# Single Nucleotide Polymorphism–based Genetic Diversity in the Reference Set of Peanut (*Arachis* spp.) by Developing and Applying Cost-Effective Kompetitive Allele Specific Polymerase Chain Reaction Genotyping Assays

Pawan Khera, Hari D. Upadhyaya, Manish K. Pandey, Manish Roorkiwal, Manda Sriswathi, Pasupuleti Janila, Yufang Guo, Michael R. McKain, Ervin D. Nagy, Steven J. Knapp, James Leebens-Mack, Joann A. Conner, Peggy Ozias-Akins, and Rajeev K. Varshney*

## Abstract

Kompetitive allele-specific polymerase chain reaction (KASP) assays have emerged as cost-effective marker assays especially for molecular breeding applications. Therefore, a set of 96 informative single nucleotide polymorphisms (SNPs) was used to develop KASP assays in groundnut or peanut (*Arachis* spp.). Developed assays were designated as groundnut KASP assay markers (GKAMs) and screened on 94 genotypes (validation set) that included parental lines of 27 mapping populations, seven synthetic autotetraploid and amphidiploid lines, and 19 wild species accessions. As a result, 90 GKAMs could be validated and 73 GKAMs showed polymorphism in the validation set. Validated GKAMs were screened on 280 diverse genotypes of the reference set for estimating diversity features and elucidating genetic relationships. Cluster analysis of marker allelic data grouped accessions according to their genome type, subspecies, and botanical variety. The subspecies *Arachis hypogaea* L. subsp. *fastigiata* Waldron and *A. hypogaea* subsp. *hypogaea* formed distinct cluster; however, some overlaps were found indicating their frequent intercrossing during the course of evolution. The wild species, having diploid genomes, were grouped into a single cluster. The average polymorphism information content value for polymorphic GKAMs was 0.32 in the validation set and 0.31 in the reference set. These validated and highly informative GKAMs may be useful for genetics and breeding applications in *Arachis* species.

Published in The Plant Genome 6
doi: 10.3835/plantgenome2013.06.0019
© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

PEANUT (*Arachis hypogaea* L.), an allotetraploid species, is a highly domesticated oilseed crop that originated through hybridization and chromosome doubling of two different genomes. Furthermore, domestication created a genetic bottleneck as no close relatives are available to share favorable alleles, resulting in an extremely narrow cultivated gene pool in the case of peanut. As a consequence, breeders have been obligated to harness available favorable alleles limited to elite breeding lines and some local landraces for developing improved cultivars. The above circumstances have greatly hampered peanut genetic improvement through conventional approaches as well as development of optimum genetic and genomic resources for molecular breeding in peanut (see Varshney et al., 2012). Nevertheless, rapid developments in availability of large scale genomic resources have now provided a well-set platform to conduct several genetic studies such as molecular diversity, genetic mapping, marker–trait association, and use of modern breeding strategies in peanut (see Pandey et al., 2012; see Varshney et al., 2013). Availability of

P. Khera, H.D. Upadhyaya, M.K. Pandey, M. Roorkiwal, M. Sriswathi, P. Janila, and R.K. Varshney, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India; Y. Guo and E.D. Nagy, Center for Applied Genetic Technologies, The Univ. of Georgia, Athens, GA 30602; Y. Guo, J.A. Conner, and P. Ozias-Akins, Dep. of Horticulture, The Univ. of Georgia, Tifton, GA 31973; M.R. McKain and J. Leebens-Mack, Dep. of Plant Biology, The Univ. of Georgia, Athens, GA 30602; S.J. Knapp, The Institute of Plant Breeding, Genetics & Genomics, The Univ. of Georgia, Athens, GA 30602; R.K. Varshney, CGIAR Generation Challenge Programme, c/o CIMMYT, Mexico DF, Mexico. Received 5 June 2013. *Corresponding author (r.k.varshney@cgiar.org).

Abbreviations: Cl, cluster; DArT, Diversity Array Technology; EST, expressed sequence tag; GKAM, groundnut KASP assay marker; KASP, Kompetitive allele specific polymerase chain reaction; PCR, polymerase chain reaction; PIC, polymorphism information content; QTL, quantitative trait loci; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; UPGMA, unweighted pair-group method with arithmetic mean.

genetic and genomic resources allow their integration with conventional breeding approaches for efficient use of existing variability in the primary gene pool and harnessing of useful alleles from synthetics.

International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) has one of the largest peanut germplasm collections with 15,445 accessions from 92 countries (Upadhyaya et al., 2010). Based on agronomic, morphological, and physiological traits and a few molecular markers, collections of manageable sizes such as a mini core and a reference set were constituted in peanut. The reference set consists of a miniscule representation (300 genotypes) of the overall germplasm diversity available in peanut representing 48 countries (Upadhyaya et al., 2002, 2003, 2010). It comprises the cultivated species *A. hypogaea*, which is further classified into subspecies *fastigiata* and *hypogaea*. The subspecies *fastigiata* was additionally classified into four botanical varieties *fastigiata* (Waldron) Krapov. & W. C. Greg., *peruviana* Krapov. & W. C. Greg., *vulgaris* Harz, and *aequatoriana* Krapov. & W. C. Greg. Similarly, the subspecies *hypogaea* was further classified into botanical varieties *hirsuta* J. Kohler and *hypogaea* (Holbrook and Stalker, 2003). Details about accessions present in the reference set are given in Supplemental Table S1.

A range of molecular markers such as restriction fragment length polymorphism, random amplified polymorphic DNA, and amplified fragment length polymorphism were used for germplasm characterization in the past but simple sequence repeat (SSR) markers have prevailed (Hilu and Stalker, 1995; Kochert et al., 1996; Subramanian et al., 2000; Dwivedi et al., 2001; He and Prakash, 2001; Herselman, 2003; Bravo et al., 2006). The SSR markers, which are known for their high polymorphism information content (PIC) and resourcefulness as molecular tools (Gupta and Varshney, 2000), have enabled discrimination and assessment of genetic variation in peanut germplasm revealing very low to moderate levels of polymorphism in cultivated germplasm as compared to wild relatives (Gautami et al., 2009; Varshney et al., 2009a; Koppolu et al., 2010). Realizing the importance of SSRs considering the advantages described above, the peanut community has developed >13,000 SSR markers through different approaches (see Pandey et al., 2012; see Varshney et al., 2013). Several genetic maps with moderate marker density were developed and some of these were used for conducting quantitative trait loci (QTL) analysis using the phenotyping data. The above developments facilitated initiation of molecular breeding in peanut for disease and oil quality traits using trait-linked markers (see Varshney et al., 2013). However, all the above marker systems failed in providing high polymorphism for constructing dense genetic maps and conducting QTL analysis in addition to being more costly and time consuming.

To generate fast genotyping data to achieve dense genetic maps and use in other genetic and breeding applications, the Diversity Array Technology (DArT) marker platform consisting of approximately 15,000 features has also been developed for peanut (see Varshney et al., 2012). However, on assessing polymorphism in a set of germplasm containing tetraploid ($2n = 2x = 40$ with AABB genome) and diploid ($2n = 2x = 20$ with AA and BB genome) genotypes, DArT markers showed very low polymorphism in $4x$ while moderate in $2x$ genotypes. Hence, DArT markers may not be a very useful and cost-effective marker system for practical applications in peanut breeding. Nevertheless, these markers could be used in tracking alien genome introgression from wild into cultivated peanut genotypes (Mallikarjuna et al., 2011).

