Genetic diversity of cultivated and wild-type peanuts evaluated with M13-tailed SSR markers and sequencing

NOELLE A. BARKLEY[1], ROB E. DEAN[1], ROY N. PITTMAN[1], MING. L. WANG[1]*,
CORLEY C. HOLBROOK[2] AND GARY A. PEDERSON[1]

[1] *USDA-ARS, Plant Genetic Resources Conservation Unit, 1109 Experiment Street, Griffin, GA 30223, USA*
[2] *USDA-ARS, Crop and Breeding Research Unit, Coastal Plain Experiment Station, P. O. Box 748, Tifton, GA 31793, USA*

Summary

Thirty-one genomic SSR markers with a M13 tail attached were used to assess the genetic diversity of the peanut mini core collection. The M13-tailed method was effective in discriminating almost all the cultivated and wild accessions. A total of 477 alleles were detected with an average of 15·4 alleles per locus. The mean polymorphic information content (PIC) score was 0·687. The cultivated peanut (*Arachis hypogaea* L.) mini core produced a total of 312 alleles with an average of 10·1 alleles per locus. A neighbour-joining tree was constructed to determine the interspecific and intraspecific relationships in this data set. Almost all the peanut accessions in this data set classified into subspecies and botanical varieties such as subsp. *hypogaea* var. *hypogaea*, subsp. *fastigiata* var. *fastigiata*, and subsp. *fastigiata* var. *vulgaris* clustered with other accessions with the same classification, which lends further support to their current taxonomy. Alleles were sequenced from one of the SSR markers used in this study, which demonstrated that the repeat motif is conserved when transferring the marker across species borders. This study allowed the examination of the diversity and phylogenetic relationships in the peanut mini core which has not been previously reported.

1. Introduction

Peanut (*Arachis hypogaea* L.) is an important oilseed crop which is widely consumed by humans. In the United States the public consumes approximately 2·4 billion pounds (1·1 billion kg) of peanuts per year (The Peanut Institute, www.peanut-institute.org). Argentina, China and the United States are the largest exporters of peanuts while the European Union and Asia tend to be the largest peanut importers (Revoredo & Fletcher, 2002). Peanut oil, peanut butter and peanut seeds are highly nutritious for human consumption. The seeds contain approximately 45–51% oil and most of their oil content consists of oleic and linoleic acids (Lopez *et al.*, 2000). Humans whom consume a high amount of mono-unsaturated fat, such as that found in peanut, tend to have reduced LDL cholesterol, lower triglycerides and improved HDL cholesterol. Peanuts are also a good source of protein, vitamin E, folate, fibre and phytochemicals.

Even though peanut is known to be an important agricultural commodity, genomic and molecular studies have lagged behind those on other legumes such as soybean. Only recently have a fairly abundant number of SSR markers become available for genotyping peanut accessions (Hopkins *et al.*, 1999; He *et al.*, 2003; Ferguson *et al.*, 2004; Moretzsohn Mde *et al.*, 2004). Additionally, previous studies have reported some difficulties in employing markers such as AFLPs, RFLPs and RAPDs that could distinguish cultivated peanut varieties (Kochert *et al.*, 1991; Subramanian *et al.*, 2000; Gimenes *et al.*, 2002; Herselman, 2003). The deficiency of molecular variation in cultivated peanut is somewhat surprising since there is a large range of variation in morphological characters such as seed size, seed coat colour, maturation time and disease resistance (Hopkins *et al.*, 1999). The lack of polymorphic markers has hindered cultivated peanut, in comparison with other crops,

* Corresponding author. Tel: +1 (770) 229342. Fax: +1 (770) 2293323. e-mail: mwang@ars-grin.gov

being enhanced by molecular techniques such as marker-assisted selection (MAS), resistance gene cloning, genetic mapping and evolutionary studies (He *et al.*, 2003).

Peanuts are native to South America and are classified in the legume family (*Fabaceae*) in the genus *Arachis*, which consists of about 70 species (Krapovickas & Gregory, 1994). Almost all the wild species are diploid ($2n = 20$) with the exception of *A. monticola*, while cultivated peanut (*Arachis hypogaea* L.) is an allotetraploid ($2n = 4x = 40$). *Arachis hypogaea* is further classified into two subspecies (*hypogaea* and *fastigiata*) and six botanical varieties (*hypogaea*, *hirsuta*, *fastigiata*, *vulgaris*, *aequatoriana* and *peruviana*) (Krapovickas & Gregory, 1994). These distinctions are made based on the presence (subsp. *fastigiata*) or absence (subsp. *hypogaea*) of flowers on the main axis and various other morphological traits such as growth habit, pod shape and pod reticulation.

The putative ancestral progenitors of cultivated allotetraploid peanuts have been of great interest to breeders and peanut researchers. *A. hypogaea*'s genome is characterized as AABB. The AA genome contains a pair of chromosomes (AA) that are considerably smaller than the other chromosomes, whereas a species characterized with a BB genome lacks these small chromosomes (Moretzsohn Mde *et al.*, 2004). Currently, it is believed that cultivated peanut originated from a single hybridization event between wild species *A. duranensis* (AA genome) and *A. ipaensis* (BB genome) followed by a chromosome duplication to produce an AABB genome (Kochert *et al.*, 1996; Hopkins *et al.*, 1999). This hybridization and chromosome duplication event isolated cultivated peanut from sharing genes with its wild relatives, and natural introgression from wild species has not been demonstrated (Hopkins *et al.*, 1999).

The USDA-ARS Plant Genetic Resources Conservation Unit (USDA-ARS PGRCU) maintains a large peanut collection ($\sim 10\,000$ accessions). A core collection (831 accessions) was constructed to represent the majority of genetic diversity with minimal redundancy (Holbrook *et al.*, 1993) and to help researchers evaluate germplasm and screen for traits of interest more efficiently. This core was designed by using information available on country of origin and morphological characters to select approximately 10% of the samples from the entire collection that would maximize genetic diversity. A mini core was subsequently constructed so that traits which are difficult or expensive to measure could be assayed on a small scale (Holbrook & Dong, 2005). To date, there have been no published studies on the molecular characterization/diversity of the peanut mini core and only recently have an abundant number of SSR markers became available for researchers to assay genetic

diversity in peanut. Therefore, the objectives of this study were to: assay the genetic diversity of the mini core utilizing the M13-tailed SSR method, determine whether SSR markers would produce PCR products in the wild relatives, sequence a few alleles from an SSR marker to determine whether a repeat motif is present in wild peanuts, use molecular and observational data to putatively classify cultivated peanut accessions into subspecies and/or botanical varieties, and determine the relationships between and among various peanut species.

2. Materials and methods

(i) *Plant material and DNA extraction*

Taxonomic classifications of the peanut accessions used in this study were determined by using GRIN (Germplasm Resources Information Network) taxonomy (http://www.ars-grin.gov/npgs/). Peanut seeds were obtained from the USDA-ARS PGRCU peanut germplasm collection located in Griffin, GA (Table 1). The seeds were germinated by wrapping them in a wet paper towel which was exposed to ethylene gas. Once germinated the seedlings were transferred to 1 gallon (4·5 l) pots filled with soil and allowed to develop for approximately 8 weeks. Young unopened leaves (approximately 100 mg) were harvested from each plant. The leaves were placed in a 2 ml screw-cap microcentrifuge tube with an o-ring along with two 3 mm tungsten carbide beads (Qiagen Valencia, CA). The tissue was pulverized with a Retsch Mixer Mill 301 (Leeds, UK) for 3 min at 30 hertz. All DNA samples were extracted using EZNA Plant DNA kit from Omega Bio-Tek (Doraville, GA). Peanut is a self-pollinated crop and plants within the same accession are uniform. To save greenhouse space, tissues from a single plant were used for DNA extraction. DNA concentration was measured with a DyNA Quant 200 fluorometer purchased from Hoefer Pharmacia Biotech (San Francisco, CA). Samples were also loaded on a 1% agarose gel along with a quantitative marker from Invitrogen (Carlsbad, CA) to confirm DNA concentration obtained with the fluorometer. All samples were subsequently diluted to 10 ng/μl for PCR.

