Turkish Journal of Botany

Volume 48
Number 7 *SI-1*

Article 5

12-17-2024

Selective InDel Marker Identification Across the Peanut (Arachis hypogaea L.) Genome Using ddRADSeq

MOIN QURESHI

BİRGÜL GUDEN

HİLAL ŞULE TOSUN

MOJTABA KORDROSTAMI

SUNIL SHIWAJI GANGURDE

See next page for additional authors

Recommended Citation

QURESHI, MOIN; GUDEN, BİRGÜL; TOSUN, HİLAL ŞULE; KORDROSTAMI, MOJTABA; GANGURDE, SUNIL SHIWAJI; UZUN, BÜLENT; and YOL, ENGİN (2024) "Selective InDel Marker Identification Across the Peanut (Arachis hypogaea L.) Genome Using ddRADSeq," *Turkish Journal of Botany*: Vol. 48: No. 7, Article 5.
https://doi.org/10.55730/1300-008X.2826
Available at: https://journals.tubitak.gov.tr/botany/vol48/iss7/5

TÜBİTAK

Turkish Journal of Botany

http://journals.tubitak.gov.tr/botany/

Research Article

Turk J Bot
(2024) 48: 441-453

doi:10.55730/1300-008X.2826

Selective InDel marker identification across the peanut (*Arachis hypogaea* L.) genome using ddRADSeq

Moin QURESHI[1], Birgül GUDEN[1], Hilal Şule TOSUN[2], Mojtaba KORDROSTAMI[3], Sunil S. GANGURDE[4], Bülent UZUN[1], Engin YOL[1,*]

[1]Department of Field Crops, Faculty of Agriculture, Akdeniz University, Antalya, Turkiye
[2]Department of Plant Protection, Faculty of Agriculture, Akdeniz University, Antalya, Turkiye
[3]Nuclear Agriculture Research School, Nuclear Science and Technology Research Institute (NSTRI), Karaj, Iran
[4]International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

Received: 04.06.2024 • **Accepted/Published Online:** 10.09.2024 • **Final Version:** 17.12.2024

Abstract: The advent of next-generation sequencing technologies, and particularly double digest restriction site-associated DNA sequencing (ddRADSeq), has significantly advanced the development of molecular markers for crop genetics. This study used ddRADSeq to identify and develop insertion–deletion (InDel) markers in 25 peanut genotypes from diverse geographical regions. The bioinformatic analysis revealed 62,728 InDels across the peanut genome, predominantly between 1 and 5 bp, which constituted 96% of the total, while InDels of ≥6 bp accounted for 3.96%. We focused on 1013 InDels of at least 10 bp for further analysis, representing 1.61% of the total reads, with a distribution of 832 insertions and 181 deletions. Of those, 21 InDels were selected for primer design and successfully amplified to produce markers within the range of 150–400 bp. Approximately 9.5% of the InDels were located in coding sequences, enhancing their potential utility in genomics-led breeding. These markers' polymorphic information content varied from 0 to 0.371, demonstrating substantial genetic diversity with an average value of 0.163. These findings confirm the effectiveness of ddRADSeq for InDel marker development in peanuts, illustrating its potential to enhance marker-assisted breeding programs by providing robust tools for assessing genetic diversity.

Key words: Groundnut, peanut, InDel markers, ddRADSeq, genetic diversity, marker-assisted breeding

1. Introduction

Peanut (*Arachis hypogaea* L.), also known as groundnut, is a self-pollinated crop that belongs to the genus *Arachis* and the family Fabaceae. Unlike other flowering plants, this genus uniquely produces fruits underground, although its flowers, leaves, and stems emerge above the soil (Krapovickas and Gregory, 1994). Peanut is cultivated in more than 120 countries, primarily in low-input farming systems between 40°N and 40°S in the world's semiarid tropical and subtropical regions (Sarkar et al., 2014). With over 30 million hectares under cultivation and annual production exceeding 49.4 million tons, it is the fourth-largest oilseed crop worldwide (USDA, https://ipad.fas.usda.gov).

Peanut seeds are rich in oil (31%–57%) (Yol et al., 2017), protein (22%–30%) (Savage and Keenan, 1994), and carbohydrates (10%–20%) (Taha et al., 2019). The mature seeds predominantly contain palmitic acid (C16:0), oleic acid (C18:1), and linoleic acid (C18:2), which together account for about 90% of the total fatty acids (Young and Waller, 1972). Seeds with high oleic acid contents exhibit a reduced rate of oxidation and a less pronounced paint-like flavor during storage, enhancing their market acceptability (Mozingo et al., 2004). Peanuts are also a valuable source of vitamins E, K, and B and particularly thiamine, niacin, and essential minerals (Kassa et al., 2009). The peanut plant is a significant cash crop, utilized in direct consumption, confectionary preparations, cooking oil production, and animal protein feed (Pandey et al., 2012). Studies have shown that peanut consumption can reduce the risk of coronary heart disease by 37% (Suchoszek-Lukaniuk et al., 2011) and inhibit leukemia cell growth by 50%, suggesting anticancer properties (Hwang et al., 2008).

Despite its high nutritional and economic importance, the production and productivity of peanuts in various regions are constrained by multiple abiotic stresses such as temperature and drought (Hamidou et al., 2013), salinity (El-Akhal et al., 2013), and biotic factors like early leaf spot fungi (Rathod et al., 2020). These challenges are exacerbated by the menace of climate change. Therefore,

* Correspondence: enginyol@akdeniz.edu.tr

developing stress-tolerant and high-yielding peanut cultivars necessitates more efficient breeding techniques.

Despite extensive classical breeding efforts, the on-farm yield of peanut remains below its potential due to the complex interplay between genotype, environment, and management factors. The integration of modern genomics tools into crop breeding has become essential for overcoming these limitations and accelerating the breeding process (Majid et al., 2017). Genomics-assisted breeding (GAB) offers a promising approach to fast-track genetic improvements in peanuts. The availability of high-quality reference genomes for both wild diploid progenitors and cultivated peanuts has significantly advanced the discovery of genes and quantitative trait loci (QTLs), enabling the development of molecular markers and genotyping assays. Recent advancements such as genomic selection, haplotype-based breeding, speed breeding, high-throughput phenotyping, and genome editing are poised to further enhance genetic gains in peanut breeding. These tools, when combined with traditional breeding methods, hold great potential in developing high-yielding and stress-resistant peanut cultivars. Moreover, the efficient selection and targeted use of genetic resources are crucial for designing peanut cultivars with superior adaptation traits. The synergy between GAB, genome editing, and speed breeding is expected to play a pivotal role in meeting future market demands and ensuring food security through the development of improved peanut varieties (Majid et al., 2017; Raza et al., 2024).

Marker-assisted breeding (MAB) enhances agricultural output by identifying trait-related genes. Modern techniques including amplified fragment length polymorphism (AFLP) (He and Prakash, 1997), simple sequence repeats (SSRs) (Liang et al., 2009), single-nucleotide polymorphisms (SNPs) (Zhou et al., 2014), and AhTE markers (Hake et al., 2017) have significantly improved the accuracy and efficiency of diversification research, genetic mapping, and QTL analysis (Roorkiwal et al., 2017). Recent efforts by Bhat et al. (2022) entailing the whole-genome resequencing of 179 *Arachis* accessions significantly contributed to our understanding of the genome-wide structural and functional features of SNPs. That study identified a large number of SNPs, particularly highlighting their distribution across various chromosomal contexts including intergenic, intronic, and exonic regions. Moreover, the study emphasized the importance of SNPs with high functional impacts, such as those affecting stop codons, splice sites, and start codons, which are crucial for advancing GAB in peanut. Additionally, Bhat et al. (2023) explored the distribution and structural features of genes, repeat elements, and transposable elements (TEs) to identify genomic differences between two subspecies of peanut (subsp. *hypogaea* and subsp. *fastigiata*). That study analyzed 67,128 predicted genes and over 2.7 million copies of TEs from the Tifrunner reference genome. Notably, gene density was highest in the telomeric regions, with B03 having the highest number of genes and A08 showing the highest gene density (53 genes/Mb). It was found that 66% of the genes exhibited two or more SNPs, indicating relatively high allelic variation, while 15,731 genes were monomorphic, showing no SNP variation across the 179 accessions. That study also revealed that retroelements, CACTA, and Mu TEs were predominant among the 101 unique types of TEs identified, with an average of 1.8 copies of TEs per gene. Incorporating recent research on the genome-wide landscapes of genes and the repeatome provides valuable context for understanding genome structure and function in peanuts and related species. A comparative repeatome analysis of wild diploid *Arachis* species revealed significant variations in the abundance of satellite DNA and LTR retroelements, emphasizing their roles in genome evolution and heterochromatin content. This study highlights the dynamic nature of repetitive elements and their influence on genome architecture, particularly in cultivated peanut and its wild relatives. Additionally, advancements in genotyping technologies, such as the development of LongTR, have enabled more accurate profiling of tandem repeats using long-read sequencing. The ability of LongTR to detect large repeat expansions, often missed by short-read sequencing, underscores the importance of comprehensive repeatome analysis in understanding genetic variation and its implications for plant breeding (Samoluk et al., 2022; Ziaei Jam et al., 2024).