In recent years, due to the advent of next-generation sequencing and faster genotyping technologies, new marker systems such as single nucleotide polymorphisms (SNPs) have attracted substantial attention and have emerged as the marker of choice in crop breeding (Varshney et al., 2009b). In the case of diploid peanut, a limited number of SNPs (5) were included in an AA genome interspecific map (Alves et al., 2008) while a dense intraspecific AA genome map was enabled through transcriptome sequencing, SNP discovery, and Illumina GoldenGate (Illumina, Inc.) genotyping (Nagy et al., 2012). Recently >2000 SNPs have been identified from tetraploid peanut sequencing (Chen et al., 2013). Two Illumina GoldenGate assays for genotyping 1536 SNPs and 768 SNPs have been developed at the University of Georgia and University of California-Davis (see Pandey et al., 2012). Such platforms are cost effective only when a large number of SNPs (e.g., 768, 1536 as mentioned above) are genotyped with a large number of samples (e.g., >500) (Mir et al., 2013). In cases where a few SNPs are required for genotyping a varying number of samples, KASP assay seems to be the most cost effective (Hiremath et al., 2012; Saxena et al., 2012). It is a fluorescence-based SNP genotyping system developed by LGC Genomics, Ltd. (Semagn et al., 2013). Due to its flexible nature and robustness, KASP assays have been implemented in maize (*Zea mays* L.) (Mammadov et al., 2011), and wheat (*Triticum aestivum* L.) (Neelam et al., 2013), among others. By using KASP assays, second generation maps have been developed recently in chickpea (*Cicer arietinum* L.) (Hiremath et al., 2012) and pigeonpea (*Cajanus cajan* L.) (Saxena et al., 2012).

Therefore, the present study was undertaken with the following objectives: (i) develop cost-effective and flexible KASP assays for informative SNPs in peanut, (ii) validate the newly developed KASP assays on a validation set, and (iii) assess genetic diversity in the peanut reference set.

## Material and Methods

### Plant Material and DNA Isolation

Two different germplasm sets consisting of diverse peanut genotypes were used in the present study. First, a panel of 94 peanut genotypes (Supplemental Table S2), referred to as the validation set, which included elite genotypes, synthetics (autotetraploids and amphidiploids), wild accessions, and parents of 27 mapping populations (Table 1), was used for validation of KASP assays.

Second, the reference set of peanut was used for genotyping with the validated KASP assays to demonstrate its application in deciphering genetic diversity. Although the reference set consists of 300 genotypes representing 48 countries (Upadhyaya et al., 2003, 2010), only 280 genotypes were used in the present study (Fig. 1).

Total genomic DNA was isolated from young leaves (2-wk-old seedlings) using the standard DNA isolation protocol following Cuc et al. (2008). The DNA quantity for all samples was estimated on 0.8% agarose gels and the DNA was then diluted to 5 ng μL$^{-1}$ for genotyping.

## Identification and Selection of Single Nucleotide Polymorphisms

Massively parallel 454 GS FLX or FLX Titanium (Roche Diagnostics Corporation) sequencing of normalized root, leaf, and seed (pericarp) RNA from 19 tetraploid genotypes, including two tetraploid species (*A. hypogaea* and *A. monticola* Krapov. & Rigoni) and four subspecies and varieties (subsp. *hypogaea* var. *hypogaea*, subsp. *hypogaea* var. *hirsuta*, subsp. *fastigiata* var. *fastigiata*, and subsp. *fastigiata* var. *vulgaris*), resulted in 1.2 Gb sequence deposited in National Center for Biotechnology Information's Sequence Read Archive (NCBI-SRA) (Table 2). The sequence assembly (Nagy et al., 2012) combined 454 GS FLX sequence and previously deposited long-read expressed sequence tags (ESTs) (Guo et al., 2012) and was used as a reference for SNP discovery through an NGMagic MySQL database (http://ngmagic.sourceforge.net/; accessed 6 May 2013) and customized bioinformatics pipeline. Single nucleotide polymorphisms were called based on a minimum of two reads, a minor allele frequency of at least 0.01, and no additional SNPs or indels within 20 bp. Of the 8486 putative SNPs identified, 6702 returned an Illumina GoldenGate design score of 0.8 or greater and 1536 were selected for a GoldenGate array after predicting intron and exon boundaries using alignments with soybean [*Glycine max* (L.) Merr.], *Medicago truncatula* Gaertn., and *Arabidopsis thaliana* (L.) Heynh. genome sequences. The GoldenGate assay was used for genotyping of 80 tetraploid inbred lines, three amphidiploids, and several diploid accessions of *Arachis*. Based on genotyping data and polymorphism in 80 tetraploid lines, a set of 96 informative SNPs was selected.

## Kompetitive Allele Specific Polymerase Chain Reaction Genotyping Assays

The KASP assay is a unique form of Kompetitive allele-specific PCR (KASP)-based homogeneous fluorescent SNP genotyping system to determine the alleles at a specific locus within genomic DNA for SNP typing. To use the SNPs shown to be polymorphic for tetraploids in the GoldenGate assay in a cost effective manner, KASP assay for these SNPs were developed. Flanking sequences (50 bp each upstream and downstream) around the SNP position were used for designing primers to develop the KASP assay (Supplemental Table S3).

**Table 1. Groundnut Kompetitive allele specific polymerase chain reaction assay marker–based polymorphisms in some segregating populations of peanut.**

| Parental genotypes of mapping populations | Segregating trait or traits[†] | Polymorphic markers | Polymorphism rate (%) |
|---|---|---|---|
| Interspecific mapping populations | | | |
| TMV 2 × TxAG 6 | Agronomic traits | 40 | 44.4 |
| ICGV 87846 × ISATGR 265-5 | Agronomic traits | 36 | 40.0 |
| ICG 00350 × ISATGR 184 | Agronomic traits | 37 | 41.1 |
| ICG 00350 × ISATGR 9B | Agronomic traits | 36 | 40.0 |
| ICG 00350 × ISATGR 5B | Agronomic traits | 44 | 48.9 |
| ICG 00350 × ISATGR 90B | Agronomic traits | 36 | 40.0 |
| Intraspecific mapping populations | | | |
| TG 26 × GPBD 4 | Rust and LLS resistance | 19 | 21.1 |
| TAG 24 × GPBD 4[‡] | Rust and LLS resistance | 18 | 20.0 |
| ICG 11337 × JL 24 | LLS resistance | 22 | 24.4 |
| ICGV 93437 × ICGV 95714 | ELS resistance | 20 | 22.2 |
| Robut 33-1 × ICGV 95714 | ELS resistance | 23 | 25.6 |
| ICGV 93437 × ICGV 91114 | Rust resistance | 9 | 10.0 |
| ICGV 93437 × ICGVSM 95342 | Rust resistance | 23 | 25.6 |
| ICGS 76 × CSMG 84-1 | Drought tolerance | 9 | 10.0 |
| ICGS 44 × ICGS 76 | Drought tolerance | 5 | 5.6 |
| TAG 24 × ICGV 86031 | Drought tolerance | 0 | 0.0 |
| Chalimbana × ICGVSM 90704 | Resistance to GRD | 2 | 2.2 |
| CG 7 × ICGVSM 90704 | Resistance to GRD | 6 | 6.7 |
| ICGV 07368 × ICGV 06420 | High and low oil content | 12 | 13.3 |
| ICGV 07166 × ICGV 06188 | High and low oil content | 10 | 11.1 |
| ICGV 06420 × SunOleic 95R[‡] | O:L ratio | 13 | 14.4 |
| Intraspecific marker-assisted backcrossing (MABC)[§] populations | | | |
| ICGV 91114 × GPBD 4 | Rust resistance | 15 | 16.7 |
| JL 24 × GPBD 4 | Rust resistance | 17 | 18.9 |
| ICGV 03042 × SunOleic 95R | O:L ratio | 12 | 13.3 |
| ICGV 02411 × SunOleic 95R | O:L ratio | 15 | 16.7 |
| ICGV 05141 × SunOleic 95R | O:L ratio | 12 | 13.3 |
| ICGV 05100 × SunOleic 95R | O:L ratio | 10 | 11.1 |

[†] LLS, late leaf spot; ELS, early leaf spot; GRD, groundnut rosette disease; O:L, oleic:linoleic acid.

[‡] Used for marker-assisted backcrossing also.

[§] MABC, marker-assisted backcrossing.

Genotyping with KASP assays was performed following essentially the same protocol given in our earlier studies (Hiremath et al., 2012; Saxena et al. 2012) at LGC Genomics Ltd., Middlesex, UK. Single nucleotide polymorphisms that could be successfully validated using the KASP assays on the validation set were then used for genotyping on the peanut reference set to demonstrate their utility.

## Analysis of Genotypic Data

Genotyping data obtained based on the florescence detected from the KASP assay was graphically viewed through the SNPviewer2 version 3.2.0.9 software (LGC Genomics, 2013). Also, the consistency between the predicted SNP and assayed ones was checked for each SNP individually. Polymorphism information content value, major allele frequency, and heterozygosity were calculated for all the SNPs using PowerMarker version 3.25 software

Figure 1. Graphical representation showing global coverage of the reference set in peanut across 48 countries.

