Genetics and Molecular Research

Development and characterization of novel EST-SSR markers and their application for genetic diversity analysis of Jerusalem artichoke (*Helianthus tuberosus* L.)

T. Mornkham[1], P.P. Wangsomnuk[1], X.C. Mo[2], F.O. Francisco[3], L.Z. Gao[4] and H. Kurzweil[5]

[1]Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand
[2]Department of Life Science and Technology, Lijiang Teachers College, Lijiang, Yunnan, China
[3]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brasil
[4]Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China
[5]Herbarium, Singapore Botanic Gardens, National Parks Board, Singapore

Corresponding author: P.P. Wangsomnuk
E-mail: prepua@kku.ac.th

Genet. Mol. Res. 15 (4): gmr15048857
Received June 6, 2016
Accepted July 20, 2016
Published October 24, 2016
DOI http://dx.doi.org/10.4238/gmr15048857

ABSTRACT. Jerusalem artichoke (*Helianthus tuberosus* L.) is a perennial tuberous plant and a traditional inulin-rich crop in Thailand. It has become the most important source of inulin and has great potential for use in chemical and food industries. In this study, expressed

sequence tag (EST)-based simple sequence repeat (SSR) markers were developed from 40,362 Jerusalem artichoke ESTs retrieved from the NCBI database. Among 23,691 non-redundant identified ESTs, 1949 SSR motifs harboring 2 to 6 nucleotides with varied repeat motifs were discovered from 1676 assembled sequences. Seventy-nine primer pairs were generated from EST sequences harboring SSR motifs. Our results show that 43 primers are polymorphic for the six studied populations, while the remaining 36 were either monomorphic or failed to amplify. These 43 SSR loci exhibited a high level of genetic diversity among populations, with allele numbers varying from 2 to 7, with an average of 3.95 alleles per loci. Heterozygosity ranged from 0.096 to 0.774, with an average of 0.536; polymorphic index content ranged from 0.096 to 0.854, with an average of 0.568. Principal component analysis and neighbor-joining analysis revealed that the six populations could be divided into six clusters. Our results indicate that these newly characterized EST-SSR markers may be useful in the exploration of genetic diversity and range expansion of the Jerusalem artichoke, and in cross-species application for the genus *Helianthus*.

Key words: Jerusalem artichoke; EST-SSR markers; Development; Genetic diversity

INTRODUCTION

Jerusalem artichoke (*Helianthus tuberosus* L.), a perennial member of the family Asteraceae; it is native to eastern North America and was introduced to Thailand decades ago. Its tubers are rich in inulin making it a healthy choice for individuals with diabetes (Kays and Nottingham, 2008; Alla et al., 2014). In general, Jerusalem artichoke has 2n (6x) = 102 chromosomes, similar to the species *Helianthus annuus*, which is commonly known as sunflower. Jerusalem artichoke has a long history of cultivation as a food supplement all over the world, and this is attributed to its adaptability to varied climates making it easy to plant for local people (Bock et al., 2014). Although Jerusalem artichoke has a very long planting history, international germplasm collections still focus on commercial breeding with the aim of developing both yield and tuber form (Kiru and Nasenko, 2010). Furthermore, plant breeding programs of Jerusalem artichokes still rely heavily on the inner genetic resources, which are essential to accurately identify genotypes and to delineate the various genetic relationships between available accessions in germplasm collections. These resources can then be utilized effectively to preserve and develop the species and to enhance its applications (Debnath, 2014). Although a number of international plant germplasm collections of Jerusalem artichokes have been established, which contain several hundred genotypes, including hybrids and landraces, a standard reference germplasm is still lacking (Kays and Nottingham, 2008). Levels and patterns of genetic diversity and the range expansion of Jerusalem artichoke remain largely unknown. DNA markers were first developed in the 1980s to evaluate variation between accessions within a germplasm or population and also variation at the DNA level between populations arising due to differences within the DNA (Park et al., 2009; Mondini et al., 2009). The markers most commonly used are simple sequence repeat (SSR) or microsatellite markers. These can

be derived from polymerase chain reactions (PCR) (Mullis et al., 1986) and represent the second generation of molecular markers. Their particular strength lies in the fact that they are spread throughout the genome, do not require large amounts of DNA for analysis, are reliable and generate multiple markers, and their use does not require any prior genome information. SSRs are specific regions of DNA that contain either simple sequences or short tandem repeats (STRs). These STRs typically range from one to six base pairs and comprise repeated tandem short sequence motifs (Merritt et al., 2015). A number of public expressed sequence tag (EST) databases now exist. This is a helpful development in the identification of functional markers in suitable candidate genes (Poczai et al., 2013) that can support gene expression analysis and assist in the detection of genetic diversity (Andersen and Lübberstedt, 2003). ESTs are short transcribed sequences that have permitted the development of SSR markers in several species of plants (Gadaleta et al., 2011; Kumari et al., 2013; Şelale et al., 2013; Zhang et al., 2014; Ju et al., 2015). This has resulted in the subsequent identification of genetic diversity through the use of EST-SSR markers (Mujaju et al., 2013; Ramu et al., 2013; Malfa et al., 2014). Although Jerusalem artichoke is an important crop with both economic and cultural significance, very few informative molecular markers have been isolated from its genome. This is unusual in comparison with other crops of similar economic importance. A limited number of markers has been used to examine genetic diversity in the Jerusalem artichoke. These include random amplified polymorphic DNA (RAPD) (Wangsomnuk et al., 2011a,b), sequence-related amplified polymorphism (SRAP), and inter simple sequence repeat (ISSR) markers (Wangsomnuk et al., 2011a). Additionally, Kou et al. (2014) noted that accessions are available in China that are based on amplified fragment length polymorphism (AFLP). However, this method is not adequate to distinguish homozygous alleles from heterozygous alleles using these markers. Compared to other DNA markers like RAPD, ISSR, SRAP, and AFLP, SSR markers have the advantage of co-dominance, reproducibility, hyper-variability, and high coverage in the genome. The development of reliable co-dominant and multi-allelic markers is thus particularly important if cultivar or parental identifications are to be made. They are also important for breeding programs and in studies genetic diversity, conservation genetics, or population structure, which demand marker-assisted selection. Therefore, this study aimed to develop EST-SSR markers from public EST sequences of Jerusalem artichoke, available in the National Center for Biotechnology Information (NCBI) database, and to monitor their performance in the assessment of polymorphism of 25 accessions from five different sources provided by the Plant Genes Resources of Canada, and 35 open-pollinated lines from Thailand.

MATERIAL AND METHODS

Materials

Plant samples

A total of 60 Jerusalem artichoke genotypes were obtained from six different sources (Table 1). Sets of five accessions were obtained from Canada, the United States, Russia, Germany, and France, along with 35 open-pollination lines from Thailand. The 35 selected open-pollinated accessions of Jerusalem artichoke used in this study were derived from *in vitro* culture based on the protocol described by Wangsomnuk et al. (2015) to enrich clonal

diversity. Young leaves were collected from all chosen genotypes of Jerusalem artichoke and dried in silica gel until use.

Table 1. List of 60 accessions of Jerusalem artichoke used in this study and their origin/source and average dissimilarity (AD).