The progress in understanding the molecular genetics and genomics of peanuts is relatively slower than that for major crops such as rice, soybean, and chickpea (Mishra et al., 2015) due to peanut's large genome of 2.7 Gb (Bertioli et al., 2016) and its allotetraploid nature, which includes a significant amount of repetitive DNA (Singh et al., 2024). These factors contribute to its low genetic diversity and DNA polymorphism rates (Liu et al., 2015a). Although cultivated varieties of peanuts demonstrate considerable diversity in morphological, physiological, and agronomic traits, traditional DNA-level methods such as RFLP, SSRs, and SNPs have revealed only minimal genetic variation (Varshney, 2016).

In response to these challenges, researchers increasingly turn to insertion–deletion (InDel) markers, which offer several advantages over other molecular markers. These include their codominant nature, abundance, accurate identification capabilities, reduced requirements for high-quality and large-volume DNA samples, and a simplified experimental process using commonly available equipment (Gao et al., 2012). InDels that arise from mechanisms such as transposable elements, slippage during replication

of simple sequences, and unequal crossover (Britten et al., 2003) have been employed for analyzing genetic relationships, genetic diversity, genotyping, genetic research, and marker-assisted selection breeding in various crops. These markers have been developed for crops such as cotton (Lu et al., 2015), rice (Liu et al., 2015b), rapeseed (Mahmood et al., 2016), and cucumber (Shen et al., 2013) and they have also been successfully utilized for peanuts, such as InDel02, linked to purple testa color (Huang et al., 2020), and SUC.InDel.A08, linked to sucrose content (Li et al., 2023a). However, the application of double digest restriction site-associated DNA sequencing (ddRADSeq) technology has yet to lead to the development of InDel markers for assessing genetic variation and selecting materials in peanut breeding research.

Therefore, the objectives of this study were to use ddRADSeq technology to detect InDels in 25 peanut genotypes compared to the reference genome sequence, to develop and validate InDel markers, and to evaluate their potential use in genetic diversity studies for cultivated peanuts. These InDel markers might serve as valuable genetic resources for analyzing genetic diversity and enhancing breeding applications in peanuts.

2. Materials and methods

2.1. Plant material

This study used plant material consisting of 25 peanut genotypes obtained from the ICRISAT and USDA gene banks (Table 1). The collection included peanut genotypes from eight regions across Asia, Africa, Europe, and the Americas. The United States had the highest representation of genetic material with six genotypes. The rest were sourced from Bolivia, China, Cyprus, India, Israel, Japan, and Zambia.

2.2. DNA extraction and genotyping with ddRADSeq

Seeds from the peanut genotypes were germinated under controlled greenhouse conditions. Genomic DNA was extracted from fresh leaves using the CTAB method. Subsequently, DNA concentration and quality were assessed using 1% agarose gel electrophoresis. The DNA concentration was standardized to 100 ng/μL using lambda DNA as a reference.

This study utilized the ddRADSeq protocol described by Peterson et al. (2012), with minor modifications as outlined by Basak et al. (2019). Genomic DNA from the 25 genotypes was digested with the restriction enzymes *Vsp*I and *Msp*I. The digested products were separated by

Table 1. List of peanut genotypes used in this study.

GenBank Code	GenBank
ICG 10870	ICRISAT
ICG 10991	ICRISAT
ICG 12635	ICRISAT
ICG 13300	ICRISAT
ICG 13338	ICRISAT
ICG 13357	ICRISAT
ICG 14335	ICRISAT
ICG 5344	ICRISAT
ICG 5687	ICRISAT
ICG 6671	ICRISAT
PI 565459 (NC7)	USDA
PI 269081	USDA
PI 289620	USDA
PI 433350	USDA
PI 442579	USDA
PI 442583	USDA
PI 459094	USDA
PI 512279	USDA
PI 561568	USDA
PI 578304	USDA
PI 584772	USDA
PI 596514	USDA
PI 602062	USDA
PI 614083	USDA
PI 614084	USDA

electrophoresis on 2% agarose gel to determine their size. Only products within the size range of 400–500 bp were selected for Illumina 150-base pair paired-end sequencing (Illumina, San Diego, CA, USA).

2.3. Bioinformatic analysis

In the first step of bioinformatic analysis, raw data were demultiplexed using Je-Demultiplex software (Galaxy Version 1.2.1). Quality control assessment was performed on each FASTQ file using fastp (Galaxy Version 0.23.4+galaxy0). Filtered reads were mapped to the peanut reference genome obtained from the NCBI GenBank (arahy.Tifrunner.gnm1.KYV3) using Bowtie2 software (Galaxy Version 2.5.3+galaxy0), resulting in BAM files for each genotype. Variants were identified using FreeBayes software (Galaxy Version 1.3.1) with a coverage value of 10X. All variant files were combined with the BCFtools merge (Galaxy Version 1.15.1+galaxy3). This combined variant file was transferred to an Excel file to filter insertion and deletion regions. Ultimately, we classified the regions of insertion and deletion based on their sizes and locations within the genome. To ensure that indels located within known minisatellite or microsatellite regions were excluded, the indel regions identified were cross-referenced against databases of known repetitive elements, such as the Tandem Repeats Database (TRDB). Indels that overlapped with or were in close proximity to known minisatellites or microsatellites were filtered out to prevent erroneous marker development. Furthermore, the maximum indel length identified among the 62,728 loci was 26 bp, which was considered in the final analysis to ensure the reliability of the developed markers.

2.4. PCR-based marker design and validation

BAM files and the reference genome were imported into the Integrated Genome Browser (IGB) to detect and confirm all InDel regions for potential marker regions. Primers for the selected regions were designed using Primer3 software. The criteria for primer design included a length of 18–27 nucleotides, melting temperatures ranging from 49 to 60 °C, GC content within the range of 30%–70%, and predicted PCR product lengths between 100 and 600 bp. Notably, since indels of 10 bp were considered in the design, the resulting PCR products could differ by approximately 10 nucleotides in length, enabling the differentiation of alleles based on these size variations. Marker IDs (names) were defined as Groundnut [G]-[deletion (D)/Insertion (I)]-[Chromosome number]-[Chromosomal location].

PCR analyses were performed using a reaction mixture of 20 µL containing 1 µL of DNA (20 ng/µL), 1 µL of Taq buffer (10X), 1 µL of $MgCl_2$ (25 mM), 0.3 µL of dNTPs (10 mM), 0.5 µL of primer (10 µM), 0.2 µL of Taq polymerase (5 U/µl), and 16 µL of H_2O. The PCR program included initial denaturation at 95 °C for 3 min, followed by 35 cycles of denaturation at 95 °C for 45 s, annealing at 50 °C for 20 s, and extension at 60 °C for 50 s. The PCR products were separated on 2% agarose gel and visualized under UV light.

