African Journal of Biotechnology Vol. 12(9), pp. 921-935, 27 February, 2013
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB12.918
ISSN 1684–5315 ©2013 Academic Journals

Full Length Research Paper

Development, characterization and application of novel expressed sequence tag- simple sequence repeat (EST-SSR) markers in radish (*Raphanus sativus* L.)

Lulu Zhai[1], Liwang Liu[1]*, Xianwen Zhu[2], Liang Xu[1], Lina Jiang[1] and Yiqin Gong[1]

[1]National Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Horticulture, Nanjing Agricultural University, Nanjing 210095, P.R. China.
[2]Department of Plant Sciences, North Dakota State University, Fargo, ND 58108, USA.

Accepted 26 September, 2012

In this study, 51,625 unique expressed sequence tag (ESTs) from a total of 289,621 radish ESTs in the National Center for Biotechnology Information (NCBI) database were used to search for simple sequence repeat (SSRs) by SSRLocate, and 2,917 SSRs in 2,891 ESTs were identified. The SSR marker motifs contained di-, tri-, tetra-, penta-, and hexa- nucleotide repeats, and the number was 945 (32.40%), 1,300 (44.57%), 179 (6.14%), 262 (8.98%) and 231 (7.92%), respectively. The motifs AG/CT (16.15%) and GA/TC (13.58%) were the most abundant type. Among 20 amino acids encoded by trinucleotide, Ser (16.17%) was the most common transcript, followed by Leu (13.00%) and Glu (9.90%). A total of 1,082 EST-SSR primers in *Raphanus sativus* were designed and synthesized, and 864 (79.85%) EST-SSRs were successfully amplified. Polymorphism of these loci was evaluated in a panel of 48 genotypes in *Raphanus*. The polymorphism information content (PIC) values of those primers varied from 0.33 to 0.84, with a mean value of 0.58. Moreover, it was found that 25 out of 53 EST-SSR primers could be successfully amplified in *Brassica* species. For its polymorphism, reproducibility and functionality, these novel EST-SSR markers could be used as a powerful tool for marker-assisted selection (MAS) and genetic mapping in radish.

Key words: *Raphanus sativus* L., expressed sequence tag (EST), simple sequence repeat (SSR), genetic diversity, polymorphism information content.

INTRODUCTION

With the rapid development in modern genetics, several newer molecular marker systems, such as microsatellite and single nucleotide polymorphism (SNP) analysis, are valuable approaches for genetic mapping, comparative mapping, gene tagging and genetic diversity analysis (Varshney et al., 2005; Yadav et al., 2011). Microsatellites, or simple sequence repeats (SSRs), are a group of tandemly repeated short motifs with 2–6 nucleotides distributed throughout the genome both in coding and non-coding regions (Gasic et al., 2009; Cloutier et al., 2012). Due to their high abundance, multi-allelic nature, codominant inheritance, reproducibility, extensive genome coverage and transferability over genotypes, SSR makers have been entensively used as an appropriate and suitable maker in plant genetics and

*Corresponding author. E-mail: nauliulw@njau.edu.cn. Tel: +86-25-84395563. Fax: +86-25-84395266.

Abbreviations: SNP, Single nucleotide polymorphism; **SSRs,** simple sequence repeats; **ESTs,** expressed sequence tags; **RAPD,** random amplified polymorphic DNA; **SRAP,** sequence-related amplified polymorphism; **MAS,** marker-assisted selection; **GO,** gene ontology; **PIC**, polymorphism information content.

breeding (Gasic et al., 2009; Wang et al., 2011).

In general, SSR markers could be developed from either genomic sequences or expressed sequence tags (ESTs). Increasing studies have shown that EST databases could provide reliable and robust sequence resources for gene discovery, genome annotation, and comparative genomics analysis (Kantety et al., 2002; Varshney et al., 2005). The availability of ESTs provides powerful approach for accelerating the systematic identification and development of SSR markers based on computer analytical approaches (Huang et al., 2011). Recently, with the increased available EST sequences in public database, EST-SSR markers have been successfully developed by mining EST databases in several important plant species, including rice (La Rota et al., 2005), wheat (Fu et al., 2006), peanut (Liang et al., 2009), barley (Varshney et al., 2006), cassava (Zou et al., 2011) and potato (Feingold et al., 2005). Moreover, the EST-SSR markers have been widely employed on plant genotyping, cultivar finger printing, genetic linkage mapping and gene tagging (Cloutier et al., 2012). Additionally, several studies have reported that EST-SSRs have higher interspecific transferability between related species or different genera, which could facilitate comparative genomic analyses (Wöhrmann and Weising, 2011; Mishra et al., 2011; Aggarwal et al., 2007).

Radish (*Raphanus sativus* L., 2n=2x=18), belonging to the *Brassicaceae* family, is one of the most important root vegetable crops in the world, especially in East Asia. In recent years, a few molecular marker systems including random amplified polymorphic DNA (RAPD) and sequence-related amplified polymorphism (SRAP) markers have been employed on genetic diversity analysis in radish (Liu et al., 2008; Kong et al., 2011). To date, however, only limited number of EST-SSRs has been successfully developed in radish (Shirasawa et al., 2011; Wang et al., 2007). Shirasawa et al. (2011) constructed a cDNA library consisting of 26,606 *R. sativus* ESTs and developed 3800 EST-SSR markers. There are abundant radish ESTs in the public National Center for Biotechnology Information (NCBI) database, while no studies have been conducted on developing SSR markers from these available ESTs. In order to increase the efficiency of genetic mapping, gene tagging and marker-assisted selection (MAS) in *R. sativus*, a great number of EST-SSR markers still remains to be developed from radish EST databases *in silico* approaches.

The objectives of the present study are: (1) To analyze the distribution and frequency of EST-SSRs in the expressed portion of the radish genome; (2) to develop a comprehensive set of novel EST-SSR markers for radish; (3) to assess their polymorphism in a set of 48 radish accessions and perform a phylogenetic analysis using a large number of EST-SSRs. In this study, the characteristics of totally 2,917 EST-SSRs in *R. sativus* were well demonstrated and a set of novel EST-SSR markers were successfully developed. These newly developed EST-SSR markers would provide a powerful tool for genetic diversity analysis, genetic mapping, gene tagging and MAS in radish.

MATERIALS AND METHODS

Plant materials

Two radish advanced inbred lines, NAU-YH and NAU-DY13 from the germplasm center of Nanjing Agricultural University were chosen for primer verification. A total of 48 genotype of radish with different root colors and origins, and 10 related species in *Brassica* were selected for genetic diversity and transferability study, respectively (Tables 1 and 2).

Sources of ESTs

A total of 289,621 radish ESTs，which comprised of 123,708 in *R. sativus*, 81,524 in *Raphanus raphanistrum* subsp*. Raphanistrum*, 41,935 in *R. raphanistrum* subsp*. Landra*, 40,660 in *R. raphanistrum* subsp*. Maritimus* and 40,329 in *R. sativus* var. *oleiforms,* were downloaded and saved as FASTA format from the NCBI database (http://www.ncbi.nlm.nih.gov/projects/dbEST/) for SSR analysis. The sequences such as vectors and poly-A tail and poly-T stretches of the obtained ESTs were trimmed by cross-match (www.phrap.org) and EST-trimmer (http://pgrc.ipk-gatersleben.de/misa), respectively. After that, the pre-processed ESTs were merged into the same index class if they are from the same gene and finally the non-redundant sequences were obtained by cd-hit software program (http://www.bioinformatics.org/cd-hit/) for further study.

EST-SSRs identification and functional annotation

The non-redundant ESTs in radish were used to search for SSRs by SSR Locate I (http://www.ufpel.tche.br/ faem/ fitotecnia/ fitomelhoramento/ falecoosco.html) with the standards of di-, tri-, tetra-, penta-, and hexa-nucleotides with 10, 7, 5, 4 and 4 repeats, respectively.

Gene ontology (GO) annotations for consensus and singleton sequences were assigned using the program Blast2GO. These sequences were submitted for GO annotation to the BLAST2GO program (Götz et al., 2008). Blast2go (www.blast2GO.com; Conesa et al., 2005), a comprehensive bioinformatics tool kit, was used to functionally annotate and analyze the gene or protein sequences. A total of 1,170 ESTs of *R. sativus,* which could be used for primer design, were blasted in the NCBI nr (non-redundant) protein database using the BLASTX algorithm (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). The ESTs were allocated to the corresponding functional categories on the basis of the BLAST searches by GO annotation with 1.0E-25 of E-Value Hit Filter and 2 of GO Weight (Huang et al., 2011; Moccia et al., 2009).

SSRs primer design and polymerase chain reaction (PCR) amplification

A total of 1,082 primer pairs were designed in accordance with the criteria of predicted product size ranging from 100 to 500 bp, the GC percentage between 40 and 60%, primer length of 21±3 bp, of which 22 bp was optimum and melting temperature at 50 to 60°C. All the primer pairs were synthesized by Genscript Biotech Co., Ltd. For PCR amplification, genotyping, polymorphism analysis and transferability study, genomic DNA of two radish advanced inbred

Table 1. Genotypes used for EST-SSR analysis.