**Table 2. Germplasm lines used to generate transcriptome sequence deposited in the National Center for Biotechnology Information (NCBI) sequence read archive (http://www.ncbi.nlm.nih.gov/sra/; accessed 15 Apr. 2013).**

| Cultivar or germplasm name | *Arachis hypogaea* subspecies and variety | Plant Introduction number | NCBI Accession | Megabases |
|---|---|---|---|---|
| N08082olJCT | subsp. *hypogaea* var. *hypogaea* | | SRX022031 | 13.5 |
| Basse | subsp. *hypogaea* var. *hypogaea* | PI 229553 | SRX022121 | 54.6 |
| C76-16 | subsp. *hypogaea* var. *hypogaea* | | SRX022122 | 51.9 |
| Dixie Giant | subsp. *hypogaea* var. *hypogaea* | PI 290676 | SRX022108 | 13.1 |
| Florida 07 | subsp. *hypogaea* var. *hypogaea* | PI 652938 | SRX022109 | 23.2 |
| Florunner | subsp. *hypogaea* var. *hypogaea* | PI 565448 | SRX022110 | 25.7 |
| Georgia Green | subsp. *hypogaea* var. *hypogaea* | PI 587093 | SRX022111 | 54.5 |
| Georgia Valencia | subsp. *fastigiata* var. *fastigiata* | PI 617040 | SRX022112 | 27.4 |
| GKBSPSc 30062 | *Arachis monticola* | PI 468196 | SRX022113 | 29.2 |
| Gregory | subsp. *hypogaea* var. *hypogaea* | PI 608666 | SRX022114 | 62.1 |
| NC 3033 | subsp. *hypogaea* var. *hypogaea* | | SRX022115 | 26.2 |
| NC-12C | subsp. *hypogaea* var. *hypogaea* | PI 596406 | SRX019971 and SRX019979 | 21.5 and 15.9 |
| New Mexico Valencia A | subsp. *fastigiata* var. *fastigiata* | PI 565452 | SRX022116, SRX020012, and SRX019972 | 28.1, 23.7, and 22.4 |
| OLin | subsp. *fastigiata* var. *vulgaris* | PI 631176 | SRX022117 | 27.2 |
| Overo Chiquitano | subsp. *hypogaea* var. *hypogaea* | PI 313949 | SRX022107 | 55.6 |
| SPT 06-06 | interspecific | | SRX022118 | 40.6 |
| SSD 6 | subsp. *hypogaea* var. *hirsuta* | PI 576638 | SRX022119 | 35.8 |
| SunOleic 97R | subsp. *hypogaea* var. *hypogaea* | PI 596800 | SRX022120 | 38.6 |
| Tifrunner | subsp. *hypogaea* var. *hypogaea* | PI 644011 | SRX020014 | 513 |

(Liu and Muse, 2005). Number of polymorphic SNPs and polymorphism rate were also computed between parents of mapping populations. Furthermore, heterozygosity percentage in genotypes was calculated as a percentage of total heterozygous alleles observed from total alleles amplified for each marker and PIC for the genome types AA, BB, EE, EX, PP, and AABB in the reference set.

Unweighted pair-group method with arithmetic mean (UPGMA) method of clustering in the neighbor-joining procedure as proposed by Saitou and Nei (1987) was used on KASP SNP data for cluster analysis using the DARwin version 5.0.158 software (Perrier and Jacquemound-Collet, 2006). The dissimilarity matrix obtained between the lines was considered as an indicator of the relative diversity of the genetic base. Furthermore, tree construction was also accomplished using the software Dendroscope version 3.2.2 for better visual representation as detailed in Huson et al. (2007). The lines were then analyzed with respect to the three categories, namely botanical variety, species, and genome type.

# Results

## Development and Validation of Kompetitive Allele Specific Polymerase Chain Reaction Markers

A set of 96 SNPs showing high-confidence calls on the GoldenGate assay was used for developing robust KASP assays. These KASP assay-based markers have been designated as groundnut KASP assay markers (GKAMs). For validation, all 96 GKAMs were genotyped on the validation set (Supplemental Table S2). This panel includes parental genotypes of the 27 mapping populations segregating for various disease resistance traits (rust, groundnut rosette disease, late leaf spot, and early leaf spot), drought tolerance related traits, oil quality (high oil content and oleic:linoleic acid ratio), and also for other agronomic traits (Table 1). A total of 90 GKAMs (93.8%) were successfully validated. There was 100% consistency between the SNPs predicted in silico and allele calls obtained in the KASP assays.

Out of the 90 validated GKAMs, 73 GKAMs (81.1%) showed polymorphism across 94 genotypes. Among these markers, 71 GKAMs showed polymorphism between parental lines of at least one of 27 mapping populations, except TAG 24 × ICGV 86031. Of the 27 populations, six were interspecific populations where the number of polymorphic GKAMs ranged from 36 (ICGV 87846 × ISATGR 265-5, ICG 00350 × ISATGR 9B, and ICG 00350 × ISATGR 90B) to 44 (ICG 00350 × ISATGR 5B) with an average of 38 GKAMs per population. Among 21 intraspecific populations, the least polymorphism was found for the population Chalimbana × ICGVSM 90704 (two GKAMs) and the maximum for Robut 33-1 × ICGV 95714 and ICGV 93437 × ICGVSM 95342 (23 GKAMs each) with an average of 14 GKAMs per population (Table 1). In general, interspecific mapping populations have shown more polymorphism (42.4%) than the intraspecific mapping populations (14.1%).

The PIC values for the polymorphic GKAMs in the above mentioned genotypes varied between 0.02 (GKAM0044) to 0.38 (GKAM0060 and GKAM0090) with an average of 0.32. Heterozygosity percentage ranged from 0.00 (29 GKAMs) to 82.22% (GKAM0040) with an average of 35.36%. Major allele frequency varied from 0.50 (GKAM0060 and GKAM0090) to 0.99 (GKAM0044) with an average of 0.67 (Table 3). Different probable scenarios of SNP genotyping observed in the validation set are shown in Fig. 2.

## Genetic Diversity Analysis in the Reference Set

Screening of 90 GKAMs on the reference set provided high quality data for 88 GKAMs and 72 GKAMs were polymorphic. After comparing the genotypic data of the validation set with the reference set, it was found that the same 16 GKAMs were monomorphic in both the sets. Interestingly, a marker GKAM0036, which was monomorphic in the validation set, was found to be polymorphic in the reference set. In addition, SNP calling could not be achieved in the reference set for two markers (GKAM0006 and GKAM0087), which were polymorphic in the validation set (Table 3).

High-quality genotyping data obtained for all 72 polymorphic GKAMs on the reference set was used for assessment of genetic diversity and understanding genetic relationships. Mean genetic dissimilarity between different pairs of genotypes was found to be 0.13 and maximum between ICG 8200 and ICG 8206 at 0.45. Dissimilarity matrix and hierarchical cluster analysis performed using UPGMA method grouped genotypes into four clusters (Fig. 3). The cluster (Cl) I largely comprised genotypes from subspecies *fastigiata* (120) and its botanical varieties *fastigiata* (35), *peruviana* (one), and *vulgaris* (62); however, a few genotypes from subspecies *hypogaea* (12) and eight genotypes of *A. hypogaea* for which subspecies information is not known were also present in this cluster. In the Cl II, most of the genotypes were from subspecies *hypogaea* (72) and its botanical varieties *hirsuta* (two) and *hypogaea* (66) including nine genotypes from subspecies *fastigiata* and six genotypes of *A. hypogaea* for which subspecies information is not known. Interestingly, the small Cl III had genotypes from both the subspecies (*fastigiata* and *hypogaea*) and also three wild genotypes, representing *A. monticola* (AABB genome) (ICG 8135 and ICG 13177) and *A. cardenasii* Krapov. & W. C. Greg. (AA genome) (ICG 13164). The Cl IV included 31 accessions from 13 wild species from South America, its center of origin, representing five diploid genomes (AA, BB, EE, EX, and PP). By and large, grouping pattern exhibited discrete clustering of genotypes based on subspecies, botanical variety, and genome type.