2.5. Genetic diversity

Population genetic parameters, including the number of alleles (Na), effective number of alleles (Ne), Shannon diversity index (I), expected heterozygosity (He), and observed heterozygosity (Ho), were computed using GenAlEx version 6.5. Principal coordinate analysis (PCoA) was performed using PAST software version 3.23. Polymorphism was measured using the Excel Microsatellite Toolkit.

3. Results

After Illumina paired-end sequencing, the ddRADSeq analysis of 25 peanut genotypes yielded 3.69 Gb of data and 70.3 million paired-end reads. Following the cleaning of the raw data, a total of 65.8 million reads were obtained (Figure 1). The average number of reads recorded was 2.63 M with GC content of 37.6%. Genotype ICG5344 had the highest number of reads at 3.8 million, whereas genotype ICG14335 had the lowest at 1.6 million.

A total of 62,728 InDels were identified across 20 chromosomes in the peanut genome (Table 2). The most InDels were found on chromosome Arahy.08, accounting for 9.08% of the total, while the lowest was observed on chromosome Arahy.07, accounting for only 3.2%. Notably, the number of deletions (35,140) surpassed the number of insertions (27,588), indicating a bias towards deletion events in the peanut genome. Specifically, chromosome Arahy.03 exhibited the highest number of deletions while chromosome Arahy.08 had the most insertions, suggesting regional genomic variability in InDel formation (Table 2). Among all InDels identified, the majority were 1 bp in size, constituting 80.13% of the total InDels. Following this, indels of 2 bp constituted 10.33% of the total, while indels of 3 bp represented 3.31%. Furthermore, indels of 4 bp accounted for 1.39% and indels of 5 bp for 0.85%. The remaining portion, consisting of InDels of ≥6 bp in size, accounted for 3.96% of the total InDels generated (Table S1).

Our study focused on larger InDels (≥10 bp) to establish PCR-based markers. These more considerable genetic variations are interesting due to their potential impact on gene function and regulatory regions. A total of 1013 InDels, constituting 1.61% of all reads, were identified (Figure 2). Notably, chromosome Arahy.03 displayed the highest frequency of deletions and insertions, totaling 74 InDels. This suggests a region of potential genomic instability or active evolution. In contrast, chromosomes Arahy.07 and Arahy.08 exhibited comparatively lower quantities of InDels, with 38 and 23 InDels, respectively (Figure 2).

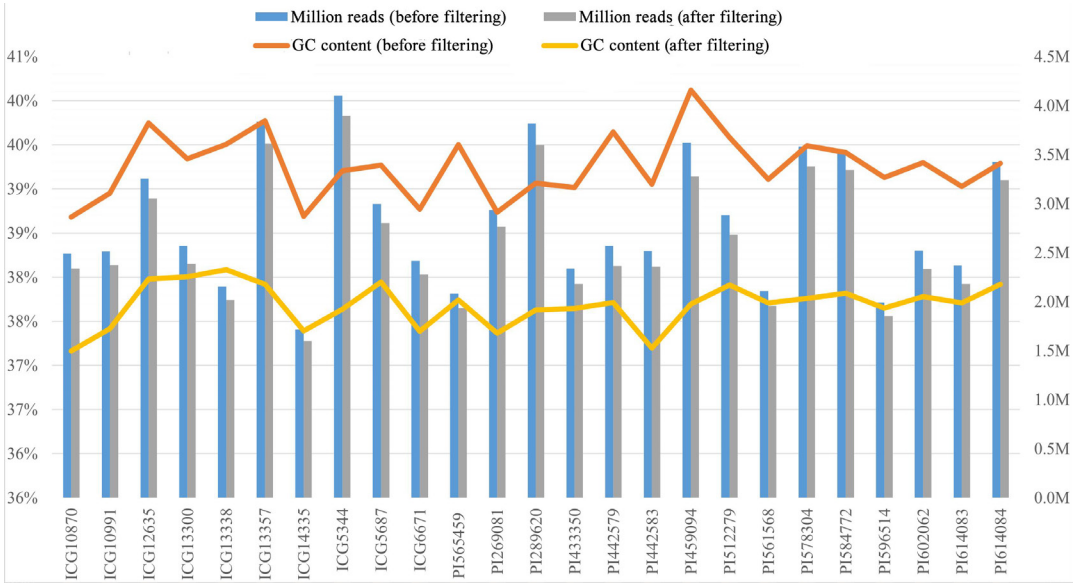

Figure 1. Total number of reads and GC content (%) per accession.

Table 2. Distribution and numbers of insertions and deletions (size of ≥1 bp) in the peanut genome.

Reference genome chromosome	Chromosome	Insertion-deletion	Deletion	Insertion
NC_037618.1	Arahy.01	3507	1811	1696
NC_037619.1	Arahy.02	2828	1571	1257
NC_037620.1	Arahy.03	4313	2382	1931
NC_037621.1	Arahy.04	4763	2185	2578
NC_037622.1	Arahy.05	2654	1599	1055
NC_037623.1	Arahy.06	2862	1635	1227
NC_037624.1	Arahy.07	2018	1219	799
NC_037625.1	Arahy.08	5696	2204	3492
NC_037626.1	Arahy.09	2468	1521	947
NC_037627.1	Arahy.10	3195	1799	1396
NC_037628.1	Arahy.11	2738	1630	1108
NC_037629.1	Arahy.12	2594	1600	994
NC_037630.1	Arahy.13	3123	1893	1230
NC_037631.1	Arahy.14	2852	1709	1143
NC_037632.1	Arahy.15	2907	1816	1091
NC_037633.1	Arahy.16	3087	1849	1238
NC_037634.1	Arahy.17	2642	1600	1042
NC_037635.1	Arahy.18	2927	1714	1213
NC_037636.1	Arahy.19	2972	1832	1140
NC_037637.1	Arahy.20	2582	1571	1011
TOTAL		62728	35140	27588

A total of 21 InDel regions were identified and effectively utilized for primer development using Primer3 software. These InDels were predominantly sourced from chromosomes Arahy.02, Arahy.03, Arahy.04, Arahy.05, Arahy.06, Arahy.09, Arahy.10, Arahy.12, Arahy.15, Arahy.16, Arahy.17, Arahy.18, and Arahy.19, showcasing a wide distribution across the peanut genome (Table 3). The developed markers, ranging from 150 to 400 bp in size, resulted in successful amplification and were found to be highly polymorphic, producing the expected band sizes on agarose gels (Figure S1). The genomic positioning of these InDels revealed that approximately 80.9% were located in intergenic regions, which were

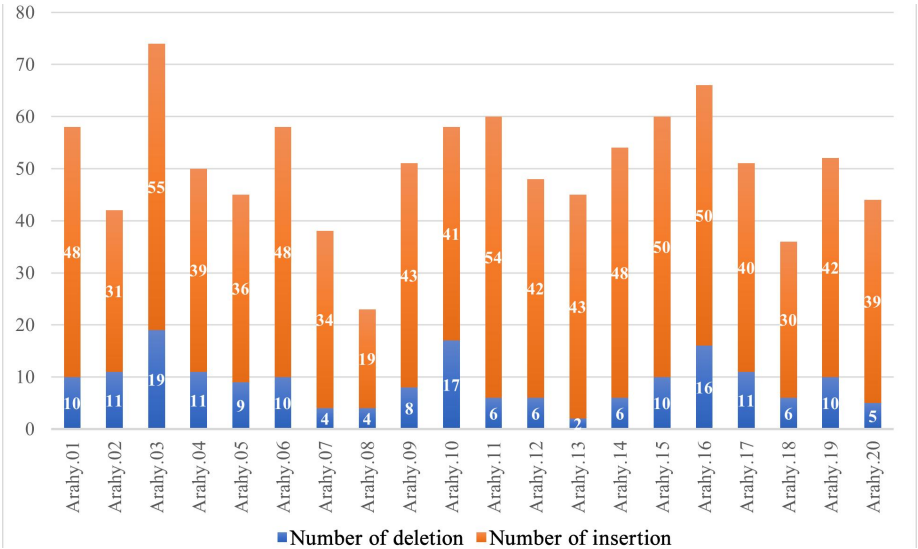

Figure 2. Distribution of insertions and deletions with size of ≥10 bp in the genome.

typically noncoding areas of the genome but could influence gene regulation and expression. Furthermore, our analysis identified three regions classified as exon/CDS, directly involved in gene coding sequences. This highlights the potential for these markers to affect protein function directly. Another region was associated with mRNA, suggesting its role in gene expression regulation (Table 3).