(ii) *M13-tailed SSR method*

A M13 primer 5′ CGTTGTAAAACGACGGCC-AGT 3′ with 6-FAM covalently bound to the 5′ end for detection on the ABI 377 was purchased from Qiagen (Valencia, CA). The two unlabelled primers in each reaction consisted of a specific SSR-targeting forward primer with a 5′ M13 tail (CGTTGTAAA-ACGACGGCCAGT) and a specific SSR-targeting reverse primer. All PCR reactions consisted of three

Table 1. *Current taxonomic classification for all peanut accessions that were used in this study as determined by GRIN*

ID no.	PI no.	Market type	Ploidy	Genus and species	Subspecies	Variety	Origin
1	295730	Valencia	4X	*A. hypogaea*			Burma
2	493329	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Argentina
3	493356	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Argentina
4	493547	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Argentina
5	493581	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Argentina
6	493631	Mixed	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Argentina
7	493693	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Argentina
8	493717	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Argentina
9	493729	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Argentina
10	493880	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Argentina
11	493938	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Argentina
12	494018	Mixed	4X	*A. hypogaea*	*fastigiata*	*vulgaris*	Argentina
13	494034	Mixed	4X	*A. hypogaea*	*fastigiata*	*vulgaris*	Argentina
14	475863	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Bolivia
15	497318	Runner	4X	*A. hypogaea*	*hypogaea*	*hypogaea*	Bolivia
16	497395	Runner	4X	*A. hypogaea*	*hypogaea*	*hypogaea*	Bolivia
17	497517	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Brazil
18	496401	Valencia	4X	*A. hypogaea*			Burkina Faso
19	496448	Valencia	4X	*A. hypogaea*			Burkina Faso
20	504614	Mixed	4X	*A. hypogaea*			Colombia
21	497639	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Ecuador
22	497668	Mixed	4X	*A. hypogaea*			Ecuador
23	502037	Valencia	4X	*A. hypogaea*			Peru
24	502040	Valencia	4X	*A. hypogaea*			Peru
25	502111	Valencia	4X	*A. hypogaea*			Peru
26	502120	Valencia	4X	*A. hypogaea*			Peru
27	494795	Virginia	4X	*A. hypogaea*			Zambia
28	331314	Mixed	4X	*A. hypogaea*			Argentina
29	339960	Valencia	4X	*A. hypogaea*			Argentina
30	331297	Mixed	4X	*A. hypogaea*			Argentina
31	274193	Runner	4X	*A. hypogaea*			Bolivia
32	290560	Spanish	4X	*A. hypogaea*			India
33	290620	Mixed	4X	*A. hypogaea*			India
34	290566	Valencia	4X	*A. hypogaea*			India
35	290594	Mixed	4X	*A. hypogaea*			India
36	290536	Valencia	4X	*A. hypogaea*			India
37	343398	Virginia	4X	*A. hypogaea*			Israel
38	343384	Mixed	4X	*A. hypogaea*			Israel
39	371521	Runner	4X	*A. hypogaea*			Israel
40	200441	Runner	4X	*A. hypogaea*			Japan
41	196635	Spanish	4X	*A. hypogaea*			Madagascar
42	259851	Runner	4X	*A. hypogaea*			Malawi
43	355271	Runner	4X	*A. hypogaea*			Mexico
44	372271	Mixed	4X	*A. hypogaea*			Nigeria
45	399581	Virginia	4X	*A. hypogaea*			Nigeria
46	337406	Valencia	4X	*A. hypogaea*			Paraguay
47	159786	Spanish	4X	*A. hypogaea*			Senegal
48	268696	Runner	4X	*A. hypogaea*			South Africa
49	298854	Runner	4X	*A. hypogaea*			South Africa
50	268868	Valencia	4X	*A. hypogaea*			Sudan
51	313129	Mixed	4X	*A. hypogaea*			Taiwan
52	155107	Mixed	4X	*A. hypogaea*			Uruguay
53	162655	Mixed	4X	*A. hypogaea*			Uruguay
54	152146	Mixed	4X	*A. hypogaea*			Brazil
55	262038	Valencia	4X	*A. hypogaea*			Brazil
56	337293	Runner	4X	*A. hypogaea*			Brazil
57	270907	Mixed	4X	*A. hypogaea*			Zambia
58	270905	Mixed	4X	*A. hypogaea*			Zambia
59	268996	Runner	4X	*A. hypogaea*			Zambia
60	270998	Mixed	4X	*A. hypogaea*			Zambia
61	268806	Mixed	4X	*A. hypogaea*			Zambia

Table 1. (*Cont.*)

ID no.	PI no.	Market type	Ploidy	Genus and species	Subspecies	Variety	Origin
62	268755	Runner	4X	*A. hypogaea*			Zambia
63	270786	Mixed	4X	*A. hypogaea*			Zambia
64	356004	Mixed	4X	*A. hypogaea*			Argentina
65	259658	Mixed	4X	*A. hypogaea*			Canada
66	259617	Mixed	4X	*A. hypogaea*			Canada
67	288146	Spanish	4X	*A. hypogaea*			India
68	288210	Runner	4X	*A. hypogaea*			India
69	319768	Runner	4X	*A. hypogaea*			Israel
70	296550	Runner	4X	*A. hypogaea*			Israel
71	296558	Runner	4X	*A. hypogaea*			Israel
72	295250	Mixed	4X	*A. hypogaea*			Israel
73	295309	Mixed	4X	*A. hypogaea*			Israel
74	370331	Mixed	4X	*A. hypogaea*			Israel
75	259836	Spanish	4X	*A. hypogaea*			Malawi
76	325943	Valencia	4X	*A. hypogaea*			Venezuela
77	338338	Valencia	4X	*A. hypogaea*			Venezuela
78	157542	Runner	4X	*A. hypogaea*			China
79	158854	Valencia	4X	*A. hypogaea*			China
80	271019	Mixed	4X	*A. hypogaea*			Zambia
81	268586	Mixed	4X	*A. hypogaea*			Zambia
82	403813	Valencia	4X	*A. hypogaea*			Argentina
83	475918	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Bolivia
84	475931	Mixed	4X	*A. hypogaea*	*hypogaea*	*hypogaea*	Bolivia
85	408743	Mixed	4X	*A. hypogaea*			Brazil
86	461427	Valencia	4X	*A. hypogaea*			China (PRC)
87	478819	Mixed	4X	*A. hypogaea*			India
88	481795	Mixed	4X	*A. hypogaea*			Mozambique
89	476636	Mixed	4X	*A. hypogaea*			Nigeria
90	476596	Runner	4X	*A. hypogaea*			Nigeria
91	372305	Runner	4X	*A. hypogaea*			Nigeria
92	476432	Runner	4X	*A. hypogaea*			Nigeria
93	476025	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Peru
94	240560	Runner	4X	*A. hypogaea*			South Africa
95	292950	Mixed	4X	*A. hypogaea*			South Africa
96	162857	Virginia	4X	*A. hypogaea*			Sudan
97	407667	Mixed	4X	*A. hypogaea*			Thailand
98	478850	Valencia	4X	*A. hypogaea*			Uganda
99	482189	Mixed	4X	*A. hypogaea*			Zimbabwe
100	471952	Mixed	4X	*A. hypogaea*			Zimbabwe
101	442768	Runner	4X	*A. hypogaea*			Zimbabwe
102	482120	Runner	4X	*A. hypogaea*			Zimbabwe
103	471954	Valencia	4X	*A. hypogaea*			Zimbabwe
104	429420	Valencia	4X	*A. hypogaea*			Zimbabwe
105	468271	Runner	4X	*A. hypogaea*	*hypogaea*	*hypogaea*	Bolivia
106	461434	Runner	4X	*A. hypogaea*			China (PRC)
107	319770	Runner	4X	*A. hypogaea*			Israel
108	196622	Runner	4X	*A. hypogaea*			Ivory Coast
109	355268	Mixed	4X	*A. hypogaea*			Mexico
110	337399	Runner	4X	*A. hypogaea*			Morocco
111	323268	Mixed	4X	*A. hypogaea*			Pakistan
112	475914	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Bolivia
113	468191	Virginia	4X	*A. hypogaea*	*hypogaea*	*hypogaea*	Argentina
114	628577	Valencia	4X	*A. hypogaea*	*fastigiata*	*peruviana*	Guatemala
115	497615	Valencia	4X	*A. hypogaea*	*fastigiata*	*aequatoriana*	Ecuador
116	576638	Valencia	4X	*A. hypogaea*	*hypogaea*	*hirsuta*	Mexico
117	502096	Valencia	4X	*A. hypogaea*	*fastigiata*	*peruviana*	Peru
118	602090	Spanish	4X	*A. hypogaea*	*fastigiata*	*vulgaris*	Sri Lanka
119	560927	Valencia	4X	*A. hypogaea*	*fastigiata*	*fastigiata*	Bolivia
120	536180	Spanish	4X	*A. hypogaea*	*fastigiata*	*vulgaris*	Brazil
121	536276	Virginia	4X	*A. hypogaea*	*hypogaea*	*hypogaea*	Brazil
122	497484	wild	2X	*A. duranensis*			Bolivia
123	468203	wild	2X	*A. duranensis*			Argentina
124	468321	wild	2X	*A. duranensis*			Bolivia

