Electronic Journal of Biotechnology 44 (2020) 25–32

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

Research article

Genome-wide development of polymorphic microsatellite markers and their application in peanut breeding program

Junjie Ma [a,b], Yuhan Zhao [a,c], Hua Chen [d], Chun Fu [e], Lin Zhu [a,c], Ximeng Zhou [a,c], Han Xia [a,c], Lei Hou [a], Guanghui Li [a], Weijian Zhuang [d], Xingjun Wang [a,b,c,*], Chuanzhi Zhao [a,c,*]

[a] Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan 250100, PR China
[b] College of Life Sciences, Shandong University, Qingdao 266237, PR China
[c] College of Life Sciences, Shandong Normal University, Jinan 250014, PR China
[d] Fujian Provincial Key Laboratory of Plant Molecular and Cell Biology, Oil Crops Research Institute, State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, Fujian Agriculture and Forestry University, Fuzhou, PR China
[e] Weifang Academy of Agricultural Sciences, Weifang 261071, China

ARTICLE INFO

Article history:
Received 9 September 2019
Accepted 3 January 2020
Available online 14 January 2020

Keywords:
Arachis hypogaea. L
Crops
Genome sequence
Microsatellites
Molecular markers
Peanut breeding program
Peanut
Polymorphic microsatellite
Polymorphism
SSR

ABSTRACT

Background: Cultivated peanut (*Arachis hypogaea.* L) represents one of the most important oil crops in the world. Although much effort has been expended to characterize microsatellites or Simple Sequence Repeats (SSRs) in peanut, the quantity and quality of the markers in breeding applications remain limited. Here, genome-wide SSR characterization and marker development were performed using the recently assembled genome of the cultivar Tifrunner.

Results: In total, 512,900 microsatellites were identified from 2556.9-Mb genomic sequences. Based on the flanking sequences of the identified microsatellites, 7757 primer pairs (markers) were designed, and further evaluated in the assembled genomic sequences of the tetraploid *Arachis* cultivars, Tifrunner and Shitouqi, and the diploid ancestral species, *A. duranensis* and *A. ipaensis*. *In silico* PCR analysis showed that the SSR markers had high amplification efficiency and polymorphism in four *Arachis* genotypes. Notably, nearly 60% of these markers were single-locus SSRs in tetraploid *Arachis* species, indicating they are more specific in distinguishing the alleles of the A and B sub-genomes of peanut. In addition, two markers closely related with purple testa color and 27 markers near to *FAD2* genes were identified, which could be used for breeding varieties with purple testa and high-oleic acid content, respectively. Moreover, the potential application of these SSR markers in tracking introgressions from *Arachis* wild relatives was discussed.

Conclusions: This study reported the development of genomic SSRs from assembled genomic sequences of the tetraploid *Arachis* Tifrunner, which will be useful for diversity analysis, genetic mapping and functional genomics studies in peanut.

How to cite: Ma J, Zhao Y, Chen H, et al. Genome-wide development of polymorphic microsatellite markers and their application in peanut breeding program. Electron J Biotechnol 2020;44. https://doi.org/10.1016/j.ejbt.2020.01.004.

1. Introduction

Microsatellites (or simple sequence repeats, SSRs), are a group of DNA sequences consisting of tandemly repeated motifs [1,2]. SSRs are widely distributed in the genome of animals and plants, with large number, rich polymorphism, codominant inheritance and relatively conservative flanking sequences. The polymorphism of SSRs derives from the repeat times of motifs, which can be easily detected by PCR amplification [3]. Thus, SSR markers are considered as a class of cost-effective, reliable, and highly polymorphic molecular markers, and widely used in genetics and breeding researches [4]. Even today, although the third-generation molecular markers are developing rapidly such as single nucleotide polymorphism (SNP) and insertion–deletion (InDel), SSR markers are still indispensable in crop genetic analysis, quantitative trait loci (QTL) mapping, and molecular breeding.

SSR markers can be developed from sequences of expressed sequence tags (EST), transcriptome and genome, corresponding to EST-SSRs, transcriptome-SSRs and g-SSRs, respectively [5]. In comparison to EST-SSRs and transcriptome-SSRs, g-SSRs showed the

* Corresponding authors.
E-mail addresses: xingjunw@hotmail.com (X. Wang), chuanzhiz@126.com (C. Zhao).
Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

https://doi.org/10.1016/j.ejbt.2020.01.004

higher polymorphism rate [6,7]. Furthermore, most of the g-SSRs come from a certain position in the genome and with the known physical position in the chromosome, which can provide more accurate information in genetic research especially in gene/QTL mapping. Recently, the completion of genome sequencing of more and more plants has provided opportunities to explore genome-wide SSR markers from the whole genome, resulting in the rapid increasing of g-SSR markers in plants. For example, the assembled genomic sequences of *Brassica rapa*, *B. oleracea* and *B. napus* were used for SSRs scanning, and resulted in the identification of 140,998, 229,389 and 420,991 perfect mono- to hexanucleotide repeat microsatellites, respectively [8]. In cotton, a total of 100,290, 83,160 and 56,937 microsatellites were identified using three sequenced *Gossypium* species: *G. hirsutum*, *G. arboreum*, and *G. raimondii*, respectively [9]. In wheat, a total of 364,347 SSRs were reported using the genome sequences of Chinese spring [7]. Moreover, genome-wide SSR characterization or marker development has been also conducted in other plants, such as *Boswellia papyrifera* [10], Chinese cabbage [11], bamboo [12], chickpea [13], *Sargassum thunbergii* [14], Jute [15] and Flax [16]. These reported SSRs have become an important resource in the genetic research for these plant species.

Peanut, which contains high-quality cooking oil (40–60%), protein (20–40%), and many other nutritional ingredients, is one of the most important oil crops in the world. Cultivated peanut is allotetraploid (AABB, 2n = 4× = 40), while most wild peanut species are diploid including *A. duranensis* (AA, 2n = 2× = 20) and *A. ipaensis* (BB, 2n = 2× = 20), the ancestral donor species of A and B sub-genomes, respectively [17]. The development of SSRs from peanut was first reported 20 years ago, when the probes labeled with ^{32}P were used to screen the SSR motifs of $(GT)_{10}$ and $(CT)_{10}$ from the peanut genome DNA library [18]. Soon afterwards, SSRs with other motifs were developed, such as (GA/CT)n, (ATT)n and (GA)n [19,20,21]. In addition, ESTs from the cDNA libraries sequencing have been the important sources for SSR development [22,23,24]. Recently, the completed whole genome sequence of the diploid ancestral species of peanut [25] provided an opportunity for peanut identification. In our previous studies, we developed 51,354 and 60,893 SSR markers through scanning the genome sequences of *A. duranensis* and *A. ipaensis*, respectively [5]. These SSRs were used in QTL mapping and identification of true F_1 individuals [26,27]. However, because these SSRs were from wild species, the amplification efficiency and polymorphism are limited in cultivated peanut species. In this year, the whole genome sequences of the cultivated peanuts have been completed [28,29]. The assembled genome sequences of cultivated peanuts provided new opportunity for developing more high-quality g-SSRs.

