

Research article

Open Access

Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species

Xuanqiang Liang*¹, Xiaoping Chen¹, Yanbin Hong¹, Haiyan Liu¹,
Guiyuan Zhou¹, Shaoxiong Li¹ and Baozhu Guo²

Address: ¹Crops Research Institute, Guangdong Academy of Agricultural Sciences, Wushan 510640, Guangzhou, PR China and ²USDA-ARS, Crop Protection and Management Research Unit, Tifton, Georgia, USA

Email: Xuanqiang Liang* - Liang-804@163.com; Xiaoping Chen - xpchen@uga.edu; Yanbin Hong - Hongyanbin1979@yahoo.com.cn; Haiyan Liu - Liu_Haiyan001@126.com; Guiyuan Zhou - zhuguyuan418@163.com; Shaoxiong Li - lishaoxiong@vip.sohu.com; Baozhu Guo - Baozhu.Guo@ars.usda.gov

* Corresponding author

Published: 24 March 2009

Received: 13 October 2008

BMC Plant Biology 2009, 9:35 doi:10.1186/1471-2229-9-35

Accepted: 24 March 2009

This article is available from: <http://www.biomedcentral.com/1471-2229/9/35>

© 2009 Liang et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Lack of sufficient molecular markers hinders current genetic research in peanuts (*Arachis hypogaea* L.). It is necessary to develop more molecular markers for potential use in peanut genetic research. With the development of peanut EST projects, a vast amount of available EST sequence data has been generated. These data offered an opportunity to identify SSR in ESTs by data mining.

Results: In this study, we investigated 24,238 ESTs for the identification and development of SSR markers. In total, 881 SSRs were identified from 780 SSR-containing unique ESTs. On an average, one SSR was found per 7.3 kb of EST sequence with tri-nucleotide motifs (63.9%) being the most abundant followed by di- (32.7%), tetra- (1.7%), hexa- (1.0%) and penta-nucleotide (0.7%) repeat types. The top six motifs included AG/TC (27.7%), AAG/TTC (17.4%), AAT/TTA (11.9%), ACC/TGG (7.72%), ACT/TGA (7.26%) and AT/TA (6.3%). Based on the 780 SSR-containing ESTs, a total of 290 primer pairs were successfully designed and used for validation of the amplification and assessment of the polymorphism among 22 genotypes of cultivated peanuts and 16 accessions of wild species. The results showed that 251 primer pairs yielded amplification products, of which 26 and 221 primer pairs exhibited polymorphism among the cultivated and wild species examined, respectively. Two to four alleles were found in cultivated peanuts, while 3–8 alleles presented in wild species. The apparent broad polymorphism was further confirmed by cloning and sequencing of amplified alleles. Sequence analysis of selected amplified alleles revealed that allelic diversity could be attributed mainly to differences in repeat type and length in the microsatellite regions. In addition, a few single base mutations were observed in the microsatellite flanking regions.

Conclusion: This study gives an insight into the frequency, type and distribution of peanut EST-SSRs and demonstrates successful development of EST-SSR markers in cultivated peanut. These EST-SSR markers could enrich the current resource of molecular markers for the peanut community and would be useful for qualitative and quantitative trait mapping, marker-assisted selection, and genetic diversity studies in cultivated peanut as well as related *Arachis* species. All of the 251 working primer pairs with names, motifs, repeat types, primer sequences, and alleles tested in cultivated and wild species are listed in Additional File 1.

Background

Cultivated peanut (*Arachis hypogaea* L.) is grown on 25.5 million hectares with a total global production of about 35 million tons. It is an allotetraploid ($2n = 4 \times = 40$) and belongs to *Arachis* genus, which can be grouped into nine sections and includes approximately 80 species [1]. A large amount of morphological and agronomic variation is evident among accessions of cultivated peanuts, but extremely low levels of polymorphism were observed using restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD) and amplified fragment length polymorphisms (AFLP) [2-5]. Only simple-sequence repeats (SSRs) showed a potential for use in genetic studies of cultivated peanuts [6-11]. However it is expensive, labor-intensive and time-consuming to develop SSR markers from genomic DNA libraries [12]. To date, the number of available SSRs is grossly inadequate for mapping studies. Although several peanut genetic maps have been published [13-16], the existing maps do not have sufficient markers to be highly useful for genetic studies. Thus, there is great need for development of novel SSR markers.