Table 1. (*Cont.*)

ID no.	PI no.	Market type	Ploidy	Genus and species	Subspecies	Variety	Origin
125	468322	wild	2X	*A. ipaensis*			Bolivia
126	AT201	Runner	4X	*A. hypogaea*	*hypogaea*	*hypogaea*	USA
127	632380	Runner	4X	*A. hypogaea*	*hypogaea*	*hypogaea*	USA
128	DP-1	Runner	4X	*A. hypogaea*	*hypogaea*	*hypogaea*	USA
129	Hull	Runner	4X	*A. hypogaea*	*hypogaea*	*hypogaea*	USA
131	476011	wild	2X	*A. cardenasii*			Bolivia
132	476012	wild	2X	*A. cardenasii*			Bolivia
133	468354	wild	2X	*A. diogoi*			Paraguay
134	468354	wild	2X	*A. diogoi*			Paraguay
136	468369	wild	4X	*A. glabrata*		*glabrata*	Paraguay
137	338305	wild	4X	*A. glabrata*		*hagenbeckii*	Argentina
138	276194	wild	2X	*A. guaranitca*			Brazil
139	476142	wild	2X	*A. tuberosa*			Brazil
140	468181	wild	2X	*A. paraguariensis*	*paraguariensis*		Brazil
141	468196	wild	4X	*A. monticola*			Argentina

Identification numbers 1–111 are from the peanut mini core designed by Holbrook & Dong (2005). The remaining samples are botanical varieties and wild accessions. Three accessions do not currently have an assigned PI number.

primers in which the M13-labelled primer and reverse primer were in excess of the forward primer, which was limited. This allows the forward M13-tailed primer and reverse primer to initiate the reaction and, when the limited primer is depleted, the labelled primer takes the place of the limited forward primer in the remaining PCR cycles (Schuelke, 2000). Forward and reverse SSR primer sequences (Table 2) were obtained from previously published peanut studies (Hopkins *et al.*, 1999; He *et al.*, 2003; Ferguson *et al.*, 2004; Moretzsohn Mde *et al.*, 2004).

The PCR reaction consisted of $1 \times$ PCR buffer, 1·5 mM $MgCl_2$, 0·2 mM dNTPs, 0·04 units *Taq* DNA polymerase (Promega Madison, WI), 0·04 μM forward primer, 0·16 μM reverse primer, 0·16 μM M13-labelled primer (Qiagen Valencia, CA), 0·6 ng DNA and dH_2O. The total volume of the reaction was 12·5 μl. A few primer sets required additional $MgCl_2$ in the reaction to produce clear uniform bands. A final concentration of 2·25 mM $MgCl_2$ was used for the following primer sets: pPGSseq3A8, pPGSseq3A1, pPGSseq16F1, pPGSseq2F5, pPGSseq2G4, pPGS-seq11G3, pPGSseq19F4 and PM210. Primer sets pPGSseq19D9, pPGSseq7G2, pPGSseq8E12 and PM50 had a final concentration of 3 mM $MgCl_2$. All PCR reactions were performed in a Perkin Elmer 9600 thermocycler in a 96-well plate. All primer sets used in this study were amplified using the same cycling conditions. The programme consisted of 1 cycle at 94 °C for 5 min for the initial denaturing, 30 cycles of 94 °C for 30 s, 56 °C for 45 s and 72 °C for 45 s, 8 cycles of 94 °C for 30 s, 53 °C for 45 s and 72 °C for 45 s, 1 cycle of 72 °C for 10 min for final extension and a 4 °C hold for temporal storage.

(iii) *PCR product separation*

Before loading samples on a gel, PCR products (1·0 μl) were mixed with 1·25 μl loading dye and 0·75 μl of GeneScan 500 TAMRA internal lane standard containing 16 fragments for size analysis (ABI, Foster City, CA) and the samples were denatured at 95 °C for 3 min. All PCR products were separated on polyacrylamide gels connected to an ABI 377 Automated DNA Sequencer (Foster City, CA). The gels were 36 cm in length and 0·20 mm thick. Each gel contained 9 g of urea, 2·5 ml of $10 \times$ TBE, 2·5 ml Long Ranger Gel Solution (Cambrex Rockland, ME) and water to obtain a final volume of 25 ml. Polymerization occurred by the addition of 125 μl of 10 % ammonium persulfate and 17·5 μl of TEMED to the gel solution. The gel was polymerized for 2 h before it was used for electrophoresis.

(iv) *Allele sequencing*

PCR products were amplified as described above and run on a 3 % agarose gel to verify that only a single band was produced. The PCR product was treated with 1 μl of Exonuclease I (10 U/μl) and 1 μl of shrimp alkaline phosphatase (1 U/μl) (GE Healthcare, Piscataway, NJ) for every 12 μl of PCR reaction to digest single-stranded DNA and cleave the 5′ phosphate. The PCR product was also cleaned with a Qiagen PCR cleanup kit (Valencia, CA) to remove excess nucleotides, primers, enzymes and other impurities. Then, 1 μl of the cleaned product was run on an agarose gel with a quantitative marker (Invitrogen Carlsbad, CA) to determine concentration and thus

Table 2. *List of primer sequences used in this study*

Primer name	Forward sequence	Reverse sequence
pPGSseq12B6	CGTTGTAAAACGACGGCCAGTGGCAGGCATGCTCAGATATT	AAAGAATGCTTGTGTATATCATCCC
pPGSseq13A10	CGTTGTAAAACGACGGCCAGTAACTCGCTTGTACCGGCTAA	AGGAATAATAACAATACCAACAGCA
pPGSseq15C12	CGTTGTAAAACGACGGCCAGTACAATGCAATGACCGTTGTT	TTGTTGCATGAGAACGTGAA
pPGSseq5E11	CGTTGTAAAACGACGGCCAGTACATGACAGAGCACAATGGC	TTGCTCAAAGAGAACACCAA
pPGSseq15F12	CGTTGTAAAACGACGGCCAGTAAAGTCAACCGCTCACACTG	AGGGTTAGGATTTTGGGTGG
pPGSseq16F1	CGTTGTAAAACGACGGCCAGTTGCTTCCATCAGCTTTTCCT	AAATGAGGGCCTCCAAAGTT
pPGSseq19D6	CGTTGTAAAACGACGGCCAGTTTTGTTATGCTCACACCCCA	AAAAATGAAGCAATATTTTGTTGTTAG
pPGSseq19D9	CGTTGTAAAACGACGGCCAGTTGTTGCCCACTGTTCTAATCA	TCAAATGGCATAGTCTCCCC
pPGPseq8D9	CGTTGTAAAACGACGGCCAGTTGAGTTTCCCCAAAAGGAGA	CAACAACAATACGGCCAACA
pPGPseq2E6	CGTTGTAAAACGACGGCCAGTTACAGCATTGCCTTCTGGTG	CCTGGGCTGGGGTATTATTT
pPGPseq2D12B	CGTTGTAAAACGACGGCCAGTAAGCTGAACGAACTCAAGGC	TGCAATGGGTACAATGCTAGA
pPGPseq3A8	CGTTGTAAAACGACGGCCAGTATACGTGACTTGGGCCAGAC	AGTGAAAAATACACCCAACGAA
pPGPseq7G2	CGTTGTAAAACGACGGCCAGTACTCCCGATGCACTTGAAAT	AACCTCTGTGCACTGTCCCT
pPGPseq3A1	CGTTGTAAAACGACGGCCAGTATCATTGTGCTGAGGGAAGG	CACCATTTTTCTTTTTCACCG
pPGPseq2G4	CGTTGTAAAACGACGGCCAGTTTCTTGGTTCCTTTGGCTTC	TGCTCAAGTGTCCTTATTGGTG
pPGPseq2F5	CGTTGTAAAACGACGGCCAGTTGACCAAAGTGATGAAGGGA	AAGTTGTTTGTACATCTGTCATCG
pPGSseq11G3	CGTTGTAAAACGACGGCCAGTCCGCGTTGTTAAACCAGAAC	ATGGAGGATGTGAGTGGGAA
pPGPseq8E12	CGTTGTAAAACGACGGCCAGTTCTGTTGAGAACCACCAGCA	GTGCTAGTTGCTTGACGCAC
pPGSseq19F4	CGTTGTAAAACGACGGCCAGTCCCATGATAAATTTGTATATCAAGCA	TCAACCACAGAAGACGACGA
PM036	CGTTGTAAAACGACGGCCAGTACTCGCCATAGCCAACAAAC	CATTCCCACAACTCCCACAT
PM050	CGTTGTAAAACGACGGCCAGTCAATTCATGATAGTATTTTATTGGACA	CTTTCTCCTCCCCAATTTGA
PM003	CGTTGTAAAACGACGGCCAGTGAAAGAAATTATACACTCCAATTATGC	CGGCATGACAGCTCTATGTT
PM183	CGTTGTAAAACGACGGCCAGTTTCTAATGAAAACCGACAAGTTT	CGTGCCAATAGAGTTTTATACGG
PM032	CGTTGTAAAACGACGGCCAGTAGTGTTGGGTGTGAAAGTGG	GGGACTCGGAACAGTGTTTATC
PM238	CGTTGTAAAACGACGGCCAGTCTCTCCTCTGCTCTGCACTG	ACAAGAACATGGGGATGAAGA
PM210	CGTTGTAAAACGACGGCCAGTCCGCAGATCTTCTCCTGTGT	CCTCCTCATCCTCTAAACTCTGC
PM137	CGTTGTAAAACGACGGCCAGTAACCAATTCAACAAACCCAGT	GAAGATGGATGAAACGGATG
Ah-041	CGTTGTAAAACGACGGCCAGTCGCCACAAGATTAACAAGCACC	GCTGGGATCATTGTAGGGAAGG
AH558	CGTTGTAAAACGACGGCCAGTTGTGACACCATCAATCAAAGGG	CAAAACCCAAATCATCACCACC
Ah4-26	CGTTGTAAAACGACGGCCAGTTGGAATCTATTGCTCATCGGCTCTG	CTCACCCATCATCATCGTCACATT
Ah4-24	CGTTGTAAAACGACGGCCAGTTTCTGATTTTAGTAGTCTTCTTTCACT	CTCCTTAGCCACGGTTCT

