JIPB Journal of Integrative Plant Biology

Analysis of genetic diversity and population structure of peanut cultivars and breeding lines from China, India and the US using simple sequence repeat markers

Research Article

Hui Wang[1,2,3,4], Pawan Khera[2,5], Bingyan Huang[6], Mei Yuan[4], Ramesh Katam[7], Weijian Zhuang[8], Karen Harris-Shultz[9], Kim M. Moore[10], Albert K. Culbreath[2], Xinyou Zhang[6], Rajeev K. Varshney[5], Lianhui Xie[1]* and Baozhu Guo[3]*

[1]Fujian Agricultural and Forestry University, College of Plant Protection, Fuzhou 350002, China, [2]Department of Plant Pathology, University of Georgia, Tifton, GA 31794, USA, [3]USDA-ARS, Crop Protection and Management Research Unit, Tifton, GA 31794, USA, [4]Shandong Peanut Research Institute, Qingdao 266100, China, [5]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, India, [6]Henan Academy of Agricultural Sciences, Cash Crops Research Institute, Zhengzhou 450002, China, [7]Department of Biological Sciences, Florida A&M University, Tallahassee, FL 32307, USA, [8]Fujian Agricultural and Forestry University, College of Crop Science, Fuzhou 350002, China, [9]USDA-ARS, Crop Genetics and Breeding Research Unit, Tifton, GA 31794, USA, [10]AgResearch Consultants, Shingler Little River Road, Sumner, GA 31789, USA. *Correspondences: fjxlh@126.com; baozhu.guo@ars.usda.gov (Baozhu Guo is fully responsible for distributions of all materials associated with this article)

Abstract Cultivated peanut is grown worldwide as rich-source of oil and protein. A broad genetic base is needed for cultivar improvement. The objectives of this study were to develop highly informative simple sequence repeat (SSR) markers and to assess the genetic diversity and population structure of peanut cultivars and breeding lines from different breeding programs in China, India and the US. A total of 111 SSR markers were selected for this study, resulting in a total of 472 alleles. The mean values of gene diversity and polymorphic information content (PIC) were 0.480 and 0.429, respectively. Country-wise analysis revealed that alleles per locus in three countries were similar. The mean gene diversity in the US, China and India was 0.363, 0.489 and 0.47 with an average PIC of 0.323, 0.43 and 0.412, respectively. Genetic analysis using the STRUCTURE divided these peanut lines into two populations (P1, P2), which was consistent with the dendrogram based on genetic distance (G1, G2) and the clustering of principal component analysis. The groupings were related to peanut market types and the geographic origin with a few admixtures. The results could be used by breeding programs to assess the genetic diversity of breeding materials to broaden the genetic base and for molecular genetics studies.

Keywords: *Arachis hypogaea*; breeding; genetic diversity; population structure; simple sequence repeat
Citation: Wang H, Khera P, Huang B, Yuan M, Katam R, Zhuang W, Harris-Shultz K, Moore KM, Culbreath AK, Zhang X, Varshney RK, Xie L, Guo B (2016) Analysis of genetic diversity and population structure of peanut cultivars and breeding lines from China, India and the US using simple sequence repeat markers. **J Integr Plant Biol** 58: 452–465 doi: 10.1111/jipb.12380
Edited by: Sanwen Huang, The Institute of Vegetable and Flowers, CAAS, China
Received May 25, 2015; **Accepted** Jul. 13, 2015
Available online on Jul. 15, 2015 at www.wileyonlinelibrary.com/journal/jipb

INTRODUCTION

Cultivated peanut (*Arachis hypogaea* L.) is an allotetraploid ($2n = 4x = 40$) that originated from hybridization of two ancient diploid species, *A. duranensis* (A-genome) and *A. ipaënsis* (B-genome), followed by chromosome duplication (Seijo et al. 2004; Leal-Bertioli et al. 2009). Peanut is one of the major economically-important legumes grown worldwide in semiarid environments and needs relatively low inputs of chemical fertilizer. Peanut is a major source of protein and vegetable oil for human nutrition, containing about 25% protein and 50% oil. The largest producers of peanut are China and India, followed by Nigeria and USA. All four countries, together, account for 69.2% of global production (FAO 2014). *A. hypogaea* belongs to genus *Arachis* of family *Fabaceae* and is divided into two subspecies, *hypogaea* and *fastigiata*; *A. hypogaea* subsp. *hypogaea* lacks flowers on the main axis, which can be divided into two botanical varieties var. *hypogaea* and var. *hirsuta*, while *A. hypogaea* subsp. *fastigiata* have flowers on the main axis and it can be divided into four botanical varieties var. *fastigiata*, var. *vulgaris*, var. *aequatoriana* and var. *peruviana*. Furthermore, modern cultivars are generally classified into four market types Virginia, Spanish, Valencia and Runner (Krapovickas and Gregory 1994). Virginia and Runner types are in the hypogaea subspecies, while Spanish and Valencia are in the fastigiata subspecies.

Nine billion people are estimated to live on the earth by the mid 21st century and farmers face the challenge of feeding them (Nature Editorial 2010). Advances in food production will require greater efforts in agricultural research to increase crop

yield along with improvements for plant protection from biotic and abiotic stresses (Guo et al. 2012). Peanut is grown mostly in the semiarid area of the world that utilizes self-sufficiency farming, and there is a requirement to increase the production and productivity in order to meet the demand of the increasing world population. During the years, efforts have been made to improve peanut yield, quality, disease resistance, and other agronomic traits resulting in great improvement of many traits in peanut (Guo et al. 2013; Varshney et al. 2013). Cultivar improvement by conventional breeding selection, intra-specific or inter-specific hybridization remain to play an important role in the increase of peanut yield (Tang et al. 2007; Guo et al. 2012). Genetic improvement in yield by hybridization and selection has more than doubled in the past. However, yield improvements have been slow in recent years. One possibility may be related to the limited genetic diversity among the high-yielding cultivars used as parents for hybridization in breeding programs. There is evidence that the allelic diversity in Runner-type peanut has significantly increased over the years and the diversity in Virginia type fluctuated greatly over the years (Milla-Lewis et al. 2010a, 2010b). Parents with diverse origins have a higher probability to produce superior progenies than those of similar ancestry, but it has become increasingly difficult to find high-yielding genotypes that do not have common parentage. Isleib et al. (2001) reported that in the Runner and Virginia market types, the average PI (plant inventory) ancestry of all cultivars was 17.9%. The parents of all Runner-type cultivars released in the US can be traced back to 13 PIs. Only three PIs occur in the pedigrees of Virginia-type cultivars and seven PIs are in the pedigrees of Spanish-type cultivars. Over the past 30 years, great efforts have been made to integrate unique germplasm into the breeding populations in the US in order to expand the peanut genetic base.

However, it is sometimes difficult to choose parents for hybridization solely based on their morphological features. Hence the knowledge of genetic diversity among released and pre-released peanut cultures is very important for germplasm utilization in peanut breeding. To broaden the genetic variation of cultivated peanut in future breeding programs, it is necessary to perform a more comprehensive study of genetic diversity and population structure of the peanut varieties (Ren et al. 2014). Therefore, research on the genetic diversity among cultivars is necessary for a more efficient and successful hybridization program because crosses involving more diverse parents are likely to produce higher heterotic effects and also more variability in the segregating generations (Suneetha et al. 2013). Furthermore, the development of reliable methods is necessary to allow for the assessment of genetic variability in germplasm collections for breeding programs. Among the numerous methods, DNA-based molecular markers technologies are the most reliable tools allowing for the assessment of genetic variability because they are not influenced by environment (Naito et al. 2008).

During the past two decades, molecular markers have become an indispensable and powerful tool for enabling breeding in many crops such as rice (Zhao et al. 2009; Babu et al. 2014), sorghum (Ramu et al. 2013), maize (Frascaroli et al. 2013), and wheat (Shakeel et al. 2013). Many types of molecular markers have also been developed for peanut, including restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and inter simple sequence repeat (ISSR) markers (Kochert et al. 1991; He and Prakash 2001; Raina et al. 2001; Herselman 2003; Vyas et al. 2014). In addition, intron-exon splice junction (ISJ), universal rice primer (URP), and directed amplification of minisatellite region DNA (DAMD) techniques also have been used for detecting peanut polymorphisms (Xiong et al. 2013). Among the different types of molecular markers used in peanut, SSRs have been effective and more informative in producing considerable higher polymorphic information content (PIC), and to date it is the most commonly used molecular marker in the peanut community (Guo et al. 2013). It is notable here that one of the key factors for crop improvement efforts depends on the amount and the use of genetic variability in breeding programs.

The objectives of this study were to select a set of highly polymorphic SSR markers that could be used in the assessment of genetic diversity among peanut germplasm and breeding lines in each breeding program, and to assess the levels of genetic diversity and population structure of peanut lines used in each breeding program in China, India and the US. In total, 111 highly polymorphic SSR markers were selected and used to analyze the genetic diversity levels and the population structure of 79 peanut genotypes (Table 1) from five different breeding programs in three countries. This assessment could provide more information for a breeding program's long-range plans and help in breeding strategies of selecting additional genetic sources.