Recently, EST-SSRs have received much attention as the increasing amounts of ESTs being deposited in databases for various plants [17-19]. EST-SSR can be rapidly developed from EST database by data mining at low cost, and due to their existence in transcribed region of genome, they can lead to the development of gene-based maps which may help to identify candidate function genes and increase the efficiency of marker-assisted selection [20]. In addition, EST-SSRs show a higher level of transferability to closely related species than genomic SSR markers [17,21] and can be served as anchor markers for comparative mapping and evolutionary studies [22]. Similar advantages of EST-SSRs have been reported for a number of plant species, such as grape [17], Medicago species [23], soybean [24], sugarcane [25], maize [18,19,24,26], rice [18,27-29], rye [29-31], and wheat [27,32,33], indicating that EST-SSR markers have potential for use in peanut genetic studies.

In peanut, only two studies described the development of EST-SSR in cultivated peanut and wild species [34,35]. Luo et al (2005) developed 44 EST-SSR markers from 1,350 cultivated peanut ESTs, nine of which exhibited polymorphism among 24 cultivated peanut lines. Proite et al (2007) developed 188 EST-SSRs from 8,785 *A. stenosperma* (*Arachis* species) ESTs, of which, 21 were polymorphic for an AA genome mapping population and 4 for a range of cultivated peanut genotypes. In this study, we screened a much larger number of ESTs (24, 238) from cultivated peanut with the following objectives: (1) to analyze the frequency and distribution of SSRs in transcribed regions of cultivated peanut genome; (2) to assess the validity of developed EST-SSR markers for detection of the polymorphism in cultivated peanut genotypes and their transferability to related wild species; (3) to develop new EST-SSR markers for both cultivated peanut and wild species.

Results

Type and frequency of peanut EST-SSRs

A total of 24,238 ESTs with an average length of 550 bp were used to evaluate the presence of SSR motifs. To eliminate redundant sequences and improve the sequence quality, the TIGR Gene Indices Clustering Tools (TGICL) [36] was employed to obtain consensus sequences from overlapping clusters of ESTs. A cluster was defined here as a group of overlapping EST sequences (at least 50 nucleotides with 90% identity and unmatched length less than 20 nucleotides). Totally, 11,431 potential unique ESTs including 1,434 contigs and 9,997 singletons were generated. As shown in Table 1, a total of 881 SSRs were identified from 780 unique ESTs, with an average of one SSR per 7.3 kb. Of those, 85 (about 10.9%) ESTs contained more than one SSR and 59 (about 7.6%) were compound SSRs that have more than one repeat type. Analysis of SSR motifs revealed that the proportion of SSR unit sizes was not evenly distributed. The occurrences of different repeat units were tri- (63.9%), di- (32.7%), tetra- (1.7%), penta- (0.7%), and hexa-nucleotide (1.0%). The mean SSR length of each unit varied between 18 and 37 bp. The

Table 1: Summary of SSR search after sequences assembled and categorized

	Contigs(bp)	Singlets(bp)	Total (bp)
EST after assembled	1434(12372129)	9997(5197116)	11431(6434245)
Identified SSRs	180	701	881
ESTs having SSRs	156	624	780
ESTs having more than 1SSR	19	66	85
Compound SSRs	16	43	59
Bi-type	73	215	288
Tri-type	98	465	563
Tetra-type	7	8	15
Penta-type	1	5	6
Hexa-type	1	8	9

overall average SSR length was 20 bp with a maximum of 86 bp di-nucleotide repeat (AG/CT). A total of 27 SSR motifs were listed in Table 2. The AG/CT was the most frequent motif and accounted for 27.7%, followed by AAG/TTC (17.37%), AAT/TTA (11.9%), ACC/TGG (7.7%), ACT/TGA (7.26%) and AT/TA (6.3%). The remaining motifs presented a frequency of 23.3%. GC-only repeat was not observed.

Primer design and validation

Among the 780 SSR-containing ESTs, 490 did not qualify for primer design as the flanking sequences were too short or poor quality. Therefore, only 290 primer pairs were designed and employed for validation of genic SSR markers (Table 3). Of these EST-SSRs, 65, 178 and 47 were observed in 5' untranslated terminal regions (UTR), translated regions and 3' UTR, respectively. After optimization, 251 primer pairs (86.5%) were successfully amplified in all cultivated peanut and wild species tested (Table 3), while the rest failed to give PCR products at various annealing temperature and Mg²⁺ concentrations. Out of 251 working primer pairs, 182 amplified the expected size

of amplicons, 41 yielded PCR products larger than expected, revealing that an intron is inside the amplicons, and the amplified products of the remaining 28 primer pairs were smaller than expected, suggesting the occurrence of deletion within the genomic sequences or a lack of specificity (Additional File 1).

