Theor Appl Genet (2005) 111: 1060–1071
DOI 10.1007/s00122-005-0028-x

ORIGINAL PAPER

M. C. Moretzsohn · L. Leoi · K. Proite
P. M. Guimarães · S. C. M. Leal-Bertioli
M. A. Gimenes · W. S. Martins
J. F. M. Valls · D. Grattapaglia · D. J. Bertioli

A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae)

Received: 23 March 2005 / Accepted: 23 June 2005 / Published online: 9 August 2005

Abstract Cultivated peanut (*Arachis hypogaea*) is an important crop, widely grown in tropical and subtropical regions of the world. It is highly susceptible to several biotic and abiotic stresses to which wild species are resistant. As a first step towards the introgression of these resistance genes into cultivated peanut, a linkage map based on microsatellite markers was constructed, using an F_2 population obtained from a cross between two diploid wild species with AA genome (*A. duranensis* and *A. stenosperma*). A total of 271 new microsatellite markers were developed in the present study from SSR-enriched genomic libraries, expressed sequence tags (ESTs), and by "data-mining" sequences available in GenBank. Of these, 66 were polymorphic for cultivated peanut. The 271 new markers plus another 162 published for peanut were screened against both progenitors and 204 of these (47.1%) were polymorphic, with 170 codominant and 34 dominant markers. The 80 codominant markers segregating 1:2:1 ($P < 0.05$) were initially used to establish the linkage groups. Distorted and dominant markers were subsequently included in the map. The resulting linkage map consists of 11 linkage groups covering 1,230.89 cM of total map distance, with an average distance of 7.24 cM between markers. This is the first microsatellite-based map published for *Arachis*, and the first map based on sequences that are all currently publicly available. Because most markers used were derived from ESTs and genomic libraries made using methylation-sensitive restriction enzymes, about one-third of the mapped markers are genic. Linkage group ordering is being validated in other mapping populations, with the aim of constructing a transferable reference map for *Arachis*.

Communicated by S. J. Knapp

Electronic supplementary material is available for this at http://dx.doi.org/10.1007/s00122-005-0028-x.

M. C. Moretzsohn (✉) · K. Proite · P. M. Guimarães
S. C. M. Leal-Bertioli · J. F. M. Valls · D. Grattapaglia
Embrapa Recursos Genéticos e Biotecnologia, C.P. 02372,
CEP 70.770-900 Brasília, DF, Brazil
E-mail: marciocm@cenargen.embrapa.br
Tel.: +55-61-4484615
Fax: +55-61-3403624

M. C. Moretzsohn · K. Proite · D. Grattapaglia
Departamento de Biologia Celular,
IB - Universidade de Brasília (UnB), Campus Universitário,
CEP 70.910-900 Brasília, DF, Brazil

L. Leoi · D. J. Bertioli · D. Grattapaglia
Universidade Católica de Brasília, Campus II, SGAN 916,
CEP 70.790-160 Brasília, DF, Brazil

M. A. Gimenes
Departamento de Genética, IB-UNESP, Rubião Jr.,
CEP 18618-000 Botucatu, SP, Brazil

W. S. Martins
Departamento de Ciência da Computação, Universidade Católica de Goiás, Goiânia, GO, Brazil

Introduction

The genus *Arachis*, family Fabaceae, is native to South America, and contains 69 described species assembled into nine sections, according to their morphology, geographic distribution, and crossability (Krapovickas and Gregory 1994). The cultivated species, peanut (*Arachis hypogaea*), is a key crop for small farmers throughout the tropics, but especially in Africa and Asia, it is a valuable source of dietary protein and oil, and of fodder for livestock. In addition, the peanut's ability to fix nitrogen is important in the maintenance of soil fertility especially where farmers are too poor to afford chemical fertilizers. Peanut world production is increasing and has reached 35.6 million tons (in the shell) and 5.8 million tons of oil per year (FAO 2003). Currently peanut is the fifth most important oilseed in the world in terms of volume of oil production.

Cultivated peanut is an allotetraploid (genome AABB). It is a member of the section *Arachis* that also includes about 25 diploid and one other tetraploid wild species (*A. monticola*) (Krapovickas and Gregory 1994).

Its origin was via hybridization of two diploid wild species followed by a rare spontaneous duplication of chromosomes (Halward et al. 1991; Young et al. 1996). The resulting tetraploid plant would have had hybrid vigor, but have been reproductively isolated from its wild relatives. Therefore, all land races of peanuts are probably derived from one, or a few plants. This led to a very limited genetic diversity, which has constrained advances in genetics that are necessary for efficient modern breeding, and led to low diversity for some traits of agricultural interest. In contrast, wild diploid *Arachis* species are genetically very diverse and have been selected during evolution by a range of environments and biotic stresses, providing a rich source of variation in agronomically important traits.

In *Arachis*, few linkage maps have been published. Halward et al. (1993) developed an RFLP-based map using an F_2 population derived from a cross between two diploid wild species with AA genome (*A. stenosperma* and *A. cardenasii*). A total of 117 loci were mapped in 11 linkage groups. Another RFLP-based map was published for a tetraploid population (Burow et al. 2001). A synthetic amphidiploid *A. batizocoi* × (*A. cardenasii* × *A. diogoi*)^{4x}, used as a donor parent, was crossed with *A. hypogaea* to generate a BC_1 population of 78 progeny. Three hundred and seventy RFLP loci were mapped onto 23 linkage groups. A BC_4F_2 population obtained through this crossing schedule was used for the identification of RAPD markers linked to root-knot nematode resistance genes (Burow et al. 1996). RFLP and RAPD markers have also been used to follow the introgression of *A. cardenasii* (AA genome) chromosome segments into *A. hypogaea* genome (Garcia et al. 1995). This analysis has shown that chromosome recombination, and not chromosome substitution, was the mechanism of introgression, which is an important result for the use of wild *Arachis* species in peanut-breeding programs. RAPD, SCAR, and RFLP markers have proved to be useful for the identification of markers tightly linked to nematode resistance genes introgressed from *A. cardenasii* into *A. hypogaea* (Garcia et al. 1996). Recently, AFLP markers were used to map genes conferring resistance to the aphid vector of groundnut rosette disease, which is the most destructive viral disease of peanut in Africa (Herselman et al. 2004).

RFLP analysis is time consuming and labor intensive, while RAPD and AFLP are dominant markers of low information content. Consequently, these are not the ideal tools for high throughput analysis and for the implementation of marker assisted selection (MAS) in peanut-breeding programs. In contrast, microsatellites are multiallelic, highly polymorphic, typically codominant genetic markers, and require small amounts of DNA, since they are based on PCR and gel electrophoresis. Their main disadvantage is the high initial cost for the development of markers using genomic DNA libraries enriched for simple sequence repeat (SSR) sequences. With the availability of expressed sequence tags (ESTs) and other DNA sequence data, development of SSR markers through data mining has become a fast, efficient, and low-cost option for many plant species. Furthermore, the use of markers developed from different sources should enable a better coverage of the whole genome.

Although the number of microsatellite markers published for peanut has increased considerably in the last 5 years (Hopkins et al. 1999; He et al. 2003; Ferguson et al. 2004; Moretzsohn et al. 2004), it is still not sufficient for the construction of saturated linkage maps. Thus, the objectives of the current study were to (1) develop a large set of microsatellite markers from SSR-enriched genomic DNA libraries, from ESTs, and by data mining published *Arachis* spp. sequences; (2) characterize the newly developed markers in a set of *A. hypogaea* accessions, and (3) construct a microsatellite-based, gene-rich map using a diploid AA genome F_2 population.

