Genetic diversity of peanut (Arachis hypogaea L.) and its wild relatives based on the analysis of hypervariable regions of the genome

Marcio de Carvalho Moretzsohn
Mark S. Hopkins
Sharon E. Mitchell
Stephen Kresovich
University of South Carolina - Columbia, sk@sc.edu
Jose Francisco Montenegro Valls

See next page for additional authors

Follow this and additional works at: https://scholarcommons.sc.edu/biol_facpub

Part of the Biology Commons

Publication Info
DOI: 10.1186/1471-2229-4-11
© BMC Plant Biology, 2004, BioMed Central
http://creativecommons.org/licenses/by/2.0/
Research article

Genetic diversity of peanut (Arachis hypogaea L.) and its wild relatives based on the analysis of hypervariable regions of the genome

Marcio de Carvalho Moretzsohn*1, Mark S Hopkins2, Sharon E Mitchell3, Stephen Kresovich3, Jose Francisco Montenegro Valls1 and Marcio Elias Ferreira1,4

Abstract

Background: The genus Arachis is native to a region that includes Central Brazil and neighboring countries. Little is known about the genetic variability of the Brazilian cultivated peanut (Arachis hypogaea, genome AABB) germplasm collection at the DNA level. The understanding of the genetic diversity of cultivated and wild species of peanut (Arachis spp.) is essential to develop strategies of collection, conservation and use of the germplasm in variety development. The identity of the ancestor progenitor species of cultivated peanut has also been of great interest. Several species have been suggested as putative AA and BB genome donors to allotetraploid A. hypogaea. Microsatellite or SSR (Simple Sequence Repeat) markers are co-dominant, multiallelic, and highly polymorphic genetic markers, appropriate for genetic diversity studies. Microsatellite markers may also, to some extent, support phylogenetic inferences. Here we report the use of a set of microsatellite markers, including newly developed ones, for phylogenetic inferences and the analysis of genetic variation of accessions of A. hypogaea and its wild relatives.

Results: A total of 67 new microsatellite markers (mainly TTG motif) were developed for Arachis. Only three of these markers, however, were polymorphic in cultivated peanut. These three new markers plus five other markers characterized previously were evaluated for number of alleles per locus and gene diversity using 60 accessions of A. hypogaea. Genetic relationships among these 60 accessions and a sample of 36 wild accessions representative of section Arachis were estimated using allelic variation observed in a selected set of 12 SSR markers. Results showed that the Brazilian peanut germplasm collection has considerable levels of genetic diversity detected by SSR markers. Similarity groups for A. hypogaea accessions were established, which is a useful criteria for selecting parental plants for crop improvement. Microsatellite marker transferability was up to 76% for species of the section Arachis, but only 45% for species from the other eight Arachis sections tested. A new marker (Ah-041) presented a 100% transferability and could be used to classify the peanut accessions in AA and non-AA genome carriers.
**Background**

The genus *Arachis*, family Fabaceae, is native to South America, probably from a region including Central Brazil and Paraguay [1]. There are 69 species described in the genus, assembled into nine sections, according to the morphology, geographic distribution and crossability [2]. Some of these species have been used for forage in South America, but the most important species is the cultivated peanut, *Arachis hypogaea* L. This crop is widely grown in more than 80 countries in the Americas, Asia and Africa [3] and is used for both human consumption and as a source of oil. World production is increasing and has reached 37 million tons (in the shell) and 5.8 million tons of oil, according to FAO estimates [4]. Peanut now ranks fifth among the world oil production crops.

Nearly all *Arachis* species are diploid, but cultivated peanut is an allotetraploid (genome AABB). It is a member of the section *Arachis*, which also includes about 25 diploid and one tetraploid wild species (*A. monticola*) [2]. *Arachis hypogaea* is classified based on the presence or absence of flowers on the main axis into two subspecies, *hypogaea* and *fastigiata*. These two subspecies were further classified into six botanical varieties based on morphology and growth habits [2]. Subspecies *hypogaea* was divided in two botanical varieties, *hypogaea* and *hirsuta*, while ssp. *fastigiata* in the varieties *fastigiata*, *vulgaris*, *aequatoriana* and *peruviana*. The identity of the progenitor species of cultivated peanut has been of great interest. Several species have been suggested as putative A and B genome donors [reviewed by [5-7]]. RFLP analysis that included 17 diploid species of the section *Arachis* and three *A. hypogaea* accessions suggested a single origin for domesticated peanut and ancestral species related to *A. duranensis* (A genome) and *A. ipaënsis* (B genome) as the most likely progenitors of *A. hypogaea* [5]. *In situ* hybridization analysis of six diploid species and one *A. hypogaea* accession, and RAPD and ISSR (Inter-simple sequence repeat) analyses of 13 *A. hypogaea* accessions and 15 wild species, however, suggested *A. villosa* (A genome) and *A. ipaënsis* (B genome) as the progenitors of cultivated peanut [6,7].