EST-SSR polymorphism

In the present study, 251 valid EST-SSR primer pairs were used for assessment of the polymorphism among cultivated and wild *Arachis* species. Within cultivated peanuts, 26 (10.3%) EST-SSRs exhibited polymorphism (Table 3). A total of 55 alleles were detected and the average number of alleles per SSR marker was 2.1 with a range of 2–4 alleles based on the dominant scoring of the SSR bands characterized by the presence or absence of a particular band (Additional File 1). The PIC values ranged from 0.09 to 0.69 with an average value of 0.33. The greatest variation of SSR alleles was found for EM-78, which interacted with 4 alleles in 22 cultivated peanuts genotypes and the PIC value was 0.69.

Table 2: Occurrence and number of repeats of 27 SSR motifs in cultivated peanut (*Arachis hypogaea* L.)

Repeats	Number of repeat units										Total repeat
	5	6	7	8	9	10	11	12	13	Above	
AC/GT	-	-	6	2	3	2		1			14
AG/CT	-	-	56	43	31	11	13	15	2	47	218
AT/AT	-	-	15	11	5	3	2	2	1	17	56
AAC/GTT	22	7	1	2	5						37
AAG/CTT	71	44	13	10	7	4	2			2	153
AAT/ATT	54	26	8	3	3	2		1	3	5	105
ACC/GGT	31	22	9	5	1						68
ACG/CTG	10	4	2	1							17
ACT/ATG	35	17	9	1			2				64
AGC/CGT	17	8	1	2	1						29
AGG/CCT	19	8	1	1							29
AGT/ATC	25	11	4	2		1					43
CCG/CGG	13	3	1	1							18
AAAG/CTTT	3	3		1							7
AAAT/ATTT	2										2
AATC/AGTT	2	1									3
AATT/AATT	1										1
ACAT/ATGT	1	1									2
AAAAG/CTTTT	1										1
AAAAT/ATTTT	2										2
AGTAT/ATATC	3										3
AAAAAG/CTTTTT	1										1
AAGACG/CTGCTT	2										2
AATAGT/ATCATT				1							1
AATGAT/ACTATT			1	2							3
AGCAGT/ATCGTC	1										1
AGCTCC/AGGTCG	1										1
Total	317	155	127	88	56	23	19	19	6	71	881

Table 3: Characteristics of cultivated peanut (*Arachis hypogaea* L.) EST-SSR and efficiency of markers development

Motif	No. of EST-SSRs	No. of designed primers	No. amplified EST-SSRs (%)		No. polymorphic EST-SSRs (%)	
			Cultivated peanut	Wild species	Cultivated peanut	Wild species
Di	288	55	42	42	10	34
AC/GT	14	6	5	5	2	4
AG/CT	218	39	29	29	8	24
AT/AT	56	10	8	8	0	6
Tri	563	221	196	196	14	174
AAC/GTT	37	14	11	11	0	9
AAG/CTT	153	59	51	51	2	43
AAT/ATT	105	27	24	24	4	23
ACC/GGT	68	32	29	29	2	28
ACG/CTG	17	4	3	3	0	3
ACT/ATG	64	26	24	24	2	21
AGC/CGT	29	10	9	9	3	9
AGG/CCT	29	16	15	15	0	11
AGT/ATC	43	23	21	21	1	19
CCG/CGG	18	10	9	9	0	8
Tetra	15	5	5	5	1	5
AAAG/CTTT	7	1	1	1	0	1
AAAT/ATTT	2	2	2	2	1	2
AATC/AGTT	3	1	1	1	0	1
AATT/AATT	1	0	0	0	0	0
ACAT/ATGT	2	1	1	1	0	1
Penta-type	6	3	3	3	0	3
AAAAG/CTTTT	1	1	1	1	0	1
AAAAT/ATTTT	2	0	0	0	0	0
AGTAT/ATATC	3	2	2	2	0	2
Hexa-type	9	6	5	5	1	5
AAAAAG/CTTTTT	1	1	1	1	0	1
AAGACG/CTGCTT	2	1	1	1	1	1
AATAGT/ATCATT	2	2	1	1	0	1
AATGAT/ACTATT	3	1	1	1	0	1
AGCAGT/ATCGTC	1	0	0	0	0	0
AGTCC/AGGTCC	1	1	1	1	0	1
Total	881	290	251	251	26	221

The polymorphism of 251 cultivated peanut-derived EST-SSR in 16 accessions of wild species was evaluated. The results showed that 221 of 251 EST-SSR loci (88%) were polymorphic (Table 3), with a total of 867 alleles (Additional File 1). Allelic diversity was estimated for those polymorphic EST-SSR markers. The number of alleles detected among 16 wild species ranged from 2 to 9, with an average of 3.9 alleles per locus (Additional File 1). A maximum of 9 alleles were observed for primer EM-71. The PIC values varied between 0.594 and 0.820 with an average value of 0.721.

