaceae, *M. esculenta*, enabled the design of 112 primer pairs for ILP markers in *J. curcas*. This primer design strategy explicitly focused on single-copy genes due to complexities generated by the presence of paralogous copies. Sequencing revealed the presence of variable-size indels and several point mutations in one intron. It is expected that the polymorphism-generating efficiency of EST-specific markers can be further enhanced using restriction enzyme digestion, heteroduplex analysis, denaturing gradient gel electrophoresis, or single-stranded conformational polymorphism (SSCP) gels (Zhang and Hewitt 2003).

A comparison between the two types of markers is of great importance. Recently, the accessibility of EST data has resulted in greater emphasis on EST-SSRs, which are derived directly from transcribed regions of the genome and are expected to be more conserved. They are therefore expected to have a higher transferability rate across species than genomic SSR (gSSR) markers. Several genetic analyses have been conducted to compare the efficiency of EST-SSR versus gSSR markers (Wen et al. 2010; Gadaleta et al. 2011). EST-SSR markers are reported to have a lower rate of polymorphism compared with the gSSR markers in many species such as rice, wheat, and *J. curcas* (Cho et al. 2000; Eujayl et al. 2001; Wen et al. 2010). Comparisons between EST-SSR and conserved-intron scanning primers (CISP) markers in grass species (Zeid et al. 2010) and between gSSR and ILP markers in rice (Wang et al. 2006) show that ILP markers are less polymorphic (based on fragment size differentiation) compared with SSR markers.

The results of the present study showed that similar allelic variation was observed in ILP and EST-SSR markers on polyacrylamide gels. EST-SSR markers detect polymorphism based on variations in the microsatellite repeat length and/or indels in transcribed regions (Varshney et al. 2005). By contrast, ILP markers are designed based on exon sequences flanking at least one intron region. Introns have less evolutionary constraints than exons, and are likely to be selectively neutral (Lessa, 1992), which makes the identification of polymorphisms that are useful for the analysis of population genetic structure more likely. In addition, a number of reports suggest that recombination rates have a strong influence on intron length (Comeron and Kreitman 2000; Lynch 2002). Here, we report the first set of ILP markers in *J. curcas* and show that the ILP marker is a valid alternative EST-derived marker in *J. curcas* with a similar rate of polymorphism to that of the EST-SSR marker.

Amplification of multiple loci using EST-SSR markers has been reported in previous studies (Varshney et al. 2005; Sim et al. 2009). In the present study, more primers generating multiple bands of PCR product were observed with EST-SSRs (36.8 %) than with ILPs (3.6 %) after PCR optimization. Amplification of more than one locus using EST-SSR primers was attributed to the possible amplification of orthologous and paralogous copies of the target region (Zeid et al. 2010). For the ILP primers, a single locus was the norm, and only rarely were two or more loci observed. This result was due to a greater tendency towards loss of specificity of the EST-SSR primers for the targeted sequence compared with ILPs. The development of EST-derived markers in the present study shows that a greater proportion of the EST-SSR primers (78.9 %) were located in 5′ or 3′ UTRs rather than in coding regions (21.1 %), while most ILP markers (68.6 %) had priming sites for both primers located within coding sequence. Previous studies reported that UTRs have a very broad distribution of nucleotide substitution rates, with some regions so highly divergent that, even between relatively close organisms, coding regions are highly conserved (Jareborg et al. 1999; Pesole et al. 2001). The tendency towards single locus amplification was a distinctive character of ILP primers compared with the EST-SSR primers tested here.

One way to obtain EST-derived markers for specific genes and genomic regions in a given species is to use primers developed from the sequence of the closest relative with a sequenced genome. However, the cross-amplification success of EST-derived markers varies strongly between species, and shows a highly significant negative relationship with genetic distance and amplification success. This is attributed to sequence conservation at the primer binding sites (Raji et al. 2009; Zeid et al. 2010). By comparing the two types of markers for PCR amplification and cross species/genera transferability, ILP markers, for which both primers were found in a coding sequence, were significantly more transferable in all four species tested. On average, a scorable PCR product was obtained with ILP markers in 80.3 % of cross-species amplifications (Table 4) compared with EST-SSR markers, for which one or both primers were located in the UTRs and which yielded an average scorable PCR product in 51.5 % of cross-species amplifications. These results are in agreement with a previous report (Fraser et al. 2005) that identified less sequence homology in UTRs and more homology in coding regions in related species. Many reports suggest that EST-SSRs are interesting because of their amplification of conserved sequences in related species (Varshney et al. 2005; Feltus et al. 2006). Others have observed losses of sequence homology when markers developed from one species were screened in distantly-related species (Asp et al. 2007; Sim et al. 2009). The remarkably high cross-amplification rate of ILPs observed in the present study adds another distinctive characteristic to these markers besides their tendency towards single locus amplification.