Materials and methods

Plant material

A single *A. hypogaea* cv. Tatu plant was used for the construction of the two genomic DNA libraries. EST libraries were made from leaf and root tissues from *A. stenosperma* (accession V10309). Sixteen *A. hypogaea* samples (Table 1) were used for the characterization of the newly developed SSR markers. For map construction, an F_2 population was produced by selfing a unique F_1 plant derived from a cross between *A. duranensis* (accession K7988), used as the female parent and *A. stenosperma* (accession V10309), used as the male. In order to obtain enough seeds for the generation of the F_2 population, the F_1 plant was cloned by cuttings. These species were chosen because they are easily crossed and their progeny are fertile, they are contrasting at the DNA level (Moretzsohn et al. 2004), related to the cultivated peanut, and they segregate for several traits of interest. All plants were obtained from the Brazilian *Arachis* germplasm collection, maintained at Embrapa Genetic Resources and Biotechnology — CENARGEN (Brasília-DF, Brazil).

DNA extraction

Total genomic DNA was extracted from young leaves using the protocol of Grattapaglia and Sederoff (1994) modified by the inclusion of an additional precipitation step with 1.2 M NaCl. DNA concentration was estimated by agarose gel electrophoresis comparing the fluorescence intensities of the ethidium bromide stained samples to those of λ DNA standards.

Genomic libraries construction

Total genomic DNA libraries enriched for dinucleotide repeats were constructed following a standard protocol

Table 1 *Arachis hypogaea* accessions analyzed in this study

Number	Accession	Species/subspecies/variety (Type)	Collection site
1	cv. BR-1	A. hypogaea ssp. fastigiata var. fastigiata (Valencia)	Paraíba, Brazil
2	cv. Runner886	A. hypogaea ssp. hypogaea var. hypogaea (Virginia)	São Paulo, Brazil
3	cv. Tatu	A. hypogaea ssp. fastigiata var. fastigiata (Valencia)	São Paulo, Brazil
4	Mf 1538	A. hypogaea ssp. hypogaea var. hirsuta	Pichincha, Ecuador
5	Mf 1560	A. hypogaea ssp. fastigiata var. peruviana	Pastaza, Ecuador
6	Mf 1678	A. hypogaea ssp. fastigiata var. aequatoriana	Sucurubios, Ecuador
7	Mf 2389	A. hypogaea ssp. fastigiata var. aequatoriana	Unknown
8	Mf 3207	A. hypogaea ssp. fastigiata var. vulgaris (Spanish)	Unknown
9	Mf 3618	A. hypogaea ssp. hypogaea var. hirsuta	Unknown
10	Of 109	A. hypogaea ssp. hypogaea var. hypogaea (Xingu type)	Mato Grosso, Brazil
11	Of 128	A. hypogaea ssp. hypogaea var. hypogaea (Xingu type)	Mato Grosso, Brazil
12	Sv 429	A. hypogaea ssp. fastigiata var. peruviana	Yurimaguas, Peru
13	V 10067	A. hypogaea ssp. hypogaea var. hypogaea (Virginia)	Santa Catarina, Brazil
14	V 13009	A. hypogaea ssp. hypogaea var. hypogaea (Virginia)	Tocantins, Brazil
15	V 13115	A. hypogaea ssp. fastigiata var. vulgaris (Spanish)	Minas Gerais, Brazil
16	V 6265	A. hypogaea ssp. fastigiata var. fastigiata (Valencia)	Bahia, Brazil

(Rafalski et al. 1996) with minor modifications. Peanut total genomic DNA, was isolated from a single *A. hypogaea* plant (cv. Tatu) and aliquots were digested separately with three different enzymes (*Mse*I, *Sau*3AI, and *Tsp*509I) to identify which one produced the most adequate fragment profile for library development. Fragments ranging in size from 200 to 800 bp were transferred to NA 45 DEAE cellulose membranes (Schleicher & Schuell BioScience, NH, USA), after electrophoresis in 1.5% low-melting agarose gel. DNA was resuspended in TE buffer and ligated to *Sau*3AI adaptors (Short adaptor: 5′-CAGCCTAGAGCCGAATTCACC-3′ and Long adaptor: 5′-pGATCGGTGAATTCGGCTCTAGGCTG-3′). Two genomic libraries enriched for dinucleotide repeats were constructed. Fragments were selected by hybridization with biotinylated dinucleotide repeats ($(TC)_{13}$ and $(AC)_{13}$) and recovered by magnetic beads linked to streptavidin (Dynabeads Streptavidin, Dynal Biotech., Norway). Enriched fractions were amplified by PCR, using the Short *Sau* adaptor as a primer. Amplification products were purified using QIAquick PCR purification system (Qiagen, CA, USA), cloned into plasmid pGEM-T (Promega, WI, USA), and then transformed into *E. coli* strain XL1-Blue (Promega) by electroporation. Transformed cells were grown on LB-Amp plates (50 μg/ml ampicillin) containing X-Gal and IPTG at 37°C. Plasmid DNA of the white colonies was extracted by alcaline lysis (Mini-preps) and prepared for sequencing. Miniprep followed the protocol described by Sambrook et al. (1989), with some minor adaptations for extraction in 96-well plates.

EST libraries construction

Two cDNA libraries were made, each with either young leaves or roots of *A. stenosperma*. Total RNA was extracted using Trizol (Invitrogen, CA, USA) and mRNA purified using the Poly A^+ RNA-Oligotex (Qiagen, CA, USA) according to the manufacturer's protocol. Full-length cDNA libraries were constructed using the SMART cDNA synthesis kit in λTriplEx2 (Clontech, CA, USA). The resulting cDNA was packed into λ phages using the Gigapack III Gold packaging kit (Stratagene, CA, USA). From the initial plating, the libraries were estimated to contain 10^7 (roots) and 10^8 (leaves) recombinant clones. The pTriplEx2 phagemid clones in *E. coli* were obtained using the mass in-vivo excision protocol according to the manufacturer's instructions (Clontech).

DNA sequencing and primer design

Sequencing reactions were performed with T7 and SP6 primers and the Big-Dye Terminator Cycle Sequencing Kit, version 3.1 (Applied Biosystems, CA, USA) or the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Pharmacia Biotech) using the Applied Biosystems automated DNA 3100 and 377 sequencers. For the ESTs, the cDNA inserts were sequenced from the 5′end using specific PCR primers for the λTriplEx2. To facilitate the selection of SSR-containing sequences from all libraries and the efficient assembly of redundant sequences and forward and reverse sequencing runs, a dedicated module (available from WSM — wsmartins@wsmartins.net — on request) has been developed which integrates the Staden package interface for sequence assembly (Staden et al. 2003) and Troll for finding SSRs (Castelo et al. 2002). In addition to the ESTs and the enriched genomic libraries, the module was used to "data-mine" SSR markers from data available in GenBank. The new sequences identified in this study were submitted to the National Center for Biotechnology Information database (NCBI, USA; GenBank accession #s DQ099037-DQ099242). Primers complementary to unique DNA sequences flanking the SSRs were designed using the computer program Primer 3 (Whitehead Institute of Biomedical Research — http://www-

genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). To reduce problems with spurious banding patterns, some stringent criteria in primer sequence design were applied: (1) primer T_m ranging from 57 to 65°C; (2) 1°C difference in T_m between primer pairs; (3) GC content ranging from 40 to 60%; (4) absence of complementarity between primers. Default values were used for the other parameters, except the "product size ranges" that was between 100 and 400 bp.