Sequence comparison of SSR bands

For further understanding of the EST-SSR polymorphism at the nucleotide level, the amplified products of primer EM-31 from two genotypes of cultivated peanuts and three accessions of wild species were cloned and sequenced (Figure 1, Figure 2). All the sequenced alleles

from both cultivars and wild species were highly identical to the original locus (EST sequence) from which the EST-SSR marker EM-31 was mined. Sequence alignment showed that all the primer-binding regions are highly conserved. Allelic diversity could be attributed mainly to differences in repeat type and length in the microsatellite regions, although some variations such as repeat number or insertions of additional motifs were observed in the microsatellite regions. In addition, a few single base substitutions were observed in the microsatellite flanking regions. Out of them, one occurred in *A. cardenasii*, one in *A. duranensis*, and two in *A. pintoi*.

Discussion

Frequency and distribution of EST-SSRs

The frequency of SSRs in SSR-ESTs more accurately reflects the density of SSRs in the transcribed region of the genome. However, random sequencing within cDNA

various plant species, tri-nucleotide was also observed to be most frequent [26], regardless of the EST-SSR search criteria. Until now, only one report described that di-nucleotide repeats were most abundant followed by tri- or mono-nucleotide repeats in dicots [38]. In the present investigation, tri-nucleotide repeat was found to be abundant followed by di-nucleotide. In term of single SSR motif, the di-nucleotide motif (AG/TC)_n was highest frequent [18,39]. Among the di-nucleotide motifs, the two most dominant motif types were AG and AT, representing an average frequency of 24.7% and 6.4%, respectively. This was in agreement with recent studies in cultivated peanut (*Arachis hypogaea* L.) [35] and wild *Arachis* species [34]. In this work, the AAG with 17.4% of frequency following di-nucleotide motif AG was the most abundant in the ten tri-nucleotide motifs. In other plant species, the most frequent tri-nucleotide repeat motifs were (AAC/TTG)_n in wheat, (AGG/TCC)_n in rice, (CCG/GGC)_n in maize, (AAG/TTC)_n in soybean, and (CCG/GGC)_n in barley and sorghum [18,19,39,40]. The previous studies of *Arabidopsis* [37] and soybean [24] also suggested that the tri-nucleotide AAG motif may be common motif in dicots. In contrast, the abundance of the tri-nucleotide CCG repeat motif was favored overwhelmingly in cereal species [18,19,32] and also considered as a specific feature of monocot genome, which may be due to increasing the G + C content [26].

Validation and polymorphism of EST-SSR markers

In this study, a total of 290 designed primer pairs were used for validation of the EST-SSR markers. Of these, 251 (86.5%) yielded amplicons in both cultivated peanut and wild species. This result was similar to previous studies in which a success rate of 60–90% amplification has been reported [21,25,40–42]. In those studies, they also reported a similar success rate of amplification for both genomic SSRs and EST-SSRs. However, EST-SSRs were reported to be less polymorphic than genomic SSRs in crop plants due to greater DNA sequence conservation in transcribed regions [17,28,43–46]. Previous studies highlighted the fact that EST-SSR markers have higher transferability and better applicability than genomic SSR markers [17,47–49]. In addition to high transferability, EST-SSRs were good candidates for the development of conserved orthologous markers for genetic analysis and breeding of different species [22]. Previous reports showed that the transferability of EST-SSRs from one species to another ranged from 40–89% [21,23,24,27,29,40,41,50,51]. Our results indicated that 100% of EST-SSR amplifiable primers for cultivated peanut can produce amplicons in *Arachis* wild species.