In this study, we conducted whole-genome microsatellite characterization and marker development using the recently assembled genome sequences of Tifrunner. The number, frequency, distribution, and types of repeat motifs in Tifrunner were analyzed. Furthermore, we evaluated the amplification efficiency and polymorphism of SSR markers in four sequenced peanut genotypes, including Tifrunner, *A. duranensis*, *A. ipaensis* and Shitouqi. Moreover, the potential application of SSR markers in QTL/gene mapping, breeding high-oleic peanuts using the MAS method and utility of *Arachis* wild relatives was analyzed.

2. Materials and methods

2.1. Plant material and DNA extraction

Cultivated peanut variety Shitouqi was provided by Dr. Weijian Zhuang of Fujian Agriculture and Forestry University (FAFU). Two diploid ancestors of peanut *A. duranensis* (V14167), *A. ipaensis* (K30076) were provided by Soraya CM Leal-Bertioli of University of Georgia. Tifrunner, Weihua 10 (WH10), YH29 and F_2 individual of (WH10 × YH29) were available in our laboratory, and the segregation population was constructed as described in our previous studies [30]. All the peanut materials were planted at Jinan, Shandong, China. Genomic DNA was isolated from leaflets using Plant Genomic DNA Extraction Kit (TIANGEN, Beijing, China).

2.2. Identification of SSRs using MISA

The genome sequences of Tifrunner, *A. duranensis* and *A. ipaensis* were downloaded from PeanutBase (http://peanutbase.org/) [25,28]. The genome sequences of Shitouqi were obtained by Dr. Weijian Zhuang of FAFU [29]. Perl scripts software MISA (http://pgrc.ipk-gatersleben.de/misa/) was used to scan SSR loci from the genome of Tifrunner, and generate the FASTA file *x.misa*. The MISA was performed using the following parameters [5,31]. For mono-nucleotide repeat motifs, repeat times ≥12; for di-nucleotide with repeats motifs, repeat times ≥6; for tri- and quad-nucleotide repeats motifs, repeat times ≥5; for penta- and hexanucleotide repeats motifs, repeat times ≥4; for compound microsatellites, the interval between two repeats motifs <100 nt.

2.3. Design and filter of SSR primers

SSR primers were designed from the FASTA file *x.misa* generated from MISA and analyzed by perl scripts *p3_in.pl* (http://pgrc.ipk-gatersleben.de/misa/primer3.html). The FASTA file *x.p3in* was obtained which contained the flanking sequences of the microsatellites. Then, the *x.p3in* file was analyzed in primer3 program (https://primer3.org/) for designing the primers, and the result was obtained as the FASTA file *x.p3out*. The parameters of primer3 were as follows. Primer length > 18 and <27, melting temperature > 55°C and <63°C, GC content >30% and <70%, the length of PCR product >80 bp and <300 bp. Finally, the FASTA file *x.p3out* was invoked by perl scripts *p3_out.pl* (http://pgrc.ipk-gatersleben.de/misa/primer3.html), and the result was made available in excel.

2.4. Evaluation of SSRs in diploid and tetraploid peanut genotypes

The amplification efficiency of newly developed SSRs was evaluated using e-PCR program (Version: 2.3.12), which has been replaced by an up-dated program "Primer-BLAST" (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi). The genome sequences of *A. duranensis*, *A. ipaensis*, Tifrunner and Shitouqi were used as templates for primers blast with the following parameters: no more than 2-bp mismatch, no more than 1-bp gap, and the PCR product >50 bp and <1000 bp [8]. Through e-PCR, the size of PCR product for each pair of primers in each template (*Arachis* species) was scored. If the size of PCR products was different in two *Arachis* species, this pair of primers was known as the polymorphic markers.

2.5. Verification using PCR amplification

To verify the e-PCR results, polymerase chain reactions (PCRs) were performed as described in our previous studies [5]. In brief, the PCR reaction was performed in sterile 200-μl tubes using 20-μl final reaction volumes containing dNTP Mix, primers, rTaq (DR100A, TaKaRa, Dalian, China), Mg^{2+}, PCR buffer and template DNA. The amplification program consisted of 94°C (4 min), 35 cycles of 94°C/55–63°C (Table S1)/72°C (30 s each), followed by 72°C for 7 min, and then terminated at 16°C. PCR reaction was performed in Thermal Cycler Dice TP600 (TaKaRa, Dalian, China). The PCR products were detected through polyacrylamide gel electrophoresis [5]. The fragment sizes of the PCR products were estimated by a comparison with DL2000 DNA ladder (TaKaRa, Dalian, China).

3. Results

3.1. Characterization of the SSRs in cultivated peanut Tifrunner genomes

MISA results showed that the total length of the assembled genome of Tifrunner was 2556.9 Mb, in which a total of 512,900 microsatellites were identified with average 200.6 SSRs per Mb. The distance between SSRs was average 5.0 Kb.

A more detailed investigation of individual motifs (repeat types) was performed and a total of 410 types of SSR motifs were detected. Among them, there were 2, 4, 10, 28, 94, 272 types of mono- to hexanucleotide repeats in Tifrunner, respectively. A total of 196,389 dinucleotide repeats (38.3%) were identified, representing the most abundant type of SSRs, followed by mono- (153,866, 30.0%), tri- (119,361, 23.3%), penta- (21,525, 4.2%), tetra- (14,115, 2.8%) and hexa- (7644, 1.5%) nucleotide SSRs (Fig. 1A). We then analyzed the distribution of top 40 types of SSR motifs (Fig. 1B). As expected, A/T had the highest occurrence (28.6% in Tifrunner), followed by AT/AT (22.4% in Tifrunner), AG/CT (12.3% in Tifrunner) and AAT/ATT (11.3% in Tifrunner) (Fig. 1B), which was in consistence with the results of diploid peanut species [5]. In comparison with C/G containing motifs, more A/T enriched motifs were identified in Tifrunner genome. For example, A, AT, AAT, AAAT, AAAAT and AAAAAT represented the most abundant motifs in mono- to hexanucleotide repeats, respectively (Fig. 2).

3.2. Development of genome wide SSR primers

Using primer3 package, a total of 7757 SSRs amplification primers were designed (Table 1). The annealing temperature (Tm) ranged from 55.0 to 62.5°C with average of 57.5°C. The expected length of PCR product ranged from 80 to 300 bp with average length of 168.5 bp (Table 1). For each pair of primers, the SSR repeat motif, position in chromosome, primer sequences, annealing temperature (Tm), and the length of PCR production, *etc.* were listed (Table S1). All these SSR primers were unequally distributed in the 20 chromosomes of Tifrunner with average 372 SSRs per chromosome (Table 2, Table S1). Among them, chromosome 19 had the largest number (520) of SSR markers, followed by chromosome 13 and 03, containing 510 and 462 SSR markers, respectively. Chromosome 07 (247) and 05 (343) had fewer SSR markers in Tifrunner (Table 2, Table S1). These datasets provided a platform for marker assisted breeding, gene and QTL mapping in peanut.