Sample	Accession	Origin/Source	AD	Sample	Accession	Origin/Source	AD
1	JA4	Canada	0.273	31	KK123	Thailand	0.299
2	JA6	Canada	0.300	32	KK126	Thailand	0.327
3	JA37	Canada	0.310	33	KK128	Thailand	0.302
4	JA42	Canada	0.294	34	KK133	Thailand	0.271
5	JA134	Canada	0.298	35	KK137	Thailand	0.304
6	JA55	USA	0.321	36	KK139	Thailand	0.304
7	AMES2736	USA	0.316	37	KK145	Thailand	0.286
8	AMES2722	USA	0.338	38	KK148	Thailand	0.332
9	PI547241	USA	0.341	39	KK154	Thailand	0.302
10	PI503260	USA	0.345	40	KK157	Thailand	0.314
11	JA59	Russia	0.300	41	KK166	Thailand	0.289
12	JA95	Russia	0.299	42	KK169	Thailand	0.290
13	JA105	Russia	0.285	43	KK176	Thailand	0.298
14	HEL65	Russia	0.281	44	KK182	Thailand	0.293
15	CN52867	Russia	0.301	45	KK185	Thailand	0.309
16	JA102	Germany	0.262	46	KK191	Thailand	0.303
17	HEL53	Germany	0.272	47	KK199	Thailand	0.279
18	HEL231	Germany	0.292	48	KK203	Thailand	0.336
19	HEL243	Germany	0.271	49	KK205	Thailand	0.292
20	HEL248	Germany	0.310	50	KK212	Thailand	0.314
21	JA78	France	0.332	51	KK216	Thailand	0.307
22	JA89	France	0.317	52	KK224	Thailand	0.286
23	JA97	France	0.303	53	KK243	Thailand	0.319
24	JA98	France	0.293	54	KK250	Thailand	0.343
25	HEL250	France	0.317	55	KK261	Thailand	0.308
26	KK101	Thailand	0.293	56	KK264	Thailand	0.290
27	KK105	Thailand	0.305	57	KK277	Thailand	0.257
28	KK112	Thailand	0.271	58	KK279	Thailand	0.320
29	KK115	Thailand	0.288	59	KK283	Thailand	0.290
30	KK121	Thailand	0.288	60	KK299	Thailand	0.297

Methods

DNA extraction

Genomic DNA was extracted from the dried leaves of different individuals of selected Jerusalem artichoke genotypes. A sample of 100 mg dried leaf tissue was ground using a pestle and mortar in liquid nitrogen. Next, the powder was suspended in 700 µL extraction buffer comprising 100 mM Tris-HCl, pH 7.5, 0.35 M mannitol, 50 mM EDTA, pH 8.0, and 0.3% β-mercaptoethanol. The mixture was gently vortexed and then incubated at 65°C for 1 h, during which the mixture was shaken gently several times during the incubation, followed by chloroform clean up. DNA was precipitated with isopropanol, and the DNA pellet was washed in ethanol (70%), air dried, and re-suspended in 100 µL TE buffer (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA, pH 8.0). The DNA was quantified by gel electrophoresis and NanoDrop™ (Thermo Scientific), and stored at -20°C until use.

EST-SSR and PCR amplification

A total of 40,362 ESTs of Jerusalem artichoke were retrieved from the NCBI nucleotide

database. They were subsequently arranged into 6563 contigs and 17,128 single sequences using CAP3 (Huang and Madan, 1999). In order to identify the SSR motifs harboring two to six nucleotides, arranged sequences with a minimum of six, five, four, four, and four repeated units in the unigenes were detected by the SSRIT software (Temnykh et al., 2001). When the ESTs contained appropriate flanking sequences to the SSR, this EST was selected as a candidate and used to design the over-flanking amplified primer. With this purpose, 79 sequences were identified and used to design primer pairs in the Primer3 Plus software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/). Primers were designed with a GC content not exceeding 40% and a melting temperature between 55° and 60°C, before DNA synthesis and amplification testing.

Genomic DNA of eight Jerusalem artichoke accessions (named CN52867, HEL53, HEL65, HEL250, JA6, JA37, JA102, and AMES2722) was used to test the efficiency of the novel designed EST-SSR primers. PCR amplification was performed on an Agilent Technology Sure Cycler 8800 (Germany) and carried out in a 10-µL reaction mixture with 30 ng DNA, 0.2 mM dNTPs, 0.2 mM each primer (Bio Basic Inc.), 0.4 U *Taq* DNA polymerase, 1X Buffer A [160 mM $(NH_4)_2SO_4$, 500 mM Tris-HCl, pH 9.1, 17.5 mM $MgCl_2$, and 0.1% Triton x-100; Vivantis], and 1.5 mM $MgCl_2$. PCR involved one cycle of 3 min at 96°C, 37 cycles of 30 s at 93°C, 30 s at the exact annealing temperature for each locus, and 1 min at 72°C, and finally, one cycle for 5 min at 72°C. Primers were initially screened for eight individuals of Jerusalem artichoke, and the fragments obtained were visualized on 2% agarose. The successfully amplified EST-SSRs and clear fragments were validated via sequencing, and were further used to detect polymorphisms in all 60 individuals from the six populations. They were screened on 10% denaturing polyacrylamide gels and visualized by silver staining with a 100-bp DNA ladder Plus (Vivantis) as a size standard. Observation was facilitated by silver staining conducted in accordance with the modified approach of Bassam et al. (1991).

Data analysis

The amplified bands, which were each considered to be an allele, were examined using UVITEC (Topac Inc. Instrumentation, USA) to clarify the allele size. At each point within a gel, the designated EST-SSR alleles were evaluated manually in terms of their presence or absence, and coded as 1 or 0, respectively. The efficiency of the EST-SSR markers was evaluated by investigating the genetic distinctions resulting from the diversity found within Jerusalem artichoke genotypes. This assessment relied upon the examination of each genotype and use of the simple matching coefficient (S) (Sokal and Michener, 1958). The dissimilarity (D) observed at each loci was denoted as 1 - S, so that a measure of the mean dissimilarity among the genotypes could be obtained by taking an average of all the n - 1 EST-SSR dissimilarities (AD) associated with each genotype (Wangsomnuk et al., 2011b). Allelic diversity for each loci was quantified using the concept of polymorphism information content (PIC) described by Botstein et al. (1980), whose equation held that PIC = 1 - $\sum\left(P_{ij}\right)^2$ when P_{ij} is the frequency of the j^{th} allele of the i^{th} locus. GenAlEx 6.5 (Peakall and Smouse, 2006; Peakall and Smouse, 2012) was used to analyze observed and expected heterozygosity (H_O and H_E, respectively), I, effective alleles (N_E) per locus, and number of alleles (N_A), based on the polymorphic markers. The calculations were performed using the data obtained from a sample of 60 Jerusalem artichoke individuals. AMOVA was performed as described by Excoffier et al. (1992) in the software package GenAlex 6.5 (Peakall and Smouse, 2006; Peakall and Smouse, 2012).

RESULTS AND DISCUSSION

Jerusalem artichoke is used in several chemical and food industries. However, its tuber contains inulin and there is wide variation in its genotypes (Johansson et al., 2015). Evaluation of the genetic diversity of Jerusalem artichoke germplasm facilitates conservation and provides knowledge for the selection of parental clones, which is essential for cultivar improvement in order to improve yield, nutritional and commercial value to farmers and consumers, and is needed in the breeding program (Moose and Mumm, 2008).