For the reference set, the PIC values for the polymorphic GKAMs ranged from 0.01 (GKAM0036) to 0.37 (38 GKAMs) with an average of 0.31. Similarly, the heterozygosity percentage varied from 0.00 (eight GKAMs) to 92.16% (GKAM0073), with an average of 59.42%. Major allele frequency ranged from 0.50 (GKAM0090) to 0.99 (five GKAMs) with an average of 0.66 (Table 3). Different probable scenarios of SNP genotyping observed in the reference set are shown in Fig. 2. An attempt was made to calculate genomewise polymorphism, major allele frequency, heterozygosity, and PIC for tetraploid (AABB) and diploid genome groups (AA, BB, and EE). These features could not be calculated for the EX and PP genome, since these genome groups had only one accession each. Polymorphic markers were highest for the AABB genome (70 GKAMs among 248 accessions) followed by AA genome (51 GKAMs among 23 accessions) and BB genome (30 GKAMs among five accessions) while least for EE genome (12 GKAMs among two accessions). Average major allele frequency was maximum in AA genome (0.81) and minimum in EE genome (0.50) while for BB and AABB genome, it was found to be 0.71 and 0.63, respectively. The average heterozygosity observed was highest at 68.04% (AABB genome) followed by 5.94% (AA genome), 3.83% (BB genome), and 0.00% (EE genome). Average PIC ranged from 0.21 (AA genome)

**Table 3. Comparative polymorphism features of groundnut Kompetitive allele specific polymerase chain reaction assay markers (GKAMs) in the validation and reference sets.**

| GKAM marker | Validation set | | | Reference set | | |
|---|---|---|---|---|---|---|
| | Major allele frequency | Heterozygosity % | PIC[†] value | Major allele frequency | Heterozygosity % | PIC value |
| GKAM0001 | 0.70 | 1.09 | 0.33 | 0.84 | 31.39 | 0.23 |
| GKAM0002 | 0.56 | 80.65 | 0.37 | 0.53 | 89.57 | 0.37 |
| GKAM0003 | 0.75 | 0.00 | 0.30 | 0.55 | 89.86 | 0.37 |
| GKAM0004 | 0.94 | 1.11 | 0.11 | 0.86 | 0.00 | 0.21 |
| GKAM0005 | 0.57 | 77.17 | 0.37 | 0.51 | 89.21 | 0.37 |
| GKAM0006 | 0.70 | 50.54 | 0.33 | NA[‡] | NA | NA |
| GKAM0007 | 0.76 | 0.00 | 0.30 | 0.55 | 89.78 | 0.37 |
| GKAM0009 | 0.76 | 0.00 | 0.30 | 0.55 | 89.82 | 0.37 |
| GKAM0010 | 0.62 | 71.11 | 0.36 | 0.53 | 84.73 | 0.37 |
| GKAM0012 | 0.74 | 0.00 | 0.31 | 0.56 | 88.89 | 0.37 |
| GKAM0013 | 0.56 | 81.11 | 0.37 | 0.53 | 89.38 | 0.37 |
| GKAM0014 | 0.75 | 0.00 | 0.31 | 0.85 | 28.78 | 0.22 |
| GKAM0015 | 0.79 | 0.00 | 0.27 | 0.54 | 89.93 | 0.37 |
| GKAM0018 | 0.64 | 0.00 | 0.35 | 0.51 | 3.62 | 0.37 |
| GKAM0020 | 0.69 | 48.35 | 0.34 | 0.68 | 58.61 | 0.34 |
| GKAM0021 | 0.55 | 80.43 | 0.37 | 0.53 | 87.82 | 0.37 |
| GKAM0022 | 0.58 | 33.70 | 0.37 | 0.69 | 43.23 | 0.33 |
| GKAM0023 | 0.75 | 0.00 | 0.30 | 0.55 | 89.71 | 0.37 |
| GKAM0024 | 0.69 | 0.00 | 0.34 | 0.77 | 45.62 | 0.29 |
| GKAM0026 | 0.57 | 55.06 | 0.37 | 0.68 | 61.69 | 0.34 |
| GKAM0028 | 0.81 | 0.00 | 0.26 | 0.88 | 23.47 | 0.19 |
| GKAM0029 | 0.75 | 0.00 | 0.30 | 0.55 | 90.18 | 0.37 |
| GKAM0030 | 0.57 | 53.93 | 0.37 | 0.63 | 57.40 | 0.36 |
| GKAM0031 | 0.59 | 0.00 | 0.37 | 0.64 | 71.79 | 0.36 |
| GKAM0033 | 0.65 | 0.00 | 0.35 | 0.82 | 36.10 | 0.25 |
| GKAM0034 | 0.55 | 81.32 | 0.37 | 0.53 | 89.38 | 0.37 |
| GKAM0035 | 0.75 | 46.15 | 0.31 | 0.66 | 66.42 | 0.35 |
| GKAM0036 | MONO[§] | MONO | MONO | 0.99 | 0.00 | 0.01 |
| GKAM0037 | 0.59 | 76.34 | 0.37 | 0.53 | 89.13 | 0.37 |
| GKAM0038 | 0.62 | 2.20 | 0.36 | 0.84 | 32.60 | 0.24 |
| GKAM0039 | 0.71 | 0.00 | 0.33 | 0.86 | 27.11 | 0.21 |
| GKAM0040 | 0.51 | 82.22 | 0.37 | 0.51 | 87.96 | 0.37 |
| GKAM0041 | 0.57 | 80.22 | 0.37 | 0.54 | 88.41 | 0.37 |
| GKAM0042 | 0.61 | 74.44 | 0.36 | 0.99 | 0.00 | 0.02 |
| GKAM0044 | 0.99 | 0.00 | 0.02 | 0.99 | 0.00 | 0.03 |
| GKAM0045 | 0.64 | 1.19 | 0.36 | 0.53 | 6.08 | 0.37 |
| GKAM0046 | 0.65 | 0.00 | 0.35 | 0.62 | 73.72 | 0.36 |
| GKAM0047 | 0.86 | 1.45 | 0.21 | 0.88 | 0.00 | 0.19 |
| GKAM0050 | 0.71 | 0.00 | 0.33 | 0.75 | 49.63 | 0.30 |
| GKAM0052 | 0.51 | 79.57 | 0.37 | 0.51 | 90.55 | 0.37 |
| GKAM0053 | 0.75 | 43.48 | 0.30 | 0.73 | 51.28 | 0.32 |
| GKAM0057 | 0.75 | 0.00 | 0.30 | 0.55 | 89.21 | 0.37 |
| GKAM0058 | 0.65 | 0.00 | 0.35 | 0.84 | 31.52 | 0.23 |
| GKAM0060 | 0.50 | 80.43 | 0.38 | 0.51 | 89.45 | 0.37 |
| GKAM0062 | 0.78 | 0.00 | 0.28 | 0.55 | 90.51 | 0.37 |
| GKAM0063 | 0.91 | 0.00 | 0.14 | 0.92 | 15.94 | 0.14 |
| GKAM0064 | 0.57 | 79.35 | 0.37 | 0.54 | 89.53 | 0.37 |
| GKAM0065 | 0.95 | 0.00 | 0.10 | 0.89 | 0.00 | 0.17 |
| GKAM0066 | 0.62 | 68.89 | 0.36 | 0.52 | 89.10 | 0.37 |
| GKAM0067 | 0.59 | 78.26 | 0.37 | 0.53 | 89.57 | 0.37 |
| GKAM0069 | 0.94 | 0.00 | 0.10 | 0.95 | 0.00 | 0.09 |
| GKAM0070 | 0.62 | 74.19 | 0.36 | 0.55 | 88.17 | 0.37 |
| GKAM0071 | 0.71 | 1.11 | 0.33 | 0.85 | 30.15 | 0.22 |
| GKAM0072 | 0.57 | 80.43 | 0.37 | 0.53 | 90.58 | 0.37 |
| GKAM0073 | 0.53 | 80.00 | 0.37 | 0.51 | 92.16 | 0.37 |
| GKAM0074 | 0.60 | 75.00 | 0.36 | 0.55 | 88.85 | 0.37 |
| GKAM0075 | 0.55 | 79.35 | 0.37 | 0.51 | 89.53 | 0.37 |
| GKAM0076 | 0.62 | 66.29 | 0.36 | 0.52 | 83.58 | 0.37 |
| GKAM0077 | 0.88 | 0.00 | 0.19 | 0.93 | 1.09 | 0.12 |
| GKAM0078 | 0.61 | 0.00 | 0.36 | 0.74 | 51.44 | 0.31 |
| GKAM0079 | 0.59 | 75.27 | 0.37 | 0.54 | 89.21 | 0.37 |
| GKAM0080 | 0.89 | 0.00 | 0.17 | 0.99 | 2.91 | 0.03 |
| GKAM0082 | 0.73 | 47.31 | 0.32 | 0.76 | 44.89 | 0.30 |
| GKAM0085 | 0.55 | 78.02 | 0.37 | 0.53 | 89.57 | 0.37 |
| GKAM0086 | 0.76 | 0.00 | 0.30 | 0.55 | 88.04 | 0.37 |
| GKAM0087 | 0.58 | 76.34 | 0.37 | NA | NA | NA |
| GKAM0088 | 0.80 | 0.00 | 0.27 | 0.53 | 89.89 | 0.37 |
| GKAM0089 | 0.57 | 1.22 | 0.37 | 0.67 | 0.74 | 0.34 |
| GKAM0090 | 0.50 | 75.56 | 0.38 | 0.50 | 89.45 | 0.37 |
| GKAM0092 | 0.77 | 0.00 | 0.29 | 0.90 | 0.00 | 0.16 |
| GKAM0093 | 0.54 | 79.35 | 0.37 | 0.51 | 89.78 | 0.37 |
| GKAM0094 | 0.55 | 48.91 | 0.37 | 0.60 | 60.89 | 0.36 |
| GKAM0095 | 0.72 | 52.22 | 0.32 | 0.69 | 61.22 | 0.34 |
| GKAM0096 | 0.56 | 80.65 | 0.37 | 0.53 | 88.13 | 0.37 |
| Mean | 0.67 | 35.36 | 0.32 | 0.66 | 59.42 | 0.31 |