The polymorphic information content (PIC) values for these 21 markers varied from 0 to 0.371, with an average value of 0.163, indicating a moderate level of genetic diversity within our marker set. The average number of alleles (Na) observed ranged from 1.0 to 1.75, and the expected number of alleles (Ne) varied from 1.0 to 1.44. These metrics demonstrate the breadth of allelic variation captured by our markers, which is critical for assessing genetic diversity and structure within populations. The highest expected heterozygosity (He), a measure of genetic variability within a population, was recorded for marker G-I-A03-132, reaching 0.267, suggesting relatively higher genetic diversity at this locus. In contrast, the lowest values were recorded for G-I-A06-260 and G-I-A10-124, both at 0.00, indicating no heterozygosity and suggesting that these loci are fixed within the population. The average Shannon diversity index (I), which quantifies the entropy or diversity represented by these markers, ranged from 0.00 to 0.404, further underscoring the variability among the markers studied (Table 4). PCoA based on the genotypes derived from these markers revealed three distinct genetic clusters within the population (Figure 3).

4. Discussion

The utilization of next-generation sequencing technologies such as ddRADSeq marks a significant transformation in the accessibility of genetic markers and their application in identifying genes associated with desirable traits in crop species, including peanuts (Pandey et al., 2016). The ddRADSeq method, which employs two restriction enzymes to reduce genome complexity (Peterson et al., 2012), has proven especially effective in our study, generating extensive marker data from 25 peanut genotypes and facilitating the detection of 62,728 InDel sites. This approach has demonstrated its cost-effectiveness and capability to enhance our understanding of crop genetics through the detailed mapping of SNP, SSR, and InDel markers (Hasan et al., 2021). Our ddRADSeq analysis yielded a considerable data volume of 3.69 Gb and 65.8 million quality-filtered paired-end reads, underscoring the efficiency of this technology in capturing a broad representation of the peanut genome. The observed variance in read numbers across genotypes might be attributed to differences in genomic complexity or DNA quality (Weissensteiner et al., 2020), with genotype ICG5344 having the highest and genotype ICG14335 the lowest. Such disparities highlight the sensitivity of ddRADSeq to genomic nuances, which is crucial for accurately assessing genetic diversity and structure within populations.

The identification of a pronounced bias toward deletions over insertions across the peanut genome is particularly telling, suggesting selection pressures that may favor genome compactness for metabolic efficiency or genomic stability (Peterson and Arick, 2019). This pattern is critical as deletions can lead to loss-of-function mutations, impacting gene expression and potentially leading to significant phenotypic variations essential for adaptation (Olsen and Wendel, 2013). Moreover, the high frequency of InDels on chromosome Arahy.08

Table 3. Primer sequences for the 21 InDel markers developed and utilized in this study.

Marker name	Chromosome	Position	Sequence information	InDel type	Forward primer (5′ to 3′)	Reverse primer (5′ to 3′)	Product size (bp)	Locus location
G-D-A02-920	Arahy.02	92098153	TGTCTCTCGTACT	Deletion	CCAACCATGGGG TTTAGAGA	CGGAAGTGGAT TTTTCTGTCA	170	Intergenic region
G-D-A04-419	Arahy.04	41939059	TTTATCCAAGATATCAC	Deletion	AAGGTGTGGAGCT CTGCTGT	CACTAGAGCCT TGATCGCTTG	201	Intergenic region
G-D-A05-220	Arahy.05	22082915	TCGGTTGTCAGGC GAGCAC	Deletion	GAATCCAGCGATT CTTTTGC	GCCATGCCTAG ACCTTGTTC	372	Intergenic region
G-D-A06-927	Arahy.06	92721401	TGGTTTTCCAACT TGGGG	Deletion	AAAAGGCTCACC TACACCACA	CGTCGTCATTG TGTTGGTTC	163	Intergenic region
G-D-A09-113	Arahy.09	113167343	TCCAAATTAA	Deletion	GAACGGCGTTAC TTTGCACT	AGCAAGTACCC CCAATTCCT	238	Intergenic region
G-D-A10-113	Arahy.10	113243482	CACTCAGAGA	Deletion	CTCGAAGTTGGA GTGAGATGC	AACGTTGAACC ACACAACCA	183	Intergenic region
G-D-A15-136	Arahy.15	136997172	CTTCTGGCTTAACT GGATGT	Deletion	CCATCTTGCAGAAAG TTGAGAA	AAAGATCCCAT TGCCAGAAG	182	EXON/CDS
G-D-A15-928	Arahy.15	9282925	TGCTTCCCTTCC AGT	Deletion	GGCCAATTCCAT GGTGTACT	AAATGAAACAC ATGGGAGGTG	163	Intergenic region
G-D-A16-139	Arahy.16	139347571	CCTTAGCACTGT GCA	Deletion	TCACTCATCATT GTGTGT	GTTGCCGTTTT TGGACATTT	212	CDS*
G-D-A18-127	Arahy.18	127015562	TGCGTGCGTTCG CACACA	Deletion	GTAGAG CACGTGCAGGG ACCT GTG	CCAGCCCTGAC CTCTCTGTA	237	Intergenic region
G-D-A19-113	Arahy.19	113592941	CGGTTGAACCC	Deletion	GAACAACTAGA TAGTGATCAAGCATT	TTAAAAGCAAG GTTGCGAAAA	170	Intergenic region
G-D-A19-140	Arahy.19	140059586	GATTTGGTCTTAT	Deletion	ACGCCAAACAA CAAAACGA	GGTCGAAAAAC CACTAGTGTCC	177	Intergenic region
G-I-A02-535	Arahy.02	53582891	TTTCAAAAATTT	Insertion	TCACAATTTCGT CCACCAAG	TTCGCAAATAA AGTTTGGAAAAA	160	Intergenic region
G-I-A03-132	Arahy.03	132892945	GGTTCTATTCG	Insertion	TTTCGGTTTTG GAACGAGAC	AAGCAAACAA AATACTGAGAATCTAAT	152	Intergenic region
G-I-A06-260	Arahy.06	26011143	GACCTTGTTCTTA	Insertion	GGACCTTTTGC TATTGTATTTTCA	ACAGCAGAGC TCCACACCTT	181	Intergenic region
G-I-A10-124	Arahy.10	12462960	AGAAGAAGGTG GAAAAGA TAACGA	Insertion	GGGGGAAGAAG AAGAAGAAGG	CGCAGAAAGTT CCTCTACCG	155	Intergenic region
G-I-A12-755	Arahy.12	75579146	TAAGTAAAAGA GAAACTAAGT	Insertion	ACATCACACAAT GGCAAAGG	GGTTTTGCTTC TTCCATTCC	232	Intergenic region
G-I-A15-148	Arahy.15	14862692	ATGCCTTAGT TTTC	Insertion	TCTGTGCAGCAT CTCAGCTC	TCCACCTACTT ATAATTGCTGTCA	190	EXON
G-I-A15-983	Arahy.15	98318479	GCGTGCGCG CGAGGTGC	Insertion	CCGGTGTGGAT TTTCCTAAT	AAGCATGCACA AGAGGAAGG	154	Intergenic region
G-I-A17-827	Arahy.17	82793935	AGTAAAAGA GAAACTAAGTA	Insertion	CCATGGAGAAA TAAAGAAAAAGGA	CCGGTAGCACT ACACTTCTCC	236	Intergenic region
G-I-A19-375	Arahy.19	3759976	TCTCTATAG GCAATG	Insertion	TCATTAATATTT TCAATGGTGTGAGA	GTCACCGGTG GATTGGTAAG	160	gene/mRNA

* Coding DNA sequence (CDS)

Table 4. Summary of genetic diversity statistics for selected markers.