Cultivated peanut exhibits a considerable amount of variability for various morphological, physiological, and agronomic traits. However, little variation has been detected at the DNA level using techniques such as RAPDs (Random Amplified Polymorphic DNAs), AFLPs (Amplified Fragment Length Polymorphisms) and RFLPs (Restriction Fragment Length Polymorphisms) [5,8-15]. The low level of variation in cultivated peanut has been attributed to three causes or to combinations of them: (1) barriers to gene flow from related diploid species to domesticated peanut as a consequence of the polyploidization event [16]; (2) recent polyploidization, from one or a few individual(s) of each diploid parental species, combined with self-pollination [8]; or (3) use of few elite breeding lines and little exotic germplasm in breeding programs, resulting in a narrow genetic base [17,18]. Recently, some studies revealed DNA polymorphism in *A. hypogaea* using SSRs (Simple Sequence Repeats), AFLPs, RAPDs and ISSRs [7,19-23].

Little is known about the genetic variability of the Brazilian *A. hypogaea* germplasm collection at the DNA level. Knowledge of the genetic variation of peanut accessions is important for their efficient use in breeding programs, for studies on crop evolution, and for conservation purposes. Molecular marker analysis, joined to phenotypic evaluation, is a powerful tool for grouping of genotypes based on genetic distance data and for selection of progenitors that might constitute new breeding populations. In the context of *ex situ* conservation, it has been demonstrated that molecular markers are very useful for the management of germplasm collections [24]. SSRs are ideal tools for such studies as they are PCR-based markers, genetically defined, typically co-dominant, multiallelic, and uniformly dispersed throughout plant genomes. SSRs have been used in plants for many genetic applications, including the assessment of genetic variability in germplasm collections or pedigree reconstruction [reviewed by [25]]. In *Arachis*, SSR markers have been recently developed and proved to be useful for accession discrimination and assessment of genetic variation [19,22,23]. Moreover,
since little genetic variability has been detected in cultivated peanut, the use of a polymorphic marker, such as SSRs, in addition to distinguishing closely related genotypes, should also be useful for phylogenetic studies, as demonstrated in other crops, such as wheat [26], melon [27], potato [28], and coffee [29].

The objectives of the present work were: (1) to develop new SSR markers for genetic analysis of cultivated peanut (A. hypogaea), (2) to employ a set of SSR markers to analyze the genetic variation among wild and cultivated peanut accessions of the Brazilian germplasm collection, and (3) to evaluate the cross-species transferability of SSR markers and their usefulness in phylogenetic studies of the genus Arachis.

Results and Discussion

SSR development and screening of markers

Microsatellite enriched genomic libraries of A. hypogaea were constructed in order to develop new SSR markers for the species. Digestion of the A. hypogaea genomic DNA with five different enzymes (AluI, MseI, Rsal, Sau3AI, and Tsp509I) revealed that Tsp509I produced the most adequate profile for library development, with fragments ranging from 200 to 800 bp in size. Four libraries were initially constructed based on trinucleotide repeat motifs (TTG, TGG, ATG, and ATC). Hybridization analysis revealed that the TTG/AAC repeats were more abundant in the peanut genome than the other tested motifs (TTG > TGG > ATG > ATC). This is in agreement with a survey of the peanut genome than the other tested motifs (TTG > TGG > ATG > ATC). Therefore, the TTG library was used for SSR development for A. hypogaea. Screening of 750 clones by anchored PCR showed that 162 had SSRs (21.6%) and these were sequenced. Out of the 162 positive clones sequenced, there were 91 unique (non-redundant) sequences (56.2% of the sequenced clones) but only 67 of them were suitable for primer designing (41.4% of the positive clones). The design of primers for the other 24 unique SSRs identified was not possible due to the occurrence of very short tandem repeats (< 5 units) or low GC content of the regions flanking the SSR. The anchored PCR screening prior to sequencing significantly improved the yield of useful clones. Few false-positive clones were obtained (less than 10% of the sequenced clones). The percentage of primers designed, in relation to the number of clones sequenced (41.4%), indicated that the method used was relatively efficient for the discovery and development of SSR markers for peanut.