S/N	Accession	Origin	Skin color[a]	Root shape[b]	Maturity
1	‘RG’	Nantong, China	WH	GL	Medium
2	‘LLYB’	China	WH	GL	Medium
3	PI358483	Macedonia	WH	TO	Late
4	PI140428	Iran	WH	TO	Early
5	PI177065	Maryland, USA	WH	GLTO	Medium
6	PI177063	Maryland, USA	WH	TOLO	Early
7	PI183242	Egypt	WH	LO	Early
8	PI135922	Maryland, USA	WH	LO	Early
9	‘DY13’	Nanchang, China	WH	LO	Medium-late
10	‘XBY’	Shanxi, China	WH	LO	Medium-late
11	'Summer Top'	Korea, South	WH	TO	Medium
12	‘LLZDC’	Nanjing, China	WH	LO	Late
13	PI263262	Osaka, Japan	WH	LO	Late
14	Late Spring Top'	Korea, South	WH	LO	Medium-late
15	'All Season White'	Korea, South	WH	TO	Medium
16	‘YDH ’	China	WHRD	CO	Early-medium
17	‘XHT	Nanjing, China	WHRD	GLTOHL	Early-medium
18	‘YH’	Nanjing, China	RD	GL	Early
19	PI436536	Guatemala	RD	TOP	Early
20	PI262942	Leningrad, Russian	RDPIWH	GLTOHL	Medium
21	PI262941	Russian Federation	RDPU	TOCOHL	Medium
22	PI109562	Columbia, USA	RDPU	GLLO	Medium
23	PI121018	Turkey	PU	TOLO	Medium
24	‘XZH’	China	RD	GL	Medium
25	'Sharlokhov'	Alma-Ata, azakhstan	RD	GL	Early
26	‘NJH’	Nanjing, China	RD	GL	Medium
27	‘WH’	China	RD	TO	Medium
28	PI391633	China	RD	TO	Early
29	‘DHP’	Wuhan, China	RD	GLLO	Late
30	‘ZSYZH’	Mianyang, China	RD	GLLO	Medium
31	‘ZQH’	Nanjing, China	RD	GLLO	Medium
32	‘QTCXH’	Nanjing, China	RD	GLLO	Medium
33	'Red King'	United States	RD	GLLO	Medium
34	‘BJXLM’	Beijing, China	GR	GL	Medium
35	‘MTH’	China	GR	GL	Medium
36	‘QTC’	Nanjing, China	GR	TO	Medium
37	‘QTBJQ’	Nanjing, China	GR	TO	Medium
38	‘WXQ’	Weifang, China	GR	HL	Medium
39	‘MTHS-1’	Germany	BL	GL	Medium
40	‘MTHS-2’	Germany	BL	GL	Medium
41	'Winter Black'	Bulgaria	BL	GLTO	Late
42	PI179982	Maryland, USA	WH	LO	Early
43	PI647064	India	WH	LO	Early
44	PI647070	-	WH	LO	Early
45	PI271456	India	WHPU	LO	Early
46	PI177533	Maryland, USA	PIWH	TOGLLO	Medium
47	PI271451	Gujarat, USA	WH	LO	Early
48	PI381011	Washington, USA	WH	LO	Early

[a]BK, Black; BL, blue; BR, brown; GN, green; PI, pink; PU, purple; RD, red; WH, white; YE, yellow; [b]CO, conical; GL, globe; HL, half-long; LO, long; TO, top.

Table 2. Related species in Brassica used for transferability analysis of radish EST-SSR primers.

S/N	Cultivar	Scientific name	Genome
1	Radish 'ZDC'	*R. sativus* L.	RR
2	Nausa-110 PI381011.	*R . raphanistrum* L.	RR
3	Pakchoi	*B.* ssp.*chinensis* (L.) Makino	AA
4	cabbage'SG21'	*B. oleracea* var. *capitata* L.	CC
5	Broccoli 'YX'	*B. oleracea* var. *italica* Planch	CC
6	cauliflower 'YG'	*B. oleracea var. botrytis L.*	CC
7	HJ	*B. nigra L.*	BB
8	'LS'	*B. napus*	AABB
9	'Ethiopia'	*B. juncea*	BBCC
10	Oil rape 'ZY'	*B. napus*	AACC

line NAU-YH and NAU-DY13, 48 genotype of radish with different root colors and origins and 10 related species in *Brassica* were extracted from young leaves using a modified cetyl trimethylammonium bromide (CTAB) protocol (Liu et al., 2003).

PCR amplifications were performed on a Thermal Cycler (SensoQuest Labcycler, Germany) with a total volume of 15 µl containing 10 ng of genomic DNA, 2.0 mM $MgCl_2$, 0.2 mM dNTP, 0.75 unit of Taq DNA polymerase (TaKaRa) and 0.1 µM forward and reverse primer, and the program was pre-denaturation of 2 min at 94°C, 35 cycles of 40 s denaturation at 94°C, 55 to 60°C (varying with the Tm of the different primer sets) of annealing for 45 s, and 1 min at 72°C, followed by 7min extension at 72°C. The PCR products were runned on 8% non-denaturing polyacrylamide gels at 160 V for 2 h and visualized with a rapid silver staining method (Bassam et al., 1991; Liu et al., 2008). Partial amplification segments were recovered from polyacrylamide gel electrophoresis (PAGE) gel with AxyPrep DNA gel extraction kit (Axygen Biotechnology Co., Ltd.) and cloned with T-A cloning kit (TaKaRa Biotechnology Co., Ltd.), the positive clones were sequenced with ABI 7330 (Beijing Genomics Institute Co., Ltd.).

Analysis of polymorphism and genetic diversity

Total of 48 different radish genotypes consisting of *R. sativus* and its wild relatives (Table 1) were applied to value polymorphism of these loci and clustered for phylogenetic analysis basing on the estimated genetic distance. The EST-SSR alleles were converted into a binary matrix where the presence of an allele was scored as 1, its absence was scored as 0 as per NTSYS data format. Through the data matrix, a dendrogram was constructed by the unweighted pair group method using arithmetic averages (UPGMA) methodology.

For primers that produced the exact fragments after PCR reactions, the amount of alleles was noted and the polymorphism information content (PIC) value (Smith et al., 1997) of an SSR locus was calculated according to the following equation:

$$PIC = 1 - \sum_{i=1}^{k} Pi^2$$

Where, *pi* is the frequency of the *i*th allele out of the total number of alleles at an SSR locus, and k stands for the total number of different alleles at the locus.

Detection of the transferability of EST-SSR primers

To assess the transferability for cross-species amplification of the EST-SSR markers, 12 related species crops in Brassicaeae family were used (Table 2). In addition, DNA bands, the expectant size fragments in the non-denaturing polyacrylamide gel were recovered. Furthermore, the products obtained after further PCR of the recovered fragments as template DNA were purified and sequenced. And sequence alignment was conducted using ClustalX software (http://www.seekbio.com/DownloadShow.asp?id=2247) (Aggarwal et al., 2007).

RESULTS

Identification and characteristics of EST-SSRs in radish

After removal of the redundant sequences, no repeat sequences or repeat with less than four motif frequency, a total of 51,625 unique EST sequences were obtained. Basing on the computer searches with SSRLocatel software, a total of 2,891 ESTs were identified to have at least one putative SSR and 2,917 SSRs were identified, indicating that merely 5.60% ESTs contained microsatellites. Among the non-redundant EST-derived SSR repeats, tri-nucleotide, accounting for 44.57% of total SSRs, was the highest frequency repeat unit, which followed by di- (32.40%), penta- (8.98%), tetra- (7.92%) and hexa- nucleotide (6.14%), respectively (Figure 1). The lengths of SSRs were varied from 20 to 114 bp among all repeat types with a mean of 24.94 bp (Table 3). The maximum motif with GAACCA/TGGTTC (3.46%), AGAGA/TCTCT (4.58%), AAAG/CTTT (7.61%), GAA/TTC (14.58%), and AG/CT (49.84%) was found in hexa-, pente-, tetra-, tri-, and di-nucleotide, respectively.

The twenty types of major motifs, the frequencies and total length of individual SSR units were exhibited in Table 4. Among the targeted motifs of EST-SSRs, the di-nucleotide motif AG/CT with a total length of 12,338 bp was the most abundant with a frequency of 16.15% (Figure 1), followed by GA/TC with a total length of 10,306 bp had a frequency of 13.58%; while the tri-nucleotide motif GAA/TTC, AGA/TCT and ATC/GAT had a frequency of 6.62, 6.24 and 4.56%, respectively (Table 4).

Table 3. Characteristics of different repeat motifs in the radish (*R. sativus*) SSRs.

Repeat type	No.	Maximum length (bp)	Total length	Average length (bp)	No. and percentage of the maximum repeat motif
Dinucleotide	945	108	24404	25.82	AG/CT(471,49.84)
Trinucleotide	1300	114	31965	24.59	GAA/TTC(193,14.85)
Tetranucleotide	179	56	4116	22.99	AAAG/CTTT(15,7.61)
Pentanucleotide	262	70	5740	21.91	AGAGA/TCTCT(12,4.58)
Hexanucleotide	231	90	6534	28.29	GAACCA/TGGTTC(8,3.46)
Total	2917	114	72759	24.94	

Table 4. Frequency distribution of major SSRs based on main motif sequence type.

