Euphytica (2006) 152:317–330
DOI 10.1007/s10681-006-9218-0

SSR analysis of cultivated groundnut (*Arachis hypogaea* L.) germplasm resistant to rust and late leaf spot diseases

E. S. Mace · D. T. Phong · H. D. Upadhyaya · S. Chandra · J. H. Crouch

Received: 3 March 2006 / Accepted: 14 June 2006 / Published online: 23 November 2006

Abstract Cultivated groundnut (*Arachis hypogaea* L.) is an agronomically and economically important oilseed crop grown extensively throughout the semi-arid tropics of Asia, Africa and Latin America. Rust (*Puccinia arachidis*) and late leaf spot (LLS, *Phaseoisariopsis personata*) are among the major diseases causing significant yield loss in groundnut. The development of varieties with high levels of resistance has been constrained by adaptation of disease isolates to resistance sources and incomplete resistance in resistant sources. Despite the wide range of morphological diversity observed in the cultivated groundnut gene pool, molecular marker analyses have thus far been unable to detect a parallel level of genetic diversity. However, the recent development of simple sequence repeat (SSR) markers presents new opportunities for molecular diversity analysis of cultivate groundnut. The current study was conducted to identify diverse disease resistant germplasm for the development of mapping populations and for their introduction into breeding programs. Twenty-three SSRs were screened across 22 groundnut genotypes with differing levels of resistance to rust and LLS. Overall, 135 alleles across 23 loci were observed in the 22 genotypes screened. Twelve of the 23 SSRs (52%) showed a high level of polymorphism, with PIC values ≥0.5. This is the first report detecting such high levels of genetic polymorphism in cultivated groundnut. Multi-dimensional scaling and cluster analyses revealed three well-separated groups of genotypes. Locus by locus AMOVA and Kruskal–Wallis one-way ANOVA identified candidate SSR loci that may be valuable for mapping rust and LLS resistance. The molecular diversity analysis presented here provides valuable information for groundnut breeders designing strategies for incorporating and pyramiding rust and late leaf spot resistances and for molecular biologists wishing to create recombinant inbred line populations to map these traits.

Keywords Groundnut · SSR loci · Rust resistance · Late leaf spot resistance · Diversity analysis

E.S. Mace and D.T. Phong contributed equally to this work.

E. S. Mace · H. D. Upadhyaya · S. Chandra · J. H. Crouch
International Crop Research Institute for the Semi Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh 502 324, India

D. T. Phong
Department of Plant Cell Technology, Institute of Biotechnology, Hoang Quoc Viet 18, Hanoi, Vietnam

E. S. Mace (✉)
Department of Primary Industries & Fisheries, Hermitage Research Station, 604 Yangan Road,

Introduction

Cultivated groundnut, *Arachis hypogaea* L., is an important oilseed crop grown as a major source of vegetable oil and protein, both for human consumption and as a fodder crop. Groundnut is cultivated in over 100 countries across Asia, Africa and the Americas with around 25 million hectares generating an annual production of nearly 35 million tonnes (FAO 2004). India, China, Nigeria and Sudan are the top producers but more than 20 other countries, mainly in Asia and Africa, each have 1–800,000 ha of groundnut production. Although groundnut is an important multipurpose crop for resource-poor farmers in the semi-arid tropics (SAT), due to environmental stresses and disease pressure, average productivity is often below 1 tonne per hectare. The major disease constraints to groundnut production are rust (causal agent *Puccinia arachidis* Speg) and late leaf spot (LLS, causal agent *Phaseoisariopsis personata* (Berk. & Curtis.) Deighton), resulting in annual economic losses of US$467 m and US$599 m, respectively (FAO 2004). Reducing groundnut yield losses due to these diseases has relied principally on breeding foliar disease resistant germplasm.

Groundnut breeding is a complex endeavour due to the allotetraploid ($2n = 4x = 40$) nature of the crop and the inheritance of agronomic traits being largely oligogenic or polygenic (Halward et al. 1991). Different sources of resistance to LLS have been reported as having a digenic recessive basis (Tiwari et al. 1984) or being conferred by a five-gene model (Nevill 1981), whereas resistance to rust has been consistently reported as genetically recessive, governed by only a few genes (Paramasivam et al. 1990; Bromfield and Bailey 1972; Tiwari et al. 1984). The development of cultivars with enhanced levels of disease resistance has had some success; a few cultivars with moderate levels of resistance to rust and LLS have been released in China, India, Mauritius and the USA (Pande et al. 2002). In the SAT the adoption of rust and LLS resistant cultivars has been low mainly because of their relatively late maturity and poor shelling characteristics. In addition, there are only moderate levels of LLS resistance available in the cultivated groundnut gene pool. In contrast, several wild *Arachis* species possess very high levels of resistance to LLS. There has been limited success in transferring LLS resistance from wild *Arachis* to cultivated groundnut, mainly because of interspecific compatibility barriers, resistance being linked with many undesirable pod/seed characteristics, and the long periods required for developing stable tetraploid interspecific derivatives (Murty and Jahnavi 1983). Marker-assisted selection (MAS) may be able to break the linkage drag to deleterious traits, increase the speed and efficiency of creating acceptable interspecific derivatives and facilitate the pyramiding of different sources of resistance from the cultivated and wild gene pools in order to create highly resistant varieties.

The *Arachis* genus comprises over 20 highly diverse species representing eight distinct genomes. However, the evolution of cultivated groundnut in South America, through a limited number of interspecific hybridizations and polyploidization, has resulted in a very narrow cultivated gene pool (Halward et al. 1991, 1992). This has been compounded by limited introductions and selection pressures from traditional breeding, as the main production areas shifted to Asia and Africa. It is therefore critical to determine the levels of genetic diversity available within sources of disease resistant germplasm in an attempt to broaden the genetic base of the crop and maximize opportunities for combining different mechanisms of resistance (Singh et al. 1997).

The use of molecular markers has become widely accepted as a valuable tool for plant breeding programs as well as for diversity, evolutionary and conservation studies (Mohan et al. 1997). Identification of DNA markers associated with resistance to rust and LLS and their location on a genetic linkage map are pre-requisites to facilitate MAS in groundnut. MAS is potentially useful in disease resistance breeding to accelerate the recovery of the genotype of the recurrent parent during backcrossing, to select for disease resistance quantitative trait loci, and to pyramid different resistance genes (Witcome and Hash 2000). A variety of molecular markers have been used to characterize the genetic diversity in groundnut, e.g. RFLPs (Halward et al. 1991),

RAPDs (Dwivedi et al. 2001; Subramanian et al. 2000), and AFLPs (He and Prakash 1997, 2001; Gimenes et al. 2002). All these studies have reported low levels of polymorphism within the cultivated gene pool.

