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

Leaf transcriptome analysis and development of SSR markers in water chestnut (*Eleocharis dulcis*)

H.B. Liu[1]*, Y.N. You[3]*, Z.X. Zhu[4], X.F. Zheng[3], J.B. Huang[4], Z.L. Hu[3] and Y. Diao[2]

[1]Hubei Key Laboratory of Quality Control of Characteristic Fruits and Vegetables, College of Agricultural, Hubei Engineering University, Xiaogan, Hubei Province, China
[2]College of Forestry and Life Sciences, Chongqing University of Arts and Sciences, Yongchuan District, Chongqing, China
[3]State Key Laboratory of Hybrid Rice, College of Life Science, Wuhan University, Wuhan, Hubei Province, China
[4]College of Plant Science & Technology, Huazhong Agricultural University, Hubei Province, China

*These authors contributed equally to this study.
Corresponding author: Y. Diao
E-mail: diaoy78@163.com

Genet. Mol. Res. 14 (3): 8314-8325 (2015)
Received October 22, 2014
Accepted March 26, 2015
Published July 27, 2015
DOI http://dx.doi.org/10.4238/2015.July.27.20

ABSTRACT. Water chestnut (*Eleocharis dulcis*) is an important aquatic crop in China; however, transcriptomic and genomic data in public databases are limited. To identify genes and development molecular markers, high-throughput transcriptome sequencing was applied to generate transcript sequences from water chestnut leaf. More than 24 million reads were obtained, trimmed, and assembled into 40,796 contigs with an average length of 616.6 bp. Sequence similarity analyses against 4 public databases (NR, GO, KEGG, KOG) revealed 17,628 contigs that could be annotated with gene descriptions,

conserved protein domains, or gene ontology terms. Among the important metabolic pathways, 27 genes were related to starch synthesis and 13 genes were in the steroid synthetic pathway. In addition, 2570 cDNA simple sequence repeats were identified as potential molecular markers in our contigs. One hundred pairs of polymerase chain reaction primers were designed and used for validation of the amplification. The results revealed that 87 primer pairs were successfully amplified in initial screening tests. Overall, this transcriptome dataset and these markers can serve as a platform for further gene expression studies, functional genomic studies, and marker-assisted selection in *E. dulcis*.

Key words: *Eleocharis dulcis*; High-throughput sequencing; Simple sequence repeat; Transcriptome

INTRODUCTION

Eleocharis dulcis (Burm. f.) Trin. ex Henschel has various names, including *Eleocharis tuberosa* (Roxb.) Roem. et Schult., which is recognized and cited by most scholars (Li et al., 2006). *E. dulcis* (Cyperaceae) is commonly known as the Chinese water chestnut. It is a perennial herbaceous plant that grows in shallow waters and is mainly distributed in low-lying areas, such as pools and mudflats, in China, Southeast Asia, the Americas, Europe, and Oceania. The planting history of water chestnuts in China spans more than 2000 years, with the main production sites in the Yangtze River valley and southern China (Wang, 2005). Water chestnuts are also cultured in North Korea, Japan, Vietnam, India, Australia, and the USA. The underground bulbs of water chestnuts are edible as vegetables or fruits; most of the water chestnuts produced in China are supplied to fresh markets. The bulbs of water chestnuts are rich in nutrients and possess medicinal properties. Every 100 g fresh bulb contains 21.8 g carbohydrate, 1.5 g protein, 0.1 g fat, and 0.6 g coarse fiber (Kong, 2004). Water chestnuts are effective in clearing heat, dissipating phlegm, and removing food retention. Water chestnuts are also used to treat various symptoms, such as thirst, jaundice, abdominal mass, conjunctival congestion, throat swelling and pain, excrescence, bloody diarrhea, and massive metrorrhagia. A recent study found that some compounds in water chestnuts have anticancer, antibacterial, and antioxidant properties (Liu et al., 2010).

Recent studies of water chestnuts have mainly focused on cultivation techniques, tissue culture, physiological and biochemical characteristics, and processing, rather than on genetics (Li et al., 2006). Only reports regarding the phylogenetic relationships among germplasm resources using random amplification polymorphic DNA (RAPD) markers are available (Jiang et al., 2012). Previous studies have reported that small genetic differences exist between the known varieties of water chestnuts. Only 19 nucleotide sequences in water chestnuts have been published in GenBank. The lack of genetic data limits the genetic breeding of water chestnuts as well as the research and utilization of its characteristic functional genes. Next-generation sequencing is a high-throughput technique that has a wide variety of applications. This technique allows for hundreds of thousands of DNA strands or even several millions to be sequenced at the same time. Thus, next-generation sequencing makes transcriptome sequencing or deep sequencing of the genome convenient and feasible (Varshney et al., 2009; Metzker, 2010). In this study, high-throughput sequencing was conducted

to analyze the transcriptome of water chestnut leaves. A considerable amount of expressed sequenced tag (EST) information on functional genes has been obtained. EST information provides basic data for the cloning of full-length genes and the study of their functions. Moreover, a large number of simple sequence repeat (SSR) markers based on EST database development can be used to study the genetic biodiversity of water chestnuts and for marker-assisted selection.

