ORIGINAL PAPER

**M. E. Ferguson · M. D. Burow · S. R. Schulze ·
P. J. Bramel · A. H. Paterson · S. Kresovich ·
S. Mitchell**

Microsatellite identification and characterization in peanut (*A. hypogaea* L.)

Abstract A major constraint to the application of biotechnology to the improvement of the allotetraploid peanut, or groundnut (*Arachis hypogaea* L.), has been the paucity of polymorphism among germplasm lines using biochemical (seed proteins, isozymes) and DNA markers (RFLPs and RAPDs). Six sequence-tagged microsatellite (STMS) markers were previously available that revealed polymorphism in cultivated peanut. Here, we identify and characterize 110 STMS markers that reveal genetic variation in a diverse array of 24 peanut landraces. The simple-sequence repeats (SSRs) were identified with a probe of two 27,648-clone genomic libraries: one constructed using *Pst*I and the other using *Sau*3AI/*Bam*HI. The most frequent, repeat motifs identified were ATT and GA, which represented 29% and 28%, respectively, of all SSRs identified. These were followed by AT, CTT, and GT. Of the amplifiable primers, 81% of ATT and 70.8% of GA repeats were polymorphic in the cultivated peanut test array. The repeat motif AT showed the maximum number of alleles per locus (5.7). Motifs ATT, GT, and GA had a mean number of alleles per locus of 4.8, 3.8, and 3.6, respectively. The high mean number of alleles per polymorphic locus, combined with their relative frequency in the genome and amenability to probing, make ATT and GA the most useful and appropriate motifs to target to generate further SSR markers for peanut.

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-003-1535-2.

M. E. Ferguson (✉)
International Crops Research Institute
for the Semi-Arid Tropics (ICRISAT),
Patancheru, 502324, Andhra Pradesh, India
e-mail: m.ferguson@cgiar.org
Tel.: +254-2-631499

M. D. Burow
Department of Crop and Soil Science,
University of Georgia,
Athens, GA 30602, USA

S. R. Schulze · A. H. Paterson
Departments of Crop and Soil Science, Botany and Genetics,
Center for Applied Genetic Technologies,
University of Georgia,
Athens, GA 30602, USA

P. J. Bramel
KISR, c/o Dr. John Peacock, P.O. Box 24885,
13109 Safat, Kuwait

S. Kresovich · S. Mitchell
Institute for Genomic Diversity,
Cornell University,
Ithaca, NY 14853, USA

Present address:
M. D. Burow, Texas Agricultural Experiment Station,
Texas A&M University,
Rte. 3, Box 219, Lubbock, TX 79403, USA

Present address:
M. E. Ferguson, International Institute for Tropical Agriculture
(IITA), c/o ILRI, P.O. Box 30709,
Nairobi, Kenya

Introduction

The application of biotechnology to the improvement of the allotetraploid ($2n=4x=40$) peanut, or groundnut (*Arachis hypogaea* L.), has been hampered by an inability to visualize genetic variation in germplasm lines. Grieshammer and Wynne (1990) found that only 3 of 25 isozymes detected polymorphism among 68 *A. hypogaea* genotypes; each of these three isoenzymes had only two forms. Lacks and Stalker (1993) found similar results in 33 *A. hypogaea* accessions from the center of origin in South America. Lanham et al. (1994) found just two polymorphic proteins in 28 *A. hypogaea* accessions. Limited genetic variation has also been reported from evaluations of seed proteins (Tombs

1963; Bianchi-Hall et al. 1991) and total proteins (Savoy 1976).

Halward et al. (1991, 1992) found little or no variation using RFLP and RAPD markers in more than 25 unadapted germplasm lines that represented four of the six botanical varieties of peanut from five of the six centers of genetic diversity in South America and several introductions from Africa and China. Kochert et al. (1991) reported little or no polymorphism among eight peanut cultivars from the United States using RFLPs. Paik-Ro et al. (1992) reported that amongst the 7–16 accessions of botanical varieties of subspecies *hypogaea* and *fastigiata* studied, none showed polymorphism with 32 endonuclease and genomic probe combinations. Due to this lack of identifiable polymorphism in the species, the only tetraploid genetic linkage map currently available is based on an interspecific cross (Burow et al. 2001).

