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

Polymorphisms in different EST-SSR types derived from the Chinese bayberry (*Myrica rubra*, Myricaceae) transcriptome

S.Y. Zhang[1,2], C. Feng[1,3], C.J. Xu[1], C.Q. Zhu[1] and K.S. Chen[1]

[1]Laboratory of Fruit Quality Biology/The State Agriculture Ministry Laboratory of Horticultural Plant Growth, Development, and Quality Improvement, Zhejiang University, Zijingang Campus, Hangzhou, China
[2]Key Laboratory of Special Wildlife Resources in Fanjing Mountain, Research Center for Conservation and Utilization of Wildlife Resources in Fanjing Mountain, College of Biology and Agro-forestry Engineering, Tongren University, Tongren, China
[3]Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

Corresponding author: C.Q. Zhu
E-mail: zcq1236@zju.edu.cn

Genet. Mol. Res. 14 (2): 6037-6041 (2015)
Received September 30, 2014
Accepted February 26, 2015
Published June 1, 2015
DOI http://dx.doi.org/10.4238/2015.June.1.21

ABSTRACT. Most plant expressed sequence tag-simple sequence repeats (EST-SSRs) are not polymorphic, and it is important to learn the characteristics of highly polymorphic EST-SSRs. In this study, 357 compound and 5557 non-compound EST-SSRs, identified from the transcriptome of the Chinese bayberry (*Myrica rubra* 'Biqi'), were divided into 11 types based on their characteristics. Polymorphisms in all 11 EST-SSR types were investigated in 10 cultivars. The percentages of polymorphic loci ranged from 12.9 to 87.5%, with 2-ntL having the highest, followed by 3-ntL, Compound B, and Compound A. The

number of alleles and the polymorphic information content of 2-ntL and Compound B were the highest, followed by 2-ntM and Compound A. Therefore, we recommend that 2-ntL, Compound B, and Compound A EST-SSRs should be preferentially selected for the screening of polymorphic EST-SSRs in the Chinese bayberry. Our results should facilitate genetic and breeding studies of this species, and provide a reference for similar study in other plant species.

Key words: Chinese bayberry; *Myrica rubra*; EST-SSR; Polymorphism; RNA-Seq

INTRODUCTION

RNA sequencing (RNA-Seq) has been widely used to develop expressed sequence tag-simple sequence repeats (EST-SSRs), and can generate thousands of EST-SSRs in a plant (Zalapa et al., 2012). However, only about 10 to 30% of plant EST-SSRs are polymorphic (Yi et al., 2006; Dutta et al., 2011), and it is usually laborious and time-consuming to screen polymorphic EST-SSRs out (Lu et al., 2014). It has been reported that the possibility of an EST-SSR being polymorphic is directly related to its repeat number and inversely related to its motif size (Yi et al., 2006). However, the characteristics of highly polymorphic EST-SSRs, including various compound EST-SSRs derived from the transcriptome of an individual plant, have not been fully documented.

The Chinese bayberry (*Myrica rubra* Sieb. et Zucc.) is a subtropical perennial plant belonging to the family Myricaceae and an economically important fruit crop in China (Chen et al., 2004; Zhu et al., 2013). Recently, RNA-Seq of the Biqi cultivar was conducted, and some genetic features, including codon usage, have been characterized (Feng et al., 2012, 2013). In this study, EST-SSRs from the Chinese bayberry transcriptome were characterized and divided into 11 types; those that were highly polymorphic were then identified.

METHODS AND RESULTS

By using MISA (http://pgrc.ipk-gatersleben.de/misa/), 6883 di- and hexa-nucleotide EST-SSR loci have been identified from all 41,239 UniGenes assembled *de novo* from the Chinese bayberry transcriptome (Feng et al., 2012) (Table 1). Among them, two to five EST-SSR loci located in a UniGene with less than 100 bases between adjacent loci were defined as compound EST-SSRs. Based on their characteristics, 5557 non-compound EST-SSRs were classified into six types, and 357 compound EST-SSRs with all their motifs in either two or three nucleotides were divided into five types (Table 2). From each type of EST-SSR, a minimum of 20 EST-SSRs with known flanking sequences longer than 40 bp were selected to investigate their polymorphisms in 10 Chinese bayberry cultivars, as reported in our previous study (Zhang et al., 2012). From 5914 EST-SSRs, 412 were selected for polymorphism analysis and 109 polymorphic EST-SSRs, including 35 compound and 74 non-compound EST-SSRs, were screened out (Table 2).