The 67 SSR markers (see Additional File 1) were screened for polymorphism on seven samples, including five varieties of A. hypogaea, one accession of A. ipaënsis and one accession of A. duranensis. Of these, 62 markers (92.5%) generated clearly interpretable PCR products, but only three markers (Ah-041, Ah-193 and Ah-558) were polymorphic in cultivated peanut (Table 1). Four other markers (Ah-075, Ah-229, Ah-522 and Ah-657) showed to be invariant for the five A. hypogaea accessions, but were polymorphic in A. ipaënsis and A. duranensis accessions. Since ancestral species related to these two species are considered potential progenitors of the AABB genome, these seven markers were included in the analysis of the genetic relationships between accessions of the Brazilian germplasm collection (see below). The other 55 markers were not polymorphic in A. hypogaea accessions and were not considered for further analysis. They have been useful, however, for studies with diploid wild species of peanut currently under development (data not shown).

SSR marker characterization

Marker locus duplication

Four (Ah-193, Ah-229, Ah-522, and Ah-657) of the seven newly developed markers produced single fragments both in the tetraploid accessions and in the diploid species belonging to the Arachis section (see Additional File 2), what indicates that each of them amplified alleles at single loci. Two markers (Ah-075 and Ah-558) produced one or two fragments in the diploid species and three (Ah-075) or four (Ah-558) fragments in the tetraploid accessions. These results are consistent with allele amplification on duplicated loci. The presence of the three alleles (150/144/138 bp) amplified by marker Ah-075 in the tetraploid accessions was confirmed by the identification of diploid species homozygous for each of these three alleles (for example, accessions K 30006, Sv 3806, V 6389, V 10309) or heterozygous for the three pairs of alleles (for example, accessions K 30076, K 30097, V 14167). Similar results were observed for marker Ah-558, but one (244 bp) of the six alleles (244/241/235/232/229/226 bp) detected in the tetraploid accessions was not found in any of the tested diploid wild species. The remaining marker (Ah-041) amplified one (292 bp) or two of three alleles (300/292/280 bp) in accessions of cultivated peanut and in the diploid species (Figure 1). This indicates that either the individuals tested are highly heterozygous for this locus or this marker locus is duplicated in the cultivated genome. Since A. hypogaea is an allotetraploid and preferentially an autogamous species, with only 2.5% outcrossing [31], the latter hypothesis is more probable. Five other markers previously developed [19] were also analyzed for marker locus duplication, and two of them (Ah4-04 and Ah4-24) seem to represent a single locus in the genome. Therefore, five out of 12 markers (41.7%) probably amplified duplicated loci in most of the cultivated peanut accessions. This proportion, although not representative, indicates that locus duplication in peanut might be higher than found for other polyploid species, such as wheat [32], apple [33], cassava [34], and sweet
Table 1: SSR loci characterization. Primer pairs, repeat motifs, range of fragment sizes, total number of alleles (A), and gene diversity (h) estimates based on the analysis of 60 Arachis hypogaea accessions for the eight polymorphic SSR markers.

<table>
<thead>
<tr>
<th>Identification</th>
<th>Motif</th>
<th>Size range (bp)</th>
<th>A**</th>
<th>h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ah-041*</td>
<td>(AAC)<em>{12} (CAA)</em>{10}</td>
<td>280–292</td>
<td>2</td>
<td>0.468</td>
</tr>
<tr>
<td>Ah-193*</td>
<td>(AAC)<em>{16} (GA)</em>{14}</td>
<td>444–462</td>
<td>8</td>
<td>0.718</td>
</tr>
<tr>
<td>Ah-558*</td>
<td>(AAC)_{6}</td>
<td>226–244</td>
<td>6</td>
<td>0.520</td>
</tr>
<tr>
<td>Ah4-04</td>
<td>(GA)_{19}</td>
<td>84–108</td>
<td>7 (6)</td>
<td>0.702</td>
</tr>
<tr>
<td>Ah4-20</td>
<td>(GA)_{19}</td>
<td>201–215</td>
<td>4 (3)</td>
<td>0.646</td>
</tr>
<tr>
<td>Ah4-24</td>
<td>(ATA)_{12}</td>
<td>406–418</td>
<td>5 (5)</td>
<td>0.761</td>
</tr>
<tr>
<td>Ah4-26</td>
<td>(CT)_{12} imperfect</td>
<td>156–213</td>
<td>15 (8)</td>
<td>0.743</td>
</tr>
<tr>
<td>Lec-1</td>
<td>(AT)_{18}</td>
<td>212–294</td>
<td>27 (14)</td>
<td>0.929</td>
</tr>
</tbody>
</table>