Marker / locus	Na	Ne	I	He	uHe	PIC
G-I-A02-535	1.750	1.307	0.318	0.199	0.225	0.189
G-I-A03-132	1.750	1.428	0.404	0.267	0.297	0.298
G-I-A06-260	1.000	1.000	0.000	0.000	0.000	0.000
G-I-A10-124	1.000	1.000	0.000	0.000	0.000	0.000
G-I-A12-755	1.750	1.307	0.318	0.199	0.225	0.233
G-I-A15-148	1.250	1.045	0.072	0.038	0.040	0.074
G-I-A15-983	1.250	1.062	0.087	0.049	0.052	0.074
G-I-A17-827	1.250	1.150	0.141	0.094	0.098	0.189
G-I-A19-375	1.500	1.332	0.292	0.198	0.207	0.298
G-D-A02-920	1.500	1.158	0.200	0.119	0.125	0.189
G-D-A04-419	1.500	1.107	0.159	0.088	0.092	0.136
G-D-A05-220	1.500	1.436	0.329	0.233	0.244	0.371
G-D-A06-927	1.500	1.158	0.200	0.119	0.125	0.189
G-D-A09-113	1.500	1.436	0.329	0.233	0.244	0.341
G-D-A10-113	1.500	1.107	0.159	0.088	0.092	0.136
G-D-A15-136	1.250	1.045	0.072	0.038	0.040	0.074
G-D-A15-928	1.250	1.045	0.072	0.038	0.040	0.074
G-D-A16-139	1.500	1.444	0.331	0.235	0.247	0.341
G-D-A18-127	1.250	1.062	0.087	0.049	0.052	0.074
G-D-A19-140	1.250	1.062	0.087	0.049	0.052	0.074
G-D-A19-113	1.250	1.045	0.072	0.038	0.040	0.074
Mean	1.392	1.177	0.177	0.112	0.120	0.163

* Number of alleles (Na), effective number of alleles (Ne), Shannon diversity index (I), expected heterozygosity (He), Unbiased expected heterozygosity (uHe) polymorphic information content (PIC).

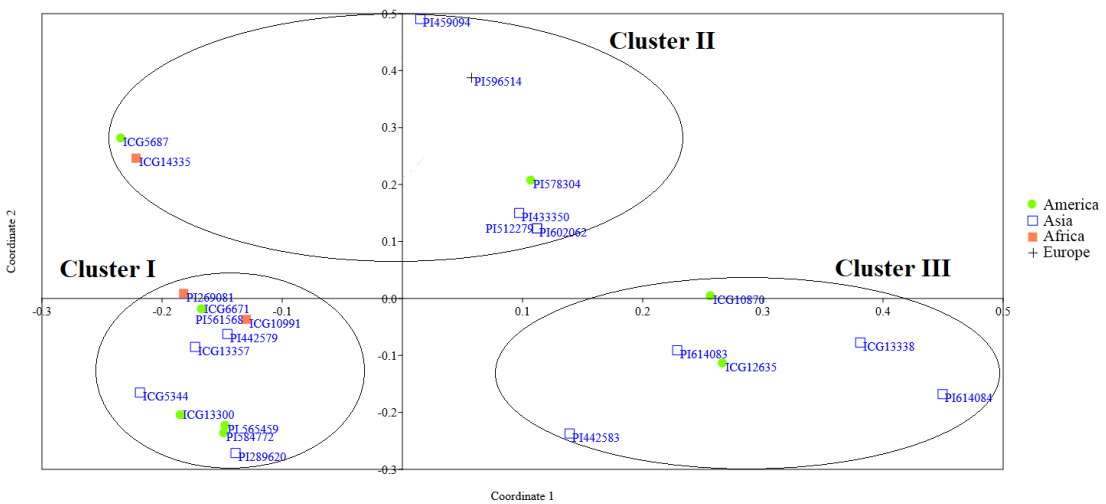

Figure 3. Principal coordinate analysis (PCoA) results for the 25 peanut accessions genotyped using 21 InDel markers.

compared to the lowest on Arahy.07 suggests differential genomic evolution or selection pressures across the peanut genome, which could influence gene functions and be linked to adaptations to various environmental conditions or resistance to specific pathogens (Cuc et al., 2008). Furthermore, the predominance of 1-bp InDels and the progression in the size distribution of these genetic variations highlight the complexity of mutational mechanisms. Smaller InDels are likely due to simple slippage or point mutations during DNA replication

(Vaughn and Bennetzen, 2014). In comparison, larger InDels may result from more complex recombination events or the activity of transposable elements within the genome (Bennetzen and Wang, 2014). These findings are consistent with those observed in other plant genomes, suggesting that the mechanisms of InDel formation are conserved across species but may have distinct functional implications depending on the genomic context and environmental interactions (Wang et al., 2021).

Our study revealed a negative correlation between the length of base pairs and the abundance of InDels, a pattern also observed in other studies across various crops, including rice (Liu et al., 2015b), mango (Cortaga et al., 2022), and radish (Li et al., 2023b). Consistently, the maximum numbers of InDels were identified as single-nucleotide InDels. This trend, whereby the highest ratio of single- and binucleotide InDels has been noted in crops like maize (Batley et al., 2003), *Brassica rapa* (Liu et al., 2013), tea (Liu et al., 2019), and chickpea (Jain et al., 2019), underscores the fundamental genetic mechanisms that favor the formation of smaller InDels. These findings align with known patterns of the formation of small InDels across various plant genomes, where single nucleotide polymorphisms often occur more frequently than more extensive mutations due to the simplicity of their mutational processes. The higher frequency of deletions over insertions could be associated with the evolutionary pressures on the peanut genome, potentially reflecting mechanisms that favor genome compactness or stability. This detailed analysis of InDel distribution provides crucial insights into the genetic architecture of the peanut genome and highlights potential areas for further genetic study and crop improvement strategies. These slight genetic variations, typically resulting from replication errors or minor recombination events, are more prevalent due to the molecular dynamics of DNA replication and repair processes (Hao et al., 2023). Chen and Zhang (2015) further observed that single-nucleotide InDels are predominant during DNA replication, which can disrupt the reading frames of genes, leading to frameshift mutations that alter amino acid sequences and significantly impact protein functionality. In contrast, our focus on larger InDels (≥10 bp) aimed to establish user-friendly PCR-based markers and provided crucial insights into the structural variations significantly influencing the genetic architecture of peanut. The identification of 1013 InDels across the peanut genome, with a particularly high concentration on chromosome Arahy.03, highlights this region as a potential hotspot for genomic activity. This concentration may be indicative of adaptive evolutionary processes or localized genomic stress factors, as suggested by Pandey et al. (2016). This pattern of variability might reflect differing levels of evolutionary pressure or functional constraints, which are evident from the varied InDel frequencies across different chromosomes (Chen et al., 2020). The larger size of these InDels has substantial biological implications, as they are likely to disrupt functional domains within proteins or regulatory elements that control gene expression, potentially leading to phenotypic variations crucial for survival and adaptation (Graham et al., 2000). One significant advantage of InDels over SNPs is their variable length, providing additional flexibility in marker development and applications (Liu et al., 2023). Moreover, PCR-based InDels provide benefits such as codominance, cost-efficiency, and significant polymorphism, enabling their simple identification using gel electrophoresis (Jain et al., 2019). In our study, 21 InDel markers that can be resolved on agarose gel were developed and successful amplifications were obtained. A success rate of 100% in PCR assays was demonstrated by the efficacy of the ddRADSeq library approach and the InDel filtering pipeline. Moreover, the most significant observed deletions and insertions, measuring 26 and 24 bp, respectively, highlight the potential for significant structural changes within the genome. These modifications can introduce new or altered gene functions, enhancing the plant's adaptability to environmental challenges or impacting agronomically essential traits (Pourkheirandish et al., 2020). Therefore, our findings underscore the necessity of further explorations into regions enriched with large InDels, which could provide deeper insights into how these genomic variations influence phenotypic traits. This comprehensive analysis enhances our understanding of the dynamics within the peanut genome. It lays a robust foundation for applying genomic information to improve breeding strategies, ultimately developing peanut varieties with superior performance and resilience.