SSR distribution

A total of 40,362 ESTs from Jerusalem artichoke were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/), and were assembled into 6563 contigs and 17,128 singletons. A total of 1949 SSR motifs two to six nucleotides in length with varying repeated units were further identified from 1676 assembled sequences (Table 2). Of these, 86.99% of sequences harbored one SSR locus, whereas two to three SSR loci accounted for 12.41% of the total identified sequences, with the remaining sequences containing four to six SSR repeat units, respectively. Investigation of SSRs revealed that trinucleotide repeats (48.08%) represented the highest proportion, whereas hexanucleotides (2.51%) represented the lowest proportion. Di-, tetra-, and pentanucleotides accounted for 34.89, 11.80, and 2.72%, respectively (Table 3). Jung et al. (2014) identified 10,778 SSRs from 8746 loci, with 18.34% sequences containing more than one SSR in the Jerusalem artichoke transcriptome. These can be searched to increase the number of informative SSR markers for the breeding of Jerusalem artichoke.

Table 2. Expressed sequence tags (ESTs) and simple sequence repeats (SSRs) identified from Jerusalem artichoke.

Parameter	Number
Total ESTs	40,362
Contigs	6,563
Singletons	17,128
Total number of sequences examined	23,691
Total number of sequences containing SSRs	1,676
Total number of SSRs discovered	1,949
Number of sequences containing one SSR	1,458
Number of sequences containing two SSRs	177
Number of sequences containing three SSRs	31
Number of sequences containing four SSRs	7
Number of sequences containing five SSRs	2
Number of sequences containing six SSRs	1
Number of primers designed	79
Number of informative primers	43

Table 3. Types and frequencies of EST-SSRs in Jerusalem artichoke.

Type	4	5	6	7	8	9	10	>10	Total
Dinucleotides	nd	nd	218	120	80	91	53	118	680
Trinucleotides	nd	420	289	105	53	24	17	29	937
Tetranucleotides	121	56	21	15	7	3	3	4	230
Pentanucleotides	41	9	1	1	0	0	1	0	53
Hexanucleotides	36	7	2	2	1	0	0	1	49

nd = not detected.

Primer development and validation

Seventy-nine ESTs containing SSRs were selected, and amplified primer pairs were generated and applied to clarify the performance of eight Jerusalem artichoke genotypes. This might increase the limited number of genetic markers for this species, where only 357 RAPD, 92 ISSR, and 194 SRAP markers have been found previously (Wangsomnuk et al., 2011a,b). Next, all designed primers were used on whole individuals to analyze polymorphism levels. Twenty-eight of 79 novel designed primers failed to amplify, or amplified only a few genotypes (35.44%). Eight loci showed monomorphic bands (10.13%) across all samples investigated. Those markers were excluded from further studies. Forty-three loci (54.43%) were informative for 60 genotypes and were further analyzed for genetic diversity. Examples of polymorphic loci are shown in Figure 1 and Table 4. The number of informative markers found in the present study was higher than that found for EST-SSR markers developed from olive EST sequences by Adawy et al. (2015). In that study, 10 of 25 randomly selected primers showed polymorphism across nine genotypes. Chen et al. (2015) reported that the number of informative EST-SSR primers validated for Adzuki bean included 38 polymorphic markers of 296 markers, which produced amplicons; this was lower than the results obtained for EST-SSR primers here.

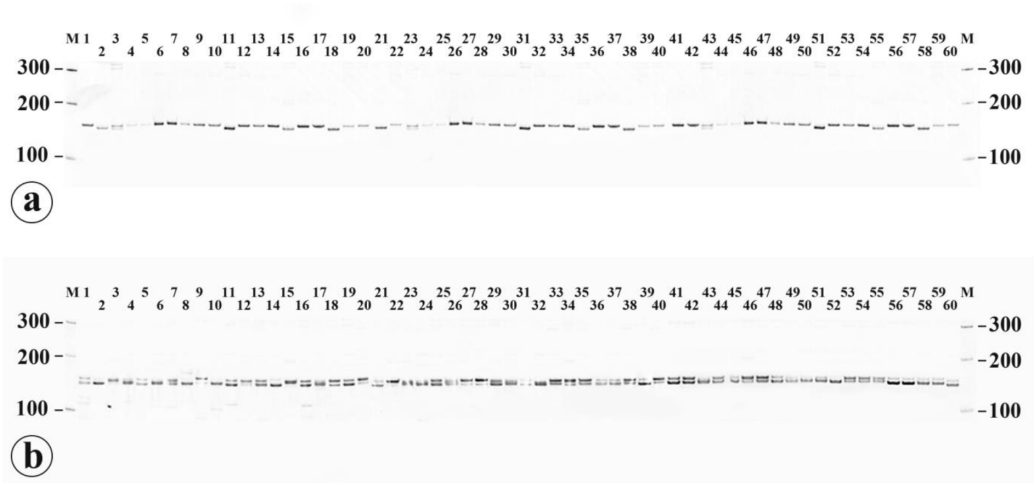

Figure 1. Polymorphism at the LP10 and LP65 loci of 60 accessions. See Table 1 detailed information of Jerusalem artichoke accessions. *Lane M* = 100-bp ladder plus (Vivantis).

Overall, 170 alleles were found among 43 loci characterized in 60 accessions. N_A, N_E per locus, H_O, and H_E of the 43 polymorphic EST-SSR loci are presented in Table 5. N_A per locus differed from two to seven with an average number of 3.95. The most frequent N_A detected per marker were three (32.56%) and four (34.88%). This is consistent with the highly conserved nature of the primer sequences flanking the SSR region, which is higher than that previously reported in some plant species such as *Phaseolus vulgaris* (Garcia et al., 2011). H_O and H_E ranged from 0.0 to 0.983 (mean 0.458) and 0.096 to 0.774 (mean 0.536), respectively. The distribution of H_E values revealed the presence of high heterozygosity within Jerusalem

artichoke populations, and showed that 67% of the markers are within the range 0.5 to 0.8 (Figure 2). This might contribute to the auto-tetraploidy and cross-pollination observed for this species (Zhou et al., 2014).

Table 4. Characteristics of 43 polymorphic SSR markers developed for genetic diversity analyses.