PCR amplification of SSR marker loci

PCR reactions were performed in 13 µl volumes, containing 1X PCR buffer (10 mM Tris–HCl pH 8.3, 50 mM KCl), 2.0 mM $MgCl_2$, 0.2 mM of each dNTP, 1 U of *Taq* DNA polymerase, 1 µg/µl of purified BSA (New England Biolabs, MA, USA), 5 pmol of each primer and 10 ng of genomic DNA. Amplifications were performed using either a 9700 System (Applied Biosystems, CA, USA) or a PTC-100 (MJ Research, MA, USA) thermal cycler, with the following conditions: 94°C for 5 min (1 cycle), 94°C for 1 min, 48–64°C for 1 min, 72°C for 1 min (30 cycles); and 72°C for 7 min (1 cycle). The annealing temperature was optimized for each primer pair to produce clear DNA band amplification, without spurious fragments. For genotype determination, the amplified products were separated on 5% denaturing polyacrylamide gels stained with silver nitrate (Creste et al. 2001). Some SSR markers highly contrasting between the progenitors of the mapping population were run on 3.5% agarose gels stained with ethidium bromide. Fragment sizes were estimated by comparing with a 10-bp DNA ladder standard (Gibco/BRL, MD, USA).

Characterization of the newly developed SSR markers

The 271 newly developed SSR markers were first tested on a set of six *A. hypogaea* accessions, one from each of the six described peanut varieties (Table 1 — accessions 1, 4, 5, 6, 8, and 14). The expected size and the quality of the amplified fragments were observed on 3.5% agarose gels. When necessary, the annealing temperatures were adjusted to obtain better specificity of PCR amplification. Primers developed from genomic libraries are named by Ah1TC or Ah2TC (for *A. hypogaea* library enriched for TC repeats) or Ah2AC (for AC repeats) followed by the plate number (TC1–TC11 or AC1–AC3) and the clone identification (A01–H12). Primers developed from cDNA libraries are named by AS1 (for *A. stenosperma* libraries), followed by two letters (ML, for mature leaf, RN or RI for root) and the plate number (from 1 to 22) and the clone identification (A01–H12). Primers obtained by data-mining GenBank were denominated by the name of the submitted sequence. These marker names start with GI, pPGP or pPGS. A subjective score was given to each marker from 1 to 3 denoting the fragment amplification quality, 3 being the best. The number of alleles per locus, the product size ranges, and gene diversity (GD) were estimated for the polymorphic markers using a set of 16 peanut accessions (Table 1). Gene diversity was calculated according to the formula: $GD = 1 - (\sum p_i^2)$, where p_i is the frequency of the ith allele in the 16 samples (Nei 1973). For the characterization of the sequences used for marker development, BLAST databases were formatted from all putative proteins encoded by the *Arabidopsis* genome and all Fabaceae proteins in GenBank. Similarity searches were done using BlastX (Altschul et al. 1990), and hits were considered significant if they had an E-value of less than 1×10^{-8}.

Map construction

The 271 newly developed SSR markers plus another 162 already published SSR markers (Hopkins et al. 1999; Palmieri et al. 2002; He et al. 2003; Moretzsohn et al. 2004; Palmieri et al. 2005; Bravo et al. 2005; Gimenes et al. 2005) were screened against the two progenitors. Polymorphic markers were analyzed on the mapping population consisting of 93 F_2 individuals. A χ^2 test was performed to test the null hypothesis of 1:2:1 or 3:1 segregations on all scored markers. Only marker loci that did not show segregation distortion ($P > 0.05$) from the expected 1:2:1 were used for the initial map construction, in order to eliminate spurious linkages. The linkage analysis was done using Mapmaker Macintosh version 2.0 (Lander et al. 1987). A minimum LOD score of 12.0 and maximum recombination fraction (θ) of 0.3 were set as thresholds for linkage groups (LG) determination with the "group" command. The most likely marker order within each LG was estimated by the matrix correlation method using the "first order" command or by the three-point analysis ("three point" command) for groups containing more than seven markers. Marker orders were confirmed by comparing the log-likelihood of the possible orders using multipoint analysis ("compare" command) and by permuting all adjacent triple orders ("ripple" command). Loci that could be ordered with a likelihood ratio support ≥1,000:1 were used to establish a framework map. Codominant marker loci that did not show segregation distortion at $P > 0.01$ were used in this analysis. In a second step, the LOD score was set to 6.0 in order to include new markers in the groups. The "try" command was then used to determine the exact position of the new markers within each group. In a following step, distorted markers and markers segregating 3:1 (present in one parent and absent in the other) were included using the "group" command. The new marker orders were again confirmed with the "first order", "compare", and/or "ripple" commands. Recombination fractions were converted into map distances in centimorgans (cM) using the Kosambi's mapping function. Genome map sizes were estimated according to Hulbert et al. (1988) taking into consideration only pairwise comparisons between markers placed on the framework map.

Results and discussion

Development of microsatellite markers

Digestion of the *A. hypogaea* genomic DNA with three different enzymes (*Mse*I, *Sau*3AI, and *Tsp*509I) revealed that *Sau*3AI produced the most adequate profile for library development, with fragments ranging from 200 to 800 bp in size. *Sau*3AI is sensitive to cytosine methylation and therefore is expected to preferentially cut in hypomethylated gene-rich DNA, producing libraries enriched for low copy sequences that are usually gene-rich. Two libraries were constructed based on dinucleotide repeat motifs (TC and AC). These repeats were chosen because they are among the most abundant SSR repeats in plant species (Wang et al. 1994; Gupta and Varshney 2000). Out of the 1,152 positive clones (white colonies) sequenced, 224 (19.4%) had SSRs, 210 being non-redundant. When necessary, these clones were sequenced on the reverse strand to enable the construction of high-quality contigs. Of the 210 unique SSR positive clones, 121 (Electronic supplementary material — ESM 1) were suitable for primer design (10.5% of the clones sequenced). The design of primers for the other 89 unique SSRs identified was not possible due to the low GC content of the regions flanking the SSR, or the close proximity of the SSR to the end of the cloned insert DNA. The percentage of primers designed, in relation to the number of clones sequenced (10.5%), is significantly lower than the proportion found (41.4%) for peanut in a previous study that included an additional selection of positive clones by anchored-PCR (Moretzsohn et al. 2004). This shows that this step is effective in significantly improving the yield of useful clones sequenced and reducing costs. However, it is time consuming and its omission did accelerate the process of SSR marker development.

A total of 2,740 EST sequences were screened for SSRs. Ninety-eight showed di-, tri-, tetra- or pentanucleotide repeats (3.6%) and 81 (3.0%) were suitable for primer design (ESM 1). These numbers are similar to the percentages reported for other plant species, ranging from 1.7 in wheat to 7.5 in barley (Varshney et al. 2002; Thiel et al. 2003; Eujayl et al. 2004; Gao et al. 2004; Khlestkina et al. 2004).

Finally, sequences available in GenBank, for which SSR primers had not been designed, were screened for the presence of repeats (>5 core units). A total of 69 new microsatellite primer pairs were designed (ESM 1), 38 from genomic sequences, 25 from ESTs, 1 from a patented sequence, and 5 from *A. hypogaea* genes, such as cold stress responsive protein, desaturase, and seed galactose-binding lectin.

Screening for SSR sequences was done using the Staden-TROLL module, and was quick and convenient with complete integration into the Staden Package. All SSR containing sequences, including those for which primers had been previously published, were used to construct a peanut SSR database in the Staden Gap4 interface. Hundreds of contigs were established thus eliminating redundant sequences. By doing this, we could be sure that all the 271 newly developed SSR markers were unique. Additionally a BlastN search of these sequences against a database of all published *Arachis* SSR containing sequences further ensured that all markers were new.

