

## Development of EST-SSR markers in peanut (*Arachis hypogaea* L.)

Xinyan WEI, Lifeng LIU (✉), Shunli CUI, Huanying CHEN, Jingjing ZHANG

*Agricultural University of Hebei, Baoding 071001, China*

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2011

**Abstract** More molecular markers for potential use in peanut genetic research were developed. A total of 92403 *EST* sequences of peanut (*Arachis*) in the NCBI database were downloaded and analyzed. 2594 SSRs distributed in 2267 non-redundant *EST* sequences were detected, with tri-nucleotide motif (65.54%) as the most abundant motif type followed by di-nucleotide motif (28.10%). Among the 92 repeat types, the top eight motif types were AG/TC (20.1%), AAG/TTC (11.8%), AAT/TTA (10.1%), AGG/TCC (6.6%), AGA/TCT (6.3%), AT/TA (5.9%), ACT/TGA (3.8%), and ATG/TAC (3.7%) with higher frequency. A total of 237 primer pairs were successfully designed based on the 2267 SSR-ESTs using DNA star software.

**Keywords** peanut, EST-SSR, development

### Introduction

Peanut (*Arachis hypogaea* L.) is an important vegetable oil and protein crop and is planted in most countries. Different types of peanut varieties were different in morphology and physiological and agronomic characters, but there were not enough useful markers for quantitative trait locus (QTL) analysis and marker-assisted selection (Raina et al., 2001). Current genetic research in peanuts is hindered for lack of sufficient molecular markers. Microsatellite or simple sequence repeat (SSR) markers are co-dominant, multiallelic, and highly polymorphic genetic markers (Powell et al., 1996). EST-SSRs have received a lot of attention when the amount of ESTs deposited in databases for various plants is increasing (Varshney et al., 2005). EST-SSRs exist in many plants like barley, maize, rice, sorghum and wheat (Kantety et al., 2002), and this indicates that EST-SSR markers have potential use in peanut genetic studies. SSR marker showed more polymorphism than anyone of the other markers in peanut (Han et

al., 2004; Tang et al., 2004; Gimenes et al., 2007; Liang et al., 2009). The application of SSR marker is not only to find the genetic diversity of peanut (Hopkins et al., 1999; He et al., 2003; Moretzsohn et al., 2004; Tang et al., 2007; Hong et al., 2008) but also to draw the genetic linkage map of peanut (Moretzsohn et al., 2005; Varshney et al., 2009; Hong et al., 2010).

A vast amount of available EST sequence data have been generated from the development of peanut EST projects, and ESTs have been established from a wild *Arachis* species for gene discovery and marker development (Proite et al., 2007). Therefore, these data offer an opportunity to identify SSR in ESTs by data mining. ESTs of peanut in the NCBI database before September 1, 2009 were downloaded and analyzed. By designing primers, more molecular markers for potential use in peanut genetic research were developed in our study.

### Materials and methods

#### Sequence data

ESTs of peanut in the NCBI database (<http://www.ncbi.nlm.nih.gov>) ended before September 1, 2009 and were downloaded to obtain a total of 92403 ESTs sequences. DNASTar

program was used for sequence splicing and redundancy analysis.

### Screening of SSRs

The SSR loci were searched using the online software of Simple Sequence Repeat Identification Tool (SSRIT). The SSR motifs, with repeat units more than seven times in dinucleotides, and more than five times in tri-, tetra-, penta- and hexa-nucleotides, were considered as SSR search criteria.

### Primer design

The Primer Select of DNASTAR Program was used for designing the primer pairs for SSRs, with the primer length from 18 to 24 bp, T<sub>m</sub> from 40 to 65°C and amplification rate of more than 70%.

## Results

### Occurrence of SSRs in non-redundant peanut ESTs

92403 EST sequences of peanut in the NCBI database were downloaded; 5834 SSR-EST sequences were screened (6.3%). After all redundancy 2267 SSR-containing EST sequences remained and 2594 SSR loci included 729 dinucleotide loci (28.1%) and 1700 trinucleotide loci (65.54%) (Table 1).

**Table 1** Occurrence of SSRs in non-redundant peanut ESTs

Repeat type	SSR number	Proportion in all SSRs (%)
Dinucleotide	729	28.10
Trinucleotide	1700	65.54
Tetranucleotide	108	4.16
Pentanucleotide	30	1.16
Hexanucleotide	22	0.85
Octamer	5	0.19
Total	2594	100

### Occurrence of different SSR of peanut

The total number of motif sequence types in SSR-EST was 92, with di-, tri-, tetra-, penta- and hexa- nucleotides of 3, 18, 26, 18 and 22. Di- and tri- nucleotides were predominant (93.64%) and others were negligible (6.36%).