Simple sequence repeats (SSRs), also known as microsatellites, are a class of molecular markers based on tandem repeats of short (2–6 bp) DNA sequences (Litt and Lutty 1989). These repeat sequences are often highly polymorphic, even among closely related cultivars, due to slippage mutations during DNA replication causing variation in the number of repeating units. Different alleles of a given locus can be readily detected using primers designed from the conserved DNA sequences flanking the SSR and the polymerase chain reaction (PCR). SSR markers are generally reported to detect higher levels of polymorphism than RFLPs, RAPDs and AFLPs (Powell et al. 1996b; Milbourne et al. 1997; Russell et al. 1997; Crouch et al. 1999), and have been widely adopted for genetic analysis in plants (Rongwen et al. 1995; Panaud et al. 1996; Powell et al. 1996a). Thus, it is believed that SSR markers will provide the molecular genetic differentiation to facilitate routine diversity analysis and molecular breeding applications (Dwivedi et al. 2003). However, the first SSRs to be developed in groundnut detected disappointing levels of polymorphism in cultivated germplasm (Hopkins et al. 1999). Nevertheless, additional SSRs developed more recently through a different approach appear to be much more promising in cultivated groundnut genotypes (He et al. 2003; Ferguson et al. 2003). In this study, we report on the evaluation of recently developed SSRs for molecular breeding of groundnut and their use to identify diverse parental genotypes for breeding and mapping of rust and LLS resistance.

Materials and methods

Plant material and DNA extraction

A total of 21 groundnut genotypes with varying levels of known resistance to rust and/or LLS and one variety susceptible to both diseases were selected for this study (see Table 1 for full details).

Total genomic DNA was isolated from newly expanded leaves of glasshouse grown plants using a CTAB-based procedure modified from Saghai-Maroof et al. (1984) and Doyle and Doyle (1987) as reported previously (Mace et al. 2003). The quality of DNA was determined spectrophotometrically at 260/280 nm. DNA concentrations were determined electrophoretically through comparison with known concentrations of uncut λ DNA standards.

SSR marker amplification

Twenty-three SSRs developed by Ferguson et al. (2003) (Table 2) were screened across the 22 genotypes. The 23 SSRs were selected on the basis of pre-screening approximately 200 groundnut SSRs (Mace, personal communication), based on the level of polymorphism revealed between LLS and rust resistant and susceptible genotypes and the reliability and quality of amplicon detection, the latter based on a quality rating modified from Smulders et al. (1997) where a quality rating between 1 and 3 indicates a good electrophoretic pattern; 1 = Weak stutter bands only giving unambiguous product; 2 = Stutter bands relatively strong but product still scorable; 3 = Appearance of bands of unexpected size but product still scorable; 4 = Appearance of bands of unexpected size and product not scorable; 5 = Ladders of bands of unequal intensity but product still scorable; 6 = Ladders of bands of equal intensity and product not scorable; 7 = Very weak bands or no amplification. Only amplicons with a quality rating between 1 and 3 were selected.

PCR amplifications were performed in 20 µl volumes using a PTC-100^{TM} Programmable Thermal Controller (MJ Research, Inc). The reaction mixtures contained 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 5–15 ng genomic DNA, 10–30 pmol of each primer, 2–4 mM $MgCl_2$, 300–400 µM of each dNTP, and 0.8–1.2 units of *Taq* DNA polymerase (Amersham). The concentrations were optimized individually for each SSR. The temperature profile consisted of an initial denaturation step of DNA at 94°C for 2 min, followed by 35 cycles: 94°C for 45 s, 57–65°C for 1 min, and 72°C for 1 min 30 s. Annealing

Table 1 Groundnut genotypes employed in this study with levels of resistance and susceptibility to rust and LLS detailed, as determined during field trials at 2 sites in Vietnam in 2001

Code	Genotype	Origin	Biological status	Botanical variety	Rust score[a]		LLS score [a]
					VASI (HN[b])	OPI (HCM[c])	OPI (HCM[c])
1	ICGV99001	ICRISAT	Interspecific derivative	Spanish	2.0	1.0	1.7
2	ICGV99003	ICRISAT	Interspecific derivative	Virginia	1.0	1.0	1.3
3	ICGV99004	ICRISAT	Interspecific derivative	Spanish	3.7	2.3	2.3
4	ICGV99005	ICRISAT	Interspecific derivative	Virginia	1.0	1.0	2.0
5	ICG 86699	ICRISAT	Cultivar	Virginia	1.3	1.0	1.7
6	ICG 87165	ICRISAT	Interspecific derivative	Spanish	4.3	3.1	6.6
7	ICG 87157	ICRISAT	Cultivar	Valencia	2.3	1.3	3.3
8	ICG 99051	ICRISAT	Breeding material	Virginia	1.0	1.0	1.3
9	ICG 99019	ICRISAT	Breeding material	Spanish	1.0	1.0	5.0
10	ICGx950084	ICRISAT	Breeding material	Spanish	1.0	1.0	3.0
11	ICGx950166	ICRISAT	Breeding material	Spanish	1.0	1.0	3.0
12	ICG 10931	Peru	Landrace	Spanish	2.7	2.0	6.0
13	ICG 10975	Peru	Landrace	Spanish	2.6	ND	ND
14	ICG 1185	Argentina	Breeding material	Spanish	1.7	1.3	4.3
15	ICG 11312	India	Breeding material	Spanish	1.0	1.3	3.0
16	ICG 11325	India	Breeding material	Spanish	1.3	1.3	2.8
17	ICG 11331	India	Breeding material	Spanish	2.0	1.7	2.0
18	ICG 11485	Peru	Landrace	Spanish	2.3	ND	4.3
19	ICG 12720	Ecuador	Breeding material	Spanish	2.3	ND	ND
20	ICG 13917	ICRISAT	Breeding material	Spanish	2.3	ND	ND
21	ICG 99052	ICRISAT	Breeding material	Virginia	1.0	1.0	2.0
22	TMV2	India	Cultivar	Spanish	4.0	5.7	6.0

[a] Average disease-response scores reported (from three replications) using a 1–9 scale where 1 = complete resistance and 9 = very susceptible; ND = not determined

[b] HN: Vietnam Agricultural Scientific Institute (VASI) (Hanoi, Vietnam)

[c] HCM: Plant Oil Institute (OPI) (Ho Chi Minh City, Vietnam)

temperatures were optimized individually for each SSR (listed in Table 2). After the final cycle, samples were incubated for 10 min to ensure complete extension.