RESULTS

SSR polymorphism

The 111 highly informative genetic and genomic SSR markers were selected from a previous screening of over 5,000 SSRs used in genetic mapping and QTL studies (Li et al. 2011; Qin et al. 2012; Pandey et al. 2014) which were distributed on the genetic maps (Tables S1, S2). The 79 peanut breeding lines used in this study were provided from each program and used in molecular marker analysis. In total, the 111 SSR markers produced 472 polymorphic bands. The number of alleles per locus varied widely among the markers, ranging from 2 to 9 alleles with an average of 4.25 alleles per marker (Table 2). The major allele frequency varied from 0.291 (TC28E09) to 0.987 (AS1R12A06) with the average of 0.644. The gene diversity index ranged from 0.025 (AS1R12A06) to 0.803 (TC4H07), with an average value of 0.480. The polymorphism information content (PIC) for all the SSR loci ranged from 0.025 (AS1R12A06) to 0.778 (TC4H07) with a mean of 0.429 (Tables 2, S1).

Genetic diversity analysis

Genetic diversity was calculated for the peanut breeding lines from each breeding program of the three countries (Table 3). Among the five breeding programs (SDPRI-China, HAAS-China, FAMU-India, ACI-US, CEL-US), HAAS had the highest number of all categories, total alleles, average alleles per locus, gene diversity, and PIC. There were a total of 366 alleles with an average of 3.41 per locus in HAAS, and the mean gene diversity and PIC values were 0.504 and 0.445, respectively, followed by FAMU, it had 353 total alleles with an average of

Table 1. Peanut breeding lines used in this study, the number used in dendrogram and structure grouping analysis, and the market type and origin

No.	Genotype	Market type	Origin†	Grouping based on dendrogram	Grouping based on structure
1	LHA01	Virginia	SDPRI-CN	G2d	P2b
2	LHA03	Virginia	SDPRI-CN	G2d	P2b
3	LHA04	Virginia	SDPRI-CN	G2b	P2b
4	LHA05	Virginia	SDPRI-CN	G2b	P2b
5	LHA07	Virginia	SDPRI-CN	G2d	P1b
6	LHA08	Virginia	SDPRI-CN	G2b	P2b
7	LHA09	Virginia	SDPRI-CN	G2d	P1b
8	LHB01	Virginia	SDPRI-CN	G2b	P2b
9	LHB02	Spanish	SDPRI-CN	G1	P1
10	NF13	Spanish	HAAS-CN	G2e	P2e
11	NF16	Spanish	HAAS-CN	G1	P1
12	NF19	Spanish	HAAS-CN	G2e	P2e
13	NF721	Virginia	HAAS-CN	G2d	P2b
14	NF74	Spanish	HAAS-CN	G1	P1
15	KNX260	Spanish	HAAS-CN	G2a	P2c
16	BS1016	Spanish	HAAS-CN	G1	P1
17	KNX	Spanish	HAAS-CN	G2a	P2c
18	KNX4566-1	Virginia	HAAS-CN	G2b	P2c
19	KNX4566-2	Virginia	HAAS-CN	G2b	P2c
20	Z1214	Spanish	HAAS-CN	G2b	P2c
21	Z1216	Virginia	HAAS-CN	G2b	P2b
22	Z1235	Virginia	HAAS-CN	G2d	P1
23	Z1239	Spanish	HAAS-CN	G2d	P1
24	Z1308	Virginia	HAAS-CN	G2b	P2c
25	Z4233	Spanish	HAAS-CN	G2a	P2c
26	KN61	Virginia	HAAS-CN	G2d	P2c
27	LH9	Virginia	SDPRI-CN	G2b	P2b
28	LH14	Virginia	SDPRI-CN	G2b	P2b
29	Kadiri-5	Spanish bunch	FAMU-IN	G1	P1
30	Vemana	Spanish bunch	FAMU-IN	G1	P1
31	Kadiri-3	Virginia bunch	FAMU-IN	G2a	P2a
32	K-1341-1	Spanish bunch	FAMU-IN	G2a	P2a
33	K-1341-2	Spanish bunch	FAMU-IN	G2a	P2a
34	K-1319	Spanish	FAMU-IN	G2a	P2a
35	K-1375	Spanish	FAMU-IN	G2a	P2a
36	JL-220	Spanish bunch	FAMU-IN	G2a	P2a
37	CSMG-84-1	Virginia runner	FAMU-IN	G2a	P2a
38	ICGS-76	Virginia bunch	FAMU-IN	G2a	P2a
39	TAG-24	Spanish bunch	FAMU-IN	G2c	P1
40	TPT-2	Spanish	FAMU-IN	G1	P1
41	GG-2	Spanish bunch	FAMU-IN	G1	P1
42	Kadiri-6	Spanish bunch	FAMU-IN	G1	P1
43	TKG-19A	Virginia bunch	FAMU-IN	G2c	P2b
44	M-13	Virginia runner	FAMU-IN	G2a	P2a
45	TIR-31	Virginia runner	FAMU-IN	G2c	P2a
46	yueyou20-5	Spanish	HAAS-CN	G1	P1
47	yueyou20-D	Spanish	HAAS-CN	G1	P1
48	FL fancy	Virginia	CEL-US	G2b	P2a
49	FI07	Runner	CEL-US	G2b	P2a
50	GA06	Runner	CEL-US	G2a	P2a
51	Georgia 06G	Runner	CEL-US	G2a	P2a
52	Gregory	Virginia	CEL-US	G2b	P2b
53	Hanoch	Virginia	ACI-US	G2b	P2b
54	M04-0149	Runner	ACI-US	G2a	P2a

(Continued)

Table 1. (Continued)

No.	Genotype	Market type	Origin†	Grouping based on dendrogram	Grouping based on structure
55	M08-0305	Runner	ACI-US	G2a	P2a
56	M09-0232	Runner	ACI-US	G2a	P2a
57	N10-0076	Runner	ACI-US	G2a	P2a
58	NC94022	Virginia × *hirsuta*	CEL-US	G2e	P2e
59	NC94022-4	Virginia × *hirsuta*	CEL-US	G2e	P2e
60	TamRunOL07	Spanish	CEL-US	G1	P1
61	Red River runner	Runner	CEL-US	G2a	P2a
62	SA Jewel	Spanish	ACI-US	G1	P1
63	SunOleic97R	Runner	CEL-US	G2a	P2a
64	SunOleic97R-8	Runner	CEL-US	G2a	P2a
65	TamRun	Runner	CEL-US	G2a	P2a
66	TamRun7	Runner	CEL-US	G2a	P2a
67	TamRun11	Runner	CEL-US	G2a	P2a
68	Tiftgard	Runner	CEL-US	G2a	P2a
69	Tifrunner-4	Runner	CEL-US	G2a	P2a
70	WT08-0883	Runner	ACI-US	G2a	P2a
71	WT09-0046	Virginia	ACI-US	G2b	P2b
72	WT09-0236	Spanish	ACI-US	G2a	P2a
73	WT09-0761	Virginia	ACI-US	G2b	P2b
74	WT09-0808	Runner	ACI-US	G2b	P2b
75	WT09-0814	Runner	ACI-US	G2a	P2a
76	WT11-0351	Virginia	ACI-US	G2b	P2a
77	WT11-1082	Runner	ACI-US	G2c	P2b
78	WT11-1083	Runner	ACI-US	G2b	P2b
79	M12-1314	Runner	ACI-US	G2a	P2a

Origin†: ACI, AgResearch consultants Inc.; CEL, commercial elite lines from US; CN, China; FAMU, Florida Agricultural & Mechanical University; HAAS, Henan Academy of Agricultural Science; IN, India; SDPRI, Shandong Peanut Research Institute; US, United States.

3.12 per locus, and the mean gene diversity and PIC values were 0.470 and 0.412, respectively. The least number of total alleles and average number of alleles per locus were from SDPRI, with 259 and 2.29, respectively. The lowest average gene diversity index and PIC were from ACI, with 0.306 and 0.269, respectively. By country, the 111 SSR markers revealed a total of 401 alleles in Chinese peanut lines, and the average number of alleles per locus was 3.55, whereas the lines from India and USA peanut lines contained a total of 353 and 362 alleles with a mean of 3.12 and 3.21 per locus, respectively (Table 3). The mean gene diversity for each SSR locus in China, India and USA peanut lines were 0.488, 0.470, and 0.363, respectively, and the mean PIC values for each SSR locus were 0.430, 0.412 and 0.323, respectively.

The results showed that the peanut lines from China had the highest total allele number, the highest gene diversity index and the highest PIC value (Table 3). The genetic diversity showed the following tendency, the order of total alleles and mean alleles per locus from high to low was China, then USA, and India, but the order of average gene diversity and PIC value was China, then India, and USA. These results indicated that the Chinese peanut breeding lines in whole had a higher level of genetic variation than Indian and US peanut breeding lines and the breeding program of HAAS in China had a higher level of genetic variation than the other programs. In addition, the parameter of major allele frequency did not change much among the breeding programs, and it did not exhibit significant correlation with allelic richness and the gene diversity.

