PLANT GENETIC RESOURCES

Gene Diversity among Botanical Varieties in Peanut (*Arachis hypogaea* L.)

M. E. Ferguson,* P. J. Bramel, and S. Chandra

ABSTRACT

For the first time, sufficient numbers of molecular markers that reveal polymorphism in cultivated peanut (*Arachis hypogaea* L.) have become available for diversity assessments. In this study, the amount and distribution of genetic variation within and among six peanut botanical varieties, as well as its partitioning among three continents of origin, was assessed at 12 simple sequence repeat (SSR) loci by means of 10 sequence-tagged microsatellite site (STMS) primers. Eighty-nine alleles were revealed, varying from 2 to 17 per locus with an average of 7.4 alleles per locus. Greater differentiation was observed between varieties (Fst = 0.33), compared with between continents (Fst = 0.016). However, maximum differentiation was observed among continents within varieties (Fst = 0.366) for three varieties. Rogers' modified distance among varieties revealed the similarity of three varieties of subspecies *fastigiata* Waldron, namely *fastigiata*, *vulgaris* C. Harz, and *aequatoriana* Krapov. & W.C. Gregory. It did not support the inclusion of var. *peruviana* Krapov. & W.C. Gregory in this grouping. In addition, the results suggest that subsp. *hypogaea* var. *hypogaea* and var. *hirsuta* Köhler are not closely related and therefore should not hold the same subspecific ranking. Discriminant function analysis reveals a high degree of accordance between variety delimitation on the basis of morphological and molecular characters. Landraces from Africa and Asia were more closely related to each other than to those from South America. Nei's unbiased estimate of gene diversity revealed very similar levels of diversity within botanical varieties. Landraces from South America had the highest diversity, and possessed 90% of alleles, compared with Africa (63%) and Asia (67%).

An understanding of the distribution of genetic variation among peanut botanical varieties in relation to the geographical origin will enhance the efficiency of the conservation and utilization of diverse genetic resources for crop improvement. This is particularly relevant in a crop like peanut, which appears to have an extremely narrow genetic base, despite extensive morphological variation. The recent development of sequence-tagged microsatellite (STMS) markers that reveal substantial variation within cultivated peanut will make this type of diversity assessment possible (He et al, 2003; Ferguson et al, 2004).

A number of hypotheses exist in regard to the site of domestication of peanut from its most likely wild allotetraploid progenitor *A. monticola* Krapov. & Rigoni. The current geographical ranges of the two most likely donors of the A and B genomes, *A. duranensis* Krapov. & W.C. Gregory and *A. ipaënsis* Krapov. & W.C. Gregory, and *A. monticola*, overlap in the eastern foothills of the Andes in the region of northwestern Argentina and southern Bolivia. This, together with archaeological and morphological diversity evidence, indicates that this may be a region of origin (Krapovickas and Rigoni, 1957; Hammons 1994; Kochert et al., 1996). Simpson et al. (2002) alternatively suggests the West Coast of Peru or a more environmentally conducive region on the eastern slopes of the Cordillera as regions of origin on the basis of early archaeological evidence (3800–3500 BC).

After domestication, peanut dispersed and evolved into a number of morphologically distinct botanical varieties that now predominate and show high levels of diversity in particular geographical areas. Seven such "gene centers" have been identified (Fig. 1) (Krapovickas, 1969; Gregory and Gregory, 1976; Singh and Simpson, 1994).

The most recent classification by Krapovickas and Gregory (1994) divides groundnut into two subspecies (see Singh and Nigam, 1997, for an English translation), subsp. *hypogaea* and subsp. *fastigiata* Waldron, with two and four botanical varieties respectively. Classification into subspecies is based on the presence (subsp. *fastigiata*) or absence (subsp. *hypogaea*) of flowers and lateral branches on the central axis and sequential (subsp. *fastigiata*) as opposed to alternate (subsp. *hypogaea*) branching of vegetative and reproductive lateral branches. The differences are likely to be due to variation in a few major genes (Wynne and Coffelt 1982; Kochert et al., 1996). Subspecies *hypogaea* tends to be medium to long duration and usually exhibits dormancy. It can either exhibit a prostrate (runner) or erect (bunch) growth habit. In contrast, subsp. *fastigiata* is generally short duration with no dormancy and is entirely erect in growth habit (Bhapkar et al., 1986). Intermediates between the subspecies are rare, but do exist, particularly in South America. This sometimes makes taxonomy of the cultivated species unclear.

Arachis hypogaea subsp. *hypogaea* is further divided into varieties *hypogaea* and *hirsuta* Köhler. Subsp. *hypogaea* var. *hypogaea*, also known as the Virginia market type, is prostrate to erect with 2 to 3 seeds per pod. It is the predominant type and has a center of diversity in the most likely region of domestication, stretching up to Rondônia and north-west Mato-Grosso in Brazil where there are few examples of subsp. *fastigiata* (Fig. 1). It is thought to have been dispersed from the Bolivian and Amazonian geographic regions (Gregory et al., 1980),

M.E. Ferguson, International Institute for Tropical Agriculture (IITA), c/o ILRI, P.O. Box 30709, Nairobi, Kenya; P.J. Bramel, c/o Dr. John Peacock, KISR, P.O. Box 24885, 13109 Safat, Kuwait; S. Chandra, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, Andhra Pradesh, India. Received 3 July 2003. *Corresponding author (m.ferguson@cgiar.org).

Published in Crop Sci. 44:1847–1854 (2004).