The polymorphic rate, indicated by the percentage of polymorphic loci (PPL) of different types of EST-SSR, ranged from 12.9 to 87.5%, with 2-ntL having the highest followed by 3-ntL, Compound B, and Compound A (Table 2). The PPL of these four types was

over 70%, which was much higher than the mean PPL of all the EST-SSRs derived from the transcriptome, which was 29.4% (Table 2). This result indicates that the efficiency of screening polymorphic EST-SSRs in these four types could be at least 1.98-fold higher than that from randomly screening all EST-SSRs. The number of alleles (N_A), polymorphic information content (PIC), the number of genotypes, expected heterozygosity, and observed heterozygosity were used to evaluate the polymorphic level of a polymorphic SSR, and we found that the mean N_A and PIC values of 2-ntL and Compound B were both higher than those of the other types of EST-SSR, followed by 2-ntM and Compound A (Table 2). The polymorphic levels of di-nucleotide EST-SSRs were generally higher than those of EST-SSRs with larger motifs (including 3-ntL) (Table 2). The numbers of polymorphisms in the compound EST-SSRs were slightly lower than those in the non-compound EST-SSRs (Table 2). These results are consistent with what has previously been reported: that SSR polymorphisms are closely associated with their forms, motif sizes, repeat numbers, and repeat lengths (Temnykh et al., 2001; Buschiazzo and Gemmell, 2006; Yi et al., 2006; Kelkar et al., 2008; Cavagnaro et al., 2010).

Table 1. Constitutions of 6883 EST-SSR loci derived from the transcriptome of Chinese bayberry.

Motif	Number of repeats										Total (Percentage)
	5	6	7	8	9	10	11	12	13	>13	
AG/CT	1092	390	269	232	201	134	136	141	115	38	2748 (39.9%)
Other 2-nt[a]	214	31	21	11	4	6	4	4	1	3	299 (4.3%)
Total	1306	421	290	243	205	140	140	145	116	41	3047 (44.3%)
	4	5	6	7	8	9	>9				
AAG/CTT	465	147	61	37	21	17	3				751 (10.9%)
AGC/CTG	234	61	28	7	3	1					334 (4.9%)
AGG/CCT	240	80	30	11	3	3					367 (5.3%)
ACC/GGT	192	63	22	9	6	2					294 (4.3%)
ATC/ATG	188	40	19	5							252 (3.7%)
Other 3-nt	248	64	20	7	4	4					347 (5.0%)
Total	1567	455	180	76	37	27	3				2345 (34.1%)
	3	4	5	6	7						
4-nt	724	91	13	4	1						833 (12.1%)
5-nt	175	17	5	1							198 (2.9%)
6-nt	375	75	7	3							460 (6.7%)
Total	1359	183	25	8	1						1491 (21.7%)

[a]Nucleotide.

CONCLUSIONS

For the Chinese bayberry, 2-ntL, Compound B, and Compound A EST-SSRs should be preferentially selected for the screening of polymorphic EST-SSRs, due to their high polymorphic rates and levels. After selecting specific EST-SSRs from the different types, the efficiency of screening polymorphic EST-SSRs in Chinese bayberry should be significantly increased. The results of the present study should facilitate genetic and breeding studies on the Chinese bayberry, and provide a reference for similar study on other plant species.

Table 2. Types of EST-SSRs derived from the transcriptome of Chinese bayberry and their polymorphism.