Underlined sequences indicate the 5′ M13 tail.

Table 3. *Number of alleles, size range and PIC score obtained with SSR primers for the entire data set and the accessions in the mini core*

Marker name	Repeat motif	No. of alleles	Size range (bp)	PIC score	% of wilds producing a band	No. of alleles in mini core	Size range in mini core (bp)
pPGSseq12B6	CTA	7	245–300	0·083	66·6	1	260
pPGSseq13A10	TAA	9	286–349	0·427	73·3	5	286–304
pPGSseq15C12	TAA	22	231–315	0·888	66·6	18	261–315
pPGSseq15E11	TTG	8	310–340	0·128	53·3	5	310–328
pPGSseq15F12	GA	12	287–319	0·811	40·0	7	293–305
pPGSseq16F1	TAA	17	262–412	0·789	86·6	12	271–307
pPGSseq19D6	TAA	13	240–278	0·847	13·3	12	240–276
pPGSseq19D9	TAA	15	285–366	0·862	53·3	13	285–321
pPGPseq8D9	CTT	14	127–172	0·666	100·0	7	136–169
pPGPseq2E6	GA	24	263–325	0·907	100·0	20	271–325
pPGPseq2D12B	TAA	16	277–327	0·810	13·3	16	277–327
pPGPseq3A8	TAA	13	156–366	0·799	60·0	11	156–186
pPGPseq7G2	TATC	10	230–266	0·748	20·0	9	230–266
pPGPseq3A1	TAA	17	137–275	0·706	100·0	10	137–266
pPGPseq2G4	TAA	25	194–413	0·911	13·3	19	278–335
pPGPseq2F5	TAA	17	255–315	0·669	100·0	8	273–297
pPGSseq11G3	CTT	29	123–401	0·796	93·3	16	168–265
pPGPseq8E12	TTG/TAA	11	194–365	0·586	86·6	9	203–248
pPGSseq19F4	GA	18	180–376	0·624	100·0	10	180–312
PM036	GA	20	195–255	0·630	100·0	7	221–249
PM050	GA	17	109–365	0·855	60·0	12	109–139
PM003	GA	23	201–253	0·883	100·0	16	209–245
PM183	CT	23	110–166	0·904	100·0	19	114–166
PM032	CT	15	108–138	0·461	100·0	5	110–130
PM238	CT	15	161–249	0·770	100·0	6	169–183
PM210	CT	13	196–240	0·823	73·3	10	218–240
PM137	GA	15	104–366	0·722	86·6	6	164–174
Ah-041	GTT/GAG	3	274–286	0·385	100·0	2	280–286
Ah-558	AAC	6	247–265	0·489	100·0	3	256–265
Ah4-26	CT	18	174–232	0·706	100·0	12	174–218
Ah4-24	ATA	12	320–446	0·616	46·6	6	428–446
Total		477				312	
Average		15·4		0·687	74·3	10·1	

prepare the sample for sequencing. Sequencing reactions were prepared by following the instructions from the DTCS quick start sequencing kit (Beckman Coulter, Fullerton, CA). The sample was sequenced bi-directionally and pUC18 was also sequenced as a positive control. Each sample was sequenced twice to verify fidelity of the sequenced bases. Samples were injected and sequenced on a Beckman CEQ 8000 using the LFR-1 method. The sequence module of the software package CEQ 8000 Genetic Analysis System version 8.0.52 from Beckman was used to call the bases after the sequencing was performed. The forward and reverse strands were aligned using AlignIR version 2.0 (LI-COR, Lincoln, NE).

(v) *Data analysis*

Gel images were scored and the data were formatted using GeneScan version 3.1.2 and Genotyper version 2.5 (ABI, Foster City, CA). A distance matrix was constructed with Microsat version 1.5 (Minch *et al.*, 1997) using the proportion of shared alleles ($D = 1 - p_s$) as a genetic distance measure (Bowcock *et al.*, 1994). The distance matrix was imported into the neighbour program part of the software package Phylip (Felsenstein, 1996) to construct a neighbour-joining tree and perform bootstrapping. *A. ipaensis* has diverged from cultivated peanut but is also closely related to cultivated peanut and was therefore chosen as the outgroup in our study. Polymorphic information content (PIC) scores were calculated to determine maker diversity using Botstein's formula:

$$1 - \sum_{i-1} p_i^2 - \sum_{i=1}^{n} \sum_{j=i+1}^{n} 2(p_i p_j)^2$$

(Botstein *et al.*, 1980). Botstein originally defined the PIC score as the probability of a given marker being informative in a random mating.
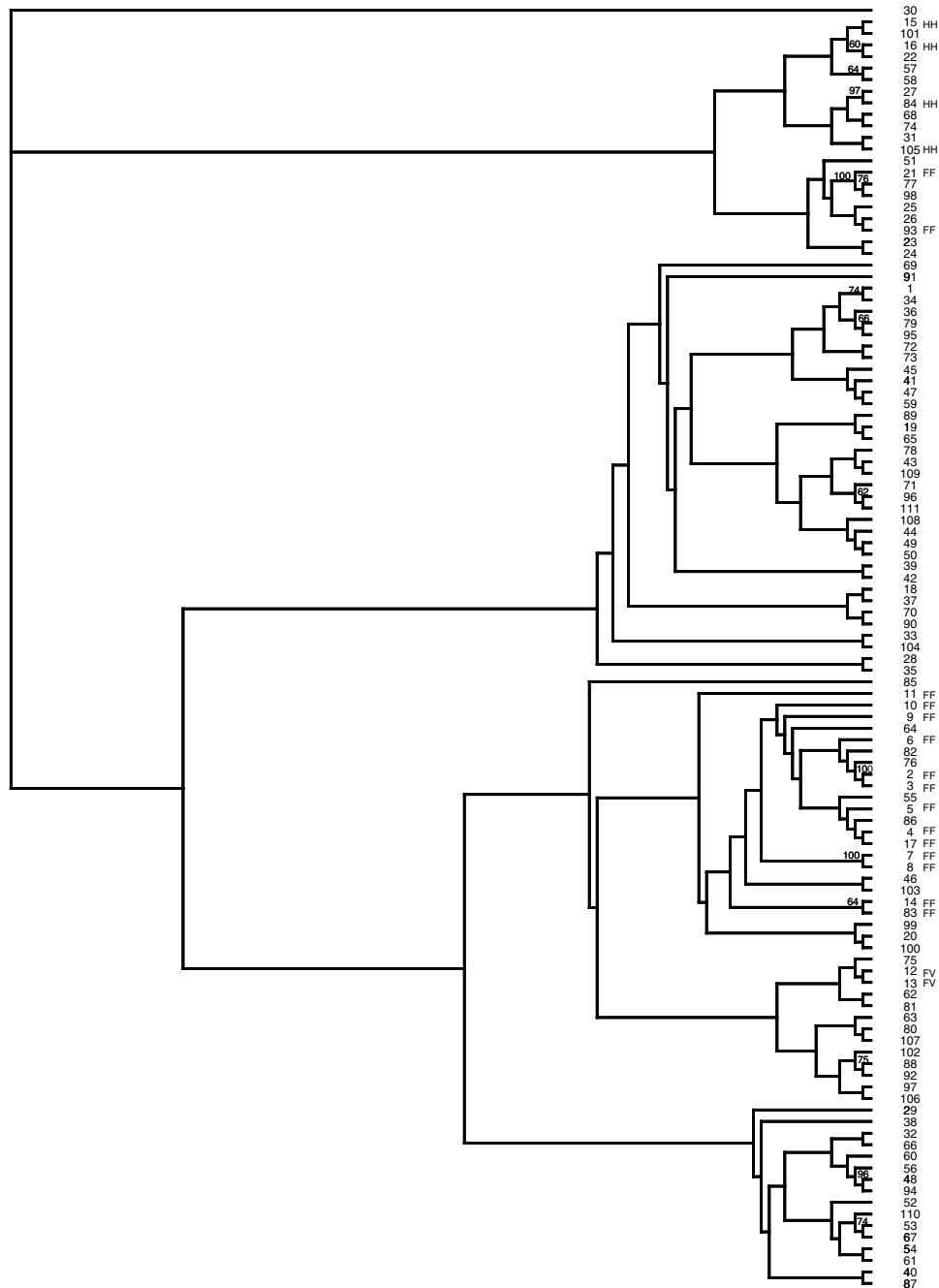
Fig. 1. Neighbour-joining tree of peanut mini core (*Arachis hypogaea*) generated from SSR data. Bootstrap values greater than 50 % are placed on the branches. FF, *fastigiata fastigiata*; FV, *fastigiata vulgaris*; HH, *hypogaea hypogaea*.

3. Results and discussion

(i) SSR markers and PIC scores

Thirty-one M13-tailed SSR markers were used to assess diversity in a collection of cultivated peanut and some near wild relatives. The number of alleles ranged from 3 to 29 with a mean of 15·4 alleles per locus. A total of 312 alleles were produced in accessions from the mini core, with the alleles per marker ranging from 1 to 20 alleles with a mean of 10·1 alleles per locus. The dinucleotide repeat markers in this data set detected more polymorphisms than the tri-nucleotide repeat markers, with an average of 17·75 and 14·11 alleles per marker, respectively. PIC scores were calculated for all markers and ranged from 0·083 to 0·911 with a mean of 0·687 (Table 3). The most informative markers in this data set were pPGPseq2G4, pPGPseq2E6 and PM183. The least informative markers were pPGSseq15E11 and pPGSseq12B6.
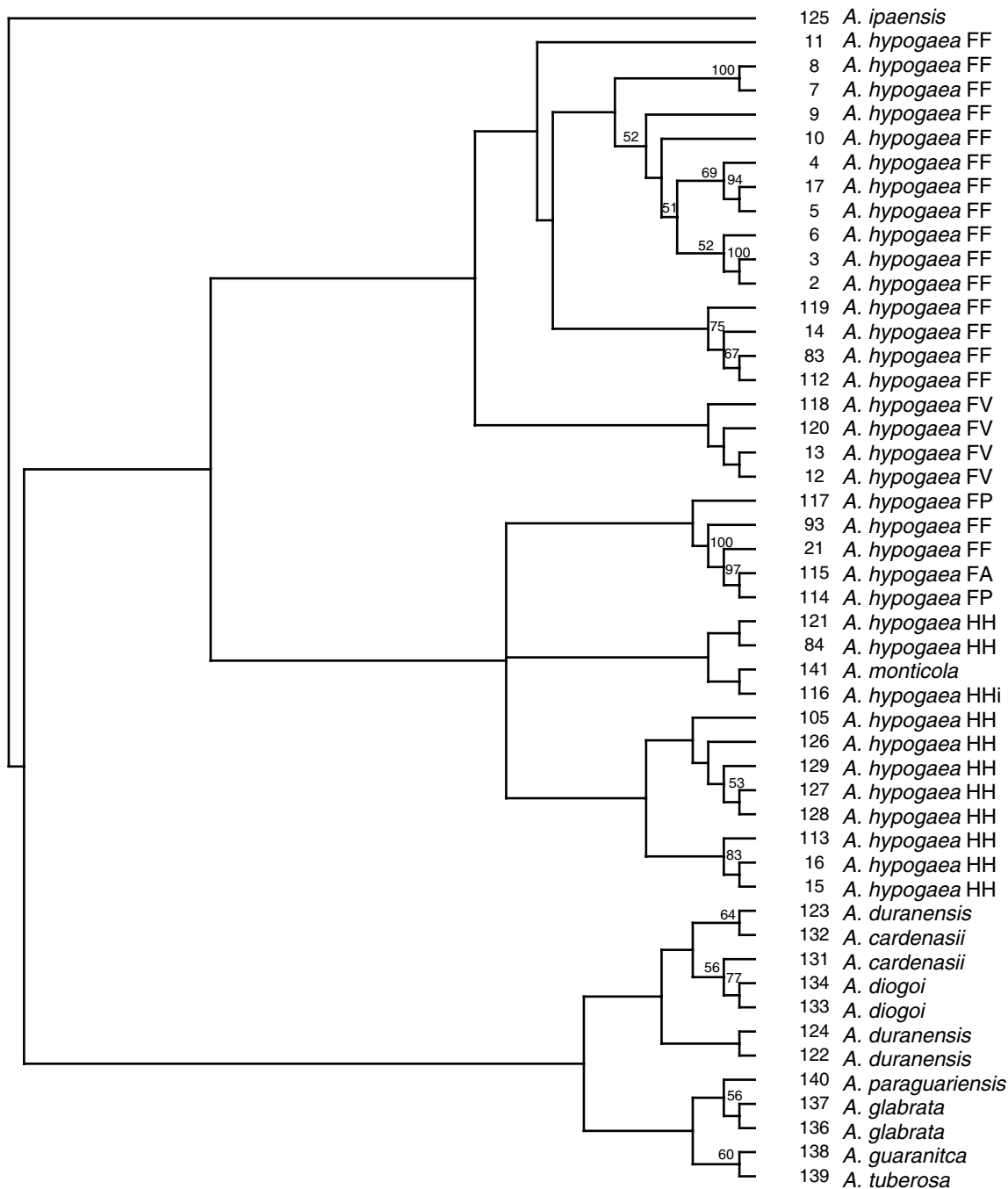
Fig. 2. Neighbour-joining tree of botanical varieties and wild relatives. Bootstrapping was performed with 100 replicates and values greater than 50% were placed on the branches. FF, *fastigiata fastigiata*; FV, *fastigiata vulgaris*; FP, *fastigiata peruviana*; FA, *fastigiata aequatoriana*; HH, *hypogaea hypogaea*; Hhi, *hypogaea hirsuta*.