677 S. Segoe Rd., Madison, WI 53711 USA

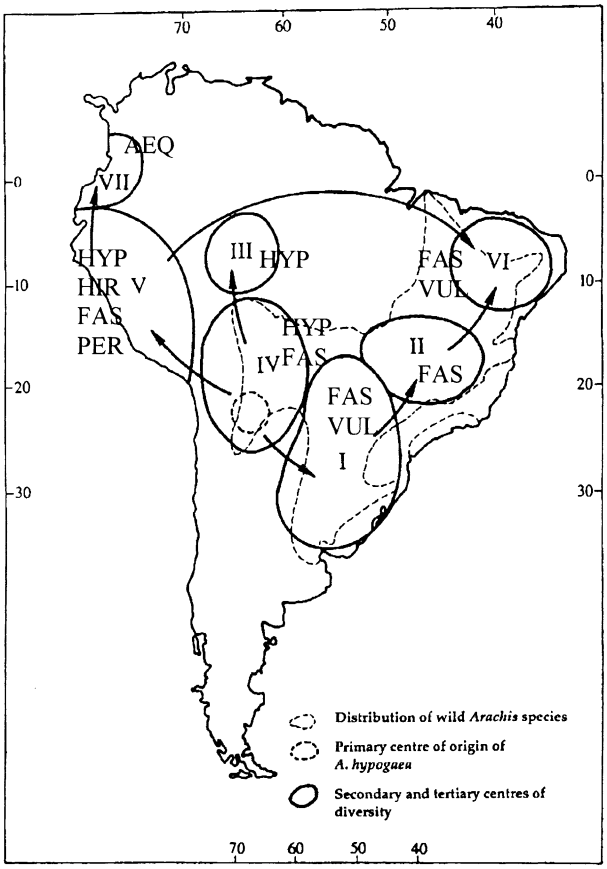

Fig. 1. Centers of origin and diversity of *A. hypogaea* in South America [adapted from Singh and Nigam (1997), from Gregory and Gregory (1976)]. HYP—subsp. *hypogaea* var. *hypogaea*; HIR—subsp. *hypogaea* var. *hirsuta*; FAS—subsp. *fastigiata* var. *fastigiata*; VUL—subsp. *fastigiata* var. *vulgaris*; PER—subsp. *fastigiata* var. *peruviana*; AEQ—subsp. *fastigiata* var. *aequatoriana*. I—Guaraní, II—Goiás and Minas Gerais (Brazil), III—Rondônia and north-west Mato Grosso (Brazil), IV—The eastern foothills of the Andes in Boliva, V—Peru, IV—North-eastern Brazil, VII—Ecuador.

but there is a need for further clarification. Variety *hirsuta*, consists of a distinctive group of large (main axis up to a meter in length), hairy, prostrate forms, whose pods are coarsely marked, have a parrot-like beak and contain 3 to 4 seeds. It has a center of diversity in Peru and was widely dispersed from there to Indonesia (Java), China, and Madagascar in the early 16th century (see Higgins, 1951, Merrill 1954; Dubard, 1906). It has not been reported recently from these regions, its decline in popularity being attributed to its extreme susceptibility to Cercospora leafspots [caused by *Cercospora arachidicola* (Hori) Deighton] (Gibbons et al., 1972).

Subspecies *fastigiata* includes varieties *fastigiata, vulgaris* C. Harz, *peruviana* Krapov. & W.C. Gregory, and *aequatoriana* Krapov. & W.C. Gregory. Variety *fastigiata*, the Valencia type, is generally erect with 3 to 4 seeds per pod. It is widely distributed in South America and predominates in five of the six centers of diversity (Fig. 1). It is thought to have been dispersed from Paraguay (Guaraní region) and central Brazil (Goiás and Minas Gerais) (Krapovickas, 1968) but Singh and Simpson (1994) propose a more likely point as the north-east coast of Brazil. These regions are dominated by distinct but different Valencia types, subsp. *hypogaea* being rare. Variety *vulgaris*, the Spanish type, is generally erect but with two seeds per pod. Its center of variation is in the Guaraní region and it was probably also disseminated from there (Krapovickas, 1969; Gregory and Gregory, 1976). Varieties *peruviana* and *aequatoriana* have narrow distributional ranges and centers of diversity in Peru and Ecuador respectively. Variety *peruviana* is distinguished from var. *fastigiata* by fruits with very marked reticulation and prominent longitudinal ribs and long strong reproductive branches (5–10 cm). Variety *aequatoriana* is characterized by leaflets with a hairy adaxial surface, long reproductive branches from the lateral branches and large, rough-looking pods containing 3 to 5 colorful seeds. Only three botanical varieties, subsp. *hypogaea* var. *hypogaea*, subsp. *fastigiata* var. *fastigiata* and var. *vulgaris*, are now widely cultivated in the Americas, Africa, and Asia.

When the early Spanish and Portuguese explorers arrived in the New World, peanut was being grown extensively in South America, Mexico, and the Caribbean Islands (Hammons, 1982). Widespread dispersion of different botanical varieties to Europe, Africa, and Asia occurred only after the discovery of the New World (Hammons 1994). Most authorities credit the Portuguese with taking the peanut to West Africa from Brazil, and via southern and East Africa to the Malabar coast of south-west India. Peanuts, particularly var. *hirsuta*, were also thought to have been distributed from the west coast of South America, across the Pacific to the Philippines, China and India (Singh and Simpson, 1994). By the middle of the 16th century, smaller-seeded runner types were introduced to North America possibly with the slaves from West Africa, but equally as likely through Mexico, Central America, and the Caribbean.

Previous studies that assessed diversity in cultivated peanut were constrained by an inability to visualize polymorphism adequately and thus, reported conflicting results. Lanham et al. (1994) found that seed proteins distinguished the two subspecies of *A. hypogaea* in 27 of 28 cases and Lu and Pickersgill (1993) found consistent differences in just two of 13 putative isozyme loci. Lacks and Stalker (1993) found that isozyme variation could not be associated with subspecies or botanical variety. Thus the objectives of our study is to quantify the degree and distribution of genetic variation within and among peanut botanical varieties as well as by continent of origin based upon DNA level diversity as indicated by 12 recently developed STMS markers.