MATERIAL AND METHODS

Plant materials

E. dulcis cv. "Tuanfeng", a famous local cultivar in Hubei Province, China, was selected as the plant material. Each tuber of *E. dulcis* cv. "Tuanfeng" was separately planted in a pot, which was placed inside a greenhouse at Wuhan University. Fresh leaves were collected from 5 individual plants and mixed together. These leaves were then frozen in liquid nitrogen and stored at -70°C until RNA extraction.

Total RNA extraction and mRNA purification

Total RNA extraction of the samples was conducted using TriZol reagent according to manufacturer instructions (Invitrogen, Carlsbad, CA, USA). The total RNA extracted was initially digested with DNase I (Ambion, Austin, TX, USA), and the mRNA was purified using the Micropoly(A)Purist™ mRNA Purification Kit (Ambion) according to manufacturer instructions. The obtained mRNA was eluted with 100 μL pre-heated elution buffer and quantified using a NanoDrop spectrometer.

cDNA synthesis

The mRNAs were first reversed into first-strand cDNA fragments using Superscript II reverse transcriptase (Invitrogen) and the *Gsu*I-oligodT primer. The 5'-cap structure of mRNA was oxidized by $NaIO_4$ (Sigma, St. Louis, MO, USA) and connected with biotin. The mRNA/cDNA connected with biotin was screened with magnetic beads (Dynal M280; Invitrogen), and first-strand cDNA was released by alkaline lysis. DNA ligase (TaKaRa, Shiga, Japan) was used to add an adaptor to the 5'-end of the first-strand cDNA. Second-strand cDNA was subsequently synthesized by Ex Taq polymerase (TaKaRa). Finally, the polyA and 5'-end adaptors were removed by *Gsu*I digestion.

Construction of cDNA library, sequencing, and EST assembly

The synthesized cDNA was fragmented into 300-500 bp using a Sonic Dismembrator (Fisher, Waltham, MA, USA) and purified using Ampure beads (Agencourt, Brea, CA, USA). A TruSeq™ DNA XXmple Prep Kit-Set A (Illumina, San Diego, CA, USA) was used to prepare the library with the purified cDNA, and amplification was conducted using the TruSeq PE Cluster Kit (Illumina). The sequencing reaction was carried out on the Illumina sequencer. The clean reads were assembled using Trinity (http://trinityrnaseq.sourceforge.net/) to generate the EST cluster (contigs) (Grabherr et al., 2011).

Gene prediction and annotation

The assembled contigs were predicted using "GetORF" of EMBOSS (Rice et al., 2000) to search for protein-coding sequences in different contigs. The predicted protein-coding sequences were aligned in the NR of GenBank, KEGG, KOG, and UniProt using BLASTp (E value <1e-5). The alignment with the highest matching value was considered to contain annotation information.

Functional classification

GO is an internationally standardized classification system of gene functions. GO provides a set of standard vocabulary with dynamic updating to describe the properties of genes and gene products of the organisms. GO analysis was carried out using GoPipe (Chen et al., 2005). The predicted proteins were first aligned in Swiss-Prot and TrEMBL using BLASTp (E value <1e-5). According to gene2go, the GO information of the predicted proteins was obtained from the alignment results using GoPipe.

Construction of metabolic pathway

KEGG is a database that systematically analyzes the metabolic pathways of gene products in cells and the functions of gene products. The predicted proteins were aligned in KEGG database using reciprocal BLAST (E value <1e-3). The KO number of the predicted protein was obtained. According to the KO number, information regarding the metabolic pathway related to the predicted protein was acquired.

Searching and analysis of SSR

SSR sites were searched in EST using MISA (http://pgrc.ipk-gatersleben.de/misa/), and the parameters were set as follows. The total length of repetitive sequences was equal to or larger than 12 bp. The least numbers of repeat of dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide were 6, 5, 4, 4, and 4, respectively. SSR primer sequences were designed using Primer 3 (Rozen and Skaletsky, 2000). A total of 100 pairs of primers were designed, and the primer sets were tested for successful polymerase chain reaction amplification in the initial screening test.

RESULTS

Sequencing and EST assembly

The transcriptome of water chestnuts was sequenced using Illumina Solexa sequencing technology, and 24,008,765 raw reads were obtained. After removing low-quality data, 23,552,387 clean reads (98.1%) were obtained, with an average length of 100 bp. The data from high-throughput sequencing were submitted to GenBank (Sequence Read Archive). After assembly, 40,796 sequence contigs were obtained, with a length of 201 to 14,363 bp (average: 616.6 bp) (Figure 1).

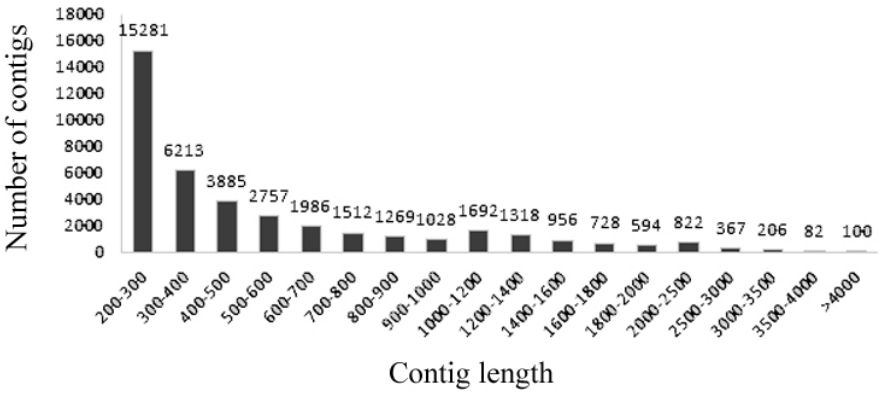

Figure 1. Length distribution of assembled contigs.