[†]PIC, polymorphism information content.

[‡]NA, not available; single nucleotide polymorphism calling could not be achieved.

[§]MONO, monomorphic marker.

to 0.38 (EE genome), while BB and AABB genomes recorded 0.31 and 0.32, respectively.

## Discussion

Currently SSR markers are considered as the markers of choice for genetics research and breeding applications in peanut. Although >13,000 SSR markers have been developed so far in peanut, low levels of polymorphism pose a serious challenge for progress towards construction of dense genetic maps for conducting QTL analysis to identify linked markers for use in molecular breeding. In addition, nonavailability of high-throughput cost-effective genotyping marker assays further decelerates the effort towards germplasm use in genetic mapping and breeding applications in tetraploid peanut. Nevertheless, in other legumes, several high-throughput and cost-effective marker assays have been developed in recent times such as DArTs in chickpea (Thudi et al., 2011), pigeonpea (Yang et al., 2011), and peanut (Mallikarjuna et al., 2011) along with KASP assays in chickpea (Hiremath et al., 2012) and pigeonpea (Saxena et al., 2012). Furthermore, GoldenGate assays with 768 SNPs have also been developed for chickpea and pigeonpea (see Varshney et al., 2012). In the case of peanut, an Illumina GoldenGate assay for 1536 SNPs

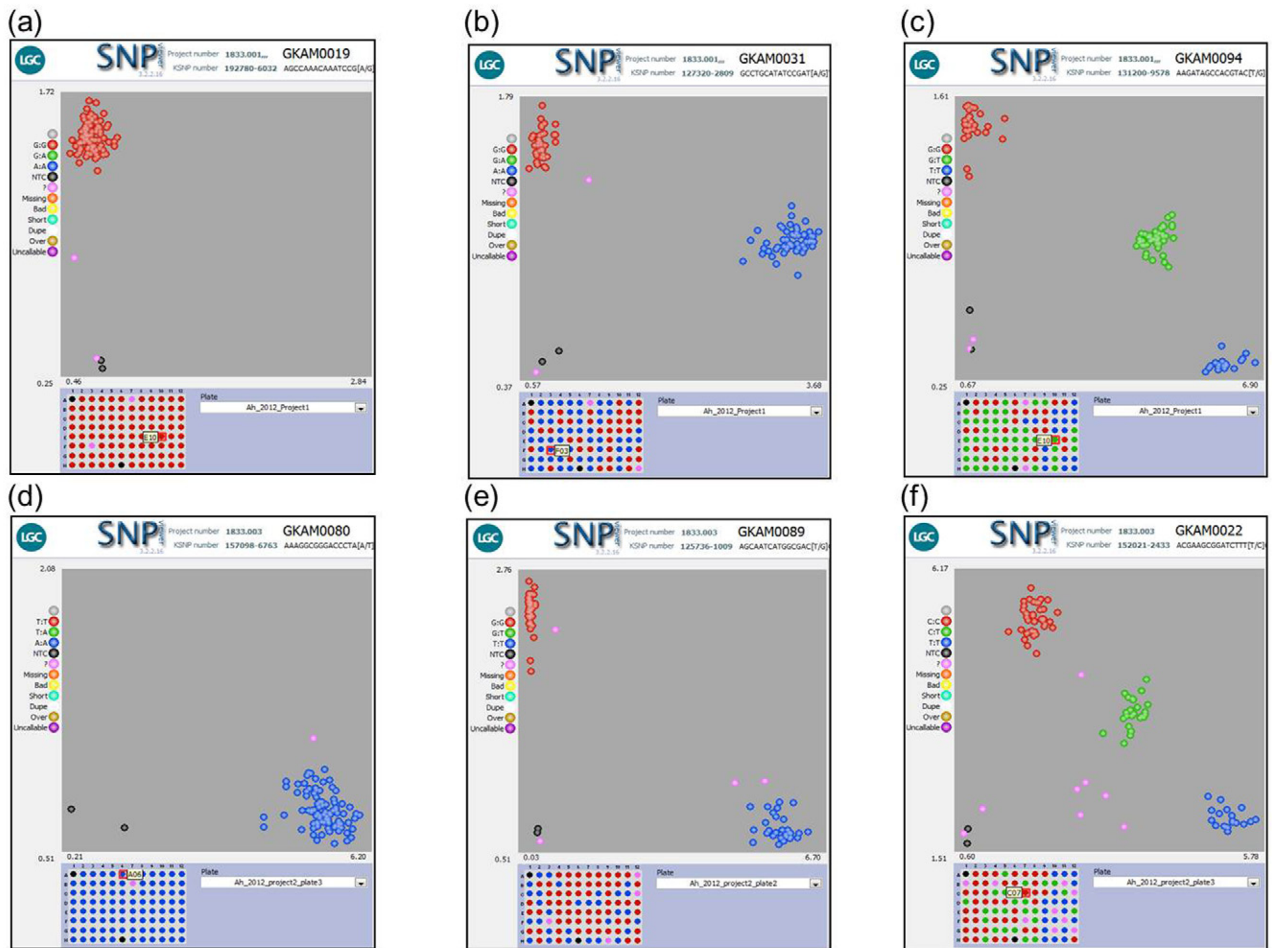


Figure 2. Snapshot displaying single nucleotide polymorphism (SNP) genotyping with Kompetitive allele specific polymerase chain reaction (KASP) assays. Different probable scenarios of SNP genotyping observed in the validation set (a–c) and the reference set (d–f) have been shown. Marker genotyping data generated for each genotype were used for allele calling using automatic allele calling option and viewed using the SNPviewer software (LGC Genomics, 2013). The scatter plot with axes *x* and *y* represents allelic discrimination for a particular marker in the genotypes examined. For validation set, the snapshot (a) shows monomorphic pattern, that is, occurrence of only one allele (red spots) for GKAM0019, snapshot (b) represents a polymorphic pattern, that is, occurrence of two alleles (red and blue spots) for GKAM0031 in nearly equal proportion, and the snapshot (c) shows heterozygosity, that is, occurrence of two alleles (green spots) for GKAM0094. Similarly for the reference set, the snapshots (d), (e), and (f) represent SNP genotyping pattern for monomorphic (GKAM0080), polymorphic (GKAM0089), and heterozygous (GKAM0022) markers.

has been developed at the University of Georgia (Guo et al., 2012; Nagy et al., 2012); however, these assays are not cost effective especially when small numbers of samples need to be genotyped and the polymorphism rate is very low (Mir et al., 2013).