MATERIALS AND METHODS

The germplasm used in the study consisted of 188 peanut landraces, which represented six botanical varieties from 10 countries in three continents. Of the 188 landraces, 80 were from South America (Bolivia, Ecuador, Paraguay, and Peru), 54 from Asia (India, Nepal, and China), and 54 from Africa (Burkina Faso, Nigeria, and Chad). Among the three continents, 54 landraces were sampled from var. *hypogaea* (18 each from the three continents), two from var. *hirsuta* (both from

Table 1. Sequences, length of amplified fragment from sequence from which primer was designed, repeat motif, and empirically defined annealing temperature of primers used in this study.

Primer ID	Forward Primer	Reverse primer	Fragment length (base pairs)	Repeat motif family	Optimum annealing temperature (°C)
pPGPseq-1B09	CGTTCTTTGCCGTTGATTCT	AGCACGCTCGTTCTCTCATT	282	ga	64
pPGPseq-2A05	GGGAATAGCGAGATACATGTCAG	CAGGAGAGAAGGATTGTGCC	252	taa	60
pPGPseq-2B10	AATGCATGAGCTTCCATCAA	AACCCCATCTTAAAATCTTACCAA	259	taa	58
pPGPseq-2C11	TGACCTCAATTTTGGGGAAG	GCCACTATTCATCGCGGTA	264	taa/cac	58
pPGPseq-2D12B	AAGCTGAACGAACTCAAGGC	TGCAATGGGTACAATGCTAGA	265	taa	60
pPGPseq-2G03	ATTCACAAGGGGACAGTTGC	ATTCAAGCCTGGGAAACAGA	215	taa	64
pPGPseq-2G04	TTCTTGGTTCCTTTGGCTTC	TGCTCAAGTGTCCTTATTGGTG	289	taa	60
pPGPseq-3A01	ATCATTGTGCTGAGGGAAGG	CACCATTTTTCTTTTTCACCG	238	taa	64
pPGPseq-4G02	TCAACTTTGGCTGCTTCCTT	TCAACCGTTTTTCACTTCCA	285	ga	60
pPGPseq-4H11	ATCACCATCAGAACGATCCC	TTTGTAGCCTTCTGGCGAGT	269	ga	60

South America), 54 from var. *vulgaris* (18 each from three continents), 20 from var. *peruviana* (all from Peru in South America), 54 from var. *fastigiata* (18 each from the three continents), and four from var. *aequatoriana* (all four from Ecuador in South America). A limited number of accessions of var. *hirsuta* and *aequatoriana* were used because of the lack of availability of germplasm and the predominance of the other varieties in commercial production. Accessions of vars. *hirsuta, peruviana,* and *aequitoriana* were only available from South America.

The 188 peanut landraces were screened for polymorphism at 12 loci, by means of 10 SSR primers. Primer sequence information, fragment length, repeat motif, and empirically defined optimum annealing temperature are given in Table 1; further information can be found in Ferguson et al. (2004). DNA was extracted from young, folded leaflets with Qiagen miniprep kits (Qiagen, Valencia, CA), and amplified by means of 10 pmol primer, 5 ng template DNA, 2 m*M* $MgCl_2$, 0.15 m*M* dNTPs and 1U Taq polymerase, 1× reaction buffer in a total reaction volume of 20 μL. Reaction conditions were 94°C for 2 min, 35 cycles of 94°C for 45 s, empirically defined annealing temperature (between 58 and 64°C; Table 1) for 1 min, 72°C for 90 s, then a final extension of 10 min at 72°C. Amplification products were visualized on nondenaturing 9% 29:1 (w/w) polyacrylamide/bisacrylamide gels followed by silver staining. Silver staining consisted of 3 min in water, 20 min in 0.1% (w/v) CTAB, 15 min in 0.3% ammonia solution, 15 min in a solution of 1 *M* NaOH, 0.1% silver nitrate and a few drops of a 25% ammonia solution, and a rinse in water (5–10 s) followed by development in a 1.5% sodium carbonate solution with 0.02% by volume formaldehyde solution. Gels were rinsed in water and fixed in a 1.5% glycerol solution. Amplification products were scored as present or absent.

Genetic polymorphism was measured in terms of the number of alleles per locus, Nei's unbiased estimate of gene diversity (*H*) (Nei 1987), and observed heterozygosity. As variation in a selfing species like peanut results mainly from the presence of different homozygotes, with heterozygotes being rare, the use of the term *gene diversity*, rather than *heterozygosity*, is more appropriate (Weir 1996) to describe genetic variation. The difference in value between *H* and observed heterozygosity can provide an indication of deviations from random mating in relation to Hardy-Weinberg equilibrium (Weir 1996).

Wright's *F*-statistics were computed to assess the degree of population differentiation among botanical varieties as well as among continents of origin. Resampling procedures of Jacknife (5000 replications) and bootstrap (2000 replications) over loci were used to obtain, respectively, the standard error (SE) and 95% confidence intervals for *F*-statistics. Matrices were calculated on the basis of Rogers' modified distance (Wright 1978) among varieties and continents. The distance matrices were subjected to UPGMA cluster analysis to assess the grouping of varieties and continents. All data, which represented six botanical varieties, were used to determine intervariety similarities and differences. Data from only three varieties *hypogaea*, *vulgaris,* and *fastigiata* (162 accessions) were used to determine intercontinent differences, as the other three varieties are endemic to South America. All analyses were done with TFPGA software (Miller 1997). In addition, discriminant function analysis (Mardia et al., 1979) was conducted by the DISCRIMINATE procedure of GenStat (version 6) to determine the degree to which morphological delimitation into botanical varieties was supported at the molecular level.