Functional annotation by searching against public database

A total of 17,628 contigs were annotated by searching and alignment using BLAST (Table 1). The largest number of annotations was found in KEGG (17,185 contigs, 42.1%), whereas the lowest number of annotations was found in KOG (6750 contigs, 1.7%). Additionally, 23,168 contigs (56.8%) were not annotated and may represent new genes.

Table 1. Summary of contigs annotated in the main published database.

	NR	KOG	KEGG	UniProt
Annotated	12,755 (31.2%)	6,075 (1.7%)	17,185 (42.1%)	9,549 (23.4%)
Un-annotated	28,041 (68.8%)	34,046 (98.3%)	23,611 (57.9%)	31,247 (76.6%)

Functional classification by GO

The contigs of water chestnuts were mapped to 52,714 GO terms, among which 17,978 (34.1%), 17,274 (32.8%), and 17,462 (33.1%) were related to molecular function, biological processes, and cellular components, respectively (Figure 2). In the GO classification system, the 3 major categories, including molecular function, biological process, and cellular component, were divided into 56 smaller categories. The 3 major sub-categories shown in Figure 2 were "catalytic activity" (GO: 0003824), "binding" (GO: 0005488), and "hydrolase activity" (GO: 0016787), which were in the cluster of molecular function. The 3 sub-categories of "cell" (GO: 0005623), "intracellular" (GO: 0005622), and "cytoplasm" (GO: 0005737) were in the cluster of cellular component, and the 2 sub-categories of "cellular process" (GO: 0009987) and "macromolecule metabolism" (GO: 0052174) were in the cluster of biological process. The classification results revealed the global expression profiles of water chestnut leaves.

Metabolic pathway analysis

The EST library of water chestnut leaves was annotated by 304 metabolic pathways in

KEGG. A total of 15 metabolic pathways related to the metabolism of carbohydrates and 289 contigs were identified. Starch and sucrose metabolism was the most important pathway, and 57 contigs were related to 27 enzymes involved in this pathway (Table 2). In addition to dietary value, water chestnuts also have important medical uses. The TCM Dictionary indicates that chestnuts contain puchiin, a type of antibacterial component, which can inhibit *Staphylococcus aureus*, *Escherichia coli*, and *Enterobacter aerogenes* (Li et al., 2003). The main pharmacologically active components of puchiin are steroids. Through homology searching, we identified 14 contigs that may encode 13 key enzymes in the steroid synthetic pathway (Table 2).

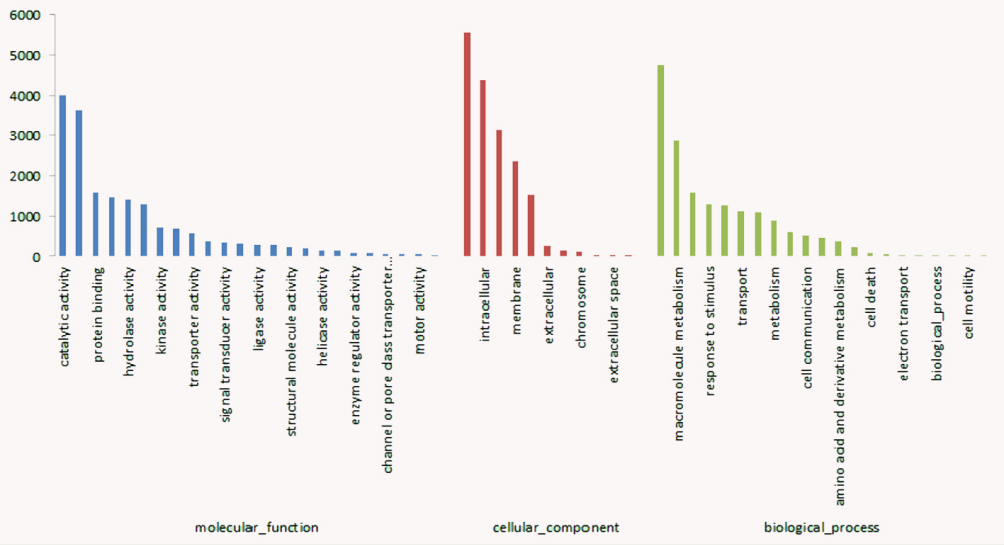

Figure 2. Gene Ontology classification of assembled contigs. The results are summarized in 3 main categories: biological process, cellular component, and molecular function.

Development and characterization of SSR markers

In the transcripts of water chestnuts, 2570 SSR sites were distributed in 2379 contigs, accounting for 5.8% of contig sequences. SSR types were diverse, including the repeat of 2-6 nucleotides (Table 3). The repetition of sequences also varied. Among the detected SSRs, 175 types of motifs were identified. The 5 types with the highest frequencies were GA/CT (734), AG/TC (677), TA/AT (152), TG/AC (84), and GT/AC (61).