Motif type	No.	Total length (bp)	Percentage (%)
AG/CT	471	12338	16.15
GA/TC	215	10306	13.58
AC/GT	15	982	1.51
CA/TG	27	634	0.93
GAA/TTC	193	5088	6.62
AGA/TCT	182	4542	6.24
ATC/GAT	133	3360	4.56
AAG/CTT	124	3087	4.25
ATG/CAT	113	2895	3.87
TCA/TGA	104	2604	3.57
CAA/TTG	73	1698	2.50
CTC/GAG	60	1326	2.06
AAC/GTT	41	990	1.41
ACA/TGT	41	1041	1.41
AGC/GCT	38	843	1.30
AGG/CCT	23	504	0.79
CGA/CTG	23	234	0.79
GGA/TCC	21	465	0.72
CAC/GTG	17	399	0.58
AAAG/CTTT	15	460	0.51

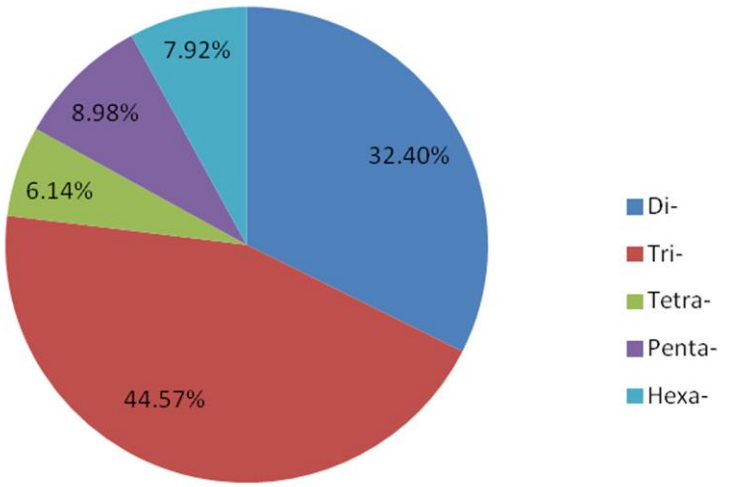

Figure 1. Distribution of nucleotides in radish ESTs.

A total of 20 amino acids were coded by the tri-nucleotide repeat sequences, and their distributions were successfully detected (Figure 2). It was found that the amount of Ser (16.84%) was the top one, which could be coded by AGC, AGU, UCA, UCC, UCG or UCU, and followed by Leu (13.53%) and Glu (10.28%), while Cys (1.36%), Trp (1.00%) and Tyr (0.59%) were much less than any other amino acids.

ESTs annotation

EST sequences of *R. sativus* were first analyzed using Blast2GO, the computational software regularly used for functional annotation. In this study, best hits in BLASTX searches were mainly to *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Thellungiella halophila*, *Brassica rapa* and *Brassica napus* with 1-20 hits per EST sequence (Figure 3). A total of 1,170 ESTs of *R. sativus* containing identified microsatellites were blasted in the NCBI nr (non-redundant) protein database with an expectation value of 1e-25 or less, while about 834 (71.28%) of the set with significant matches were classified into more than 30 categories on level 2. The annotations of each EST were assigned to the GO categories for three main categories consisting of biological processes, cellular component and molecular function (Figure 4).

According to the GO terms retrieved, the most abundant genes were involved in metabolism (380, 25.97%) under biological processes (Figure 4A). When grouped according to cellular component (Figure 4B), the ESTs could be assigned into five categories and a major subset of ESTs (46.09%) was covered by the GO term of 'Cell' and 34.73% were linked to organelle, while the remaining groups involved macromolecular complex, membrane-enclosed lumen and extracellular region. For molecular function (Figure 4C), the *R. sativus* ESTs could be assigned into 11 categories. The majority (48.11%) of the ESTs was assigned to 'binding' and 'catalytic activity' ranked in the second.

Development and validation of EST-SSRs primers

Based on SSRs flanking sequences, a total of 1,082 EST-SSR primers from 1,170 ESTs were designed and synthesized, from which 864 SSRs were successfully amplified with genomic DNA of two radish advanced inbred lines, 'NAU-YH' and 'NAU-DY13'. Out of 864 SSR primers, 592 (68.52%) primers could generate stable and repeatable bands with expected sizes, while the other EST-SSR primers failed to amplify the expected bands, which mainly attributed to the regions covering intron and extron of the specific gene, or their origin EST sequences with compound SSRs structures. To further confirm the reality and positivity of the polymorphic microsatellite-containing sequence, 20 co-dominant segregation segments were recovered and sequenced after T-A cloning. The results show that the sequences were in high accordance with the original EST sequences, indicating that the developed SSR primers were highly specific.

Application of radish EST-SSRs in genetic diversity analysis

Ninety-seven informative EST-SSR primers were selected for genetic diversity analysis in a panel of 48 radish genotypes. A total of 72 (74.23%) primers showed polymorphism in 48 radish lines and totally 263 alleles were detected (Figure 5). The number of alleles per locus varied from 1 to 7, with the average of 2.6. The PIC values varied from 0.33 to 0.84, with a mean value of 0.58. The sequence of the forward and reverse primers, the repeat motif, annealing temperature, PIC value and expected size of the PCR products from the 97 EST-SSR primers in *R. sativus* are listed in Table 5.

Based on the EST-SSR analysis, a phylogenetic tree was constructed with the similarity coefficient between 0.61 and 0.83 by UPGMA methodology (Figure 6). All the accessions could be clustered into four main groups according to their origins at 0.62 of similarity coefficient. Group I contained 29 accessions collected from China and other Asia countries including Japan and Korea. In this group, the 'QTC' was quite close to 'QTBJQ' and a strong relation existed between 'WXQ' and 'XLM'. Moreover, all these four genotypes had green skin root. Additionally, 'Summer Top', 'All Season White' collected from Korea and 'PI263262' obtained from Japan have white roots and clustered into the same subgroup. The second main cluster (II) was comprised of 11 wild accessions obtained from different countries in the word. The wild accessions have purple long siliques (more than 20 cm) and deep purple flowers, which was extremely different from the local radish cultivars. In addition, all accessions in the group II were in early maturity. The third group, consisting of two accessions of 'PI109562' and 'red king' were collected from USA, indicating there is close relationship between these two accessions with red skin root. Group IV contained six accessions mainly collected from Europe. Accession No. 22 was much closely related to accession 'Red King', both of which were originated from German with black roots. Moreover, the 'PI262941' and 'Sharlokhov', collected from different countries, were clustered into one group, indicating that their parental lines may be involved in the same germplasm pool. The results from cluster analysis suggested that the genotypes of the *Raphanus* accessions were correlated with their geographic origins, although the genetic variation exists among different plant species.

Transferability of radish EST-SSR markers across *Brassicacaea* species

A set of 53 randomly selected EST–SSRs were evaluated

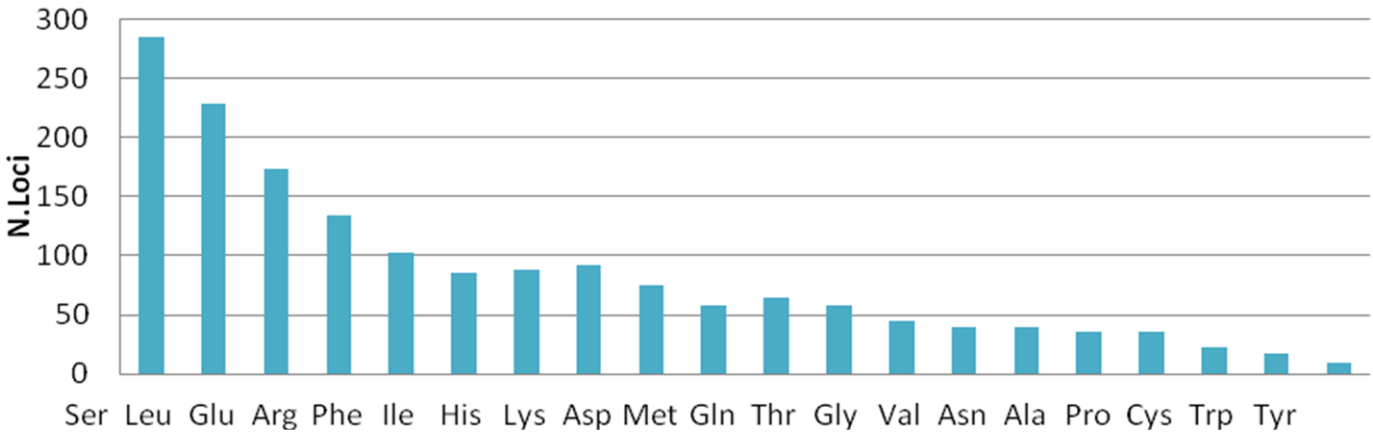

Figure 2. Distribution of amino acids synthesized by trinucleotide.