Our comprehensive study has significantly advanced our understanding of the genomic positioning and functional implications of InDels within the peanut genome. We observed that InDels were most abundant in intergenic regions, consistent with findings from other species such as radish (Li et al., 2023b) and sesame (Kizil et al., 2020), where similar distributions were noted using transcriptome data and ddRADSeq analysis, respectively. Traditionally considered noncoding, these regions are increasingly recognized for their crucial roles in regulating gene expression and influencing plant physiology and adaptation. This is particularly pertinent in peanuts, where approximately 80.9% of the identified InDels are located in intergenic areas, underscoring their potential beyond mere structural genomic elements. Moreover, the observed coding sequence (CDS) regions accounted for only 9.5% of the InDels, attributed to the higher degree of conservation in these regions than others within the genome (Liu et al., 2019). This conservation is crucial as

InDels within CDS regions can profoundly impact protein structure and function, often more significantly than single base alterations. For example, an 11-bp deletion in chickpea's early flowering-three gene (*ELF3*) has been effectively utilized as an InDel marker (Ridge et al., 2017), highlighting the utility of such variations in genomics-led breeding applications.

Our study's utilization of 21 InDel regions for primer development marks a significant advancement in peanut genomic research and highlights the genetic diversity inherent within the peanut genome. These markers' successful amplification and high polymorphism confirm their utility in capturing genetic variability, which is crucial for effective breeding programs. Furthermore, the presence of InDels in exon/CDS regions and regions associated with mRNA transcription presents significant implications for altering amino acid sequences and potentially producing novel protein variants (Li et al., 2017). Such genetic modifications can lead to new phenotypes that may offer adaptive advantages under specific environmental conditions, such as enhanced drought tolerance or disease resistance (Tan, 2021). By leveraging the genetic markers developed in this study, future research can more effectively link phenotypic traits with their genetic bases, paving the way for more precise genetic interventions to boost agricultural output and sustainability. This comprehensive analysis not only enhances our understanding of the peanut genome's structural diversity but also underscores the potential of these genetic variations in improving crop resilience and productivity through targeted genetic modifications (Majid et al., 2017).

Our study reveals a diverse spectrum of PIC values ranging from 0 to 0.371 across 21 genetic markers, with an average value of 0.163. This indicates a moderate level of genetic diversity within our marker set, which is essential for effective population genetic studies and breeding programs. Interestingly, the highest PIC value observed at marker locus G-D-A16-139, located in the CDS region, emphasizing its potential for significant genetic variation and utility in genetic studies or breeding applications. The distribution of these InDels, predominantly found in intergenic regions, highlights their potential role beyond mere structural genomic elements. Intergenic regions, traditionally considered noncoding, are increasingly recognized for their influence on gene regulation and expression, which can profoundly affect plant physiology and adaptation. Moreover, the use of PCoA on these markers revealed three distinct classes, suggesting a lack of definitive geographical origin patterns, which could be attributed to the migration of various genotypes by individuals and interregional trade spanning several centuries (Basak et al., 2019). The variations in PIC values, coupled with the differences in the number of alleles (Na) from 1.0 to 1.75 and the expected number of alleles (Ne) from 1.0 to 1.44, substantiate the genetic variability accessible within the peanut genome. This allelic variation is crucial for assessing the genetic diversity and structure within peanut populations, enabling the identification of genetic bottlenecks or extensive outcrossing events that might have occurred in the species' evolutionary history. Marker G-I-A03-132 exhibited the highest expected heterozygosity (He) at 0.267, suggesting a locus potentially under less selective pressure or more recently evolved, maintaining a variety of alleles that could be advantageous under varying environmental conditions. In contrast, the absence of heterozygosity at loci G-I-A06-260 and G-I-A10-124, fixed within the population, might result from intense selective pressures favoring specific alleles over others, possibly due to their adaptive advantages. These findings underscore the importance of further exploring the regions enriched with large InDels and using markers located within or near genes associated with desirable traits such as disease resistance, drought tolerance, or increased yield. By leveraging the genetic markers developed in this study, future research can more effectively link phenotypic traits with their genetic bases, paving the way for more precise genetic interventions to boost agricultural output and sustainability.

5. Conclusion

This study has provided pivotal insights into genetic diversity and structural variations within the peanut genome through the advanced application of ddRADSeq technology. We identified 62,728 InDel regions across all chromosomes, demonstrating the complex nature of the peanut genome and its evolutionary dynamics. A broad spectrum of PIC values among 21 InDel markers was observed, ranging from 0 to 0.371, with an average of 0.163. This moderate genetic diversity is crucial for effectively implementing breeding programs and population genetic studies. The discovery of InDels predominantly in intergenic regions and significant occurrences within coding sequences underscores their potential regulatory roles and contributions to functional research.

Notably, this study constitutes the first report of the development of InDel markers in peanuts using ddRADSeq technology. The genetic markers developed in this study will be valuable tools for targeted breeding strategies aiming to combine favorable traits from diverse genetic backgrounds to improve crop resilience and productivity. Specifically, the highest PIC value observed at specific loci indicates potential hotspots for genetic variation, which is instrumental in identifying genes associated with important agronomic traits. Moreover, PCoA revealed distinct genetic clusters within the population, emphasizing the intricate relationship between genetic

makeup and geographical dispersion, likely influenced by human activities over centuries.

InDels, the second most abundant structural variation in the genome after SNPs, offer significant opportunities for labs with small to medium genotyping facilities, especially those focusing on longer polymorphisms. Our findings reveal a diverse distribution of InDels, with a notable abundance in intergenic regions suggesting potential regulatory roles. Monitoring InDels in CDS and exon regions has also significantly contributed to genomics-driven breeding endeavors. Future studies should focus on linking these genetic variations with phenotypic outcomes and exploring the potential of these markers in other crops to enhance global food security. By leveraging detailed genetic markers and understanding their distribution across the genome, we can pave the way for more precise and efficient genetic interventions, ultimately leading to the development of peanut varieties with improved performance and resilience. This study advances our knowledge of the peanut genome and sets a precedent for using genomic information to enhance breeding strategies, marking a significant step forward in agricultural biotechnology.

Acknowledgments

This study was supported by the Scientific Research Projects Coordination Unit of Akdeniz University with Project FYL-2022-5981. We are grateful to the USDA and ICRISAT for supplying genetic material several times.

References

Basak M, Uzun B, Yol E (2019). Genetic diversity and population structure of the Mediterranean sesame core collection with use of genome-wide SNPs developed by double digest RAD-Seq. PLoS One 14 (10): e0223757. https://doi.org/10.1371/journal.pone.0223757

Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D (2003). Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiology 132 (1): 84-91. https://doi.org/10.1104/pp.102.019422

Bennetzen JL, Wang H (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annual Review of Plant Biology 65: 505-530. https://doi.org/10.1146/annurev-arplant-050213-035811

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nature Genetics 48 (4): 438-446. https://doi.org/10.1038/ng.3517

Bhat RS, Shirasawa K, Chavadi SD (2022). Genome-wide structural and functional features of single nucleotide polymorphisms revealed from the whole genome resequencing of 179 accessions of *Arachis*. Physiologia Plantarum 174 (1): e13623. https://doi.org/10.1111/ppl.13623

Bhat RS, Shirasawa K, Gangurde SS, Rashmi MG, Sahana K et al. (2023). Genome-wide landscapes of genes and repeatome reveal the genomic differences between the two subspecies of peanut (*Arachis hypogaea*). Crop Design 2 (1): 100029. https://doi.org/10.1016/j.cropd.2023.100029

Britten RJ, Rowen L, Williams J, Cameron RA (2003). Majority of divergence between closely related DNA samples is due to indels. Proceedings of the National Academy of Sciences 100 (8): 4661-4665. https://doi.org/10.1073/pnas.0330964100

Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM et al. (2020). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. Nature Genetics 52 (5): 525-533. https://doi.org/10.1038/s41588-020-0614-5

Chen W, Zhang L (2015). The pattern of DNA cleavage intensity around indels. Scientific Reports 5 (1): 8333. https://doi.org/10.1038/srep08333

Cortaga CQ, Lachica JAP, Lantican DV, Ocampo ETM (2022). Genome-wide SNP and InDel analysis of three Philippine mango species inferred from whole-genome sequencing. Journal of Genetic Engineering and Biotechnology 20 (1): 46. https://doi.org/10.1186/s43141-022-00326-3

Cuc LM, Mace ES, Crouch JH, Quang VD, Long TD et al. (2008). Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). BMC Plant Biology 8: 1-11. https://doi.org/10.1186/1471-2229-8-55

El-Akhal MR, RincónA, Coba de la Peña T, Lucas MM, El Mourabit N et al. (2013). Effects of salt stress and rhizobial inoculation on growth and nitrogen fixation of three peanut cultivars. Plant Biology 15 (2): 415-421. https://doi.org/10.1111/j.1438-8677.2012.00634.x

Gao Q, Yue G, Li W, Wang J, Xu J et al. (2012). Recent progress using high-throughput sequencing technologies in plant molecular breeding. Journal of Integrative Plant Biology 54 (4): 215-227. https://doi.org/10.1111/j.1744-7909.2012.01115.x

Graham SW, Reeves PA, Burns AC, Olmstead RG (2000). Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. International Journal of Plant Sciences 161 (S6): S83-S96.

