INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
15–776/2016/18–3–589–600
DOI: 10.17957/IJAB/15.0130
http://www.fspublishers.org

Full Length Article

SSR Development and Utilization with *Rehmannia glutinosa* Transcriptome

Mingjie Li[1],Yanhui Yang[3], Fajie Feng[1], Xinyu Li[2], Li Gu[1], Yunhe Tian[1], Jianming Wang[1], Jiankai Wang[1], Zhenfang Li[1], Aiguo Chen[1], Xinjian Chen[2*] and Zhongyi Zhang[1*]

[1]*College of Crop Sciences, Fujian Agriculture and Forestry University, Fuzhou, China, 350002*
[2]*College of Life Sciences, Henan Agricultural University, Zhengzhou, China, 450002*
[3]*College of Bioengineering, Henan University of Technology, Zhengzhou, China, 450001*
[*]For correspondence: zyzhang@fafu.edu.cn; xinjian1@yahoo.com

Abstract

Rehmannia glutinosa Libosch is a well-known medicinal plant in China. Despite its common use, very few molecular markers are available for this species, severely impeding any improvement of key agronomic traits. In this study, gene-based SSR (simple sequence repeat) markers were mined from the *R. glutinosa* transcriptome, assembled from leaf and root libraries, to obtain polymorphic information content (PIC). A total of 1,812 SSR (≥18 bp) loci were identified from 1,747 transcripts, in which 1,018 transcripts had significant homology to known proteins, and 35 transcripts were identified as non-coding RNA. Of all SSR loci, AG/GT was the most frequent SSRs repeat type, and bias of base composition presented SSRs loci from non-coding RNA. A set of 279 SSR primers (≥20 bp) were synthesized, of which 204 were successfully amplified in the Wen85-5 genome and 91 had a polymorphic information content between 0.33 and 0.90. Seven pairs of polymorphic SSR primers were selected to evaluate the genetic diversity of 36 *R. glutinosa* germplasms, and results demonstrated that cultivars are differ considerably from wild strains and that distinct genetic relationships exist among wild strains. The SSRs identified in this study provide critical information for future *R. glutinosa* breeding. © 2016 Friends Science Publishers

Keywords: *Rehmannia glutinosa;*·EST-SSR;·Transcriptomes;·Germplasms;·Polymorphism

Introduction

Rehmannia glutinosa Libosch, a perennial herbaceous medicinal plant belongs to the *Rehmannia* specie of *Scrophulariaceae* family and is considered a top-grade Chinese herb. The tuberous root is its main medicinal tissue, containing various bioactive components including sugars, amino acids, vitamins, iridoid, aucubin, approximately 70 of which have been identified to date (Zhang *et al*., 2008). More than 50 cultivars of *R. glutinosa* are common in Henan, Shandong, Jiangsu, and other provinces in China, at total acreage of more than 8,600 to 9,300 hectares with an annual output of approximately 80 thousand ton (Zhou *et al*., 2010).

R. glutinosa production currently suffers several major issues, however. The genetic relationship among *R. glutinosa* germplasms remains unclear, due to their complex genetic backgrounds, continuous vegetative propagation and the high heterozygosity of each cultivar (Zhou *et al*., 2010), making it difficult to select varieties of *R. glutinosa* with high heterosis. Moreover, biological characteristics for most *R. glutinosa* cultivars lack definition and among them there is no obvious difference in appearance (Wu *et al*., 2007); thus, a cultivar is easily mixed with another one in the main production regions (Zhou *et al*., 2007; Zhou *et al*., 2010), significantly increasing rate of genetic variation. The cultivars of *R. glutinosa* in production are thus gradually graded with prolonged planting times. In addition, *R. glutinosa* contains a large numbers of wild germplasms that are distributed worldwide, many of which have favorable traits, such as medicinal components, and tolerance to biological or abiotic stresss (Wang *et al*., 2008; Zhao and Li, 2009). As such, wild strains are very valuable resources for improving certain traits of existing cultivars. Wild germplasms, however, are unnamed, and their genetic backgrounds are unclear in *R. glutinosa*. Due to general lack of effective identification methods, some landraces and good germplasms have been lost from *R. glutinosa* genetic resources.

Therefore, it is imperative to develop efficient molecular markers that identify genetic relationship of different cultivars and distinguish valuable wild germplasms. Although a few molecular markers, including AFLP (Yuan and Hong, 2003), RAPD (Chen *et al*., 2002; Zhang *et al*., 2006; Wu *et al*., 2007; Zhou *et al*., 2007; Wang *et al*., 2008), ISSR (Zhou *et al*., 2007, Wang *et al*., 2008), SRAP (Zhou *et al*., 2010) have been reported to identify genetic relationships in various *R. glutinosa* germplasms, their utilization has been restricted by limited polymorphic information content (Zhang *et al*., 2012). A

To cite this paper: Li, M.,Y. Yang, F. Feng, X. Li, L. Gu, Y. Tian, J. Wang, J. Wang, Z. Li, A. Chen, X. Chen and Z. Zhang, 2016. SSR development and utilization with *Rehmannia glutinosa* transcriptome. *Int. J. Agric. Biol*., 18: 589–600

recently published set of 15 polymorphic EST-SSR markers that were identified based on *Rehmannia* species are available (Liu *et al*., 2015). However, little polymorphic SSR markers specifically suited to identify intraspecific variation were mined in *R. glutinosa*. Since very few SSR markers are available for this species, which severely hampers its breeding progress and the improvement of key agronomic traits.

In this study, a high-capacity *R. glutinosa* transcriptome derived from different organs using high-throughput sequencing technology (Li *et al.*, 2013) was built to provide basic support for the development. Meanwhile, Polymorphism level of novel *R. glutinosa* EST-SSR markers was finely identified based on same *R. glutinosa* species from different ecological regions. The genetic diversity among *R. glutinosa* strains are evaluated and genetic relationship between wilds strains and cultivars were tested through using newly mined EST-SSR markers. Furthermore, the relationships of SSR motif types and gene function were in detail clarified. These SSR markers developed in this study will proved a powerful tool for identification of genetic diversity among *R. glutinosa* germplasms, thereby ensuring favorable characteristics of the germplasm, broaden of breeding resources and finally breeding of superior *R. glutinosa* cultivars.

Materials and Methods

Collection of Germplasms and DNA Extraction

A total of 18 cultivars and 18 wild strains were analyzed. A mixture of leaves from five independent plants of each germplasm was collected at the 7–8 leaf. Genomic DNA from each sample was extracted using a CTAB (Hexadecyl trimethyl ammonium bromide) method. DNA concentrations were measured using a spectrophotometer, and DNA integrity was determined by analysis on a 1.5% (w/v) agarose gel. A serious of analyses was conducted as shown in Fig. 1.

SSR Mining and Functional Analysis

A MISA script (http://pgrc.ipk-gatersleben.de/misa/) was used to search SSR loci in the *R. glutinosa* transcriptome. Only perfect motifs of 2–6 nucleotides were considered. Minimum SSR length was defined as at least 9 repetitions for dinucleotides, 6 for trinucleotides, 5 for tetranucleotides, 4 for pentanucleotides, and 3 for hexanucleotides. To analyze the function of the transcripts containing SSRs, the sequences were aligned using BlastX against Nr databases, applying a threshold E-value of 10-5. GO and KEGG annotation were assigned by homology with known genes. Bestorf scripts in Molequest software were used to predict CDS (coding) regions. Transcripts without any annotation in the Nr, GO, or KEGG database were mapped into *R. glutinosa* miRNA precursor generated in previous study (Li *et al.* 2013), Rfam 10.0 (rfam.sanger.ac.uk/), NONCODE V3.0 (http://www.noncode.org/), and Repbase 20.0 (http://www.girinst.org/repbase/) databases using BlastN with an E-value cutoff of 10-5. Sequences recognized as long non-coding RNAs (lncRNA) in the Rfam and NONCODE databases were further distinguished from transcript-coding proteins (Wu *et al*., 2012).

SSR primer design. Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/) was used to design the primer of each SSR locus. SSR loci with flanking sequence shorter than 100 bp were removed. Primers were designed under the following parameters: primer length from 18 to 24 bp (20 bp was optimal length); product from 100 to 300 bp (150 bp optimum); annealing temperature from 59 to 60°C (59°C optimum); and (G+C) content from 30 to 70%. PCR amplification was performed in a final volume of 15 μL in a reaction mixture containing 80 ng of genomic DNA, 1× PCR buffer (20 mM Tris-HCl pH 9.0, 100 mM KCl, 1.5 mM MgCl2), 150 μM dNTP, 0.1 μM of each primer, and 1.5 U of Taq polymerase (TaKaRa, Tokyo, Japan). The PCR reaction program was set as follows: DNA denaturation at 94°C for 3 min, followed by 30 cycles of 94°C for 1 min, 55–65°C for 30 seconds, 72°C for 1 min, and 72°C for a final extension of 7 min.