In general, the more evolutionarily distant the taxa, the less success of cross amplification (Reddy et al. 2010; Zeid et al. 2010). The potential for *J. curcas*-derived EST-SSR primers to amplify across species, expressed as successful amplification with a scorable PCR product, was highest in *J. gossypifolia,* (76.5 %), moderate in *M. esculenta* (54.3 %) and *H. brasiliensis* (49.4 %; a higher percentage than achieved in a previous report (Whankaew et al, 2011)), and relatively low in *R. communis* (25.9 %). These results suggest that *J. curcas* has a more distant relationship with *R. communis* than with *H. brasiliensis* and *M. esculenta,* and an expectedly close relationship with *J. gossypifolia*, in agreement with the phylogenetic distances obtained by cluster analysis (Fig. 4a). For cassava-derived ILP primers, the amplification rate of *M. esculenta* (95.8 %) was almost equal to the transfer rate of *H. brasiliensis* (93.0 %), but decreased with *R. communis* (69.0 %) and *J. gossypifolia* (58.8 %). The high transfer rate of cassava-derived ILP primers to *H. brasiliensis* suggests that *M. esculenta* has a closer relationship with *H. brasiliensis* than with *R. communis* and *J. gossypifolia*, a suggestion supported by cluster analysis (Fig. 4b). Previous phylogenetic studies in Euphorbiaceae (Chase et al. 1993; Kenneth et al. 2005; Whankaew et al. 2011) also support the construction of phylogenetic trees calculated based on ILPs in this study.

Comparison of diversity estimates made using different marker systems in several plant species frequently revealed incongruent estimates for different types of markers (Nybom 2004; Woodhead et al. 2005). Even comparison between dendrograms revealed a significant correlation between similarity matrix data. In the present study, the

average similarity coefficient values based on EST-SSR markers (0.69) were lower than those based on ILPs (0.75), suggesting that EST-SSRs discriminate between same-species alleles more clearly than ILPs. This may be especially applicable when fingerprinting of closely related individuals or subgroups is required to identify accessions within Asia (similarity coefficients of 0.33 and 0.68 as determined by EST-SSRs and ILPs in this study, respectively). However, both dendrograms revealed expected genetic relationships between *J. curcas* accessions based on their geographic distributions, in agreement with a previous genetic diversity study based on 224 Inter Simple Sequence Repeat (ISSR) markers (Sirithunya and Ukoskit, 2010). These results suggested that EST-SSR markers are relatively better at detecting genetic differentiation within *J. curcas* germplasm collections than ILPs. Although the sensitivity of ILPs is lower than that of EST-SSRs within the *Jatropha* species, it is still high enough to allow the development of sufficient ILP markers for practical use without substantial difficulty. On the other hand, ILPs showed a broader range of similarity coefficient values (0 to 0.79) than EST-SSRs (0.04 to 0.67), and had larger proportions of transferable primers across Euphorbiaceae species and genera (Table 4). These results suggested that highly species-specific markers, e.g. the ILP indel-type markers developed in this study, seem to be more reliable for identifying a particular species. They therefore may be especially suitable for phylogenetic studies involving distant species or higher taxonomic ranks. This is in agreement with previous diversity studies reporting that ILPs are suitable for studying genetic variation between species, or at higher taxonomic ranks, particularly in the construction of phylogenetic trees of relatively distantly related organisms (Tamura et al. 2009; Chen et al. 2010).

In summary, EST-SSR and ILP markers developed from transcribed regions of *J. curcas* and cassava, respectively, yielded comparable results. Compared with EST-SSR markers, ILP markers had similar advantages, including the rate of polymorphism, ease of use, interpretation, and recording. However, the ILP markers had higher cross-amplification rates than the EST-SSR markers, with fewer multiple bands in the PCR product. These results suggest that ILP markers are more appropriate than EST-SSR markers for phylogenetic studies involving distant species or higher taxonomic ranks, while EST-SSRs are more efficient at discriminating between same-species alleles. EST-SSRs, in combination with the ILP markers developed in the present study, enriched the number of molecular markers not only in *J. curcas,* but also in *M. esculenta*, *H. brasiliensis,* and *R. communis*. Therefore they show strong potential for use in different applications, such as DNA fingerprinting, marker-assisted selection, and comparative genetic mapping, in all four species.