In peanut, Hopkins et al. (1999) designed 26 primers from a size-fractioned (200–500 bp) *Taq*I genomic library probed for $(GT)_{10}$ and $(CT)_{10}$ oligonucleotides and found that five pairs of primers, plus one designed from a search of publicly available DNA sequences, gave polymorphism in a diverse array of 22 *A. hypogaea* germplasm lines. More than 70 primer pairs from a range of different genomic libraries were later screened, which did not reveal any further polymorphic loci (S. Kresovich, personal communication). Thus, the specific objective of this study was to locate polymorphic SSRs in cultivated peanut germplasm, and to characterize the SSRs by motif and polymorphism to enhance the efficiency of future STMS development in peanut. The strategy was to probe two relatively large genomic libraries with up to 12 oligonucleotide SSR probes using high-throughput robotics and sequencing.

Materials and methods

DNA was extracted from Florunner (Burow et al. 2001), and after single colony purification, the vector pBluescript (Stratagene) was isolated from *Eschericia coli* using the alkali-detergent method of Sambrook et al. (1989). Both peanut DNA and the vector DNA were cesium chloride purified (Sambrook et al. 1989). For the first library, 5 μg peanut DNA was digested with *Pst*1 and size fractioned on a 1.5% agarose gel into the following molecular weight (MW) ranges: 300 bp–600 bp [L(ow)MW], 600 bp–1 kb [M(id)MW], and 1 kb–1.3 kb [H(igh)MW]. The vector pBluescript was digested with *Pst*I, dephosphorylated using alkaline phosphatase (CIAP) (Stratagene), and gel purified. Peanut DNA fragments were ligated into the vector and then, after drop dialysis for desalting, transformed into DH-10B cells by electroporation at 1.9 kV/cm, and blue-white selected for those containing recombinant plasmids. A Q-bot (Genetix) high-speed, multi-task robot was used to pick 27,648 white colonies. The library was gridded onto two sets of Hybond Nylon membranes in a 2×2 pattern, to be probed for SSR elements. The library consisted of 15,360 HMW clones (56%), 7,296 MMW clones (26%), and 4,992 LMW clones (18%). A *Sau*3 AI/*Bam*HI library was also constructed using the same methodology. This library consisted of 3,072 HMW (11%), 14,208 MMW (51%), and 10,368 (38%) LMW fragments.

Both libraries were probed using γ-[^{32}P]-dATP-labeled oligonucleotide probes. The filters were exposed to X-ray film for 8 days and developed. Both libraries were probed with $(GT)_{15}$, $(GA)_{15}$, $(AAC)_{10}$, $(ATC)_{10}$, $(AGT)_{10}$, $(ATT)_{10}$, $(CAC)_{10}$, $(CTT)_{10}$, and $(CTG)_{10}$, with a separate hybridization reaction for each probe. In addition, the *Pst*I library was probed with $(ATCC)_{10}$, $(GATA)_{10}$, and $(AAAT)_{10}$. The intensity of hybridization was scored as high, medium, or low. Filters were also screened with genomic DNA to eliminate those clones that contained repetitive DNA.

Plasmids were isolated from positive clones using an alkaline lysis method with modifications made for the 96-well plate format (Marra et al. 1997). Cycle-sequencing reactions were performed using the BigDye Terminator Cycle Sequencing Kit Version 2 (Applied Biosystems, Foster City, Calif.) and MJ Research (Watertown, Pa.) PTC-100 thermocycler. Finished cycle-sequencing reactions were filtered through Sephadex filter plates (Krakowski et al. 1995) directly into Perkin-Elmer MicroAmp Optical 96-well reaction plates, and sequencing was performed using an ABI3700 automated DNA Analyzer (Applied Biosystems, Foster City, Calif.). All sequences were submitted to GenBank.

Sequences were searched for all possible repeat motifs using a program developed by Sam Cartinhour (USDA-ARS Center for Agricultural Bioinformatics). Di-, tri-, and longer motifs (tetra-, penta-, and hexa-nucleotides) containing $\geq$ six, five and four repeat units, respectively, were identified. Optimal primer sequences for amplifying each SSR locus were obtained with Primer 0.5 (Daly et al. 1991) (http://www-genome.wi.mit.edu/ftp/pub/software/primer0.5), with a minimum product size of 100 bp, CG content of between 40% and 60% T 60–65°C, ΔT 1–2°C, and primer length of 18–25 bp. Possible hairpins, dimers, and cross dimers were identified on each primer combination using Netprimer (http://www.premierbiosoft.com/netprimer/netprimer.html).