SSR Validation and Polymorphism Examination

The efficiency of 279 SSR (≥ 20 bp) primers was identified in the Wen85-5 genome. Clear amplification was considered to indicate a validated SSR primer; other primers were not further analyzed. Polymorphic information content of validated SSRs was identified in 12 *R. glutinosa* germplasms, including 7 cultivars and 5 wild strains (Table 1). The PCR-amplified products were separated by 6% polyacrylamide gel electrophoresis. For each of the SSRs in the 12 germplasms, each allele was scored as present (1), absent (0) or unclear (9) in the same position. The polymorphism index of the SSR primers was calculated according to the methods described by Botstein *et al*. (1980).

Genetic Diversity Evaluation and Data Analysis

In total, 5–10 pairs of polymorphic primers were selected to detect the genetic distance of 36 germplasms using the presence/absence notation described above, and 1-s and 0-s matrices were established accordingly. The 0 and 1 matrices were then imported into NTSYSpc 2.0 software (Rohlf, 2000), which used SIMQUAL subroutines to calculate genetic similarity based on the simple matching coefficient. The SHAN clustering procedure was also utilized for cluster analysis of all *R. glutinosa* germplasms based on their UPGMA (Unweighted Pair-Group Average).

Results

Assembly of *R. glutinosa* Transcriptome and SSR Loci Mining

A previous study constructed two *R. glutinosa* transcriptome

Table 1: *R. glutinosa* germplasms collected in this study

ID in germplasm garden	Provisional name for wild strains and ordinary name for cultivars	Wild or Cultivar	Culster
Reh1	wild1	wild strain	wild-Cluster2
Reh2	wild2*	wild strain	wild-Cluster1
Reh3	wild3*	wild strain	wild-Cluster1
Reh4	wild4	wild strain	wild-Cluster2
Reh5	wild5	wild strain	wild-Cluster3
Reh6	wild6	wild strain	wild-Cluster2
Reh7	wild7	wild strain	wild-Cluster3
Reh8	wild8	wild strain	wild-Cluster1
Reh9	wild9	wild strain	wild-Cluster1
Reh10	wild10*	wild strain	wild-Cluster3
Reh11	wild11	wild strain	wild-Cluster2
Reh12	wild12	wild strain	wild-Cluster2
Reh13	wild13	wild strain	wild-Cluster2
Reh14	wild14	wild strain	wild-Cluster2
Reh15	wild15	wild strain	wild-Cluster3
Reh16	wild16	Relative wild strain	Cul-Cluster2
Reh17	wild17*	Relative wild strain	Cul-Cluster2
Reh18	wild18*	Relative wild strain	Cul-Cluster2
Reh19	JingZhuangyuan	Cultivar	Cul-Cluster3
Reh20	Hongshuwang*	Cultivar	Cul-Cluster1
Reh21	Zeng2*	Cultivar	Cul-Cluster4
Reh22	Shengjing1	Cultivar	Cul-Cluster2
Reh23	85-5 (tissue culture)*	Cultivar	Cul-Cluster2
Reh24	9302	Cultivar	Cul-Cluster3
Reh25	Sankuai	Cultivar	Cul-Cluster4
Reh26	Wen85-5	Cultivar	Cul-Cluster2
Reh27	Zeng1*	Cultivar	Cul-Cluster3
Reh28	Shandong Cultivars*	Cultivar	Cul-Cluster2
Reh29	Qinghuai1*	Cultivar	Cul-Cluster3
Reh30	Beijing1	Cultivar	Cul-Cluster3
Reh31	Beijing3	Cultivar	Cul-Cluster3
Reh32	9302-1	Cultivar	Cul-Cluster3
Reh33	858*	Cultivar	Cul-Cluster2
Reh34	Zeng3	Cultivar	Cul-Cluster4
Reh35	Shanxibeixiang	Cultivar	Cul-Cluster3
Reh36	Shanxi Cultivars	Cultivar	Cul-Cluster3

*in column 2 represent strains or cultivars used to identify polymorphic information index (PIC) of SSR primers

libraries (SRA No: SRX269425 and SRX269426) from leaves and roots, assembled into 87,665 transcripts sequences. These sequences were used to mine *R. glutinosa* SSR loci. As a result, 1,812 SSRs (≥ 18 bp) were identified from 1,747 transcripts. Out of these transcripts, 957 contained a single SSR locus and 61 had more than two SSR loci. Out of all SSR loci, 44 loci from 188 transcripts were considered compound SSRs (Table 2). As far as repeat motifs of SSR loci, the dinucleotide was the most abundant repeat type, followed by hexanucleotide, trinucleotide, tetranucleotide, and pentanucleotide (Fig. 2a). The numbers of reiterations of any specific repeat unit varied, mostly ranged from 5 to 22. SSR length was about 18–24 bp, with a notable maximum of 74 bp dinucleotides (AG/CT) (Fig. 2b). Among 1,812 SSR loci, 205 distinct motif types were identified, of which AG/CT was the most frequent repeat type, followed by AC/GT, AAG/CCT, ATC/ATG, ACC/GGT, AAAAAG/CTTTTT, etc. The full repeat motifs are listed in Table 3.

Of the 1,747 transcripts carried SSR loci, 1,018 had at least one annotation in the Nr database, while 729 did not. The transcripts were further annotated in the GO and KEGG databases, as well. GO analysis showed that during biological processes, most transcripts were involved in the metabolic process, cellular process. In molecular function, transcripts carrying SSR loci were assigned to binding and catalytic activities. As far as cellular components, most transcripts containing SSR loci were related to cells and organelles (Fig. 3a). KEGG analysis showed that genes with SSR loci fell into 68 pathways, most of which involved transcription and environmental adaption (Fig. 3b). Moreover, 33% SSR loci were located in CDS, 62% in untranslated Regions (UTR) (Fig. 4a). For different repeat types of SSR motifs, dinucleotides were mostly located in

Table 2: Summary of SSR loci mined from *R. glutinosa* transcriptome

Items	Numbers
Total number of sequences examined:	87,665
Total size of examined sequences (bp):	41,829,880
Total number of identified SSRs:	1,018
Number of SSR containing sequences:	1,747
Number of sequences containing more than 1 SSR:	61
Number of SSRs present in compound formation:	41

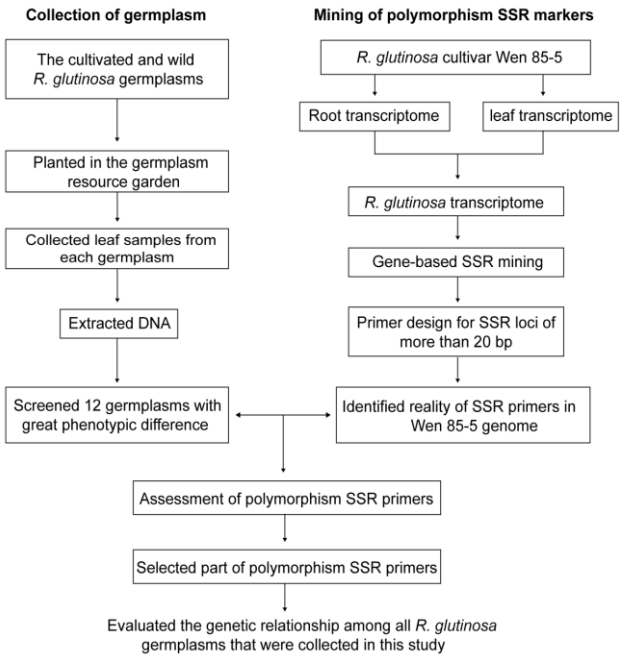

Fig. 1: Protocol for the mining and identification of *R. glutinosa* SSR loci

UTR regions; most trinucleotides and hexanucleotides in CDS regions (Fig. 4b). Transcripts contained SSR loci without any annotation were analyzed with non-coding database to determine whether they belonged, in fact, to non-coding RNA. 35 transcripts were identified as non-coding RNA, including lncRNA, tRNA, and miRNA, and four transcripts were considered repeats from retrotransposons. Of these, SSR loci on non-coding RNA were mainly dinucleotide, while those on repeats were hexanucleotide. Notably, the bases of SSR on non-coding transcripts were mainly composed of 'AG', 'TG' and 'AC' (Fig. 3c and d; Table 4), suggesting that a bias of base composition might be present in non-coding RNA-derived SSRs.