3.3. In silico e-PCR analysis

To test the amplification efficiency, specificity and polymorphism of the newly developed SSR primers, the genome sequences of *A. duranensis*, *A. ipaensis*, Tifrunner and Shitouqi were used as the templates for *in silico* e-PCR analysis. The number, length, and location of the *in silico* PCR product (s) were recorded and summarized (Table 3). In Tifrunner, a total of 4612 (59.46%), 1305 (16.82%), 334 (4.31%) and 1506 (19.41%) of these primers could generate 1, 2, 3, >3 PCR products in Tifrunner, respectively, and the amplification efficiency of these primers was 100%. In Shitouqi, a total 4562 (58.81%), 1278 (16.48%), 327 (4.22%) and 1507 (19.43%) of the primers could generate1, 2, 3, >3 PCR products, respectively. Only 83 (1.07%) of the primers could not generate PCR product, and the amplification efficiency was 99% (Table 3). Notably, nearly 59.46% and 58.81% of these SSRs are single-locus SSR markers which generated only one PCR product in Tifrunner and Shitouqi, respectively (Table 3).

In *A. duranensis*, the A sub-genome donor of cultivated peanut, a total of 4948 (63.79%), 1734 (22.35%), 266 (3.43%), 117 (1.51%) and 692 (8.92%) of these primers could generate 0, 1, 2, 3, >3 PCR products, respectively. In *A. ipaensis*, the B sub-genome donor of cultivated peanut, a total of 5400 (69.61%), 1127 (14.53%), 267 (3.44%), 125 (1.61%) and 838 (10.80%) of them could generate 0, 1, 2, 3, >3 PCR products, respectively. The amplification efficiency was 36.21% and 30.39% in *A. duranensis* and *A. ipaensis*, respectively, which was significantly lower than that in tetraploid peanut species (Table 3).

Venn analysis showed that a total of 1287 (16.6%) of these primers could amplify in both *A. duranensis*, *A. ipaensis*, Tifrunner

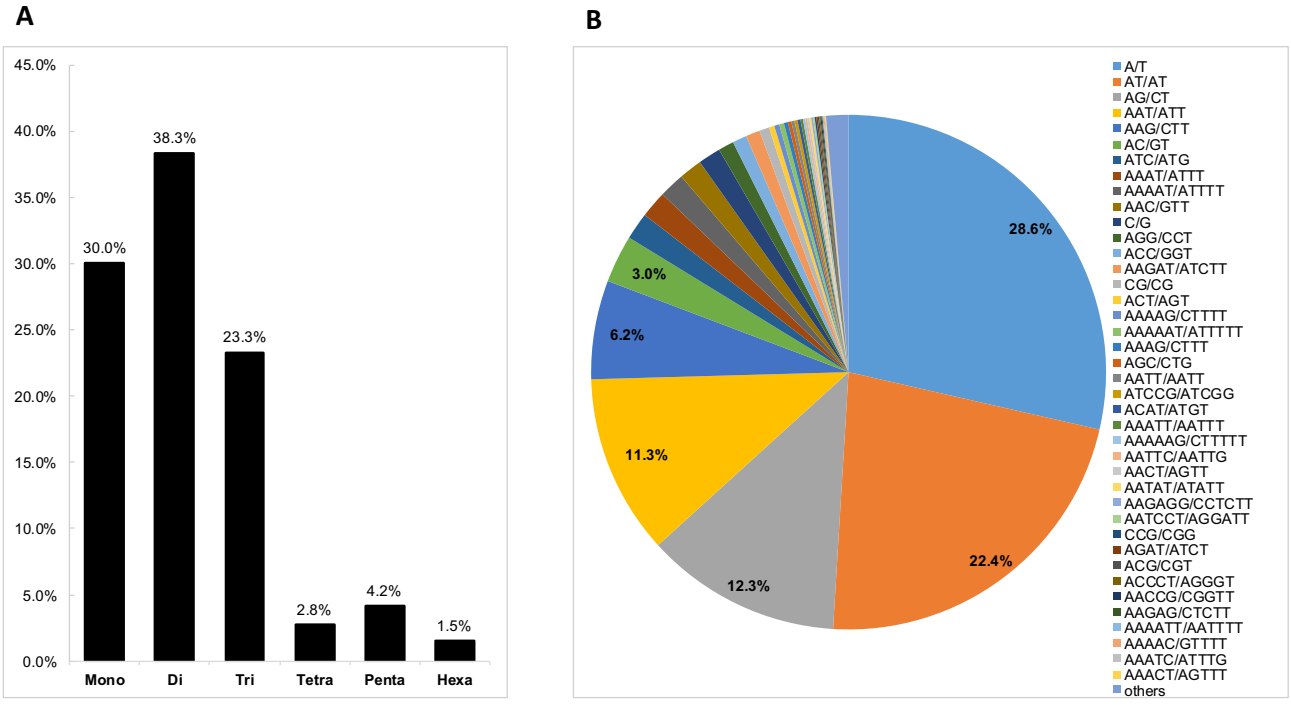

Fig. 1. Distribution of microsatellite motifs in Tifrunner. (A) Distribution of SSR motif length from mono- to hexanucleotide repeats in Tifrunner. Flexible relaxed criteria were used to identify SSRs with minimum repeats number of 12, 6, 5, 5, 4, and 4 for mono- to hexanucleotide repeat motifs. (B) The most abundant SSRs motifs.

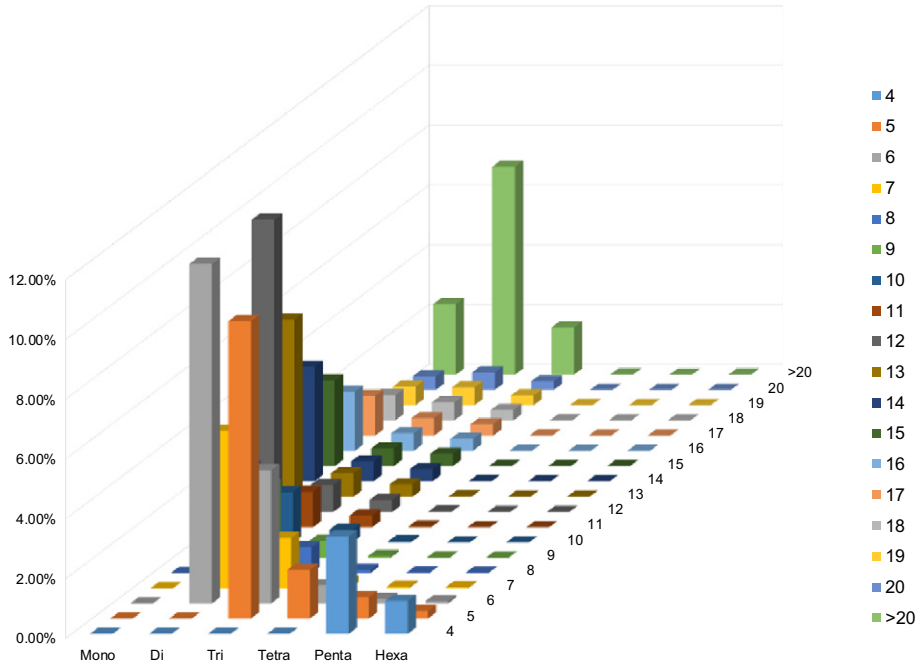

Fig. 2. Distribution of SSR motif repeat number from mono- to hexanucleotide. The vertical axis shows the abundance of SSRs with different motif repeat number (from 4 to >20), which are discriminated by legends of different colors.

and Shitouqi, suggesting these SSR markers had transferability between cultivated and wild *Arachis* species. In addition, we found that 1070 (13.8%), 1522 (19.6%) of these primers were specific in *A. duranensis* and *A. ipaensis*, respectively, implying that the primers can be used to distinguish the *A. duranensis* and *A. ipaensis* (Fig. 3).