The markers used in this study were originally designed for cultivated peanut (*Arachis hypogaea* L.) and not for the accessions in this study that are wild relatives, which are classified in nine separate species. On average, 74·3% of the wild accessions produced a PCR product when using these markers. Several of these markers, including pPGSseq19D6, pPGPseq2D12B and pPGPseq2G4, did not produce a PCR product in most of the wild relatives (Table 3)

and thus had a low rate of success. However, 13 of the markers transferred and produced a band in all the wild relative accessions in this study. This suggests that these markers may be beneficial in future studies assaying genetic diversity of wild species. Sequencing would need to be preformed to ensure that when transferring these primers to a different species the repeat motif is conserved, since repeat motifs do not always transfer across genus and sometimes species

```
113       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCACAGGAAA GAGCGCATTA GAACTGAGAG
122       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCACAGGAAA GAGCGCATTA GAACTGAGAG
123       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCACAGGAAA GAGCGCATTA GAACTGAGAG
124       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCACAGGAAA GAGCGCATTA GAACTGAGAG
131       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCACAGGAAA GAGCGCATTA GAACTGAGAG
132       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCACAGGAAA GAGCGCATTA GAACTGAGAG
133       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCACAGGAAA GAGCGCATTA GAACTGAGAG
134       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCACAGGAAA GAGCGCATTA GAACTGAGAG
136       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCAGAGGAAA GAGCGCATTA GAACTGAAAA
137       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCAGAGGAAA GAGCGCATTA GAACTGAAAA
139       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCAGAGGAAA GAGCGCATTA GAACTGAAAA
138       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCAGAGGAAA GAGCGCATTA GAACTGAAAA
140       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCAGAGGAAA GAGCGCATTA GAACTGAAAA
125       GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCACAGGAAA GAGCGCATTA GAACTGAGAG
AH041f    GCTGGGATC ATTGTAGGGA AGG
      1   --------- ---------- ---------- ---------- ---------- ----------
Consensus GCTGGGATC ATTGTAGGGA AGGTGAAGAT GCASAGGAAA GAGCGCATTA GAACTGARAR
                                                 *                     *  *

113       TTTGTTGTTA TTGTTG---G TTCTTGTTCT TCTTGTTGTT GTTGGTGGTG GTGATGATTG
122       TTTGTTGTTA TTGTTG---G TTCTTGTTCT TCTTGTTGTT GTTGGTGGTG GTGATGATTG
123       TTTGTTGTTA TTGTTG---G TTCTTGTTCT TCTTGTTGTT GTTGGTGGTG GTGATGATTG
124       TTTGTTGTTA TTGTTG---G TTCTTGTTCT TCTTGTTGTT GTTGGTGGTG GTGATGATTG
131       TTTGTTGTTA TTGTTG---G TTCTTGTTCT TCTTGTTGTT GTTGGTGGTG GTGATGATTG
132       TTTGTTGTTA TTGTTG---G TTCTTGTTCT TCTTGTTGTT GTTGGTGGTG GTGATGATTG
133       TTTGTTGTTA TTGTTG---G TTCTTGTTCT TCTTGTTGTT GTTGGTGGTG GTGATGATTG
134       TTTGTTGTTA TTGTTG---G TTCTTGTTCT TCTTGTTGTT GTTGGTGGTG GTGATGATTG
136       TTTGTTGTTG TTGTTGTTGG TTCTTGTTCT T--------- GTTGGTGGTG GTGATGATTG
137       TTTGTTGTTG TTGTTGTTGG TTCTTGTTCT T--------- GTTGGTGGTG GTGATGATTG
139       TTTGTTGTTG TTGTTGTTGG TTCTTGTTCT T--------- GTTGGTGGTG GTGATGATTG
138       TTTGTTGTTG TTGTTG---G TTCTTGTTCT T--------- GTTGGTGGTG GTGATGATTG
140       TTTGTTGTTG TTGTTG---G TTCTTGTTCT T--------- GTTGGTGGTG GTGATGATTG
125       TTTGTTGTTG TTGTTG---G TTCTTGTTCT T------GTT GTTGGTGGTG GTGATGATTG
      61  ---------- ---------- ---------- ---------- ---------- ----------
Consensus TTTGTTGTTR TTGTTGTTGG TTCTTGTTCT TCTTGTTGTT GTTGGTGGTG GTGATGATTG
               *            ***           ********

113       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GAGGAACCTC GCCGTACGAG
122       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GAGGAACCTC GCCGTACGAG
123       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GAGGAACCTC GCCGTACGAG
124       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GAGGAACCTC GCCGTACGAG
131       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GAGGAACCTC GCCGTACGAG
132       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GAGGAACCTC GCCGTACGAG
133       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GAGGAACCTC GCCGTACGAG
134       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GAGGAACCTC GCCGTACGAG
136       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GATGAACCTC GCCGTACGAG
137       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GATGAACCTC GCCGTACGAG
139       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GATGAACCTC GCCGTACGAG
138       GTTATTCTTG AAGGAGGAGG AGGAGGAGGA GATGGAGGGG GAGGAACCTC GCCGTACGAG
140       GTTATTCTTG AAGGAGGAGG AGGAG----- -ATGGAGGGG GAGGAATCTC GCCGTACGAG
125       GTTATTCTTG AAGGAGGAGG AGGAGGAG-- -ATGGAGGGG GAGGAACCTC GCCGTACGAG
     121  ---------- ---------- ---------- ---------- ---------- ----------
Consensus GTTATTCTTG AAGGAGGAGG AGGAGGAGGA GATGGAGGGG GAKGAAYCTC GCCGTACGAG
                                 ***** *              *     *
```
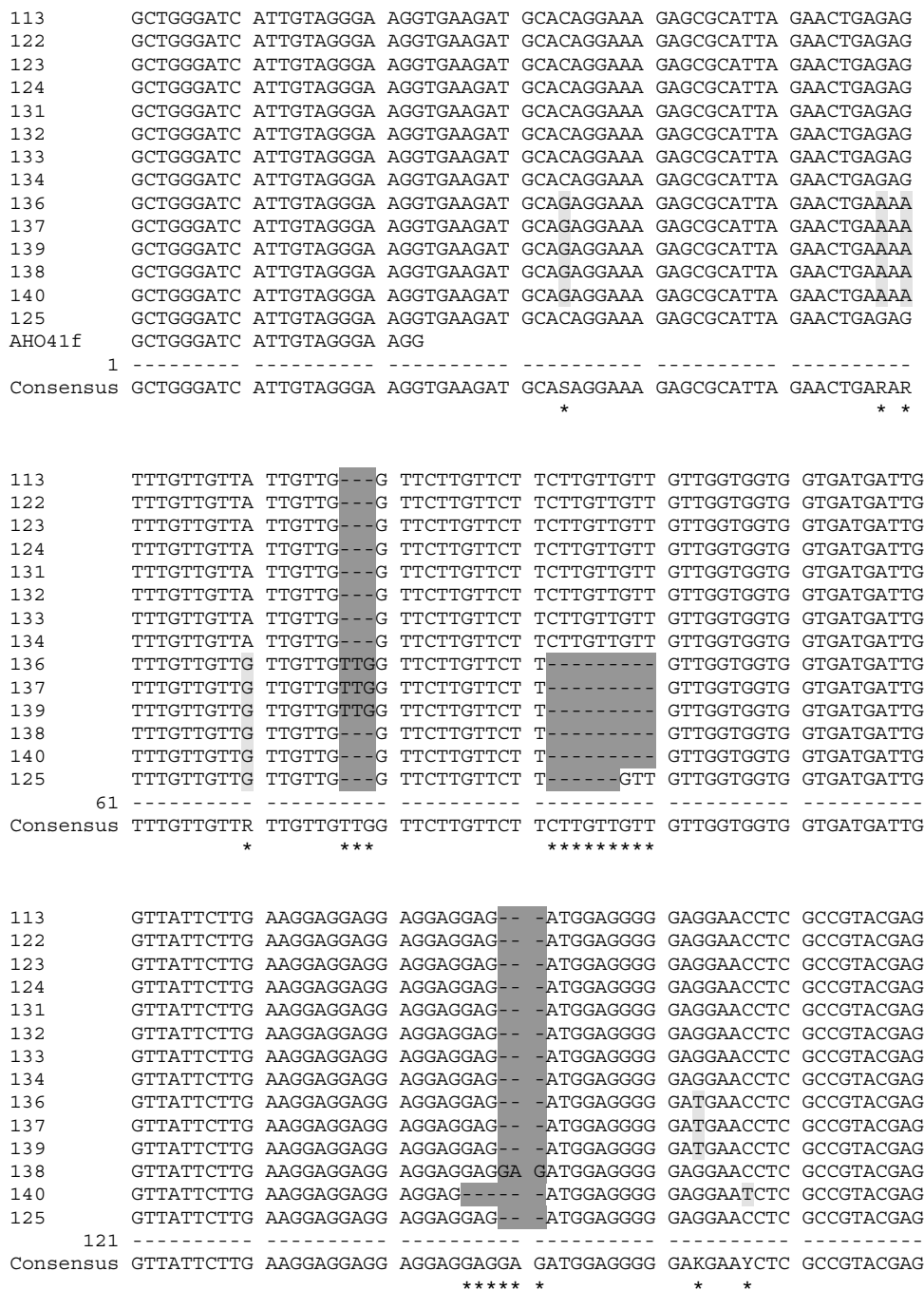
Fig. 3. Sequence alignment of Ah041 microsatellite alleles generated from cultivated and wild peanuts produced by using AlignIR version 2.0.

borders (Noor *et al.*, 2001; Chen *et al.*, 2002; Rossetto *et al.*, 2002).