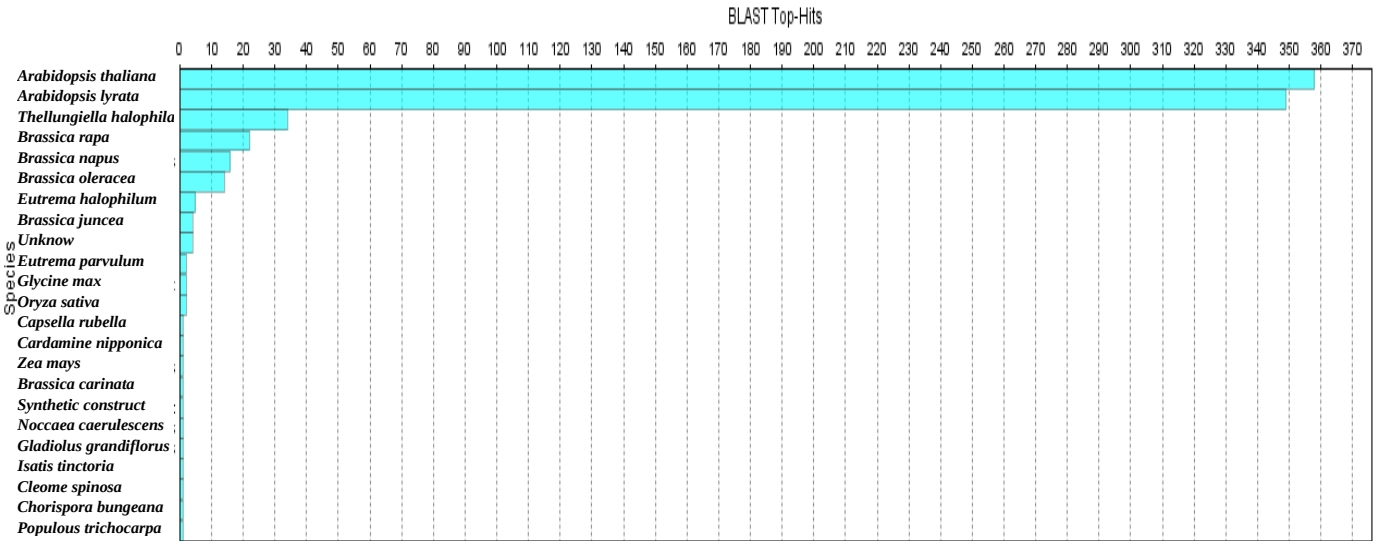

Figure 3. Top-species distribution of *Raphanus sativus* ESTs blasted on NCBI database.

using genomic DNA of 10 plant species belonging to Brassicacaea family (Table 2). Overall, a total of 25 (47.17%) primers could produce repeatable and stable bands in all 10 plant species while the other 28 primers could merely be amplified in some of these species. Furthermore, the specific markers that showed polymorphic in *Brassica* species were further sequenced. The DNA sequences were subject to alignment analysis and the SSR motifs with different length were identified (Figure 7). The results showed that a SSR primer could amplify fragments with different sizes visualized in the polyacrylamide gel, indicating that different copies of motifs were observed among different species of Brassicacaea family. Moreover, the regions flanking the SSRs were sufficiently conserved (Figure 7). In conclusion, the high level of transferability of these EST-SSRs demonstrated that the newly developed radish EST-SSR markers were applicable and reliable in some related plant species in Brassicacaea family as well.

DISCUSSION

Identification and characterization of EST-SSRs in radish

In recent years, the EST sequences of several plant species have been widely employed for identification and characterization of SSRs in various genetic studies (Varshney et al., 2005; Narina et al., 2011). In this study,

Table 5. Details of newly synthesized EST-SSR primer pairs in *R. sativus* L.

S/N	Primer	Core motif	Primer sequence (5′-3′) F	Tm	Primer sequence (5′~3′) R	Tm	Allele No.	Expected size (bp)	PIC
1	NRSr-2	(CTC)9	CGAGGAGAGGAAAGAAGAGTAT	56	CTATTACTAACCGCCACTTCAG	56	2	168	0.48
2	NRSr-3	(CTC)7	CCAAACAAACGATACAGAAAAT	57	AGAAGAGAAGGAGGAGGAGACT	56	3	217	0.62
3	NRSr-7	(GATTC)4-(AG)10	AGAGCGAAAGAAAATACAAACA	56	GGAGTAGAGCTTGAAGGAGATT	56	3	174	0.66
4	NRSr-9	(CTT)7	TAACAAACAAGATCTTCCCAAC	56	CAGTTTTGAGACTGAGAATTGA	55	2	255	0.50
5	NRSr-14	(GATTC)4	GAAATCTATAGCTTCTCCGACC	56	ATCAAATTTCCCTTTCTCATCT	56	1	262	
6	NRSr-15	(GA)10	AGAAACAATCTTTCGTTGAAAA	56	GTGAGCTACGGTGATGAGTATT	56	1	184	
7	NRSr-17	(AG)10	GCATAGACCCTACACTCAGAAA	56	TCATTAGTCTAAAATGGTTGGG	56	2	468	0.49
8	NRSr-25	(ACACAT)6	CGATCGGAAATAGAGGTAATACT	56	TCTCCAGTTCCTATCAGTGTTC	56	3	352	0.66
9	NRSr-30	(GATCG)4	TCTTAGGGTTTTGATCGAATTA	56	ACAAACAATCTAAAAACCTCCC	56	2	274	0.47
10	NRSr-39	(GTTT)5	TCAGCCAATTAATACCAAAATC	56	TTATTTGGAAATGTTGAACACA	56	2	222	0.50
11	NRSr-40	(CAACAG)4	AGATGATGATGGCTCAGTTCT	56	AGGAGGAAGATAAGACGATGAT	56	5	200	0.74
12	NRSr-47	(TCG)7	GACGGTCATAACCAGAACACT	57	TGCAAGAGAGTTCTTCAGAGAT	56	2	437	0.49
13	NRSr-62	(TCA)7	GTGTTTGATAAAATGGCTCAGT	56	TTACACATGCAGAGAGAGACAA	56	1	497	
14	NRSr-68	(TGA)8	AAACAAACAAAGTAAAAGCCAA	56	TAATGAAATGGTCCAAGAAGAG	56	4	276	0.71
15	NRSr-71	(ACACC)5	TTTCTTCTCCTCAGTGGTTAGA	56	AACAGGATGAAGTCAGAGAATG	56	4	305	0.68
16	NRSr-73	(GAA)7	GGAAGAACAAAAACAAAAACAA	56	ACAAGGCATAGAAACATGGTAG	56	7	163	0.79
17	NRSr-77	(GAGAGG)4	CTTTGTTGTTTGGTAAACCTGT	56	CTTAATTGATGCCCTTATGTGT	56	5	276	0.79
18	NRSr-80	(TGA)8	CAAAGAGACTCAAAGAGACCAG	56	TGCTCTCTCTCTTCTGTCAAAT	56	1	406	
19	NRSr-84	(GAA)7	GCACTATGTAAAGAAGACGCTC	56	AAGTTCTTCTTCCTCCTCCTCT	57	4	319	0.75
20	NRSr-86	(TCC)8	GATAAACAAAAGTCACACAGCC	56	GGATAGTTATGAAAGAGGGAGG	56	4	195	0.73
21	NRSr-88	(AAAG)5	AAATTACAAAACAAGCAACACC	56	CTGCTAAAGCCTAAAGCAAATA	56	5	138	0.78
22	NRSr-103	(GA)16	ACAGATTAACCCCAAAAAGAAT	56	ACGACGACGACTTAAGTAGAGA	56	2	156	0.49
23	NRSr-105	(TCT)7	CTGTTCCTGACATTGTTCTTCT	56	ACTTCACCAAAATCTCTCCTTT	56	2	264	0.50
24	NRSr-127	(AG)10	GAAAGATCAAACATCATCCAAT	56	TACTTGGAGGGTTTTACAGTTG	56	5	202	0.76
25	NRSr-133	(TTACCG)4	AAACACCTTCCTCCACTAATCT	56	GAGGGTTTAGAGAAGAGCTGAC	56	6	162	0.71
26	NRSr-136	(CAT)9	GCCACTAAATAGTAACAATGGC	56	GCTCATGGACGTAAACTAACAT	56	5	425	0.68
27	NRSr-138	(AGT)8	GGATTTAGTCGCTTGCATATAG	56	AAAGAAGAGAGAACTCCCAAAC	56	2	186	0.44
28	NRSr-148	(GAATC)4	ACTTCGTGAAGGTGTTGTAGAC	56	ATGAACATTTCAGATTAGCAGG	56	4	407	0.63
29	NRSr-173	(GAC)8	GGAGATCCTAGCTAATTCGTC	55	TATCTGAGGGAGAGATTTGTTG	56	4	120	0.67
30	NRSr-189	(TC)10	ATCTCGCTGTAGGGGTTTCTA	58	AGAGGATTTGAAGGAGAGTGAT	56	2	173	0.33
31	NRSr-193	(CAT)9	CACAGTAATTTCACACATGGAG	56	AGAGAGACTACTTTGGTGCTCA	56	1	214	
32	NRSr-226	(AG)11	TCAAATCTCTTCACACTCCTTT	56	GAGAGTGTTTTTAGCTCCACTG	56	4	139	0.75
33	NRSr-230	(TG)11	CATACAAATCACAGTGCTGAAG	56	ATGGTTGATGTCAATCTCAAAT	56	7	257	0.84
34	NRSr-240	(TGATG)4	CAATGATCTGACTGCTGTAGTG	56	TTCTCTGGTAATGACAAATTCA	55	4	198	0.71
35	NRSr-292	(ATC)11	CTTGCCCACTACTAACATCATT	56	AACTTGTAAAAGGCTTGGAAAT	56	1	185	

Table 5. Contd.