**Enclosed in parenthesis, the number of fragments detected by Hopkins et al. [19].

Figure 1
Microsatellite polymorphism (marker Ah-041) in Arachis species visualized in silver-stained denaturing polyacrilamide gel. Samples are: (1) A. duranensis-V14167; (2) A. ipaënsis; (3) A. magna-V13760; (4) A. batizocoi-K9484; (5) A. cardenasii; (6) A. stenosperma-V10229; (7) A. magna-K30097; (8) A. helodes; (9) A. hoehnei; (10) A. batizocoi-K9484m; (11) A. villosa; (12) A. microsperma; (13) A. simpsonii; (14) A. monticola; (15) A. hypogaea fastigiata fastigiata; (16) A. hypogaea fastigiata vulgaris; (17) A. hypogaea fastigiata peruiana; (18) A. hypogaea hypogaea hypogaea; (19) A. hypogaea hypogaea hypogaea, Xingu type; (20) A. stenosperma-V10309; (21–23) Hybrids K7988 × V10309; (24) A. duranensis-K7988; (25) A. diogo; (26) A. duranensis-G10038; (27) A. kempf-merca-; (28) Arachis sp. 6389; (29) A. cruziana; (30) A. praecox; (31) A. valida.

Microsatellite polymorphism (marker Ah-041) in Arachis species visualized in silver-stained denaturing polyacrilamide gel. Samples are: (1) A. duranensis-V14167; (2) A. ipaënsis; (3) A. magna-V13760; (4) A. batizocoi-K9484; (5) A. cardenasii; (6) A. stenosperma-V10229; (7) A. magna-K30097; (8) A. helodes; (9) A. hoehnei; (10) A. batizocoi-K9484m; (11) A. villosa; (12) A. microsperma; (13) A. simpsonii; (14) A. monticola; (15) A. hypogaea fastigiata fastigiata; (16) A. hypogaea fastigiata vulgaris; (17) A. hypogaea fastigiata peruiana; (18) A. hypogaea hypogaea hypogaea; (19) A. hypogaea hypogaea hypogaea, Xingu type; (20) A. stenosperma-V10309; (21–23) Hybrids K7988 × V10309; (24) A. duranensis-K7988; (25) A. diogo; (26) A. duranensis-G10038; (27) A. kempf-merca-; (28) Arachis sp. 6389; (29) A. cruziana; (30) A. praecox; (31) A. valida.

Marker allele diversity
Allelic diversity for a sample of 60 A. hypogaea accessions (Additional File 2) was estimated for the three newly developed polymorphic markers (Ah-041, Ah-193, Ah-558) and five other SSR markers (Ah4-04, Ah4-20, Ah4-24, Ah4-26 e Lec-1), previously developed for cultivated peanut [19] (Table 1). Marker Ah6-125 [19] was not included in the analysis since it was found to be invariant for the 60 accessions of cultivated peanut. The number of alleles detected ranged from 2 to 27 at each of the eight polymorphic loci, with an average of 8.4 alleles per locus. Gene diversity values ranged from 0.468 to 0.929, with an average value of 0.683. For polyploid organisms, gene diversity is an estimate of the probability that two randomly chosen genes from a population are different and it is not related to the expected proportion of heterozygous, as for diploid organisms in Hardy-Weinberg equilibrium [36]. The values obtained in the present study are probably overestimated, since some of the markers amplified alleles on duplicated loci. Thus, they cannot be compared to those obtained for diploid species, but they are similar to values found for other polyploid species such as cassava [34], potato [37] and wheat [38]. Moreover, average h values estimated by RFLP markers were considerably lower for A. hypogaea [10]. These results demonstrate the usefulness of microsatellite markers for genetic diversity analysis of cultivated peanut.