In the present investigation, the mean percentage of polymorphic loci of EST-SSR markers was 9.96% in cultivated

peanuts. This value was lower than those of genomic SSR found in earlier studies [7,12,47], but higher than the percentage of polymorphic loci in cultivated peanut observed using RAPD (6.6%) [5] and AFLP (6.7%) [4]. No major difference was observed in terms of allele numbers and PIC values for the EST-SSR markers among the cultivated genotypes, while significant difference was observed among wild species. Therefore, the low level of EST-SSR polymorphism detected in cultivated peanuts may be compensated by their higher potential for cross-species transferability to wild species. In the present study, 100% transferability of EST-SSR with 86.6% polymorphism from cultivated peanut to wild *Arachis* species was observed. The value is higher than that of genomic SSR cross-transferability [10]. The high level of transferability indicated that these markers would be highly effective for molecular study of *Arachis* species. Since current molecular markers display a low level of genetic polymorphism in cultivated peanuts [2–4,6,52,53], it is difficult to construct a high-density genetic linkage map for cultivated peanut which could be used in breeding programs. However, a genetic map constructed using wild species together with transferable molecular markers derived from cultivated peanuts would contribute to understanding the introgression of genes from wild species to cultivated peanuts [10,13]. Therefore, the development of a set of transferable EST-SSR markers from cultivated peanuts will be a great benefit to construct a high-density genetic map of wild species. The map would allow the identification of markers, especially transferable EST-SSR markers, associated with resistance or other agronomic traits in wild species, and in turn, help to discover corresponding markers or genes in cultivated peanuts.

Additionally, a comparison of sequences of cross-species amplicons generated by primer EM-31 further confirmed the conservation and transferability of the developed EST-SSR loci. In general, the amplified regions were found to be similar to the original peanut EST sequences from which the SSRs were developed and their comparisons across species (Figure 2) correlated the observed 'cross-species alleles' precisely with the expected length variations. Furthermore, in addition to the variation of the number of SSR repeat, the allele sequences also indicated that a few additional point mutation in the SSR motifs flanking regions. Similar variation has been reported in earlier studies [39,47,54,55]. This phenomenon is supposed to be the innate evolving nature of the genome, and thus can be indicative of the evolutionary relationships of the tested taxa [47].

Conclusion

EST-SSR markers developed in this study will complement the genomic SSR markers and provide a valuable resource

for linkage mapping, gene and QTL identification, and marker-assisted selection in peanut genetic study. Since these markers were developed based on expressed sequence and they are conserved across *Arachis* genus, they may be valuable for comparative genome mapping and functional analysis of candidate genes. In addition, these markers may be potentially useful for study of pods traits because majority of these EST sequences were derived from pods at three developmental stages.

Methods

Plant materials and DNA extraction

In the present study, twenty-two accessions of cultivated peanut (*A. hypogaea* L.) corresponding to two subspecies (*hypogaea* and *fastigiata*) and sixteen accessions of wild species from seven sections of the genus *Arachis* were used (Additional File 2). The leaf samples of each accession were collected from Peanut Germplasm Bank located in Crops Research Institute, Guangdong Academy of Agriculture, Guangzhou, China. The genomic DNA was extracted as described by Sharma [56].

Data mining for SSR marker

A total of 24,238 EST sequences including 20,160 developed by Guo et al (2008)[57] and 4078 retrieved from National Center of Biotechnology Information (NCBI) were used in this study. These ESTs were assembled using the TGICL program [36]. A Perl script known as MicroSATellite (MISA <http://pgrc.ipk-gatersleben.de/misa/>) was used to mine microsatellites. In this work, SSRs were considered for primer design that fitted the following criteria: a minimum length of 14 bp, excluding polyA and polyT repeat, at least 7 repeat units in case of di-nucleotide and at least 5 repeat units for tri-, tetra-, penta- and hexa-nucleotide SSRs. Therefore, the paired numbers representing SSR motif length and the minimum repeat number in the MISA configuration file (misa.ini) were modified to 2-7, 3-5, 4-5, 5-5 and 6-5 (mono-type excluded).

Primer design and PCR amplification

Using Primer Premier 5 program (Whitehead Institute for Biomedical Research, Cambridge, Mass), primers were designed based on the following core criteria: (1) melting temperature (T_m) between 52°C and 63°C with 60°C as optimum; (2) product size ranging from 100 bp to 350 bp; (3) primer length ranging from 18 bp to 24 bp with amplification rate larger than 80%; (4) GC% content between 40% and 60%. The parameters were modified when unsuitable primer pairs were retrieved by the program. PCR analysis was performed in a total volume of 20 μ l with the following cycling profile: 1 cycle of 5 min at 94°C, an annealing temperature of 55°C for 35 cycles (1 min at 94°C, 30 s at 55°C, 45 s at 72°C) and an additional cycle of 10 min at 72°C. Each of the primer pairs was screened twice to confirm the repeatability of the

observed bands in each genotype. PCR products were separated on 6% polyacrylamide denaturing gels. The gels were silver stained for SSR bands detection.

