GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Development and characterization of novel expressed sequence tag-derived simple sequence repeat markers in *Hevea brasiliensis* (rubber tree)

Z.W. An[1,2], Y.C. Li[1], Q.L. Zhai[1], L.L. Xie[1], Y.H. Zhao[3] and H.S. Huang[2]

[1]College of Agriculture, Hainan University, Haikou, Hainan, China
[2]State Centre for Rubber Breeding & Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan, China
[3]College of Life Sciences, Ludong University, Yantai, Shangdong, China

Corresponding author: H.S. Huang
E-mail: azwcn@126.com

Genet. Mol. Res. 12 (4): 5905-5910 (2013)
Received March 13, 2013
Accepted September 8, 2013
Published November 22, 2013
DOI http://dx.doi.org/10.4238/2013.November.22.18

ABSTRACT. Cultivated clones of *Hevea brasiliensis* have a narrow genetic base. In order to broaden the genetic base, it is first necessary to investigate the genetic diversity of wild populations. Expressed sequence tag-simple sequence repeat (EST-SSR) markers were developed to investigate the genetic diversity of *Hevea* populations. Four hundred and thirty microsatellites were identified and 148 primers were designed to amplify the loci. Twenty-nine primer pairs were synthesized and evaluated for their ability to detect genetic polymorphisms among 40 wild accessions of *H. brasiliensis*. Twenty-one of the 29 loci were polymorphic. The number of alleles per locus in the 40 accessions ranged from 2 to 7. H_O and H_E at each locus ranged from 0.0000 to 0.9000 and from 0.0000 to 0.8704, respectively. All 21 loci could amplify in *H. brasiliensis*, *H. pauciflora*, *H. nitida*, *H. spruceana*, and *H. camargoana*. The EST-SSR primers developed herein can be used in genetic diversity and structure studies in *H. brasiliensis*.

Key words: *Hevea brasiliensis*; Expressed sequence tags; Microsatellite; Diversity

INTRODUCTION

Hevea brasiliensis (rubber tree), the main source of natural rubber, is indigenous to the Amazon basin of South America. Today, the rubber tree is extensively cultivated in Southeast Asia, where it comprises more than 90% of the world's natural rubber production (Priyadarshan and Gonçalves, 2003). The cultivated clones of Southeast Asia and China were all derived from 22 seedlings surviving from seeds collected by Henry Wickham in 1876, which showed little genetic variation (Tan, 1987).

Expressed sequence tag-derived simple sequence repeat (EST-SSR) markers are powerful tools for many purposes, such as genetic diversity, linkage map, and marker-assisted selection, and can be used across a number of related species. EST-SSRs are developed from ESTs. Because mRNA expressions differ among different tissues or developmental phases, several different EST-SSRs can be identified. Although EST-SSR markers have been developed for rubber tree (Feng et al., 2009; Triwitayakorn et al., 2011; Li et al., 2012), novel SSR markers are still needed for rubber tree research. Rubber trees show little genetic variation (Tan, 1987), and therefore, a lot of SSR markers that are distributed throughout the rubber tree genome are needed to accurately analyze genetic structure, construct a high-density genetic map, or identify linkage markers, which can then facilitate the establishment of rubber tree improvement and conservation programs.

MATERIAL AND METHODS

Forty wild *H. brasiliensis* accessions (Table 1), originating from the Brazilian States of Acre (AC), Rondônia (RO), and Mato Grosso (MT), and four related species (*H. pauciflora*, *H. nitida*, *H. spruceana*, and *H. camargoana*) were selected from the State Rubber Tree Germplasm Repository of China. *Hevea* ESTs were obtained via the ENTREZ search tool of the EST database of the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/dbEST/). The EST-trimmer software (http://pgrc.ipk-gatersleben.de/misa/download/est_trimmer.pl) was used to remove poly-A or poly-T stretches of the 5'- or 3'-end until there were no $(T)_6$ or $(A)_6$ within the range of 50 bp. ESTs were clustered and assembled into unigenes, including contigs and singletons, using the CAP3 software (Huang and Madan, 1999). After pre-treatment, the SSRIT program (Temnykh et al., 2001) was used to search for di-, tri-, tetra-, penta-, and hexa-nucleotide motifs in rubber unigenes, with a minimum number of five repeats. Primers were designed to amplify these repeats using the Primer Premier 6.1 program (PREMIER Biosoft International, Palo Alto, CA, USA) if the microsatellite repeat sequences were longer than 20 bp.