Electrophoresis and data analysis

The PCR products were separated on 6% non-denaturing polyacrylamide gels, and amplification products were revealed using the silver staining procedure based on a histologically derived procedure using ammoniacal solutions of silver, modified from Kolodny (1984). Amplicons were scored as present (1) or absent (0).

Estimates of interindividual genetic similarity were obtained according to Nei and Li's (1979) as $S_{ij} = S_{ij} = 2a/(2a + b + c)$, where S_{ij} is the similarity between two individuals i and j, a_{ij} is the number of bands present in both individuals i and j, b_i is the number of bands present in individual i but absent in individual j, and c_j is the number of bands present in individual j but absent in individual i. The resulting 22×22 similarity matrix was subjected to multi-dimensional scaling (MDS) (Kruskal and Wish 1978) to assess whether the observed molecular variation indicated any evidence of clustering among accessions. The unweighted pair-group method with arithmetic average (UPGMA) was used to independently confirm the clustering indicated by the two-dimensional MDS plot. The UPGMA-based dendrogram was constructed using the NTSYS 2.1 software, version 2.1 (Rohlf 2000). The WinBoot software (Yap and Nelson 1996) was used to compute bootstrap-based P-values to assess the strength of evidence for clustering obtained.

The polymorphism information content (PIC) of each microsatellite locus was determined as described by Weir (1990): $\text{PIC} = 1 - \Sigma P_i^2$, where P_i is the frequency of the ith allele in the examined

Table 2 Numbers of alleles per locus and PIC values of 23 polymorphic SSR loci based on levels of diversity revealed across 22 groundnut genotypes

SSR ID	Annealing temperature (°C)	Expected size of PCR product (bp)	Observed size range (bp)	Number of alleles	Number of repeats	PIC values
pPGPseq-3A08	64	152	152–200	6	$taa_{(20)}$	0.606
pPGPseq-8E12	59	198	200–230	2	$ttg_{(6)}$, $taa_{(15)}$	0.294
pPGPseq-14H6	59	285	280–380	10	$gt_{(31)}$	0.543
pPGPseq-4G02	60	285	300–420	6	$ga_{(9)}$, $gt_{(9)}$	0.398
pPGPseq-17F6	58	152	120–300	9	$Ga_{(35)}$	0.415
pPGPseq-13A7	58	265	280–350	2	$taa_{(10)}$	0.292
pPGPseq-13A10	57	264	270–350	5	$taa_{(12)}$	0.464
pPGPseq-15E8	58	298	300–380	3	$taa_{(14)}$	0.444
pPGPseq-4H11	60	269	250–350	7	$ga_{(26)}$, $gt_{(26)}$	0.294
pPGPseq-2D12B	60	265	300–390	9	$taa_{(16)}$	0.619
pPGPseq-18C5	60	281	270–320	7	$taa_{(23)}$	0.519
pPGPseq-2B10	58	259	260–310	4	$taa_{(16)}$	0.509
pPGPseq-2F05	58	262	270–300	4	$taa_{(19)}$	0.556
pPGPseq-12F7	57	290	230–270	5	$taa_{(12)}$	0.566
pPGPseq-8D9	61	132	120–170	6	$Ctt_{(13)}$	0.481
pPGPseq-3A01	64	238	250–300	6	$taa_{(22)}$	0.428
pPGPseq-15C10	64	203	220–300	7	$taa_{(16)}$	0.557
pPGPseq-3D09	63	292	290–300	6	$ga_{(19)}$, $gt_{(9)}$	0.557
pPGPseq-7G2	65	225	220–280	4	$tatc_{(12)}$	0.531
pPGPseq-16C6	65	230	250–300	4	$ga_{(18)}$	0.317
pPGPseq-10D4	62	203	200–300	8	$ga_{(24)}$	0.557
pPGPseq-16G8	60	194	230–290	7	$taa_{(19)}$	0.509
pPGPseq-1B09	64	282	260–340	8	$ga_{(19)}$	0.303

genotypes. An analysis of molecular variance (AMOVA) was used to partition genetic variability using Arlequin software version 2.0 (Schneider et al. 2000). Significance of estimated variance components was assessed based on 10,000 random permutations.

Single-marker analysis was used to detect potential associations between marker (genotypic) classes (presence or absence of the band) and their respective phenotypic values (disease score). The data on each marker were subjected to the non-parametric Kruskal–Wallis one-way analysis of variance (K–W ANOVA), using the KRUSKAL procedure in GenStat, to identify markers potentially linked to the disease. This was done by splitting the disease phenotyping data into two classes corresponding to the presence and the absence of band at each marker. We chose this non-parametric method instead of the usually adopted parametric ANOVA in view of the ordinal nature of the disease phenotyping data. The Kruskal–Wallis one-way analysis of variance was used to test the hypothesis that several (K) samples come from distributions with the same mean. The test statistic H, is formed by ranking the combined data set, then considering the sum of these ranks within each sample:

$$H=[(12/N*(N+1))*\Sigma j=1\cdots K\{Rj*Rj/nj\}] - 3*(N+1)$$

where R_j is the sum of ranks for the jth sample, n_j is the size of the jth sample, and N is the size of the combined data set. If ties are present in the data, then an adjustment to the statistic H is required:

$$\text{adjusted } H = H/(1-\Sigma k\{tk^3-tk\}/(N^3-N))$$

where tk is the number of observations with rank k. When there are at least five cases in each of the samples, H has approximately a Chi-square distribution on K–1 degrees of freedom. When this condition is not satisfied, and there are three samples, KRUSKAL uses a table of calculated values of the distribution of the statistic.

Results

A total of 135 alleles were revealed across 23 polymorphic SSR loci in the 22 cultivated groundnut genotypes. On average, each locus revealed approximately 6 alleles with 5 loci detecting 8 or more alleles. The PIC values of the 23 polymorphic SSR loci reveal a high level of polymorphism (Table 2) with 12 of the 23 SSR loci having a PIC value of ≥0.5. An example of the SSR polymorphism detected amongst the 22 groundnut genotypes is shown in Fig. 1.