The frequency of various repeat units varied greatly and the top eight repeat units were AG/TC (20.1%), AAG/TTC (11.8%), AAT/TTA (10.1%), AGG/TCC (6.6%), AGA/TCT (6.3%), AT/TA (5.9%), ACT/TGA (3.8%), and ATG/TAC (3.7%) and other 84 repeat units were 31.7% frequency. The frequency of AG/TC and AAG/TTC was 71.47% among the

di-nucleotide motifs and 17.94% among the tri-nucleotide motifs, respectively (Table 2).

### Primer design

Only 237 prime pairs were successfully designed based on the 2267 SSR-EST sequences. These prime pairs contained five kinds of repeat types, one of which had 213 pairs, with di-, tri-, tetra-, penta- and hexa-nucleotides of 56, 143, 9, 2, 3 pairs, respectively. The remaining 24 pairs included two kinds of repeat types.

## Discussion

EST-SSR markers have more advantages than the general molecular markers, e.g. RFLP-, RAPD-, AFLP- and gSSR-marker, with high information and excellent transferability (Decroocq et al., 2003; Shangguan et al., 2010). If an EST is tagged with a genetic traits chain, this EST may be related with the gene controlling the traits. The developing EST-SSR marker is simple, quick and low in cost.

A high proportion of redundant ESTs usually appeared as the result of random sequencing within cDNA libraries. In this study, SSR search was performed following redundancy elimination to reduce the data set size and avoid over-estimation of the EST-SSR frequency. 92403 EST sequences of peanut in the NCBI database were downloaded for SSR search, 6.3% of which (5834) contained SSR motifs, higher than that of rice (3.4%) (Cardle et al., 2000), bread wheat (5.4%) (Gupta et al., 2003), maize (1.4%), and barley (3.4%) (Kantety et al., 2002) but lower than soybean (Song et al., 2004). It was similar to the previous reports for cultivated peanuts (Liang et al., 2009). This indicated that there were relatively higher abundant SSRs in peanut ESTs. The different abundance of SSRs was known to be dependent on different species, SSR search criteria, and database-mining tools (Varshney et al., 2005).

In earlier reports, both di-nucleotide repeat and trinucleotide repeat were generally the most common motif (Li et al., 2004; Varshney et al., 2005), but the repeat types were usually different. In this study, the total number of motif sequence types in SSR-EST was 92 including 65.54% trinucleotide repeats and 28.10% di-nucleotide repeats. Therefore, the AG/TC with 20.1% frequency was the most abundant motif for most plants, while AAG/TTC with 11.8% frequency was similar to that of *Arabidopsis thaliana* (14%) (Cardle et al., 2000) and soybean (Gao et al., 2003).

A total of 237 pairs of primers were successfully designed based on the 2267 SSR-EST sequences and had 10.5% frequency. The polymorphism of these primers was detected using local peanut varieties of Hebei Province, China.





(Continued)

Repeat unit	Repeat time											Total				
	5	6	7	8	9	10	11	12	13	14	15		16			
TTATC/AATAG		1														1
TTATT/AATAA	1															1
TTCCC/AAGGG		1														1
AAAAAG/TTTTTC	1															1
AACAIG/TTGTAC	1															1
AAGAGG/TTCTCC	1															1
AGAAAG/TCCTCC	1															1
ATCACATAGTGT	1															1
ATCGTC/TAGCAG	1															1
ATCTCT/TAGAGA	1															1
CACCAG/GTGGTC	1															1
CACCGC/GTGGCG	1															1
CATCAC/GTAGTG		1														1
CATGAG/GTACTC	1															1
CGAAGA/GCTTCT	1															1
CGTCAC/GCAGTA	1															1
CTTTCC/GAAAAGG	1															1
GAAGAT/CGGCTA	1															1
GAGAAG/CTCTTC	1															1
GTGGCG/CACCGC	1															1
TCCAGC/AGGTCTG	1															1
TCTCTT/AGAGAA	1															1
TCTGGT/AGACCA	1									1						1
TCTTCA/AGAAAGT	1															1
TTACTA/AATGAT							1									1
GAGTAGT/ATCATCA			1													1
TAATCAA/ATTAGTT	1															1
TCACCAG/AGTGGTC	1															1
TCTTCTT/AGAAGAA		1														1
TTAGAAC/AATCTTG	1															1
Total	995	428	461	202	144	83	58	47	41	32	18	85	2594			

## Acknowledgements

This work was financially supported by the Natural Science Fund of Hebei Province, China (No. C2009000591), Project for the study of Application Foundation of Hebei Province, China (No. 10960120D) and the special fund for Modern Agro-industry Technology Research System, China.