Most genetics and breeding applications require flexibility in genotyping, that is, varying numbers of SNPs with varying numbers of samples. In such cases, SNP genotyping such as primer extension followed by matrix-assisted laser desorption ionization–time of flight (MALDI-TOF), an alternative to Sequenom assays (Sequenom, Inc.) (Sauer et al., 2000), dynamic allele-specific hybridization (Podder et al., 2008), and arrayed primer extension reaction (Podder et al., 2008) are available. However, more recently LGC Genomics KASP (see Semagn et al., 2013) has emerged as a marker of choice given its robustness, amenability to flexibility, and multiplexing and cost-effective nature for screening small to large populations with tens to hundreds of markers (Hiremath et al., 2012; Saxena et al., 2012; Semagn et al., 2013). This is the first study in which the cost-effective and flexible KASP assays have been developed in peanut.

Analysis of KASP assays with the validation set showed 93.75% assay conversion rate that is generally higher than many other SNP genotyping platforms (87.8% for SNPstream (Beckman Coulter, Inc.), >80% for iPlex assays (Sequenom, Inc.) on the MassARRAY platform (Sequenom, Inc.) and Molecular Inversion Probes (Affymetrix, Inc.), and around 80% for GoldenGate and Infinium (Illumina, Inc.) assays; see Ragoussis, 2009). This rate is above the average of 90% when compared with a wide variety of organisms (Semagn et al., 2013) and other legumes studied recently such as pigeonpea (88.4%) (Saxena et al., 2012) and chickpea (80.6%) (Hiremath et al., 2012). This

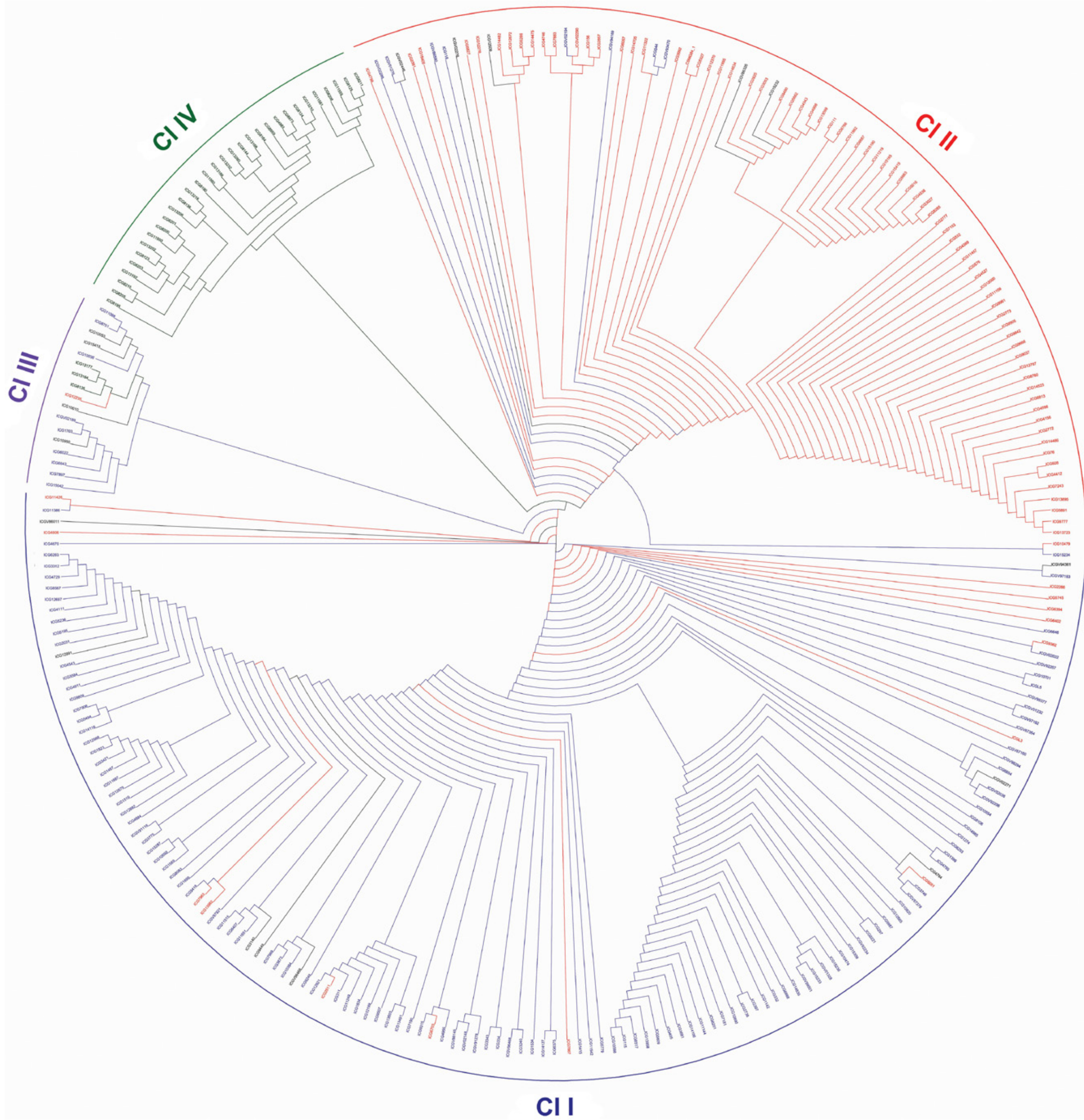


Figure 3. Clustering pattern for *Arachis* species among peanut reference set. Genotypes with blue, red, green, and grey line color indicates *Arachis hypogaea* subsp. *fastigiata*, *A. hypogaea* subsp. *hypogaea*, wild species, and unknown subspecies, respectively. Cl, cluster.

may be attributed to following two reasons: (i) use of stringent criteria and deep sequencing data for SNP discovery and (ii) selection of SNPs after their first round of validation on Illumina's GoldenGate assays. A higher polymorphism rate (81.1%) was observed in the peanut lines of the validation set, as compared to other validation and diversity studies in legumes, for example, pigeonpea (77.4%) (Saxena et al., 2012) and chickpea (66.8%) (Hiremath et al., 2012). This may be attributed to use of seven synthetic (autotetraploid and amphidiploid) lines and 19 wild accessions in the validation set. For instance, after excluding these lines from the validation set, the polymorphism rate drops down from 81.1 to 71.1%, which is a little higher than chickpea while lower than the pigeonpea that reconfirms much narrower genetic diversity in all the three legumes.

For the reference set, the polymorphism rate was 80.0%, which was slightly lower than the validation set (81.1%). Similarly, the PIC values of the polymorphic GKAMs were also found to be slightly lower in the reference set (average 0.31) than the validation set (average 0.32) even though the reference set has threefold more genotypes than the validation set. This is because both the sets had representation of both diploids as well

as tetraploids and share several common genotypes. In terms of heterozygosity, the average heterozygosity is lower (35.36%) in the validation set than the reference set (59.42%). This may be due to the presence of elite lines in validation set, which have been selfed for several generations leading to high level of homozygosity. On the other side the reference set has a mix of genotypes including landraces and wilds. Although only single plant was used for DNA isolation, still there are chances of heterozygosity in landraces and wilds owing to their genetic nature during domestication. In the context of dissecting the heterozygosity percentage at different genome levels in the reference set, it was found that the tetraploid genome AABB was highly heterozygous (68.04%) followed by AA (5.94%), BB (3.83%), and zero in EE genome. This may be attributed to the tetraploid genome as well as presence of homeo-SNPs (between A- and B- genome), which eventually make the interpretation of true heterozygosity detected through SNP genotyping data more difficult. Therefore, GKAMs without heterozygosity are of greater relevance than the GKAMs with heterozygosity, particularly for deployment in early generations of breeding populations.