Validation of SSR Primers, Collection of *R. glutinosa* Germplasm and SSR Polymorphism Identification

The availability of SSR primers was determined by flanking sequences of sufficient length on both sides of an SSR locus. Among the 1,771 SSR loci, 969 transcripts were removed due to inappropriate flanking length, and 802 were retained to design primers. Some studies have shown that SSR loci length is closely related to polymorphism. In general, SSRs of more than 20 bp has a high probability of polymorphic information content. To obtain more efficient polymorphic SSR markers, only SSRs lengths above 20 bp were selected to design primers in this study. A total of 377 pairs of primers (279 pairs of synthesized primers) were generated. The primers were tested in the Wen85-5 genome

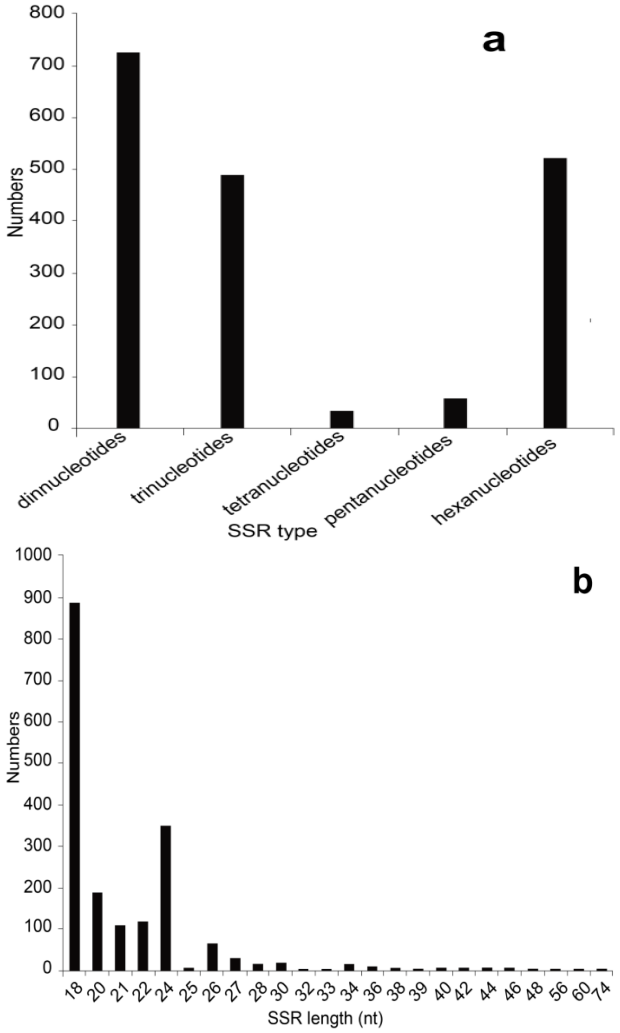

Fig. 2: (a) Type and (b) length distribution of SSRs motifs identified from the *R. glutinosa* transcriptome

for validation. A total of 204 pairs of primers were successfully amplified in the genome with expected products. To evaluate SSR polymorphism, and identify the genetic relationships among *R. glutinosa* germplasms, both cultivated and wild *R. glutinosa* were collected across different regions. A total of 18 cultivars and 18 wild strains were obtained, as shown in Table 1. The collection of germplasm provides an important basic resource for the evaluation of SSR polymorphisms. Seven cultivars and five wild strains were selected to identify polymorphisms for 204 pairs of SSRs primers (Table 1). 91 pairs of primers showed higher polymorphism (0.33 to 0.90) with an average value of 0.71 (Fig. 5; Table 5).

Evaluation of *R. glutinosa* Germplasm Genetic Diversity

Seven primer pairs (RSSR651, RSSR1500, RSSR1504, RSSR971, RSSR945, RSSR282, and RSSR1438) were randomly selected from the polymorphic SSR loci to

Table 3: Distribution of SSR motif types in *R. glutinosa*

Repeats	3	4	5	6	7	8	9	10	11	12	13	14	15+	total
AG/CT	-	-	-	-	-	-	93	50	72	153	37	8	44	457
AC/GT	-	-	-	-	-	-	73	62	41	48	25	2	5	256
AAG/CTT	-	-	-	70	27	21	10	5	1	-	2	-	1	137
ATC/ATG	-	-	-	43	29	14	10	-	1	-	-	-	-	97
ACC/GGT	-	-	-	38	19	8	2	1	-	-	-	-	-	68
AAAAAG/CTTTTT	58	10	-	-	-	-	-	-	-	-	-	-	-	68
AAT/ATT	-	-	-	34	7	4	2	-	-	-	-	-	-	47
AAC/GTT	-	-	-	27	10	6	2	-	-	-	-	-	-	45
CCG/CGG	-	-	-	30	3	-	-	-	-	-	-	-	-	33
AAAAAT/ATTTTT	28	2	-	-	-	-	-	-	-	-	-	-	-	30
AGG/CCT	-	-	-	20	7	2	-	-	-	-	-	-	-	29
ACCTCC/AGGTGG	17	8	-	-	-	-	-	-	-	-	-	-	-	25
AGC/CTG	-	-	-	17	3	2	-	-	-	-	-	-	-	22
ACCGCC/CGGTGG	19	2	-	-	-	-	-	-	-	-	-	-	-	21
AAAAG/CTTTT	-	19	1	-	-	-	-	-	-	-	-	-	-	20
AAGATG/ATCTTC	15	3	-	-	-	-	-	-	-	-	-	-	-	18
AAGAGG/CCTCTT	10	4	-	-	-	-	-	-	-	-	-	-	-	14
AAAAT/ATTTT	-	10	2	-	-	-	-	-	-	-	-	-	-	12
others	285	78	24	13	4	0	7	1	1	-	-	-	-	413
Total	432	136	27	292	109	57	199	119	116	201	64	10	50	1,018

Table 4: The SSRs loci derived from transcripts that have identifed as non-coding RNA and repeat sequences

Unigene id	Type of SSRs	SSR motif and repetitions time	ID in Rfam, NONOCODE and Repbase	Type of Noncoding RNA or repeats sequences
Unigene425_All	p2	(GA)10	n334879	mRNAlike lncRNA
Unigene51112_All	p2	(GA)22	n264864	mRNAlike lncRNA
Unigene38544_All	p2	(GA)12	n276503	mRNAlike lncRNA
Unigene43035_All	p2	(GA)21	n275433	mRNAlike lncRNA
Unigene68059_All	p2	(AG)19	n275065	mRNAlike lncRNA
Unigene78703_All	p2	(AG)11	RF00026	U6
Unigene613_All	p2	(TG)9	n287474	mRNAlike lncRNA
Unigene3240_All	p2	(TG)12	n297183	mRNAlike lncRNA
Unigene70021_All	p2	(TG)9	n290208	mRNAlike lncRNA
Unigene47222_All	p2	(TG)13	n288999	mRNAlike lncRNA
Unigene25813_All	p2	(GT)17	n275381	mRNAlike lncRNA
Unigene18687_All	p2	(AC)12	n283784	mRNAlike lncRNA
Unigene21069_All	p2	(AC)12	n287142	mRNAlike lncRNA
Unigene16327_All	p2	(AC)10	RF00100	7SK
Unigene42481_All	p2	(AC)13	n338062	mRNAlike lncRNA
Unigene78272_All	p2	(CA)11	RF00100	7SK
Unigene85611_All	p2	(CA)10	RF00001	5S_rRNA
Unigene26566_All	p2	(CA)10	RF00315	snoJ33
Unigene41185_All	p2	(CA)11	n283634	mRNAlike lncRNA
Unigene23508_All	p2	(TC)10	n281857	mRNAlike lncRNA
Unigene25233_All	p2	(TC)9	n281799	mRNAlike lncRNA
Unigene17868_All	p2	(CT)15	RF00005	tRNA
Unigene8186_All	p2	(CT)11	n417493	lncRNA
Unigene68397_All	p2	(CT)19	n269631	mRNAlike lncRNA
Unigene72964_All	p2	(CT)37	RF00004	U2
Unigene43528_All	p2	(TA)9	RF00028	Intron_gpI
Unigene495_All	p3	(CAT)6	RF01059	mir-598
Unigene41481_All	p3	(GCC)6	RF01766	cspA
Unigene85097_All	p5	(TGATT)4	RF00016	SNORD14
Unigene35550_All	p5	(CCTCT)4	RF00029	Intron_gpII
Unigene85097_All	p5	(TGATT)4	n1400	snoRNA
Unigene35042_All	p6	(AAAAAT)3	RF00638	MIR159
Unigene35042_All	p6	(AAAAAT)3	n361659	miRNA
Unigene25675_All	p6	(GGGATC)3	n267364	mRNAlike lncRNA
Unigene42021_All	p6	(AGAGGA)4	n293725	mRNAlike lncRNA
Unigene35444_All	p6	(GGAGCA)4	Copia1-I_DM	Partial sequences of retrotransposons
Unigene35444_All	p6	(GGAGCA)4	Copia-1_BM-I	Partial sequences of retrotransposons
Unigene35444_All	p6	(GGAGCA)4	YOYOI	Partial sequences of retrotransposons
Unigene11534_All	p6	(GAGCTC)3	Gypsy-3_PPc-I	Partial sequences of retrotransposons

identify the genetic relationships among 36 *R. glutinosa* germplasms, 900 polymorphic bands were produced in 36 germplasms by seven pairs of primers on average; each primer pair generated 180 allele fragments. The genetic similarity coefficient among the 36 *R. glutinosa* germplasms ranged from 0.575 to 1.000 with an average value of 0.672. The clustering analysis revealed that 36 *R. glutinosa* germplasms were clustered into 2 categories: wild strains