Marker Ah-041 and genome specific alleles
One of the markers developed in the present study (Ah-041) presented interesting banding patterns (Figure 1) on AA and non-AA genomes of diploid wild species. Genome AA is characterized by the presence of a pair of chromosomes (called AA) that are considerably smaller than the other chromosomes [39], while genome BB lacks this small chromosome pair. All AA genome species, previously characterized by cytological analysis, showed a 292 bp allele when genotyped at the Ah-041 locus, while all BB genome species and A. glandulifera (DD genome) showed a 280 bp allele. Arachis hypogaea accessions showed both alleles (292/280) or they had only the 292 bp allele. Arachis monticola, the other tetraploid species of section Arachis, also had both 292 and 280 bp alleles. Accessions of five AA genome species (A. stenosperma, A. villosa, A. microsperma, A. simpsonii and A. kempf-merca-
one BB genome species (A. valida) and A. glandulifera showed a 300 bp band. Marker Ah-041 allowed allele detection on all the 114 accessions tested, with a 100% transferability in the genus. Species from the sections Erectoides, Extranervosae, and Triseminatae had different alleles (277 and 274 bp). Allele 280 bp was also found in species from the sections Trirectoides, Heteranthae, Procumbentes, and Rhizomatoseae. Only section Arachis species showed the allele 292 bp (AA genome). Although allelic segregation and genome mapping of this marker locus has not yet been examined in controlled crosses, our results suggest that the 292 bp allele is specific to AA genome species. If so, this marker may be valuable for genetic studies of peanut, including phylogenetic inferences in the genus Arachis and studies about genome origin (see below). It should be noticed that the absolute transferability of marker Ah-041 to all Arachis species suggests that this SSR sequence is positioned in or near coding regions. However, search on the National Center for Biotechnology Information (NCBI) database http://www.ncbi.nlm.nih.gov/BLAST/ did not find any significantly complete or partially homologous sequence.

Transferability of SSR markers

Twelve SSR markers developed for Arachis hypogaea (Ah-041, Ah-193, Ah-558, Ah-075, Ah-229, Ah-522, Ah-657, Ah-4-04, Ah-4-20, Ah-4-24, Ah-4-26 and Lec-1) were tested on 54 accessions of wild species of Arachis, representing the nine sections of the genus. The transferability of the twelve markers was up to 76% for species of the section Arachis, but ranged from 23% (Triseminatae) to 62% (Procumbentes), with an average of 45%, for species of the other eight Arachis sections (see Additional File 3). Transferability of SSR markers between related species is a consequence of the homology of flanking regions of the microsatellites and the size of the region between the primer pair amenable to amplification by PCR. Other studies have demonstrated the conservation of SSR sequences in plants, as reviewed by Gupta and Varshney [25]. The possibility of using microsatellite markers developed for one species in genetic evaluation of other species is controversial, since the development of microsatellite markers is still expensive and time consuming. The SSR markers developed in this study could be very useful for genetic analysis of wild species of Arachis, including comparative genome mapping, population genetic structure and phylogenetic inferences among species.

Relationships between Arachis spp. accessions

Genetic similarities among accessions were estimated by shared allele distance in pairwise comparisons of 60 A. hypogaea accessions and 36 other accessions of wild species belonging to section Arachis (accessions 1 to 96 – Additional File 2), using the same set of 12 markers described above. Among the 60 A. hypogaea accessions, the average genetic distance was 0.336. Some accessions, such as Tatu, Tatu2 e Pd2622, shared all the alleles detected with the 12 SSR markers.

A dendrogram based on the neighbor joining method was constructed for the 96 accessions of the section Arachis (Figure 2). Considering initially the A. hypogaea accessions, two main clusters were evident. Group I contained all 32 fastigiata/fastigiata, all four fastigiata/vulgaris accessions, the only hypogaea/hirsuta accession (Mf 1538) and one accession not identified to the subspecies level (Wi 632). The two fastigiata/aequatoriana accessions included in the study (Mf 1678 and Mf1640) were also clustered in this group. The two fastigiata/peruviana accessions (Mf 1560 e Sv 429) formed a separate subgroup within Group I. Another subgroup was formed by a fastigiata/aequatoriana accession (Mf 1640), two variant forms of hypogaea/hypogaea accessions (Nambiquarae, accession As 436 and Malhado, accession V 12577) and V 12549, identified as "ssp hypogaea Xingu type", because of its morphological differences from the Virginia type. All 45 accessions of Group I had the alleles 292 bp and 280 bp at the Ah-041 locus. The other group (Group II) contained solely the 15 Virginia type (hypogaea/hypogaea) accessions examined in the study. The accessions of this group had only the 292 bp allele (Figure 1). Polymorphism at the Ah-041 locus could be caused by mutation at the primer binding site in the Group II accessions, preventing the detection of the 280 bp PCR product, or by an insertion/deletion event. If the Ah-041 locus is duplicated in allotetraploid peanut, these results could indicate that cultivated peanut did not have a single origin since Group I accessions could be derived from hybridization of AA and non-AA ancestral diploid species followed by genome duplication, while Group II accessions could be derived from straight genome duplication of an AA diploid species. Multiple origin hypotheses, based on cytological, crossing and isozyme studies, have been previously suggested to explain the cultivated peanut origin [40-42].