Genetic diversity amongst cultivated groundnut germplasm

The cultivated groundnut collections included a total 22 accessions, encompassing 16 Spanish types (subspecies *fastigiata* var. *vulgaris*), 5 Virginia types (subspecies *hypogaea* var. *hypogaea*) and 1 valencia type (subspecies *fastigiata* var. *fastigiata*). The dendrogram constructed using Nei and Li's similarity coefficient and UPGMA clustering is presented with bootstrap values obtained from 2000 replicates of the data set are indicated at each branch point (Fig. 2).

The first two dimensions of the MDS plot indicate the presence of three well-separated clusters (Fig. 3) that correspond to the groupings identified at a similarity threshold of approximately 50% in the dendrogram (Fig. 2). The separation of cluster (a) from (b) and (c) is well supported, with a bootstrap value of 75%, however although the bootstrap values supporting the separation of clusters (b) and (c) is less robust, based on the combined MDS and cluster analyses, three clusters have been identified and the genotypes grouping in these three clusters will be detailed further. Both analyses reveal the following relationships among the cultivated groundnut genotypes studied. Cluster (a), comprises three genotypes: ICGV99001, ICGV99004 and ICG13917, all Spanish types and all resistant to LLS (however, there are also LLS resistant Spanish types present in clusters (b) and (c)). Cluster (b) includes 10 genotypes of two botanical types, Spanish and Virginia. Five of these genotypes, ICGV99003, ICGV86699, ICGV99005, ICGV99051 and ICGV99052, are varieties developed at ICRISAT. Varieties ICGV99003 and ICGV99005 are both interspecific derivatives, which are cytologically stable tetraploids (Dwivedi et al. 2001). Cluster (b) also includes two breeding lines with highly complex pedigrees, ICGV950084 and ICGV950166, the former having eight different genotypes and the latter with six different genotypes in their pedigree, respectively, all of which are lines of *Arachis hypogaea*. The three remaining genotypes in cluster (b), ICG11325, ICG11312 and ICG11331 are all breeding lines originating from Indian national breeding programs. Cluster (c) contains nine genotypes, five of which are landraces from South America, which group together separately within cluster (c) at a similarity level of 66%. Three genotypes within cluster (c), ICGV87165, ICGV99019 and ICGV87157 are varieties developed at ICRISAT; the first being an interspecific derivative. The rust and LLS susceptible control genotype, TMV2, is also present within this cluster, grouping with the other members of the cluster (c) at 56% similarity.

Analysis of molecular variation (AMOVA) of the 22 groundnut genotypes revealed a very low proportion of the total genetic variation associated with level of disease resistance and botanical type (Table 3). Table 3 indicates 6.67% of the total variation observed is accounted for by between resistance/susceptible groups, whereas the majority of the variation (91.65%) is accounted for by within botanical type groups.

Fig. 1 Polymorphism detected by pPBPseq-2D12B across 22 cultivated groundnut genotypes (see Table 1 for explanation of genotype codes 1–22); M: 100 bp molecular ladder

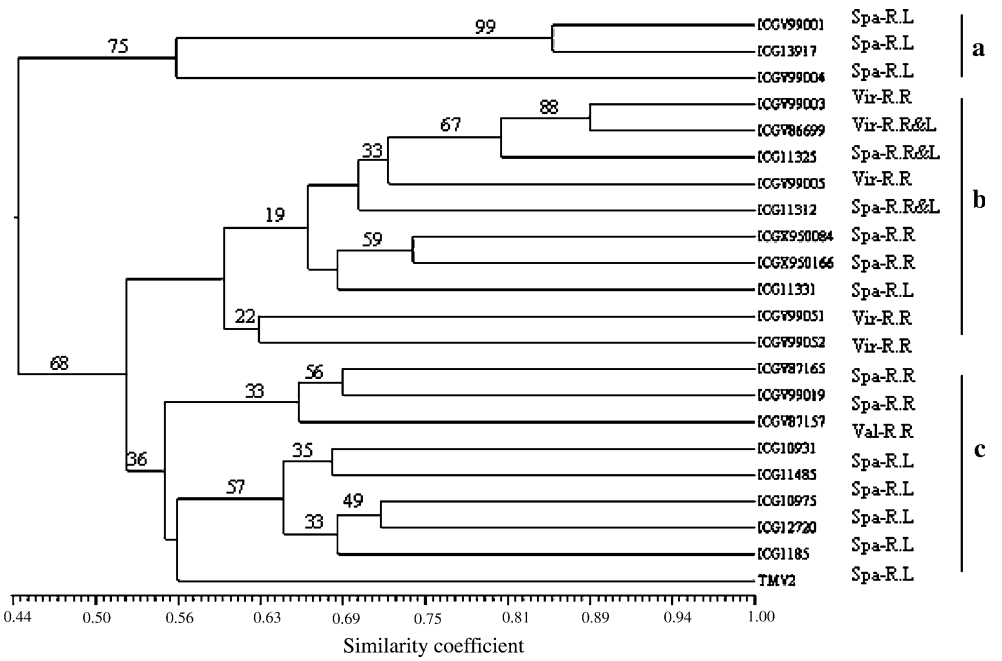

Fig. 2 Dendrogram constructed using Nei and Li's similarity coefficient and UPGMA clustering for 22 groundnut genotypes (Spa: Spanish; Vir: Virginia; Val: Valencia; R.R: resistant to rust; R.L: resistant to LLS; R&L: resistant to rust and LLS; S-R&L: susceptible to rust and LLS). Bootstraps values are given