3.4. Polymorphism analysis of SSR primers in different cultivars and wild species

To evaluate the polymorphism of the developed SSR primers, the length of e-PCR product was compared between two cultivars, Tifrunner and Shitouqi, and a total of 4562 primers could generate one specific PCR products, and 2286 (50.11%) of them showed polymorphism. Between two wild peanut species, *A. duranensis* and *A. ipaensis*, 1416 out of 1417 (99.93%) of primers showed polymorphism. Between cultivars and wild peanut species, including Tifrunner and *A. duranensis*, Tifrunner and *A. ipaensis*, Shitouqi and *A. duranensis*, Shitouqi and *A. ipaensis*, 4314 out of 4584 (94.11%), 4220 out of 4591 (91.92%), 4204 out of 4465 (91.15%) and 4126 out of 4471 (92.28%) primers showed polymorphism, respectively (Fig. 4). To validate the polymorphism predicted by electric PCR, 20 SSRs were selected for amplification in *A. duranensis*, *A. ipaensis*, Tifrunner and Shitouqi. Our results showed all these primers could produce clear amplification products. Most of the SSRs displayed polymorphism in the tetraploid and diploid peanuts which is consistent with the results of e-PCR (Fig. 5). These results demonstrated that these SSR primers have high polymorphisms, especially between *A. duranensis* and *A. ipaensis*, indicating that these SSRs have a high application value in peanuts, and can be used for distinguishing the alleles from A or B sub-genome.

Table 1
Summary of SSR markers in Tifrunner.

Repeats	Number	Types	Proportion (%)	Frequency (per Mb)
Mono-	66	3	0.85%	0.03
Di-	4743	10	61.14%	1.87
Tri-	2358	36	30.40%	0.93
Tetra-	229	29	2.95%	0.09
Penta-	264	58	3.40%	0.10
Hexa-	97	54	1.25%	0.04
Total	7757	190	100.00%	3.06

3.5. Identification SSR markers linked with purple testa and high-oleic acid

In our previous study, a candidate gene *AhTc1* that controls purple seed coat color was mapped in a 4.7-Mb region (108.0 Mb to 112.7 Mb) in chromosome A10 [30]. In this study, a total of 33 SSR markers were selected between 108.0 to 112.70 Mb of chromosome A10 (Table S1). After screening, we found that two SSR markers *Arahy.10_191964_3* and *Arahy.10_191425_1* showed polymorphism

Table 2
Distribution of SSR primers in chromosomes of Tifrunner.

Chr.	Chromosome size (Mb)	Number	Frequency (per Mb)	Frequency (one every kb)
01	112.42	364	3.24	308.83
02	102.97	346	3.36	297.59
03	143.79	462	3.21	311.24
04	128.80	368	2.86	350.00
05	115.90	343	2.96	337.90
06	115.49	303	2.62	381.16
07	81.10	247	3.05	328.36
08	51.86	321	6.19	161.54
09	120.50	387	3.21	311.36
10	117.08	343	2.93	341.35
11	149.27	381	2.55	391.79
12	120.57	359	2.98	335.85
13	146.71	510	3.48	287.67
14	143.20	353	2.47	405.68
15	160.87	430	2.67	374.11
16	154.79	444	2.87	348.63
17	134.92	435	3.22	310.16
18	135.13	422	3.12	320.22
19	158.59	520	3.28	304.97
20	143.96	407	2.83	353.71
Others	1.98	12	6.05	165.17
Total	2539.90	7757	3.29	320.35

Table 3
Generated number of *in silico* PCR products by genome-wide SSR markers in *A. duranensis*, *A. ipaensis*, Tifrunner and Shitouqi genomes.

In silico PCR in	Zero	One	Two	Three	>Three	Total
A. duranensis	4948 (63.79%)	1734 (22.35%)	266 (3.43%)	117 (1.51%)	692 (8.92%)	7757 (100.00%)
A. ipaënsis	5400 (69.61%)	1127 (14.53%)	267 (3.44%)	125 (1.61%)	838 (10.80%)	7757 (100.00%)
Tifrunner	0 (0.00%)	4612 (59.46%)	1305 (16.82%)	334 (4.31%)	1506 (19.41%)	7757 (100.00%)
Shitouqi	83 (1.07%)	4562 (58.81%)	1278 (16.48%)	327 (4.22%)	1507 (19.43%)	7757 (100.00%)
						7757 (100.00%)
						7757 (100.00%)

between the purple and pink testa parents. To test the linkage between these two markers with the purple testa, 10 purple and 10 pink extreme trait F_2 individuals were selected for verification (Fig. 6A). Results showed that these two SSR markers were all co-segregated with the testa color, which confirmed the QTL-seq results of our previous studies. In addition, this result suggested that these two markers were closely related with the purple testa (Fig. 6B).

In peanut, high-oleic acid trait is jointly controlled by two recessive genes, *FAD2A* and *FAD2B*, which are homologous alleles from A sub-genome and B sub-genome, respectively [32,33]. *FAD2A* was located at 114,194,562- to 114,195,701-bp map position on chromosome 09 and the *FAD2B* was located at 154,049,683- to 154,048,544-bp map position on chromosome 19. In our dataset, there are 11 and 16 SSR markers near to *FAD2A* and *FAD2B*, respectively (Table S1). These SSR markers located on Chr09: 113,236,238~115,199,158 and Chr19: 153,152,701~155,030,189, and the distances between SSR markers with each *FAD2* genes are no more than 1 Mb, indicating that the probabilities of segregation and recombination between these SSR markers with the *FAD* genes are rare. Thus, these SSR markers have potential application value in screening the genotype of *FAD* genes.