Primer	GenBank accession No.	Forward primer (5'-3')	Reverse primer (5'-3')	Size (bp)	Ta (°C)
LP1	EL463197	GAAATTGAAGTAGGTGTTGTA	GATTCTCGGCCTCGTCTCTG	225	55
LP2	EL452772	GATCATCCATGGCTATTGCA	AACAAAGAGTGCAGAACTCAG	113	55
LP3	EL453058	CTTTAATACTTTGCCGGATT	GTTGTGAATTAGGGTTTGTGA	150	55
LP4	EL453095	CCTGAAATTCAACTCCAACT	CTTTCTTTCACCGTTCTCTC	115	55
LP5	EL453410	GGGTGCATCCAAATATATAAC	TAGCTCGACGTCTTGTTTTT	150	55
LP6	EL453427	CAACCTCCACAAGAATCCTA	TAAACCCTGAGGGGTGTAAA	130	55
LP7	EL453432	ATCCTCTGCTGGTGTTGTAG	GTTTCACTAGCAGCATGTCC	165	55
LP8	EL453460	ATCCATTTTGTTTGGAATTG	TCAACAGATGTCGTTTCTGA	168	55
LP10	EL453502	TTCTCATCATCGTCTCAACA	CCTTCTTCATCGTCTACCAA	153	55
LP12	EL453840	GGGGTATTCCCTACTTAACG	GAAAATTGACATGCTTACGG	169	55
LP15	EL454207	CCGAACTGGTATATTCGGTA	TGGATTGGATTGGAGTTG	153	55
LP16	EL454269	ACCAAAAGTCTCAAACAAAGT	ATTTGTTCTTCCTGTAAATTGG	150	55
LP18	EL454165	CCACAAGAATCATCATCAAC	GGATCATCTCTGATCAAAATCT	157	55
LP20	EL435002	TCCTGCAACTTCTCTCTCTC	GATAAGAGTGCTTTGTGTTCC	150	55
LP24	EL435734	CAACTGCTGATTCAGATGG	TACAGATCGCGATTGATAGG	155	55
LP25	EL437978	CGGACCCTTTTGAATCCTTC	ATCCTTAAACGCTTCACGAG	197	55
LP27	EL443017	ATAAACAGCCACGACGGCTA	CTGATGATGATAGGGGCAAA	103	55
LP29	EL443443	CATATTGGCTACTTCCATCTC	GAATGTTGTTGAATGGAAAGAG	150	55
LP30	EL453557	GCAAGTGGTGGCAAGAAGAT	AAAGGAAGAAACCCTCTTCG	121	55
LP34	EL447983	TCGATCGTCGTTATCCTAGA	CTCGGACTTCCTTCTTCACT	167	55
LP35	EL447932	AGCCTCATATTTCATGATCG	AGCTTAGTGAGGGCAATGA	213	55
LP37	EL448436	GTCCCCATTTCTAAGCTACC	GTCATATCCCCGACGTCTAC	224	55
LP39	EL456599	GTGTTGTACCCGTTTTGTTT	GTTTTCCCTCTTCTCTCACAA	160	55
LP41	EL455655	GCAGTTTCACCAACAACTTT	CTTGAGGAGAAACCCTCCTA	182	55
LP42	EL455225	CTTCAAGCCCAGTAAAAACA	ACTCCCTACAAGTGGTGGTT	162	55
LP45	EL463403	TAGGGTTCACATCGCATAAG	TGCGTGGTACACAACTAACA	189	55
LP46	EL466180	GTCTCGCATTATTGGAGTTG	CACAAGATCGGTGAATCAAT	211	55
LP48	EL465498	CAAGGGGGACAATTTTAAGT	AACACAATCCAGGCGATAA	157	55
LP50	EL471983	TCGGAACCATAAATTCTGC	ACTCTCGCAATCAGAAACTG	206	55
LP51	EL440203	CTACCACCAGACCTTCTCAA	CCCAATTCTCCAGTTTCTTT	246	55
LP54	EL466909	GTTCTTAACAATCCCGCTTT	CTGACAAAATTTGGGAACAG	218	55
LP55	EL440203	CTACCACCAGACCTTCTCAA	CCCAATTCTCCAGTTTCTTT	246	55
LP59	EL435465	CAAGTCAACCTGATCAGAAA	AACCATTCCAGGGCACTAT	153	55
LP61	EL454381	TCAGTCCAAGATATCGATGTT	CTTCATCATCATCACTATCAG	128	58
LP63	EL455401	CTTGCAGCATACGAAGACGAA	TCCAAAACCTGGCGATTAAC	198	58
LP65	EL462012	GGACTCACGTTATGGGTACA	GCCCCACGTATTTAATTTCT	152	55
LP66	EL459585	GTTGAACCACTTCTAGTTTG	CAATCTTTAACATACCCATGT	156	58
LP68	EL460172	TTGAATTCCACACGATTAGGG	CAACCTTAGGCTGTGAAAAATTG	105	58
LP70	EL463104	GAATGGCCGGATAAACTCAA	GCGAGCATAATGTGCAAAAA	204	60
LP72	EL452769	TCTGACCATTCAATCTCCTC	GGGTACACTGTAACTGTAAAGAA	114	55
LP73	EL464624	GCTCAACTAAACGGCTTGCA	AAAACAGGCAGTAATCACCG	170	58
LP75	EL464737	CGTTAAATCGATGGGGAGAA	TTTGCCACCTTTCCACTACC	180	58
LP78	EL464751	CAAACCACATCCCCACACTT	TGTTGAACCACCGTCAGGAC	166	55

Ta = annealing temperature.

The number of effective alleles (N_E) per polymorphic locus varied from 1.106 to 4.425 with an average of 2.505. The locus LP5 possessed the highest effective number of alleles (4.425) and the highest expected heterozygosity (0.774), and harbored repeat motifs of $(ACAT)_5$. Locus LP8 possesses the lowest number of effective alleles (1.106) and the lowest expected heterozygosity (0.096) with repeat motifs of $(TCA)_5$ (Table 5).

Informativeness of markers was measured by the PIC. Markers with many alleles or those that are highly polymorphic tend to be highly informative. The degree of polymorphism can be classified into three levels: high (PIC > 0.5), medium (0.5 > PIC > 0.25), and low (PIC

< 0.25) (Hildebrand et al., 1992). PIC analysis revealed that 43 loci have values ranging from 0.096 (LP8) to 0.854 (LP27) (Table 5 and Figure 3), with an average value of 0.568. The largest group of loci (27.91%) ranged from 0.611 to 0.689, followed by the group with PIC values ranging from 0.714 to 0.788 (20.93%). Nearly three-quarters of all loci possess PIC values higher than 0.5, meaning that the majority of loci studied here possess high levels of polymorphism. Only one-tenth of the loci possessed low polymorphism (PIC < 0.25) (Figure 3). The average PIC values reported here are higher than the allelic variation at 32 loci detected in cowpea (Gupta and Gopalakrishna, 2010).

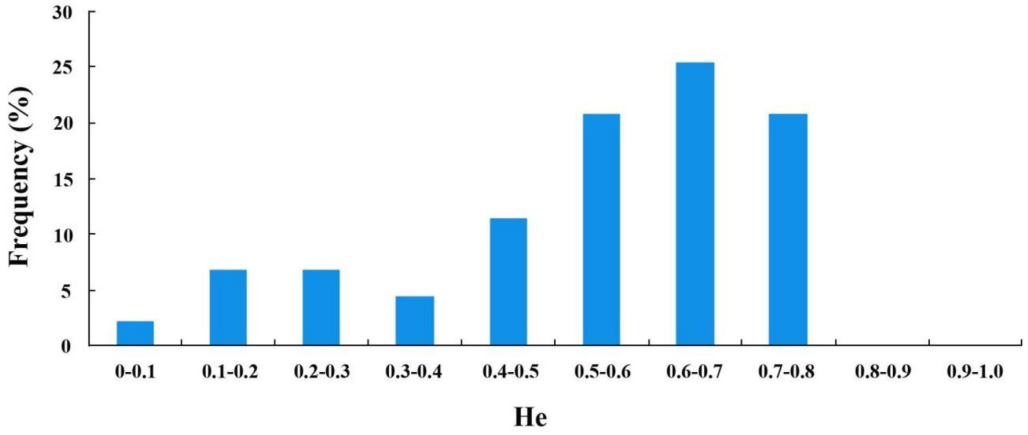

Figure 2. Distribution of estimates of genetic heterozygosity.