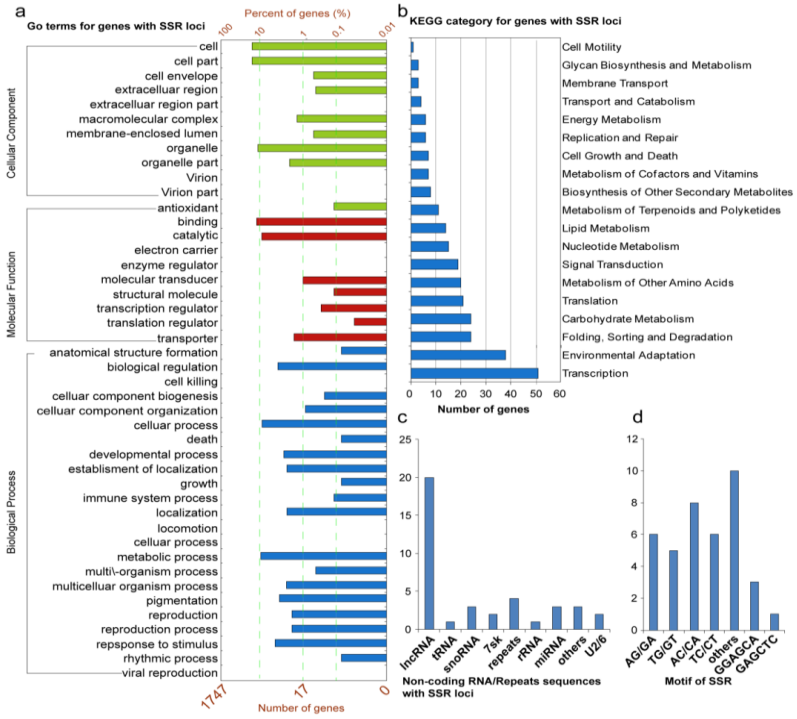

Fig. 3: (a) GO and KEGG, (b) category of genes with SSR loci, (c) types of non-coding RNA and repeat sequences with SSR loci, and (d) distribution of the SSR motifs located on these sequences

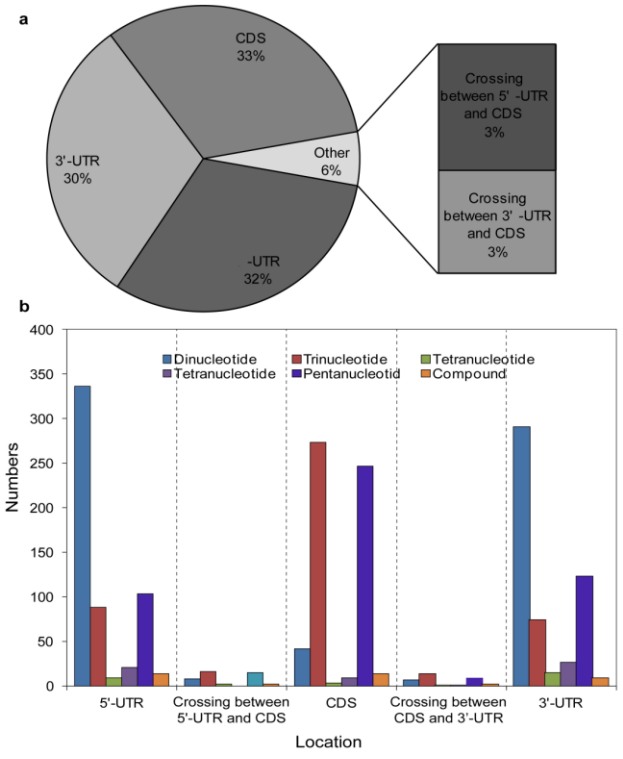

Fig. 4: (a) Distribution of *R. glutinosa* SSR loci locations on transcripts, and (b) comparison of SSR locations of different SSR motifs on transcripts

and cultivated strains. The wild strain was further divided into three groups: the first group (wild-Cluster1) included four types, wild2, wild8, wild9 and wild3, which were more closely related to "Hongshuwang", conventional cultivars. The second group (wild-Cluster2) contained wild1, wild4, wild6, wild13, wild14, wild11 and wild12, with distant relatives of cultivated *R. glutinosa* from the first group but relatively close relation to the wilds. The third group (wild-Cluster3) included wild7, wild5, wild15 and wild10, further relatives from the other two groups. The cultivars were further categorized into four groups (Fig. 6, Table 1). Notably, three wild strains (wild16, wild17 and wild18) were also observed in the cultivar category.

Discussion

EST-SSRs take not only full advantage of available datasets, but it captured from transcript sequences represent genes associated with phenotypic traits, making them highly conserved among relative species (Pashley *et al*., 2006). These advantages have made EST-SSR markers popularly applied to non-model plants (Gupta and Gopalakrishna, 2010; Wang *et al*., 2010; Blanca *et al*., 2011; Dutta *et al*., 2011; Asadi and Monfared, 2014); especially for species with unavailable genomes or complex genetic backgrounds. Although some EST-SSRs have been mined in *R. glutinosa* in a previous study, but polymorphism of these SSRs were tested based mainly on different *Rehmannia* species including *R. glutinosa*, *Rehmannia chingii*, *Rehmannia*

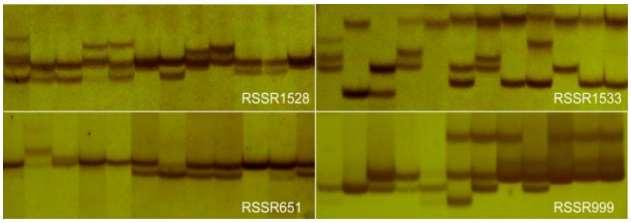

Fig. 5: Typical polymorphic information content for RSSR1528, RSSR1533, RSSR651, and RSSR999 from 12 *R. glutinosa* germplasm

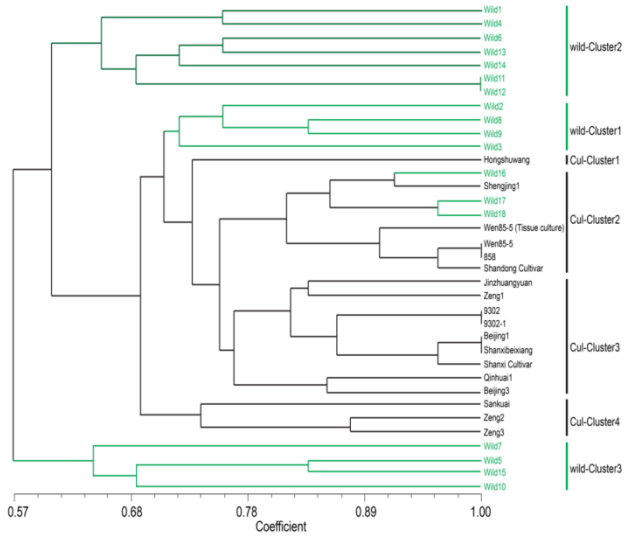

Fig. 6: Genetic relationships between *R. glutinosa* cultivars (Cul) and wild strains (wild) identified in 7 *R. glutinosa* polymorphic SSR markers

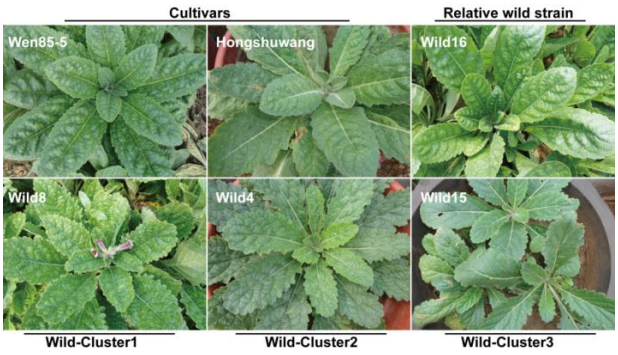

Fig. 7: Phenotypic comparison of aboveground parts for representative *R. glutinosa* cultivar and wild strains

piasezkii, *Rehmannia solanifolia*, *Rehmannia henryi* (Liu *et al*., 2015). These EST-SSRs have thus limited amplification for genetic relationship among same *R. glutinosa* species with relatively small genetic difference and variety breeding in agriculture.