RESULTS

The 12 loci revealed a total of 89 alleles, varying from 2 to 17 alleles per locus with an average of 7.4 alleles per locus (Table 2). Accession numbers, used as stan-

Table 2. Number of alleles per locus and Fst statistics by locus, among varieties, among continents, and among continents within varieties.

Locus	Number of alleles	Fst over alleles among varieties	Fst over alleles among continents	Fst over alleles among continents within varieties	Nei's unbiased estimate of gene diversity (*H*)	Observed heterozygosity (direct count)
pPGPseq-1B09	3	0.603	0.000	0.6391	0.6440	0.0000
pPGPseq-2A05 (290 bp)	8	0.356	0.009	0.4106	0.7540	0.0053
pPGPseq-2A05 (310 bp)	4	0.418	0.022	0.0882	0.1880	0.0000
pPGPseq-2B10	4	0.491	0.001	0.5992	0.6610	0.0053
pPGPseq-2C11 (310 bp)	2	0.000	0.020	0.0588	0.0210	0.0000
pPGPseq-2C11 (340 bp)	11	0.369	0.005	0.3865	0.7321	0.0000
pPGPseq-2D12B	17	0.162	0.016	0.1968	0.8880	0.0053
pPGPseq-2G03	7	0.374	0.018	0.4393	0.6380	0.0053
pPGPseq-2G04	13	0.230	0.008	0.2551	0.8610	0.0000
pPGPseq-3A01	5	0.282	0.050	0.3181	0.6790	0.0000
pPGPseq-4G02	6	0.220	0.017	0.1556	0.5940	0.0106
pPGPseq-4H11	9	0.240	0.048	0.2635	0.7510	0.0000
Total	89	0.330	0.016	0.3665	0.6170	0.0027

Table 3. Alleles identified at each locus and the accessions that were used as a standard to represent that allele. The first number in parenthesis is the ICRISAT identifier (ICG), followed by the USDA Plant Introduction (PI) number where available.

Locus	Allele with accession representative
pPGPseq-1B09	a(10362); b(13422); c(11606, 476209)
pPGPseq-2A05 (310 bp)	a(10950); b(13124); c(10362); d(10037, 476174)
pPGPseq-2A05 (290 bp)	a(2858); b(2232); c(12112, 476177); d(10362); e(9987, 475867); f(10037, 476174); g(11183, 476020); h(12178, 476038)
pPGPseq-2B10	a(10362); b(13422); c(10037, 476174); d(12112, 476177)
pPGPseq-2D12B	a(10399); b(20002, 501296); c(12235, 468281); d(10044, 476177); e(10362); f(13422); g(15093); h(13124); i(12112, 476177); j(7283, 262012); k(7898, 407454); l(11183, 476020); m(11606, 476209); n(10359); o(10074, 476198); p(8253); q(14193, 540839)
pPGPseq-2C11 (310 bp)	a(14214, 540860); b(10362)
pPGPseq-2C11 (340 bp)	a(13263, 468283); b(10044, 476177); c(9596); d(20002, 501296); e(10074, 476198); f(12178, 476038); g(10950, 476178); h(14214, 540860); i(13422); j(5149, 261965); k(421, 152143)
pPGPseq-2G3	a(13901); b(12178, 476038); c(10950, 476178); d(11505); e(13422); f(20002, 501296); g(4210)
pPGPseq-2G4	a(12759, 476164); b(11183, 476020); c(12112, 476177); d(10037, 476174); e(10362); f(10399); g(3204); h(2232); i(15093); j(9987, 475867); k(11606, 476209); l(20002, 501296); m(14193, 540839)
pPGPseq3-A1	a(12235, 468281); b(12053, 475954); c(10950, 476178); d(11505); e(20002, 501296)
pPGPseq4-G2	a(10362); b(8253); c(10037, 476174); d(7283, 262012); e(10950, 476178); f(13263, 468283)
pPGPseq4-H11	a(10044, 476177); b(12112, 476177); c(12053, 475954); d(10362); e(15093); f(13422); g(20002, 501296); h(12665, 476025); i(9987, 475867)

dards, with the allele they represent are given in Table 3. Thirty-three accessions were required to represent all alleles, 23 of these are in the ICRISAT 'core' collection. Associated passport information, including collector number of these accessions can be viewed at http://singer.cgiar.org/index.htm; verified 13 April 2004.

Observed average heterozygosity across all loci and all accessions was extremely low (0.0027) and ranged from 0.0106 to 0.0053 (Table 2). There was a large difference between observed heterozygosity and Nei's unbiased estimate of gene diversity, *H*, which suggest nonrandom mating. According to Weir (1996), gene diversity will be close to the value for heterozygosity for randomly mating populations.

Differentiation among varieties had an Fst = 0.330 (with a Jacknife SE = 0.0425 and bootstrap 95% CI of 0.2578–0.4131) over all loci (Table 2). This is greater than the differentiation among three continents for three varieties, which was extremely low with an Fst of 0.0163 (with Jacknife SE = 0.006 and bootstrap 95% CI of 0.0279–0.0061) over all loci. Maximum differentiation was observed among continents within varieties where Fst was 0.366 (with Jacknife SE = 0.051 and bootstrap 95% CI of 0.2773–0.4648) over all loci, whereas differentiation between the three varieties alone was Fst = 0.309.

UPGMA clustering based on Rogers' modified distance (*D*) among varieties showed that the botanical varieties most closely related were var. *vulgaris* and var. *aequatoriana*, (D = 0.325), var. *vulgaris* and var. *fastigiata* (D = 0.368), and var. *aequatoriana* and var. *fastigiata* (D = 0.409), all within subsp. *fastigiata* (Fig. 2). The most distantly related varieties were var. *peruviana* and var. *fastigiata* (D = 0.562), also within subsp. *fastigiata*. This was followed by var. *hypogaea* and var. *hirsuta* (D = 0.561) within subsp. *hypogaea*. Botanical variety allocation as determined by discriminate function analysis reveals a high degree of accordance between botanical variety delimitation based on morpho-taxonomic classification and molecular estimates of DNA level diversity (Table 4). Maximum misclassification occurred in subsp. *aequatoriana* where 25% of accessions were classified as subsp. *vulgaris*.