36	NRSr-301	(TTC)8	CACAACCCAACTTACTTTGACT	56	GTGTTGAGATTGAACCTTGATT	56	3	174	0.66
37	NRSr-304	(TC)13	ATGGTGATTATTATTTGCCTTG	56	TACGAGAAGAAAGAAGGATTCA	56	2	156	0.45
38	NRSr-319	(AATC)5	CTGCTTTTCTCTCTCTCTCTCA	56	CTTACAACTTCACTTCGCTCTC	56	2	100	0.50
39	NRSr-332	(TATCA)4	TGGACAAGAGTACACATTTTGA	56	TGATGATGAAGAAGAAGAGGAC	56	2	191	0.44
40	NRSr-341	(ACA)7	GAGGGTCGACGCTTGAAC	59	CCTCTTGGATTTTGACAATAGA	56	6	245	0.83
41	NRSr-375	(TTG)8	CAATCCACGAGACAAAAAGTAT	56	TGAACAGAGAAGAGAAGGAAGA	56	2	393	0.50
42	NRSr-382	(GTT)8	GGATACCCGATTAAGAGACTTC	56	CATGTTCAGCTCTCTTCTTCTC	56	1	471	
43	NRSr-383	(CT)12	GAAAAGCCAAAGGCCAAG	58	TCAAAGTAAAGAGAAAGACCGA	56	2	270	0.50
44	NRSr-387	(AAC)7	GGTGATGAGAGGAGGGTTAC	56	TGAGTGTTGTATCCTTGCAGTA	56	2	139	0.47
45	NRSr-388	(AG)15	AGTCTATCTTCTCTCCGTCGTT	56	CCATAGATTCTTCAGAACAGGA	56	1	443	
46	NRSr-389	(ATTGG)6	GGACAATCTCTTTCCACAGTTA	56	AGGTTAAGTTCAGTTCGGTTTT	56	2	341	0.49
47	NRSr-402	(GA)12	AAAGATAAGAAGGAAAGGATCG	56	GAACAATTGAAGATGGAGAGAA	56	1	221	
48	NRSr-417	(ATCCA)6	AAAAGAAGGAAAAGCATCTCTC	56	GGACAATCTCTTTCCACAGTTA	56	3	180	0.54
49	NRSr-423	(TC)10	ACATCTATCCATCTTCTCCGAT	57	TGTACTGTTTAAGACCTTGGCT	56	1	441	
50	NRSr-444	(AGT)9	AGAGACTCCATCTCCTAAAAGC	56	TTACCATACAGATAGGCAAAGG	56	1	446	
51	NRSr-449	(TCC)7	GCAAGGACCTAATTAACATCAA	56	AGCTTCTTTGTTGAGAACACAT	56	3	212	0.65
52	NRSr-450	(TC)13	ATTTACTCCATTTTGATCATCG	56	AACCAAATCTGTTTGTGACAGT	56	3	185	0.66
53	NRSr-481	(CAT)9	TATTTTCTCCATTTCCTGTCTG	56	AGATTTGAATTTTGTTGTCTGG	56	3	223	0.64
54	NRSr-484	(ACA)9	CTGCAGTATCCATCTTGACAAT	57	AAGAGTTCATGTCACTCGTCTC	56	5	152	0.78
55	NRSr-491	(CAT)7	GACTGAGGTTTATGTCTCGGTA	56	AGAAGAACGTTAGCAACTTTGA	56	4	193	0.75
56	NRSr-493	(AGAAG)6	GTTGCATTTGGATTTTGTAAAC	56	TGTAACTTTGATGTGGATGATG	56	2	288	0.48
57	NRSr-505	(TG)11	AAGATCATCGAGGAGGTAATCT	56	TCTGTGACAAAGATAAAGCAAA	55	2	354	0.50
58	NRSr-508	(CT)12	TACCAAGGTCTCAAGATGTCTC	56	CGATTCCAAATCAAAGTTACAT	56	2	322	o.49
59	NRSr-522	(CTCTTC)5	GTTGGGTCAAGTAGAAGACAGA	56	AGAAAAAGGAAGAGACTTTTGG	56	3	196	0.63
60	NRSr-534	(AGAGGG)4	GACGTAACACTGGGAATAAAGA	56	GCTCCCTCTTCTTTCAGTAGAT	56	1	312	
61	NRSr-546	(TGA)7	AAACATCTACGTGAACCTTTTG	56	GATATCATTGACGATGGAACTC	56	3	361	0.63
62	NRSr-558	(GAACA)4	CCAAAACAAGACTGAAACAAAT	56	AAGAGATCAAAAACACCAAATG	56	2	222	0.37
63	NRSr-559	(CTTTTA)4	TTTAGCGGTGGATTTTACATAC	56	GAAAGTTGAAGTTGTGACGAGT	56	2	213	0.50
64	NRSr-562	(TCA)7	AGGAAGAGTATCTCATCAGGGT	56	TCAGAGGCATCTTCTGTGTC	56	1	170	
65	NRSr-564	(CAA)7-(GAAGAG)4	AAACTTTGCTTTTCATTTTCAA	56	CTCATCCTCGTCTAGTTCTTCA	56	3	249	0.67
66	NRSr-573	(TCT)16	AGAGAGCTTTATCACAACCAAA	56	ATGAGAATGTCTCCAAGAAGTG	56	1	451	
67	NRSr-631	(TTC)7	AACCTAAGATTACACCGGAAAT	56	AAGATGATGATGATGAAGAAGC	56	1	109	
68	NRSr-635	(TTC)7	TCTCTAGTGGCAAGTGAATTTT	56	TATGTTAGAAGGGAAAAGCAAA	56	2	475	0.49
69	NRSr-659	(CT)13	GATTGAACTCGTACTTGAGGAG	56	ATCAAAAACAATCGATCAAAAG	56	1	332	
70	NRSr-663	(TCT)8	TATAATCAATGCCCAAAGAAAG	56	AGAGATTCAAGCTTTGTTCAGA	56	2	179	0.50
71	NRSr-666	(CATA)5	CCTGTTTACTGCTACAAAGCTC	56	AGCAGAGATGCTTATGAGGTTA	56	2	200	0.48
72	NRSr-667	(TGTA)5	AGCAGAGATGCTTATGAGGTTA	56	GAGCTTACTTTGACACAAGGAG	56	1	133	

Table 5. Contd.

73	NRSr-669	(TAGT)5	ATATCACCATCTTGTGGAATTG	56	AATCATATTGGAGTGCATGTCT	56	2	409	0.50
74	NRSr-730	(CTT)11	AGACAAGGTGTTTCTTCATCAG	56	GAATGATAAGTTACCAGAGGCA	56	4	238	0.70
75	NRSr-734	(GA)10	GTTGTTATACCACAGCGACTTT	56	ATACCTCGTAGACCTTCTCCAT	56	2	435	0.50
76	NRSr-739	(ACAAA)4	AAAAATTGTTTCCTTTTGTCGT	57	GAGCATGCTTACAAGTCTCTTC	56	2	223	0.50
77	NRSr-755	(AAGGAG)7	GCTTCGTCCTCCTTACTATTTT	56	TAGCGATACCTAAACCTTGAGA	56	2	479	0.50
78	NRSr-767	(AAG)7	CGTATACAATCTTATGCGTGTG	56	GTGAAACCATTCTTCTCTTGAA	56	1	164	
79	NRSr-825	(CAA)10	CCCAAACAACATTACATAAACC	56	AGGTATTCTGCAACAACTCATC	56	1	391	
80	NRSr-840	(AG)10	TGTGGGATAGAGATAGTTTTGG	56	GACCAAGTAACAATATGACGCT	56	1	201	
81	NRSr-845	(GAA)7	CGTAACTCTTTTGATTTTCTCG	56	TCCATTTAATGATAACAAAGGC	56	1	388	
82	NRSr-856	(AG)20	AGAAGGGAAGAAGAAAACAAAC	56	GGACTAGAGATGAAATCAACGA	56	1	119	
83	NRSr-859	(AGG)8	AGGGAGGTTCTACGATGAGTAT	56	TGTAAGGAGTTGTTCCTTTGAC	56	4	119	0.68
84	NRSr-866	(CTC)7	TCTCATCTGAGACCCACTAAAC	56	TAACCAATCATCTGTATCGGAG	57	2	439	0.50
85	NRSr-869	(AAAAAC)4	AAAACTCACTTTAAAGAAAGAAAA	53	ATACAATTTTCATGGGAAATCA	56	2	281	0.36
86	NRSr-883	(GAG)7	TTGGAGGTGTAACAAGACTCTC	56	GGGAAAGAGAAAGGTGTGTAT	55	2	386	0.39
87	NRSr-908	(GAG)7	TCTAAAGCTATTTGGATTCAGC	56	CAAAGCTTGAGAGTTACCAAAC	56	6	208	0.83
88	NRSr-909	(CTT)8	GGGCTTGTAACTAGCTACTTTTCT	57	TAGATATGGCAATCCAAGACTC	56	1	311	
89	NRSr-949	(ATC)7	GTTCTTGATGAAAGATACGGAA	56	TCGATTGATGAGACCATGTAT	55	2	400	0.33
90	NRSr-959	(CAT)7	GTAGCTCGGTAAGTGAAGTTTG	56	GATGGTGAGGAAGTTGAAAGT	56	5	119	0.78
91	NRSr-1012	(TTG)7	ATGGTTAGACTTGCACATTTTC	56	AAAACCTTGACCAGTCCTATCT	56	2	447	0.49
92	NRSr-1014	(GTT)8	GAGTAACGGAACGATTGTTTTA	56	AGCCTCAAAATCCTCTCTAGTC	56	1	290	
93	NRSr-1015	(ACTC)6	ACCAAACATTGAAGAACAAGAG	56	TTGAAGAGAGTGAGAAGGAAGA	56	3	227	0.66
94	NRSr-1019	(AAAGAA)4	CAATCTCGGGTTATTCTCTACA	56	CCAAAGGAAGAGATTTGTGATA	56	4	296	0.74
95	NRSr-1025	(TCA)9	ATGCAAAACATACACACAACAC	56	GTTCAAGGATAACGGTTACTGA	56	2	449	0.50
96	NRSr-1037	(TCCTC)4	AAAAGAAACATCTGATACCTC	50	CGAGCCAGTAGTTGTAGACTTC	56	3	133	0.63
97	NRSr-1076	(ACA)8	ATCTTTCCTTAAACGACTCCAT	56	CTTGTCATCTTCCTTTGTTGTT	56	1	470	