Among the contigs containing SSR, 1606 were suitable for SSR primer design. A total of 4818 pairs of SSR primer sequences were obtained using Primer 5.0. Up to 100 pairs of randomly selected primers were synthesized and tested by polymerase chain reaction (Table 4). The results indicated that 87 pairs were effectively amplified, whereas 13 pairs failed. The effective rate was 87%. Sixty-four pairs showed the expected target band size. The amplified bands of 11 pairs were slightly larger than the target band, whereas those of the 2 pairs were slightly smaller. The products of the 9 pairs exceeded 500 bp, and 1 pair produced multiple bands without a target band. These variations suggest the presence of an intron or deletion within the amplicons, a lack of specificity, or assembly errors.

Table 2. Summary of water chestnut contigs annotated in steroid biosynthesis and starch and sucrose metabolism pathways in KEGG database.

Pathway	Contig	Enzyme
Steroid biosynthesis	CYP51; cytochrome P450, family 51 (sterol 14-demethylase) [EC:1.14.13.70]	comp422479_c0_seq1_4
	NSDHL, ERG26; sterol-4alpha-carboxylate	comp11991_c0_seq1_14
	3-dehydrogenase (decarboxylating) [EC:1.1.1.170]	comp14886_c1_seq1_4
	EBP; cholestenol delta-isomerase [EC:5.3.3.5]	comp16395_c0_seq1_5
	DHCR24; delta24-sterol reductase [EC:1.3.1.72]	comp89873_c0_seq1_10
	SC5DL, ERG3; lathosterol oxidase [EC:1.14.21.6]	comp1646_c1_seq1_4
	DHCR7; 7-dehydrocholesterol reductase [EC:1.3.1.21]	comp25636_c0_seq1_21
	LIPA; lysosomal acid lipase/cholesteryl ester hydrolase [EC:3.1.1.13]	comp25698_c0_seq1_5
	TGL4; TAG lipase/steryl ester hydrolase/phospholipase A2/ LPA acyltransferase [EC:3.1.1.3 3.1.1.13 3.1.1.4 2.3.1.51]	comp24776_c1_seq1_14
	CYP2R1; cytochrome P450, family 2, subfamily R, polypeptide 1 (vitamin D 25-hydroxylase) [EC:1.14.13.159]	comp20316_c0_seq1_7
	SMT1, ERG6; sterol 24-C-methyltransferase [EC:2.1.1.41]	comp28699_c5_seq1_15
	cycloartenol synthase [EC:5.4.99.8]	comp24111_c0_seq1_19
	SMO1; 4-dimethyl-9beta,19-cyclopropylsterol-4alpha-methyl oxidase [EC:1.14.13.72]	comp17180_c0_seq1_2
	24-methylenesterol C-methyltransferase [EC:2.1.1.143]	comp27351_c0_seq1_8
Starch and sucrose metabolism	E3.2.1.26, sacA; beta-ructouranosidase [EC:3.2.1.26]	comp26858_c0_seq1_29
	E2.4.1.14; sucrose-phosphate synthase [EC:2.4.1.14]	comp29444_c0_seq1_21
	E2.4.1.13; sucrose synthase [EC:2.4.1.13]	comp27523_c0_seq1_14
		comp28823_c0_seq1_16
	E3.2.1.20, malZ; alpha-glucosidase [EC:3.2.1.20]	comp21527_c1_seq1_11
	E3.2.1.58; glucan 1,3-beta-glucosidase [EC:3.2.1.58]	comp27552_c0_seq1_8
	beta-glucosidase [EC:3.2.1.21]	comp20184_c0_seq1_25
		comp23205_c2_seq1_7
		comp28597_c1_seq1_14
	otsB; trehalose 6-phosphate phosphatase [EC:3.1.3.12]	comp20984_c0_seq1_24
	TPS; trehalose 6-phosphate synthase/phosphatase [EC:2.4.1.15 3.1.3.12]	comp25984_c0_seq1_56
		comp27077_c0_seq1_22
		comp28762_c1_seq1_53
		comp29282_c2_seq1_15
		comp29341_c0_seq1_47
	E3.2.1.28, treA, tre; alpha,alpha-trehalase [EC:3.2.1.28]	comp16908_c0_seq1_11
		comp16908_c1_seq1_9
	treX, glgX; glycogen operon protein [EC:3.2.1.-]	comp18906_c0_seq1_19
		comp28816_c1_seq1_13
	UGDH, ugd; UDPglucose 6-dehydrogenase [EC:1.1.1.22]	comp24578_c0_seq1_9
	UGT; glucuronosyltransferase [EC:2.4.1.17]	comp16074_c0_seq1_4
		comp16383_c0_seq1_30
		comp17247_c1_seq1_4
		comp19142_c0_seq1_8
		comp20042_c0_seq1_25
		comp21047_c0_seq1_5
		comp21613_c0_seq1_2
		comp25221_c1_seq1_7
		comp25699_c0_seq1_28
	UXS1; UDP-glucuronate decarboxylase [EC:4.1.1.35]	comp27223_c0_seq1_26
	UDP-glucuronate 4-epimerase [EC:5.1.3.6]	comp677_c0_seq1_3
		comp20588_c2_seq1_7
	GAUT; alpha-1,4-galacturonosyltransferase [EC:2.4.1.43]	comp20939_c1_seq1_7
		comp24841_c1_seq1_28
		comp25106_c1_seq1_16
		comp25267_c0_seq1_4
	pectinesterase [EC:3.1.1.11]	comp24761_c1_seq1_7
	UGP2, galU, gal; UTP-glucose-1-phosphate uridylyltransferase [EC:2.7.7.9]	comp29232_c0_seq1_4
	pgm; phosphoglucomutase [EC:5.4.2.2]	comp28111_c0_seq1_33
	HK; hexokinase [EC:2.7.1.1]	comp20724_c0_seq1_9
		comp28872_c0_seq1_8
	GPI, pgi; glucose-6-phosphate isomerase [EC:5.3.1.9]	comp27919_c0_seq1_17
		comp28536_c0_seq1_11

Continued on next page

Table 2. Continued.