Rogers' modified distance among continents across all botanical varieties was calculated. Africa and Asia were closely related with a genetic distance of 0.094. South America was distantly related to both Asia (D = 0.159) and Africa (D = 0.164) to nearly the same degree. Rogers' modified distance among botanical varieties within continents is given in Table 5, with the UPGMA dendrogram given in Fig. 3. When the relationships among the three continents were compared separately for each of the botanical varieties, var. *hypogaea* was more distinctly partitioned between South America and Asia or Africa than the other two varieties. In var. *vulgaris*, the three continents are nearly equal distance

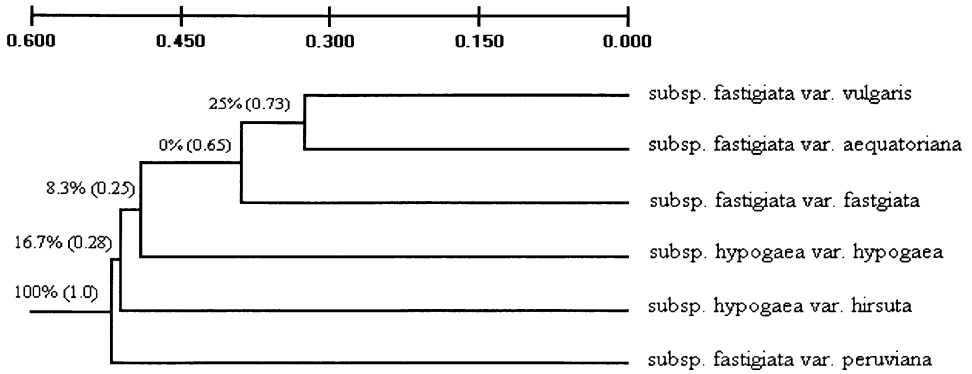

Fig. 2. Relationship between botanical varieties of peanut on the basis of Wright's (1978) modification of Rogers' distance using UPGMA clustering. At each node, the parentheses indicate the percentage of loci supporting the node and the proportion of similar replicates in 2000 bootstrap replications.

Table 4. Percentage of allocation of accessions to botanical varieties as a result of discriminant function analysis.

Taxon allocation based on SSRs		*A. hypogaea*		*A. fastigiata*			
Taxonomic classification		*hypogaea*	*hirsuta*	*vulgaris*	*peruviana*	*fastigiata*	aequatoriana
hypogaea	*hypogaea*	98	0	0	0	2	0
	hirsuta	0	100	0	0	0	0
fastigiata	*vulgaris*	7	0	91	0	2	0
	peruviana	0	0	0	100	0	0
	fastigiata	0	0	6	0	94	0
	aequatoriana	0	0	25	0	0	75

from each other. This result is illustrated in Fig. 3 where the continents within the three botanical varieties form the most distinct clusters.

When the distances between varieties were compared within each continent, var. *hypogaea* was distantly related to the other two varieties within Africa (Table 5). All varieties were most closely related in South America. In Asia, var. *hypogaea* was distantly related to var. *fastigiata* but not var. *vulgaris*. Also, vars. *vulgaris* and *fastigiata* were most distantly related in Asia.

In terms of allelic richness, South America harbored the largest number of alleles (90% of all observed alleles), followed by Asia (67%), and Africa (63%) (Table 6). Nei's unbiased estimate of gene diversity revealed very similar levels of diversity within the varieties. In South America varieties *fastigiata* and *aequatoriana* had the highest estimates of gene diversity ($H = 0.547$ and $H = 0.583$, respectively). When the levels of diversity are compared within each continent separately, the level of diversity was twice as much for South America in var. *hypogaea* compared to the other two continents. The level of diversity was very similar across all three continents in var. *vulgaris*, while the level of diversity was much higher for South America for var. *fastigiata* but lower for Asia.

DISCUSSION

The number and frequency of alleles at different DNA marker loci found in this study has not been previously reported in peanut. This has allowed for the first time, an analysis of the distribution of diversity at the molecular level within and among botanical varieties from three continents. Genetic diversity assessments are often more informative if they are considered in relation to diversity assessment in other studies. Reference accessions that harbor representative alleles for the 12 loci are listed to allow comparison of this study with future diversity assessments. In this study, the low level of heterozygosity and deviation from random-mating based on Hardy-Weinberg equilibrium were as expected for a highly inbreeding species.

Genetic Distance within and among Subspecies and Continent of Origin

When all six botanical varieties were used, greater differentiation was observed among botanical varieties (Fst = 0.33) than was observed among continents (Fst = 0.0163). However, the maximum differentiation occurred among continents within varieties (Fst = 0.366) when the analysis included the three varieties, *vulgaris*, *hypogaea*, and *fastigiata.* Both results indicate that the primary classification is at the botanical variety level, followed by a secondary classification at the continent of origin level. Discriminant function analysis highlights the relative integrity of botanical varieties, suggesting little introgression among varieties, and reveals that taxon delimitation on the basis of morphological characters is supported at the molecular level (Table 4). The distinction between subspecies found here is supported by a study of morphological traits and associated heterosis among different groups using principal component analysis by Isleib and Wynne (1983). Factors that may contribute to this observed maintenance of definition among subspecies include (i) the inbreeding nature of peanut, leading to reproductive isolation and rare hybridization events among subspecies; (ii) the fact that var. *hypogaea* is characterized by a medium to long duration and seed dormancy, whereas var. *fastigiata* is much earlier to mature and does not exhibit seed dormancy, thereby exerting a temporal barrier to hybridization; and (iii) the dominance of different subspecies and botanical varieties in different geographical regions (South America (Fig. 1) and only three of the six varieties widely distributed in Africa and Asia) leading to spatial separation. In addition in South America peanuts are grown in a variety of niches, including relatively