Approximately 82% of genetic variation was detected within individuals of accessions from a given country, with a much smaller amount of variation occurring among individuals (13%) or populations (5%) (Table 6). All the components of differentiation determined by AMOVA were statistically significant at P < 0.001.

Pairwise differentiation (F_{ST}) was calculated for all accessions. According to a previous study, F_{ST} of 0.00-0.05 indicates low differentiation, 0.05-0.15 indicates moderate differentiation, while F_{ST} > 0.15 indicates high levels of differentiation (Hartl and Clark, 1997). Variation in F_{ST} in the present study ranged from 0 to 0.096, which implies low-to-moderate genotypic differentiation across loci between six countries. There was no diversity of genetic subdivision of populations from Canada, Germany, and Russia (Table 7). A large pairwise F_{ST} value was observed between the populations from the USA and Russia (F_{ST} = 0.096), followed by accessions from the USA and Canada (F_{ST} = 0.093). These data indicate that accessions from the USA are more differentiated from those of Russia and Canada. This finding was also supported by the unweighted pair-group method based on arithmetic average (UPGMA) of Nei's unbiased genetic distance analysis among accessions from different countries (resources) (Figure 4).

All pairwise group F_{ST} values were statistically significant at P < 0.05, except for those that were non-significant and are highlighted in bold and italic.

Table 5. Informativeness of SSR loci according to the amplification from 60 Jerusalem artichoke accessions.

Locus	N_A	N_E	I	H_O	H_E	PIC
LP1	4	2.811	1.157	0.267	0.644	0.656
LP2	3	1.126	0.262	0.083	0.112	0.171
LP3	5	4.011	1.467	0.95	0.751	0.749
LP4	3	1.49	0.576	0.367	0.329	0.416
LP5	6	4.425	1.577	0.75	0.774	0.761
LP6	4	2.19	0.98	0.7	0.543	0.593
LP7	3	1.854	0.709	0.683	0.461	0.498
LP8	3	1.106	0.23	0	0.096	0.096
LP10	2	1.444	0.486	0.069	0.307	0.346
LP12	6	2.449	1.196	0.579	0.592	0.635
LP15	6	4.278	1.583	0.517	0.766	0.782
LP16	4	1.795	0.808	0.2	0.443	0.478
LP18	4	2.803	1.13	0.983	0.643	0.638
LP20	3	2.158	0.888	0.267	0.537	0.576
LP24	3	1.206	0.37	0.117	0.171	0.241
LP25	3	2.1	0.845	0.169	0.524	0.509
LP27	5	3.967	1.48	0.31	0.748	0.854
LP29	3	2.2	0.921	0.283	0.545	0.595
LP30	4	2.171	1.018	0.183	0.539	0.576
LP34	2	1.301	0.393	0.267	0.231	0.332
LP35	6	3.076	1.397	0.683	0.675	0.730
LP37	3	1.224	0.37	0.033	0.183	0.206
LP39	3	1.826	0.754	0.617	0.452	0.511
LP41	5	3.512	1.371	0.9	0.715	0.714
LP42	3	1.462	0.603	0.167	0.316	0.465
LP45	6	3.172	1.308	0.783	0.685	0.689
LP46	4	3.121	1.231	0.967	0.68	0.680
LP48	4	2.071	0.913	0.466	0.517	0.611
LP50	7	3.391	1.508	0.542	0.705	0.744
LP51	3	2.182	0.888	0.667	0.542	0.535
LP54	2	1.835	0.647	0	0.455	0.455
LP55	3	2.026	0.734	0.95	0.507	0.508
LP59	4	1.839	0.837	0.22	0.456	0.489
LP61	4	3.11	1.251	0.3	0.678	0.684
LP63	4	2.631	1.146	0.433	0.62	0.670
LP65	4	3.048	1.24	0.883	0.672	0.667
LP66	5	4.243	1.511	0.567	0.764	0.788
LP68	4	3.524	1.314	0.417	0.716	0.724
LP70	4	2.777	1.182	0.45	0.64	0.658
LP72	5	3.206	1.299	0.917	0.688	0.673
LP73	3	2.751	1.049	0.617	0.637	0.643
LP75	4	3.468	1.317	0.283	0.712	0.738
LP78	4	1.342	0.544	0.1	0.255	0.341

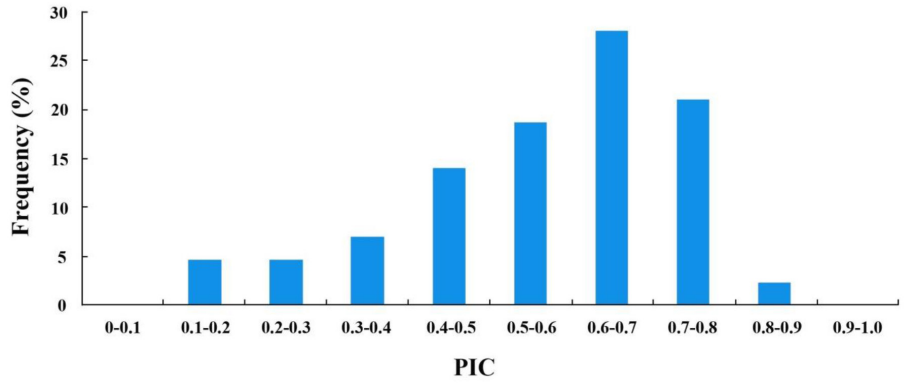

Figure 3. Distribution of polymorphic information content (PIC) values for 170 simple sequence repeat (SSR) markers.

Table 6. Analyses of molecular variance (AMOVA) of Jerusalem artichoke by simple sequence repeat (SSR) loci.

Source of variation	d.f.	Sum of squares	Variance component	Percentage of variance (%)	Pvalue
Among Pops	5	105.679	0.543	5	<0.001
Among Individual	54	701.771	1.59	13	<0.001
Within Individual	60	589	9.817	82	
Total	119	1396.45	11.949	100	

Table 7. Proportional SSR variation among Jerusalem artichoke accessions of different origin/sources estimated from the analysis of molecular variance of 43 SSR loci.