Genotyping data were used for estimating genetic dissimilarity and relationships in the reference set. Maximum dissimilarity of 45% was observed between the genotypes ICG 8200 (AA genome) (*A. duranensis* Krapov. & W. C. Greg.) and ICG 8206 (BB genome) (*A. ipaensis* Krapov. & W. C. Greg.). However, an earlier study on genetic relationships determined that the accessions ICG 8200 (AA genome) and ICG 8206 (BB genome) were most closely related to tetraploid *A. hypogaea* (Koppolu et al., 2010). Cluster analysis based on the genetic dissimilarity index revealed clear grouping of 280 peanut accessions into four clusters as per their subspecies and species type. The Cl I, Cl II, Cl III, and Cl IV contained accessions from subspecies *fastigiata*, subspecies *hypogaea*, and both subspecies and wild species, respectively. During cluster analysis, some genotypes of subspecies *fastigiata* and *hypogaea* overlapped between the Cl I and Cl II; this is understandable as during the course of evolution, there has been considerable intercrossing among the subspecies *fastigiata* and *hypogaea*. Two tetraploid genotypes from *A. monticola* (ICG 8135 and ICG 13177) and one diploid (AA genome) genotype of *A. cardenasii* (ICG 13164) grouped into Cl III, along with 19 accessions of *A. hypogaea*. Interestingly, four of the five accessions from the botanical variety *peruviana* clustered in Cl III. Furthermore, the fourth cluster (Cl IV) exclusively included 31 accessions from 13 wild species. Several earlier studies on genetic diversity reported *A. monticola* to be genetically closely related to *A. hypogaea* (Koppolu et al., 2010; Moretzsohn et al., 2013). The distinct clustering pattern of wild and cultivated genotypes was also observed in the earlier genetic diversity studies through using SSR and EST-derived SSR marker systems (Moretzsohn et al., 2004; Kottapalli et al., 2007; Koppolu et al., 2010). It clearly reveals that the level of diversity between wild species and cultivated peanut is very high. By and large, the grouping pattern in our study exhibited discrete clustering of genotypes based on subspecies, botanical variety, and genome type. Sensible use of the available genetic diversity in germplasm is the pillar of crop improvement. The narrow genetic variation observed in cultivated tetraploid peanut may be due to its very recent origin in evolutionary time as compared to other crops and is a serious genetic bottleneck towards modern breeding efforts. Hence, tapping the maximum genetic variation available in the primary genepool is vital for peanut crop improvement.

In summary, the present study adds a new type of SNP marker system to the marker repertoire of peanut, which is cost effective and highly flexible. In addition to understanding diversity features and genetic relationships among the genotypes of the reference set, this study also provides polymorphic markers between parental genotypes of different mapping populations. The polymorphic markers can be used for genotyping the respective populations for developing and/or enriching the genetic maps and trait mapping. Therefore, GKAMs developed here are expected to enhance adoption of SNP marker technology for genetics and breeding applications in peanut.

## Supplemental Information Available

Supplemental material is available at http://www.crops.org/publications/tpg.

Supplemental Table S1. List of reference set genotypes used for diversity analysis.

Supplemental Table S2. List of genotypes included in validation set.

Supplemental Table S3. The groundnut Kompetitive allele specific polymerase chain reaction (KASP) assay marker (GKAM) identification (ID), clone ID, and sequence used for designing Kompetitive allele specific polymerase chain reaction (KASP) assay in peanut.


## Acknowledgments

This work has been undertaken as part of the CGIAR Research Program on Grain Legumes. ICRISAT is a member of CGIAR Consortium. The work presented in this article is contribution from research projects sponsored by CGIAR Generation Challenge Programme (Theme Leader Discretionary Grant), The Bill & Melinda Gates Foundation (Tropical Legumes I & II).


## References


Alves, D.M.T., R.W. Pereira, S.C. Leal-Bertioli, M.C. Moretzsohn, P.M. Guimarães, and D.J. Bertioli. 2008. Development and use of single nucleotide polymorphism markers for candidate resistance genes in wild peanuts (*Arachis* spp). Genet. Mol. Res. 7:631–642. doi:10.4238/vol7-3gmr453

Bravo, J.P., A.A. Hoshino, C. Angelici, C.R. Lopes, and M.A. Gimenes. 2006. Transferability and use of microsatellite markers for the genetic analysis of the germplasm of some *Arachis* section species of the genus *Arachis*. Genet. Mol. Biol. 29:516–524. doi:10.1590/S1415-47572006000300021

Chen, X., W. Zhu, S. Azam, H. Li, F. Zhu, H. Li, Y. Hong, H. Liu, E. Zhang, H. Wu, S. Yu, G. Zhou, S. Li, N. Zhong, S. Wen, X. Li, S.J. Knapp, P. Ozias-Akins, R.K. Varshney, and X. Liang. 2013. Deep sequencing analysis of the transcriptomes of peanut aerial and subterranean young pods identifies candidate genes related to early embryo abortion. Plant Biotech. J. 11: 115–127.

Cuc, L.M., E.S. Mace, J.H. Crouch, V.D. Quang, T.D. Long, and R.K. Varshney. 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. doi:10.1186/1471-2229-8-55

Dwivedi, S.L., S. Gurtu, S. Chandra, W. Yuejin, and S.N. Nigam. 2001. Assessment of genetic diversity among selected groundnut germplasm I: RAPD analysis. Plant Breed. 120:345–349. doi:10.1046/j.1439-0523.2001.00613.x

Gautami, B., K. Ravi, M. LakshimiNarasu, D.A. Hoisington, and R.K. Varshney. 2009. Novel set of groundnut SSR markers for germplasm analysis and interspecific transferability. Int. J. Integr. Biol. 7:100–106.

Guo, Y., S. Khanal, S. Tang, J.E. Bowers, A.F. Heesacker, N. Khalilian, E.D. Nagy, D. Zhang, C.A. Taylor, H.T. Stalker, P. Ozias-Akins, and S.J. Knapp. 2012. Comparative mapping in intraspecific populations uncovers a high degree of macrosynteny between A- and B- genome diploid species of peanut. BMC Genomics 13:608. doi:10.1186/1471-2164-13-608

Gupta, P.K., and R.K. Varshney. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185. doi:10.1023/A:1003910819967

He, G., and S.H. Prakash. 2001. Evaluation of genetic relationship among botanical varieties of cultivated peanut (*Arachis hypogaea* L.) using AFLP markers. Genet. Resour. Crop Evol. 48:347–352. doi:10.1023/A:1012019600318

Herselman, L. 2003. Genetic variation among southern African cultivated peanut (*Arachis hypogaea* L.) genotypes as revealed by AFLP analysis. Euphytica 133:319–327. doi:10.1023/A:1025769212187

Hilu, K.W., and H.T. Stalker. 1995. Genetic relationship between peanuts and wild species of *Arachis* section *Arachis*: Evidence from RAPD. Plant Syst. Evol. 198:167–178. doi:10.1007/BF00984735

Hiremath, P.J., A. Kumar, R.V. Penmetsa, A. Farmer, J.A. Schlueter, K. Siva, A.M. Whaley, N. Carrasquilla-Garcia, P.M. Gaur, H.D. Upadhyaya, P.B.K. Kishor, T.M. Shah, D.R. Cook, and R.K. Varshney. 2012. Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnol. J. 10:716–732. doi:10.1111/j.1467-7652.2012.00710.x

Holbrook, C.C., and H.T. Stalker. 2003. Peanut breeding and genetic resources. Plant Breed. Rev. 22:297–356.

Huson, D.H., D.C. Richter, C. Rausch, T. Dezulian, M. Franz, and R. Rupp. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinf. 8:460. doi:10.1186/1471-2105-8-460

Kochert, G., H.T. Stalker, M. Gimenes, L. Galgaro, and K. Moore. 1996. RFLP and cytogenetic evidence for the progenitor species of allotetraploid cultivated peanut (*Arachis hypogaea* L.). Am. J. Bot. 83:1282–1291. doi:10.2307/2446112

Koppolu, R., U. Hari, D. Sangam, H. David, and R.K. Varshney. 2010. Genetic relationships among seven sections of genus *Arachis* studied by using SSR markers. BMC Plant Biol. 10:15. doi:10.1186/1471-2229-10-15

Kottapalli, K.R., M.D. Burow, G. Burow, J. Burke, and N. Puppala. 2007. Molecular characterization of the U.S. peanut mini core collection using microsatellite markers. Crop Sci. 47:1718–1727. doi:10.2135/cropsci2006.06.0407

LGC Genomics. 2013. SNPviewer. SNPviewer. LGC Limited, Teddington, UK. http://www.lgcgenomics.com/services/instruments-and-software/software/?id=78 (accessed 3 Mar. 2013).