Candidate markers for rust and /late leaf spot resistance

A locus-by-locus AMOVA was performed in order to obtain an estimate of how each locus contributes to the differentiation between disease-response type groups and compared with a Kruskal–Wallis one-way analysis of variance (ANOVA). The loci contributing greater than 60% of the total differentiation between the disease-response groups (based on locus-by-locus AMOVA) and also having a probability value of less than 0.05 from the Kruskal–Wallis one-way ANOVA were compared with results of QTL mapping studies (based on F_2 and F_6 populations for each disease, Mace pers. comm.). There is a high level of similarity between loci identified as linked to disease resistance genes by the different methods (Table 4). In particular, there is significant association between the three alleles detailed of SSR locus pPGPseq-17F6 and LLS and rust resistance as supported through all 3 comparative methods, with an AMOVA differentiation value of up to 74.96% and P values from the Kruskal–Wallis one-way ANOVA of $P = 0.001$ for rust resistance and $P = 0.030$ for LLS resistance, in addition to this locus being linked to both LLS and rust resistance QTLs as identified in a parallel study. Table 4 details thirteen additional loci, with a total of 25 alleles, with significant association to LLS or rust resistance loci as supported by a minimum of 1 of the 3 comparative statistical methods employed; of these, 11 alleles from 9 loci listed below are supported by a minimum of 2 of the comparative methods; pPGPseq-2B10, pPGPseq-2F05, pPGPseq-3A01, pPGPseq-8E12, pPGPseq-10D4, pPGPseq-12F7, pPGPseq-13A10, pPGPseq-13A7q-17 and pPGPseq-16C6. Table 5 details the polymorphism revealed by the 14 alleles from 10 loci detailed above, putatively associated with LLS and rust resistance genes as supported by a minimum of 2 of the 3 comparative statistical methods, between the parental genotypes of the ICRISAT LLS and rust mapping populations; ICGV99003 and ICGV99005 (rust resistance genotypes) and ICGV99001 and ICGV99004 (LLS resistant genotypes) and TMV2 (rust and LLS susceptible

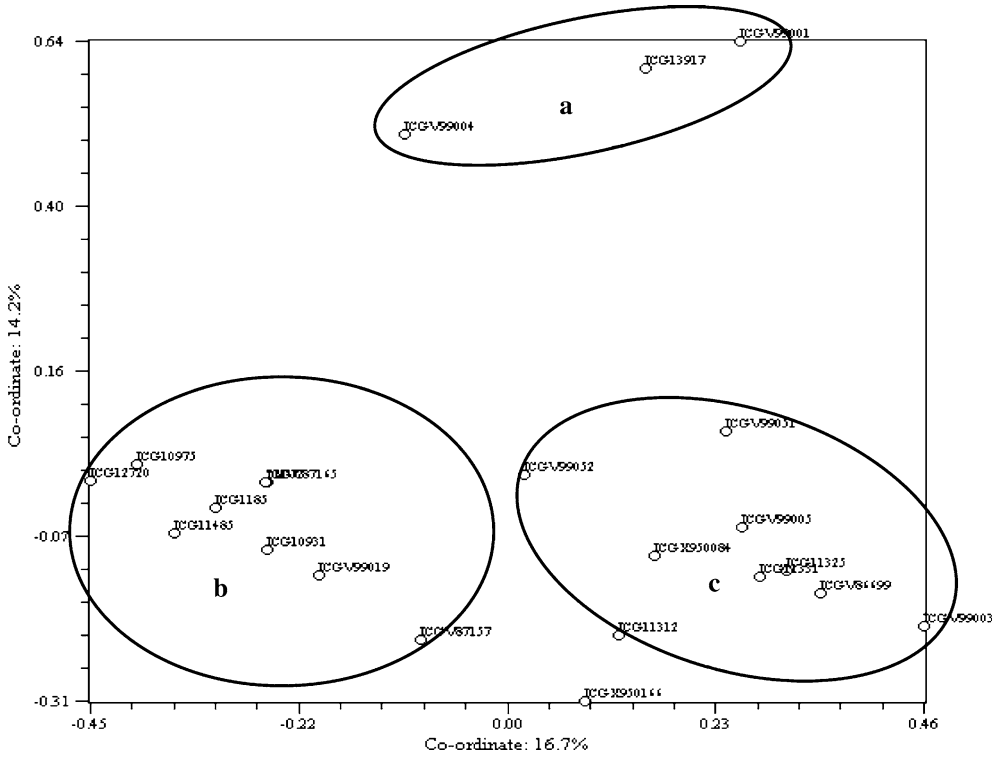

Fig. 3 MDS of SSR marker analysis across 22 cultivated groundnut genotypes. Three clusters of genotypes are indicated as 'a', 'b' and 'c'

variety). The polymorphisms revealed between the mapping population parental genotypes provide additional support for a few key loci; specifically pPGPseq-17F6 (alleles at 120, 140 and 150 bp), pPGPseq-2F05 (280 bp allele), pPGPseq-8E12 (200 bp allele) and pPGPseq-16C6 (263 bp allele) associated with rust resistance, with the alleles present in both rust resistant genotypes and absent in the susceptible genotype, and additionally pPGPseq-8E12 (210 bp allele) and pPGPseq-13A10 (250 bp allele) also associated with rust with the alleles absent in both rust resistant genotypes and present in the susceptible genotype. Table 5 also provides additional support for three loci putatively associated with LLS resistance genes; pPGPseq-2B10 (290 bp allele) and pPGPseq-2F05 (280 bp allele) with the alleles present in both LLS resistant genotypes and absent in the susceptible

Table 3 Analysis of molecular variance (AMOVA; Excoffier et al. 1992) for 22 genotypes grouped as four disease-response types within botanical type; employing 135 SSR alleles

Source of variation	d.f.[a]	SSD[b]	Variance components	Percentage of total variance	*P* values[c]
Between disease response groups	3	82.03	1.40	6.67	0.25
Between botanical type groups within disease response groups	3	60.04	0.35	1.68	0.23
Within botanical type groups	15	289.19	19.27	91.65	0.03*

[a] Degrees of freedom

[b] Sum of squared deviations

[c] Significance (*P*) of the variance components

* Significant at *P*=0.05

Table 4 List of alleles putatively linked with LLS and/or rust resistance as determined through locus-by-locus AMOVA % differentiation between disease resistant and susceptible groups and Kruskal–Wallis one-way analysis of variance, compared to results from QTL analyses of LLS and rust mapping populations