Origin/source	Pairwise F_{ST}				
	Canada	USA	Russia	Germany	France
Canada					
USA	0.093				
Russia	***0.000***	0.096			
Germany	***0.000***	0.086	***0.000***		
France	0.089	0.087	***0.050***	0.073	
Thailand	***0.010***	0.085	***0.018***	***0.018***	0.074

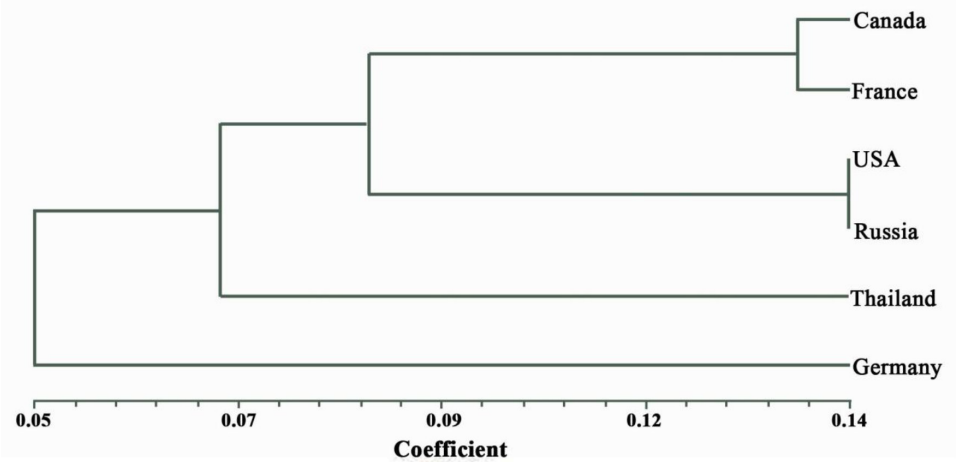

Figure 4. Unweighted pair-group method based on arithmetic average (UPGMA) dendrogram showing genetic relationships among Jerusalem artichoke origins/sources based on Nei's unbiased genetic distances.

Analysis of genetic diversity

Genetic diversity parameters for the 43 microsatellite loci of the 60 Jerusalem artichoke accessions were calculated. Polymorphism among genotypes within each country of origin was as follows: Canada (58.24%), the USA (61.77%), Russia (56.47%), Germany (51.18%), France (55.29%), and Thailand (91.18%). The highest number of polymorphic bands was observed for accessions from Thailand, which may contribute to the large number of accessions (35) compared to other origins (5). It is important to note that increasing the number of samples from other countries or analyzing the same set of samples using more informative primers developed from other available ESTs of Jerusalem artichoke (Jung et al., 2014) may change the genetic diversity information of each population.

A total of 3739 alleles were detected from populations of different sources with an average of 62.32 allele per genotype. The minimum number of alleles was 58, which was observed in four accessions from Thailand, namely, KK101, KK166, KK243, and KK283 (Figure 5). The maximum number of alleles was presented in AMES2736 from the USA. Accessions from Russia and France possess between 60 and 64 alleles, with mean values of 62.20 and 61.80, respectively. Within the accessions from Canada, the number of alleles was between 60 and 63 with an average number of 62. In five accessions from the USA, the number of alleles ranged from 59 to 69 with a mean number of 63. Within the accessions from Germany, the number of alleles ranged from 60 to 67, with an average of 62. Within accessions from Thailand, the number of alleles ranged from 58 to 68, with an average number of 61.80 alleles per accession.

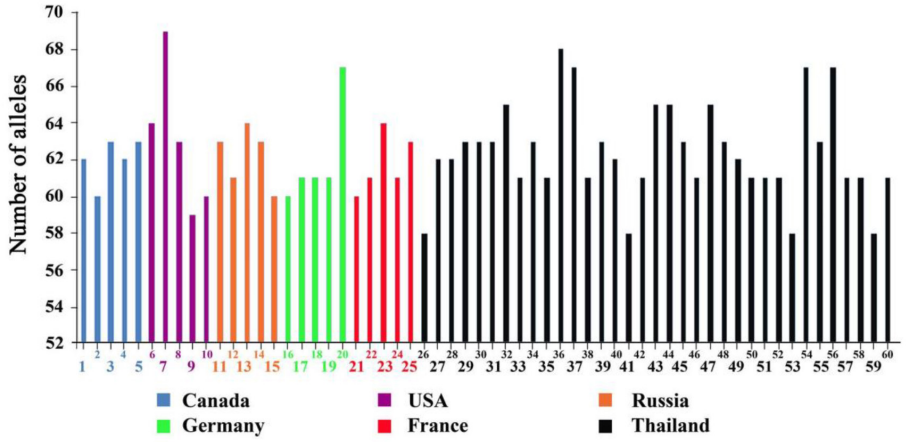

Figure 5. Number of alleles detected in 60 Jerusalem artichoke accessions based on 43 SSR loci.

Genetic differentiation and cluster analysis

AD of Jerusalem artichoke accessions ranged from 0.257 (KK277) to 0.345 (PI503260) (Table 1) with a mean AD of 0.301. The 10 most distinct accessions with an AD of 0.319 or higher included PI503260, KK250, PI547241, AMES2722, KK203, KK148, JA78, KK126, JA55, and KK279. Of note, five open-pollinated lines produced in Thailand are among these 10 accessions, in addition to four wild accessions from the USA and one accession from France.

It is worth noting that the largest genetic distance calculated using the simple matching coefficient (0.45) was observed between KK133 (Thailand) and AMES2722 (USA), which can be used as potential parental sources for further breeding programs. The lowest genetic distance (0.11) was found between KK121 and KK112, and also between KK176 and KK212 which are breeding lines from Thailand, suggesting that EST-SSR markers could be used successfully to distinguish between closely related genotypes. The average genetic distance of accessions from Canada, the USA, Russia, Germany, France, and Thailand was 0.28, 0.30, 0.28, 0.25, 0.28, and 0.24, respectively. These results suggest that accessions from the USA possess higher levels of genetic diversity and might serve as a valuable resource. Overall, 54.58% of the genetic distance between any two accessions of six origins was at least 0.30 (Figure 6).

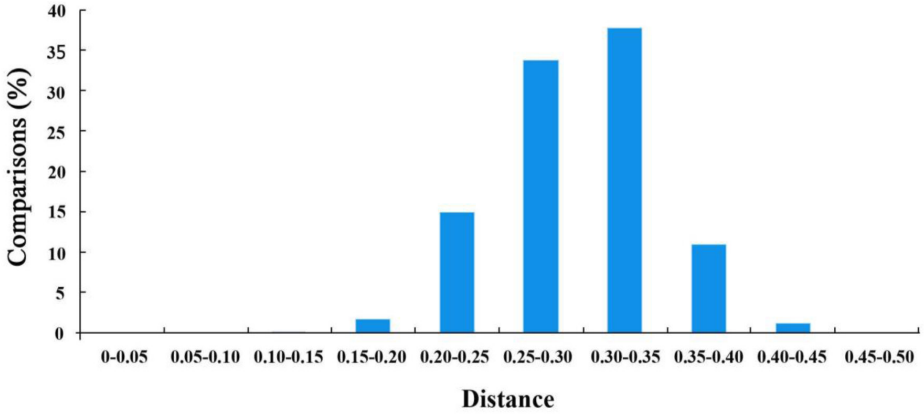

Figure 6. Distribution of pairwise genetic distances based on 170 SSR markers of 60 accessions.