Pathway	Contig	Enzyme
	scrK; ructokinase [EC:2.7.1.4]	comp15209_c0_seq1_7
		comp26426_c0_seq1_20
	glgC; glucose-1-phosphate adenylyltransferase [EC:2.7.7.27]	comp23698_c0_seq1_14
		comp28067_c0_seq1_36
		comp28307_c0_seq1_11
	E2.4.1.21, glgA; starch synthase [EC:2.4.1.21]	comp19342_c1_seq1_6
		comp28655_c0_seq1_35
	glgB; 1,4-alpha-glucan branching enzyme [EC:2.4.1.18]	comp29367_c0_seq1_58
	E2.4.1.1, glgP, PYG; starch phosphorylase [EC:2.4.1.1]	comp28261_c0_seq1_16
		comp28261_c1_seq1_30
		comp29102_c0_seq1_9
	beta-amylase [EC:3.2.1.2]	comp29037_c0_seq1_34
	malQ; 4-alpha-glucanotransferase [EC:2.4.1.25]	comp22564_c0_seq1_35
		comp29472_c0_seq1_46

Table 3. Summary of SSR search results.

SSR type	Repeat unit				Total
	<6	6-10	11-20	>20	
Dinucleotide	0	1654	56	0	1710
Trinucleotide	399	318	0	0	717
Tetranucleotide	74	19	0	0	93
Pentanucleotide	16	2	0	0	18
Hexanucleotide	16	2	1	0	19
Others					13

DISCUSSION

This is the first study to apply high-throughput sequencing technology (Solexa) to water chestnuts for sequencing of the leaf transcriptome, functional analysis, and mining of important functional genes. A large number of ESTs related to the functional genes of water chestnuts were identified by homology searching. Starch is the main biomass synthesized by water chestnuts. The current studies on water chestnut starch focus on the properties and processing quality (Huang, 1994; Liu, 1999; Wu et al., 2003; Jiang et al., 2009; Kong et al., 2011), but the synthesis pathway and relevant genes of water chestnuts have not been investigated. Our results present the complete synthesis pathways of starch in water chestnuts, the enzymes involved in the metabolism, and the corresponding functional genes. These findings provide basic data for gene cloning and studies of gene functions. Phytosterol (or plant sterol) is an active component in plants that can be directly applied for anti-inflammation, reducing blood lipid levels, and ulcer and cancer treatment (Wu and Zhang, 2007). For example, 24-ethyl-Δ7-cholesterol, one of the steroids present in water chestnut, is an active ingredient that has significant antibacterial, anti-inflammatory, and analgesic effects *in vitro* and *in vivo* (Hao et al., 2005; Liu et al., 2006). We identified important genes related to the synthesis of phytosterol in water chestnuts. The discovery of these genes provides a basis for the study of biosynthetic pathways of active components and regulation mechanisms. The results of our study also provide a theoretical foundation for the promotion of effective components in water chestnuts or production of effective components and their intermediates using biological techniques.

Table 4. SSR primers for polymerase chain reaction tests.