4. Discussion

4.1. Quality and specificity of the SSR markers

Despite the rapid development of new generation of molecular markers, such as SNPs and Indels, SSRs are found to be an indispensable marker for genotyping, diversity estimation, QTL mapping and other related genetic and genomic studies. Thus, many efforts have been made for large-scale identification of the SSR markers from ESTs, transcriptomes and genome sequences of diploid peanut [34,35,36]. Cultivated peanut is an allotetraploid (AABB) originated from a single hybridization event between *A. duranensis* (AA) and *A. ipaensis* (BB). Many studies have revealed the presence of repeating sequences in both the genome and the transcriptome of peanut because of their polyploid nature. Hence, many of the previously identified SSR markers come from the homologous DNA sequences. In this study, we developed a total of 7757 SSRs markers using recently assembled genome of cultivated peanuts. Considering the potential repetitive sequences in peanut genomes, strict parameters were applied and low-quality primers were removed including the repeat primers, length of PCR product >2000 bp, mismatched bases > 1 in peanut genome. The e-PCR results demonstrated that nearly 60% of them are single-locus SSR markers, which have advantages in genetics and molecular breeding studies compared with multi-locus markers, such as the specificity in distinguishing the alleles of A and B sub-genomes (Fig. 3).

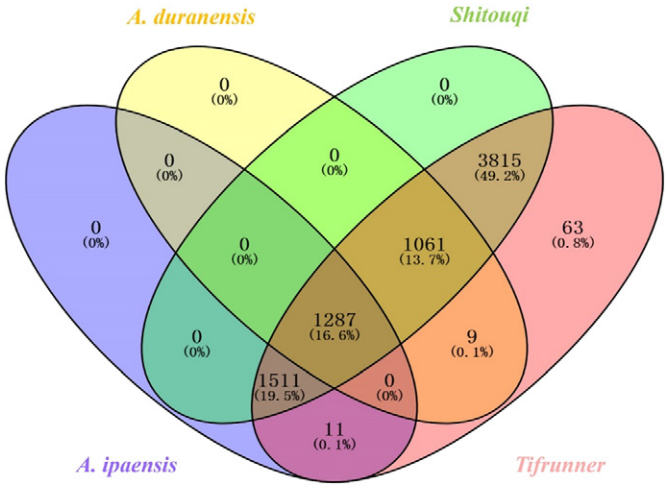

Fig. 3. Transferability of SSR markers between different *Arachis*.

4.2. Application of SSR markers in mapping QTLs/genes and molecular marker-assisted breeding

In the last decades, great achievements have been made in mapping QTLs/genes related to important agronomic traits of peanut, such as purple testa color, oleic acid content, branch habit, seed size, and resistance to nematode, leaf spot and *Tomato spotted wilt virus* (TSWV) [30,37,38,39,40,41]. Using the recently released genome sequences of peanuts [28,29,42], many of these markers, QTL and candidate genes could be re-mapped in specific physical region in peanut genome. In the present study, we developed large-scale genome-wide SSR markers and their positions are specific in peanut genome which can be used for re-mapping these QTLs and even cloning the candidate genes. These SSR markers will be an important genomic resource for the peanut research community.

Testa color is an important agronomic trait for peanut. Recently, peanuts with purple testa have attracted increasing attention of customers due to their higher anthocyanin content and better antioxidant effect [43]. Here, we identified two SSR markers, *Arahy.10_191964_3* and *Arahy.10_191425_1* which were closely related with the purple testa (Fig. 6). In comparison with SNPs, SSR markers can easily be detected cost-effectively, and can be used for breeding the purple testa peanut varieties using molecular marker-assisted breeding (MAS) approach. Moreover, these SSR markers can be used for genotyping all the F_2 lines, which would assist in further fine mapping and cloning candidate genes.

High-oleic acid is one of the important objectives of peanut breeding. To breed high-oleic acid cultivars using MAS together with backcrossing approach, several kinds of molecular markers has been developed for detecting the mutant of *FAD2A* and *FAD2B*, including Cleaved Amplified Polymorphic Sequences (CAPS), Direct Sequencing for PCR products, and Kompetitive allele specific PCR (KASP) [44,45,46]. However, the stability of these methods was sometimes not satisfactory. In this study, we identified 11 and 16 SSR markers near to *FAD2A* and *FAD2B* (Table S1), respectively, which have potential application value in screening the genotype of *FAD* genes, and could be an alternative way to in high oleic acid peanut breeding using MAS.

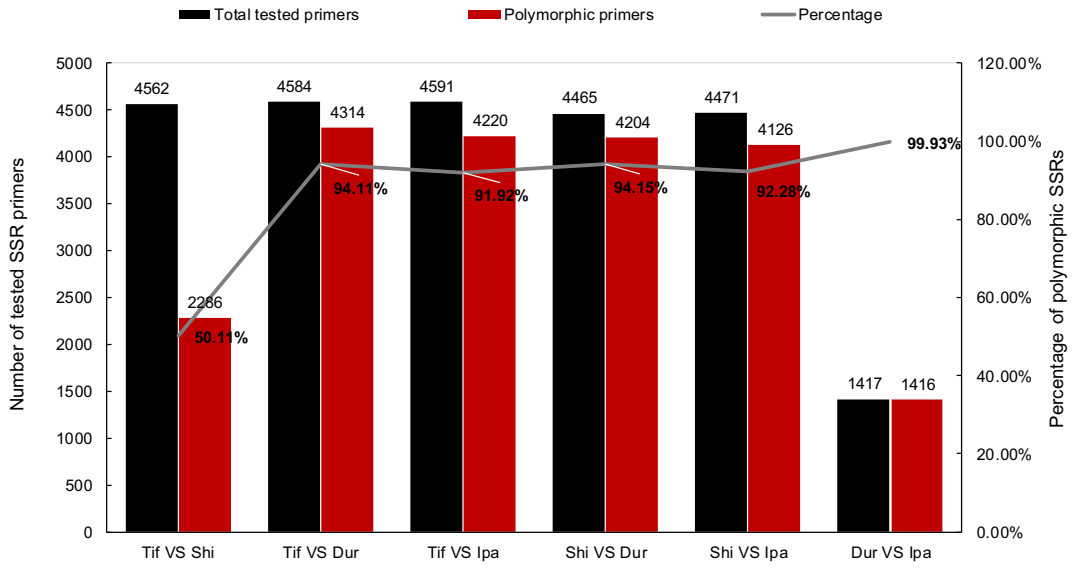

Fig. 4. Characterization of the polymorphism of identified SSRs through pairwise comparison in different *Arachis*. Abbreviation: Tif, Tifrunner; Shi, Shitouqi; Dur, *A. duranensis*; Ipa, *A. ipaensis*.