The taxonomic status of A. hypogaea ssp. fastigiata var. peruviana is controversial. Studies based on AFLP [21] as well as RAPD and ISSR [7] markers have shown that accessions of this variety have greater similarity to ssp. hypogaea rather than ssp. fastigiata, to which they are currently thought to belong [2]. In the present study, the two fastigiata/peruviana accessions grouped with the other fastigiata accessions and not with the 15 hypogaea/hypogaea accessions clustered in Group II. The data, therefore, corroborates the current classification of this botanical variety. However, the classification of hypogaea/hirsuta and the three variant forms of hypogaea/hypogaea (Nambiquarae, Malhado and Xingu types) remains controversial, since these four accessions did not cluster in Group II.
Figure 2

Dendrogram based on allele sharing genetic distances of 60 *A. hypogaea* and 36 *Arachis* wild species generated by the neighbor joining method. The letters, after each *A. hypogaea* accession number, refer to the subspecies, varieties, and types: FF-fastigiata/fastigiata; FV-fastigiata/vulgaris; FA-fastigiata/aequatoriana; FP-fastigiata/peruviana; HH-hypogaea/hypogaea; HHN-hypogaea/hirsuta; HHN-hypogaea/hypogaea Nambiquarae type; HX-hypogaea/hypogaea "Malhado" type; HX-hypogaea Xingu type. Ni, corresponds to a non-identified accession to the subspecies level. The letters after the wild species accession names refer to the genome type (A or B), according to the patterns obtained with marker Ah-041. The D genome of *A. glandulifera* was defined by Stalker, 1991.
The SSR markers used in the present study detected genetic variability in *A. hypogaea* accessions, despite the known narrow genetic base of cultivated peanut. In the two main clusters (Group I and Group II), some similarity groups, genetically distant from each other, were formed. These groups were not related to collection sites, seed color, or any other known trait. However, the establishment of similarity groups, jointly with morphologic and agronomic data, can be used in the selection and crossing of superior plants, allowing for a more efficient use of the existing genetic variability conserved in the collection.

Considering the wild species, the analysis showed that, in general, the species containing the same genome clustered together. However, accessions of *A. duranensis* clustered with *A. hypogaea* and *A. monticola* accessions (Figure 2). This result corroborates the hypothesis that an ancestral species related to *A. duranensis* or *A. duranensis* itself could be the donor of the AA genome to *A. hypogaea*. In the group composed of BB genome species there is a cluster of four species (*A. ipaënsis, A. magna, A. williamsii* and *Arachis* sp. 6389). It has recently been noticed that plants of these four species are morphologically very similar and produce fertile hybrids. A revision of their taxonomic status is currently under evaluation by JFM Valls and CE Simpson.

The genetic distance analysis based on SSR marker polymorphism supports previous estimates of high genetic similarity between *A. monticola* accessions and *A. hypogaea* accessions (Figure 2). Only one (Ah-522) out of the 12 SSR loci included in the present study showed alleles detected in *A. hypogaea* not observed in *A. monticola*. Other studies with molecular data have also shown the little differentiation between these two taxa [8,9,11,42]. Moreover, hybrids between *A. monticola* and *A. hypogaea* are fertile [2]. These facts have led some researchers to suggest that *A. monticola* and *A. hypogaea* should not be considered as separate species [11]. Our results, based on SSR markers and only one accession of *A. monticola*, are certainly insufficient to define the taxonomic status of *A. monticola*. However, they indicated that this species could be directly related to the allotetraploid ancestral progenitor of *A. hypogaea*, as suggested by several studies [6,7,43].