the tri-nucleotide (44.57%) was the most abundant repeat among the 2,917 SSRs from 2,891 non-redundant ESTs, which was in highly accordance with previous reports in sweet potato (Wang et al., 2011), pineapple (Ong et al., 2012) and pepper (Yi et al., 2006). Gong et al. (2010) found that the most frequently occurring microsatellite motif in ESTs of pea was GAA/TTC, which constituted 3.3% of all types of di- to hexa-nucleotide repeats. Among the targeted motifs of radish EST-SSRs, the di-nucleotide motif AG/CT was the most abundant with a frequency of 16.15% (Figure 1), followed by GA/TC with a frequency of 13.58%; while the tri-nucleotide motif GAA/TTC, AGA/TCT and ATC/GAT had a frequency of 6.62, 6.24 and 4.56%, respectively (Table 4). It was also reported that AT/TA (18.4%) was the most frequent repeat followed by CTT/GAA, AG/TC, and AGA/TCT in *Gossypium hirsutum*, while the GA/CT repeat was the most abundant one (24.16%) in *Fragaria* (Han et al., 2006; Folta et al., 2005). The distribution, frequency and abundance of SSRs were highly variable between various plant species, which mainly depend on the size of the dataset in the SSR search criteria and the database mining tool (Varshney et al., 2005).

Additionally, the Gene Ontology annotations were assigned basing on similarities to *A. thaliana* and *A. lyrata* (Berardini et al., 2004; Narina et al., 2011).

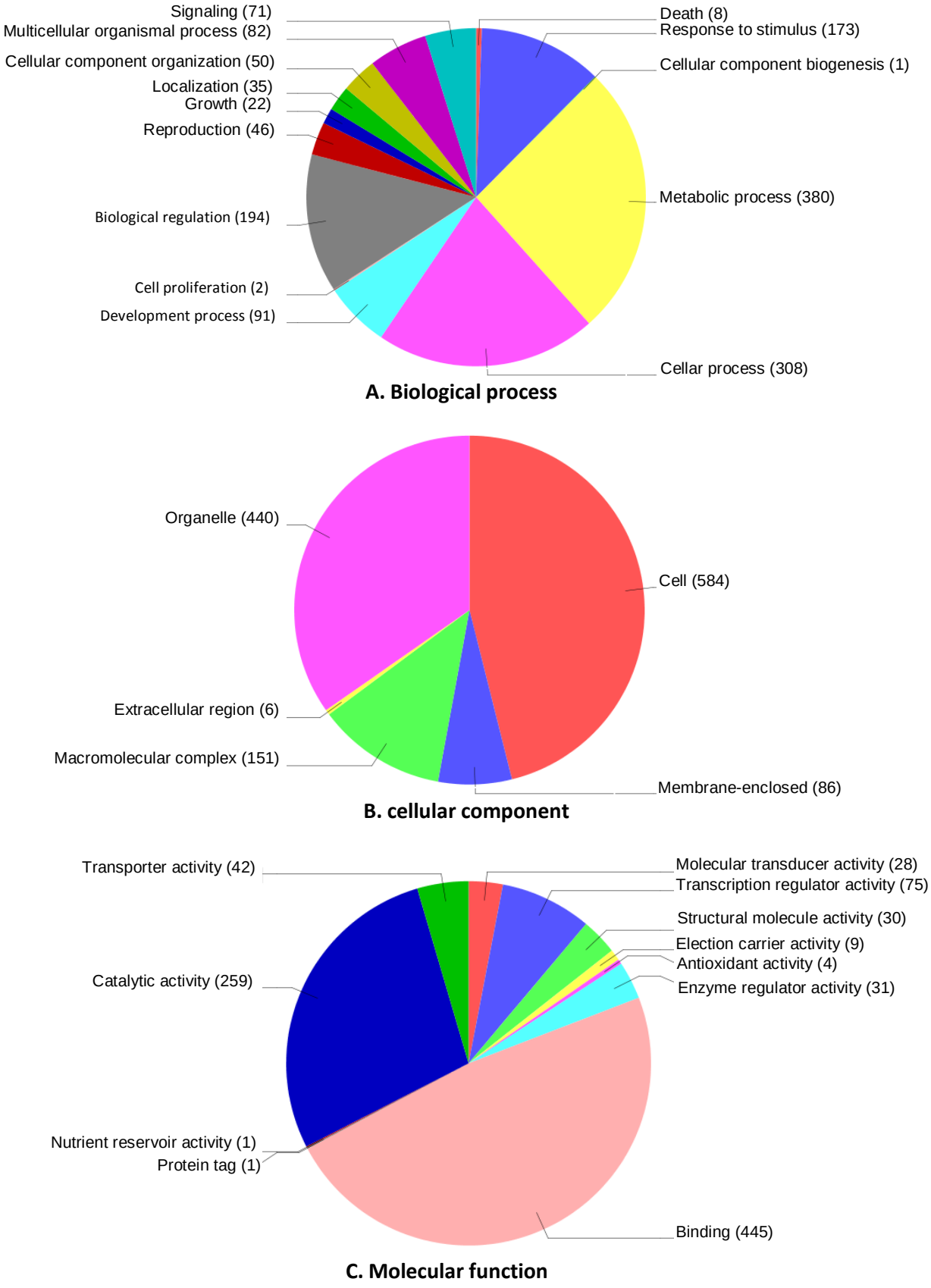

Figure 4. Gene Ontology (GO) classification of the *R. sativus* EST library. The relative frequencies of GO hits for *R. sativus* ESTs assigned to the GO functional categories A. Biological process, B. Cellular component, and C. Molecular function.

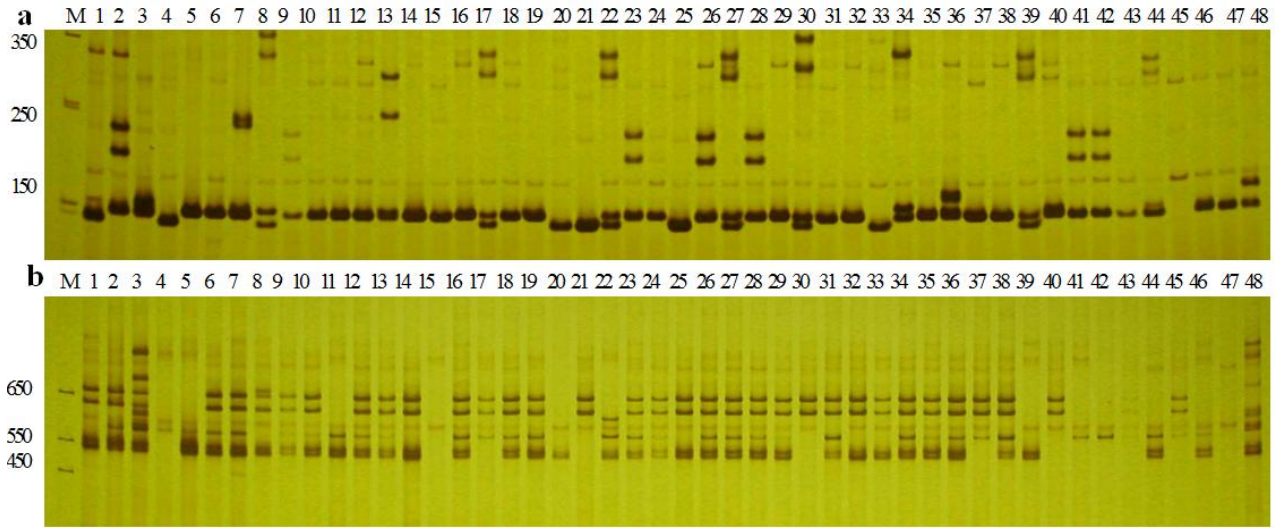

Figure 5. Polyacrylamide gel electrophoresis patterns of microsatellite alleles amplified in radish using EST-SSR markers NRSr-88(a) and NRSr-77(b). M: Maker 100bp; 1-48: Accession numbers of radish, representing cultivar names were listed in Table 1.