Germplasm

A diverse set of 24 peanut accessions consisting of two accessions each from South America, Africa, and Asia of subspecies *fastigiata* Waldron var. *fastigiata* and *vulgaris* C. Harz., and *hypogaea* var. *hypogaea*, plus two each of the remaining three botanical varieties, were selected as a test set to determine primer polymorphism within *A. hypogaea* (Table 1). DNA was extracted from young, folded leaflets using Qiagen miniprep kits (Qiagen, Valencia, Calif.).

Optimization of amplification

Primers were optimized using a modified Taguchi method (Cobb and Clarkson 1994; H. Buhariwallah, personal communication) that is designed to reveal the effects and interactions of specific reaction components simultaneously, using a few reactions. In the Taguchi method, reaction components that are likely to affect the PCR process are arranged in an orthogonal array. Each component occurs at one of three predetermined levels (A, B, and C), each of which occurs an equal number of times within the orthogonal array (Table 2). Here, primer concentrations of 10 pM (A), 20 pM (B), and 30 pM (C), DNA amounts of 5 ng (A), 10 ng (B), and 15 ng (C); Mg^{2+} concentrations of 2 mM (A), 3 mM (B), and 4 mM (C); dNTP concentrations of 0.15 mM (A), 0.2 mM (B), and 0.25 mM (C); and *Taq* polymerase of 0.8 U (A), 1 U (B), and 1.2 U (C) were used. This resulted in nine different reaction combinations. Reactions were carried out in a final volume of 20 µl with 1× reaction buffer. Where necessary, Q-solution (Qiagen, Valencia, Calif.) was added to improve specificity. Reaction conditions were 94°C for 2 min, 35 cycles of 94°C for 45 s, annealing temperature (between 58°C and 64°C) for 1 min, 72°C for 90 s, and then a final extension of 10 min at 72°C. Protocol optimization was undertaken at 60°C. Empirical annealing temperatures were determined using a gradient thermocycler between 55°C and 65°C. In some cases "hot-start PCR" was used to reduce primer-dimer formation.

Table 1 Accessions used to characterize polymorphism

FAO-designated ICRISAT GeneBank accession number (ICG)	Botanical variety	Country of origin	Market type
10362	hypogaea	Nigeria	Virginia, runner or bunch
4389	hypogaea	India	Virginia, runner or bunch
12621	hypogaea	India	Virginia, runner or bunch
10971	hypogaea	Peru	Virginia, runner or bunch
13420	hypogaea	Chad	Virginia, runner or bunch
12235	hypogaea	Bolivia	Virginia, runner or bunch
20002	hirsuta	Peru	Peruvian runner
20016	hirsuta	Ecuador	Peruvian runner
3204	fastigiata	China	Valencia
10704	fastigiata	China	Valencia
14421	fastigiata	Nigeria	Valencia
9987	fastigiata	Bolivia	Valencia
13430	fastigiata	Chad	Valencia
11605	fastigiata	Bolivia	Valencia
10384	vulgaris	Nigeria	Spanish
11505	vulgaris	China	Spanish
12483	vulgaris	Peru	Spanish
11515	vulgaris	China	Spanish
13415	vulgaris	Chad	Spanish
11175	vulgaris	Bolivia	Spanish
10074	peruviana	Peru	
10911	peruviana	Peru	
7898	equatoriana	Ecuador	
12553	equatoriana	Ecuador	

Table 2 Orthogonal arrays used for three levels of reaction components for five different variables

Reaction number	Primer[a]	DNA[b]	Mg^{2+}[c]	dNTP[d]	Taq polymerase[e]
1	A	A	A	A	A
2	A	B	B	B	B
3	A	C	C	C	C
4	B	A	B	C	C
5	B	B	C	A	B
6	B	C	A	B	A
7	C	A	C	B	B
8	C	B	A	C	C
9	C	C	B	A	A

[a] Primer concentration: A = 10 pM, B = 20 pM, C = 30 pM
[b] DNA concentration: A = 5 ng, B = 10 ng, C = 15 ng
[c] Mg^{++} concentration: A = 2 mM, B = 3 mM, C = 4 mM
[d] dNTP concentration: A = 0.15 mM, B = 0.2 mM, C = 0.25 mM
[e] Taq polymerase concentration: A = 0.8 U, B = 1 U, C = 1.2 U