**Conclusions**

The present study showed the utility of microsatellite markers for the detection of polymorphisms among cultivated peanut accessions and for the genetic relationship analysis between *A. hypogaea* accessions and wild species of section *Arachis*. New SSR markers were developed for genetic analysis of peanut, but the percentage of markers that detected polymorphism among accessions of *A. hypogaea* was low, corroborating the notion that cultivated peanut presents a relatively reduced variation at the DNA level. However, a set of informative SSR markers detected considerable levels of genetic variability in the Brazilian Peanut Germplasm Collection. Based on this data, cultivated peanut was classified into two major groups, one composed of solely the Virginia type (*hypogaea/hypogaea*) accessions and another defined by *fastigiata/fastigiata, fastigiata/vulgaris, hypogaea/hirsuta, fastigiata/aequatoriana, fastigiata/peruviana* and two variant forms of *hypogaea/hypogaea* accessions. The identification of similarity groups could be useful for the selection of parental plants to be used in the breeding programs. Marker transferability was up to 76% for species of the section *Arachis*, but ranged from 23% (*Triseminatae*) to 62% (*Procumbentes*), with an average of 45%, for species of the other eight *Arachis* sections. The possibility of using microsatellite markers developed for one species in genetic evaluation of other species greatly reduces the cost of the analysis, since the development of microsatellite markers is still expensive and time consuming. The SSR markers developed in this study could be very useful for genetic analysis of wild species of *Arachis*, including comparative genome mapping, population genetic structure and phylogenetic inferences among species. A marker (Ah-041) was identified that allows the discrimination of AA from non-AA genome accessions of peanut.

**Methods**

**Plant material**

A single *A. hypogaea* plant was used for DNA library construction (accession UF 91108). Five *A. hypogaea* samples (accessions 3274, 3294, 3310, 3324, and UF 91108), one *A. ipaënsis* (accession 3309) and one *A. duranensis* (accession 4454) were used to test the effectiveness of SSR markers to detect polymorphism at the DNA level. These plants were obtained from USDA-ARS Plant Genetic Resources Conservation Unit (Griffin, GA) peanut collection.

A total of 114 accessions of wild and cultivated peanut were included in the study (Additional File 2). Ninety-six accessions belonging to section *Arachis* were used for genetic diversity analysis. Of these, 60 are *A. hypogaea* accessions, 47 of which collected in Brazil, representing both *A. hypogaea* ssp. *hypogaea* and ssp. *fastigiata*, and their six varieties. The other 36 accessions belong to 27 wild species, including representatives of four new taxa that are being described by JFM Valls and CE Simpson. The wild species include tetraploid *A. monticola* (AABB), 13 diploid (2n = 20) species showing the small chromosome pair [44] which are associated to the AA genome of *A. hypogaea*, the diploid *A. glandulifera*, for which a DD genome has been suggested [45], two diploids with 2n = 18 [46] and ten species with 20 chromosomes lacking the small chromosome pair typical of the AA genome, which are classified as BB genome species (Additional File 2). Other 18 accessions of wild species belonging to eight *Arachis* sections were included in the analysis to test the
transferability of SSR markers. Three hybrid plants of the 
A. stenosperma (V 10309) × A. duranensis (K 7988) cross and one hybrid plant of the A. appressipila (G 10002) × A. repens (Nc 1579) cross were also used to verify the heritability of alleles in interspecific crosses. The 114 accessions and the four hybrid plants were obtained from the Brazilian Peanut Germplasm Collection, maintained at Embrapa Genetic Resources and Biotechnology – CENARGEN (Brasilia-DF, Brazil). All plants were grown from seed or cuttings (sections Rhizomatosa and Caulorrhiza) under greenhouse conditions at CENARGEN prior to DNA extraction.

**DNA extraction**

Total genomic DNA used for library construction was extracted and purified on cesium chloride gradients following standard protocols [47]. For the other samples, total genomic DNA was extracted according to published protocol [48], with some modifications: Proteinase K (20 mg/ml) was added to the extraction buffer and an additional step of polysaccharide precipitation with 2 M NaCl was included.

**Library construction**

Total genomic DNA libraries enriched for trinucleotide repeats were constructed following a standard protocol with minor modifications [49]. Peanut total genomic DNA was digested with five different enzymes (AluI, MseI, Rsal, Sau3AI, and Tsp509I) 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 S&S NA 45 DEAE cellulose membranes, after electrophoresis in 1.5% low-melting agarose gel. DNA was resuspended in TE buffer and ligated to Tsp509I adapters. Four genomic libraries enriched for trinucleotide repeats were constructed. Fragments were selected by hybridization with biotinylated trinucleotide repeats ((TTG)$_o$ (TTG)$_o$ (ATG)$_o$ (ATC)$_o$) and recovered by magnetic beads linked to streptavidine. Enriched fractions were amplified by PCR, using a primer complementary to the adapters. Amplification products were purified using Wizard PCR Prep DNA purification system (Promega, Madison, WI), cloned into λ ZAP II phagemid vector (Stratagene) and then transformed into E. coli strain XL1-Blue. Transformed cells were grown on LB-Amp plates (50 µg/ml ampicillin) at 37°C, until plaques were between 0.5 and 1.0 mm in diameter. Individual clones were picked up from the plates and phage were eluted in SM buffer (500 µl) with a drop of chloroform. Clone DNAs were then amplified by PCR using primers complementary to the vector DNA flanking the insert (T3 or SK and T7 or KS primers, Stratagene, CA), and electrophoresed in 1.5% agarose gels in 0.5X TBE buffer to determine insert sizes. To determine the position of the SSRs relative to the vector cloning site, an anchored PCR strategy was performed for each positive clone [49], using T3 (or SK) and T7 (or KS) primers separately plus another primer complementary to the target SSR.