Table 2. Summary of statistics of the 111 simple sequence repeat (SSR) markers used in this study

	MAF	Average alleles	GD	PIC
Max	0.9873	9.0	0.8028	0.7780
Min	0.2913	2.0	0.0250	0.0247
Mean	0.6439	4.25	0.4805	0.4293

GD, gene diversity; MAF, major allele frequency; PIC, polymorphism information content.

Genetic relationship among the peanut lines

Genetic distance was also used to evaluate the genetic diversity among the lines. The genetic distance among 79 peanut lines ranged from 2.4% to 86.0% with an average of 51.4% (Table 4). The largest genetic distance was between BS1016 and Gregory; BS1016 is a Spanish market type line from China and Gregory is a Virginia type from US NC94022 and NC94022-4 showed the least genetic distance, as they were selected from a breeding population (Qin et al. 2012) in the US. The average genetic distance among the peanut lines from China was the largest of 51.6%, ranging from 2.7% (NF13 and NF16) to 83.9% (NF19 and BS1016). Overall 53.79% peanut lines from China had a genetic distance greater than 50%. The

Table 3. Summary of genetic statistics of peanut breeding lines from different breeding program

Country	Program	NL	Total alleles	Average alleles	MAF	GD	PIC
China	SDPRI	11	259	2.29	0.743	0.348	0.294
	HAAS	19	366	3.41	0.646	0.504	0.445
	Total/Average	30	401	3.55	0.621	0.488	0.430
India	FAMU	17	353	3.12	0.650	0.470	0.412
USA	ACI	16	277	2.45	0.772	0.306	0.269
	CEL	16	322	2.85	0.720	0.386	0.341
	Total/Average	32	362	3.21	0.735	0.363	0.323

ACI, AgResearch consultants Inc.; CEL, commercial elite lines; FAMU, Florida Agricultural & Mechanical University; GD, gene diversity; HAAS, Henan Academy of Agricultural Science; MAF, major allele frequency; NL, number of peanut lines; PIC, polymorphism information content; SDPRI, Shandong Peanut Research Institute.

peanut lines from India had an average genetic distance of 50.76%, ranging from 3.33% (CSMG-84-1 and ICGS-76) to 78.01% (K-1341-1 and Kadiri-6); and 51.47% of the lines had a genetic distance greater than 50%. Finally, the average genetic distance among the peanut lines from the US was the lowest with 41.71%, ranging from 2.38% (NC94022 and NC94022-4) to 79.12% (TamRunOL07 and WT09-0808); and 24.60% of the lines had a genetic distance greater than 50%.

When comparing the average genetic distance among the breeding programs, it was found that HAAS had the largest average genetic distance of 53.41; and 62.57% of the lines had a genetic distance greater than 50%. The average genetic distance from SAPRI and ACI were 38.4 and 36.58%, respectively; and 12.73 and 15.00% of the lines had a genetic distance greater than 50%.

Grouping of peanut lines based on neighbor-joining

A dendrogram based on neighbor-joining was created from data of the 111 SSR markers, based country of origin (Figure 1A) or market types (Figure 1B). The 79 peanut lines were divided into two major groups, here denoted as G1 and G2. The smaller group G1 had 13 peanut lines, among which five were from India; six were from China and two from the US (Figure 1A). Interestingly, all of the peanut lines in G1 were Spanish market types (Figure 1B). The bigger group G2 included 66 peanut lines from all three countries, and grouped into G2a, G2b, G2c, G2d and G2e. Subgroup G2a had 30 peanut lines, in which 18 were from the US, nine from India and three from China. Based on the market type, the G2a subgroup consisted of 16 Runner market type lines, 10 Spanish type lines and four Virginia type (Figure 1B). G2b contained 20 peanut lines, among which nine were from the US, and 11 were from China. Among this subgroup, most peanut lines were Virginia market types, except three Runner types and one Spanish type. G2c was a small group with four peanut lines, three from India and one from the US. The G2d subgroup contained eight peanut lines, all from China. G2e had four peanut lines, two from China and two from the US. In total, 88% of the peanut lines from the

Table 4. Genetic distance based on 111 simple sequence repeat (SSR) markers among different peanut breeding programs

			Genetic distance (GD) (%)				
Country	Program	NL	Max./Lines†	Min./Lines	Average	No. of GD≥50%‡	Ratio of GD≥50% (%)§
China	SDPRI	11	62.4 2,5	6 9,27	38.4	7	12.73
	HAAS	19	83.9 12,16	2.7 10,12	53.4	107	62.57
	Total	30	83.9 12,16	2.7 10,12	51.6	234	53.79
India	FAMU	17	78 32,42	3.3 37,38	50.5	70	51.47
USA	ACI	16	65.7 62,78	6.9 53,71	36.6	18	15.00
	CEL	16	78.4 58,60	2.4 58,59	45.8	45	37.50
	Total	32	79.1 60,74	2.4 58,59	41.7	122	24.60
Total		79	86 16,52	2.4 58,59	51.4	1623	52.68

ACI, AgResearch consultants Inc.; CEL, commercial elite lines; FAMU, Florida Agricultural & Mechanical University; HAAS, Henan Academy of Agricultural Science; NL, number of peanut lines; SDPRI, Shandong Peanut Research Institute. †Lines, the line number corresponds to the serial number in Table 1. ‡The number of pairs with Genetic distance greater than 50% in each group. §The percentage of pairs with genetic distance greater than 50% in each group.

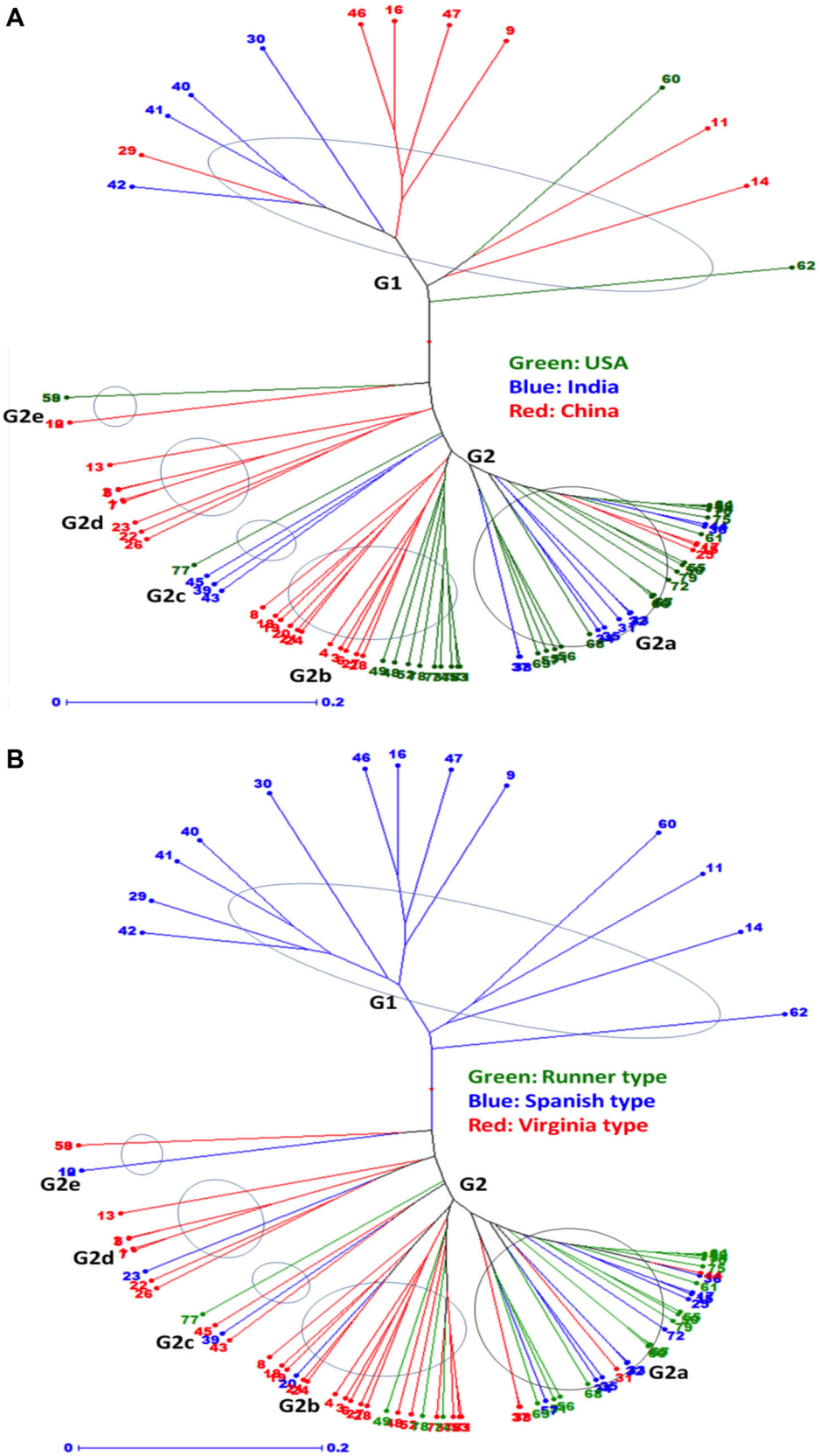

Figure 1. Dendrogram based on the genetic distance matrix among the peanut lines from three countries
(A) The colors of the branches indicate the origin of the peanut lines. **(B)** The colors of the branches indicate the market type of the peanut lines. The peanut lines were sorted according to the serial number from Table 1.