Moreover, a great amount of hydrolases, transferases kinases and peptidase were annotated which provide a powerful approach for identification of novel genes involved in the secondary metabolite synthesis pathways (Li, 2010).

Development and application of radish EST-SSRs

For representing coding regions of the genome, the development of EST-SSR markers was particularly attractive for multi-allelic detection, high-transferability across related species and high reproducibility across various laboratories (Narina et al., 2011). Up to now, a great number of EST-SSR markers have been extensively developed in several important plant species, including rice (La Rota et al., 2005), wheat (Fu et al., 2006), cotton (Han et al., 2006), peanut (Liang et al., 2009), walnut (Zhang et al., 2010) and potato (Feingold et al., 2005). In this study, a total of 2,891 ESTs (5.60%) were identified to have at least one putative SSR, from which 2,917 SSRs were successfully identified. The rate was much lower than 10.2% in pepper (Yi et al., 2006), 7.4% in durum wheat (Chabane et al., 2008), and 7.15% in cotton (Han et al., 2006), which is possibly due to the fact that the majority of the ESTs were short sequences or the simple sequence repeats located at either the 5′ or 3′ end. In addition, the genome sizes of those species mentioned above were much bigger than that of radish, which might count for the higher proportion of SSRs in those EST sequences.

To date, only limited number of EST-SSRs has been successfully developed in radish species (Shirasawa et al., 2011; Wang et al., 2007). In this study, a total of 1,082 EST-SSR primers were finally designed from 1,170 non-abundant radish ESTs, from which 864 SSRs were successfully amplified with radish genomic DNAs. The major of *R. sativus* EST-SSRs could generate high quality amplification products, suggesting that ESTs are available for specific primer design. Gasic et al. (2009) reported that some differences may exist between the expected and observed sizes of the amplification products. In this study, the sizes of PCR products amplified by few EST-SSRs were larger or smaller than predicted, which possibly is due to simultaneous amplification of an intron during the PCR, the lack of specificity of some primer pairs or deletions within the genomic sequences (Senthilvel et al., 2008).

Though the EST-SSR markers exhibited lower level of polymorphism comparing with some other genome origin molecular markers, they could also detect different loci across radish accessions as well as other related Brassicacaea species. The application of EST-SSR for fingerprinting and diversity analyses has been reported in several plant species (Zhu et al., 2012; Hu et al., 2011; Chapman et al., 2009). In this study, EST-SSR markers were demonstrated to be efficient for differentiating all 48 radish cultivars mainly based on the geographical distribution. Classification of the examined germplasm, in more or less separated clusters, showed a clear geographical repartition (growth habitat) of the different populations of radish, which were in accordance with those obtained in rice (Sagnard et al., 2011) and wheat (Chabane et al., 2008). The EST-SSR markers were appropriate and superior markers for the discrimination between cultivated landraces and wild species (Chabane

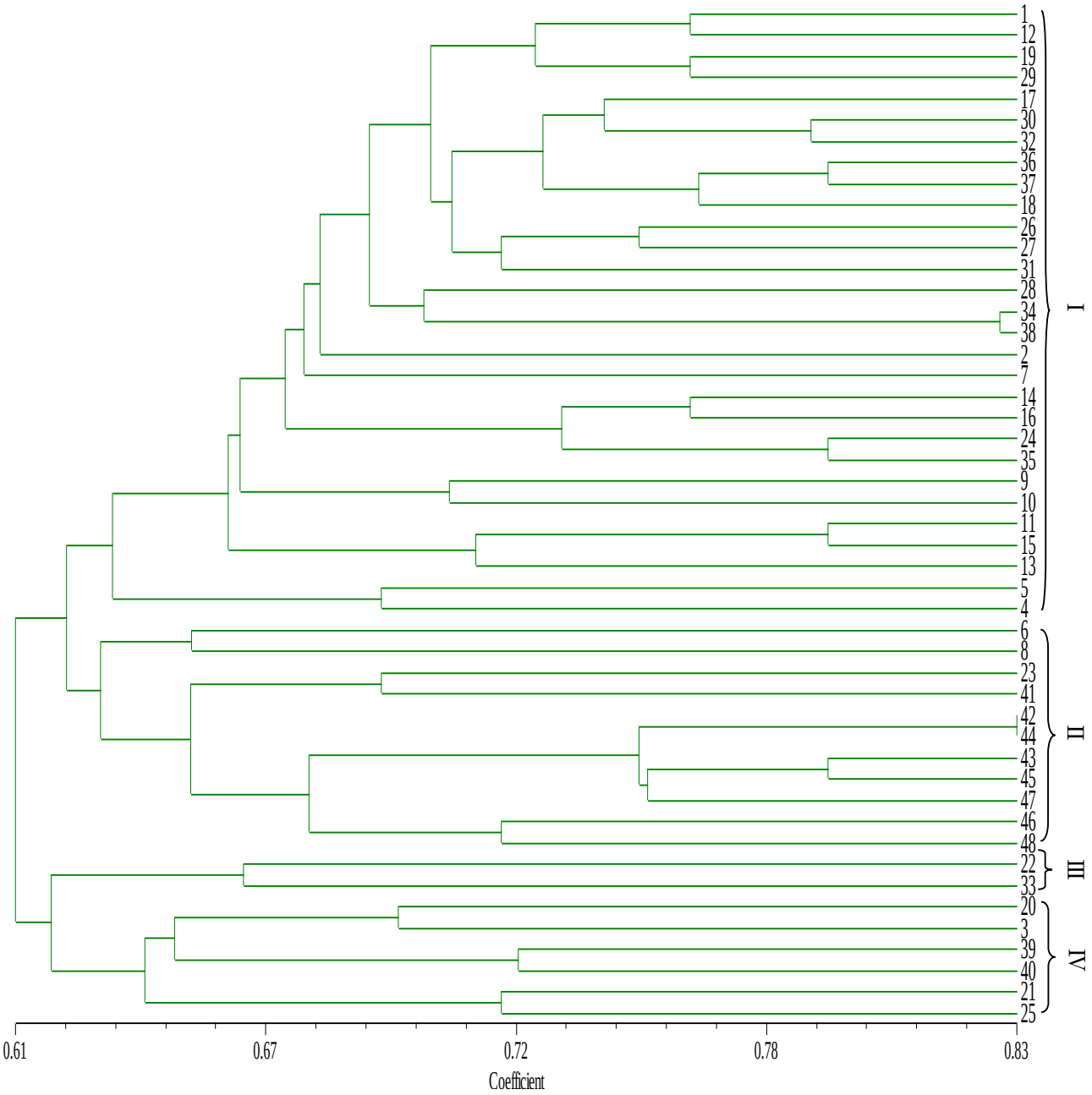

Figure 6. The UPGMA dendrogram of 48 *Raphanus* accessions constructed from EST-SSR marker analysis. The genotype numbers correspond to those listed in Table 1.

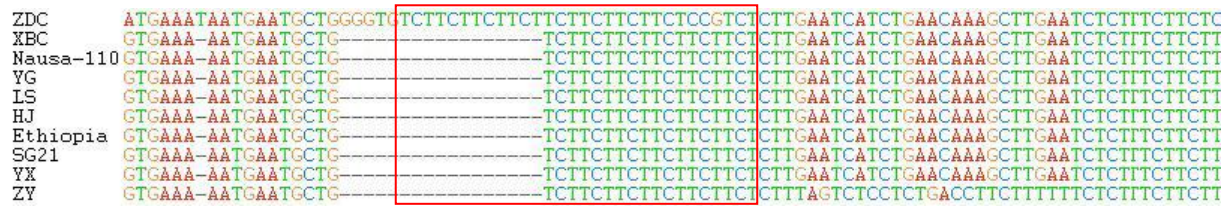

Figure 7. Multiple sequence alignment of alleles amplified from related species of radish in Brassica at locus by primer pair NRSr-663. The details of representing cultivars were listed in Table 2.

et al., 2008).