Here, a non-redundant *R. glutinosa* transcriptome contained 87,665 transcripts from two sets of roots and leaves have constructed in our previous study (Li *et al*., 2013). Approximately 1.97% transcript sequences were found to carry at least one SSR locus (⩾18 bp), similar to the 2.8% prevalence in barley (Fraser *et al*., 2004), 3.2% in wheat (Parida *et al*., 2006), 3.83% in Ramie (Liu *et al*., 2013). Some studies had pointed out that majority of transcript sequences with SSR loci originated from the functional genes associated various cellular functions (Gupta *et al*., 2010; Dutta *et al*., 2011). In this study, 58.27% of the transcripts with SSR loci had at least an functional annotation, majority of which are closely related to transcription, metabolic process and environment adaption; 41.73% of the transcripts with SSRs had no homologous. This reflects that SSRs loci in *R. glutinosa* obviously prefer to particular category of genes. Additionally, 35 SSRs were found to locate in non-coding transcripts, in who's derived SSRs a bias of base composition was obviously observed, that was also found in other animals and plants (Chen *et al*., 2010; Joy *et al*., 2013; Asadi and Monfared, 2014). It was reported that EST-SSR markers were identified on transcribed regions of DNA that were often considered as putative functional sequences (Jiang *et al*., 2012). EST-SSRs thus owned higher transferability than genome-SSRs (Wen *et al*., 2010). The function information of specific transcripts with SSRs in this study provided a potential values for their further appraisal of transferability level among relative species.

Of SSR repeat types, dinucleotide repeats were the most common SSR repeat type, followed by tri-, hexa-, tetra- and penta-nucleotides, consistent with some plants including spruce, pumpkin (Gong *et al*., 2008), chickpea (Choudhary *et al*., 2009) and pigeon pea (Dutta *et al*., 2011). But that is different from many crop plants including barley, wheat, maize, rice and peanut, in which tri-nucleotide repeats generally exhibit a higher abundance. Moreover, of all repeat motifs types, AG/CT, AC/GT and AAG/CCT were the most abundant types, similar to sweet potato (Wang *et al*., 2011), peanut (Liang *et al*., 2009), and *Epimedium* (Zeng *et al*., 2010). AAG motifs have been observed to commonly present in dicots, such as *Arabidopsis* (Cardle *et al*., 2000), Soybean (Gao *et al*., 2003) and *Medicago* (Liang *et al*., 2009). Conversely, CCG motifs are the most common for monocots, such as maize, barley, and sorghum (Kantety *et al*., 2002; Varshney *et al*., 2002). The relatively high frequency of AAG (7.56%) with low CCG (1.82%) in this study, show usage bias of *R. glutinosa* SSR type is fully consistent with most dicotyledonous plants.

The polymorphic level of SSR markers directly determines its potential value applied in identification of genetic relationships and genetic mapping construction (Liang *et al*., 2009). To measure the level of polymorphism above set of developed SSR markers, validation of 279 primer pairs were firstly determined in the Wen 85-5 genome, followed by their polymorphisms were identified by 12 *R. glutinosa* strains.

Table 5: Polymorphic SSR loci that were identified based on 12 *R. glutinosa* germplasms including 7 cultivars and 5 wild strains

Marker ID	Unigene ID contained SSR sequences	SSR type	SSR size	Reverse primer	Annealing temperature (℃)	Forward primer	Annealing Temperature (℃)	Production (bp)	Alleles among 12 gemplasms	Polymorphsims (PIC)
RSSR1317	Unigene55658_All	(TTGT)7	28	TGGCCTATGTTGTC CATTTTT	59.33	CCGCAACTTGTAT TTGGTTG	59.08	127	80	0.33
RSSR1143	Unigene47222_All	(TG)13	26	GAACAAACTCGTG TGGGGTT	59.87	CTGCAACAATCAA CCAATCG	60.11	269	12	0.45
RSSR677	Unigene27359_All	(TTTTTG)4	24	TGTCAAATCCTTCC GTGGAT	60.32	CTTACTCCTCGTCG TCGGTC	59.87	179	12	0.45
RSSR169	Unigene7167_All	(TCGGAA)5	30	TTTTCTACCACGCC GTTTTC	60.11	CAATGGATGTCAA TGAAGAGGA	59.93	118	12	0.45
RSSR501	Unigene20311_All	(AAAAT)4	20	CACACGATTCTAAT CGAAGTGGTA	60.41	CACCACAACAAGG TGAATGC	60.01	214	12	0.50
RSSR256	Unigene11047_All	(CAA)7	21	CTGAGGGATGTTGT GACTATGG	59.47	GCAGGGGAAGACC TAAAAAGA	59.71	202	24	0.55
RSSR896	Unigene35550_All	(CCTCT)4	20	ACACGTATAGACG CGAACCC	60.02	GATAGGCGAAGGA AAGACCC	60.04	187	24	0.55
RSSR955	Unigene37465_All	(AAAAG)4	20	CCCTGCTCCACTCT TGCTAC	60.01	CGCTCTTCTGGAA ATCTTGG	59.95	147	24	0.55
RSSR1000	Unigene39280_All	(TGAA)6	24	GCCTCTCCATGGTA ATCCAA	59.89	TCCAAGAGTATGG GTTGAAGG	59.04	158	24	0.55
RSSR2	Unigene9_All	(TCGGGT)4	24	AGAATGGTCGAGA GGAACGA	59.80	TTCCAATTTAATTC CGCCAG	59.90	152	24	0.58
RSSR779	Unigene30448_All	(TCG)7	21	TCGAAGCACACCA TTTTCAA	60.23	AAAAGTGCTTTCG TGAAGGG	59.36	128	24	0.58
RSSR1633	Unigene75783_All	(GGA)7	21	GCCATGACTGGTTT CTAGGG	59.55	CGTACCCCATTGA TTATGCC	60.04	112	24	0.58
RSSR1646	Unigene77243_All	(TGA)8	24	ACTGGAATGCGAC AAAAAGG	60.11	TGCTACTCCCTTTG AACAGATG	59.38	207	24	0.59
RSSR1471	Unigene66533_All	(GAT)7	21	CATTAAAGGAGTT GGAGGGG	58.50	TGCCAACGAGTGA GACAAAG	60.03	195	24	0.59
RSSR1421	Unigene63517_All	(TG)10	20	ATTGAATGTCCCAC CTTTGC	59.80	GGTATGTTTGGTG CCTGAGAA	59.99	166	24	0.59
RSSR1136	Unigene47018_All	(GAT)7	21	AAGAACTGCTTCC GAACCG	60.38	AGTCCGGTGAATT TGCTCAC	60.12	277	24	0.59
RSSR1438	Unigene64623_All	(AAACGA)4	24	CCCCCTCCCTCAAT TTAACA	61.03	CCTATAGGTGGCG GAGATGA	60.05	236	26	0.62
RSSR1236	Unigene50943_All	(CGTCAC)4	24	CGAAGAAGCAGCA GAGAGGT	59.89	AAAGAGGGCAAG AATGGATG	59.13	250	24	0.62
RSSR1261	Unigene52112_All	(AAAAT)5	25	TGATTGACCCACTT GCATAAA	59.03	TGATTAGAAGCCA TTTTGGG	57.70	280	24	0.62
RSSR228	Unigene9835_All	(TG)13	26	GGCTAAAATCATC ACCCTTCC	59.79	GGTTCCAATAAGT TCCCCTGA	60.17	216	24	0.62
RSSR460	Unigene18986_All	(CTT)8	24	GGGATGCACCATA GGAGAGA	60.03	GATGGCGAGTATT CCACGTT	59.96	216	24	0.62
RSSR943	Unigene37012_All	(AGA)7	21	CTTTCCAGTATTTG GCCCTG	59.56	TCAGAGGGCTCAG GTTGAGT	59.99	184	24	0.63
RSSR1059	Unigene42299_All	(ATC)9	27	AAGTGCCCTACCC AACTCCT	59.99	TTACGTGTTTGGCT GTCTCG	59.90	235	24	0.63
RSSR703	Unigene28042_All	(GAA)11	33	ACCGACTCGAATA GAAGCGA	59.98	TTAAGAAGAGCCT CCCACTGA	59.06	192	24	0.63
RSSR1530	Unigene69753_All	(TG)12	24	ATGATCCAAACCC GAATGAG	59.75	CCACAGCTGTAAA TCCCACC	60.38	251	24	0.63
RSSR491	Unigene20076_All	(TGATGG)4	24	ACCCACATCAACC ACTGTCA	59.85	CGTTTTCCAAGCA CGACTC	59.43	282	24	0.63
RSSR880	Unigene35279_All	(TAAAT)4	20	GGATGAGTCATGG GCCTAAC	59.37	AGACATCCTGCAC CCTATCAA	59.57	111	24	0.64
RSSR738	Unigene29059_All	(CCAATC)4	24	GATGGAGGAAACT GCGTTGT	60.12	GAGGTTTGGGATG GTTGAGA	59.90	255	24	0.64
RSSR220	Unigene9455_All	(TTC)20	60	CGTCCCGAACTCAT TCAA	58.10	TCAGAAGGGTCAT TGTTGTTG	58.61	240	24	0.65
RSSR21	Unigene425_All	(GA)10	20	CACGATGCATTCCC TAAGAA	58.72	TTGATTCATTTCCC ACCACA	59.75	123	24	0.66
RSSR982	Unigene38293_All	(AC)12	24	TACGTCTAGGCGCC TTCATT	59.87	CGATGACGGCTGA TGTTGTA	60.69	187	24	0.66
RSSR1481	Unigene67013_All	(CCA)8	24	TAATTTCGCTTCCC CACAAG	60.07	AAGGTGGAGAGAG TGGCAGA	59.99	251	24	0.67
RSSR1199	Unigene49494_All	(GA)10	20	CGCCATAGCTTTAA TCCCAA	60.05	ATCCCTTCTCACA GCCCTTT	60.07	259	24	0.67
RSSR870	Unigene34952_All	(TCA)11	33	ATCCAGCGGTGTTG TTCACT	60.58	CTGCCATGGAGGA GGAGATA	60.17	127	24	0.69
RSSR1694	Unigene81853_All	(TGGTGA)4	24	GACGGGGAGTAAT GGTGAGA	59.93	AAACTGAAGAAAT CAAAATCCCTC	59.08	104	24	0.70