US grouped in G2a and G2b, 78% of the peanut lines from China grouped in G2b and G2d and peanut lines from India were almost evenly grouped in three different groups G1, G2a and G2c. Based on the market types, 48% of the Spanish type peanut lines grouped in G1 and 37% grouped in G2a; 80% of the Runner-type peanut lines grouped in G2a and 20% grouped in G2b; and 76% of the Virginia type peanut lines grouped in G2b, G2c, G2d and G2e. About 10% of the Runner types and 14% of the Spanish types placed in other subgroups (Figure 1B).

Factorial analysis and PCA

Factorial analysis was performed to gain information regarding distances among the peanut lines based on marker genotypes, and this analysis was not only interested in individual effect but to give an overall representation of diversity among the tested peanut lines (Figure 2). Based on the factorial analysis, the 79 peanut lines were clustered into three different groups according to the race and geographic origin. Most of the peanut lines from China were clustered on the top quadrant 'I', in contrast most of the peanut lines from the US were clustered on the middle quadrant 'II' and the majority of peanut lines from India were clustered on the bottom quadrant 'III'. Principal component analysis (PCA) was also performed to confirm the clusters in the genetic dendrogram (Figure 1). Using Dice's coefficient of similarity and the functions in NTSYSpc based on single data, the first three components were plotted into three-dimensional scatter plots (Figure 3). The first and second principal components accounted for 16.18 and 9.10% of the total molecular variance, respectively. The first three components together explained 31.83% of the total variation. The peanut lines were clustered into six distinct groups (Figure 3), which were similar with the dendrogram analysis and generally corresponding to the six groups (G1, G2a, G2b, G2c, G2d and G2e) with a few exceptions (Figures 1B, 3).

Population structure analysis

To assess genetic population structure of the 79 peanut lines, STRUCTURE analysis was performed using the multi-allelic data from 111 SSR markers, which are distributed on the peanut genetic maps/genome (Tables S1, S2). Calculation of ΔK for the 79 peanut lines found that $K = 2$. Clustering bar plots with $K = 2$ and $K = 5$ are shown in Figure 4. At $K = 2$, all 79 peanut lines were divided into two subpopulations, here denoted as P1 and P2 and each contained 21 and 58 peanut lines, respectively, which was generally consistent with the cluster analysis based on distance matrix G1 and G2 (Figure 1). Most of the lines that were grouped in P1 population were also grouped in G1. Similarly, most of the lines that were grouped in the P2 population were also grouped in G2. When $K = 5$, structure analysis revealed five subpopulations with relatively low levels of admixture between subpopulations named P1, P2a, P2b, P2c and P2e (Figure 4). Population P1 contained 16 peanut lines, of which 13 lines were also grouped in G1 and two lines from G2d and one from G2c (Figures 1, 4; Table 1). Second, P2a contained 31 lines with 27 lines from G2a and three lines from G2b and one from G2c. Third, P2b contained 20 lines with 13 lines from G2b and two lines from G2c and five lines from G2d. Fourth, P2c contained eight lines with four from G2b, three lines from G2a and one from G2d. Finally, P2e contained four lines that were all from G2e (Table 1). The results from STRUCTURE analysis was generally consistent with the results from neighbor-joining clustering analysis based on genetic distance with a few exceptions.

DISCUSSION

Peanut is one of the most important oil seed crops. China, India and the United States are the major peanut producers, and all have very important peanut breeding programs. The peanut annual planting area is around 24 million hectares and world peanut production totals about 40.4 million tons in 2014. China is the largest peanut producer with an annual production of 17 million tons, followed by India and the United States, with the average peanut production of 7.0 and 2.1 million tons, respectively (FAO 2014). Therefore, it is important to assess the genetic diversity for genetic improvement of peanut quality and yield through breeding efforts in representing breeding programs in these countries. Cultivated peanut has a limited genetic diversity within the peanut primary gene pool because of its tetraploid nature (AABB genome) (Pandey et al. 2012a). Any crop improvement program starts with the study of variability among genotypes (Ramu et al. 2013). Identification of genetically diverse germplasm with beneficial traits and then using it in peanut breeding is the key to broaden the genetic base of peanut cultivars (Kottapalli et al. 2011). Many studies on the genetic diversity of cultivated peanut have been reported in recent years (Kottapalli et al. 2011; Koilkonda et al. 2012; Suneetha et al. 2013; Jiang et al. 2014; Ren et al. 2014; Roomi et al. 2014; Vyas et al. 2014). Significant progress has been made in peanut genetics and genomics (Chen et al. 2010; Feng et al. 2012; Liu et al. 2013; Guo et al. 2013). There was substantial diversity among cultivated peanut in morphological, physiological, agronomic traits, and at molecular levels. This study provided the resource of 111 highly informative SSR markers, which was selected from a larger available genomic pool as a community effort to develop molecular markers for peanut genetic and genomic research (Guo et al. 2013, 2015), and these markers could be used in the primary assessment of the diversity of breeding lines in each program.

Molecular marker diversity

Simple sequence repeat markers are frequently and randomly distributed throughout the eukaryotic genome and they are more advantageous over many other markers, as they are often highly polymorphic, multi-allelic, genetically codominant, and amenable to high-throughput genotyping, making them become one of the most widely used molecular markers in recent years in the peanut community (Guo et al. 2015). However, many studies including construction of linkage maps, marker assisted breeding and germplasm diversity have been hindered by the initial process of screening a large number of SSR markers to find sufficient number of polymorphic markers (Ravi et al. 2011; Sarvamangala et al. 2011; Qin et al. 2012).

A large number of SSR markers with high PIC and number of alleles are very valuable in helping peanut breeding programs to select the most informative markers to use in their studies, thus saving time and cost in the genetic enhancement in peanut

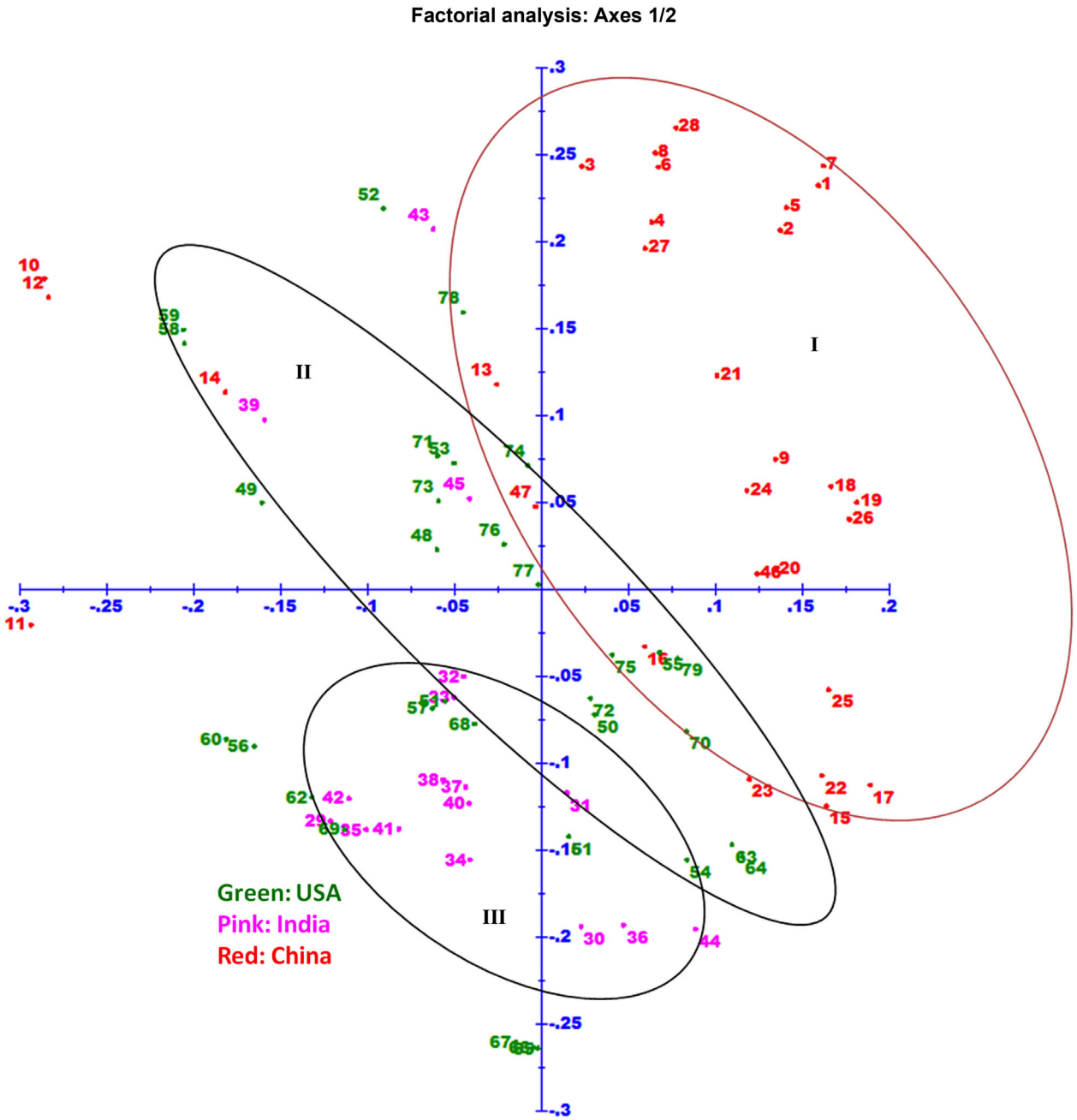

Figure 2. Factorial analysis of the peanut lines based on 111 simple sequence repeat (SSR) markers
All 79 peanut lines were clustered into three main groups. The colors indicate the country of the origin of the peanut lines, and the peanut lines were sorted according to the serial number from Table 1.