SSR ID	Allele (bp)	AMOVA % diff.	Marker linkage through QTL mapping[a]	Rust (VASI)		Rust (OPI)		LLS (OPI)	
				KW	χProb[b]	KW	χProb[b]	KW	χProb[b]
pPGPseq-8E12	200	81.93						5.31	0.021
pPGPseq-8E12	210	81.93						5.31	0.021
pPGPseq-14H6	295	19.51	Rust					4.40	0.036
pPGPseq-14H6	380	2.17	Rust					5.23	0.022
pPGPseq-17F6	120	62.27	Rust & LLS	10.98	0.001	7.12	0.008	4.69	0.030
pPGPseq-17F6	140	74.96	Rust & LLS	10.98	0.001	4.05	0.044		
pPGPseq-17F6	150	37.35	Rust & LLS	4.85	0.028	2.91	0.088	5.05	0.025
pPGPseq-13A7	305	7.79	LLS					4.60	0.032
pPGPseq- 13A10	250	5.243	LLS	4.15	0.042	4.54	0.033	5.31	0.021
pPGPseq- 13A10	260	1.66	LLS	2.96	0.085	5.07	0.024	3.20	0.074
pPGPseq-2D12B	300	10.61	LLS	2.85	0.091				
pPGPseq-2D12B	320	13.64	LLS	2.96	0.085				
pPGPseq-18C5	375	0.89		4.23	0.040	3.28	0.070		
pPGPseq-2B10	280	17.82	LLS	3.93	0.048			2.84	0.092
pPGPseq-2B10	290	63.19	LLS	3.68	0.055				
pPGPseq-2F05	280	17.06	LLS			3.05	0.081	7.20	0.007
pPGPseq-12F7	235	21.64	Rust					4.18	0.041
pPGPseq-3A01	250	2.36	Rust	5.44	0.020			4.88	0.027
pPGPseq-3A01	260	18.7	Rust					7.31	0.007
pPGPseq-3A01	370	2.36	Rust	5.44	0.020			4.88	0.027
pPGPseq-3A01	390	18.70	Rust					7.31	0.007
pPGPseq-16C6	260	0.13	Rust & LLS	2.75	0.098				
pPGPseq-16C6	263	22.72	Rust & LLS	4.39	0.036				
pPGPseq-16C6	275	4.88	Rust & LLS					3.13	0.077
pPGPseq-10D4	200	13.94		3.68	0.055			4.02	0.045
pPGPseq-10D4	235	81.70						2.97	0.085
pPGPseq-16G8	240	5.96		6.27	0.012				

[a] 'Rust' indicates the marker was found to be significantly linked to rust resistance genes in QTL analyses, 'LLS' to LLS resistance genes in QTL analyses and 'Rust & LLS' found significant linkages in both independent QTL analyses

genotype and pPGPseq-13A7 (305 bp allele) with the alleles absent in both LLS resistant genotypes and present in the susceptible genotype.

Based on the comparison of the 3 statistical methods employed to identify loci associated with LLS and rust resistance (Table 4) together with the assessment of polymorphism between the parental genotypes of the ICRISAT mapping populations (Table 5), 5 SSR loci have been identified with significant association to rust resistance genes (pPGPseq-17F6, pPGPseq-2F05, pPGPseq-8E12, pPGPseq-13A10 and pPGPseq-16C6) and 3 SSR loci have been identified with significant association to LLS resistance genes (pPGPseq-2B10, pPGPseq-2F05 and pPGP13A7); only one loci, pPGPseq-2F05, being associated with both LLS and rust resistance genes.

Discussion

Assessment of molecular diversity facilitates the identification of agronomically valuable and diverse germplasm for use in linkage mapping and genetic enhancement of specific traits in groundnut. Agronomically superior germplasm lines with relatively high DNA marker polymorphism have been identified for mapping rust and LLS disease resistance traits through the use of 23 SSRs in the present study, in addition to the

Table 5 Candidate SSR loci associated with rust/LLS resistance/susceptibility assessed through allele differentiation between genotypes used as parental lines in LLS and rust mapping populations at ICRISAT

SSR ID	Allele (bp)	LLS mapping population genotypes[a]			Rust mapping population genotypes[a]		
		ICGV99001	ICGV99004	TMV2[b]	ICGV99005	ICGV99003	TMV2[b]
pPGPseq-17F6	120	1	0	0	1	1	0
pPGPseq-17F6	140	1	0	0	1	1	0
pPGPseq-17F6	150	1	0	0	1	1	0
pPGPseq-2B10	290	1	1	0	0	0	0
pPGPseq-2F05	280	1	1	0	1	1	0
pPGPseq-3A01	250	0	0	0	0	1	0
pPGPseq-3A01	370	0	0	0	0	1	0
pPGPseq-8E12	200	1	0	0	1	1	0
pPGPseq-8E12	210	0	1	1	0	0	1
pPGPseq-10D4	235	0	0	0	0	0	0
pPGPseq-12F7	235	0	0	0	0	0	0
pPGPseq-13A10	250	0	1	1	0	0	1
pPGPseq-13A7	305	0	0	1	0	1	1
pPGPseq-16C6	263	0	0	0	1	1	0

[a] '1' indicates presence of the allele and 0 indicates absence of the allele

[b] For both mapping populations for the two traits, the susceptible parent is TMV2

identification of candidate marker-disease resistance trait associations.

The SSR marker-based genetic diversity analysis of 22 groundnut genotypes with known resistance to rust and/or LLS reported here, indicates that botanical group is a poor indicator of genetic diversity. Moreover, sources of disease resistance are available in Spanish, Virginia and Valencia types that have 60–70% genetic divergence. It is also clear that national groundnut breeding programs both in India and South America are producing highly related breeding material whereas ICRISAT disease resistant breeding lines fall into all genetic groups, as identified through cluster and MDS analyses. This suggests that ICRISAT's international groundnut breeding programs are effective vehicles for broadening the genetic base of the groundnut crop. This is highly valuable information both for the selection of genetically diverse material for use in groundnut disease resistance breeding programs and for the selection of parental genotypes for generating recombinant inbred line (RIL) mapping populations.

A maximum of 56% genetic dissimilarity across 135 alleles was observed amongst the 22 groundnut genotypes screened in this study. This is a significantly higher level of polymorphism than previously reported within the cultivated groundnut gene pool: 41% genetic dissimilarity revealed by RAPDs (Dwivedi et al. 2001) and 52% genetic dissimilarity revealed by AFLPs (He and Prakash 2001). The PIC scores revealed by the 23 loci screened are also very high with over 50% of loci having a PIC value of ≥0.5, however such high PIC values could be due to marker pre-selection and caution should be taken in interpreting the diversity revealed within the cultivated gene pool based on PIC values alone. The level of diversity revealed through cluster analyses, however, is the highest yet recorded between cultivated groundnut genotypes and supports previous observations that SSRs have a higher discriminatory power compared to other molecular markers (Powell et al. 1996b; Milbourne et al. 1997; Russell et al. 1997; Crouch et al. 1999). Our study also supports the suggestion by Singh et al. (1998) that the lack of genetic variation detected in cultivated groundnut is due to the limited range of genotypes previously used and type of molecular marker assay employed to detect molecular polymorphism.