**DNA sequencing and primer design**

Clones containing repeats that were not directly adjacent to the cloning site were amplified, treated with ExoI and shrimp alkaline phosphatase enzymes, and PCR products were sequenced on an ABI 377 automated DNA sequencer (Applied Biosystems, CA, USA), using fluorescent dye-terminator chemistry. Redundant sequences were identified using the software Sequencer (Gene Codes Corporation). 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). Primer design parameters were as follows: (i) Tm ranged from 60°C to 65°C, (ii) 1°C difference in Tm between primer pairs; (iii) GC content ranging from 45% to 55%, (iv) absence of complementarity between primers, (v) default values for the other parameters.

**Primer screening and PCR amplification of SSR loci**

Primers were initially screened for polymorphisms in a set of five cultivated and two wild peanut accessions (3274, 3294, 3310, 3324, Uf 91108, 3309, 4454). PCRs were performed in 13 µl volumes, containing 1X PCR buffer (10 mM Tris-HCl pH8.3, 50 mM KCl, 1.5 mM MgCl$_2$), 0.2 mM of each dNTP, 1 unit Taq DNA polymerase, DMSO 50% (1.3 µl), 5 pmol of each primer and 10 ng of genomic DNA. Amplifications were performed using either a 9600 System (Applied Biosystems, CA, USA) or a PTC-100 (MJ Research, MA, USA) thermal cycler, with the following conditions: 94°C for 2 min (1 cycle), 94°C for 1 min, 55–66°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. PCR products were resolved on 3.5% Metaphor agarose (FMC Bioproducts, ME, USA) gels stained with ethidium bromide. For genotype determination, the amplified products were separated on 4% denaturing polyacrylamide gels stained with silver nitrate [50]. Fragment sizes were estimated by comparison with a 10-bp DNA ladder standard (Gibco/BRL, MD, USA).

**Data analysis and transferability of SSR loci**

Eight selected SSR markers were characterized for number of alleles per locus and gene diversity [51] using 60 accesses of Arachis hypogaea (accessions 1 to 60 – Additional File 2). The markers included three developed in the present study and five other SSR markers developed for A. hypogaea [19]. Gene diversity (h) at a marker locus was estimated according to the formula: $h = 1 - \sum(p_i)^2$, where
Additional material

Additional File 1
List of the 67 microsatellite markers for genetic analysis of Arachis spp. Marker sequences, repeat motifs, and annealing temperature (Tm) for the 67 newly developed SSR markers
Click here for file

Additional File 2
Arachis spp. accessions analyzed in this study. Numbers in parenthesis, after the specific names, correspond to numbers included in the dendrogram (Figure 2). *Genome types of Arachis hypogaea accessions were defined according to the genotypes obtained with marker Ah-041.
Click here for file

Additional File 3
Transferability of SSR markers developed for A. hypogaea across 45 wild species of Arachis. This file contains a table that lists the markers that amplified visible products for each of the 49 accessions of the Arachis wild species included in the study.
Click here for file

Acknowledgements
The authors thank Dr. R. E. Dean for computational assistance and L. Zhang and T. M. Jenkins for their technical laboratory assistance. We also thank M.R. Bertozzo for generously providing the DNA samples of some wild species of Arachis. Development of SSR marker research supported by grants from the USDA. Characterization of SSR markers supported by Embrapa Recursos Genéticos e Biotecnologia, Brazil.

References
45. Lewis PO, Zaykin D: Genetic Data Analysis: Computer program for data. Version 1.0 (d1c6). 2001 [http://lewis.eeb.uconn.edu/lewis/home/software.html]. Free program distributed by the authors over the Internet.
48. Genetic Distance Calculator [http://www2biology.uibk.ac.at/].