Primer name	Forward primer (5'-3')	Reverse primer (5'-3')
ED-EST-SSR-1	TTCCCCTGTTCGATGTTTTC	ACCGGTCACTAAATTCGCAG
ED-EST-SSR-2	GACGAGGATGAGGATGAGGA	CCTTTACCACCACCACATCA
ED-EST-SSR-3	AACCAAGAGCAAACACACCC	GGAACCTATCCGAGGAGGAG
ED-EST-SSR-4	GGATTAACCATCAATCAACTTCC	TGGACAAATTGCAACTCCTG
ED-EST-SSR-5	GCACAAACCTTGCACCAGTA	TCAGAGACGTTGGTGTCGAG
ED-EST-SSR-6	ACGCCTGGAATTCAAGCTAA	TTCTATCGTACGCCAAACCC
ED-EST-SSR-7	CTAAGGCTCCCTCTTGCCTT	TGGAGAAGAGGAGAGGTGGA
ED-EST-SSR-8	TCGCAAACGCCTTTTAGTTT	GTATGGAGGTGTGGTGAGGC
ED-EST-SSR-9	GGGATTTGTGATGGGTTTTG	TATTCGCTTCGCTTCACTCA
ED-EST-SSR-10	TTAGGCTTTCAGCTCGTGGT	TGCATCACTCAGCTCTCTGG
ED-EST-SSR-11	ATGTGCCCAAGTTTTTCGAC	ATTGGTTGGAGGGGGATAAG
ED-EST-SSR-12	CCACTGCGCATTCATAGAGA	TTGACTACGCGATGGACAAT
ED-EST-SSR-13	GGAAATAAGACGCTGGATGC	CTTGTAGCCAAGGAAGGCTG
ED-EST-SSR-14	CCAAGTGACGAAATCTGTTCC	TTCGCTTAATTGCGGAGTCT
ED-EST-SSR-15	TTTCCCACGTTCCTACCTTG	CCTGGAGTTTCTGAAAAGCG
ED-EST-SSR-16	GTCGTTGGATCGCTCAATTT	CTTGCTCTTGCTCTTTCGCT
ED-EST-SSR-17	AACCCTAACCCTAACCCGTG	GATCGAGCCCTTTCTTTCC
ED-EST-SSR-18	CCCTCTCTCTTCCTTTGCCT	TTGCGCCTTCTTTTCAACTT
ED-EST-SSR-19	GAGATTGAGTTGCCTCTCCG	GAAAGTGGAGCACCAACCAT
ED-EST-SSR-20	CTTCGTTTTGCGTCCCTAAT	GACATTAACACCACCACCCC
ED-EST-SSR-21	GCCGAGACCTCGGTTTAGTA	TCACTTTTAGGAGGTGGTTGG
ED-EST-SSR-22	CGTCGACAAAATGGAGAAGA	TCCATTGTAATCGCCTTTCC
ED-EST-SSR-23	GTCCACATTCCATTCCATCC	CTTTCTCCAGTGAGGTCCGA
ED-EST-SSR-24	AAGCGTCAAAATTTCATAAACCA	GGTACGGGATAGGAATCGGT
ED-EST-SSR-25	CGGTACCAGAAGTGGAGGAA	AACTGCACAGGGAGAGAAAAA
ED-EST-SSR-26	CCCAATCATCTTTCTTTGTATTCC	GAATGGGAGGTGCGTATTGT
ED-EST-SSR-27	CCCTTTCTTTCTCCCGTTCT	TCCCTTTCGCATTTTCTGAC
ED-EST-SSR-28	AAGAGGCAAAGTGGTAGGCA	TGAAGAAAAATGGTCCCTCG
ED-EST-SSR-29	AGCCTCCGGATCATCCTATT	TTCATCAAGAGAAAACGGGG
ED-EST-SSR-30	GGGTCCCCTATGTCAAGGTT	AGTGCTTTGCACTGCATTTG
ED-EST-SSR-31	TGGACCACAAAAATTGCAGA	TTCATTTCTAGCCTATCCCAACTT
ED-EST-SSR-32	AATTCACATGGCAAAGAGCC	TAGAGCAAGGAAGAGGGCAA
ED-EST-SSR-33	CGAGTAGGCGAGCAAGATG	TCGAGGAAGCGAGGAAACTA
ED-EST-SSR-34	GCCTTCTCTGCTTTCCCTCT	CCCTTTCCCGTACTTGTTCA
ED-EST-SSR-35	TCGCCACCTCTCTTTTTCTC	CGTCCAATGGAGAACATCCT
ED-EST-SSR-36	AACTGAATCAACCATTCGCC	TCAACTATCGCATCATGCTTCT
ED-EST-SSR-37	CGGGAAATTTAGCATCTTGG	TCGAATATTTCATTTCTACACTTGC
ED-EST-SSR-38	GAGACAAAGGAGAAGGCGTG	CAATCATCAATCATCGTCCG
ED-EST-SSR-39	GGGCAAAACTGAAAAACCAA	CATCACCTTGATTGGGAGGT
ED-EST-SSR-40	GGGTTGGAGACATGAGAGGA	GAGCAGCATCTCTTTGCTCTT
ED-EST-SSR-41	CACTATTCCATTAATTCCCACTCTC	TTCGTTCTTCCTCCGCTTTA
ED-EST-SSR-42	CTCCTCCTTCCTGCCTCTCT	TGATGCTTTAATGTTTTTCCTCC
ED-EST-SSR-43	CACGGAACCCTGATTTCCTA	TCCCTCGTTATCCCCTTCTT
ED-EST-SSR-44	GCTGCTTCCAGAACTCCATC	TCGATTTGGTTTCTTCGACC
ED-EST-SSR-45	CCTCTCTCCCCTTTTTGTCC	GCTGCTCGGTGCTATAAAGG
ED-EST-SSR-46	AAACCGTTGAATCCAATCCA	TTCAAGGTAACCTCATCGGG
ED-EST-SSR-47	GAGCATCATCCTTCCTCTGC	CAACCGCAATCTCAATCCTT
ED-EST-SSR-48	TCTCAATGCCTGACAAGACG	CAACATAACCCTTCCCGCTA
ED-EST-SSR-49	AAAAAGCAGTTGCTGTGGCT	ATCCCAACCCATCAATGAAA
ED-EST-SSR-50	AAAAGGGAAAGGCTCCGTTA	GGGGATGTGAATTCAGGAGA
ED-EST-SSR-51	CAAGGTCGCACTGGTTACCT	CTGCATCTGTATCCCGGTTT
ED-EST-SSR-52	AGCAAACCACTTCCAGCACT	TTATAGCCCAAACCCACTCG
ED-EST-SSR-53	ATTTCTTCTCCCTCGCACCT	TTCCCGTTTGGGAGTATCAG
ED-EST-SSR-54	TTCAGCAAAGACACGTGGTTA	CATTTCCCTTTCCAACGCTA
ED-EST-SSR-55	CGGAATAGCAAAGAAAACGC	TTCGTATTCACGATTTGCCA
ED-EST-SSR-56	AGAGATGTTGGATCGGGTTG	GCTCTGCTCTGCTCTCGTCT
ED-EST-SSR-57	TCAGTTCAGCCTCAAAAGCC	AACAAGCATAGGAGGCCAGA
ED-EST-SSR-58	ATCTCTTCGCCAGAAACCAA	ATGGCCTTGATCTGATCGTC
ED-EST-SSR-59	GAAAGGAACGATGCCAGGTA	ACTTGAGGGAGGAACTGGGT
ED-EST-SSR-60	TTTTGGTGCCTTAACAAACG	GTCGGAGGCTGCTATTTGAG
ED-EST-SSR-61	TGCCCTTGACATGTTAGCAG	GGGATGAAGTGGGATGTGAG
ED-EST-SSR-62	TCACACAAGCTCACACACCA	AGTATTGCGTATCGGGTTCG

Continued on next page

Table 4. Continued.