(ii) *Phylogenetic analysis of the peanut mini core*

The data from 31 SSR markers were utilized to construct a neighbour-joining tree showing the relationships among the peanut mini core accessions (Fig. 1). The authors are unaware of any reports showing the intraspecific relationships of accessions in the mini core. This was accomplished by calculating a distance matrix based on the proportion of shared alleles for all pairwise combinations. Bootstrapping was performed with 100 replicates and all values greater than 50% were placed on the branches. High bootstrap support was obtained between many of the accessions that clustered closely together; however, low bootstrap support was obtained between the clades. Perhaps with the addition of more SSR markers the bootstrap support may increase; however, the low

```
113          GAGCTTTTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAAG AGGGAGTGGT GGTTGGAAGG
122          GAGCTTTTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAAG AGGGCGTGGT GGTTGGAAGG
123          GAGCTTTTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAAG AGGGCGTGGT GGTTGGAAGG
124          GAGCTTTTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAAG AGGGCGTGGT GGTTGGAAGG
131          GAGCTTTTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAAG AGGGCGTGGT GGTTGGAAGG
132          GAGCTTTTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAAG AGGGCGTGGT GGTTGGAAGG
133          GAGCTTTTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAAG AGGGCGTGGT GGTTGGAAGG
134          GAGCTTTTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAAG AGGGCGTGGT GGTTGGAAGG
136          GAGCTTCTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAGG AGGGCGTGGT GGTTGGAAGG
137          GAGCTTCTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAGG AGGGCGTGGT GGTTGGAAGG
139          GAGCTTCTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAGG AGGGCGTGGT GGTTGGAAGG
138          GAGCTTCTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAGG AGGGCGTGGT GGTTGGAAGG
140          GAGCTTCTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAGG AGGGCGTGGT GGTTGGAAGG
125          GAGCTTTTCG GCAACGTCGT CGGGGTCGGA GAGGCGGAAG AGGGCGTGGT GGTTGGAAGG
       181   ---------- ---------- ---------- ---------- ---------- ----------
Consensus    GAGCTTYTCG GCAACGTCGT CGGGGTCGGA GAGGCGGARG AGGGMGTGGT GGTTGGAAGG
                     *                                  *          *


113          GGTTATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
122          GGTTATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
123          GGTTATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
124          GGTTATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
131          GGTTATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
132          GGTTATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
133          GGTTATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
134          GGTTATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
136          GGATATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
137          GGATATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
139          GGATATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
138          GGATATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
140          GGATATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
125          GGTTATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
AHO41R                  GGTGCTTG TTAATCTTGT GGCG
AH041R TAIL             GGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
       241   ---------- ---------- ---------- ---------- ---------- -----
Consensus    GGWTATTCTT CTGGTGCTTG TTAATCTTGT GGCGACTGGC CGTCGTTTTA CAACG
                 *
```

Fig. 3. (*Cont.*)

bootstrap support could also be due to the narrow genetic base of cultivated peanut. Genetic variation in the mini core was obtained using these SSR markers; however, accessions #2 and #3 had identical banding patterns for the 31 markers used in this study, suggesting that these two accessions are genetically similar.

Two main clades were produced in this tree. The majority of all the accessions classified as botanical varieties clustered together. This work supports the current taxonomy. Two accessions (#12 and #13) classified as *Arachis hypogaea fastigiata vulgaris* clustered with one another. All the accessions classified as *Arachis hypogaea hypogaea hypogaea* clustered together. Lastly, all but two accessions (#21 and #93) of *Arachis hypogaea fastigiata fastigiata* grouped together. PI 497639 (#21) is currently listed on GRIN (http://www.ars-grin.gov/npgs/) as *Arachis hypogaea fastigiata fastigiata* but the descriptor data lists this accession as having no flowers on the main axis, a spreading and bunch growth habit, rough pod reticulation, deep strangulation of pods and a pod shape of *hirsuta*. These morphological observations suggest that #21 may be *Arachis hypogaea hypogaea*

hirsuta and not *Arachis hypogaea fastigiata fastigiata*. Additionally, #93 PI 476025, which is currently classified as *Arachis hypogaea fastigiata fastigiata*, has flowers on the main axis, dark green leaves, rough pod reticulation, deep strangulation of pods and was collected in Peru, which would suggest that this accession should be classified in subspecies *fastigiata* due to the flowers on the main axis but may be var. *peruviana* or var. *aequatoriana* rather than var. *fastigiata*. Both these accessions are now being grown out by the peanut curator to re-examine the descriptor data in detail and determine whether they are possibly misclassified.

(iii) *Phylogenetic analysis of botanical varieties and wild relatives*

To better understand the interspecific and the intraspecific relationships between the peanut botanical varieties and wild relatives in this data set, a neighbour-joining tree was constructed using the data from the 31 SSR markers with the 35 accessions classified as botanical varieties and the 14 wild relatives (Table 1). The proportion of shared alleles was

used to calculate genetic distance between accessions and bootstrapping was performed with 100 replications. All bootstrap values greater than 50% were placed on the tree (Fig. 2). Two main clades were produced consisting of the botanical varieties and the other clade consisted of wild relatives. The clade with the botanical varieties split into two subgroups. The relationships between botanical varieties in this tree were similar to the relationships obtained among botanical varieties in the mini core tree.

The first subgroup consisted of all varieties classified as *A. hypogaea fastigiata fastigiata* except #21 and #93, which appear to be misclassified based on observed morphological data listed on GRIN (discussed previously). Additionally, this subgroup also clustered all four accessions of botanical variety *A. hypogaea fastigiata vulgaris* together. The relationship in this tree suggests that *A. hypogaea fastigiata fastigiata* and *A. hypogaea fastigiata vulgaris* are similar to one another. The second subgroup consisted of all varieties classified as *A. hypogaea hypogaea hypogaea*, *A. hypogaea hypogaea hirsuta*, *A. hypogaea fastigiata peruviana* and *A. hypogaea fastigiata aequatoriana*. In previous papers, there has been some debate as to the placement of *A. hypogaea fastigiata peruviana*. Some studies have shown that this variety was more similar to subspecies *hypogaea* whereas other studies found it more similar to subspecies *fastigiata*, in which it is currently classified (He & Prakash, 2001; Raina *et al.*, 2001; Moretzsohn Mde *et al.*, 2004). Our study suggests *peruviana* is more similar to subspecies *hypogaea* than *fastigiata*.

The other clade consisted of all the wild relatives except *A. monticola*. This clade can be divided into two subgroups. The first subgroup consisted of all wild relatives with the A genome (*A. duranensis*, *A. cardenasii* and *A. diogoi*). The second subgroup consists of *A. paraguariensis*, *A. glabrata*, *A. guaranitca* and *A. tuberosa*. *A. monticola* did not cluster with the wilds but grouped within the clade consisting of all the botanical varieties, suggesting that it is more closely related to cultivated peanut than the wild relatives. The botanical varieties were removed from this data set and a tree was constructed with just the wild accessions, which produced a tree with the same phylogenetic relationships among the wilds as Fig. 2 except that *A. monticola* clustered with *A. ipaensis* as opposed to clustering with cultivated peanut (data not shown).

(iv) *Classifying peanuts into subspecies and botanical varieties*

Since only 21 of 111 (18·9%) samples in the mini core are classified in botanical varieties, one of the goals of this work was to determine whether the SSR markers would allow us to putatively classify these accessions into subspecies and botanical varieties based on molecular data and GRIN observation data. The data from 31 SSR markers from the botanical varieties and the wilds were used to construct multiple phylogenetic trees by adding a few samples from the mini core of unknown subspecies and botanical varieties to see whether they would cluster with a particular group such as *fastigiata fastigiata*, *fastigiata vulgaris* or *hypogaea hypogaea*. Based on molecular data alone, 11 accessions grouped with *fastigiata fastigiata*, 24 with *fastigiata vulgaris*, 43 with *hypogaea hypogaea* and 13 were questionable (data not shown). The morphological data that are available on GRIN seems to be consistent with the putative subspecies classification based on SSR markers of 40 accessions. Many of these accessions did not have enough morphological data to determine whether this tentative classification may be correct, and they will need to be grown in the field and evaluated further to confirm the putative classification.