Visualization of amplification products

All amplification products were visualized on non-denaturing 9% 29:1 (w/w) polyacrylamide/bisacrylamide gels followed by silver staining. Silver staining consisted of 3 min in water, 20 min in 0.1% CTAB, 15 min in 0.3% ammonia solution, 15 min in a solution of 1 M NaOH, 0.1% silver nitrate and a few drops of a 25% ammonia solution, and a rinse in water (5–10 s.), followed by development in a 1.5% sodium carbonate solution with 0.02% by volume formaldehyde solution. Gels were rinsed in water and fixed in a 1.5% glycerol solution.

Primer evaluation

A quality score was given to each primer pair according to the extent of amplification and the ease of scoring. Such quality scores have been used previously (Pepin et al. 1995; Smulders et al. 1997); however, these have tended to be based on the appearance of stutter bands. We did not observe stuttering, so the following system of quality assessment was used: 1 = unambiguous scoring; 2 = alleles close, but scoring possible; 3 = alleles too close for accurate scoring when run by polyacrylamide gel electrophoresis (PAGE); 4 = weak amplification in at least one locus; 5 = variation due to absence of band; and 6 = variation in a fainter, secondary locus, but of expected size range. Polymorphism information content (PIC) values were calculated according to the formula $1 - \Sigma p_i^2$, where p_i is the frequency of the ith allele in the 24 genotypes tested (Anderson et al. 1993). A scatter plot of the number of repeat units against number of alleles per locus was produced for perfect repeats of the six major, repeat families.

Results

Two 27,648-clone libraries were used to detect SSRs. If it is assumed that the average insert sizes were 450 bp, 800 bp, and 1,150 bp for the three selected size classes, the PstI library would be expected to contain 25,747.3 kb and the Sau3AI/BamHI would contain 24,230.4 kb, which corresponded to 0.86% and 0.8% of the peanut genome, respectively ($1C = 3 \times 10^9$ bp). When contiguous sequences (sequences that align to form contiguous blocks) are

excluded, the expected coverage of the genome is reduced to approximately 0.6% and 0.74%, respectively.

Of the 589 positive clones detected in the *Pst*I library and 1,122 clones from the *Sau*3AI/*Bam*HI library, 1.9% and 6.5% were found to be from repetitive sequences, respectively, and were eliminated from the analysis. A total of 578 clones were sequenced from the *Pst*I library and 1,049 clones from the *Sau*3AI/*Bam*HI library. All sequences can be found at GenBank (accessions: BZ999351–CC000573), through http://www. plantgenome.uga.edu, http://www.icrisat.org or http:// link.springer.com. Poor quality sequences represented 11.2% and 11.3% of the clones sequenced in the *Pst*I and *Sau*3AI/*Bam*HI libraries, respectively. Of the remainder, 30% of the *Pst*I library and 7% of the *Sau*3AI/*Bam*HI library were not unique fragments and were considered as contiguous sequences or "contigs." One hundred twelve clones contained 132 SSRs from the *Pst*I library, and 183 clones contained 216 SSRs from the *Sau*3AI/*Bam*HI library. This represented 31% and 21% of clones with good quality sequence data, excluding contigs, from the *Pst*I and *Sau*3AI/*Bam*HI libraries, respectively. These figures may under-represent the number of SSRs present, as many of the clones had large insert sizes that prevented full sequencing of the inserts with sequencing from one direction only. Information on the SSRs identified can be found through the above-mentioned Web sites. This includes repeat motifs and number of repeats.

Ninety-two and 134 primer pairs were designed from the 132 SSRs and 216 SSRs from the *Pst*1 and *Sau*3AI/ *Bam*HI libraries, respectively. In the remaining clones, the repeat was either too close to the vector, or sequence information was not sufficiently reliable. Primer information including forward- and reverse-sequence information, repeat family, fragment length, annealing temperatures (*T*), empirical annealing temperatures, amplification, and PCR conditions can be found on the associated Web sites provided above. From the *Pst*I library, 64% of the primers designed came from HMW fragments (1 kb–1.3 kb), 30% came from MMW fragments (600 bp–1 kb), and 6% came from LMW (300 bp–600 bp) fragments, representing 0.38%, 0.37%, and 0.1% of the available clones in each size class, respectively. From the *Sau*3AI/*Bam*HI library, 15% of the clones came from HMW fragments, 64% from MMW fragments, and 21% from LMW fragments, representing 0.62%, 0.59%, and 0.26% of the available clones in each size class, respectively.