Primer name	Forward primer (5'-3')	Reverse primer (5'-3')
ED-EST-SSR-63	ATTTCTCGCTCTCGGTCAAA	CCCAGGAACAGAAAGAGCAG
ED-EST-SSR-64	GAGACCAAGGGGACAAATGA	TCCTATCTGATCGCGTTTCC
ED-EST-SSR-65	GGAATAAGAGACGGAAGGGG	CAAGCAAACCCTTCTCTTGG
ED-EST-SSR-66	CTCCCTTCCTTTCTCCATCC	TCCTTCGTCGATTGAGGTTC
ED-EST-SSR-67	GGAGTCGAAGAAAGCACCAA	AAACCAGCAGGTTATGGGAA
ED-EST-SSR-68	CGTTCCGTGCTTTCATTCAT	GCGAACTATGAGGAAGGCTG
ED-EST-SSR-69	TCCATTGCAATTGGTTCAGA	AGCCTCATCAGCATTCCAGT
ED-EST-SSR-70	ATCATCATCCTGTCGCATCA	CTCAGTCAGCCCCATGAAAT
ED-EST-SSR-71	CAGAACCCAAATTAAGCAAATATACTC	AAGCTCTCTCCCAGAGTCCC
ED-EST-SSR-72	TGGTGCTTGTATTGGCTGAG	CCCACACCATTCTTCACCTT
ED-EST-SSR-73	TTCCGCCTCTTCTTCTGTTC	CCACTCCCCAAAATAAGCAA
ED-EST-SSR-74	CACCCCTCTGAAGCAAAAAG	TTGCGCCAATTCCTTACTTC
ED-EST-SSR-75	AATGCCTTAAGCCACTCAGC	GAGTGCCCTGCTCTAGTTCG
ED-EST-SSR-76	CTCAGTGGCATTTTGGGAAT	GTTTCTTTTCCTTCTCCGCC
ED-EST-SSR-77	ACCTTTCAAACGCTCTCTCG	GCTCCCACTCGAATTGAAAG
ED-EST-SSR-78	TCTCCATCACCTCCATCCTC	CTTCTGCCTTCTATTCCCCC
ED-EST-SSR-79	GTTCTTTGACAGCGTCTCCC	TCTCTTTCACGCCTTGACCT
ED-EST-SSR-80	CCAACATCATCATCGAACCA	TTCCTCGCGTGAGAAAACTT
ED-EST-SSR-81	ATGGCATTACTTGTCCCTGG	GGCAATGCTGATGCAGACTA
ED-EST-SSR-82	CCAGTCACCACATCACGAAC	TCACTCCAACCTCTCTCGCT
ED-EST-SSR-83	CATATCCCGCGAGTTGAAAT	TTTGAGTTGGGAATTGGGAG
ED-EST-SSR-84	TGGGGTTGAGTACTTGGGAG	CCACACGCTTTCTTTCTTCC
ED-EST-SSR-85	TTTCCAAGCACCTGGAGAGT	TGAATCTACTATTGTACATGGCTCC
ED-EST-SSR-86	AAGAGGGAAATGGAGCGAAT	TGCCTCCTTTTTCCTTTCAA
ED-EST-SSR-87	GGCTAGAAATGAAACCCAATCA	CGCTATACGCTATTTAAAACCTTG
ED-EST-SSR-88	CAGAAACCCCAACAATTAACAA	TGCTTAGTTGCCGAAGGAGT
ED-EST-SSR-89	TCGACTTTTCACCAGCAGAG	ATGGCGGAACCATTCTATCA
ED-EST-SSR-90	ATTGCACGCGAGATTCTCTT	TCCCTACCTCTCCCCTTGTT
ED-EST-SSR-91	GGGGTCAAATACCCATGTCA	CCACAAGCTCACGTGAAGAA
ED-EST-SSR-92	CAGATCAAAGCTCAATGCGA	TACCTCGACACCGAAAAAGG
ED-EST-SSR-93	TGCCCAAATTAACTGGAACC	CAAGCTGCTTTAGGTAGCATCA
ED-EST-SSR-94	TGAAACCTTCAGAACCTGCC	TCCAGTCACTTCGATCCTGA
ED-EST-SSR-95	ACAGCAACAGCAACAGCATC	ATTCCTAGCCACAGGTGCAG
ED-EST-SSR-96	CTCTTGCTCTCCTCGATTGG	GGAGAGATGGCAAAAGCAAG
ED-EST-SSR-97	TCTCGACGGCTGTAGAGCTT	CCGCTTTCTTCGTCTCAATC
ED-EST-SSR-98	TGCATGTGATATATGTCAACTTTTG	TTTCAAAATCCTTTTTGGCG
ED-EST-SSR-99	CGAGTTCGGATCGACTAAGG	TTTCTGAACACAAGCATGCAC
ED-EST-SSR-100	CCGAGCAAGACAACAGTCAA	CCCTAGACCCCTCGTAAAGC