Hake AA, Shirasawa K, Yadawad A, Sukruth M, Patil M et al. (2017). Mapping of important taxonomic and productivity traits using genic and non-genic transposable element markers in peanut (*Arachis hypogaea* L.). PLoS One 12 (10): e0186113. https://doi.org/10.1371/journal.pone.0186113

Hamidou F, Halilou O, Vadez V (2013). Assessment of groundnut under combined heat and drought stress. Journal of Agronomy and Crop Science 199 (1): 1-11. https://doi.org/10.1111/j.1439-037X.2012.00518.x

Hao Q, Zhan C, Lian C, Luo S, Cao W et al. (2023). DNA repair mechanisms that promote insertion-deletion events during immunoglobulin gene diversification. Science Immunology 8 (81): eade1167. https://doi.org/10.1126/sciimmunol.ade1167

Hasan N, Choudhary S, Naaz N, Sharma N, Laskar RA (2021). Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. Journal of Genetic Engineering and Biotechnology 19 (1): 128. https://doi.org/10.1186/s43141-021-00231-1

He G, Prakash CS (1997). Identification of polymorphic DNA markers in cultivated peanut (*Arachis hypogaea* L.). Euphytica 97: 143-149. https://doi.org/10.1023/A:1002949813052

Huang L, Liu X, Pandey MK, Ren X, Chen H et al. (2020). Genome-wide expression quantitative trait locus analysis in a recombinant inbred line population for trait dissection in peanut. Plant Biotechnology Journal 18 (3): 779-790. https://doi.org/10.1111/pbi.13246

Hwang JY, Wang YT, Shyu YS, Wu JSB (2008). Antimutagenic and antiproliferative effects of roasted and defatted peanut dregs on human leukemic U937 and HL-60 cells. Phytotherapy Research 22 (3): 286-290. https://doi.org/10.1002/ptr.2306

Jain A, Roorkiwal M, Kale S, Garg V, Yadala R et al. (2019). InDel markers: an extended marker resource for molecular breeding in chickpea. PLoS One 14 (3): e0213999. https://doi.org/10.1371/journal.pone.0213999

Kassa MT, Yeboah SO, Bezabih M (2009). Profiling peanut (*Arachis hypogea* L.) accessions and cultivars for oleic acid and yield in Botswana. Euphytica 167: 293-301. https://doi.org/10.1007/s10681-008-9852-9

Kizil S, Basak M, Guden B, Tosun HS, Uzun B et al. (2020). Genome-wide discovery of InDel markers in sesame (*Sesamum indicum* L.) using ddRADSeq. Plants 9 (10): 1262. https://doi.org/10.3390/plants9101262

Krapovickas A, Gregory WC (1994). Taxonomia del genero "arachis (leguminosae)." Bonplandia 8 (1/4): 1-186. (in Spanish)

Li W, Zhu Z, Chern M, Yin J, Yang C et al. (2017). A natural allele of a transcription factor in rice confers broad-spectrum blast resistance. Cell 170 (1): 114-126. https://doi.org/10.1016/j.cell.2017.06.008

Li W, Huang L, Liu N, Chen Y, Guo J et al. (2023a). Identification of a stable major sucrose-related QTL and diagnostic marker for flavor improvement in peanut. Theoretical and Applied Genetics 136 (4): 78. https://doi.org/10.1007/s00122-023-04306-0

Li Y, Luo X, Peng X, Jin Y, Tan H et al. (2023b). Development of SNP and InDel markers by genome resequencing and transcriptome sequencing in radish (*Raphanus sativus* L.). BMC Genomics 24 (1): 445. https://doi.org/10.1186/s12864-023-09528-6

Liang X, Chen X, Hong Y, Liu H, Zhou G et al. (2009). Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and Arachis wild species. BMC Plant Biology 9: 1-9. https://doi.org/10.1186/1471-2229-9-35

Liu B, Wang Y, Zhai W, Deng J, Wang H et al. (2013). Development of InDel markers for Brassica rapa based on whole-genome re-sequencing. Theoretical and Applied Genetics 126: 231-239. https://doi.org/10.1007/s00122-012-1976-6

Liu J, Li J, Qu J, Yan S (2015b). Development of genome-wide insertion and deletion polymorphism markers from next-generation sequencing data in rice. Rice 8 (1): 27. https://doi.org/10.1186/s12284-015-0063-4

Liu L, Dang PM, Chen CY (2015a). Development and utilization of InDel markers to identify peanut (*Arachis hypogaea*) disease resistance. Frontiers in Plant Science 6: 988. https://doi.org/10.3389/fpls.2015.00988

Liu S, An Y, Tong W, Qin X, Samarina L et al. (2019). Characterization of genome-wide genetic variations between two varieties of tea plant (*Camellia sinensis*) and development of InDel markers for genetic research. BMC Genomics 20: 1-16. https://doi.org/10.1186/s12864-019-6347-0

Liu Z, Zhao Y, Zhang Y, Xu L, Zhou L et al. (2023). Development of Omni InDel and supporting database for maize. Frontiers in Plant Science 14: 1216505. https://doi.org/10.3389/fpls.2023.1216505

Lu C, Zou C, Zhang Y, Yu D, Cheng H et al. (2015). Development of chromosome-specific markers with high polymorphism for allotetraploid cotton based on genome-wide characterization of simple sequence repeats in diploid cottons (*Gossypium arboreum* L. and *Gossypium raimondii* Ulbrich). BMC Genomics 16: 55. https://doi.org/10.1186/s12864-015-1265-2

Mahmood S, Li Z, Yue X, Wang B, Chen J et al. (2016). Development of INDELs markers in oilseed rape (*Brassica napus* L.) using re-sequencing data. Molecular Breeding 36: 1-13. https://doi.org/10.1007/s11032-016-0501-z

Majid A, Parray G, Wani SH, Kordostami M, Sofi N et al. (2017). Genome editing and its necessity in agriculture. International Journal of Current Microbiology and Applied Sciences 6 :5435-5443. https://doi.org/10.20546/ijcmas.2017.611.520

Mishra GP, Radhakrishnan T, Kumar A, Thirumalaisamy PP, Kumar N et al. (2015). Advancements in molecular marker development and their applications in the management of biotic stresses in peanuts. Crop Protection 77: 74-86. http://dx.doi.org/10.1016/j.cropro.2015.07.019

Mozingo RW, O'keefe SF, Sanders TH, Hendrix KW (2004). Improving shelf life of roasted and salted inshell peanuts using high oleic fatty acid chemistry. Peanut Science 31 (1): 40-45. https://doi.org/10.3146/pnut.31.1.0009

Olsen KM, Wendel JF (2013). Crop plants as models for understanding plant adaptation and diversification. Frontiers in Plant Science 4: 290. https://doi.org/10.3389/fpls.2013.00290

Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimarães P et al. (2012). Advances in *Arachis* genomics for peanut improvement. Biotechnology Advances 30 (3): 639-651. https://doi.org/10.1016/j.biotechadv.2011.11.001

Pandey MK, Roorkiwal M, Singh VK, Ramalingam A, Kudapa H et al. (2016). Emerging genomic tools for legume breeding: current status and future prospects. Frontiers in Plant Science 7: 179244. https://doi.org/10.3389/fpls.2016.00455

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS One 7 (5): e37135. https://doi.org/10.1371/journal.pone.0037135

Peterson DG, Arick M (2019). Sequencing plant genomes. Progress in Botany Vol 80: 109-193.