Genomic DNA was extracted from leaf tissues using the protocol described by An and Huang (2005). Polymerase chain reactions (PCR) were carried out with a Mastercycler thermocycler (Eppendorf AG, Hamburg, Germany) in a 10-μL volume containing 20 ng template DNA, 0.2 μM of each primer, and 1X PCR MasterMix (TianGen Biotech, Beijing, China). PCR conditions were as follows: 94°C for 5 min followed by 35 cycles of 94°C for 1 min, 45 s at the annealing temperature of each primer pair (Table 1), and 72°C for 1 min, with a final extension for 7 min at 72°C. The products were electrophoresed on 6% sequencing polyacrylamide gels (BioRad Sequi-Gen GT, USA), and bands were visualized by silver nitrate staining (An et al., 2009).

Table 1. Origins of *Hevea brasiliensis* wild accessions used in this study.

No.	Accession ID	Origin	Longitude (W)	Latitude (S)	No.	Accession ID	Origin	Longitude (W)	Latitude (S)
1	AC/S/10 37/93	AC	68° 39'	09° 04'	21	RO/C/8 24/10	RO	62° 51'	08° 02'
2	AC/X/21 64/221	AC	68° 29'	10° 38'	22	RO/C/8 24/293	RO	62° 51'	08° 02'
3	AC/F/5 21/197	AC	70° 21'	08° 10'	23	RO/PB/2 3/82	RO	63° 34'	11° 30'
4	AC/S/12 42/494	AC	68° 39'	09° 04'	24	RO/JP/3 22/189	RO	61° 56'	10° 53'
5	AC/X/21 64/177	AC	68° 29'	10° 38'	25	RO/PB/2 3/267	RO	63° 34'	11° 30'
6	AC/AB/15-54/418	AC	69° 34'	10° 56'	26	RO/C/8 24/97	RO	62° 51'	08° 02'
7	AC/F/5 21/108	AC	70° 21'	08° 10'	27	RO/PB/2 3/184	RO	63° 34'	11° 30'
8	AC/F/5 21/203	AC	70° 21'	08° 10'	28	MT/IT/18 31/125	MT	56° 55'	12° 40'
9	AC/T/1 5/8	AC	70° 45'	08° 08'	29	MT/IT/16-34/5	MT	56° 55'	12° 40'
10	AC/T/2 4/67	AC	70° 45'	08° 08'	30	MT/C/2 10/54	MT	56° 55'	12° 40'
11	AC/T/1 5/130	AC	70° 45'	08° 08'	31	MT/C/2 10/8	MT	56° 55'	12° 40'
12	AC/F/5 21/220	AC	70° 21'	08° 10'	32	MT/C/9 15/33	MT	56° 55'	12° 40'
13	AC/F/5 21/100	AC	70° 21'	08° 10'	33	MT/C/4 7/153	MT	56° 55'	12° 40'
14	AC/T/2 4/85	AC	70° 45'	08° 08'	34	MT/C/2 10/155	MT	56° 55'	12° 40'
15	RO/CM/10-44/738	RO	64° 13'	12° 24'	35	MT/C/11 9/10	MT	56° 55'	12° 40'
16	RO/CM/10/44/683	RO	64° 13'	12° 24'	36	MT/C/11 9/1	MT	56° 55'	12° 40'
17	RO/J/5-33/38	RO	62° 28'	10° 25'	37	MT/C/2 10/3	MT	56° 55'	12° 40'
18	RO/OP/4 20/125	RO	62° 15'	10° 42'	38	MT/C/11 9/66	MT	56° 55'	12° 40'
19	RO/C/9 23/238	RO	62° 51'	08° 02'	39	MT/C/11 9/70	MT	56° 55'	12° 40'
20	RO/C/8 24/272	RO	62° 51'	08° 02'	40	MT/C/11 9/67	MT	56° 55'	12° 40'

AC = Acre; RO = Rondônia; MT = Mato Grosso.