Sequencing of PCR bands

The SSR alleles amplified in two cultivars and three wild species for EM-31 primer were individually cloned and sequenced. PCR amplification products were separated by 6% polyacrylamide gel and target allele bands were excised and dipped in 10 μ l of nuclease free water for 30 min. Another round of PCR was made following the same protocol with recycled DNA as template. The second-round PCR products were separated in a 2% agarose gel and the target band was purified using TIANGEN Gel Extracting Kit (TIANGEN Inc. Beijing China). The purified PCR fragment from agarose gel was cloned using the Takara TA cloning kit pMD-18 (Takara, Dalian, China). The ligation product was transformed into competent *Escherichia coli* cells. The positive clones identified by PCR were sequenced by Invitrogen Company. The final edited sequences belonging to different genotypes were compared with the original SSR containing EST sequence using Omega program [58], and the exported multiple sequence alignment was modified by Genedoc <http://www.nrbsc.org/gfx/genedoc/index.html>.

Data scoring and statistical analysis

The allelic and genotypic frequencies were calculated for the samples analyzed. The genetic diversity of the samples as a whole was estimated based on the number of alleles per locus (total number of alleles/number of loci), the percentage of polymorphic loci (number of polymorphic loci/total number of loci analyzed) and polymorphism information content (PIC). The polymorphism was determined according to the presence or absence of the SSR locus. The value of PIC was calculated using the formula

$$PIC = 1 - \sum_{i=1}^n P_i^2$$

where P_i is the frequency of an individual genotype generated by a given EST-SSR primer pair and summation extends over n alleles.

Authors' contributions

All authors read and approved the final manuscript. XL participated in conceiving the study and drafting the manuscript. XC participated in conceiving the study, sequence analysis and drafting the manuscript. YH participated in conceiving the study, the development of SSR markers and data analysis. HL developed the SSRs and designed the SSR primers. GZ and SL planted and collected the peanut materials. BG participated in the development of SSR markers.

Additional material

Additional File 1

List of EST-SSR primers developed from cultivated peanut ESTs. The file contains a table that lists primer names, repeat motifs, primer sequences, allele number and product length for the newly developed EST-SSR markers.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2229-9-35-S1.xls>]

Additional File 2

List of cultivated peanut and wild species materials used in this study. The file includes a table that lists the name, type, ploidy and origin of 22 genotypes of cultivated peanuts and 16 accessions of wild species tested in this study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2229-9-35-S2.xls>]

Acknowledgements

This research was funded by a grant from National High Technology Research Development Project (863) of China (No 2006AA0Z156, 2006AA10A115) and Science Foundation of Guangdong province (No 07117967).