Table 5: Continued

Table 5: Continued

RSSR1500	Unigene67775_All	(AC)10	20	ACCATACGCATGCT TCACAA	60.14	CTCTCATCAACCA CCCCACT	59.96	143	24	0.70
RSSR1205	Unigene49676_All	(GA)28	56	CAAAGCTGTTGCTG CTCAAA	60.32	CCACGAGCCATTA TCTTTGG	60.46	237	36	0.71
RSSR1191	Unigene49248_All	(CT)12	24	CAGTTTTGCAAGA GTTGGGAG	59.90	GAGAGAAGAAGC ATCCACGG	59.95	225	24	0.71
RSSR1312	Unigene55417_All	(TGA)7	21	AAGGGAAAGGGAA GAGACCA	60.05	AAGCCATTCATTC ACTACTGTTG	58.33	168	24	0.71
RSSR862	Unigene34507_All	(GGT)7	21	TCTTCCTTGTTGTC ATCCCC	59.90	GAGAGAAAATGGG TGCTTGC	59.82	193	36	0.72
RSSR103	Unigene3240_All	(TG)12	24	GCCGGAATTAAAC ATCAAGG	59.41	TTGCAAGCATTGC CTACATC	59.84	251	24	0.73
RSSR445	Unigene18636_All	(GA)13	26	TGTGCGCATTTCCT ACTCTG	60.01	CGGAGAGATCAGA CGACGTA	58.98	154	36	0.74
RSSR116	Unigene4317_All	(AG)13	26	CCATTTCTCTCGCT CCTATTCT	59.00	TTACTGACCATTC ACACACGAA	59.10	192	24	0.74
RSSR1474	Unigene66681_All	(TTTC)5	20	ACGTTGGATTTGCA ACAGG	59.56	CTATCGACTCAGC ACGACCA	60.01	185	24	0.74
RSSR1315	Unigene55492_All	(TTCTCC)4	24	CACCAGGTTTTCCT CCCAT	59.76	GATAGATGGAGGA GGGGGAG	59.85	161	24	0.74
RSSR1444	Unigene65012_All	(AAG)7	21	TGGAGGAGCTTGG ACTGTTT	59.84	TCAACAAATTGAA TTGCTCAAC	57.81	214	24	0.74
RSSR983	Unigene38310_All	(GCAGAT)4	24	AGAATTGAAGGCT CTGGCAA	59.96	CAGTCATCGTCGG ACTCAAA	59.83	222	24	0.74
RSSR320	Unigene14562_All	(TTC)13	39	TGAATAAGCGGCG GAAAGTA	60.71	CGGGTAACAATGA GCGAGAT	60.10	198	36	0.74
RSSR282	Unigene12087_All	(AACA)5	20	ACACAAGGGAAAA ATGGCAG	59.97	AAGATCATTGGGC TTATGCG	60.06	211	36	0.75
RSSR1443	Unigene64991_All	(TG)10	20	GCGTTATTGCCAAC ATCCTC	60.48	TTTGCTTCCATCTC AAACCC	60.05	191	24	0.75
RSSR1613	Unigene74541_All	(TTTC)5	20	GTTTGGGGTATGTG GGATTG	59.91	AAAACGTGGCTAG CAGAACC	59.38	184	24	0.76
RSSR1590	Unigene73134_All	(TC)23	46	CAGCTATTTCCTCA GGCAGG	59.97	TGGAGGCAAGCGA AAGTTAT	59.85	269	24	0.76
RSSR1220	Unigene50174_All	(CCGCCA)4	24	TCGAATAAAATCT ATGTTGTTACTCCT	58.01	TTACCCTAAATGG AGCGGTG	59.95	189	36	0.76
RSSR1025	Unigene40645_All	(GA)11	22	CGCAAGCTCTGCTT CATGT	60.30	TCCGAACTTGCAC TTTGTTG	59.88	234	36	0.76
RSSR1676	Unigene80009_All	(GGTGGA)4	24	GCTATGATGAGGG ACGAGGA	60.18	CAGCCCCATACAC AACCTCT	59.99	106	36	0.76
RSSR14	Unigene263_All	(GT)10	20	GGCGGTTCTGCTAA AAAGC	59.97	ACACAATACCAAT CCCACCAG	59.58	300	36	0.77
RSSR694	Unigene27921_All	(TCATCC)4	24	TCTCCCATCCAGAT CCACTC	60.01	GGAAAAGGTTGAG GATGAGC	58.72	149	36	0.78
RSSR1546	Unigene70501_All	(AG)24	48	GTCGTTTCGAGGTT TTTCCA	60.09	CCCACATCCGAGT TGCTATT	59.96	137	24	0.78
RSSR651	Unigene26703_All	(GAGGGA)3	18	CAATTGAAAAGCC CTCCAAA	60.05	AAAGGGAGGCAG GTATGGAT	59.79	276	36	0.78
RSSR1528	Unigene69476_All	(TCA)9	27	TAATTCCCAATGCC AGAAGC	60.04	GAGGAATCAGTCA AGCCAGC	59.96	275	39	0.78
RSSR1200	Unigene49496_All	(CT)17	34	CCCATGCCAGTTAC CAACAT	60.64	GCGACATAGAGAG ATGGGGA	60.18	193	36	0.78
RSSR1151	Unigene47510_All	(CAC)7	21	CCTATCAAATTTTC TTCCACTCTC	57.58	TGCTCAGTGGTGC TCATAGG	60.01	122	36	0.79
RSSR282	Unigene12087_All	(AACA)5	20	ACACAAGGGAAAA ATGGCAG	59.97	AAGATCATTGGGC TTATGCG	60.06	211	36	0.79
RSSR129	Unigene5175_All	(TAATCC)4	24	CGCCACAATTTATA ACATAGATTCC	60.01	GAATTAGCCGGAG ATGTGGA	60.04	252	36	0.79
RSSR440	Unigene18537_All	(AG)16	32	CTACTCCGGTACTC CCCTCC	59.95	GTCAGAGGAGGAA GCGTTTG	59.99	270	36	0.80
RSSR447	Unigene18687_All	(AC)12	24	CCGTTTCTTCTCTC CCCTTC	60.18	GTAGTTGATGGCG GCAAAGT	60.14	129	60	0.80
RSSR1504	Unigene68059_All	(AG)19	38	CGTCTCTCTTCTTT GCCCAG	60.13	GCTGAACGCCATT TTTGATT	60.08	224	36	0.80
RSSR158	Unigene6737_All	(GAT)7	21	ACCAGTTCCACTAC CATCTTCC	59.37	CATCTCATCTCCCC AAGTTCA	60.06	215	24	0.80
RSSR392	Unigene17479_All	(CCA)8	24	ATCCACTCTCCACA CTTCACAC	59.08	ACGACGGGGACAT ACAGC	59.50	222	36	0.81
RSSR1269	Unigene52470_All	(GAATC)5	25	TTTGATGCTTGCTT TCACCA	60.38	CTTTCAAGCCCAG ACCAGAG	59.98	222	48	0.81
RSSR1011	Unigene39862_All	(CA)10	20	GGACCAATGCTTG CACAATA	59.55	TAGCACTTTCTCG GCCTCAT	59.98	240	36	0.81
RSSR1677	Unigene80078_All	(CAT)7	21	GGGTGGGGAGGGA AATAGTA	60.01	TGGCAAATTCTTA TTACTAACATCCA	60.14	133	36	0.82