Liu, K., and S. V. Muse. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. doi:10.1093/bioinformatics/bti282

Mallikarjuna, N., S. Senthilvel, and D. Hoisington. 2011. Development of new sources of tetraploid *Arachis* to broaden the genetic base of cultivated groundnut (*Arachis hypogaea* L.). Genet. Resour. Crop Evol. 58:889–907. doi:10.1007/s10722-010-9627-8

Mammadov, J., W. Chen, J. Mingus, S. Thompson, and S. Kumpatla. 2011. Development of versatile gene based SNP assays in maize (*Zea mays* L.). Mol. Breed. 3:779–790.

Mir, R.R., P.J. Hiremath, O. Riera-Lizarazu, and R.K. Varshney. 2013. Evolving molecular marker technologies in plants: From RFLPs to GBS. In: T. Lübberstedt and R.K. Varshney, editors, Diagnostics in plant breeding. Springer, Dordrecht, Netherlands. p. 229–247.

Moretzsohn, M.C., E.G. Gouves, P.W. Inglis, S.C.M. Leal-Bertioli, J.F.M. Valls, and D.J. Bertioli. 2013. A study of the relationships of cultivated peanut (*Arachis hypogaea* L.) and its most closely related wild species using intron sequences and microsatellite markers. Ann. Bot. (London) 111:113–126. doi:10.1093/aob/mcs237

Moretzsohn, M.C., M.S. Hopkins, S.E. Mitchell, S. Kresovich, J.F.M. Valls, and M.E. Ferreira. 2004. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hyper variable regions of the genome. BMC Plant Biol. 4:11. doi:10.1186/1471-2229-4-11

Nagy, E.D., Y. Guo, S. Tang, J.E. Bowers, R.A. Okashah, C.A. Taylor, D. Zhang, S. Khanal, A.F. Heesacker, N. Khalilian, A.D. Farmer, N. Carrasquilla-Garcia, R.V. Penmetsa, D. Cook, H.T. Stalker, N. Nielsen, P. Ozias-Akins, and S.J. Knapp. 2012. A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. BMC Genomics 13:469. doi:10.1186/1471-2164-13-469

Neelam, K., G. Brown-Guedira, and L. Huang. 2013. Development and validation of a breeder-friendly KASPar marker for wheat leaf rust resistance locus *Lr21*. Mol. Breed. 31:233–237. doi:10.1007/s11032-012-9773-0

Pandey, M.K., E.S. Monyo, P. Ozias-Akins, X. Liang, P. Guimarães, S.N. Nigam, H.D. Upadhyaya, P. Janila, X. Zhang, B. Guo, D.R. Cook, D.J. Bertioli, R. Michelmore, and R.K. Varshney. 2012. Advances in *Arachis* genomics for peanut improvement. J. Biotechnol. Adv. 30:639–651. doi:10.1016/j.biotechadv.2011.11.001

Perrier, X., and J.P. Jacquemoud-Collet. 2006. DARwin software. Centre de coopération internationale en recherche agronomique pour le développement (CIRAD), Paris, France. http://darwin.cirad.fr/ (accessed 17 Apr. 2013).

Podder, M., J. Ruan, B.W. Tripp, Z.E. Chu, and S.J. Tebbutt. 2008. Robust SNP genotyping by multiplex PCR and arrayed primer extension. BMC Med. Genet. 1:5.

Ragoussis, J. 2009. Genotyping technologies for genetic research. Annu. Rev. Genomics Hum. Genet. 10:117–133. doi:10.1146/annurev-genom-082908-150116

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Sauer, S., D. Lechner, K. Berlin, C. Plançon, A. Heuermann, and H. Lehrach. 2000. Full flexibility genotyping of single nucleotide polymorphisms by the good assay. Nucleic Acids Res. 28:E100. doi:10.1093/nar/28.23.e100

Saxena, R.K., R.V. Penmetsa, H.D. Upadhyaya, A. Kumar, N.C. Garcia, J.A. Schlueter, A. Farmer, A.M. Whaley, B.K. Sarma, G.D. May, D.R. Cook, and R.K. Varshney. 2012. Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeonpea and comparative mapping in legumes. DNA Res. 19:449–461. doi:10.1093/dnares/dss025

Semagn, K., R. Babu, S. Hearne, and M. Olsen. 2013. Single nucleotide polymorphism genotyping using Kompetitive allele specific PCR (KASP): Overview of the technology and its application in crop improvement. Mol. Bred. doi:10.1007/s11032-013-9917-x

Subramanian, V., S. Gurtu, R.C. Nageswara Rao, and S.N. Nigam. 2000. Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. Genome 43:656–660. doi:10.1139/g00-034

Thudi, M., A. Bohra, S.N. Nayak, N. Varghese, T.M. Shah, R.V. Penmetsa, N. Thirunavukkarasu, S. Gudipati, P.M. Gaur, P.L. Kulwal, H.D. Upadhyaya, P.B.K. Kishor, P. Winter, G. Kahl, C.D. Town, A. Kilian, D.R. Cook, and R.K. Varshney. 2011. Novel SSR markers from BAC-End sequences, DArT arrays and a comprehensive genetic map with 1291 marker loci for chickpea (*Cicer arietinum* L.). PLoS ONE 6:e27275. doi:10.1371/journal.pone.0027275

Upadhyaya, H.D., P.J. Bramel, R. Ortiz, and S. Singh. 2002. Developing a mini core of peanut for utilization of genetic resources. Crop Sci. 42:2150–2156. doi:10.2135/cropsci2002.2150

Upadhyaya, H.D., R. Ortiz, P.J. Bramel, and S. Singh. 2003. Development of a groundnut core collection using taxonomical, geographical and

morphological descriptors. Genet. Resour. Crop Evol. 50:139–148. doi:10.1023/A:1022945715628

Upadhyaya, H.D., D. Yadav, N. Dronavalli, C.L.L. Gowda, and S. Singh. 2010. Mini core germplasm collections for infusing genetic diversity in plant breeding programs. Electron. J. Plant Breed. 1:1294–1309.

Varshney, R.K., H. Kudapa, M. Roorkiwal, M. Thudi, M.K. Pandey, R.K. Saxena, S.K. Chamarthi, S. Murali Mohan, N. Mallikarjun, H.D. Upadhyaya, P.M. Gaur, L. Krishnamurthy, K.B. Saxena, S.N. Nigam, and S. Pande. 2012. Advances in genetics and molecular breeding of three legume crops of semi-arid tropics using next-generation sequencing and high-throughput genotyping technologies. J. Biosci. 37:811–820. doi:10.1007/s12038-012-9228-0

Varshney, R.K., T. Mahendar, R. Aruna, S.N. Nigam, K. Neelima, and V. Vadez. 2009a. High level of natural variation in a groundnut (*Arachis hypogaea* L.) germplasm collection assayed by selected informative SSR markers. Plant Breed. 128:486–494. doi:10.1111/j.1439-0523.2009.01638.x

Varshney, R.K., S. Murali Mohan, P.M. Gaur, N.V.P.R. Gangarao, M.K. Pandey, A. Bohra, S.L. Sawargaonkar, A. Chitikineni, P.K. Kimurto, P. Janila, K.B. Saxena, A. Fikre, M. Sharma, A. Rathore, A. Pratap, S. Tripathi, S. Datta, S.K. Chaturvedi, N. Mallikarjuna, G. Anuradha, A. Babbar, A.K. Choudhary, M.B. Mhase, Ch. Bharadwaj, D.M. Mannur, P.N. Harer, B. Guo, X. Liang, N. Nadarajan, and C.L.L. Gowda. 2013. Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol. Adv. doi:10.1016/j.biotechadv.2013.01.001

Varshney, R.K., S.N. Nayak, G.D. May, and S.A. Jackson. 2009b. Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol. 27:522–530. doi:10.1016/j.tibtech.2009.05.006

Yang, S.Y., R.K. Saxena, P.L. Kulwal, G.J. Ash, A. Dubey, J.D.I. Harper, H.D. Upadhaya, R. Gothalwal, A. Kilian, and R.K. Varshney. 2011. The first genetic map of pigeonpea based on diversity arrays technology (DArT) markers. J. Genet. 90:103–109. doi:10.1007/s12041-011-0050-5