There was no significant correlation between the number of repeats within a motif and the PIC of that SSR locus, although a positive correlation has been reported in other crops (Areshchenkova and Ganal 2002). For example, the locus pPBPseq-12F7 and pPBPseq-12G2 contain only

12 (taa) and (tatc) motifs, respectively but are highly polymorphic (PIC = 0.56 and 0.53). Other loci, e.g., pPBPseq-17F6 and pPBPseq-4H11 contain 35 (ga) and 26 (ga/gt) motifs but revealed much less molecular variation (PIC = 0.29 and 0.41, respectively) within the germplasm studied here. However, as expected there was reasonable correlation ($P = 0.005$; $r^2 = 0.32$) between the number of repeats within the motif and the number different alleles detected. Similarly, markers for dinucleotide repeats tend to detect a greater number of different alleles than trinucleotide repeat markers.

Considerable morphological variation in agronomic traits has been observed in groundnut, presumably resulting from positive breeding selection and agroecological adaptation (He and Prakash 2001). The current study is the first to report a parallel level of molecular genetic differentiation and morphological variation in the gene pool of tetraploid groundnut. The recently published first genetic linkage map of the tetraploid groundnut genome (Burow et al. 2001) was achieved by introducing variability from diploid wild species into tetraploid cultivated *A. hypogaea* through the use of a synthetic amphidiploid as a donor parent. This is clearly a highly effective means of capturing a high level of polymorphism for rapid and cost efficient linkage mapping. However, this approach may have limited value for molecular breeders as the power of selection of markers identified in this way may not be retained upon their application in breeding populations that will inevitably have very different recombination profiles. This clearly highlights the issue that careful selection of parental genotypes is critical when developing mapping populations, in order to capture the maximum amount of molecular polymorphism between the parental lines, and is particularly important in an autogamous, tetraploid, genetically impoverished crop species such as groundnut.

The diversity analysis reported in this study has also been used to identify candidate SSRs for use in mapping LLS and rust resistances and for application in MAS. In other crops, the applications of molecular markers in improving disease resistance have been extensively reported and alternative methods to the traditional QTL analyses of bi-parental segregating populations for identifying markers linked to traits of interest are increasingly being employed. For example, Sun et al. (2003) recently characterized the genetic diversity among 35 spring wheat cultivars and lines with different levels of *Fusarium* head blight (FHB) resistance using 160 RAPD markers; and found that ''association analysis between RAPD markers and the *Fusarium* index detected three RAPD markers significantly associated with FHB-resistance genotypes''. Their results suggested that a collection of unrelated genotypes can be used to identify markers linked to agronomically important traits, and that such markers can be used as candidate markers for further gene mapping. Sun et al. (2003) also highlight that this approach could have advantages over the use of mapping populations as the markers are more likely to be applicable to a large number of breeding programmes, whereas markers from traditional linkage studies require validation in diverse independent populations prior to application in molecular breeding.

The five SSR loci associated with rust resistance and the three SSR loci associated with LLS resistance identified in this study through comparison of results obtained from locus-by-locus AMOVA and Kruskal–Wallis one-way ANOVA on the diverse set of genotypes included in this study (Table 1) in contrast to the results obtained through more traditional QTL analyses undertaken on segregating F_2 and RIL populations of 4 ICRISAT mapping populations for rust resistance (ICGV99003 × TMV2 and ICGV99005 × TMV2) and LLS resistance (ICGV99001 × TMV2 and ICGV99004 × TMV2) indicate that non-traditional methodologies can be employed to associate genomic regions with traits of importance. This is particularly important in orphan crops such as groundnut which has very limited genomic resources developed to date, including genetic maps with sufficient marker coverage which are pre-requisites for a conventional mapping approach. It should be noted that the SSRs screened in this study have not yet been anchored to specific linkage groups in the groundnut genome. However, although marker-trait associations have been identified using a combination of approaches in this study, the results from the different methodologies were not always

concordant for each allele; for example pPGPseq-2F05 (280 bp allele) was only found to contribute 17.06% of the differentiation between the disease resistant and disease susceptible groups, from the locus-by-locus AMOVA result, whereas the Kruskal–Wallis one-way ANOVA and the QTL analyses significantly associated this allele with LLS resistance. Similarly, pPGPseq-13A7 (305 bp allele) was found to contribute 7.79% of the differentiation between the disease resistant and disease susceptible groups, from the locus-by-locus AMOVA result, when the Kruskal–Wallis one-way ANOVA and the QTL analyses significantly associated this allele with LLS resistance. The small sample size, both in terms of genotypes and loci screened, employed in this study may have contributed to the overall lack of concordance of results generated from different methodologies and in this particular study, the Kruskal–Wallis one-way ANOVA appeared more robust in the absence of a larger number of samples than the locus-by-locus AMOVA. In conclusion, the genetic diversity analysis undertaken has presented a valuable opportunity to additionally identify loci putatively linked to two traits of importance, LLS and rust resistance. Such genetic analysis studies to be undertaken on groundnut in the future would focus on assembling large sample sizes with adequate representation of the different states of the trait in question, bearing in mind that the putative association of loci and traits based on this approach is valid only for simply inherited, oligogenic traits, but acknowledging that this does have an important role in orphan crops such as groundnut with very limited genomic resources available to date. It is also acknowledged that the number of loci screened in the current study is low, particularly for a tetraploid species with 20 linkage groups and hence a close linkage between markers/alleles and loci controlling disease resistance cannot not be expected. Moreover, trait data are limited and are from 2 locations in 1 year only and may introduce bias. Despite these shortcomings, eight loci putatively linked to resistance loci have been identified and additionally the SSR data set generated for the 22 diverse groundnut genotypes with varying levels of resistance to rust and LLS has also provided critical information to breeders for planning future breeding strategies. It will also enable plant breeders to make informed decisions about parental selection for developing mapping populations, as demonstrated by Anderson et al. (1993) through the use of the PIC scores to select potential mapping parents with a high level of polymorphism. This type of analysis also offers a mechanism for breeders to counteract further genetic impoverishment of the cultivated groundnut gene pool.