References

- Krapovickas A, Gregory WC: **Taxonomia del genero *Arachis* (Leguminosae).** *Bonplandia* 1994, **8**:1-186.
- Halward TM, Stalker HT, Larue EA, Kochert G: **Genetic variation detectable with molecular markers among unadapted germplasm resources of cultivated peanut and related wild species.** *Genome* 1991, **34**:1013-1020.
- Kochert G, Halward T, Branch WD, Simpson CE: **RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species.** *Theor Appl Genet* 1991, **81**(5):565-570.
- He GH, Prakash CS: **Identification of polymorphic DNA markers in cultivated peanut (*Arachis hypogaea* L.).** *Euphytical* 1997, **97**(2):143-149.
- Subramanian V, Gurtu S, Nageswara Rao RC, Nigam SN: **Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay.** *Genome* 2000, **43**(4):656-660.
- Hopkins MS, Casa AM, Wang T, Mitchell SE, Dean RE, Kochert GD, Kresovich S: **Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut.** *Crop Sci* 1999, **39**(4):1243-1247.
- Ferguson ME, Burow MD, Schulze SR, Bramel PJ, Paterson AH, Kresovich S, Mitchell S: **Microsatellite identification and characterization in peanut (*A. hypogaea* L.).** *Theor Appl Genet* 2004, **108**(6):1064-1070.
- Moretzsohn Mde C, Hopkins MS, Mitchell SE, Kresovich S, Valls JF, Ferreira ME: **Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome.** *BMC Plant Biol* 2004, **4**:11.
- Tang R, Gao G, He L, Han Z, Shan S, Zhong R, Zhou C, Jiang J, Li Y, Zhuang W: **Genetic Diversity in Cultivated Groundnut Based on SSR Markers.** *J Genet Genomics* 2007, **34**(5):449-459.
- Gimenes MA, Hoshino AA, Barbosa AV, Palmieri DA, Lopes CR: **Characterization and transferability of microsatellite markers of the cultivated peanut (*Arachis hypogaea* L.).** *BMC Plant Biol* 2007, **7**:9.
- He G, Meng RH, Gao H, Guo B, Gao G, Newman M, Pittman RN, Prakash CS: **Simple sequence repeat markers for botanical varieties of cultivated peanut (*Arachis hypogaea* L.).** *Euphytical* 2005, **142**(1):131-136.
- He G, Meng RH, Newman M, Gao GQ, Pittman RN, Prakash CS: **Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.).** *BMC Plant Biol* 2003, **3**:3.
- Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SC, Gimenes MA, Martins WS, Valls JF, Grattapaglia D, Bertioli DJ: **A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae).** *Theor Appl Genet* 2005, **111**(6):1060-1071.
- Halward T, Stalker HT, Kochert G: **Development of an RFLP linkage map in diploid peanut species.** *Theor Appl Genet* 1993, **87**(3):379-384.
- Burow MD, Simpson CE, Starr JL, Paterson AH: **Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.) broadening the gene pool of a monophyletic polyploid species.** *Genetics* 2001, **159**(2):823-837.
- Varshney RK, Bertioli DJ, Moretzsohn MC, Vadez V, Krishnamurthy L, Aruna R, Nigam SN, Moss BJ, Seetha K, Ravi K, et al: **The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.).** *Theor Appl Genet* 2009, **118**(4):729-739.
- Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ: **Analysis of SSRs derived from grape ESTs.** *Theor Appl Genet* 2000, **100**(5):723-726.
- Kantety RV, La Rota M, Matthews DE, Sorrells ME: **Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat.** *Plant Mol Biol* 2002, **48**(5-6):501-510.
- Varshney RK, Thiel T, Stein N, Langridge P, Graner A: **In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species.** *Cell Mol Biol Lett* 2002, **7**(2A):537-546.
- Gupta PK, Rustgi S: **Molecular markers from the transcribed/expressed region of the genome in higher plants.** *Funct Integr Genomics* 2004, **4**(3):139-162.
- Saha MC, Mian MA, Eujayl I, Zwonitzer JC, Wang L, May GD: **Tall fescue EST-SSR markers with transferability across several grass species.** *Theor Appl Genet* 2004, **109**(4):783-791.
- Varshney RK, Graner A, Sorrells ME: **Genic microsatellite markers in plants: features and applications.** *Trends Biotechnol* 2005, **23**(1):48-55.
- Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MA: ***Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp.** *Theor Appl Genet* 2004, **108**(3):414-422.
- Gao L, Tang J, Li H, Jia J: **Analysis of microsatellites in major crops assessed by computational and experimental approaches.** *Mol Breed* 2003, **12**(3):245-261.
- Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ: **Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum.** *Plant Sci* 2001, **160**(6):1115-1123.
- Morgante M, Hanafey M, Powell W: **Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes.** *Nat Genet* 2002, **30**(2):194-200.
- Yu JK, La Rota M, Kantety RV, Sorrells ME: **EST derived SSR markers for comparative mapping in wheat and rice.** *Mol Genet Genomics* 2004, **271**(6):742-751.
- Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N: **Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.).** *Theor Appl Genet* 2000, **100**(5):713-722.
- Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A: **Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice.** *Plant Sci* 2005, **168**(1):195-202.
- Dracatos PM, Dumsday JL, Olle RS, Cogan NO, Dobrowolski MP, Fujimori M, Roderick H, Stewart AV, Smith KF, Forster JW: **Development and characterization of EST-SSR markers for the crown rust pathogen of ryegrass (*Puccinia coronata* f.sp. *lolii*).** *Genome* 2006, **49**(6):572-583.
- Kuleung C, Baenziger PS, Dweikat I: **Transferability of SSR markers among wheat, rye, and triticale.** *Theor Appl Genet* 2004, **108**(6):1147-1150.
- Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ: **One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat.** *Theor Appl Genet* 2004, **108**(7):1392-1400.