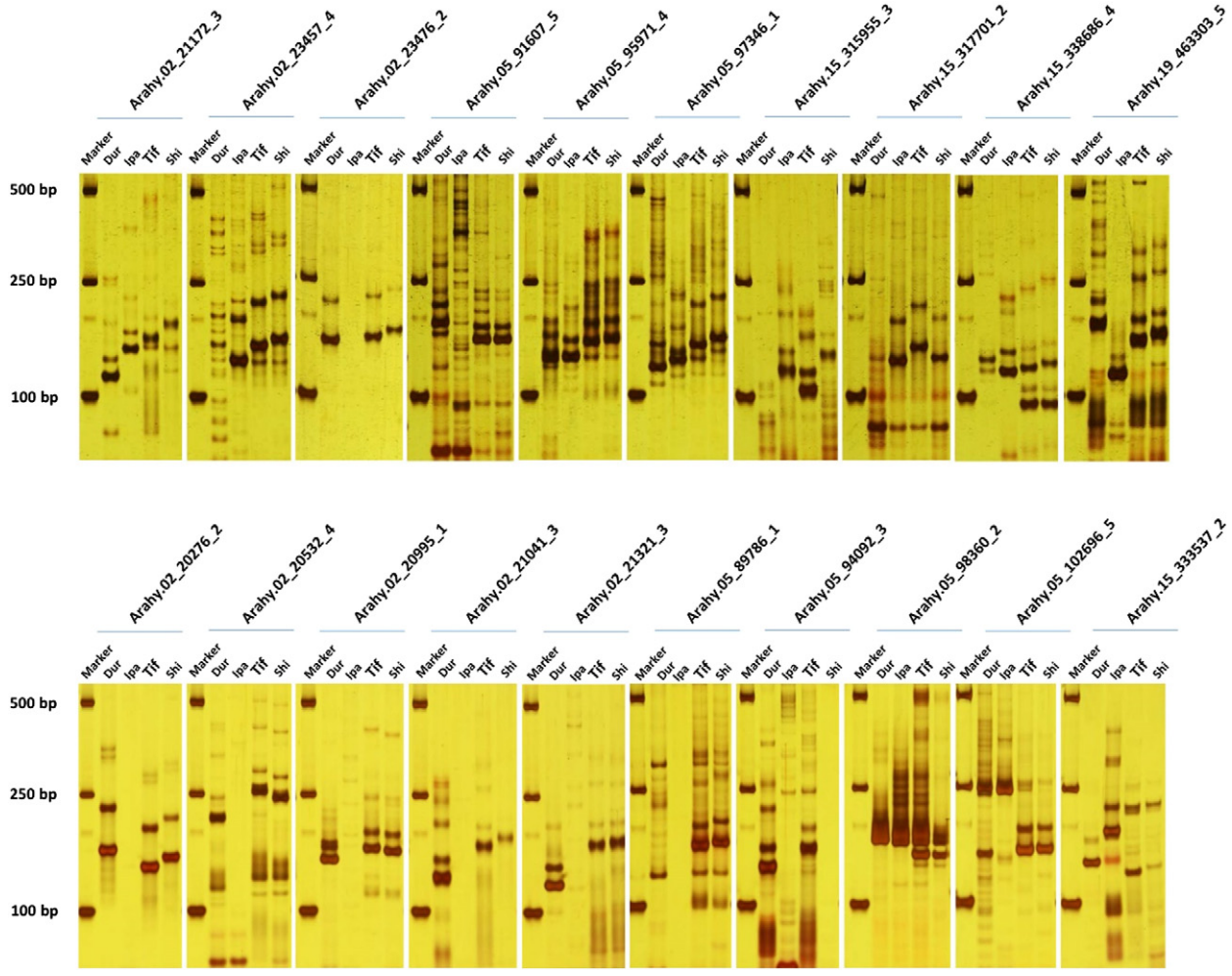

Fig. 5. Validation of the amplification efficiency and polymorphism of SSRs through PCR amplification and electrophoresis analysis. Abbreviation: Tif, Tifrunner; Shi, Shitouqi; Dur, *A. duranensis*; Ipa, *A. ipaensis*.

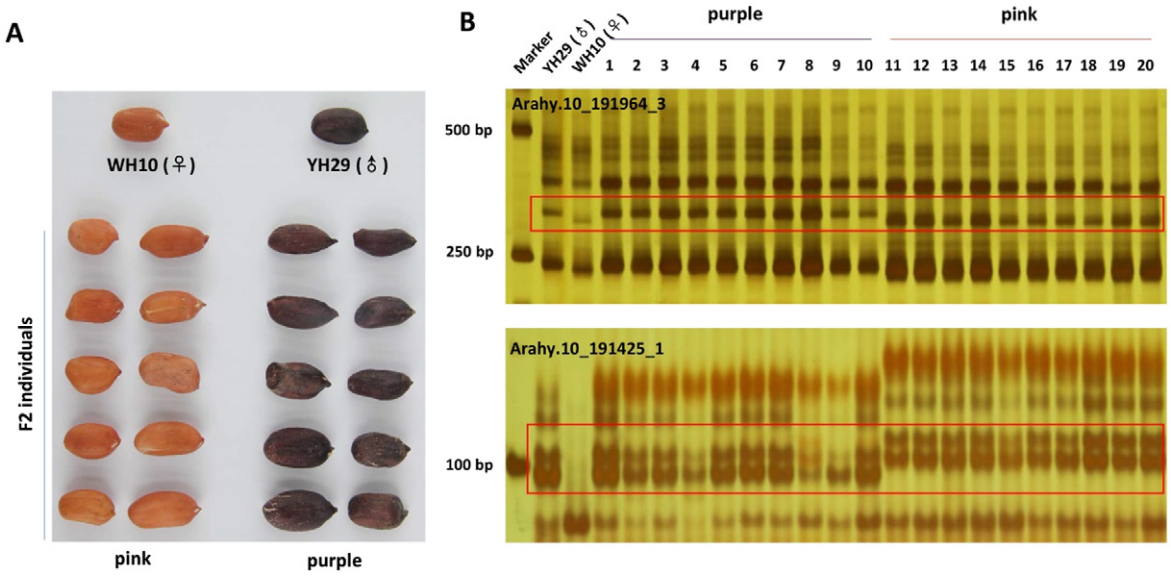

Fig. 6. Screening and identification of SSRs closely related to the purple testa color. A: Seeds of parental lines and F_2 extreme individuals. B: Genotyping the parental lines and F_2 extreme individuals using marker *Arahy.10_191964_3* and marker *Arahy.10_191425_1*.

4.3. SSR markers could be used to accelerate the application of Arachis wild relatives

Molecular marker analysis has shown that cultivated peanut has a narrow genetic base [47]. Many wild species of peanut showed high resistance to biotic and abiotic stresses, and are valuable genetic resources for cultivars improvement [48,49,50]. Resynthesis of tetraploid materials *via* the use of various diploid species have been proved an effective method to utilize the excellent traits of wild species [51]. In this study, the SSR markers showed highly polymorphism between the cultivated and wild peanut species, which could be used for both background selection and tracking target traits.

5. Conclusion

In this study, we reported the development of 7757 filtered genomic SSR markers from assembled genomic sequences of the tetraploid *Arachis* Tifrunner. These SSR markers had high amplification efficiency and polymorphism in four *Arachis* genotypes. Nearly 60% of these markers were single-locus SSRs in tetraploid *Arachis*. We demonstrated that these SSRs have valuable application value in gene/QTL mapping, MAS, conservation of *Arachis* wild relatives, and useful for peanut genetic improvement.

Conflict of interest

The authors declare no competing financial interests.

Financial support

This research is supported by the Natural Science Foundation of China (31861143009), Shandong Province Natural Science Foundation (ZR2017MC005), Shandong Province Germplasm Innovation and Utilization Project (2016LZGC025) and Shandong Key Research Projects (2019GNC106026, 2019GSF107008, 2018GSF121032), Taishan Scholar Project Funding, Agricultural Scientific and Technological Innovation Project of SAAS (CXGC2018E13, CXGC2016B02) and Young Talents Training Program of SAAS. The founders did not play any roles in the design, analysis, and interpretation of this study or relevant data.