Acknowledgements The authors gratefully acknowledge Mr. A. Gafoor for technical assistance, Mr. S. Hari Krishna for biometrical assistance, Dr. S.L. Dwivedi for selection of material, and Dr. M. Ferguson for early access to SSR markers. D.T. Phong carried out this research through a visiting scientist fellowship under the Asian Development Bank-funded project 'Rapid Crop Improvement for Poor Farmers in the Semi-Arid Tropics of Asia' (RETA 5945). Biotechnology research at ICRISAT is grateful for unrestricted grants from the governments of Japan, UK and EU.

References

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

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

Bromfield KR, Bailey WK (1972) Inheritance of resistance to *Puccinia arachidis* in peanut. Phytopath 62:748

Burow MD, Simpson CE, Starr JL, Pateron 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

Crouch JH, Crouch HK, Constandt H, Van Gysel A, Breyne P, Van Montagu M, Jarret RL, Ortiz R (1999) Comparison of PCR-based molecular marker analyses of *Musa* breeding populations. Mol Breed 5:233–244

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf material. Phytochem Bull 19:11–15

Dwivedi SL, Gurtu S, Chandra S, Yuejin W, Nigam SN (2001) Assessment of genetic diversity among selected groundnut germplasm. I: RAPD analysis. Plant Breed 120:345–349

Dwivedi SL, Crouch JH (2003) Proceedings of a Workshop for the Asian Development Bank supported project on molecular breeding of sorghum, groundnut and chickpea, ICRISAT:28–43

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

FAOSTAT (2004) Available at http://apps.fao.org/cgi-bin/nph-db.pl?subset = agriculture

Ferguson ME, Burow MD, Schultz SR, Bramel PJ, Paterson AH, Kresovich S, Mitchell S (2003) Microsatellite identification and characterization in peanut (*A. hypogaea* L.). Theor Appl Genet 108:1064–1070

Gimenes MA, Lopes CR, Valls JFM (2002) Genetic relationships among *Arachis* species based on AFLP. Genet Mol Biol 25(3):349–353

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

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

He G, Prakash CS (1997) Identification of polymorphic DNA markers in cultivated peanut (*Arachis hypogaea* L.). Euphytica 97:143–149

He G, Prakash CS (2001) Evaluation of genetic relationships among botanical varieties of cultivated peanut (*Arachis hypogaea* L.) using markers. Genet Resour Crop Evol 48:347–352

He G, Meng R, Newman M, Gao G, Pittman RN, Prakash CS (2003) Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biol 3:3–9

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

Kolodny GM (1984) An improved method for increasing the resolution and sensitivity of silver staining of nucleic acid bands in polyacrylamide gels. Anal Biochem 138(1):66–67

Kruskal JB, Wish M (1978) Multidimensional scaling. Sage, Newbury Park

Litt M, Lutty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeats within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Mace ES, Buhariwalla HK, Crouch JH (2003) A high throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol Biol Rep 21:459a–459h

Milbourne D, Meyer R, Bradshaw JE, Baird E, Bonar N, Provan J, Powell W, Waugh R (1997) Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. Mol Breed 3:127–136

Mohan M, Nair S, Bhagwat A, Krishma TG, Yano M, Bhatia CR, Sasaki T (1997) Genome mapping, molecular markers and markers-assisted selection in crop plants. Mol Breed 3:87–103

Murty UR, Jahnavi MR (1983) Breeding potential of interspecific tetraploids in *Arachis* L. In: Cytogenetics of *Arachis*, ICRISAT, (1983) 125–130

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction end nucleases. Proc Natl Acad Sci U S A 76:5269–5273

Nevill DJ (1981) Components of resistance to *Cercospora arachidicola* and *Cercosporidium personatum* in groundnut. Ann Appl Biol 99:77–86

Panaud T, Chen OX, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa L.*). Mol Gen Genet 252:597–607

Paramasivam K, Jayasekhar M, Rajasekharan R, Veerabadhiran P (1990) Inheritance of rust resistance in groundnut (*Arachis hypogaea*). Madras Agric J 77:50–52

Pande S, Rao J, Dwivedi SL (2002) Components of resistance to late leaf spot caused by *Phaeoisariopsis personata* in inter-specific derivaties of groundnut. Indian Phytopathol 55(4):444–450

Powell W, Mackray GC, Provan J (1996a) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996b) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB (1995) The use of microsatellite DNA makers for soybean genotypes identification. Theor Appl Genet 90:43–48

Rohlf FJ (2000) Numerical taxonomy and multivariate analysis system. Applied Biostatistics Inc., New York

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714–722

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci U S A 81:8014–8018

Schneider A, Roessli D, Excoffier L (2000) Software for population genetics data analysis. Arlequin ver 2000

Singh AK, Mehan VK, Nigam SN (1997) Sources of resistance to groundnut fungal and bacterial diseases: and update and appraisal. Information Bulletin No 50, ICRISAT

Singh AK, Smartt J, Simpson CE, Raina SN (1998) Genetic variation vis-à-vis molecular polymorphism in groundnut, *Arachis hypogaea* L. Genet Resour Crop Evol 45:119–126

Smulders MJM, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor Appl Genet 94:264–272

Subramanian V, Gurtu S, Rao RCN, Nigam SN (2000) Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. Genome 43:656–660

Sun G, Bond M, Nass H, Martin R, Dong Z (2003) RAPD polymorphism in spring wheat cultivars and lines with different level of Fusarium resistance. Theor Appl Genet 106:1059–1067

Tiwari SP, Ghewande MP, Misra DP (1984) Inheritance of resistance to rust and late leaf spot in groundnut (*Arachis hypogaea*). J Cytol Genet 19:97–101

Weir BS (1990) Genetic data analysis. Methods for discrete genetic data. Sinauer Associates, Inc., Sunderland. Mass. pp. 125

Witcombe JR, Hash CT (2000) Resistance gene deployment strategies in cereal hybrids using marker-assisted selection: gene pyramiding, three-way hybrids and synthetic parent populations. Euphytica 112:175–186

Yap V, Nelson RJ (1996) Winboot: A program for performing bootstrap analysis of binary data determine the confidence limits of UPGMA—based dendrograms. IRRI, Manila 1099, Philippines