33. Gadaleta A, Mangini G, Mulè G, Blanco A: **Characterization of dinucleotide and trinucleotide EST-derived microsatellites in the wheat genome.** *Euphytica* 2007, **153**(1-2):73-85.
34. Proite K, Leal-Bertioli SC, Bertioli DJ, Moretzsohn MC, da Silva FR, Martins NF, Guimaraes PM: **ESTs from a wild *Arachis* species for gene discovery and marker development.** *BMC Plant Biol* 2007, **7**:7.
35. Luo M, Dang P, Guo B, He G, Holbrook CC, Bausher MG, Lee RD: **Generation of Expressed Sequence Tags (ESTs) for Gene Discovery and Marker Development in Cultivated Peanut.** *Crop Sci* 2005, **45**:346-353.
36. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, et al.: **TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets.** *Bioinformatics* 2003, **19**(5):651-652.
37. Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R: **Computational and experimental characterization of physically clustered simple sequence repeats in plants.** *Genetics* 2000, **156**(2):847-854.
38. Kumpatla SP, Mukhopadhyay S: **Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species.** *Genome* 2005, **48**(6):985-998.
39. Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A: **Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants.** *Mol Biol Evol* 1998, **15**(10):1275-1287.
40. Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.).** *Theor Appl Genet* 2003, **106**(3):411-422.
41. Yu JK, Dake TM, Singh S, Benschler D, Li W, Gill B, Sorrells ME: **Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat.** *Genome* 2004, **47**(5):805-818.
42. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS: **Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat.** *Mol Genet Genomics* 2003, **270**(4):315-323.
43. Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K: **Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags.** *Theor Appl Genet* 2004, **109**(6):1283-1294.
44. Eujayl I, Sorrells M, Baum M, Wolters P, Powell W: **Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs.** *Euphytica* 2001, **119**:39-43.
45. Chabane K, Ablett GA, Cordeiro GM, Valkoun J, Henry RJ: **EST versus Genomic Derived Microsatellite Markers for Genotyping Wild and Cultivated Barley.** *Genet Resour Crop Evol* 2005, **5**(7):903-909.
46. Russell J, Booth A, Fuller J, Harrower B, Hedley P, Machray G, Powell W: **A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome.** *Genome* 2004, **47**(2):389-398.
47. Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L: **Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species.** *Theor Appl Genet* 2007, **114**(2):359-372.
48. Wanga ML, Barkley NA, Yua JK, Deana RE, Newman ML, Sorrells ME, Pederson GA: **Transfer of simple sequence repeat (SSR) markers from major cereal crops to minor grass species for germplasm characterization and evaluation.** *Plant Genet Res* 2005, **3**:45-57.
49. Guo W, Wang W, Zhou B, Zhang T: **Cross-species transferability of *G. arboreum*-derived EST-SSRs in the diploid species of *Gossypium*.** *Theor Appl Genet* 2006, **112**(8):1573-1581.
50. Holton TA, Christopher JT, McClure L, Harker N, Henry RJ: **Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat.** *Mol Breed* 2002, **9**(2):63-71.
51. Saha MC, Mian R, Zwonitzer JC, Chekhovskiy K, Hopkins AA: **An SSR- and AFLP-based genetic linkage map of tall fescue (*Festuca arundinacea* Schreb.).** *Theor Appl Genet* 2005, **110**(2):323-336.
52. Halward TM, Stalker HT, Larue EA, Kochert G: **Use of single primer DNA amplification in genetic studies of peanut.** *Plant Mol Biol* 1992, **84**:201-208.
53. Paik-Ro OG, Smith RL, Knauff DA: **Restriction fragment length polymorphism evaluation of six peanut species within the *Arachis* section.** *Theor Appl Genet* 1992, **84**(1-2):201-208.
54. Shepherd LD, Lambert DM: **Mutational bias in penguin microsatellite DNA.** *J Hered* 2005, **96**(5):566-571.
55. Sethy NK, Choudhary S, Shokeen B, Bhatia S: **Identification of microsatellite markers from *Cicer reticulatum*: molecular variation and phylogenetic analysis.** *Theor Appl Genet* 2006, **112**(2):347-357.
56. Sharma KK, Lavanya M, Anjaiah V: **A Method for Isolation and Purification of Peanut Genomic DNA Suitable for Analytical Applications.** *Plant Mol Biol Repr* 2000, **18**:393a-393h.
57. Guo B, Chen X, Dang P, Scully BT, Liang X, Holbrook CC, Yu J, Culbreath AK: **Peanut gene expression profiling in developing seeds at different reproduction stages during *Aspergillus parasiticus* infection.** *BMC Dev Biol* 2008, **8**:12.
58. Kramer JA: **Omiga: a PC-based sequence analysis tool.** *Mol Biotechnol* 2001, **19**(1):97-106.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