(Pandey et al. 2012b). The results in the present study showed that there are abundant polymorphic SSR molecular markers in the cultivated peanut. The peanut lines from three different countries could be distinguished from each other with the SSR markers. The 111 highly informative markers were chosen according to their location on three different peanut genetic maps (Shirasawa et al. 2013; Khera et al. 2014; Wang et al. 2014) (Table S2) in which Qin et al. (2012) contributed significantly to the initial screening of over 5,000 SSRs available at that time. All of the 111 SSR markers were polymorphic (100%) (Table S1). The PIC value of markers based on the frequencies of different alleles can provide an estimate of the discrimination power in peanut lines and also indicate the quality of the markers. In this study, the number of alleles was detected and the PIC value was calculated. The PIC value for each marker ranged from 0.025 to 0.778 with an average of 0.429. A recent study by Ren et al. (2014) had an average PIC value of 0.38, and the study of Pandey et al. (2012a) had an average PIC value of 0.31. In addition, there were some studies in peanut that observed a higher average PIC value, such as

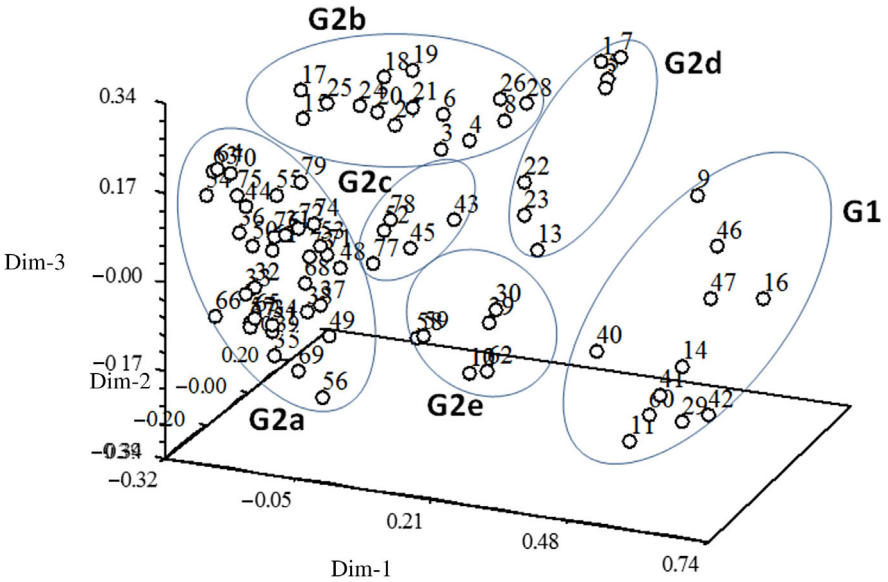

Figure 3. The principal components analysis (PCA) of 111 SSR marker data of the peanut lines from three countries
Dim-1, Dim-2 and Dim-3 refers to the first, second and third dimension of principal component, respectively. Dim-1 represents 16.18%, Dim-2 represents 9.10% and Dim-3 represents 6.55% of the variation, respectively. The peanut lines were sorted according to the serial number from Table 1.

0.64 (Geleta et al. 2006), 0.46 (Cuc et al. 2008), and 0.548 (Roomi et al. 2014).

The 111 SSRs detected a total of 472 alleles, the number of alleles per marker detected ranged from 2 to 9 with an average of 4.25 per locus and this result is in agreement with those of the previous genetic diversity studies. Ren et al. (2014) reported the results of 2 to 9 alleles with an average of 3.0 alleles per locus, Roomi et al. (2014) resulted in 2 to 4 alleles with an average of 2.6 alleles per locus, and Pandey et al. (2012a) observed 2 to 14 alleles per locus with an average of 3.2. Most of the studies reported 2–8 alleles per locus, compared to higher alleles of 2–13 alleles of Song et al. (2010), 2–20 alleles of Varshney et al. (2009) and 3–15 alleles of Naito et al. (2008). For specific peanut market type, a study reported that Valencia peanut germplasm collection uncovered an average of 13 alleles per locus (Kottapalli et al. 2011), and a study of US mini-core collection observed an average of 8.1 alleles per locus (Wang et al. 2011). There were several reasons that might result in the differences in allele numbers as reported by different groups; one may be the detection system as reported by Fountain et al. (2011) and another may be the geographic origins of peanut lines. In this study, more diverse peanut lines were used from three countries with an improved detection system (Fountain et al. 2011), which might contribute to the higher alleles detected.

In this study, most of the peanut lines were released cultivars, and some were breeding lines, so the allele number per locus would be similar to that of released cultivars (Jiang et al. 2014). There is a report suggesting that there is less genetic diversity in released cultivars than landraces (Ren et al. 2014). This study is in agreement with Ren et al. (2014) that the 111 SSR markers are effective and useful for analyzing the genetic diversity of peanut breeding lines, germplasm, and released cultivars. They have the potential to detect more alleles and more polymorphism in different sources of peanut germplasm and will be useful for the initial screening and assessment of the genetic background of the breeding materials.

Geographical diversity

When comparing genetic diversity among the three countries (China, India and USA), it was found that Chinese peanut lines had a higher number of alleles, also had a higher gene diversity and PIC value, indicating that Chinese peanut lines had a higher level of genetic variation than peanut lines from India and USA. China is the largest peanut producer in the world with over 20% of the planting area and more than 40% of production and it is also one of the centers of genetic diversity of cultivated peanut. A recent study by Ren et al. (2014) analyzed the genetic diversity of 196 Chinese peanut cultivars, suggesting that the genetic base of Chinese peanut cultivars were relatively narrow. According to their result of population structure, there were genetic differences in peanut cultivars from the south to the north of China (Ren et al. 2014). When comparing genetic diversity among the five different peanut programs in this study, it was found that HAAS-CN has a higher level of genetic variation than the other groups, among the lines tested with an average genetic distance of 0.5341 and 62.57% of the lines with a genetic distance greater than 0.50, indicating the relatively broad genetic base in the breeding program. HAAS is located in central China and provides breeding lines and cultivars for all market types. Therefore, a high level of genetic diversity is needed in this program. The peanut lines from SDPRI-CN represent the peanuts cultivated in northern China, mostly Virginia type. There were two market types (Virginia type and Spanish type) in the FAMU-IN program, and the 17 lines in this group were from all over India. Hence India peanut in this study also showed higher gene diversity and PIC value. Among the US lines tested in this study, 70% were the Runner type, and the 16 peanut lines from the ACI-US group