The genetic statistics based on three populations were calculated using POPGENE (Yeh and Boyle, 1997), including the number of alleles, observed (H_O), and expected (H_E) heterozygosities.

RESULTS AND DISCUSSION

In total, 430 SSRs were identified from 3090 unigenes, with 66 unigenes containing more than one SSR and 287 unigenes containing only one SSR. The frequency of EST-SSRs observed in the *H. brasiliensis* transcriptome was 11.42%. A total of 148 primer sets were designed. To evaluate the efficiency of primer design and utilization, 29 of the 148 primers were randomly selected and synthesized from the Sangon Company (Shanghai, China), and the genetic diversity of 40 wild accessions were investigated. Among the 29 primer pairs, 8 failed to amplify, and the remaining 21 yielded polymorphism in 40 accessions (Table 2). H_O and H_E at each locus ranged from 0.0000 to 0.9000 and from 0.0000 to 0.8704, respectively (Table 3). The 21 EST sequences containing SSR loci were blasted against the GenBank non-redundant database using BLASTx (Altschul et al., 1997). Fifteen of the EST sequences showed significant similarities to known genes (Table 2).

The number of alleles per locus in the 40 wild accessions ranged from 2 to 7, with an average of 3.43, which exceeded the average of 3.12 alleles reported by Feng et al. (2009), but was below the average of 3.85 alleles reported by Triwitayakorn et al. (2011). To test the cross-species transferability, 21 primer pairs were used for amplification in five *Hevea* species: *H. brasiliensis*, *H. pauciflora*, *H. nitida*, *H. spruceana*, and *H. camargoana.* The expected PCR products of 21 primer pairs were obtained from all five species (Figure 1).

The primers developed in this study are powerful and suitable for evaluating the genetic diversity and structure of rubber tree populations. These newly developed EST-SSRs are an additional resource for the genetic characterization of *H. brasiliensis*.

Table 2. Characteristics of 21 EST-SSRs developed in *Hevea brasiliensis*.