Acknowledgments The authors gratefully acknowledge the financial support provided by the Faculty of Science and Technology, Thammasat University.

References

Asp T, Frei UK, Didion T, Nielsen KK, Lubberstedt T (2007) Frequency, type, and distribution of EST-SSRs from three genotypes of Lolium perenne, and their conservation across orthologous sequences of Festuca arundinacea, *Brachypodium distachyon*, and *Oryza sativa*. BMC Plant Biol 7:36

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germ plasm. Theor Appl Genet 94:773–781

Bao JB, Corke HC, Sun MS (2002) Microsatellites in starch-synthesizing genes in relation to starch physicochemical properties in waxy rice (*Oryza sativa* L.). Theor Appl Genet 105:898–905

Benbouza H, Jacquemin JM, Baudoin T, Mergeai G (2006) Optimization of a reliable, fast, cheap and sensitive silver staining method to detect SSR markers in polyacrylamide gels. Biol Agron Soc Environ 10:77–81

Bierne N, Lehnert SA, Bédier E, Bonhomme F, Moore SS (2000) Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. Mol Ecol 9:233–235

Brendel V, Xing L, Zhu W (2004) Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genomic locus. Bioinformatics 20:1157–1169

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

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L et al (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann Missouri Bot Gard 80:528–580

Chen C, Zhou P, Choi Y, Huang S, Gmitter F (2006) Mining and characterizing microsatellites from citrus ESTs. Theor Appl Genet 112:1248–1257

Chen X, Zhang G, Wu W (2010) Investigation and utilization of intron length polymorphisms in conifers. New Forest 41:379–388

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Choi H-K, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O et al (2004) A sequence-based genetic map of medicago truncatula and comparison of marker colinearity with *M. sativa*. Genetics 166:1463–1502

Comeron JM, Kreitman M (2000) The correlation between intron length and recombination in drosophila. Dynamic equilibrium between mutational and selective forces. Genetics 156:1175–1190

Eujayl I, Sorrells M, Baum M, Wolters P, Powell W (2001) Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Euphytica 119:39–43

Eujayl I, Sledge M, Wang L, May G, Chekhovskiy K, Zwonitzer J, Mian M (2004) *Medicago truncatula* EST-SSRs reveal cross-