A range in intensity of signal from hybridizations was observed and scored as "strong", "medium," or "weak." Across the two libraries, SSRs were located in 39.8% of all clones with strong hits, 44.1% with medium strength hits, and 16.9% with weak hits. The majority of the signals were weak. Of the microsatellites for which primers were designed, 200 (88.5%) were perfect repeats, 20 (8.8%) were compound repeats, and six (2.7%) were imperfect repeats.

Of the 226 primers designed, 192 amplified well, and 110 pairs revealed polymorphism in a diverse array of 24
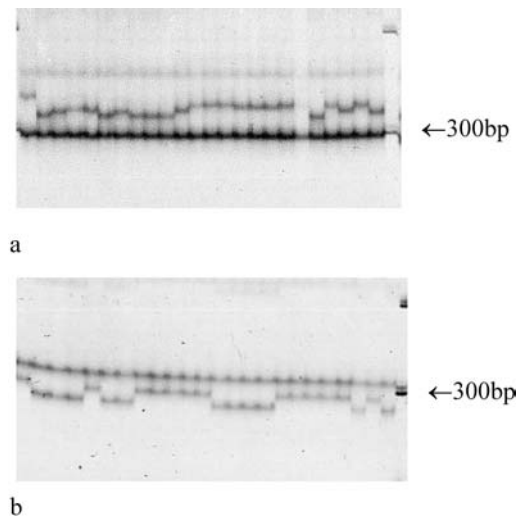

Fig. 1 Genetic variation revealed by primer pairs **a** pPGPseq2C11 and **b** pPGPseq2A5

cultivated peanut accessions. Examples of amplification products are shown in Fig. 1. The 110 primer pairs showed polymorphism at 123 loci. This is possible because of the presence of two genomes in allotetraploid peanut. Thirty-five of these primers were derived from the *Pst*I library and 74 from the *Sau*3AI/*Bam*HI library. This is 44.3% and 65.8% of all amplifiable primers, respectively, and 57.3% overall. Information regarding polymorphism can be found at the associated Web sites indicated above. This includes PCR quality score, number of polymorphic loci, number of alleles found in the 24-germplasm arrays by locus, and PIC value.

Maximum amplification occurred in the majority of primers with protocol 8 (Table 2) from the orthogonal array; this consisted of 30 pM primer, 10 ng DNA, 2 mM Mg^{2+}, 0.25 mM dNTP, 1.2 U *Taq*, and 1× reaction buffer in a total volume of 20 μl. Most primers, however, also amplified well under lower component conditions in protocols 1 and 6 (Table 2).

The most frequent repeat motifs identified were ATT and GA, which represent 29% and 28%, respectively, of all SSRs identified (Table 3). These were followed by AT, which was not probed for, and CTT. Of the amplifiable primers, 81% of ATT and 70.8% of GA repeats were polymorphic in the cultivated groundnut test array. AT showed the maximum number of alleles per locus (5.7), followed by a single SSR of TATC that showed 5 alleles. ATT, GT, and GA had a mean number of alleles per locus of 4.8, 3.8, and 3.6, respectively (Table 3). The average numbers of tandem repeat units varied considerably according to repeat family, with the highest mean repeat number in AT (19.5 repeats), GA (16.2), ATT (15.5), GT (14.7), and the lowest in AATT with just five repeat units. The two SSRs with the largest number of repeats, GA with 42 repeats (primer pPGPseq3B5) and ATT with 37 (pPGPseq2H12) repeats, did not show polymorphism in the diverse groundnut array. The scatter plot in Fig. 2