(v) *Further classifying cultivated and wild peanuts by SSR allele sequencing*

To determine whether the simple sequence repeat motif was included in the alleles produced from wild relatives and to examine how these alleles were evolving (stepwise manner or infinite allele model), a few alleles from marker Ah041 were chosen for sequencing. All the wild accessions except *A. monticola* produced a single band with marker Ah041 and were sequenced (Fig. 3). Three different-sized alleles were chosen for sequencing. The sequence produced from a cultivated peanut (#113) was used to perform a BLAST search (www.ncbi.nlm.nih.gov/BLAST/). The BLAST results showed that the cultivated peanut sequence had high homology to accession DQ099247.1, with 94% identity and an E value of $6e^{-36}$. This accession was produced from an *Arachis hypogaea* clone microsatellite sequence that was used to develop marker Ah041. There were no other sequences in the database from any other organisms that had high homology to the sequenced cultivated peanut microsatellite allele.

Several single nucleotide polymorphisms (SNPs) were observed between the alleles sequenced (Fig. 3). In general, the SNPs occurred between *A. glabrata*, *A. guaranitca*, *A. tuberosa* and *A. paraguariensis* compared with *A. hypogaea hypogaea hypogaea*, *A. duranensis*, *A. ipaensis*, *A. cardenasii* and *A. diogoi*. There are six SNPs that separate these species that are considered AA (*A. duranensis*, *A. diogoi*, *A. cardenasii*), BB (*A. ipaensis*) and AABB (cultivated peanut) genomes from the remaining wild species (*A. glabrata*, *A. guaranitca*, *A. tuberosa* and *A. paraguariensis*) included in this study. Four of these six SNPs were transitions while the remaining two were

transversions. Another SNP (G/T transversion) set apart the two *A. glabrata* accessions and one *A. tuberosa* accession from all the other accessions sequenced. Lastly, a SNP (A/C) detected in *A. hypogaea hypogaea hypogaea* was not seen in any of the other wild accessions sequenced. It is possible that further testing of additional accessions from each of the wild species would allow these SNPs to be used to design new markers to distinguish between wild peanut species.

In all the alleles sequenced the primer attachment sites were conserved with no point mutations or gaps observed when compared with the primer sequences. Additionally, the repeat motif (GTT/GAG) was conserved in all wild accessions sequenced, suggesting that this SSR marker would be suitable to evaluate diversity among various wild accessions. The variation among alleles differed by increments of the repeat motif and insertion/deletions (indels) occurring in regions near the simple sequence repeat. This would suggest that changes in allele size in peanuts are not always due to changes in the length of the repeat motif and thus a stepwise mutation model would not be appropriate for analysing peanut SSR data. Therefore, an infinite allele model or a genetic distance measure that assumes all alleles are equally related, such as the proportion of shared alleles (Bowcock *et al.*, 1994), might be appropriate to analyse a SSR data set in peanut.

(vi) *Peanut diversity*

In this study, the diversity and phylogenetic relationships of the peanut mini core, botanical varieties and some wild accessions were assessed with 31 previously published SSR markers using the M13-tailed method. The mini core was not as genetically diverse as the entire population, having fewer SSR alleles (312 produced in the mini core and 477 in the total population). This difference in alleles produced is probably due to the inclusion of wild peanuts in this data set that are generally thought to be more genetically diverse than cultivated peanut. However, even though the mini core was less diverse than the entire population, many of the accessions in the mini core were able to be distinguished from one another by using a fairly large number of markers. These SSR markers has helped identify a few accessions (#21, #93) in the mini core that appear to be misclassified based on morphological and molecular data which are currently being evaluated. Clarifying the proper classification of these accessions will help in the curation of the peanut germplasm collection. These markers have also shown that accessions classified as botanical varieties are very similar to each other genetically and phylogenetically, lending further support to their current taxonomy. Using the molecular data and descriptor data from evaluating the morphology of these accessions it is possible that this study will permit more accessions in the mini core to be classified into botanical varieties. Furthermore, sequencing some alleles in wild accessions from marker Ah041 has demonstrated that the simple sequence repeat motif was conserved when transferring across species borders and with further testing this sequence data may allow SNP markers to be produced that help distinguish accessions classified as wild species. Overall, these SSR data have allowed the examination of the diversity and phylogenetic relationships among accessions in the mini core and provided data that will be helpful in the overall collection and utilization in breeding management of this germplasm collection.

References

Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**, 314–331.

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. & Cavalli-Sforza, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457.

Chen, X., Cho, Y. G. & McCouch, S. R. (2002). Sequence divergence of rice microsatellites in *Oryza* and other plant species. *Molecular Genetics and Genomics* **268**, 331–343.

Felsenstein, J. (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Computer Methods for Macromolecular Sequence Analysis* **266**, 418–427.

Ferguson, M. E., Burow, M. D., Schulze, S. R., Bramel, P. J., Paterson, A. H., Kresovich, S. & Mitchell, S. (2004). Microsatellite identification and characterization in peanut (*A. hypogaea* L.). *Theoretical and Applied Genetics* **108**, 1064–1070.

Gimenes, M. A., Lopes, C. R. & Valls, J. F. M. (2002). Genetic relationships among *Arachis* species based on AFLP. *Genetics and Molecular Biology* **25**, 349–353.

He, G. & Prakash, C. (2001). Evaluation of genetic relationships among botanical varieties of cultivated peanut (*Arachis hypogaea* L.) using AFLP markers. *Genetic Resources and Crop Evolution* **48**, 347–352.

He, G., Meng, R., Newman, M., Gao, G., Pittman, R. N. & Prakash, C. S. (2003). Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biology* **3**, 3.

Herselman, L. (2003). Genetic variation among Southern African cultivated peanut (*Arachis hypogaea* L.)

genotypes as revealed by AFLP analysis. *Euphytica* **133**, 319–327.

Holbrook, C. C. & Dong, W. B. (2005). Development and evaluation of a mini core collection for the US peanut germplasm collection. *Crop Science* **45**, 1540–1544.

Holbrook, C. C., Anderson, W. F. & Pittman, R. N. (1993). Selection of a core collection from the United States germplasm collection of peanut. *Crop Science* **33**, 859–861.

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

Kochert, G., Halward, T., Branch, W. D. & Simpson, C. E. (1991). RFLP variability in peanut (*Arachis hypogaea* L) cultivars and wild species. *Theoretical and Applied Genetics* **81**, 565–570.

Kochert, G., Stalker, H. T., Gimenes, M., Galgaro, L., Romero Lopez, C. & Moore, K. (1996). RFLP and cytogenetic evidence on the origin and evolution of allo-tetraploid domesticated peanut, Arachis hypogaea (Leguminosae). *American Journal of Botany* **83**, 1282–1291.

Krapovickas, A. & Gregory, W. (1994). Taxonomia del genero Arachis (Leguminosae). *Bonpladia* **8**, 1–187.

Lopez, Y., Nadaf, H. L., Smith, O. D., Connell, J. P., Reddy, A. S. & Fritz, A. K. (2000). Isolation and characterization of the Delta(12)-fatty acid desaturase in peanut (*Arachis hypogaea* L.) and search for polymorphisms for the high oleate trait in Spanish market-type lines. *Theoretical and Applied Genetics* **101**, 1131–1138.

Minch, E., Ruiz-Linares, A., Goldstein, D., Feldman, M., Kidd, J. R. & Cavalli-Sforza, L. L. (1997). Microsat 1.5: a computer program for calculating various statistics on microsatellite allele data. Available via http://hpglstanfordedu/projects/microsat/programs/.

Moretzsohn Mde, C., Hopkins, M. S., Mitchell, S. E., Kresovich, S., Valls, J. F. & Ferreira, M. E. (2004). Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biology* **4**, 11.

Noor, M. A. F., Kliman, R. M. & Machado, C. A. (2001). Evolutionary history of microsatellites in the *obscura* group of *Drosophila*. *Molecular Biology and Evolution* **18**, 551–556.

Raina, S. N., Rani, V., Kojima, T., Ogihara, Y., Singh, K. P. & Devarumath, R. M. (2001). RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species. *Genome* **44**, 763–772.

Revoredo, C. L. & Fletcher, S. M. (2002). World peanut market: an overview of the past 30 years. *Georgia Agricultural Experiment Station Research Bulletin* **437**, 1–24.

Rossetto, M., McNally, J. & Henry, R. J. (2002). Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae. *Theoretical and Applied Genetics* **104**, 61–66.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**, 233–234.

Subramanian, V., Gurtu, S., Rao, R. C. N. & Nigam, S. N. (2000). Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. *Genome* **43**, 656–660.