Locus	Primer sequence (5'-3')	Repeat	Size (bp)	Ta (°C)	GenBank	GenBank BLASTx analysis		
						Putative function	GenBank No.	E-value
HESR020	F: GACTGGAGACGGAAAACCAA R: CCGCGTCTGGTAAAATGAAT	$(AG)_{23}$	231	47.2	EC607220	NC domain-containing protein	XP_002873253.1	4e-103
HESR023	F: GCCACCGTCTGCTTCTCTAC R: CTACTCCGATGGTGGTGGTT	$(CT)_{10}$	273	55.6	EC607600	Conserved hypothetical protein	XP_002531462.1	2e-27
HESR026	F: GGAAGGTGCTAGATCGCCTA R: CACAGCAAGTTTTACATCACACG	$(CT)_{15}(AT)_{19}$	250	51.3	EC609591	Heat-shock protein	XP_002512004.1	5e-51
HESR029	F: GGAGTGGCCGATGATATGAG R: AAGTGGGAAATACAAATGGACA	$(AAT)_{14}$	250	53.4	EC609279	Conserved hypothetical protein	XP_002510159.1	6e-26
HESR030	F: GGCACCTGAAAAATTGGAAA R: CTGCCACCAAAACGAAAAAT	$(TA)_{12}$	159	51.3	EC605483	No hits found	-	-
HESR031	F: GGCCAATGGATTCTCTCTCA R: CAAAACCCAAAAAGGAACGA	$(CT)_{16}$	137	51.3	EC607490	Predicted protein	XP_002309041.1	9e-37
HESR032	F: GGTGCTGGACCCAGTTGTTA R: GCCAATGCTCCTTCAATCTC	$(AG)_{12}$	191	53.4	EC608363	No hits found	-	-
HESR033	F: GTGATCGGCGTGGTAATCTT R: CCAAATCGTCCAAACGCTAT	$(TC)_{13}$	206	53.4	EC608073	No hits found	-	-
HESR034	F: GTGCCGCTTCTCCATACTTA R: TCACAGCCATCACAGTCCAT	$(CT)_{11}$	202	55.6	EC606744	Metal ion binding protein	XP_002531065.1	9e-74
HESR041	F: TCTCCATTTTTCCCCTTTT R: AGGCAAGGGTAATGCAAATG	$(TC)_{14}$	249	49.5	EC607442	No hits found	-	-
HESR042	F: TGGGAAGGTGCTGGAATTAG R: TGCATGTGTGTGGGTTTAGG	$(AAT)_{19}$	159	53.4	EC608637	Chaperone protein dnaJ 8	XP_002512737.1	4e-85
HESR043	F: TTCACACGCATGTTTTCCTC R: GGGCTTTGCTCAGAAATCAG	$(TC)_{10}$	162	51.3	EC608652	No hits found	-	-
HESR045	F: TTGATGGCCAATTATCAGCA R: GTTGCCTTCATCTTCTGCAA	$(AAG)_{10}$	212	49.5	EC607429	Conserved hypothetical protein	XP_002522897.1	9e-20
HESR046	F: TTGATGGCCAATTATCAGCA R: CTTTCTTCGTCGTTGCCTTC	$(AAG)_{10}$	223	57.8	EC605124	Hypothetical protein VITISV_001461	CAN63134.1	3e-17
HESR048	F: TTGTCTCCATCTTTCCAGCTT R: AGGCAGCTGATTGTTTCTGC	$(AT)_{18}$	213	49.5	EC607346	Conserved hypothetical protein	XP_002515051.1	8e-11
HESR049	F: TTTCATTTCCAGTAAAACCCAAA R: CTCCCTCCATCTTGAACCAA	$(TC)_{14}$	179	49.5	EC608764	Probable inositol transporter 2 isoform2	XP_002278770.1	2e-45
HESR050	F: TGCTGTTGGAGGAAGA R: CCAGGGACTTTGTTAGAC	$(CTT)_{13}$	345	51.3	EC609907	Lectin-domain containing receptor kinase A4.2	XP_003612439.1	7e-27
HESR051	F: GCAGTATCAACAACAGCCACC R: TCCCACCACGACCACGT	$(GCA)_8$	163	47.2	EC609720	No hits found	-	-
HESR053	F: AAGGGTCAGTCAGGTGG R: TTCCCAGAAGACCAATC	$(TCTA)_5$	159	51.3	EC608753	Catalytic	XP_002523793.1	7e-86
HESR054	F: TTGCTTGATGGTGGGT R: AATGGGTCCTTGTCCTT	$(TGA)_8$	242	51.3	EC608668	Transcription factor bHLH143	XP_003593247.1	3e-16
HESR055	F: TTGCTGAGATGTGATG R: GAGGTTTAACTGCTAAG	$(GAA)_9$	316	55.6	EC606684	Anthranilate phosphoribosyltransferase	XP_002527915.1	2e-27

Ta = optimal annealing temperature.

Table 3. Results of initial primer screening in three populations of *Hevea brasiliensis*.

Locus	AC (N = 14)			RO (N = 13)			MT (N = 13)		
	N_A	H_O	H_E	N_A	H_O	H_E	N_A	H_O	H_E
HESR020	2	0.2143	0.3042	2	0.4615	0.3692	2	0.5000	0.4524
HESR023	2	0.0000	0.2540	2	0.1538	0.5169	2	0.0000	0.2540
HESR026	4	0.4286	0.4683	4	0.6154	0.5169	4	0.5000	0.5476
HESR029	7	0.6429	0.8333	6	0.6154	0.8369	7	0.7857	0.8704
HESR030	3	0.7857	0.6032	3	0.3846	0.5415	3	0.5714	0.5397
HESR031	2	0.1429	0.1376	2	0.2308	0.2123	2	0.2857	0.2540
HESR032	6	0.7500	0.7572	6	0.9000	0.8474	4	0.9000	0.7737
HESR033	3	0.1429	0.3704	4	0.3846	0.5446	3	0.4286	0.5185
HESR034	4	0.6000	0.7842	4	0.6250	0.7917	4	0.7778	0.7647
HESR041	3	0.5000	0.4153	1	0.0000	0.0000	2	0.4286	0.3492
HESR042	2	0.2857	0.2540	2	0.3077	0.2708	2	0.0714	0.0714
HESR043	4	0.5714	0.5476	4	0.6923	0.6277	4	0.5714	0.5899
HESR045	2	0.6429	0.4947	2	0.6154	0.5169	2	0.4286	0.5079
HESR046	4	0.7143	0.7434	3	0.6923	0.5138	2	0.6429	0.4524
HESR048	4	0.4286	0.6667	2	0.2308	0.5200	4	0.7857	0.6958
HESR049	3	0.3571	0.4206	2	0.2308	0.2123	1	0.0000	0.0000
HESR050	3	0.2143	0.4550	2	0.0000	0.3692	1	0.0000	0.0000
HESR051	3	0.4286	0.6376	3	0.0769	0.6923	3	0.1429	0.6640
HESR053	2	0.3571	0.4524	3	0.1538	0.5815	3	0.2143	0.3148
HESR054	3	0.0714	0.3148	2	0.0000	0.3692	2	0.0714	0.1984
HESR055	2	0.2143	0.3889	2	0.3077	0.2708	2	0.1429	0.2540