Transferability of radish EST-SSR markers across Brassicacaea species

Recently, a considerable degree of cross-species transferability of microsatellite markers has been demonstrated within several plant families including leguminosae (Gutierrez et al., 2005), Rosaceae (Gasic et al., 2009) and *Silene* species (Moccia et al., 2009). Wen et al. (2010) reported that the level of transferability of EST-SSRs was higher than that of G-SSRs, mainly due to that EST-SSR markers derived from transcribed genome regions and represented putative functional sequences. In the present study, a high level of transferability from radish to some related Brassicaceae species was observed. The success of high efficient amplification suggests that the flanking regions of these SSR loci are sufficiently conserved. Overall, data mining for SSRs in EST resources and EST-SSR markers development in Radish (these EST-SSR markers are of high transferability) could be applied for facilitating the comparative mapping analyses in some plant species of Brassicaceae family.

Conclusion

The development of EST-SSR markers of the radish at large scale was conducted in the present research. With their polymorphism, reproducibility and transferability, these newly developed EST-SSR markers would be valuable molecular tools for germplasm identification, genetic mapping, gene tagging, comparative mapping and genetic diversity analysis in radish species in the future.

ACKNOWLEDGEMENTS

This work was in part supported by grants from the Program for NCET, MOE (NCET-10-047), the National Key Technologies R and D Program of China (2012BAD02B01), the National Natural Science Foundation of China (31171956), Key Technology R and D Program of Jiangsu Province (BE2010328) and the Priority Academic Program Development of Jiangsu Province (PAPD). We thank Dr Bin Xu from Virginia Polytechnic institute and state university for his critical review and helpful comments during the preparation of this manuscript.

REFERENCES

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh LJ (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.

Bassam BJ, Anollés GC, Gresshoff PM (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Biochem. 196:80-83.

Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G (2004). Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiol. 135(2):745-755.

Chabane K, Varshney RK, Graner A, Valkoun J (2008). Generation and exploitation of EST-derived SSR markers for assaying molecular diversity in durum wheat populations. Genet. Resour. Crop Evol. 55:869-881.

Chapman MA, Hvala J, Strever J, Matvienko M, Kozik A, Michelmore RW, Tang SX, Knapp SJ, Burke JM (2009). Development, polymorphism, and cross-taxon utility of EST–SSR markers from safflower (*Carthamus tinctorius* L.). Theor. Appl. Genet. 120:85-91.

Cloutier S, Niu ZX, Datla R, Duguid S (2009). Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theor. Appl. Genet. 119:53-63.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005). Blast2GO: auniversal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Feingold S, Lloyd J, Norero N, Bonierbale M, Lorenzen J (2005). Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.). Theor. Appl. Genet. 111:456-466.

Folta KM, Staton M, Stewart PJ, Jung S, Bies DH, Jesdurai C, Main D (2005). Expressed sequence tags (ESTs) and simple sequence repeat (SSR) markers from octoploid strawberry (*Fragaria* × *ananassa*). BMC Plant Biol. 5:12.

Fu YB, Peterson GW, Yu JK, Gao L, Jia J, Richards KW (2006). Impact of plant breeding on genetic diversity of the Canadian hard red spring wheat germplasm as revealed by EST-derived SSR markers. Theor. Appl. Genet. 112:1239-1247.

Gasic K, Han YP, Kertbundi S, Shulaev V, Iezzoni AF, Stover EW (2009). Characteristics and transferability of new apple EST-derived SSRs to other Rosaceae species. Mol. Breed. 23:397-411.

Gong YM, Xu SC, Mao WH, Hui QZ, Zhang GW, Ding J, Li YD (2010). Developing new SSR markers from ESTs of pea (*Pisum sativum* L.). J Zhejiang Univ-Sci B. 11(9):702-707.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids. Res. 36:3420-3435.

Gutierrez MV, Vaz Patto MC, Huguet T, Cubero JI, Moreno MT, Torres AM (2005). Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. Theor. Appl. Genet. 10(7):1210-1217.

Han ZG, Wang CB, Song XL, Guo WZ, Gou JY, Li CH, Chen XY, Zhang TZ (2006). Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. Theor. Appl. Genet. 112:430-439.

Hu J, Wang L, LI J (2011). Comparison of genomic SSR and EST-SSR markers for estimating genetic diversity in cucumber. Biologia Plantarum. 55(3):577-580.

Huang H, Lu J, Ren Z, Hunter W, Dowd SE, Dang P (2011). Mining and validating grape (*Vitis* L.) ESTs to develop EST-SSR markers for genotyping and mapping. Mol. Breed. 28(2):241-254.

Kantety R, La Rota M, Matthews D, Sorrells M (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol. Biol. 48(5):501-510.

Kong QS, Li XX, Xiang CP, Wang HP, Song JP, Zhi HY (2011). Genetic diversity of radish (*Raphanus sativus* L.) germplasm resources revealed by AFLP and RAPD markers. Plant Mol. Biol. Rep. 29:217-223.

La Rota M, Kantety RV, Yu JK, Sorrells ME (2005). Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat and barley. BMC Genomics 6:23.

Li Y, Luo H, Sun C, Song J, Sun Y, Wu Q, Wang N, Yao H, Steinmetz A, Chen S (2010). EST analysis reveals putative genes involved in glycyrrhizin biosynthesis. BMC Genomics 11:268.

Liang XQ, Chen XP, Hong YB, Liu HY, Zhou GY, Li SX, Guo BZ (2009). Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and Arachis wild species. BMC Plant Biol. 9:35.

Liu L, Guo W, Zhu X, Zhang T (2003). Inheritance and fine mapping of fertility-restoration for cytoplasmic male sterility in *Gossypium hirsutum* L. Theor. Appl. Genet. 106:461-469.

Liu LW, Zhao LP, Gong YQ, Wang MX, Chen LM, Yu FM, Wang LZ (2008). DNA fingerprinting and genetic diversity analysis of late-bolting radish cultivars with RAPD, ISSR and SRAP markers. Sci. Hortic. 116:240-247.

Mishra RK, Gangadhar BH, Yu JW, Kim DH, Park SW (2011). Development and characterization of EST based SSR markers in Madagascar periwinkle (*Catharanthus roseus*) and their transferability in other medicinal plants. Plant Omics J. 4(3):154-162.

Moccia MD, Oger-Desfeux C, Marais GAB, Widmer A (2009). A White Campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. BMC Genomics 10:243.

Narina SS, Buyyarapu R, Kottapalli KR, Sartie AM, Ali MI, Robert A, Hodeba MJD, Sayre BL, Scheffler BE (2011). Generation and analysis of expressed sequence tags (ESTs) for marker development in yam (*Dioscorea alata* L.). BMC Genomics 12:100.

Ong WD, Voo CLY, Kumar SV (2012). Development of ESTs and data mining of pineapple EST-SSRs. Mol. Biol. Rep. 39:5889-5896.

Sagnard F, Deu M, Dembélé D, Leblois R, Toure L (2011). Genetic diversity, structure, geneflow and evolutionary relationships within the Sorghum bicolor wild–weedy–crop complex in a western African region. Theor. Appl. Genet. 123:1231-1246.

Senthilvel S, Jayashree B, Mahalakshmi V, Sathish Kumar P, Nakka S, Nepolean T, Hash CT (2008). Development and mapping of Simple Sequence Repeat markers for pearl millet from data mining of Expressed Sequence Tags. BMC Plant Biol. 8:119.

Shirasawa K, Oyama M, Hirakawa H, Sato S, Tabata S, Fujioka T (2011). An EST-SSR linkage map of *Raphanus sativus* and comparative genomics of the Brassicaceae. DNA Res. pp. 1-12.

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997). An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.) comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. 95:163-173.

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

Varshney RK, Grosse I, Hahnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A (2006). Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. Theor. Appl. Genet. 113:239-250.

Wang N, Hu JJ, Ohsawa R, Ohta M, Fujimura T (2007). Identification and characterization of microsatellite markers derived from expressed sequence tags (ESTs) of radish (*Raphanus satviu*s L.). Mol. Ecology. 7(3):503-506.

Wang ZY, Li J, Luo ZX, Huang LF, Chen XL, Fang BP, Li YJ, Chen JY, Zhang XJ (2011). Characterization and development of EST-derived SSR markers in cultivated sweetpotato (*Ipomoea batatas*). BMC Plant Biol. 11:139.

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

Wöhrmann T, Weising K (2011). In silico mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across Bromeliaceae. Theor. Appl. Genet. 123:635-647.

Yadav HK, Ranjan A, Asif MH, Mantri S, Sawant SV, Tuli R (2011). EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species / genera. Tree Genet. Genomes 7:207-219.

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

Zhang R, Zhu AD, Wang XJ, Yu J, Zhang HR, Gao JS, Cheng YJ, Deng XX (2010). Development of *Juglans regia* SSR markers by data mining of the EST database. Plant Mol. Biol. Rep. 28:646-653.

Zhu XC, Raman H, Wu H, Lemerle D, Burrows GE, Stanton R (2012). Development of SSR markers for genetic analysis of silverleaf nightshade (*Solanum elaeagnifolium*) and related Species. Plant Mol. Biol. Rep. DOI10.1007/s11105-012-0473-z

Zou ML, Xia ZQ, Ling P, Zhang Y, Chen X, Wei ZS, Bo WP, Wang WQ (2011). Mining EST-derived SSR markers to assess genetic diversity in Cassava (*Manihot esculenta* Crantz). Plant Mol. Biol. Rep. 29:961-971.