Type of EST-SSRa	Model	Total No.	EST-SSRs selected for designed primers	EST-SSRs with successful PAGE[b] analysis	Polymorphic EST-SSRs	PPL[c]	Average N_A[e]	Average PIC[f]
2-ntS	$(NN)_j$ (5 ≤ j ≤ 7)	1523	32	19	5	26.3	3.2	0.41
2-ntM	$(NN)_j$ (8 ≤ j ≤ 11)	569	93	45	23	51.1	3.5	0.47
2-ntL	$(NN)_j$ (j ≥ 12)	266	51	32	28	87.5	4.3	0.60
3-ntS	$(NNN)_j$ (4 ≤ j ≤7)	1894	43	31	4	12.9	3.0	0.43
3-ntL	$(NNN)_j$ (j ≥ 8)	53	27	12	10	83.3	2.8	0.40
(4-6)-nt	$(NNNN)_j$ & $(NNNNNN)_j$ (j ≥ 3)	1252	29	16	4	25.0	2.3	0.30
Compound A	$(NN)...(NN)_k$ (j, k ≥ 5, j + k ≤ 15)	184	37	14	10	71.4	3.4	0.47
Compound B	$(NN)_j...(NN)_k$ (j, k ≥5, j + k ≥ 16)	37	20	10	8	80.0	4.3	0.58
Compound C	$(NNN)_j...(NNN)_k$ (j, k ≥ 4, j + k ≤ 9)	55	25	15	3	20.0	2.3	0.27
Compound D	$(NNN)_j...(NNN)_k$ (j, k ≥ 4, j + k ≥10)	42	31	16	11	68.8	3.1	0.44
Compound E	$(NN)_{j1}...(NN)_{j2}...(NN)_{jm}$ & $(NNN)_{k1}...(NNN)_{k2}...(NNN)_{kn}$ (3≤ m, n ≤5; j1, j2,..., jm ≥5; k1, k2, ..., kn ≥4)	39	24	9	3	33.3	3.7	0.41
Total	/	5914	412	219	109	29.4[d]	/	/

[a]nt, nucleotide; S, short; M, moderate; L, long. In compound EST-SSRs included in this study, the number of bases interrupting two adjacent EST-SSR loci is no more than 100. [b]Polyacrylamide gel electrophoresis. [c]Percentage of polymorphic loci. The values in the last column (except 29.4) are the percentage of the values in the last but one column to the corresponding ones in the last but two column. [d]The average PPL of all EST-SSRs derived from the transcriptome is the ratio of 1737.6 (the number of all potential polymorphic EST-SSRs summed with the multiplies of values in the 3rd and the last but two column for each type of EST-SSR) to 5914 (in the 3rd column, as the number of all EST-SSRs derived from the transcriptome). [e]Number of alleles. [f]Polymorphic information content.

ACKNOWLEDGMENTS

We are grateful to D. Grierson for a critical appraisal of the manuscript. Research supported by the National High Technology Research and Development Program of China (#2013AA102606), the Special Fund for Agro-scientific Research in the Public Interest (#201203089-2), the Program of International Science and Technology Cooperation (#2011DFB31580), the Science and Technology Project of Zhejiang Province (#2012C12904-3), and the Scientific Research Starting Foundation for Introduced Doctoral Teachers of Tongren University (#trxyDS1303).

REFERENCES

Buschiazzo E and Gemmell NJ (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. *Bioessays* 28: 1040-1050.

Cavagnaro PF, Senalik DA, Yang L, Simon PW, et al. (2010). Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics* 11: 569.

Chen KS, Xu CJ, Zhang B and Ferguson IB (2004). Red bayberry: botany and horticulture. *Hort. Rev.* 30: 83-114.

Dutta S, Kumawat G, Singh BP, Gupta DK, et al. (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol.* 11: 17.

Feng C, Chen M, Xu CJ, Bai L, et al. (2012). Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. *BMC Genomics* 13: 19.

Feng C, Xu CJ, Wang Y, Liu WL, et al. (2013). Codon usage patterns in Chinese bayberry (*Myrica rubra*) based on RNA-Seq data. *BMC Genomics* 14: 732.

Kelkar YD, Tyekucheva S, Chiaromonte F and Makova KD (2008). The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Res.* 18: 30-38.

Lu JJ, Kang JY, Ye SR and Wang HZ (2014). Isolation and characterization of novel EST-SSRs in the showy dendrobium, *Dendrobium nobile* (Orchidaceae). *Genet. Mol. Res.* 13: 986-991.

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.

Yi G, Lee JM, Lee S, Choi D, et al. (2006). Exploitation of pepper EST-SSRs and an SSR-based linkage map. *Theor. Appl. Genet.* 114: 113-130.

Zalapa JE, Cuevas H, Zhu H, Steffan S, et al. (2012). Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am. J. Bot.* 99: 193-208.

Zhang SY, Li X, Feng C, Zhu CQ, et al. (2012). Development and characterization of 109 polymorphic EST-SSRs derived from Chinese bayberry (*Myrica rubra*, Myricaceae) transcriptome. *Am. J. Bot.* 99: e501-e507.

Zhu CQ, Feng C, Li X, Xu CJ, et al. (2013). Analysis of expressed sequence tags from Chinese bayberry fruit (*Myrica rubra* Sieb. and Zucc.) at different ripening stages and their association with fruit quality development. *Int. J. Mol. Sci.* 14: 3110-3123.