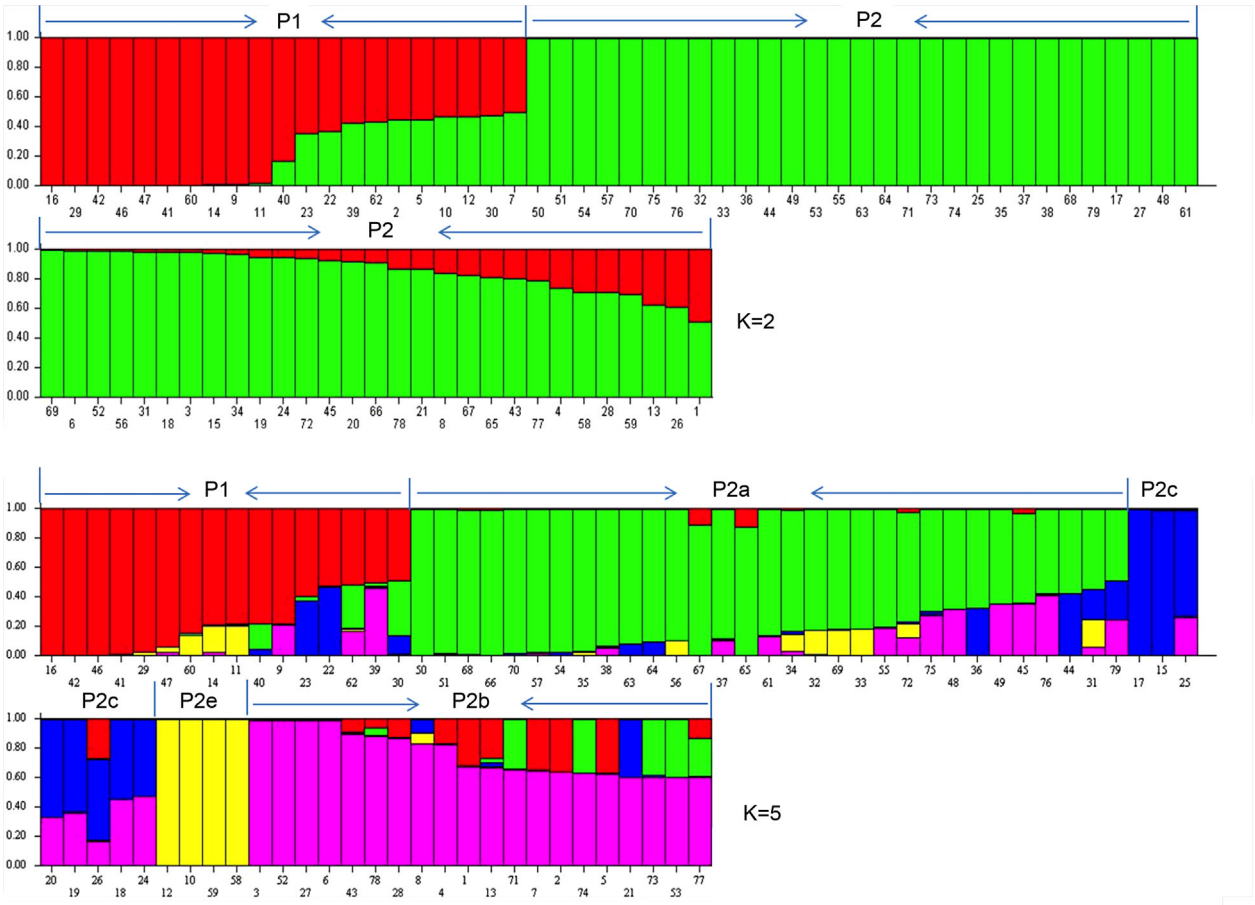

Figure 4. Population structure using STRUCTURE, a model-based clustering method to cluster these peanut lines based on the 111 SSR markers at K = 2 and K = 5
The peanut lines were sorted according to the serial number from Table 1.

were mostly advanced breeding lines. Although Runner-type peanut breeders have been successful at increasing levels of diversity among cultivars released in the last three decades of modern plant breeding, because the Runner market type was derived from fewer ancestry lines than the Virginia market type, they have a lower genetic diversity than the Virginia market type (Milla-Lewis et al. 2010a). This may be the reason why US peanut lines had lower genetic diversity than the other two countries in this study.

Genetic diversity and population structure

The SSR data were further used to compute the similarity matrix based on the alleles between peanut lines from different countries. The dendrogram was constructed based on the genetic data and classified the 79 peanut lines into two groups G1 and G2, and group G2 was further divided into five subgroups. The majority of peanut lines were clustered according to their pedigree and origin. For example, almost all of the peanut lines from the USA (including ACI-US and CEL-US) were clustered together, and 56% of the US lines were clustered in G2a and 32% lines were clustered in G2b. The peanut lines from China and India were also clustered closely within each program, respectively, but the distribution was more relatively dispersed.

It is commonly recognized that cultivated peanut has evolved into four market types, Spanish, Virginia, Valencia and Runner. Among the 79 peanut lines in this study, there were three market types, Spanish, Virginia and Runner. The results of clustering revealed some interesting groupings, and the classification has relevance with the market types of the peanut lines. All the peanut lines in groups G1 were Spanish types, while in the G2 group all the three different types were present. Among the five subgroups, G2a, G2b, G2c, G2d and G2e from G2, 80% Runner type lines were clustered in G2a, and Virginia type peanut lines were clustered in G2b, G2c, G2d and G2e. Except the Spanish lines in G1, the rest of Spanish type lines were clustered in G2a. Therefore, a clear process of transition was observed from Spanish type to Runner type, then to Virginia type (Figure 1B).

Factorial analysis and PCA were also performed to confirm the relationships seen in the genetic dendrogram. Factorial analysis and dendrogram methods are two very different approaches for the representation of diversity structure, considered complementary to each other. Dendrogram tree methods tend to represent individual relations faithfully and may be less accurate for the overall structure. But factorial analysis can give an overall representation of diversity (Perrier et al. 2006). In our study, although factorial analysis did not

separate the 79 peanut lines into distinct groups, it revealed three groups representing their pedigree and origin. Peanut lines from the same country were clustered closer with few admixtures. Using NTSYSpc, the PCA revealed six distinct groups, which were consistent with the dendrogram analysis for most of the lines.

STRUCTURE analysis has been used widely for different crops to identify the presence of subpopulations in core collections (Hasan et al. 2008; Zhao et al. 2009; Anderson et al. 2009; Wang et al. 2011; Jiang et al. 2014). In this study, we also identified the population structure using STRUCTURE, a model-based clustering method, which uses a Bayesian algorithm. STRUCTURE analysis was run using *K* from 1 to 10 and *K* was found to be 2. STRUCTURE analysis can estimate the number of subpopulations, the degree of admixture among subpopulations, and the genetic relatedness among peanut lines. On the other hand, dendrogram analysis can graphically display relationships between peanut lines (Wang et al. 2011). STRUCTURE analysis assigned the 79 peanut lines into five subpopulations (P1, P2a, P2b, P2c and P2e) when $K = 5$, and neighbor-joint tree analysis clustered these peanut into six subgroups (G1, G2a, G2b, G2c, G2d and G2e). In general, subpopulations identified by STRUCTURE corresponded to a genetic cluster on the dendrogram (P1/G1, P2a/G2a, P2b/G2b, and P2e/G2e). Although the number of peanut lines in the corresponding subgroups and subpopulation was not exactly the same, most of the peanut lines were classified into the corresponding subpopulation and subgroup with only a few exceptions.

In summary, the 111 SSR markers constructed were useful in determining the genetic relationships and diversity of 79 peanut breeding lines. The dendrogram, PCA clustering, and STRUCTURE analysis divided the 79 genotypes into distinct groups and generated similar results. Assessment of genetic diversity is important for characterization and conservation of germplasm, recognition of good germplasm resources and the collection of better cultivars for hybridization purposes. The results derived from analysis of genetic diversity could be used for designing effective breeding programs to broaden the genetic base of commercial cultivars. The molecular diversity observed in this study provides useful information for the representing breeding programs to aid selection of diverse parents for establishment of breeding populations that could be used for peanut improvement. A narrow genetic base has been reported in peanut cultivars (Stalker et al. 1994; Subramanian et al. 2000), and peanut breeders have been successful in increasing levels of diversity among cultivars released in the last three decades of modern plant breeding (Milla-Lewis et al. 2010). The results in this study showed the genetic variation in peanut is not undesirably lower, and a wider range of germplasm including wild diploid species and exotic materials should be incorporated to broaden the genetic resource and improve peanut varieties in the future breeding programs.

MATERIALS AND METHODS

Plant material

A total of 79 peanut breeding lines were provided by colleagues and collaborators (Table 1). Thirty were from China, 17 from India, and 32 from the USA and represent five peanut breeding programs including Shandong Peanut Research Institute (SDPRI) and Henan Academy of Agricultural Science (HAAS) from China, Florida Agricultural & Mechanical University (FAMU) in collaborating with Indian scientists, Commercial elite lines (CEL) from US released cultivars and AgResearch consultants Inc (ACI) from the US. All of the peanut lines were planted in the greenhouse.

DNA extraction

Several unopened young leaves of 2–3-week-old plants from each peanut line were sampled and put in liquid nitrogen immediately for DNA isolation. Total genomic DNA was extracted using the protocol of hexadecyl trimentyl ammonium bromide (CTAB) method as described by Grattapaglia and Sederoff (1994). The quality of the DNA was checked on 1% agarose gel and the DNA concentrations were estimated with the NanoDrop-1000 spectrophotometer (Nano Drop Technologies, Wilmington, DE, USA). DNA concentration was normalized to 20 ng/μL for use in polymerase chain reaction (PCR).

Genotyping with SSR markers

A set of 111 SSR markers were selected that were previously developed from the initial screening by Li et al. (2011) with 16 peanut lines with over 5,000 genic and genomic SSRs by Qin et al. (2012) and additional BAC-end sequence derived SSR markers by Wang et al. (2012). The location of the SSR markers on the peanut genetic map (Shirasawa et al. 2013; Khera et al. 2014; Wang et al. 2014) was taken into account when the markers were selected. These 111 SSR markers were distributed on to the 18 linkage groups (LGs) of Shirasawa's map (Shirasawa et al. 2013), Khera's map (Khera et al. 2014) and 20 LGs of Wang's map (Wang et al. 2014) (Table S1 and S2). PCR amplification for all markers was carried out using DNA Engine Tetrad 2 Peltier Thermal Cycler (BioRad Laboratories, Hercules, CA, USA) and Eppendorf Mastercycler Pro (Eppendorf North America, Hauppauge, NY, USA). PCR was performed in a reaction volume of 15 μL, the reaction mixtures contained 0.5 μM of each SSR forward and reverse primer, 25 ng genomic DNA, 10×PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs and 0.5 U *Taq* polymerase (Thermo Scientific). The PCR temperature profile was 95 °C for 4 min for initial denaturation, followed by 35 cycles of 94 °C for 45 s, 50–55 °C for 45 s and 72 °C for 1 min, and 72 °C for 7 min for a final extension step. The PCR products were separated in 6% non-denaturing polyacrylamide gel electrophoresis (PAGE) at 140 voltages for 2 h in 0.5×TBE using DYCZ-30B gel rigs (Beijing, China) (Fountain et al. 2011). The amplicons were visualized by silver staining (Qin et al. 2012).