Table 5. Rogers' modified distance among continents within varieties.

		hypogaea			*vulgaris*			*fastigiata*		
Variety	Cont.	SA	Asia	Africa	SA	Asia	Africa	SA	Asia	Africa
hypogaea	SA†	–								
	Asia	0.327	–							
	Africa	0.340	0.164	–						
vulgaris	SA	0.481	0.559	0.595	–					
	Asia	0.445	0.467	0.515	0.200	–				
	Africa	0.518	0.570	0.607	0.182	0.143	–			
fastigiata	SA	0.469	0.551	0.591	0.348	0.395	0.425	–		
	Asia	0.556	0.637	0.667	0.393	0.501	0.500	0.303	–	
	Africa	0.517	0.593	0.629	0.293	0.402	0.395	0.282	0.172	–

† SA = South America.

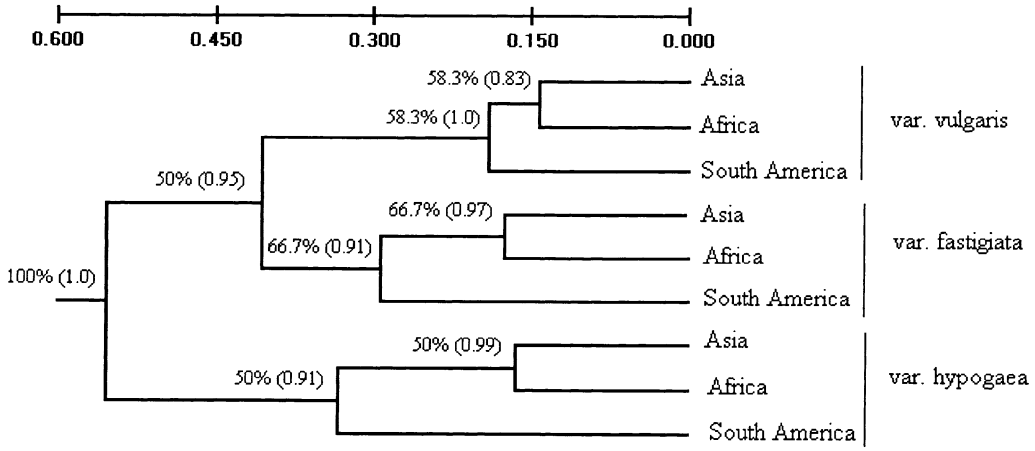

Fig. 3. Relationship among continent of origin within botanical varieties of peanut on the basis of Wright's (1978) modification of Rogers' distance using UPGMA clustering. At each node, the parentheses indicate the percentage of loci supporting the node and the proportion of similar replicates in 2000 bootstrap replication.

isolated, inaccessible slash-and-burn plots in the rainforests of the Amazonian lowlands and on sandbars left exposed by rivers at low water (Williams, 1989). This again would contribute to spatial isolation. In several reported instances farmers also try to maintain variety integrity for cultural reasons or for specific uses, for example in Ecuador, white peanuts are selected exclusively for candy (Becker, 1998) and in Bolivia peanuts are selected specifically for earliness so that they may be harvested before the rivers begin to rise again at the beginning of the rainy season (Williams, 1989).

Genetic Similarities among Botanical Varieties within Subspecies

Genetic distances (Fig. 2) reveal the relative similarity of three varieties of subspecies *fastigiata*, namely *fastigiata*, *vulgaris*, and *aequatoriana*. The close relationship of these varieties is consistent with the sympatric distribution of *fastigiata* and *vulgaris* in the Guaraní region and north-east Brazil (Fig. 1). According to the centers of diversity of botanical varieties (Fig. 1) and the close geographical proximity of var. *fastigiata* in Peru to the geographical center of diversity of var. *aequatoriana*, a closer relationship of var. *aequatoriana* to var. *fastigiata* rather than var. *vulgaris* was expected. Variety *peruviana* is quite distantly related to all other varieties, which is surprising because of the predominance of var. *fastigiata*, as well as subsp. *hypogaea* and *hirsuta* in its narrow region of diversity in Peru, Ecuador (Becker, 1998), and Bolivia (Williams, 1989). These data suggest limited introgression of var. *peruviana* with other varieties and a subspecific ranking of this taxon rather than its inclusion as a variety within subsp. f*astigiata.* In addition subsp. *hypogaea* var. *hypogaea* and var. *hirsuta* were distantly related, which suggests that they should also not hold the same subspecific designation. This result is based on only two accessions of var. *hirsuta*.

Genetic Distance among Botanical Varieties within Continents

The distinctness of varieties seems to some extent to be continent specific. In Africa, the distinction among botanical varieties is clear with var. *hypogaea* being more distantly related from var. *fastigiata* ($D = 0.629$) and var. *vulgaris* ($D = 0.607$) than they are to each other ($D = 0.395$). This is consistent with the current classification. In South America, botanical varieties appear to be less distinct than in Africa or Asia, although they do still reflect the classification based on morphological characters with *fastigiata* and *vulgaris* having a closer genetic distance ($D = 0.348$) than either has with var. *hypogaea* ($D = 0.469$ and $D = 0.481$, respectively). This lack of differentiation can be explained by the greater diversity present in South America both in terms of variety diversity (all six botanical varieties are present) and allelic diversity (see below), leading to wider

Table 6. Number of alleles and Nei's unbiased estimate of gene diversity by variety and by continent of origin.