Table 3 Characterization of libraries for SSRs, by repeat family

Repeat family[a]	Number of clones sequenced	Number of SSRs identified	Number of primers designed	Number (and percentage) of polymorphic SSRs from amplifiable primers	Mean number of alleles per locus
AT	-	35	13	6 (50%)	5.7
GA	273	97	63	26 (51%)	3.6
GT	131	28	15	5 (42%)	3.8
ATT	234	101	62	45 (76%)	4.8
CGG	-	1	1	1 (100%)	2.0
CTA	34	7	3	2 (100%)	3.0
CTC	-	2	2	-	
CTG	101	3	2	1 (50%)	2.0
CTT	170	31	21	7 (47%)	3.0
GGT	384	3	1	-	-
TCA	146	4	3	2 (50%)	3.3
TTG	138	15	13	6 (55%)	2.7
ATCC	9	-	-	-	-
CTCA	-	2	1	1 (100%)	2.0
CTTT	-	2	1	1 (100%)	2.0
TATC	7	5	3	1 (100%)	5.0
TATG	-	1	3	-	-
TATT	0	4	2	-	-
TTAA	-	1	1	-	-
TTG/ATT	-	2	1	2 (100%)	3.5
AT/GA	-	2	3	-	-
AT/GT	-	-	3	-	-
GT/GA	-	2	4	2 (75%)	2.0
AT/ATT	-	-	2	1 (100%)	3.0
GT/TATT	-	-	1	-	-
CTA/CTT	-	-	-	-	-
CTT/CTG	-	-	1	-	-
ATT/CAC	-	-	1	1 (100%)	3.5
TATG/AT	-	-	1	-	-
Totals	1,627	348	226	110	

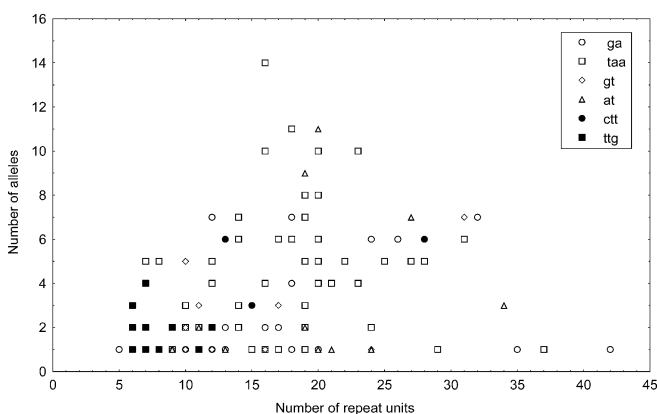[a] Results by library can be found at http://link.springer.com

Fig. 2 Number of alleles per locus for six of the most frequent repeat families

shows that over all repeats, the widest variation in number of alleles is between 12 and 32 repeats, with a lower number of alleles found in the low number or very high number repeats. However, this seems to depend upon motif, as TTG is only found in the quadrant with a low number of alleles and a low number of repeats. GA, GT, and CTT do not differ much for number of alleles across a wide range of repeat numbers. The number of alleles for AT and ATT varies widely over a broad range of number of repeat units. There is no consistent relationship between number of repeats and number of alleles, either across all motifs or individually. The relationship appears to be repeat-motif specific and may become more variable at the intermediate level of repeats.

Discussion

*Pst*I and *Sau*3AI/*Bam*HI libraries were constructed with the objective to find SSRs distributed in the gene-rich euchromatic regions of the genome, as opposed to the centromeric-heterochromatic region rich in repetitive elements. *Pst*I is a six-base pair, methylation-sensitive enzyme and does not cleave at its recognition sequence of 5′-CTGCA▼G-3′ if the 5′ cytosine is methylated. *Sau*3AI/*Bam*HI is a four-base pair cutter 5′-▼GATC-3′, and is sensitive to methylation when the restriction site overlaps the methylation sequence 5′CpNpG3′ or 5′CpG3′. The 5′CpNpG3′ is the most common site for methylation in plant genomes, and the CTG site is methylated in approximately 70% of all cases (Gruenbaum et al. 1981). Most repetitive sequences are completely methylated at this site; therefore, the enzymes cleave preferentially in, and thus enrich for, single- or low-copy-number sequences (Tanksley et al. 1987; Burr et al. 1988). *Pst*I cut the DNA very infrequently, which

resulted in few LMW fragments. Only 1.9% of clones were found from the genomic screen to be from repetitive regions, compared to 6.5% from the *Sau*3AI/*Bam*HI library; therefore, *Pst*I in particular appears to have been enriched for low-copy-number sequences. The large proportion of contigs (30%) found in the *Pst*I library compared to the *Sau*3AI/*Bam*HI library (7%) may be attributed to the rare cutting of this 6-bp enzyme. This infrequent cleavage also leads to a poor representation of the genome.