N_A = number of alleles; H_E = expected heterozygosity; H_O = observed heterozygosity; N = individuals of population; AC = Acre; RO = Rondônia; MT = Mato Grosso.

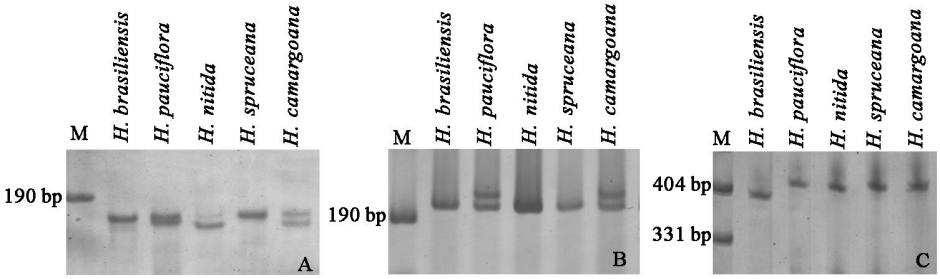

Figure 1. PCR amplification patterns showing polymorphism and transferability of EST-SSRs. **A.** Primer HESR030 amplification pattern in 5 species. **B.** Primer HESR045 amplification pattern in 5 species. **C.** Primer HESR050 amplification pattern in 5 species. *Lane M* = a pUC19 DNA/*Msp*I (*Hpa*II) marker.

ACKNOWLEDGMENTS

Research partially supported by the National Natural Science Foundation of China (#30860221, #31271796), the Fundamental Research Funds of the Rubber Research Institute, CATAS (#1630022011002), and the Earmarked Fund for Modern Agro-Industry Technology Research System (#CARS-34).

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.

An ZW and Huang HS (2005). A method for genomic DNA extraction from leaves of rubber tree (*Hevea brasiliensis* Müell. Arg.). *Plant Physiol. Commun.* 41: 513-515.

An ZW, Xie LL, Cheng H, Zhou Y, et al. (2009). A silver staining procedure for nucleic acids in polyacrylamide gels without fixation and pretreatment. *Anal. Biochem.* 391: 77-79.

Feng SP, Li WG, Huang HS, Wang JY, et al. (2009). Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). *Mol. Breed.* 23: 85-97.

Huang X and Madan A (1999). CAP3: A DNA sequence assembly program. *Genome Res.* 9: 868-877.

Li D, Deng Z, Qin B, Liu X, et al. (2012). *De novo* assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *BMC Genomics* 13: 192.

Priyadarshan PM and Gonçalves PS (2003). *Hevea* gene pool for breeding. *Genet. Resour. Crop Evol.* 50: 101-114.

Tan H (1987). Strategies in Rubber Tree Breeding. In: Improving Vegetatively Propagated Crops (Abbott AJ and Atkin RK, eds.). Academic Press, London, 27-62.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, et al. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 11: 1441-1452.

Triwitayakorn K, Chatkulkawin P, Kanjanawattanawong S, Sraphet S, et al. (2011). Transcriptome sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. *DNA Res.* 18: 471-482.

Yeh FC and Boyle TJB (1997). Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian J. Bot.* 129: 157.