Category		Total number of alleles		Nei's unbiased estimate of gene diversity (*H*)		
Subspecies	Variety		*n*†	SA‡	Asia	Africa
hypogaea	*hypogaea*	52	54	0.4952	0.2868	0.2063
	hisuta	20	2	0.4444		
fastigiata	*vulgaris*	58	54	0.4857	0.4648	0.3788
	peruviana	48	20	0.4962		
	fastigiata	57	54	0.5466	0.3503	0.4132
	aequatoriana	30	4	0.5833		
Continent	SA	71 (90%)	54	0.6279		
	Asia	53 (67%)	54		0.5565	
	Africa	49 (63%)	54			0.5336

† *n*: sample size (number of accessions).
‡ SA = South America.

variation in intermediary forms and less distinction among defined taxa. Evidence that introgressive hybridization does occur, at least among var. *fastigiata* and *vulgaris,* in the Guaraní region is provided by Krapovickas and Rigoni (1960) and Singh and Simpson (1994) who noted "hybrid swarms" of intermediate types. In Asia, subsp. *fastigiata* var. *vulgaris* is more closely related to subsp. *hypogaea* var. *hypogaea* (D = 0.467) than it is to its con-subspecific var. *fastigiata* (D = 501). This is contrary to morphological classification. Var. *hypogaea* is, as expected, distantly related to var. *fastigiata* (D = 0.637).

Genetic Distances among Continents

The low genetic differentiation observed between germplasm from Africa and Asia (D = 0.094) and their relative equidistant relationship to South America (D = 0.164 and D = 0.159, respectively) may be explained by (i) the existence particularly of Portuguese trade routes between Brazil and the African and Asian Portuguese colonies, which are thought to have facilitated both the introduction of similar germplasm to the two continents from Brazil as well as the exchange of germplasm between Africa and Asia; (ii) trade routes from Europe to the Indian subcontinent, Sri Lanka, and the Far East via the southern and eastern coasts of Africa (Gibbons et al., 1972) and possibly West Africa; (iii) in Africa, the evolution and adaptation particularly of var. *hypogaea* and var. *fastigiata* into new and distinct patterns of variation, presumably in response to different environmental conditions and specialized agricultural requirements (Gibbons et al., 1972); and (iv) well-established trade routes between the Indian subcontinent and East Africa that facilitated the distribution of new variation from Africa to Asia [although this study based on germplasm from West (Nigeria and Burkina Faso) and Central (Chad) Africa, does not necessarily reflect this route]. Singh and Simpson (1994) report that there were introductions to East Africa from Asian countries.

Genetic Distance within Varieties among Continents

When the genetic distance within the three main varieties var. *hypogaea*, var. *fastigiata*, and var. *vulgaris* is viewed in relation to continent of origin, we can see that in all varieties' germplasm from Africa and Asia are closely related, reflecting the overall relationship found among continents. Germplasm of var. *vulgaris*, however, showed less distinction among continents, and germplasm from South America of this variety was more closely related to that in both Africa and Asia in other varieties. In addition, var. *vulgaris* exhibits particularly high gene diversity in Asia (H = 0.4648) (Table 6), only slightly lower than in South America (H = 0.4857). This suggests that in var. *vulgaris*, rather than the evolution of novel variation in response to changing environmental conditions or use, variation in Asia and Africa is due, to some extent, to the introduction of variation from South America. This is consistent with the observations of Gibbons et al. (1972) of cultivar clusters of var. *vulgaris* (Spanish, Natal and Manyema) in Africa having similarities with that of three regions in South America (the Guaraní region, the region of the eastern slopes of the Andes in Bolivia, and parts of western Brazil) suggesting introductions from these regions.

Genetic Diversity among Subspecies and Varieties

Nei's unbiased estimate of gene diversity revealed very similar levels of diversity within the varieties. This is surprising considering the narrow distributional ranges of var. *hirsuta*, *aequatoriana*, and *peruviana* in relation to the other three botanical varieties.

Genetic Diversity among Continents

Landraces from South America showed the greatest diversity among continents, harboring 90% of alleles, compared with Africa (63%) and Asia (67%). This was also reflected in Nei's unbiased estimate of gene diversity with South America having a value of 0.6279, followed by Asia (0.5565) and Africa (0.5336). This is an expected result as it is the center of origin of the peanut where all wild species are endemic and all six botanical varieties originated and are cultivated. Despite the concentration of diversity in South America, germplasm from South America makes up less than 16% of the peanut collection in the ICRISAT genebank and 39% in the USDA collection. There is an urgent need to collect, conserve and utilize germplasm from South America to broaden the genetic base of peanut.

Genetic Diversity among Varieties within Continents

Varieties *hypogaea* and *vulgaris* showed the same trend of maximum diversity in South America, followed by Asia and Africa. Variety *fastigiata* had the highest diversity in South America, and slightly more diversity in Africa than in Asia. Variety *aequatoriana* was found to harbor maximum diversity in South America (H = 0.5833), with 30 alleles from four accessions. There is an urgent need to collect and conserve this underrepresented yet potentially valuable diversity. In South America, var. *hirsuta* has the lowest diversity and the smallest number of alleles. This limited diversity as well as its low harvest index, late maturity, low yield, and long fragile pegs may explain its decline in popularity after its initial dispersal to the Far East, China, India, and Madagascar from the west coast of South America. The susceptibility of var. *hirsuta* to Cercospora leafspots is possibly an indication of the observed narrow genetic base of this variety.

The low diversity of var. *hypogaea* in Africa (H = 0.2063) is surprising due to observations of high morphological variation for this variety in the continent (Gibbons et al., 1972). Our sampling in just three countries (Burkina Faso, Nigeria, and Chad) located fairly closely together may have biased the diversity assessment as diversity is reported to show geographical patterns and the spectrum of variants produced in Central

Africa to be characteristic of the region and quite distinct from that of West Africa (Smartt, 1990).