Characteristics of *Arachis* SSR repeats

Of the 271 sequences for which primers were designed, 159 (58.7%) had perfect repeats, 82 (30.2%) had compound repeats, and 30 (11.1%) had imperfect repeats. Repeats in the TC/AG- and AC/TG-enriched genomic libraries were, as expected, almost entirely TC/AG and AC/TG respectively. For the ESTs, in the 106 markers developed, a total of 153 repeats were found, since some sequences had more than one repeat motif. The trinucleotide motifs were the most abundant type of SSRs detected, with 112 (73.2%) SSRs, followed by di- (22.2%), tetra- (3.3%) and pentanucleotide repeats (1.3%). These proportions are variable in the few published studies of EST derived microsatellites in plants. For trinucleotides, percentages have ranged from 30.2% in maize to 90.3% in sugarcane; dinucleotide motifs from 9.7% in sugarcane to 39.0% in soybean; tetranucleotide motifs from 2.2% in soybean and *Arabidopsis* to 10.3% in *Medicago*; and pentanucleotide repeats from 0.6% in barley to 9.0% in *Medicago* (Cordeiro et al. 2001; Varshney et al. 2002; Gao et al. 2003; Thiel et al. 2003; Eujayl et al. 2004; Tian et al. 2004). However, all these studies have shown that the trimeric SSRs are more abundant in plant EST sequences than the other SSR unit-size classes. These findings are not unexpected and have also been reported for other higher eukaryotes including *Drosophila*, *Caenorhabditis elegans*, mammals, humans, and yeast (reviewed by Li et al. 2004). This has been attributed to negative selection against frameshift mutations in coding regions (Metzgar et al. 2000) and possibly by positive selection for specific single amino-acid stretches (Morgante et al. 2002).

AAG/TTC (27 out of 153), AAT/TTA (21/153), and AAC/TTG (14/153) motifs were the most abundant trinucleotides. Considerable differences in the frequencies of trinucleotide repeat types have also been reported, but these three classes were usually between the most common tri-motifs in other dicot species (Cardle et al. 2000; Gao et al. 2003; Eujayl et al. 2004; Tian et al. 2004). In several monocot species, CCG/GGC repeats have been shown to be the most common motif (Cordeiro et al. 2001; Morgante et al. 2002; Varshney et al. 2002; Gao et al. 2003), but were detected in only two sequences in the present study. The AAT/TTA motifs are the least common ($<1\%$) in animals and in monocot species (Li et al. 2004). In contrast, these motifs were the second most frequent motif found in *Arachis* ESTs and had also high frequencies in soybean and *Medicago* (Gao et al. 2003; Eujayl et al. 2004; Tian et al. 2004).

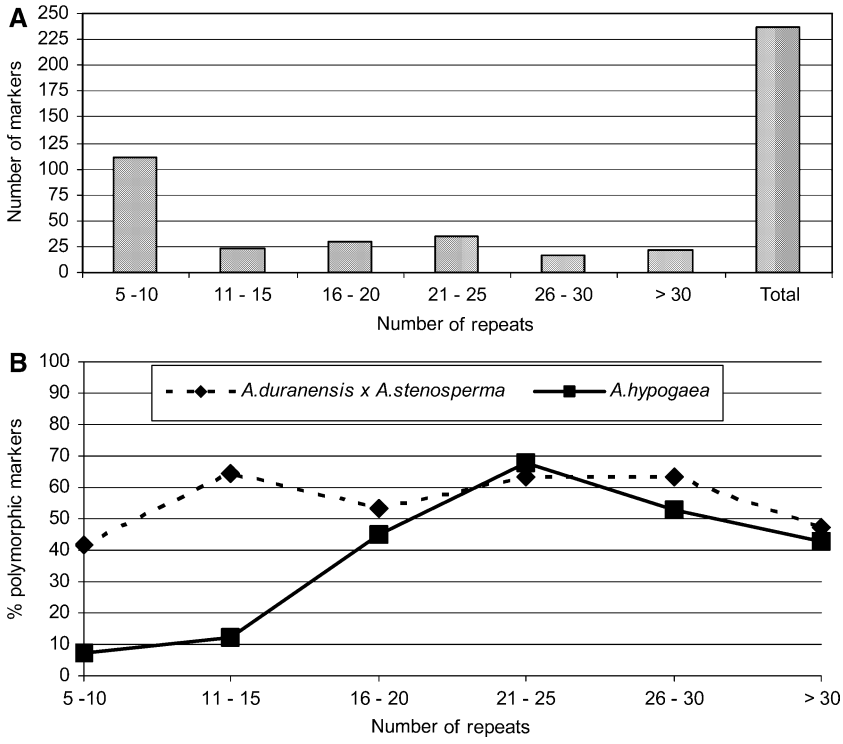

Fig. 1 Number of markers detected per repeat size class (**a**) and relationships between repeat numbers of the SSRs and the polymorphism observed when markers were screened against *A. duranensis* and *A. stenosperma*, and in six accessions of *A. hypogaea* (**b**). A total of 225 markers that amplified products with good resolution were considered in the analysis of *A. duranensis* and *A. stenosperma*, and 236 markers in the analysis of *A. hypogaea*

AG/TC and AT/TA (16 out of 153 each) had the highest frequencies among the dinucleotide repeats, followed by AC/TG (6/153). CG/GC repeats were not found in any of the sequences. Similar results have been described for several plant species, with AG/TC repeats showing the highest frequencies and CG/GC the lowest (Kantety et al. 2002; Eujayl et al. 2004; Tian et al. 2004). Although frequencies of AT/TA repeats have been usually lower than those found in the present study, high frequencies have also been described for other legume species, such as *Medicago* and soybean (Morgante et al. 2002; Gao et al. 2003; Eujayl et al. 2004). The putative functions of SSRs in ESTs are still not clear as are the reasons why repeat motif types are so variable in the different plant species. However, these findings showed that the predominance of a particular repeat motif in one species is not always an indication of representation in another plant species.

Polymorphism and transferability of the new SSR markers

The 271 new SSR markers (ESM 1) were screened for polymorphism against six *A. hypogaea* samples, being one from each of the six described varieties and against the progenitors of the diploid mapping population. As expected, polymorphism between the two accessions of different species was greater than for the six accessions within the cultivated species (Fig. 1, Table 2).

For the wild species progenitors, 27 (10.0%) primer pairs did not amplify any interpretable fragment, 11 (4.1%) amplified fragments only in *A. stenosperma* and 4 (1.5%) in *A. duranensis*. Thus, 229 (84.5%) primer pairs were fully transferable between the two wild species. For these 229 markers, 113 (49.3%) were polymorphic.

For *A. hypogaea*, 37 primer pairs (13.7%) did not amplify any clearly interpretable fragment. Of the remaining 234 primer pairs, 66 (28.2%) were polymorphic for at least one of the six samples. Sixty-two polymorphic markers were used for the analysis of allelic diversity using a sample of 16 *A. hypogaea* accessions (Table 3). The other four polymorphic markers (AC2C2, RN0-602, Seq14G3, and TC1E6) were not included in this analysis, because they resulted in very poor amplification products. The number of alleles detected ranged from 2 to 12 at each of the 62 polymorphic loci analyzed, with an average of 5.87 alleles per locus. GD values ranged from 0.121 (for loci TC11B11 and GI-385) to 0.910 (for locus TC2D6), with an average value of 0.690. GD is an estimate of the probability that two randomly chosen genes from a population are different. The values obtained in the present study are probably overestimated, since some of the markers amplified alleles at duplicated loci. Thus, they cannot be compared to those obtained for diploid species, but the average value (0.690) is similar to values found for several polyploid species, including peanut (Provan et al. 1996; Chavarriaga-Aguirre et al. 1998; Prasad et al. 2000; He et al. 2003; Ferguson et al. 2004; Moretzsohn et al. 2004). In contrast, average GD values estimated by RFLP markers were considerably lower for five *Arachis* species, including *A. hypogaea* (Paik-Ro et al. 1992). These results demonstrate the usefulness of microsatellite markers for genetic diversity analysis of cultivated peanut.