Table 5: Continued

Table 5: Continued

RSSR358	Unigene15545_All	(GA)14	28	GCCATCTCCATTGC TCTTCT	60.00	AAACCATAGCCCA ACCTTCC	60.19	151	36	0.82
RSSR975	Unigene38091_All	(CT)10	20	GGCACATGAAATT TGAGGAAA	59.93	GCAAACACACCCA TGGTAAA	59.30	110	48	0.82
RSSR178	Unigene7676_All	(CA)11	22	CACGGTCAAGAAA CACCAGA	59.72	CGTTGGCAGCTTT CAGAAC	59.58	268	48	0.82
RSSR123	Unigene4772_All	(TTG)8	24	TGTTTGATGCTTTC TTCTTCGT	59.02	CCCCCATTCTTCTT TACCACT	59.33	164	48	0.83
RSSR27	Unigene683_All	(TTTC)5	20	TTGCAATGTTCCGA CGAATA	60.07	GGTGTGAACAACA CATCCCTT	59.74	247	48	0.83
RSSR578	Unigene23882_All	(AC)17	34	TATTCCCCACCATC TTTTTGG	60.89	ACCACATCAGTCA CCCCCTA	60.24	272	48	0.83
RSSR934	Unigene36690_All	(AG)18	36	GTTTAGGGTTCTCT GCGTGC	59.88	TGCTGCTTCTCTTG CTGAAA	60.01	194	48	0.83
RSSR129	Unigene5175_All	(TAATCC)4	24	CGCCACAATTTATA ACATAGATTCC	60.01	GAATTAGCCGGAG ATGTGGA	60.04	252	48	0.85
RSSR945	Unigene37186_All	(CT)14	28	TGAGCAAACCTCA AGCAACA	60.57	GAAGTGCCTTTGC GTCTTTC	60.00	236	48	0.85
RSSR1123	Unigene46565_All	(TG)10	20	GAGACCTCTTGAG GGTGCAA	60.39	AAGCACGGAATGA GTACGCT	59.90	227	48	0.85
RSSR1438	Unigene64623_All	(AAACGA)4	24	CCCCCTCCCTCAAT TTAACA	61.03	CCTATAGGTGGCG GAGATGA	60.05	236	60	0.87
RSSR29	Unigene732_All	(GAA)10	30	AGCAGATCGTAAG CCCAAGA	59.98	AGTTTCCTCACCA CCACCTG	60.00	185	48	0.87
RSSR1603	Unigene73845_All	(TC)14	28	TGCATTGGATGAAT GAGATGA	60.03	CAAAGCCGTTTTT GGTTTTC	59.60	202	36	0.88
RSSR1089	Unigene43853_All	(AAGAAA)4	24	TTCTTCGAGGCCCC TTTAAT	60.03	AAACACATTCCTT CCTCCCC	60.17	215	24	0.89
RSSR999	Unigene39139_All	(GAA)8	24	GGAAGAAGACCAA CCGATGA	60.05	GTACTGTTATATTG CTGCTTTTATCTG	57.13	181	60	0.89
RSSR224	Unigene9661_All	(TCT)7	21	TGGAGGAGGAAAT GAAAAGAGT	59.21	AGCCTTGGTAGGG GGTTTAG	59.48	175	60	0.89
RSSR1533	Unigene69994_All	(GA)18	36	GCTTAATTCCCATC AGGCAG	59.67	TTCCACAAGGAAA AACAGGC	60.09	242	72	0.89
RSSR971	Unigene37958_All	(TGT)8	24	AATTGTCCCAAACC CTAGCC	60.19	TATCCCAACCATA AACCCCA	59.87	151	65	0.90
RSSR1116	Unigene46175_All	(TGA)8	24	TGCACTTCCAATGT TTCCAC	59.55	TGCTTGACACCCA AAAACAA	60.13	196	72	0.90

Totally 204 of 279 primer pairs were successfully amplified in Wen 85-5 at an amplification rate of 73.1%, similar to that of conventional grain crops with amplification rate of 60–90% (La Rota *et al*., 2005). Out of 204 primer pairs successfully validated, 91 primers produced high level polymorphism with 45% polymorphism frequency that was agreed with that of 40–89% in other crop plants (Yu *et al*., 2004; Varshney *et al*., 2005; Zhang *et al*., 2012). Variation of SSR polymorphism frequency among plants were affected by many factors, but that largely depended on the genetic differences of plant itself and the representativeness of germplasms sets tested SSR polymorphism (Scott *et al*., 2000; Liang *et al*., 2009). PIC value for a SSR markers reflected level variation of their polymorphism (Liang *et al*., 2009). PIC value of polymorphic EST-SSRs markers produced this study range from 0.33 to 090, in which more than 60% were over 0.70, indicating that it is feasible to capture highly polymorphic SSR markers through *R. glutinosa* transcriptome.

The strains used to evaluate genetic diversity in this study were mainly cultivars and wild strains and have a relative wide genetic background. The clustering analysis indicated cultivars and wild strains were easily distinguished. Grouped results of cultivated strains were basically consistent with that grouped based on RAPD, ISSR (Zhou *et al*., 2007; Wang *et al*., 2008), SRAP (Zhou *et al*., 2010; Zhou *et al*., 2012) markers, but with a handful of difference (Fig. 6). For example, Beijin1, 9302, Jinzhuangyuan and Wen85-5 were divided into different group using different molecular markers (Zhou *et al*., 2007; Wang *et al*., 2008; Zhou *et al*., 2010; Zhou *et al*., 2012). Notably, the genetic relationships among wild strains were dramatically different; some wild had relatively closer genetic relationship with cultivars and some much further, which was identical to results identified by ISSR and RAPD-based markers (Wang *et al*., 2008). A major reason for some wild strains having a more closely genetic relationship with cultivars was that multiple events of hybridization between wild strains and cultivars happened during artificially planting. Their some biological traits, such as leaf shape, size, and color, were thus very close to those of the cultivars (Fig. 7), which are often called wild relative strains. Overall, these wild germplasms owned rich genetic variations that provided excellent candidate to broaden the available genetic resources using for cultivar breeding.

A series of cultivars and wild strains derived from same *R. glutinosa* species were collected in this study, and their genetic relationships were firstly identified by polymorphic EST-SSR markers. These works laid a solid foundation for further determining genetic distance of a wider scale *R. glutinosa* germplasms and assist selection and breeding of new *R. glutinosa* cultivars in future.

Conclusion

In this study, a total of 1,018 SSR (≥18 bp) loci were mined from an *R. glutinosa* transcriptome assembled from leaf and root library constructed by Illumina sequencing technology, in which two nucleotides being the most abundant repeat types, followed by trinucleotides and hexanucleotides. A set of 204 were successfully amplified in the Wen85-5 genome; 91 were found to have a polymorphic content ranging from 0.33 to 0.90, with means of 0.71, among 12 *R. glutinosa* strains. Analysis of genetic diversity in 36 *R. glutinosa* germplasms demonstrated that cultivated strains are notably separated from wild strains, that different wild strains had distinct genetic relationships, with some very close to cultivated germplasms. These newly developed SSRs will provide an important information platform for accelerating *R. glutinosa* breeding and promote the application of marker-assisted selection (MAS) systems in *R. glutinosa* breeding.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (Nos. 81274022, 31271674, 30973875, 81403037), the Research Fund for the Doctoral Program of Higher Education of China (No. 20123515110005) and the Science and Technology Research Key Henan Educational Committee (No. 13A180160).