Analysis of data

For each marker, the polymorphic bands at the expected base pair size were treated as a unique characteristic and scored as '1' or '0' based on whether the band is present or absent for the same amplified fragments, respectively. Data needed to be changed to different formats according to the requirements of the analytic software. Major allele frequency (MAF), number of alleles per locus, gene diversity and polymorphism information content (PIC) were analyzed using allelic data with PowerMarker version 3.25 software (Liu and Muse 2005).

Using DARwin 6.0.5 software (Perrier and Jacquemoud-Collet 2006), the genetic distance (GD) matrix was generated for all peanut breeding lines. A dendrogram was created based on genetic distance matrix by using hierarchical clustering. Factorial analysis was also performed by using distance matrix based on DARwin software. The principal component analysis (PCA) was used to identify eigenvectors, amounts of variance, and cumulatively explained variances per component. The PCA was performed using the Dice coefficient of similarity and the functions DCENTER, EIGEN and MXPLOT in NTSYSpc version 2.20 (Rohlf 2008).

The genetic structure analysis was performed using the model-based clustering method implemented in STRUCTURE software version 2.3.1 (Pritchard et al. 2000) using a burn-in period of 10,000 and 100,000 Markov Chain Monte Carlo (MCMC) replications for all samples. The model with admixture was chosen and using the correlated allele frequency option. To determine the best *K* (the number of subpopulations), the L (*K*) procedure was used by performing a batch job form $K = 1–10$, with 10 iterations for all peanut lines. The program Structure Harvester was used to calculate ΔK (Earl and vonHoldt 2012).

ACKNOWLEDGEMENTS

We thank Liming Yang, Jake Fountain and Billy Wilson for technical assistance in the greenhouse and the laboratory. This work is partially supported by the US Department of Agriculture Agricultural Research Service (USDA-ARS), the Georgia Agricultural Commodity Commission for Peanuts, Peanut Foundation and National Peanut Board. This work has been undertaken as part of the CGIAR Research Program on Grain Legumes and USAID University Linkage Grant. ICRISAT is a member of CGIAR Consortium. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer.

AUTHOR CONTRIBUTIONS

H.W. conducted genotyping, data analysis and manuscript preparation. P.K. proformed data formating and analysis. B.H. involved planting and genotyping. M.Y., R.K., and K.M. provided planting materials. W.Z., A.C., X.Z. R.V. participated in planning and discussion. K.H. participated in data analysis. L.X. and B.G. conceived the project, mentored the student, and finalized the manuscript.

REFERENCES

Anderson WF, Maas A, Ozias-Akins P (2009) Genetic variability of a forage bermuda grass core collection. **Crop Sci** 49: 1347–1358

Babu BK, Meena V, Agarwal V, Agrawal PK (2014) Population structure and genetic diversity analysis of Indian and exotic rice (*Oryza sativa* L.) accessions using SSR markers. **Mol Biol Rep** 41: 4329–39

Chen X, Wang ML, Holbrook CC, Culbreath A, Liang X, Brenneman T, Guo B (2010) Identification and characterization of a multigene family encoding germin-like proteins in cultivated peanut (*Arachis hypogaea* L.). **Plant Mol Biol Rep** 29: 389–403

Cuc LM, Mace ES, Crouch JH, Quang VD, Long TD, Varshney RK (2008) Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea* L.). **BMC Plant Biol** 8: 55

Earl DA, vonHoldt BM (2012) Structure harvester: A website and program for visualizing Structure output and implementing the Evanno method. **Conserv Genet Resour** 4: 359–361

FAO Statistical Databases. http://faostat.fao.org/browse/

Feng S, Wang X, Zhang X, Dang PM, Holbrook CC, Culbreath AK, Wu Y, Guo BZ (2012) Peanut (*Arachis hypogaea*) expressed sequence tag project: Progress and application. **Comp Funct Genomics** 2012: 373768

Fountain J, Qin H, Chen C, Dang P, Wang ML, Guo BZ (2011) A note on development of a low-cost and high-throughput SSR-based genotyping method in peanut (*Arachis hypogaea* L.). **Peanut Sci** 38: 122–127

Frascaroli E, Schrag TA, Melchinger AE (2013) Genetic diversity analysis of elite European maize (*Zea mays* L.) inbred lines using AFLP, SSR, and SNP markers reveals ascertainment bias for a subset of SNPs. **Theor Appl Genet** 126: 133–141

Geleta N, Labuschagne MT, Viljoen CD (2006) Genetic diversity analysis in sorghum germplasm as estimated by AFLP, SSR and morpho-agronomical markers. **Biodivers Conserv** 15: 3251–3265

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

Guo BZ, Chen CY, Chu Y, Holbrook CC, Ozias-Akins P, Stalker HT (2012) Advances in genetics and genomics for sustainable peanut production. In: Benkeblia N, ed. *Sustainable Agriculture and New Biotechnologies*. CRC Press, Boca Raton, FL. pp. 341–367

Guo BZ, Khera P, Wang H, Peng Z, Sudini H, Wang X, Osiru M, Chen J, Vadez V, Yuan M, Wang C, Zhang X, Waliyar F, Wang J, Varshney RK (2015) Annotation of trait loci on integrated genetic maps of *Arachis* species. In: Stalker T, Wilson R, eds. *Peanuts: Genetics, Processing, and Utilization*. The American Oil Chemists' Society (AOCS). (in press)

Guo BZ, Pandey MK, He G, Zhang XY, Liao B, Culbreath AK, Varshney RK, Nwosu V, Wilson RF, Stalker HT (2013) Recent advances in molecular genetic linkage maps of cultivated peanut (*Arachis hypogaea* L.). **Peanut Sci** 40: 95–106

Hasan M, Friedt W, Pons-Kuhnemann J, Freitag NM, Link K, Snowdon RJ (2008) Association of gene-linked SSR markers to seed glucosinolate content in oilseed rape (*Brassica napus* ssp. *napus*). **Theor Appl Genet** 116: 1035–1049

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: 1–3

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

Herselman L (2003) Genetic variation among southern African cultivated peanut (*Arachis hypogaea* L.) genotypes as revealed by AFLP analysis. **Euphytica** 133: 319–327

Isleib TG, Holbrook CC, Gorbet DW (2001) Use of *Arachis* spp. plant introductions in peanut cultivar development. **Peanut Sci** 28: 96–113

Jiang H, Huang L, Ren X, Chen Y, Zhou X, Xia Y, Huang J, Lei Y, Yan L, Wan L, Liao B (2014) Diversity characterization and association analysis of agronomic traits in a Chinese peanut

(*Arachis hypogaea* L.) mini-core collection. **J Integr Plant Biol** 56: 159–69

Khera P, Upadhyaya HD, Pandey MK, Roorkiwal M, Sriswathi M, Janila P, Guo Y, McKain M, Nagy ED, Knapp SJ, Leebens-Mack J, Conner JA, Ozias-Akins P, Varshney RK (2013) SNP-based genetic diversity in the reference set of peanut (*Arachis* spp.) by developing and applying cost-effective KASPar genotyping assays. **Plant Gen** 6: 1–11

Khera P, Wang H, Culbreath AK, Pandey MK, Varshney RK, Wang X, Liao B, Zhang X, Wang J, Holbrook CC, Guo B (2014) QTL mapping and quantitative disease resistance to TSWV and leaf spots in a recombinant inbred line population SunOleic 97R and NC94022 of peanut (*Arachis hypogaea* L.). *Advances in Arachis through Genomics & Biotechnology*. November 11–13, 2014, Savannah, USA

Kochert G, Halward T, Branch WD, Simpson CE (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. **Theor Appl Genet** 81: 565–570

Koilkonda P, Sato S, Tabata S, Shirasawa K, Hirakawa H, Sakai H, Sasamoto S, Watanabe A, Wada T, Kishida Y, Tsuruoka H, Fujishiro T, Yamada M, Kohara M, Suzuki S, Hasegawa M, Kiyoshima H, Isobe S (2012) Large-scale development of expressed sequence tag-derived simple sequence repeat markers and diversity analysis in *Arachis* spp. **Mol Breed** 30: 125–138

Kottapalli P, Upadhyaya HD, Kottapalli KR, Payton P, Dwivedi SL, Burow M, David KO, Sanogo S, Puppala N (2011) Population structure and diversity in Valencia peanut germplasm collection. **Crop Sci** 51: 1089–1100

Krapovickas A, Gregory WC (1994) Taxonomy of the genus *Arachis* (Leguminosae). **Bonplandia** 8: 1–186

Leal-Bertioli SC, Jose AC, Alves-Freitas DM, Moretzsohn MC, Guimaraes PM, Nielen S, Vidigal BS, Pereira RW, Pike J, Favero AP, Parniske M, Varshney RK, Bertioli DJ (2009) Identification of candidate genome regions controlling disease resistance in *Arachis*. **BMC Plant Biol** 9: 112