The higher proportion of sequences that contained SSRs from which primers could be designed from the HMW (1–1.3 kb) and MMW (600 bp–1 kb) fragments compared to the LMW (300–600 bp) fragments suggests an increased efficiency from using genomic libraries with size fractions between 600 bp and 1.3 kb. This efficiency would likely also improve if larger fragment sizes were sequenced from both directions. It was hypothesized that SSRs would more likely be located in clones with strong hybridization signals. In these two libraries, 43% of primers were designed from clones having weak signals, so weak signals should not necessarily be discarded.

It is apparent that all SSR motifs are not equally abundant in eukaryotic genomes and the relative abundance of different motifs varies among species (Wang et al. 1994). The abundance of various microsatellite motifs found in this study is largely consistent with previous reports of dicotyledonous species. ATT was the most frequent motif, followed closely by GA, then AT, CTT, and GT. EMBL and GenBank database searches of plant SSRs over 20 bp in length have found the AT motif to be the most frequent, followed by A and GA (Lagercrantz et al. 1993; Morgante and Olivieri 1993; Wang et al. 1994); ATT and CTT were found to be the most frequent trinucleotide repeat motifs (Lagercrantz et al. 1993; Morgante and Olivieri 1993). In this study, both ATT and GA were more frequent than AT (Table 3), although the observed frequency of AT is likely to be highly under-represented as it was not probed for. A similar situation was found in chickpea when genomic libraries were screened (Hüttel et al. 1999). ATT was found to be the most common trinucleotide motif in soybean (Akkaya et al. 1995), tomato (Smulders 1997; Areshchenkova and Ganal 2002), and pine (Echt and May-Marquardt 1997). A GenBank-EMBL database search of *Pisum sativum* sequences revealed that the four most frequent motifs were the same as found in this study for peanut (Burstin et al. 2001). Surprisingly, from a search of GenBank sequences in *Phaseolus* and *Vigna*, only 1 of 61 SSRs identified was of the ATT motif, the most frequent being AT, followed by GA, CCA, CCT, and ATTT (Yu et al. 1999).

The high mean number of alleles per polymorphic locus in ATT (4.8) and GA (3.6), combined with their relative frequency in the genome, make them the most useful and appropriate motifs to target for generating further SSR markers for peanut. Song et al. (2002) reported that ATT was the most useful trinucleotide motif for developing SSR markers and had resulted in the development of more than 500 SSR markers in soybean (Cregan et al. 1999). Udupa et al. (1999) found ATT to be extremely polymorphic in chickpea.

In this study, the average number of repeat units varied considerably between motif types. The repeat motifs that had amongst the largest number of tandem repeat units (AT, GA, ATT, and GT) also revealed the highest number of alleles per locus. These are also the motifs that are generally most abundant in the peanut genome. It has been reported that the degree of polymorphism increases with the total length of the repeat (Weber 1990; Saghai-Maroof et al. 1994; Innan et al. 1997; Smulders et al. 1997; Hüttel et al. 1999; Udupa et al. 1999; Burstin et al. 2001); however, not all studies have supported this finding, some finding the correlation to be weak (Love et al. 1990; Bell and Ecker 1994; Yu et al. 1999). Here, there is no consistent relationship between the number of repeats and the number of alleles, either across all motifs or individually. The relationship appears to be repeat-motif specific and may become variable at the intermediate level of repeats.

In conclusion, through this study we have substantially increased the number of markers that reveal polymorphism in cultivated peanut. Sufficient markers are now available for a skeleton map of cultivated groundnut, provided that a sufficiently wide cross is used. Our results indicate that ATT and GA are the most convenient probes by which to isolate microsatellite loci in peanut by screening genomic libraries.