The genetic relationship among 60 genotypes of Jerusalem artichoke is presented based on the neighbor-joining (NJ) analysis (Figure 7). Most of the accessions from six countries are dispersed among several clusters owing to the low resolution of SSR loci. Six clusters were detected. The first cluster contained 16 accessions, including seven (KK137, KK148, KK166, KK205, KK277, KK279, and KK299) from Thailand, one from Germany (JA102), two from Russia (JA59, JA95), one from Canada (JA134), four from France (JA89, JA97, JA98, HEL250), and one from the USA (JA55). The second cluster comprised seven accessions, including one from Canada (JA42), and the rest from Thailand (KK133, KK139, KK182, KK191, and KK264). The third cluster contained two accessions, including JA105 from Russia and KK191 from Thailand. The forth cluster, which was the biggest group, contained 20 accessions, most of which were from Thailand (15 accessions), with two accessions from Germany (HEL53, HEL231), two from Canada (JA6, JA37), and one from Russia (CN52867). The fifth cluster comprised six accessions, including five accessions from Thailand (KK212, KK224, KK250, KK261, KK283) and one from Russia (HEL65). The last cluster contained nine accessions, including four from the USA (PI547241, AMES2722, PI503260, AMES2736), two from Germany (HEL243, HEL 248), and one accession each from France (JA78), Canada (JA4), and Thailand (KK157). Jerusalem artichoke is a highly self-incompatible plant, which favors cross-pollination as it generally produces wider variation than vegetative propagation. Without control of pollination, varieties can be developed for characters of interest such as high tuber yield and disease resistance. Thus, it can be inferred that the genetic background of these Jerusalem artichoke accessions does not always correlate with their geographical regions.

A PCoA was performed based on the genetic distance of the 60 accessions. The first three axes accounted for 29.78% (12.54, 9.17, and 8.07% of the distribution, respectively). The distribution of the relative contribution of each variable in the total variance of the first two axes is well represented by the projection of vectors indicating the maximum variation in the 1st and 2nd axes. The PCoA result revealed somewhat different clusters of accessions compared to those obtained by NJ cluster analysis. However, moderate agreement was detected between these two approaches.

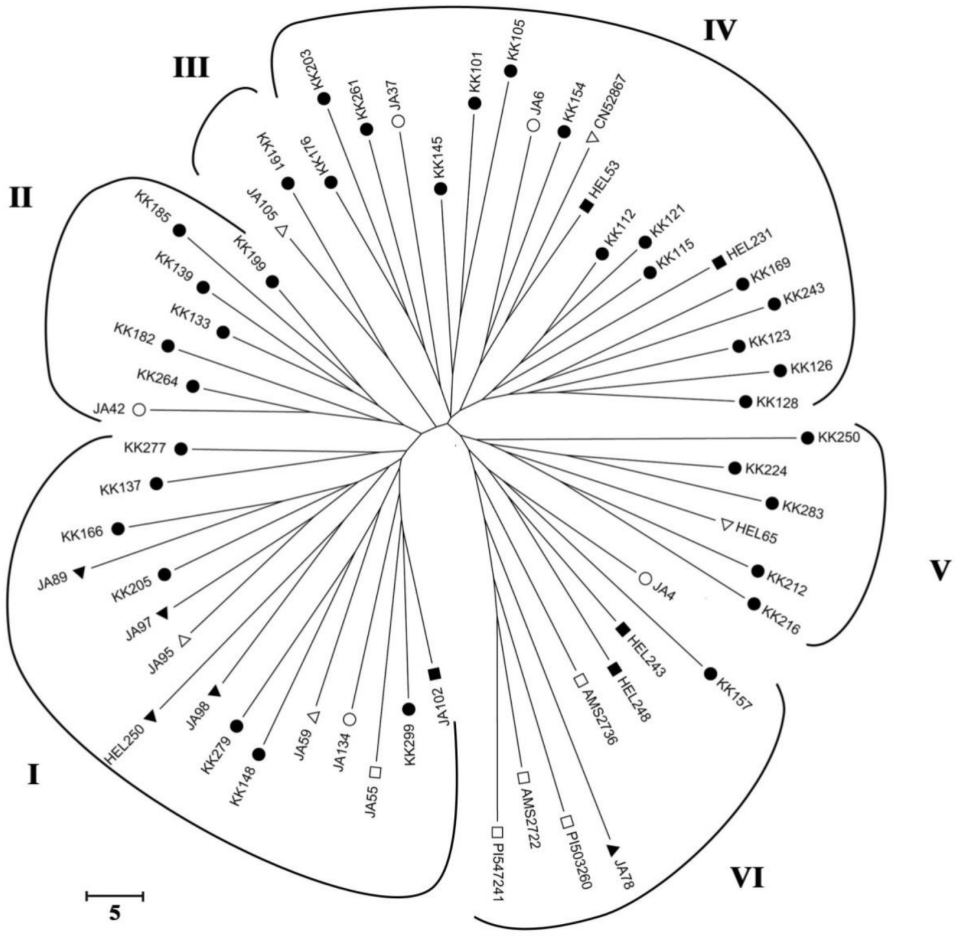

Figure 7. Neighbor-joining tree showing the genetic association of 60 Jerusalem artichoke genotypes labeled with their origin/source: open square for USA; filled square for Germany; open triangle for Russia; filled triangle for France; open circle for Canada; filled circle for Thailand.

In the present study, the genetic diversity of 60 Jerusalem artichoke was evaluated based on 43 EST-SSR loci. These markers were highly robust with high PIC values (mean 0.568), and polymorphism among accessions within each country ranging from 50.588% (Germany) to 91.764% (Thailand). These newly developed EST-SSR loci have the potential to be applied to studies on molecular breeding and genetic diversity in this species, which might help to cross species and determine genetic variation within the genus *Helianthus*.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The Higher Education Research Promotion, the National Research University Project of Thailand, the Office of the Higher Education Commission through the Food and Functional Food Research Cluster of Khon Kaen University (F-2553-Ph.d-02 and FC1.1.5 PhD), and research funding from Khon Kaen University to the corresponding author are acknowledged for financially support. We thank Assistant Professor Dr. Tawan Remsungnen for his assistance on the preliminary analysis of SSRs, and the journal reviewers for their valuable comments on this manuscript.

REFERENCES

Adawy SS, Mokhtar MM, Alsamman MA and Sakr MM (2015). Development of EST-SSR annotated database in olive (*Oleaeuropaea*). *IJSR*09.

Alla NA, Domokos-Szabolcsy É, El-Ramady H, Hodossi S, et al. (2014). Jerusalem artichoke (*Helianthus tuberosus* L.): A review of in vivo and in vitro propagation. *Int. J. Hortic. Sci.* 20: 131-136.

Andersen JR and Lübberstedt T (2003). Functional markers in plants. *Trends Plant Sci.* 8: 554-560. http://dx.doi.org/10.1016/j.tplants.2003.09.010

Bassam BJ, Caetano-Anollés G and Gresshoff PM (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* 196: 80-83. http://dx.doi.org/10.1016/0003-2697(91)90120-I

Bock DG, Kane NC, Ebert DP and Rieseberg LH (2014). Genome skimming reveals the origin of the *Jerusalem Artichoke* tuber crop species: neither from Jerusalem nor an artichoke. *New Phytol.* 201: 1021-1030. http://dx.doi.org/10.1111/nph.12560

Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.