Table 2 Number of polymorphic markers detected by a screening against accessions K7988 (*A.duranensis*) and V10309 (*A. stenosperma*), used as progenitors of the F_2 mapping population, and against six *A. hypogaea* accessions belonging to the six described varieties

Repeat type	Genomic library				cDNA library
	TC/AG	AC/TG	Various	Total	Various
K7988 × V10309					
Polymorphic markers					
Codominant	48 (48.5%)	5 (22.7%)	13 (34.2%)	66 (41.5%)	44 (41.5%)
Dominant	13 (13.1%)	–	–	13 (8.2%)	2 (1.9%)
Monomorphic markers	32 (32.3%)	15 (68.2%)	17 (44.7%)	64 (40.3%)	49 (46.2%)
No amplification	6 (6.1%)	2 (9.1%)	8 (21.1%)	16 (10.0%)	11 (10.4%)
Total	**99**	**22**	**38**	**159**	**106**
A. hypogaea					
Polymorphic markers	48 (48.5%)	5 (22.7%)	4 (10.5%)	57 (35.8%)	8 (7.5%)
Monomorphic markers	42 (42.4%)	13 (59.1%)	27 (71.1%)	82 (51.6%)	82 (77.4%)
No amplification	9 (9.1%)	4 (18.2%)	7 (18.4%)	20 (12.6%)	16 (15.1%)
Total	**99**	**22**	**38**	**159**	**106**

The number of repeat units found in the 271 sequences ranged from 5 to 54 (ESM 1). Several studies have suggested that loci with longer repeats are much more likely to be variable (Weber 1990; Thomas and Scott 1993; Udupa et al. 1999; Burstin et al. 2001). For the six accessions of *A. hypogaea* there was a trend, with maximum polymorphism of SSRs with 21–25 repeat units (Fig. 1). Polymorphism for SSRs with 5–10 repeats was low at only 7.2%. For the progenitors of the mapping population, polymorphism varied less with repeat size, and even small SSRs with 5–10 repeats showed more than 40% polymorphism. In contrast, the polymorphism of apparently genic markers derived from libraries that had longer repeats, was similar between the two wild and six cultivated accessions.

For both wild and cultivated germplasm, AG/TC repeats are more polymorphic than AC/TG repeats (Table 2), similar effects of repeat type on polymorphism were also observed by Ferguson et al. (2004).

Primer pairs were developed in the present study by using clones from SSR-enriched genomic libraries and ESTs (159 and 106, respectively). Thus, it was possible to compare polymorphism levels detected by SSR markers from these two sources (Table 2). Generally the number of repetitions was much smaller in ESTs than in genomic sequences and correspondingly the polymorphism of ESTs was much lower in *A. hypogaea* than between the two wild species (7.5 and 35.8% respectively, see Table 2). It has been reported that EST-derived SSRs are less polymorphic than those derived from genomic libraries for other plant species, such as rice (Cho et al. 2000), sugarcane (Cordeiro et al. 2001), bread wheat (Eujayl et al. 2002; Gupta et al. 2003), barley (Thiel et al. 2003), and soybean (Song et al. 2004). This has been attributed to possible selection against alterations in these conserved sequences (Scott et al. 2000). However, for *A. duranensis* and *A. stenosperma* accessions, the percentages were identical: 41.5% for ESTs and for genomic libraries, considering only the codominant markers. The reasons for these differences are not clear, and need further investigation to be elucidated.

Characterization of genic content of libraries used for marker development

A. hypogaea has a large and poorly characterized genome. Comparative genetic mapping with other legumes will allow peanut research to begin benefitting from the knowledge of other genomes, especially those of the model legumes *Lotus japonicus* and *Medicago truncatula*. For such comparisons the placement of genic markers on the *Arachis* map is essential. This map includes EST and genome-derived markers from a number of sources. EST markers are derived from transcribed sequences that are almost entirely genic. Nevertheless, not all the homologies of these markers are easily identified, and from 106 only 67 (63.2%) gave significant BlastX hits.

The gene content of genomic libraries depends upon the way that they are made. Using BlastX we estimated the genic content of different libraries that was used to develop SSR markers for peanut (Table 4). The proportion of sequences giving BlastX hits varied from more than 20% to 0%. There was a clear trend, libraries made using restriction enzymes susceptible to methylation at CG or CNG sites are more gene-rich. The single exception being the relatively high gene content of the sequences from Moretzsohn et al. (2004). There the anomaly can be explained by the library being enriched exclusively for tri-repeat SSRs. All eight BlastX hits encode poly-amino acid tracts in their open reading frames. The top BlastX hits for *Arabidopsis* and legume proteins of all markers is in ESM 1.

Map construction

Out of the 433 SSR markers screened against *A. duranensis*, accession K7988 and *A. stenosperma*, accession V10309, 204 (46.8%) were polymorphic, with 170 segregating in a codominant fashion and 34 in a dominant 3:1 ratio (Fig. 2). Of the 170 codominant markers, 80 did not show distortions from the expected 1:2:1 segregation ratio ($P < 0.05$) and were initially used to establish the LG. Distorted and dominant markers were

Table 3 Primer pairs, a quality score for the amplification products, fragment size range, number of individuals analyzed (*n*), total number of alleles amplified (*A*), and gene diversity (GD) for 62 new polymorphic SSR markers identified for *A. hypogaea* in this study

Loci	Score	Size range	*n*	*A*	GD
AC2B3	3	296–308	14	6	0.720
AC2B5	3	200–212	15	5	0.680
AC2C8	2	146–220	15	7	0.828
AC2C12	3	204–212	16	5	0.685
TC1A1	2	202–222	13	6	0.735
TC1A2	3	240–276	16	11	0.815
TC1D2	3	242–278	16	4	0.458
TC1E1	3	154–248	14	11	0.892
TC2A2	3	194–212	15	7	0.800
TC2B1	2	190–215	11	4	0.610
TC2B9	3	190–200	16	5	0.645
TC2C7	2	182–210	16	3	0.492
TC2D6	3	196–224	15	10	0.910
TC2E5	3	102–122	15	8	0.864
TC3B5	3	248–270	12	4	0.710
TC3E2	3	270–310	16	9	0.708
TC3E5	3	358–370	16	5	0.798
TC3G1	2	250–256	15	4	0.745
TC3G5	3	118–130	14	2	0.423
TC3H2	3	280–300	15	4	0.671
TC3H7	2	246–270	16	9	0.863
TC4A2	3	280–284	16	3	0.411
TC4D2	2	202–240	12	5	0.786
TC4D9	2	220–240	14	2	0.389
TC4E10	3	300–344	16	8	0.841
TC4F12	3	220–232	15	6	0.763
TC4G2	2	130–166	11	6	0.784
TC4G10	2	205–220	15	3	0.543
TC4H2	2	184–192	12	4	0.725
TC4H7	2	166–220	13	8	0.742
TC5A6	2	164–200	14	7	0.825
TC5D6	2	174–184	13	5	0.782
TC6E1	3	154–186	15	9	0.802
TC6G9	3	132–146	12	6	0.746
TC6H3	3	210–228	14	8	0.754
TC7A2	2	308–320	15	6	0.706
TC7C6	3	148–176	16	9	0.827
TC7D3	3	142–148	15	4	0.579
TC7E4	3	290–300	16	4	0.694
TC7G10	3	110–142	15	4	0.451
TC7H11	2	340–360	14	7	0.847
TC9B8	3	98–108	15	8	0.777
TC9C12	3	256–300	15	7	0.855
TC9F4	3	122–142	15	6	0.800
TC9F10	3	286–320	15	10	0.871
TC11A2	3	284–292	14	6	0.825
TC11A4	3	172–204	15	12	0.906
TC11B11	2	144–146	16	2	0.121
TC11E4	2	110–136	16	12	0.909
TC11F12	2	182–188	14	2	0.254
TC11H6	3	190–214	16	8	0.847
GI-385	3	330–340	16	2	0.121
GI-832	3	200–210	15	2	0.480
GI-1107	3	360–384	14	5	0.680
RN0-615	2	390–400	15	3	0.625
RN0-681	2	310–350	16	6	0.520
RI1F06	3	312–372	11	8	0.823
RN3E10	3	252–255	11	2	0.416
RN17F12	2	260–278	12	5	0.812
Seq4E8	3	274–286	10	6	0.842
Seq4F10	2	180–188	9	4	0.706
Seq16C7	3	242–258	14	5	0.571
Average		–	14.41	5.87	0.690