## References

- Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*, 156(2): 847–854
- Decroocq V, Favé M G, Hagen L, Bordenave L, Decroocq S (2003). Development and transferability of apricot and grape EST microsatellite markers across taxa. *Theor Appl Genet*, 106(5): 912–922
- Gao L F, Tang J F, Li H W, Jia J Z (2003). Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding*, 12(3): 245–261
- Gimenes M A, Hoshino A A, Barbosa A V, Palmieri D A, Lopes C R (2007). Characterization and transferability of microsatellite markers of the cultivated peanut (*Arachis hypogaea*). *BMC Plant Biol*, 7: 9–21
- Gupta P K, Rustgi S, Sharma S, Singh R, Kumar N, Balyan H S (2003). Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Genet Genomics*, 270(4): 315–323
- Han Z Q, Gao G Q, Wei P X, Tang R H, Zhong R C (2004). Analysis of DNA polymorphism and genetic relationships in cultivated peanut (*Arachis hypogaea* L.) using microsatellite markers. *Acta Agron Sin*, 11: 1097–1101 (in Chinese)
- He G, Meng R, Newman M, Gao G, Pittman R N, Prakash C S (2003). Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol*, 3: 3–8
- Hong Y B, Chen X P, Liang X Q, Liu H Y, Zhou G Y, Li S X, Wen S J, Holbrook C C, Guo B Z (2010). A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. *BMC Plant Biol*, 10: 17–39
- Hong Y B, Liang X Q, Chen X P, Lin K Y, Zhou G Y, Li S X, Liu H Y (2008). Genetic differences in peanut cultivated types (*Arachis hypogaea*) revealed by SSR polymorphism. *Mol Plant Breed*, 6(1): 71–78 (in Chinese)
- Hopkins M S, Casa A M, Wang T, Mitchell S E, Dean R E, Kochert G D, Kresovich S (1999). Discovery and characterization of polymorphic simple sequence repeat (SSRs) in peanut. *Crop Science*, 39(4): 1243–1248
- Kantety R V, La Rota M, Matthews D E, Sorrells M E (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol*, 48(5/6): 501–510
- Li Y Q, Li H W, Gao Li F, He B R (2004). Progress of simple sequence repeats derived from expressed sequence tags. *Journal of Plant Genetic Resources*, 5: 91–95 (in Chinese)
- Liang X Q, Chen X P, Hong Y B, Liu H, Zhou G Y, Li S X, Guo B Z (2009). Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. *BMC Plant Biol*, 9: 35–43
- Moretzsohn M C, Hopkins M S, Mitchell S E, Kresovich S, Valls J F, Ferreira M E (2004). Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biol*, 4: 11–20
- Moretzsohn M C, Leoi L, Proite K, Guimarães P M, Leal-Bertioli S C, Gimenes M A, Martins W S, Valls J F, Grattapaglia D, Bertioli D J (2005). A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (*Fabaceae*). *Theor Appl Genet*, 111(6): 1060–1071
- Powell W, Machray G C, Provan J (1996). Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*, 1: 215–222
- Proite K, Leal-Bertioli S C, Bertioli D J, Moretzsohn M C, da Silva F R, Martins N F, Guimarães P M (2007). ESTs from a wild *Arachis* species for gene discovery and marker development. *BMC Plant Biol*, 7: 7–16
- Raina S N, Rani V, Kojima T, Ogihara Y, Singh K P, Devarumath R M (2001). RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species. *Genome*, 44(5): 763–772
- Shangguan L F, Li X Y, Song C N, Wang X C, Wang Y Z, Zhang Z, Fang J G (2010). Development of EST-SSR markers in *Prunus mume* and its application. *Acta Botanica Boreali-Occidentalia Sinica*, 30(9): 1766–1772 (in Chinese)
- Song Q J, Marek L F, Shoemaker R C, Lark K G, Concibido V C, Delannay X, Specht J E, Cregan P B (2004). A new integrated genetic linkage map of the soybean. *Theor Appl Genet*, 109(1): 122–128
- Tang R H, Gao G Q, He L Q, Han Zh Q, Shan Sh H, Zhong R Ch, Zhou C Q, Jiang J, Li Y R, Zhuang W J (2007). Genetic diversity in cultivated groundnut based on SSR markers. *Journal of Genetics and Genomics*, 34(5): 449–459
- Tang R H, Zhuang W J, Gao G Q, Han Zh Q, Zhong R Ch, He L Q, Zhou C Q (2004). Simple sequence repeats polymorphism among accessions of var.vulgaris Harz in *Arachis hypogaea* L.. *Chinese Journal of Oil Crop Sciences*, 26(2): 21–26 (in Chinese)
- Varshney R K, Bertioli D J, Moretzsohn M C, Vadez V, Krishnamurthy L, Aruna R, Nigam S N, Moss B J, Seetha K, Ravi K, He G, Knapp S J, Hoisington D A (2009). The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). *Theor Appl Genet*, 118(4): 729–739
- Varshney R K, Graner A, Sorrells M E (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol*, 23(1): 48–55