Acknowledgments

The authors thank Prof. Weijian Zhuang (Fujian Agriculture and Forestry University) for providing the peanut variety Shitouqi and the genome sequences of Shitouqi. The authors also would like to thank Soraya C M Leal-Bertioli (University of Georgia) for providing the *A. duranensis* and *A. ipaensis*. The authors also acknowledge Dr. Jake C Fountain (University of Georgia) for his help in editing the manuscript.

Supplementary material

https://doi.org/10.1016/j.ejbt.2020.01.004

References

[1] Gupta PK, Balyan HS, Sharma PC, et al. Microsatellites in plants: a new class of molecular markers. Curr Sci 1996;70(1):45–54.
[2] Haq SU, Jain R, Sharma M, et al. Identification and characterization of microsatellites in expressed sequence tags and their cross transferability in different plants. Int J Genomics 2014;2014(863948). https://doi.org/10.1155/2014/863948 PMid: 25389527.
[3] Varshney RK, Thiel T, Stein N, et al. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 2002;7(2A):537–46. [PMid: 12378259].
[4] Kalia RK, Rai MK, Kalia S, et al. Microsatellite markers: an overview of the recent progress in plants. Euphytica 2011;177(3):309–34. https://doi.org/10.1007/s10681-010-0286-9.
[5] Zhao C, Qiu J, Agarwal G, et al. Genome-wide discovery of microsatellite markers from diploid progenitor species, *Arachis duranensis* and *A. ipaensis*, and their application in cultivated peanut (*A. hypogaea*). Front Plant Sci 2017;8(1209). https://doi.org/10.3389/fpls.2017.01209. PMid: 28769940.
[6] Balfourier F, Roussel V, Strelchenko P, et al. A worldwide bread wheat core collection arrayed in a 384-well plate. Theor Appl Genet 2007;114(7):1265–75. https://doi.org/10.1007/s00122-007-0517-1. PMid: 17318494.
[7] Han B, Wang C, Tang Z, et al. Genome-wide analysis of microsatellite markers based on sequenced database in Chinese spring wheat (*Triticum aestivum* L.). PLoS One 2015;10(11):e0141540. https://doi.org/10.1371/journal.pone.0141540 PMid: 26536014.
[8] Shi J, Huang S, Zhan J, et al. Genome-wide microsatellite characterization and marker development in the sequenced *Brassica* crop species. DNA Res 2013;21(1):53–68. https://doi.org/10.1093/dnares/dst040. PMid: 24130371.
[9] Wang Q, Fang L, Chen J, et al. Genome-wide mining, characterization, and development of microsatellite markers in *Gossypium* species. Sci Rep 2015;5(1):10638. https://doi.org/10.1038/srep10638. PMid: 26030481.
[10] Addisalem AB, Esselink GD, Bongers F, et al. Genomic sequencing and microsatellite marker development for *Boswellia papyrifera*, an economically important but