Chen H, Liu L, Wang L, Wang S, et al. (2015). Development and validation of EST-SSR markers from the transcriptome of Adzuki bean (*Vigna angularis*). *PLoS One* 10: e0131939. http://dx.doi.org/10.1371/journal.pone.0131939

Debnath SC (2014). Structured diversity using EST-PCR and EST-SSR markers in a set of wild blueberry clones and cultivars. *Biochem. Syst. Ecol.* 54: 337-347. http://dx.doi.org/10.1016/j.bse.2014.03.018

Excoffier L, Smouse PE and Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Gadaleta A, Giancaspro A, Zacheo S, Nigro D, et al. (2011). Comparison of genomic and EST-derived SSR markers in phylogenetic analysis of wheat. *Plant Genet. Resour.* 9: 243-246. http://dx.doi.org/10.1017/S147926211100030X

Garcia RAV, Rangel PN, Brondani C, Martins WS, et al. (2011). The characterization of a new set of EST-derived simple sequence repeat (SSR) markers as a resource for the genetic analysis of *Phaseolus vulgaris*. *BMC Genet.* 12: 41-54. http://dx.doi.org/10.1186/1471-2156-12-41

Gupta SK and Gopalakrishna T (2010). Development of unigene-derived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other *Vigna* species. *Genome* 53: 508-523. http://dx.doi.org/10.1139/G10-028

Hartl DL and Clark AG (1997). Principle of Population Genetics. 1997. Sinauer Associates, Inc.

Hildebrand CE, Torney DC and Wagner RP (1992). Informativeness of polymorphic DNA markers. *Los Alamos Sci.* 20: 100-102.

Huang X and Madan A (1999). CAP3: A DNA sequence assembly program. *Genome Res.* 9: 868-877. http://dx.doi.org/10.1101/gr.9.9.868

Johansson E, Prade T, Angelidaki I, Svensson SE, et al. (2015). Economically viable components from Jerusalem artichoke (*Helianthus tuberosus* L.) in a biorefinery concept. *Int. J. Mol. Sci.* 16: 8997-9016. http://dx.doi.org/10.3390/ijms16048997

Ju MM, Ma HC, Xin PY, Zhou ZL, et al. (2015). Development and characterization of EST-SSR markers in *Bombax ceiba* (Malvaceae). *Appl. Plant Sci.* 3: 1500001. http://dx.doi.org/10.3732/apps.1500001

Jung WY, Lee SS, Kim CW, Kim H-S, et al. (2014). RNA-seq analysis and *de novo* transcriptome assembly of Jerusalem artichoke (*Helianthus tuberosus* Linne). *PLoS One* 9: e111982. http://dx.doi.org/10.1371/journal.pone.0111982

Kays SJ and Nottingham SF (2008). Genetic resources, breeding and cultivars. In: Biology and Biochemistry of Jerusalem Artichoke (Taylor and Francis eds.). CRC Press, 149-240.

Kiru S and Nasenko I (2010). Use of genetic resources from Jerusalem artichoke collection of N. Vavilov Institute in

breeding for bioenergy and health. *Agron. Res.* 8: 625-632.
Kou YX, Zeng J, Liu JQ, Kou YX, et al. (2014). Germplasm diversity and differentiation of *Helianthus tuberosus* L. revealed by AFLP marker and phenotypic traits. *J. Agric. Sci.* 152: 779-789. http://dx.doi.org/10.1017/S0021859613000476
Kumari M, Grover A, Yadav PV, Arif M, et al. (2013). Development of EST-SSR markers through data mining and their use for genetic diversity study in Indian accessions of *Jatropha curcas* L.: a potential energy crop. *Genes Genomics* 35: 661-670. http://dx.doi.org/10.1007/s13258-013-0118-0
Malfa SL, Currò S, Douglas AB, Brugaletta M, et al. (2014). Genetic diversity revealed by EST-SSR markers in carob tree (*Ceratonia siliqua* L.). *Biochem. Syst. Ecol.* 55: 205-211. http://dx.doi.org/10.1016/j.bse.2014.03.022
Merritt BJ, Culley TM, Avanesyan A, Stokes R, et al. (2015). An empirical review: Characteristics of plant microsatellite markers that confer higher levels of genetic variation. *Appl. Plant Sci.* 3: 1500025. http://dx.doi.org/10.3732/apps.1500025
Mondini L, Noorani A and Pagnotta MA (2009). Assessing plant genetic diversity by molecular tools. *Diversity (Basel)* 1: 19-35. http://dx.doi.org/10.3390/d1010019
Moose SP and Mumm RH (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol.* 147: 969-977. http://dx.doi.org/10.1104/pp.108.118232
Mujaju C, Sehic J and Nybom H (2013). Assessment of EST-SSR markers for evaluating genetic diversity in watermelon accessions from Zimbabwe. *Am. J. Plant Sci.* 4: 1448-1456. http://dx.doi.org/10.4236/ajps.2013.47177
Mullis K, Faloona F, Scharf S, Saiki R, et al. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51: 263-273. http://dx.doi.org/10.1101/SQB.1986.051.01.032
Park YJ, Lee JK and Kim NS (2009). Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops. *Molecules* 14: 4546-4569. http://dx.doi.org/10.3390/molecules14114546
Peakall R and Smouse P (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295. http://dx.doi.org/10.1111/j.1471-8286.2005.01155.x
Peakall R and Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* 28: 2537-2539. http://dx.doi.org/10.1093/bioinformatics/bts460
Poczai P, Varga I, Laos M, Cseh A, et al. (2013). Advances in plant gene-targeted and functional markers: a review. *Plant Methods* 9: 6. http://dx.doi.org/10.1186/1746-4811-9-6
Ramu P, Billot C, Rami JF, Senthilvel S, et al. (2013). Assessment of genetic diversity in the sorghum reference set using EST-SSR markers. *Theor. Appl. Genet.* 126: 2051-2064. http://dx.doi.org/10.1007/s00122-013-2117-6
Şelale H, Çelik I, Gültekin V, Allmer J, et al. (2013). Development of EST-SSR markers for diversity and breeding studies in opium poppy. *Plant Breed.* 132: 344-351. http://dx.doi.org/10.1111/pbr.12059
Sokal RR and Michener CD (1958). A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* 38: 1409-1438.
Temnykh S, DeClerck G, Lukashova A, Lipovich L, et al. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 11: 1441-1452. http://dx.doi.org/10.1101/gr.184001
Wangsomnuk PP, Khampa S, Jogloy S, Srivong T, et al. (2011a). Assessing genetic structure and relatedness of Jerusalem Artichoke (*Helianthus tuberosus* L.) germplasm with RAPD, ISSR and SRAP Markers. *AJPS* 2: 753-764. http://dx.doi.org/10.4236/ajps.2011.26090
Wangsomnuk PP, Khampa S, Wangsomnuk P, Jogloy S, et al. (2011b). Genetic diversity of worldwide Jerusalem artichoke (*Helianthus tuberosus*) germplasm as revealed by RAPD markers. *Genet. Mol. Res.* 10: 4012-4025. http://dx.doi.org/10.4238/2011.December.12.4
Wangsomnuk PP, Khampa S and Jogloy S (2015). Exogenous supplementation of growth regulators and temperature improves germination of dormant Jerusalem Artichoke (*Helianthus tuberosus* L.) seeds under in vitro and in vivo conditions. *JABS* 9: 23-30.
Zhang M, Mao W, Zhang G and Wu F (2014). Development and characterization of polymorphic EST-SSR and genomic SSR markers for Tibetan annual wild barley. *PLoS One* 9: e94881. http://dx.doi.org/10.1371/journal.pone.0094881
Zhou Q, Chen TL, Wang YR and Liu ZP (2014). The development of 204 novel EST-SSRs and their use for genetic diversity analyses in cultivated alfalfa. *Biochem. Syst. Ecol.* 57: 227-230. http://dx.doi.org/10.1016/j.bse.2014.08.023