species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–422
Feltus FA, Singh HP, Lohithaswa HC, Schulze SR, Silva TD, Paterson AH (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. Plant Physiol 140:1183–1191
Fraser LG, McNeilage MA, Tsang GK, Harvey CF, De Silva HN (2005) Cross-species amplification of microsatellite loci within the dioecious, polyploidy genus *Actinidia* (Actinidiaceae). Theor Appl Genet 112:149–157
Fujimori S, Washio T, Higo K, Ohtomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S et al (2003) A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. FEBS Lett 554:17–22
Gadaleta A, Giancaspro A, Zacheo S, Nigro D, Giove SL, Colasuonno P, Blanco A (2011) Comparison of genomic and EST-derived SSR markers in phylogenetic analysis of wheat. Plant Genet Resour 9:243–246
Gawel N, Jarret R (1991) A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. Plant Mol Biol Rep 9:262–266
Han Z, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T (2006) Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. Theor Appl Genet 112:430–439
Hong CP, Piao ZY, Kang TW (2007) Genomic distribution of simple sequence repeats in *Brassica rapa*. Mol Cells 23:349–356
Ince A, Karaca M, Onus A (2010) Polymorphic microsatellite markers transferable across *Capsicum* species. Plant Mol Biol Rep 28:285–291
Jareborg N, Birney E, Durbin R (1999) Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. Genome Res 9:815–824
Kantety R, Rota M, Matthews D, Sorrells M (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510
Kenneth JW, Petra H, Mark MC (2005) Molecular phylogenetic analysis of uniovulate Euphorbiaceae *(Euphorbiaceae* sensu stricto*)* using plastid RBCL and TRNL-F DNA sequences. Am J Bot 92:1397–1420
Kumar YH, Ranjan A, Asif M, Mantri S, Sawant S, 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
Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al (2007) ClustalW and ClustalX version 2. Bioinformatics 23:2947–2948
Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biol 7:R14
Lessa EP (1992) Rapid surveying of DNA sequence variation in natural populations. Mol Biol Evol 9:323–330
Li D, Deng Z, Qin B, Liu X, Men Z (2012) De novo assembly and characterization of bark transcriptome using illumine sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). BMC Genomics 13:192
Liu K, Muse SV (2004) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129
Lynch M (2002) Intron evolution as a population-genetic process. Proc Natl Acad Sci USA 99:6118–6123
Martins WS, Lucas DC, Neves KF, Bertioli DJ (2009) WebSat: a web software for microsatellite marker development. Bioinformation 3:282–283
Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res 33:677–680
Ming H, Fang-min X, Li-yun C, Xiang-qian Z, Jojee L, Madonna D (2010) Comparative analysis of genetic diversity and structure in rice using ILP and SSR markers. Rice Sci 17:257–268
Morgante M (2002) Microsatellites are preferentially present with non- repetitive DNA in plant genomes. Nat Genet 30:194–200
Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273
Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 13:1143–1155
Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S (2001) Structural and functional features of eukaryotic mRNA untranslated regions. Gene 276:73–81
Raji A, Anderson J, Kolade O, Ugwu C, Dixon A, Ingelbrecht I (2009) Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms, and cross-taxa utility. BMC Plant Biol 9:118
Reddy MRK, Rathour R, Kumar N, Katoch P, Sharma TR (2010) Cross-genera legume SSR markers for analysis of genetic diversity in Lens species. Plant Breed 129:514–518
Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV (2003) Remarkable inter kingdom conservation of intron positions and massive, lineage specific intron loss and gain in eukaryotic evolution. Curr Biol 13:1512–1517
Rohlf FJ (1998) NTSYSpc: numerical taxonomy and multivariate analysis system, version 2.02. Exeter Software, Setauket
Roy SW, Gilbert W (2005) Rate of intron loss and gain: implications for early eukaryotic evolution. Proc Natl Acad Sci USA 102:5773–5778
Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386
Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N et al (2010) Sequence analysis of the genome of an oil-bearing tree *Jatropha curcas* L. DNA Res 18:65–76
Sharma A, Chauhan RS (2011) Repertoire of SSRs in the castor bean genome and their utilization in genetic diversity analysis in *Jatropha curcas*. Compt Funct Genomics. doi:10.1155/2011/286089
Sim SC, Yu JK, Yk Jo, Sorrells M, Jung G (2009) Transferability of cereal EST-SSR markers to ryegrass. Genome 52:431–437
Sirithunya P, Ukoskit K (2010) Population genetic structure and genetic diversity of *Jatropha curcas* germplasm as investigated by 5′-anchored simple sequence repeat primers. J Crop Sci Biotechnol 13:147–153
Tamura KI, Yonemaru JI, Hisano H, Kanamori H, King J, King I, Tase K, Sanada Y, Komatsu T, Yamada T (2009) Development of intron-flanking EST markers for the *Lolium Festuca* complex using rice genomic information. Theor Appl Genet 118: 1549–1560
Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST database for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422
Ukoskit K, Thipmongkolcharoen P, Chatwachirawong P (2012) Novel expressed sequence tag- simple sequence repeats (EST-SSR) markers characterized by new bioinformatic criteria reveal high genetic similarity in sugarcane (*Saccharum* spp.) breeding lines. Afr J Biotechnol 11:1337–1363
Varshney R, Graner A, Sorrells M (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Wang X, Zhao X, Zhu J, Wu W (2006) Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). DNA Res 12:417–427

Wei H, Fu Y, Arora R (2005) Intron-flanking EST–PCR markers: from genetic marker development to gene structure analysis in *Rhododendron*. Theor Appl Genet 111:1347–1356

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

Whankaew S, Kanjanawattanawong S, Phumichai C, Smith DR, Narangajavana J, Triwitayakorn K (2011) Cross-genera transferability of (simple sequence repeat) SSR markers among cassava (*Manihot esculenta* Crantz), rubber tree (*Hevea brasiliensis* Muell. Arg.) and physic nut (*Jatropha curcas* L.). Afri J Biotechnol 10:1768–1776

Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M, Powell W (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. Mol Ecol 14:1681–1695

Yang L, Jin G, Zhao X, Zheng Y, Xu Z, Wu W (2007) PIP: a database of potential intron polymorphism markers. Bioinformatics 23:2174–2177

Zeid M, Yu JK, Goldowitz I, Denton ME, Costich DE, Jayasuriya CT, Saha M, Elshire R, Benscher D, Breseghello F et al (2010) Cross-amplification of EST-derived markers among 16 grass species. F Crops Res 118:28–35

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Mol Ecol 12:563–584