threatened tree native to dry tropical forests. AoB Plants 2015;7:plu086. https://doi.org/10.1093/aobpla/plu086. PMid: 25573702.
[11] Song X, Ge T, Li Y, et al. Genome-wide identification of SSR and SNP markers from the non-heading *Chinese cabbage* for comparative genomic analyses. BMC Genomics 2015;16(1):328. https://doi.org/10.1186/s12864-015-1534-0. PMid: 25908429.
[12] Zhao H, Yang L, Peng Z, et al. Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus *Phyllostachys*. Sci Rep 2015;5:8018. https://doi.org/10.1038/srep08018. PMid: 25620112.
[13] Parida SK, Verma M, Yadav SK, et al. Development of genome-wide informative simple sequence repeat markers for large-scale genotyping applications in chickpea and development of web resource. Front Plant Sci 2015;6(645). https://doi.org/10.3389/fpls.2015.00645. PMid: 26347762.
[14] Liu F, Hu Z, Liu W, et al. Distribution, function and evolution characterization of microsatellite in *Sargassum thunbergii* (Fucales, Phaeophyta) transcriptome and their application in marker development. Sci Rep 2016;6:18947. https://doi.org/10.1038/srep18947. PMid: 26732855.
[15] Saha D, Rana RS, Chakraborty S, et al. Development of a set of SSR markers for genetic polymorphism detection and interspecific hybrid jute breeding. Crop J 2017;5(5):416–29. https://doi.org/10.1016/j.cj.2017.02.006.
[16] Saha D, Rana RS, Das S, et al. Genome-wide regulatory gene-derived SSRs reveal genetic differentiation and population structure in fiber flax genotypes. J Appl Genet 2019;60(1):13–25. https://doi.org/10.1007/s13353-018-0476-z. PMid: 30368734.
[17] Kochert G, Halward T, Branch WD, et al. RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor Appl Genet 1991;81(5):565–70. https://doi.org/10.1007/BF00226719. PMid: 24221368.
[18] Hopkins MS, Casa AM, Wang T, et al. Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. Crop Sci 1999;39(4):1243–7. https://doi.org/10.1186/1756-0500-5-86 PMid: 22305491.
[19] He G, Meng R, Newman M, et al. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biol 2003;3:3. https://doi.org/10.1186/1471-2229-3-3 PMid: 12713672.
[20] Ferguson ME, Burow MD, Schulze SR, et al. Microsatellite identification and characterization in peanut (*A. hypogaea* L.). Theor Appl Genet 2004;108(6):1064–70. https://doi.org/10.1007/s00122-003-1535-2 PMid: 15067392.
[21] Macedo SE, Moretzsohn MC, Leal-Bertioli SCM, et al. Development and characterization of highly polymorphic long TC repeat microsatellite markers for genetic analysis of peanut. BMC Res Notes 2012;5(1):86. https://doi.org/10.1186/1756-0500-5-86 PMid: 22305491.
[22] Proite K, Leal-Bertioli SCM, Bertioli DJ, et al. ESTs from a wild Arachis species for gene discovery and marker development. BMC Plant Biol 2007;7(1):7. https://doi.org/10.1186/1471-2229-7-7 PMid: 17302987.
[23] Song GQ, Li MJ, Xiao H, et al. EST sequencing and SSR marker development from cultivated peanut (*Arachis hypogaea* L.). Electron J Biotechnol 2010;13(3):7–8. https://doi.org/10.2225/vol13-issue3-fulltext-10.
[24] Wang H, Penmetsa RV, Yuan M, et al. Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biol 2012;12:10. https://doi.org/10.1186/1471-2229-12-10 PMid: 22260238.
[25] Bertioli DJ, Cannon SB, Froenicke L, et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 2016;48(4):438. https://doi.org/10.1038/ng.3517 PMid: 26901068.
[26] Khedikar Y, Pandey MK, Sujay V, et al. Identification of main effect and epistatic quantitative trait loci for morphological and yield-related traits in peanut (*Arachis hypogaea* L.). Mol Breed 2018;38:7. https://doi.org/10.1007/s11032-017-0764-z.
[27] Qiu J, Zhao Y, Deng L, et al. Molecular identification on true F1 hybrid of peanut. Shandong Agric Sci 2018;50(2):13–8.
[28] Bertioli DJ, Jenkins J, Clevenger J, et al. The genome sequence of segmental allotetraploid peanut Arachis hypogaea. Nat Genet 2019;51(5):877. https://doi.org/10.1038/s41588-019-0405-z. PMid: 31043755. https://doi.org/10.1038/s41588-019-0405-z PMid: 31043755.
[29] Zhuang W, Chen H, Yang M, et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet 2019;51(5):865. https://doi.org/10.1038/s41588-019-0402-2 PMid: 31043757.
[30] Zhao Y, Ma J, Li M, et al. Whole-genome resequencing-based QTL-seq identified *AhTc1* gene encoding a R2R3-MYB transcription factor controlling peanut purple testa colour. Plant Biotechnol J 2020;18:96–105. https://doi.org/10.1111/pbi.13175 PMid: 31131506.
[31] Deng P, Wang M, Feng K, et al. Genome-wide characterization of microsatellites in *Triticeae* species: abundance, distribution and evolution. Sci Rep 2016;6:32224. https://doi.org/10.1038/srep32224 PMid: 27561724.
[32] Jung S, Swift D, Sengoku E, et al. The high oleate trait in the cultivated peanut [*Arachis hypogaea* L.]. I. Isolation and characterization of two genes encoding microsomal oleoyl-PC desaturases. Mol Gen Genet MGG 2000;263(5):796–805. https://doi.org/10.1007/s004380000244 PMid: 10905347.
[33] Jung S, Powell G, Moore K, et al. The high oleate trait in the cultivated peanut [*Arachis hypogaea* L.]. II. Molecular basis and genetics of the trait. Mol Gen Genet MGG 2000;263(5):806–11. https://doi.org/10.1007/s004380000243 PMid: 10905348.
[34] Liang X, Chen X, Hong Y, et al. Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. BMC Plant Biol 2009;9(1):35. https://doi.org/10.1186/1471-2229-9-35 PMid: 19309524.
[35] Zhou X, Dong Y, Zhao J, et al. Genomic survey sequencing for development and validation of single-locus SSR markers in peanut (*Arachis hypogaea* L.). BMC Genomics 2016;17(1):420. https://doi.org/10.1186/s12864-016-2743-x PMid: 27251557.
[36] Zhong R, Zhou M, Zhao C, et al. SSR marker development from peanut gynophore transcriptome sequencing. Plant Breed 2016;135(1):111–7. https://doi.org/10.1111/pbr.12336.
[37] Wang H, Pandey MK, Qiao L, et al. Genetic mapping and quantitative trait loci analysis for disease resistance using F_2 and F_5 generation-based genetic maps derived from 'Tifrunner' × 'GT-C20' in peanut. Plant Genome 2013;6(3). https://doi.org/10.3835/plantgenome2013.05.0018.
[38] Burow MD, Starr JL, Park CH, et al. Introgression of homeologous quantitative trait loci (QTLs) for resistance to the root-knot nematode [*Meloidogyne arenaria* (Neal) Chitwood] in an advanced backcross-QTL population of peanut (*Arachis hypogaea* L.). Mol Breed 2014;34(2):393–406. https://doi.org/10.1007/s11032-014-0042-2.
[39] Kayam G, Brand Y, Faigenboim-Doron A, et al. Fine-mapping the branching habit trait in cultivated peanut by combining bulked segregant analysis and high-throughput sequencing. Front Plant Sci 2017;8(467). https://doi.org/10.3389/fpls.2017.00467 PMid: 28421098.
[40] Luo H, Ren X, Li Z, et al. Co-localization of major quantitative trait loci for pod size and weight to a 3.7 cM interval on chromosome A05 in cultivated peanut (*Arachis hypogaea* L.). BMC Genomics 2017;18(1):58. https://doi.org/10.1186/s12864-016-3456-x PMid: 28068921.
[41] Chu Y, Chee P, Culbreath A, et al. Major QTLs for resistance to early and late leaf spot diseases are identified on chromosomes 3 and 5 in peanut (*Arachis hypogaea*). bioRxiv 2019:567206. https://doi.org/10.3389/fpls.2019.00883 PMid: 31333711.
[42] Yin D, Ji C, Ma X, et al. Genome of an allotetraploid wild peanut *Arachis monticola*: a *de novo* assembly. GigaScience 2018;7(6):giy066. https://doi.org/10.1093/gigascience/giy066.
[43] Kuang Q, Yu Y, Attree R, et al. A comparative study on anthocyanin, saponin, and oil profiles of black and red seed coat peanut (*Arachis hypogaea*) grown in China. Int J Food Prop 2017;20(sup1):S131–40. https://doi.org/10.1080/10942912.2017.1291676.
[44] Chu Y, Ramos L, Holbrook CC, et al. Frequency of a loss-of-function mutation in oleoyl-PC desaturase (ahFAD2A) in the mini-core of the US peanut germplasm collection. Crop Sci 2007;47(6):2372–8. https://doi.org/10.2135/cropsci2007.02.0117.
[45] Chu Y, Holbrook CC, Ozias-Akins P. Two alleles of ahFAD2B control the high oleic acid trait in cultivated peanut. Crop Sci 2009;49(6):2029–36. https://doi.org/10.2135/cropsci2009.01.0021.
[46] Zhao S, Li A, Li C, et al. Development and application of KASP marker for high throughput detection of *AhFAD2* mutation in peanut. Electron J Biotechnol 2017;25:9–12. https://doi.org/10.1016/j.ejbt.2016.10.010.
[47] Halward T, Stalker T, LaRue E, et al. Use of single-primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L.). Plant Mol Biol 1992;18(2):315–25. https://doi.org/10.1007/BF00034958 PMid: 1731991.
[48] Pande S, Rao JN. Resistance of wild *Arachis* species to late leaf spot and rust in greenhouse trials. Plant Dis 2001;85(8):851–5. https://doi.org/10.1094/PDIS.2001.85.8.851 PMid: 30823052.
[49] Simpson CE. Use of wild *Arachis* species/introgression of genes into *A. hypogaea* L. Peanut Sci 2001;28(2):114–6. https://doi.org/10.3146/i0095-3679-28-2-12.
[50] Leal-Bertioli SCM, Santos SP, Dantas KM, et al. *Arachis batizocoi*: a study of its relationship to cultivated peanut (*A. hypogaea*) and its potential for introgression of wild genes into the peanut crop using induced allotetraploids. Ann Bot 2014;115(2):237–49. https://doi.org/10.1093/aob/mcu237 PMid: 25538110.
[51] Hickey LT, Hafeez AN, Robinson H, et al. Breeding crops to feed 10 billion. Nat Biotechnol 2019;37:744–54. https://doi.org/10.1038/s41587-019-0152-9 PMid: 31209375.