References

Asadi, A.A. and S.R. Monfared, 2014. Characterization of EST-SSR markers in durum wheat EST library and functional analysis of SSR-containing EST fragments. *Mol. Genet. Genom.*, 289: 625–640

Blanca, J., J. Cañizares, C. Roig, P. Ziarsolo, F. Nuez and B. Picó, 2011. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics*, 12: 104

Botstein, D., R.L. White, M. Skolnick and R.W. Davis, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Human Genet*., 32: 314–331

Cardle, L., L. Ramsay, D. Milbourne, M. Macaulay, D. Marshall and R. Waugh, 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*, 156: 847–854

Chen, J.L., L.Q. Huang, A.J. Shao and S.F. Lin, 2002. RAPD Analysis on different varieties of *Rehmannia glutinosa*. *Chin. J. Chin. Mater. Med.* 27: 505–508

Chen, M., Z. Tan, G. Zeng and J. Peng, 2010. Comprehensive analysis of simple sequence repeats in pre-miRNAs. *Mol. Bio. Evol*., 27: 2227–2232

Choudhary, S., N.K. Sethy, B. Shokeen and S. Bhatia, 2009. Development of chickpea EST-SSR markers and analysis of allelic variation across related species. *Theor. Appl. Genet.*, 118: 591–608

Dutta, S., G. Kumawat, B.P. Singh, D.K. Gupta, S. Singh, V. Dogra, K. Gaikwad, T.R. Sharma, R.S. Raje, T.K. Bandhopadhya, S. Datta, M.N. Singh, F. Bashasab, P. Kulwal, K.B. Wanjari, R.K. Varshney, D.R. Cook and N.K. Singh, 2011. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol*., 11: 17

Fraser, L.G., C.F. Harvey, R.N. Crowhurst and H.N. de Silva, 2004. EST-derived microsatellites from Actinidia species and their potential for mapping. *Theor. Appl. Genet*., 108: 1010–1016

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

Gao, L., J. Tang, H. Li and J. Jia, 2003. Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol. Breed.*, 12: 245–261

Gupta, S.K. and T. Gopalakrishna, 2010. Development of unigene-derived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other Vigna species. *Genome*, 53: 508–523

Jiang. J., L.J. Wang, LW. Liu, X.W. Zhu, L.L. Zhai and Y.Q. Gong, 2012. Development and characterization of cDNA library based novel EST-SSR marker in radish (*Raphanus sativus* L.). *Sci. Hortic.*, 140: 164–172

Joy, N., S. Asha, V. Mallika and E.V. Soniya, 2013. De novo transcriptome sequencing reveals a considerable bias in the incidence of simple sequence repeats towards the downstream of 'pre-miRNAs' of black pepper. *PloS One*, 8: e56694

Kantety, R.V., M. La Rota, D.E. Matthews and M.E. Sorrells, 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol*., 48: 501–510

La Rota, M., R.V. Kantety, J.K. Yu and M.E. Sorrells, 2005. Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genom.*, 6: 23

Liu, T., S.Y. Zhu, L.L. Fu, Q.M. Tang, Y. Yu, P. Chen, M.B. Luan, C.B. Wang and S.W. Tang, 2013. Development and characterization of 1,827 expressed sequence tag-derived simple sequence repeat markers for Ramie (*Boehmeria nivea* L. Gaud). *PloS One*, e60346

Liu, C.Y., C. Yuan, X.T. Li, Z.H. Li, P. Zhao, Z.Q. Qian, J.F. Li and Z.L. Liu, 2015. Development of microsatellites from *Rehmannia glutinosa* transcriptome and evaluation of genetic diversity among *Rehmannia* species. *Biochem. Syst. Ecol*., 59: 177–182

Liang, X., X. Chen, Y. Hong, H. Liu, G. Zhou, S. Li and B. Guo, 2009. Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and Arachis wild species. *BMC Plant Biol.*, 9: 35

Li, M.J., Y.H. Yang, X.J. Chen, F.Q. Wang, W.X. Lin, Y.J. Yi, L. Zeng, S.H. Yang and Z.Y. Zhang, 2013. Transcriptome/Degradome-wide identification of *R. glutinosa* miRNAs and their Targets: the role of miRNA activity in the Replanting Disease. *PloS One*, e68531

Pashley, C.H., J.R. Ellis, D.E. McCauley and J.M. Burke, 2006. EST databases as a source for molecular markers: lessons from Helianthus. *J. Hered.*, 97: 381–388

Parida, S.K., K. Anand Raj Kumar, V. Dalal, N.K. Singh and T. Mohapatra, 2006. Unigene derived microsatellite markers for the cereal genomes. *Theor. Appl. Genet.*, 112: 808–817

Rohlf, F.J., 2000. *NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System*, Version 2.1, User Guide, Exeter Software, New York, USA

Scott, K.D., P. Eggler, G. Seaton, M. Rossetto, E.M. Ablett, L.S. Lee and R.J. Heny, 2000. Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet*., 100: 723–726

Varshney, R.K., T. Thiel, N. Stein, P. Langridge and A. Graner, 2002. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol. Biol. Lett*., 7: 537–546

Varshney, R.K., R. Sigmund, A. Borner, V. Korzun, N. Stein, M.E. Sorrells, P. Langridge and A. Graner, 2005. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Sci.*, 168: 195–202

Wang, Z., B. Fang, J. Chen, X. Zhang, Z. Luo, L. Huang, X. Chen and Y. Li, 2010. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genom.*, 11: 726

Wang, Z., J. Li, Z. Luo, L. Huang, X. Chen, B. Fang, Y. Li, J. Chen and X. Zhang, 2011. Characterization and development of EST-derived SSR markers in cultivated sweetpotato (*Ipomoea batatas*). *BMC Plant Biol.*, 11: 139

Wang, Y., X.E. Li, X.D. Li, J.J. Qi, P. Sun, L.L. Zhou, 2008. Analysis of genetic diversity of wild *Rehmannia glutinosa* by using RAPD and ISSR markers. *Chin. J. Chin. Mater*., 33: 2591–2595

Wu, B., Y. Li, H. Yan, Y.M. Ma, H.M. Luo, L.C. Yuan, S.L. Chen and S.F. Lu, 2012. Comprehensive transcriptome analysis reveals novel genes involved in cardiac glycoside biosynthesis and mlncRNAs associated with secondary metabolism and stress response in *Digitalis purpurea*. *BMC Genom.*, 13: 15

Wu, Z.G., M. Wang, L.Q. Huang, H.Y. Liu, F. Wang and Q.J. Yuan, 2007. RAPD analysis on genetic relationship among varieties of *Rehmannia glutinosa*. *Chin. J. Chin. Mater*., 32: 1865

Yu, J.K., T.M. Dake, S. Singh, D. Benscher, W. Li, B. Gill and M.E. Sorrells, 2004. Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. *Genome*, 47: 805–818

Yuan, M.F. and Y. Hong, 2003. Heterogeneity of Chinese medical herbs in Singapore assessed by fluorescence AFLP analysis. *Amer. J. Chin. Med*., 31: 773–779

Zeng, S., G. Xiao, J. Guo, Z. Fei, Y. Xu, B.A. Roe and Y. Wang, 2010. Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. *BMC Genom.*, 11: 94

Zhang, H.Y., L.B. Wei, H.M. Miao, T. Zhang, C.Y. Wang, 2012. Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genomics*, 13: 316

Zhang, J.Z., Y.F. Wang and J.B. Wang, 2006. Analysis of Genetic Diversity of *Rehmania glutinosa* based on RAPD markers. *J. Henan Agric. Sci.*, 6: 97–100

Zhang, R.X., M.X. Li and Z.P. Jia, 2008. *Rehmannia glutinosa*: review of botany, chemistry and pharmacology. *J. Ethnopharmacol*., 117: 199–214

Zhao, N. and H.Q. Li, 2009. Population Genetic Diversity of *Rehmannia glutinosa* (*Rehmannia*). *Henan Sci*., 27: 1386–1391

Zhou, C.E., F.P. Gu and Y.Q. Zhou, 2012. Genetic Diversity of *Rehmannia glutinosa* Germplasm Resources, *Guizhou Agric. Sci.*, 40: 4–7

Zhou, Y.Q., F.P. Gu. C.E. Zhou, H.L. Yao, F.Y. Duan and F. Wang, 2010. Genetic diversity of *Rehmannia glutinosa* cultivars based on sequence-related amplified polymorphism markers. *Sci. Hort*., 125: 789–794

Zhou, Y.Q., J.J. Li, F. Wang, F.P. Gu, C.E. Zhou, Z.Y. Zhang and Z.M. Gao, 2007. Genetic diversity of different cultivars in *Rehmannia glutinosa* Libosch. f. hueichingensis (*Chao et Schih*) Hsiao. *Life Sci. J.*, 4: 69–75

(Received 15 September 2015; Accepted 30 November 2015)