Table 4 Genic content of different libraries as estimated by BlastX hits (E-values of $<10^{-8}$) against *Arabidopsis* and legume putative proteins

Library	Enzyme(s)	Cleaved methylation sites/cloned fragment CG:CNG	Selection/enrichment	Blast hits/total sequences (%)
EST	NA	NA	di-, tri-, tetra-, and penta-SSRs	67/106 (63.2%)
pPGP (Ferguson et al. 2004)	*Pst*I	0:4	GT GA AAC ATC AGT ATT CAC CTT CTG ATCC GATA AAAT	31/164 (20.7%)
TC	*Sau*3AI	1:1	TC	17/102 (16.7%)
Genomic (Moretzsohn et al. 2004)	*Tsp*509I	0:0	TTG	8/87 (9.2%)
AC	*Sau*3AI	1:1	AC	2/22 (9.1%)
PPGS (Ferguson et al. 2004)	*Sau*3AI/ *Bam*HI	0.5:0.5	GT GA AAC ATC AGT ATT CAC CTT CTG	7/288 (2.4%)
Genomic (He et al. 2003)	*Hin*dIII/*Mse*I	0:0	AT GT GA	0/56 (0.0%)

Data for ESTs are presented only for comparison to the genomic libraries. Cleaved methylation sites/cloned fragment presents the average number of methylation sites (CG and CNG sites, where cytosine methylation protects against restriction) cleaved per fragment cloned. For instance, *Pst*I cleaves CTGCAG, and so cleaves two CNG sites. Methylation at either of the cytosines protects *Pst*I sites against restriction, and since both ends of the cloned fragments are cut by *Pst*I, four CNG methylation sites are cut per cloned fragment. For *Sau*3AI that cleaves GATC, the cleavage of methylation sites depends upon the flanking DNA sequence, on average 0.5 CG and 0.5 CNG sites will be cleaved per *Sau*3AI fragment terminal. Sequence data from Ferguson et al. (2004) are the assembled SSR containing sequences made available in their supplementary data, and contains sequences that have not been used for marker development. Data from Moretzsohn et al. (2004) also includes some sequences that were not used for marker development. The data shows a clear trend, libraries made from DNA cleaved by restriction enzymes sensitive to methylation at CG and CNG sites are more gene-rich

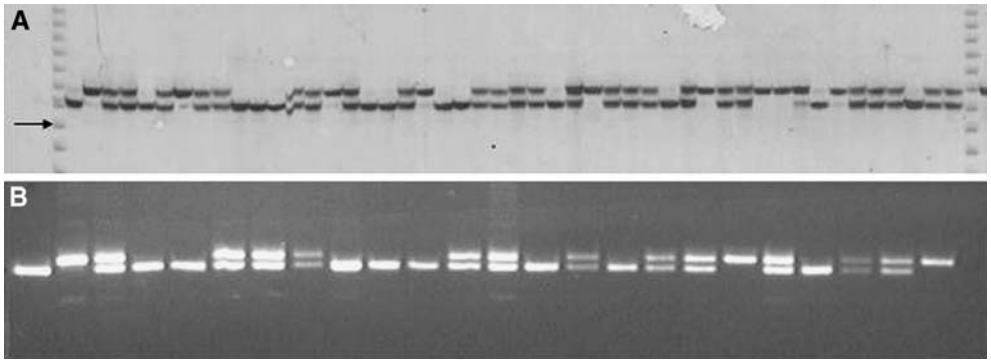

Fig. 2 Amplification patterns obtained with primer RN8C09 in 5% polyacrylamide gel electrophoresis (**a**) and with primer RN10F09 in 3.5% agarose gel (**b**). In panel A, the first and the last lanes are 10 bp DNA ladder. In both gels, from left to right: *A. duranensis* (accession K7988) used as the female progenitor, *A. stenosperma* (accession V10309) used as male, the F_1 hybrid, and 45 and 21 F_2 individuals in panels A and B, respectively. *Arrow* indicates a ladder fragment of 250 bp

further included in the map. Using a minimum LOD score of 12.0 and a maximum recombination fraction (θ) of 0.30, 170 markers mapped into 11 LG covering 1,230.89 cM of total map distance (Fig. 3). This total length is comparable to the 1,063 cM found in a map previously reported from a cross between two AA genome diploid species, *A. duranensis* and *A. cardenasii* (Halward et al. 1993) and to half of the 2,210 cM found for a published tetraploid map (Burow et al. 2001). A few markers were grouped with LOD scores <12.0 (from 6.0 to 12.0) and these are indicated in the map. Groups ranged from 20.07 cM (4 markers) to 280.10 cM (32 markers), with an average distance of 7.24 cM between adjacent markers. Linkage groups were numbered sequentially from the longest to the shortest. Eight SSR primer pairs amplified consistently two loci (AC2C2, TC1B2, TC1D1, TC4H2, TC7H9, Ah573, Ah745, and Seq15D6) and the 16 loci were included in the map. These markers were identified by the numbers −1 and −2 after the marker names (Fig. 3). The framework map, consisting of markers positioned with very high confidence, was composed of 74 markers covering 1,006.20 cM with an average distance of 13.59 cM between markers. The number of markers per linkage group in the framework map ranged from 3 (group 11) to 15 (group 2). Marker names that constitute the framework map were underlined in the map (Fig. 3). *Arachis duranensis* and *A. stenosperma* are diploid species with $2n=20$ chromosomes and, thus, 10 LG were expected instead of the 11 obtained in the present study. By reducing the stringency of the LOD score to 10.0, groups 1 and 2 merged in a single group, but linked with a very high recombination fraction (>40.0 cM). Thus, they were kept as separate groups. In addition, the upper portion of group 3 was joined with markers that constitute the framework map by decreasing the LOD score to 8.0. Thirty-four segregating markers remained unlinked, suggesting that at least parts of the genome have not been covered by this map. The inclusion of new markers would saturate this map, and probably result in the expected 10 LG.

A total of 104 (51.0%) out of the 204 markers deviated from the expected F_2 ratio of 1:2:1 (90 markers) or 3:1 (14 markers) at $P>0.05$ level. Of the 90 markers that showed distortion from the expected 1:2:1 segregation ratio, 41 markers were skewed toward the *A. duranensis* allele, 44 markers toward the *A. stenosperma* allele, and 5 markers toward the heterozygote. For the 14 dominant distorted markers, 5 showed an excess of *A. duranensis* alleles and 9 of *A. stenosperma* alleles. The proportion of distorted markers found in this study is higher than the 25% found for the two RFLP-based maps published for *Arachis* (Halward et al. 1993; Burow et al. 2001). Distorted markers were scattered on LG 1, 3, 4, 5, and 6 (Fig. 3). Groups 1, 3, and 6 had markers with an excess of both *A. duranensis* and *A. stenosperma* alleles, while all the distorted markers of group 4 and group 5 were skewed toward the *A. duranensis* and *A. stenosperma* alleles, respectively (Fig. 3). Segregation distortion has been described for most of the published molecular marker-based linkage maps in plants, and has been attributed to different causes, such as: deleterious recessive alleles (Berry et al. 1995), self-incompatibility alleles (Barzen et al. 1995), structural rearrangements (Quillet et al. 1995) or differences in DNA content (Jenczewski et al. 1997). Whatever the cause is, segregation distortion in some chromosomes, or in parts of chromosomes, must have important implications for using this map information in a MAS program. New genetic maps and cytogenetic experiments that are in progress will enable a better understanding and use of these findings.