References

Akkaya M, Shoemaker R, Specht J, Bhagwat A, Cregan P (1995) Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci 35:1439–1445

Anderson J, Churchill G, Autrique J, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Areshchenkova T, Ganal M (2002) Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. Theor Appl Genet 104:229–235

Bell C, Ecker J (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144

Bianchi-Hall C, Keys R, Stalker H (1991) Use of protein profiles to characterize peanut cultivars. Newsl Assoc Official Seed Anal 65:25–26

Burow MD, Simpson CE, Starr JL, Paterson AH (2001) Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): broadening the gene pool of a monophyletic polyploid species. Genetics 159:823–837

Burr B, Burr F, Thompson K, Albertsen M, Stuber C (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Burstin J, Deniot G, Potier J, Weinachter C, Aubert G, Baranger A (2001) Microsatellite polymorphism in *Pisum sativum*. Plant Breed 102:311–317

Cobb BD, Clarkson JM (1994) A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. Nucleic Acids Res 22:3801–3805

Cregan P, Jarvik T, Bush A, Shoemaker R, Lark K, Kahler A, Van Toai T, Lohnes D, Chung J, Specht J (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Daly MJ, Lincoln SE, Lander ES (1991) "PRIMER", unpublished Software, Whitehead Institute / MIT Center for Genome Research

Echt C, May-Marquardt P (1997) Survey of microsatellite DNA in pine. Genome 40:9–17

Grieshammer U, Wynne JC (1990) Isozyme variability in mature seeds of U.S. peanut cultivars and collections. Peanut Sci 18:72–75

Gruenbaum Y, Naveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292:860–862

Halward TM, Stalker HT, LaRue EA, Kochert G (1991) Genetic variation detectable with molecular markers among unadapted germ-plasm resources of cultivated peanut and related wild species. Genome 34:1013–1020

Halward T, Stalker T, LaRue E, Kochert G (1992) Use of single-primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L.). Plant Mol Biol 18:315–325

Hopkins M, Casa A, Wang T, Mitchell S, Dean R, Kochert G, Kresovich S (1999) Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. Crop Sci 39:1243–1247

Hüttel B, Winter P, Weising K, Choumane W, Weigand F, Kahl G (1999) Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Genome 42:210–217

Innan H, Terauchi R, Miyashita N (1997) Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics 146:1441–1452

Kochert G, Halward T, Branch WD, Simpson CE (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor Appl Genet 81:565–570

Krakowski K, Brunfille J, Seto J, Baskin D, Seto D (1995) Rapid purification of fluorescent dye-labeled products in a 96-well format for high-throughput automated DNA sequencing. Nucleic Acids Res 23:4930–4931

Lacks G, Stalker H (1993) Isozyme analyses of *Arachis* species and interspecific hybrids. Peanut Sci 20:76–81

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111–1115

Lanham P, Forster B, McNicol P, Moss J, Powell W (1994) Seed storage protein variation in *Arachis* species. Genome 37:487–496

Love J, Knight A, McAleer M, Todd J (1990) Towards construction of a high-resolution map of the mouse genome using PCR-analyzed microsatellites. Nucleic Acids Res 18:4123–4130

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterson RH (1997) High throughput fingerprinting analysis of large-insert clones. Genome Res 7:1072–1084

Morgante M, Olivieri A (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Paik-Ro OG, Smith RL, Knauft DA (1992) Restriction fragment length polymorphism evaluation of six peanut species within the *Arachis* section. Theor Appl Genet 84:201–208

Pepin L, Amigues Y, Lepingle A, Berthier J, Bensaid A, Vaiman D (1995) Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra nircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. Heredity 74:53–61

Saghai-Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 99:5466–5470

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning. A laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Savoy C (1976) Peanut (*Arachis hypogaea* L.). Seed protein characterization and genotype sample classification using polyacrylamide gel electrophoresis. Biochem Biophys Res Commun 68:886–893

Smulders M, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor Appl Genet 97:264–272

Song Q, Fickus E, Cregan P (2002) Characterization of trinucleotide SSR motifs in wheat. Theor Appl Genet 104:286–293

Tanksley S, Miller J, Paterson A, Bernatzky R (1987) Molecular mapping of plant chromosomes. In: Gustafson J, Appels R (eds) Chromosome structure and function. Plenum Press, New York, pp 157–173

Tombs MP (1963) Variant forms of arachin. Nature 200:1321–1322

Udupa S, Robertson L, Weigand F, Baum M, Kahl G (1999) Allelic variation at $(TAA)_n$ microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.). Mol Gen Genet 261:354–363

Wang Z, Weber J, Zhong G, Tanksley S (1994) Survey of short tandem DNA repeats. Theor Appl Genet 88:1–6

Weber J (1990) Informativeness of human $(dC\text{-}dA)_n \times (dG\text{-}dT)_n$ polymorphisms. Genomics 7:524–530

Yu K, Park S, Poysa V (1999) Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*). Genome 42:27–34