Given the abundance, co-dominance, conformity to Mendel's law, good technical repeatability, easy operation, and reliable results obtained using SSR markers, they are widely applied for parentage analysis, genetic diversity analysis, construction of fingerprints, and genetic linkage mapping (Brown et al., 1996; Gupta et al., 2000; Varshney et al., 2005; Aggarwal et al., 2007; Gong et al., 2008). EST-SSR is a new molecular marker that has higher value in real application because its polymorphism may be directly related to gene function and is universal between similar plants (Chabane and Varshney, 2005; Castillo et al., 2008; Zhao et al., 2011). Previous studies have only reported the application of RAPD markers in water chestnuts. Because RAPD markers are dominant or partially dominant, they fail to distinguish between homozygous and heterozygous genes and are unable to provide complete genetic information. The stability of RAPD markers is also unsatisfactory. Therefore, the large number of SSR markers developed in this study will facilitate the analysis of genome differences between the water chestnut and closely related species as well as the construction of a genetic linkage map.

ACKNOWLEDGMENTS

Research supported by the National Science and Technology Supporting Program (#2012BAD27B01).

REFERENCES

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, et al. (2007). Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor. Appl. Genet*. 114: 359-372.

Brown SW, Szewc-McFadden AK and Kresovichi S (1996). Development and Application of Simple Sequence Repeat (SSR) Loci for Plant Genome Analysis. In: Methods of Genome Analysis in Plants (Jauhar PP, ed.). CRC, Boca Raton, 147-159.

Castillo A, Budak H, Varshney RK, Dorado G, et al. (2008). Transferability and polymorphism of barley EST-SSR markers used for phylogenetic analysis in *Hordeum chilense*. *BMC Plant Biol*. 8: 97.

Chabane K, Ablett GA and Cordeiro GM (2005). EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genet. Resour. Crop Evol*. 52: 903-909.

Chen ZZ, Xue CH, Zhu S, Zhou FF, et al. (2005). GoPipe: streamlined gene ontology annotation for batch anonymous sequences with statistics. *Prog. Biochem. Biophys*. 32: 187-191.

Gong L, Stift G, Kofler R, Pachner M, et al. (2008). Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L. *Theor. Appl. Genet.* 117: 37-48.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol*. 29: 644-652.

Gupta P and Varshney R (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113: 163-185.

Hao SX, Liu X, Zhao LC and Chen YQ (2005). Investigation on the stability of bacteriostatic component from puchiin extract. *Food Sci*. 26: 71-74.

Huang LX (1994). A study on the water chestnut starch. *J. South Chin. Univ. Technol.* (*Nat. Sci.*) 22: 16-23.

Jiang W, Li YR, Yang LT, Chen LJ, et al. (2009). Study on agronomic characters and nutrition of Chinese water chestnut (*Eleocharis tuberosa*). *Chin. Veg.* 2: 51-54.

Jiang W, Cai BH, Chen LJ, Ou KP, et al. (2012). Genetic diversity analysis of 24 water chestnut (*Eleocharis tuberosa* Schulut) cultivars using RAPD markers. *J. South Agric.* 43: 895-900.

Kong JX, Han WF, Lv GY, Xiong SB, et al. (2011). Research progress of Chinese water chestnut processing. *Preserv. Process*. 11: 43-46.

Kong QD (2004). Variety Resources of Water Vegetable in China. Hubei Science and Technology Press, Wuhan.

Li DS and Wang JH (2003). Study on the bacteriostatic effect of leaf and stem extract from water chestnut. *Food Oil Process Food Machine* 10: 76-77.

Li F, Ke WD and Liu YM (2006). Advances in research of common spikesege. *J. Changjiang Veg.* 8: 39-43.

Liu JQ (1999). Achene micromorphological characters of *Eleocharis* (Cyperaceae) from China and its taxonomic significance. *Chin. J. Appl. Environ. Biol.* 5: 578-584.

Liu X, Zhao LC and Zhou AM (2006). Preliminary study on functional component and functional activities of waste slurry derived in processing water chestnut starch. *Food Sci*. 27: 251-256.

Liu X, Zhou ZG, An BS and Li BL (2010). Pharmacological research progress in water chestnut. *Info. Tradit. Chin. Med*. 27: 106-108.

Metzker ML (2010). Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11: 31-46.

Rice P, Longden I and Bleasby A (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 16: 276-277.

Rozen S and Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132: 365-386.

Varshney RK, Graner A and Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23: 48-55.

Varshney RK, Nayak SN, May GD and Jackson SA (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol*. 27: 522-530.

Wang W (2005). Health function and processing of chufa. *Food Drug* 7: 45-48.

Wu SP and Zhang Z (2007). Research status of plant sterol. *Food Nutr. Chin.* 9: 20-22.

Wu TY, Yin DL, Zhang C, Yao MC, et al. (2003). A preliminary study on the biological characteristics of wild water chestnut. *Weed Sci.* 1: 15-17.

Zhao H, Yu JY, You FM, Luo MC, et al. (2011). Transferability of microsatellite markers from *Brachypodium distachyon* to *Miscanthus sinensis*, a potential biomass crop. *J. Integr. Plant Biol.* 53: 232-245.