Thirty-six of the 170 loci placed on the map are ESTs or characterized genes. Of the other 134 that were developed from genomic sequences, 25 gave BlastX hits. Therefore, about 61 of the mapped microsatellites most likely represent genes. As expected, homologies are diverse, but they include ones of possible interest for: seed quality (seed allergen and seed maturation protein); disease resistance (lectin, protease inhibitor, cinnamoyl-CoA reductase, and wound-induced protein); and drought resistance (dehydration responsive element binding protein, and aquaporin) (ESM 1).

Total genome size was estimated using the method of Hulbert et al. (1988). Only framework markers were used in this procedure to avoid an overestimate of genome coverage. A total map distance of 1,424.70 cM was estimated, of which 1,006.20 cM (70.6%) and 1,230.89 cM (86.4%) were covered by the framework map and by the total map constructed in this study, respectively. The haploid genome size of *A. hypogaea* was estimated as being 2,813 Mbp (Bennet and Leitch 2005). If wild species have similar genome sizes and given the estimated total map distance, the average physical equivalent of 1 cM would correspond to 1.97 Mbp in the *Arachis* genome.

Conclusions

This is the first microsatellite based and gene-rich linkage map published for *Arachis*. Since *A. hypogaea* has a relatively narrow genetic base, wild species of *Arachis* constitute an important source of genes for traits of interest to peanut breeding programs. The availability of moderate coverage, only microsatellite map will certainly represent a considerable improvement in our ability to map useful genes and to implement MAS. The validation of the LG obtained in the present study, both in terms of locus ordering and recombination fraction estimates, is now in progress using two other mapping populations: a diploid F_2 generated by crossing two BB genome species (*A. ipaënsis* × *A. magna*) and a tetraploid F_2 obtained by crossing an *A. hypogaea* accession (cv. IAC Runner 886) and an amphidiploid plant resulting from a cross between two wild diploid species (*A. duranensis* — AA genome x *A. ipaënsis* — BB genome) treated with colchicine. The development of genetic

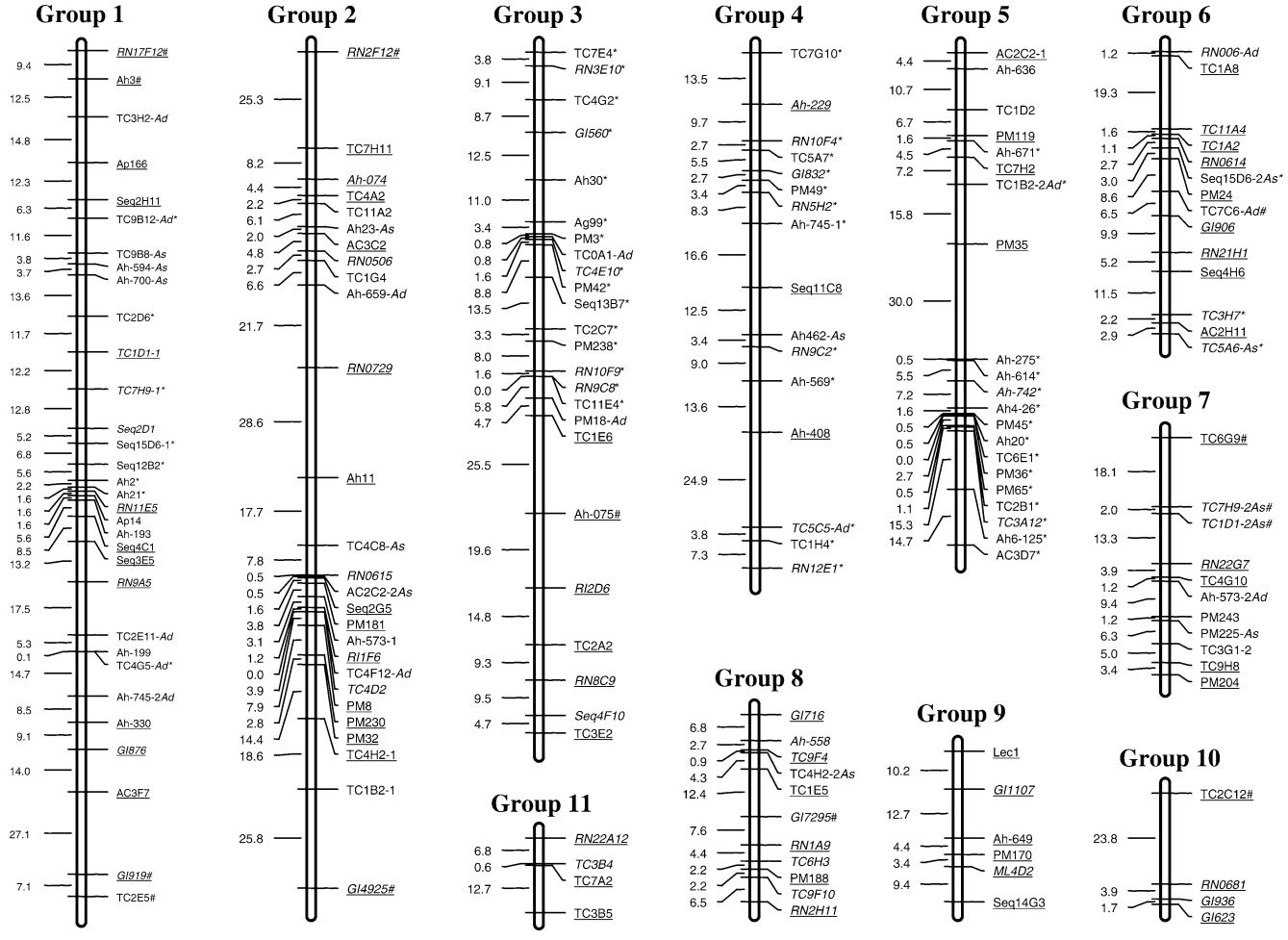

Fig. 3 Peanut genetic linkage map, based on SSR markers, obtained through the analysis of 93 F_2 plants, generated from a cross between two diploid wild species with AA genome, *A. duranensis* × *A. stenosperma*. Using a minimum LOD score of 12.0 and a maximum recombination fraction (θ) of 0.30, 170 markers mapped into 11 LG. Marker names that grouped with LOD scores < 12.0 have a symbol # after the marker name. Linkage groups were numbered sequentially from the longest to the shortest. Probable genic markers are shown *in italics*. Framework map was composed by 74 markers that could be ordered with a likelihood support > 1,000:1. Marker names that constitute the framework map were underlined in the map. *Numbers on the left* of each group are Kosambi map distances. Markers that showed significant distortions from the 1:2:1 segregation ratio are indicated by an *asterisk*. Dominant markers were identified by *Ad* or *As* after the marker name if they amplified alleles from *A. duranensis* or *A. stenosperma*, respectively. Markers that amplified two loci have numbers 1 and 2 after the marker name

maps for *Arachis* diploid wild species with AA and BB genomes will facilitate detailed inheritance studies of co-segregating microsatellites and traits of interest in the tetraploid population. Transferability of SSR loci to different species will allow the construction of a reference map for *Arachis* and the mapping of QTLs in different genetic backgrounds. Additionally, the present SSR-based map is the only published map for *Arachis* constructed with markers derived from sequences that are all currently available in GenBank. Thus, researchers from different laboratories can use these markers for map development using other mapping populations and contributing to the construction of a consensus map for *Arachis*. Markers linked to resistance genes are being identified to potentially implement MAS strategies in peanut-breeding programs, therefore accelerating the introgression of wild genes into cultivated peanut.