ACKNOWLEDGMENTS

The authors are grateful to Ms. Manisha Singh and Mr. Shiwanand Varma for technical assistance and to The World Bank and the Common Fund for Commodities for financial assistance.

REFERENCES

Becker, H. 1998. Unique peanuts bring nations together. Agricultural Research. Sept. 10–11.

Bhapkar, D.G., P.S. Patil, and V.A. Patil. 1986. Dormancy in groundnut—A review. J. Maharashtra Agric. Univ. 11:68–71.

Dubard, Monsieur. 1906. De l'origine de l'arachide. Muséum National d'Histoire Naturelle (Paris). 5:340–344.

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

Gibbons, R.W., A.H. Bunting, and J. Smartt. 1972. The classification of varieties of groundnut (*Arachis hypogaea* L.). Euphytica 21:78–85.

Gregory, W., and M. Gregory. 1976. Peanut. p. 151–154. *In* N. Simmonds (ed.) Evolution in crop plants. Longman, London.

Gregory, W.C., A. Krapovickas, and M.P. Gregory. 1980. Structures, variation, evolution and classification in *Arachis*, p. 469–481. *In* R.J. Summerfield and A.H. Bunting (ed.) Advances in legume science. Royal Botanic Gardens, Kew, London.

Hammons, R.O. 1982. Origin and early history of the peanut. p. 1–20. *In* H.E. Pattee and C.T. Young (ed.) Peanut science and technology. American Peanut Research and Education Society, Yoakum, TX.

Hammons, R.O. 1994. The origin and early history of the peanut. p. 24–42. *In* J. Smartt (ed.) The peanut crop: A scientific basis for improvement. Chapman and Hall, London.

He, G., R. Meng, M. Newman, G. Gao, R.N. Pittman, and C.S. Prakash. 2003. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biology. http://www.biomedcentral.com/1471-2229/3/3; verified 9 April 2004.

Higgins, B.B. 1951. Origin and early history of the peanut. p. 18–27. *In* The peanut—The unpredictable legume. The National Fertilizer Association, Washington DC.

Isleib, T.G., and J.C. Wynne. 1983. Heterosis in test crosses of 27 exotic peanut cultivars. Crop Sci. 23:832–841.

Kochert, G., H.T. Stalker, M. Gimenes, L. Galgaro, C. Romero Lopes, and K. Moore. 1996. RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (*Leguminosae*). Am. J. Bot. 83:1282–1291.

Krapovickas, A. 1968. Origen, variabilidad y difusión del maní (*Arachis hypogaea*). Actas y Memorias del XXXVII Congreso Internacional de Americanistas 2:517–534. see also Smartt, J. (translator). 1969. The origin, variability and spread of the groundnut (*Arachis hypogaea*). p. 427-441. *In* P.J. Ucko and I.S. Falk (ed.) The domestication and exploitation of plants and animals. Gerald Duckworth, London.

Krapovickas, A. 1969. The origin, variability and spread of the peanut (*Arachis hypogaea*). p. 427–440. *In* J. Ucko and C. Dimbleby (ed.) The domestication and exploitation of plants and animals. Duckworth, London.

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

Krapovickas, A., and V.A. Rigoni. 1957. Neuvas especies de Arachis vinculadas al problema del origen del maní. Darwiniana 11:431–455.

Krapovickas, A., and V.A. Rigoni. 1960. La nomenclature de las subspecies y variedades de *Arachis hypogaea* L. Rev. Invest. Agric. 14:197–228.

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

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

Lu, J., and B. Pickersgill. 1993. Isozyme variation and species relationships in peanut and its wild relatives (*Arachis* L.—*Leguminosae*). Theor. Appl. Genet. 85:550–560.

Mardia, K.V., J.T. Kent, and J.M. Bibby. 1979. Multivariate analysis. Academic Press, London.

Merrill, E.D. 1954. The botany of Cook's voyages. Chron. Bot. 14: 161–354.

Miller, M.P. 1997. Tools for population genetic analysis (TFPGA) 1.3. A Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by the author, see http://bioweb.usu.edu/mpmbio/index.htm; verified 9 April 2004.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Simpson, C.E., A. Krapovickas, and J.F.M. Valls. 2002. History of *Arachis* including evidence of *A. hypogaea* L. progenitors. Peanut Sci. 28:79–81.

Singh, A.K., and C.E. Simpson. 1994. Biosystematics and genetic resources. p. 96–137. *In* J.Smartt (ed.) The peanut crop: A scientific basis for improvement. Chapman and Hall, London.

Singh, A.K., and S.N. Nigam. 1997. Peanut. p. 114–127. *In* D. Fuccillo et al. (ed.) Biodiversity in trust: Conservation and use of plant genetic resources in CGIAR centers. Cambridge University Press, Cambridge, England.

Smartt, J. 1990. The groundnut, *Arachis hypogaea* L. p. 30–84. *In* Grain legumes: Evolution and genetic resources. Cambridge University Press, Cambridge, England.

Weir, B.S. 1996. Diversity. p. 141–159. *In* Genetic data analysis II. Methods for discrete population genetic data. Sinauer Associates, Inc. Publishers, Sunderland, MA.

Williams, D. 1989. Exploration of Amazonian Bolivia yields rare peanut landraces. Diversity 5:12–13.

Wright, S. 1978. Evolution and the genetics of populations. Vol 4. Variability within and among natural populations. University of Chicago Press, Chicago.

Wynne, J.C., and T.A. Coffelt. 1982. Genetics of *Arachis hypogaea* L. p. 50–94. *In* H.E. Pattee and C.T. Young (ed.) Peanut science and technology. American Peanut Research and Education Society, Yoakum, TX.