Li Y, Chen CY, Knapp SJ, Culbreath AK, Holbrook CC, Guo BZ (2011) Characterization of simple sequence repeat (SSR) markers and genetic relationships within cultivated peanut (*Arachis hypogaea* L.). **Peanut Sci** 38: 1–10

Liu K, Muse S (2005) PowerMarker: An integrated analysis environment for genetic marker analysis. **Bioinformatics Appl** 21: 2128–2129

Liu Z, Feng S, Pandey MK, Chen X, Culbreath AK, Varshney RK, Guo BZ (2013) Identification of expressed resistance gene analogs (RGAs) from peanut (*Arachis hypogaea* L.) expressed sequence tags (ESTs). **J Integr Plant Biol** 55: 453–461

Milla-Lewis SR, Zuleta CM, Isleib TG (2010a) Assessment of genetic diversity among U.S. Runner-type peanut cultivars using simple sequence repeat markers. **Crop Sci** 50: 2396–2405

Milla-Lewis SR, Zuleta MC, Isleib TG (2010b) Simple sequence repeat allelic diversity in Virginia-type peanut cultivars released from 1943 to 2006. **Crop Sci** 50: 1348–1356

Naito Y, Suzuki S, Iwata Y, Kuboyama T (2008) Genetic diversity and relationship analysis of peanut germplasm using SSR markers. **Breed Sci** 58: 293–300

Nature Editorial (2010) How to feed a hungry world. **Nature** 466: 531–532

Pandey MK, Gautami B, Jayakumar T, Sriswathi M, Upadhyaya HD, Chennabyre gowda MV, Radhakrishnan T, Bertioli DJ, Knapp SJ, Cook DR, Varshney RK (2012a) Highly informative genic and genomic SSR markers to facilitate molecular breeding in cultivated groundnut (*Arachis hypogaea*). **Plant Breed** 131: 139–147

Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimarães P, Nigam SN, Upadhyaya HD, Janila P, Zhang X, Guo B, Cook DR, Bertioli DJ, Michelmore R, Varshney RK (2012b) Advances in *Arachis* genomics for peanut improvement. **Biotechnol Adv** 30: 639–651

Pandey MK, Wang ML, Qiao L, Feng S, Khera P, Wang H, Tonnis B, Barkley NA, Wang J, Holbrook CC, Culbreath AK, Varshney RK, Guo BZ (2014) Identification of QTLs associated with oil content and mapping *FAD2* genes and their relative contribution to oil quality in peanut (*Arachis hypogaea* L.). **BMC Genetics** 15: 133

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. http://darwin.cirad.fr/

Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. **Genetics** 155: 945–959

Qin H, Feng S, Chen C, Guo Y, Knapp S, Culbreath A, He G, Wang ML, Zhang X, Holbrook CC, Ozias-Akins P, Guo BZ (2012) An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. **Theor Appl Genet** 124: 653–664

Ramu P, Billot C, Rami JF, Senthilvel S, Upadhyaya HD, Ananda Reddy L, Hash CT (2013) Assessment of genetic diversity in the sorghum reference set using EST-SSR markers. **Theor Appl Genet** 126: 2051–64

Raina SN, Rani V, Kojima T, Ogihara Y, Singh KP, Devarumath RM (2001) RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species. **Genome** 44: 763–772

Ravi K, Vadez V, Isobe S, Mir RR, Guo Y, Nigam SN, Gowda MV, Radhakrishnan T, Bertioli DJ, Knapp SJ, Varshney RK (2011) Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea* L.). **Theor Appl Genet** 122: 1119–1132

Ren X, Jiang H, Yan Z, Chen Y, Zhou X, Huang L, Lei Y, Huang J, Yan L, Qi Y, Wei W, Liao B (2014) Genetic diversity and population structure of the major peanut (*Arachis hypogaea* L.) cultivars grown in China by SSR markers. **PLoS ONE** 9: e88091

Rohlf FJ (2008) NTSYSpc: Numerical Taxonomy System, Version 2.20. Seatauker, NY: Exeter Publishing, Ltd

Roomi S, Sabiha B, Iqbal A, Suleman M, Muhammad I, Zia MA, Ahmad MZ, Rashid F, Ghafoor A, Tabbasam N (2014) SSR based genetic diversity analysis in a diverse germplasm of groundnut (*Arachis hypogaea* L.) from Pakistan. **Aust J Crop Sci** 8: 55–61

Sarvamangala, C, Gowda MV, Varshney RK (2011) Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (*Arachis hypogaea* L.). **Field Crops Res** 122: 49–59

Seijo JG, Lavia GI, Fernandez A, Krapovickas A, Ducasse DA, Moscone EA (2004) Physical mapping of the 5S and 18S-25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaënsis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae). **Am J Bot** 91: 1294–1303

Shakeel M, Ilyas M, Kazi M (2013) Evaluation of synthetic hexaploid wheats (derivative of durum wheats and *Aegilops tauschii* accessions) for studying genetic diversity using randomly amplified polymorphic DNA (RAPD) markers. **Mol Biol Rep** 40: 21–26

Shirasawa K, Bertioli DJ, Varshney RK, Moretzsohn MC, Leal-Bertioli SC, Thudi M, Pandey MK, Rami JF, Foncéka D, Gowda MV, Qin H, Guo B, Hong Y, Liang X, Hirakawa H, Tabata S, Isobe S (2013) Integrated consensus map of cultivated peanut and wild relatives

reveals structures of the A and B genomes of *Arachis* and divergence of the legume genomes. **DNA Res** 20: 173–84

Song GQ, Li MJ, Xiao H, Wang XJ, Tang RH, Xia H, Zhao CZ, Bi YP (2010) EST sequencing and SSR marker development for cultivated peanut (*Arachis hypogaea* L.). **Electronic J Biotech** 13: 1–9

Stalker HT, Phillips TD, Murphy JP, Jones TM (1994) Variation of isozyme patterns among *Arachis* species. **Theor Appl Genet** 87: 746–755

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

Suneetha N, Vasanthi RP, Sudhakar P, Raja Reddy K (2013) Genetic diversity analysis among released and pre-release cultures in groundnut. **Legume Res** 36: 208–213

Tang R, Gao G, He L, Han Z, Shan S, Zhong R, Zhou C, Jiang J, Li Y, Zhuang W (2007) Genetic diversity in cultivated groundnut based on SSR markers. **J Genet Genomics** 34: 449–459

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: Features and applications. **Trends Biotechnol** 23: 48–55

Varshney RK, Mahendar T, Aruna R, Nigam SN, Vadez V, Hoisington DA (2009) High level of natural variation in a groundnut (*Arachis hypogaea* L.) germplasm collection assayed by selected informative SSR markers. **Plant Breed** 128: 486–494

Varshney RK, Mohan SM, Gaur PM, Gangarao NV, Pandey MK, Bohra A, Sawargaonkar SL, Chitikineni A, Kimurto PK, Janila P, Saxena KB, Fikre A, Sharma M, Rathore A, Pratap A, Tripathi S, Datta S, Chaturvedi SK, Mallikarjuna N, Anuradha G, Babbar A, Choudhary AK, Mhase MB, Bharadwaj CH, Mannur DM, Harer PN, Guo B, Liang X, Nadarajan N, Gowda CL (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. **Biotechnol Adv** 31: 1120–1134

Vyas D, Munot J, Maloo SR, Dashora A, Rajpurohit D (2014) RAPD based evaluation of genetic diversity in groundnut (*Arachis hypogaea* L.) genotypes. **Legume Res** 37: 26–31

Wang H, Khera P, Culbreath AK, Pandey MK, Varshney RK, Wang X, Liao B, Zhang X, Wang J, Holbrook CC, Guo B (2014) Genetic mapping and QTL analysis of disease resistance traits in peanut population Tifrunner x GT-C20. *Advances in Arachis through Genomics & Biotechnology*. November 11–13, 2014, Savannah, USA

Wang H, Penmetsa RV, Yuan M, Gong L, Zhao Y, Guo B, Farmer AD, Rosen BD, Gao J, Isobe S, Bertioli DJ, Varshney RK, Cook DR, He G (2012) Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). **BMC Plant Biol** 12: 10

Wang ML, Sukumaran S, Barkley NA, Chen Z, Chen CY, Guo B, Pittman RN, Stalker HT, Holbrook CC, Pederson GA, Yu J (2011) Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. **Theor Appl Genet** 123: 1307–1317

Xiong F, Liu J, Jiang J, Zhong R, He L, Han Z, Li Z, Tang X, Tang R (2013) Molecular profiling of genetic variability in domesticated groundnut (*Arachis hypogaea* L.) based on ISJ, URP, and DAMD markers. **Biochem Genet** 51: 889–900

Zhao W, Chung JW, Ma KH, Kim TS, Kim SM, Shin DI, Kim CH, Koo HM, Park YJ (2009) Analysis of genetic diversity and population structure of rice cultivars from Korea, China and Japan using SSR markers. **Genes Genom** 31: 283–292

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Table S1. Summary information of 111 simple sequence repeat (SSR) markers used in this study

Table S2. List of 111 simple sequence repeat (SSR) markers, including their position in three genetic maps, number of alleles, major allele frequency, and summary of genetic diversity and polymorphic information content analysis