Acknowledgments The authors thank Raphael FTM Oliveira, Aryanne G Amaral, and Juliana Dias for their technical assistance. This research is part of the Ph.D. Thesis of Marcio de Carvalho Moretzsohn and has been supported by the European Union: INCO-DEV, Contract ICA4-CT-2001-10072 Project "ARAMAP" and by the World Bank and EMBRAPA "The Agricultural Technology Development Project for Brazil" (PRODETAB), project 004/01/01.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Barzen E, Mechelke W, Ritter E, Schulte-Kappert E, Salamini F (1995) An extended map of the sugar beet genome containing RFLP and RFLP loci. Theor Appl Genet 90:189–193

Bennet MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. Ann Bot 95:45–90

Berry ST, Leon AJ, Hanfrey CC, Challis P, Burkholz A, Barnes SJ, Rufener GK, Lee M, Caligari PDS (1995) Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. Theor Appl Genet 91:195–199

Bravo JP, Hoshino AA, Angelici CMLCD, Lopes CR, Gimenes MA (2005) Analysis of genetic variability in germplasm of Arachis section Arachis by using microsatellites. Gen Mol Biol (in press)

Burow MD, Simpson CE, Paterson AH, Starr JL (1996) Identification of peanut (*Arachis hypogaea* L.) RAPD markers diagnostic of root-knot nematode (*Meloidogyne arenaria* (Neal) Chitwood) resistance. Mol Breed 2:369–379

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

Burstin J, Deniot G, Potier J, Weinachter C, Aubert G, Baranger A (2001) Microsatellite polymorphism in *Pisum sativum*. Plant Breed 120:311–317

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 156:847–854

Castelo A, Martins WS, Gao G (2002) TROLL: tandem repeat occurrence locator. Bioinformatics J 18:634–636

Chavarriaga-Aguirre P, Maya MM, Bonierbale MW, Kresovich S, Fregene MA, Tohme J, Kochert G (1998) Microsatellites in Cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. Theor Appl Genet 97:493–501

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci 160:1115–1123

Creste S, Tulmann Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Mol Biol Rep 19:299–306

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–422

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

FAO (2003) http://faostat.fao.org/faostat/collections?subset = agriculture

Ferguson ME, Burow MD, Schulze SR, Bramel PJ, Paterson AH, Kresovich S, Mitchell S (2004) Microsatellite identification and characterization in peanut (*A. hypogaea* L.). Theor Appl Genet 108:1064–1070

Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392–1400

Gao LF, Tang JF, Li HW, Jia JZ (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 12:245–261

Garcia GM, Stalker HT, Kochert G (1995) Introgression analysis of an interspecific hybrid population in peanuts (*Arachis hypogaea* L.) using RFLP and RAPD markers. Genome 38:166–176

Garcia GM, Stalker HT, Shroeder E, Kochert G (1996) Identification of RAPD, SCAR, and RFLP markers tightly linked to nematode resistance genes introgressed from *Arachis cardenasii* into *Arachis hypogaea*. Genome 39:836–845

Gimenes MA, Hoshino AA, Barbosa AVG, Palmieri DA, Bravo JP, Lopes CR (2005) Characterization and transferability of microsatellite markers from the cultivated peanut (*Arachis hypogaea* L.). Bot J Linn Soc (in press)

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. Genetics 137:1121–1137

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genomics 270:315–323

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Halward TM, Stalker HT, Kochert G (1993) Development of an RFLP linkage map in diploid peanut species. Theor Appl Genet 87:379–384

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

He G, Meng R, Newman M, Gao G, Pittman RN, Prakash CS (2003) Microsatellites as DNA markers in cultivated peanut (A. hypogaea L.). BMC Plant Biol 3:3. http://www.biomedcentral.com/1471-2229/3/3

Herselman L, Thwaites R, Kimmins FM, Courtois B, van der Merwe PJA, Seal SE (2004) Identification and mapping of AFLP markers linked to peanut (*Arachis hypogaea* L.) resistance to the aphid vector of groundnut rosette disease. Theor Appl Genet 109:1426–1433

Hopkins MS, Casa AM, Wang T, Mitchell SE, Dean R, Kochert GD, Kresovich S (1999) Discovery and characterization of polymorphic Simple Sequence Repeats (SSRs) in peanut. Crop Sci 39:1243–1247

Hulbert SH, Ilott TW, Legg EJ, Lincoln SE, Lander ES, Michelmore RW (1988) Genetic analysis of the fungus, *Bremia lactucae*, using Restriction Fragment Length Polymorphisms. Genetics 120:947–958

Jenczewski E, Gherardi M, Bonnin I, Prosperi JM, Olivieri I, Huguet T (1997) Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). Theor Appl Genet 94:682–691

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Khlestkina EK, Than MH, Pestsova EG, Roder MS, Malyshev SV, Korzun V, Borner A (2004) Mapping of 99 new microsatellite-derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. Theor Appl Genet 109:725–732

Krapovickas A, Gregory WC (1994) Taxonomía del género Arachis (Leguminosae) Bonplandia 8:1–186

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic maps of experimental and natural populations. Genomics 1:174–181

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10:72–80

Moretzsohn MC, Hopkins MS, Mitchell SE, Kresovich S, Valls JFM, Ferreira ME (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. http://www.biomedcentral.com/1471-2229/4/11

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

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

Palmieri DA, Hoshino AA, Bravo JP, Lopes CR, Gimenes MA (2002) Isolation and characterization of microsatellite loci from the forage species *Arachis pintoi* (Genus *Arachis*). Mol Ecol Notes 2:551–553

Palmieri DA, Bechara MD, Curi RA, Gimenes MA, Lopes CR (2005) Novel polymorphic microsatellite markers in section *Caulorrhizae* (*Arachis*, Fabaceae). Mol Ecol Notes 5:77–79

Prasad M, Varshney RK, Roy JK, Balyan HS, Gupta PK (2000) The use of microsatellites for detecting DNA polymorphism genotype identification and genetic diversity in wheat. Theor Appl Genet 100:584–592

Provan J, Powell W, Waugh R (1996) Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). Theor Appl Genet 92:1078–1084

Quillet MC, Madjidian N, Griveau Y, Serieys H, Tersac M, Lorieux M, Bervillé A (1995) Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. Theor Appl Genet 91:1195–1202

Rafalski JA, Vogel JM, Morgante M, Powel W, Andre C, Tingey SV (1996) Generating and using DNA markers in plants. In: Birren B, Lai E (eds) Analysis of non-mammalian genomes — a practical guide. Academic, New York, pp 75–134

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning — A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Scott KD, Eggler P, Seaton G, Rosetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Song QJ, Marek LF, Shoemaker RC, Lara KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

Staden R, Judge DP, Bonfield JK (2003) Managing sequencing projects in the GAP4 environment. In: Krawetz SA, Womble DD (eds) Introduction to Bioinformatics. A Theoretical and practical approach. Human Press Inc, Totawa

Tian AG, Wang J, Cui P, Han YJ, Xu H, Cong LJ, Huang XG, Wang XL, Jiao YZ, Wang BJ, Wang YJ, Zhang JS, Chen SY (2004) Characterization of soybean genomic features by analysis of its expressed sequence tags. Theor Appl Genet 108:903–913

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites (STSs). Theor Appl Genet 86:985–990

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

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546

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

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

Young ND, Weeden NF, Kochert G (1996) Genome mapping in legumes (Family Fabaceae). In: Paterson AH (ed) Genome mapping in plants. RG Landes, Austin, TX