Pourkheirandish M, Golicz AA, Bhalla PL, Singh MB (2020). Global role of crop genomics in the face of climate change. Frontiers in Plant Science 11: 522800. https://doi.org/10.3389/fpls.2020.00922

Rathod V, Hamid R, Tomar RS, Padhiyar S, Kheni J et al. (2020). Peanut (*Arachis hypogaea*) transcriptome revealed the molecular interactions of the defense mechanism in response to early leaf spot fungi (*Cercospora arachidicola*). Plant Gene 23: 100243. https://doi.org/10.1016/j.plgene.2020.100243

Raza A, Chen H, Zhang C, Zhuang Y, Sharif Y et al. (2024). Designing future peanut: the power of genomics-assisted breeding. Theoretical and Applied Genetics 137 (3): 66.

Ridge S, Deokar A, Lee R, Daba K, Macknight RC et al. (2017). The chickpea *Early Flowering 1* (Efl1) locus is an ortholog of Arabidopsis *ELF3*. Plant Physiology 175 (2): 802-815. https://doi.org/10.1104/pp.17.00082

Roorkiwal M, Jain A, Thudi M, Varshney RK (2017). Advances in chickpea genomic resources for accelerating the crop improvement. The Chickpea Genome 53-67.

Samoluk SS, Vaio M, Ortíz AM, Chalup LM, Robledo G et al. (2022). Comparative repeatome analysis reveals new evidence on genome evolution in wild diploid *Arachis* (Fabaceae) species. Planta 256 (3): 50. https://doi.org/10.1007/s00425-022-03961-9

Sarkar T, Thankappan R, Kumar A, Mishra GP, Dobaria JR (2014). Heterologous expression of the *AtDREB1A* gene in transgenic peanut-conferred tolerance to drought and salinity stresses. PLoS One 9 (12): e110507. https://doi.org/10.1371/journal.pone.0110507

Savage GP, Keenan JI (1994). The composition and nutritive value of groundnut kernels. In: The Groundnut Crop: A scientific basis for improvement. Dordrecht: Springer Netherlands, pp. 173-213. http://dx.doi.org/10.1007/978-94-011-0733-4_6

Shen D, Liu B, Qiu Y, Zhang XH, Zhang ZH (2013). Development and application of cucumber InDel markers based on genome re-sequencing. Journal of Plant Genetic Resources 14 (2): 278-283. (in Chinese with an abstract in English)

Singh S, Sangh C, Kona P, Bera SK (2024). Genome editing in peanuts: advancements, challenges and applications. The Nucleus 67 (1): 127-139. https://doi.org/10.1007/s13237-024-00482-6

Suchoszek-Łukaniuk K, Jaromin A, Korycińska M, Kozubek A (2011). Health benefits of peanut (*Arachis hypogaea* L.) seeds and peanut oil consumption. In Nuts and seeds in health and disease prevention. Academic Press., pp. 873-880. https://doi.org/10.1016/B978-0-12-375688-6.10103-3

Taha MG, Yousef HM, EL-Behery SA, Mostafa H (2019). Fatty acids and chemical composition of peanut (*Arachis hypogaea* L.). Al-Azhar Journal of Agricultural Research 44 (1): 119-127. https://doi.org/10.21608/ajar.2019.59735

Tan X (2021). Comprehensive analysis of resistance genes, structural variation, and recent genome rearrangements in plant genomes. PhD, University of Georgia, Athens, Georgia, USA.

Varshney RK (2016). Exciting journey of 10 years from genomes to fields and markets: some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. Plant Science 242: 98-107. https://doi.org/10.1016/j.plantsci.2015.09.009

Vaughn JN, Bennetzen JL (2014). Natural insertions in rice commonly form tandem duplications indicative of patch-mediated double-strand break induction and repair. Proceedings of the National Academy of Sciences 111 (18): 6684-6689. https://doi.org/10.1073/pnas.1321854111

Wang J, Yan C, Shi D, Zhao X, Yuan C et al. (2021). The genetic base for peanut height-related traits revealed by a meta-analysis. Plants 10 (6): 1058. https://doi.org/10.3390/plants10061058

Weissensteiner MH, Bunikis I, Catalán A, Francoijs K-J, Knief U et al. (2020). Discovery and population genomics of structural variation in a songbird genus. Nature Communications 11 (1): 3403. https://doi.org/10.1038/s41467-020-17195-4

Yol E, Ustun R, Golukcu M, Uzun B (2017). Oil content, oil yield and fatty acid profile of groundnut germplasm in Mediterranean climates. Journal of the American Oil Chemists' Society 94 (6): 787-804. https://doi.org/10.1007/s11746-017-2981-3

Young CT, Waller GR (1972). Rapid oleic/linoleic microanalytical procedure for peanuts. Journal of Agricultural and Food Chemistry 20 (6): 1116-1118.

Zhou X, Xia Y, Ren X, Chen Y, Huang L et al. (2014). Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genomics 15: 1-14. https://doi.org/10.1186/1471-2164-15-351

Ziaei Jam H, Zook JM, Javadzadeh S, Park J, Sehgal A et al. (2024). LongTR: genome-wide profiling of genetic variation at tandem repeats from long reads. Genome Biology 25 (1): 176. https://doi.org/10.1186/s13059-024-03319-2

Supplementler materials

Table S1. Distribution of InDels across chromosomes according to length of base pair.

Chromosome	1 bp		2 bp		3 bp		4 bp		5 bp		6 bp		7 bp		8 bp		9 bp		≥10 bp	
	DEL*	INS	DEL	INS	DEL	INS	DEL	INS	DEL	INS	DEL	INS	DEL	INS	DEL	INS	DEL	INS	DEL	INS
Arahy.01	1540	1246	151	236	56	47	24	35	10	18	6	22	5	16	5	14	4	14	10	48
Arahy.02	1367	925	133	176	31	35	20	22	1	14	3	21	1	8	1	8	3	17	11	31
Arahy.03	2027	1405	211	271	78	70	23	29	6	24	9	26	3	13	5	24	1	14	19	55
Arahy.04	1848	1954	199	377	80	101	17	32	8	16	8	21	7	12	5	13	2	13	11	39
Arahy.05	1370	753	124	136	54	37	21	25	7	26	8	13	2	8	4	10	0	11	9	36
Arahy.06	1407	917	148	122	34	39	14	29	8	19	5	15	3	12	5	12	1	14	10	48
Arahy.07	1066	548	92	100	36	31	5	18	8	16	0	19	6	9	2	11	0	13	4	34
Arahy.08	1853	2685	240	534	64	116	27	42	9	28	0	19	3	22	0	15	4	12	4	19
Arahy.09	1289	643	138	116	55	38	13	26	2	18	4	21	4	11	6	18	2	13	8	43
Arahy.10	1541	1013	157	173	56	64	9	20	9	22	2	15	3	22	3	11	2	15	17	41
Arahy.11	1395	765	147	128	44	42	17	32	10	19	5	20	2	21	3	12	1	15	6	54
Arahy.12	1376	731	139	101	54	31	14	16	6	24	4	13	0	10	0	13	1	13	6	42
Arahy.13	1626	887	167	139	56	56	21	37	5	15	5	20	6	9	5	12	0	12	2	43
Arahy.14	1476	785	148	137	55	43	11	30	4	21	4	21	1	25	2	18	2	15	6	48
Arahy.15	1587	782	129	111	53	45	27	18	2	24	4	27	1	13	2	11	1	10	10	50
Arahy.16	1606	892	142	130	49	49	15	19	9	28	8	20	2	21	0	9	2	20	16	50
Arahy.17	1372	738	135	130	51	44	12	24	5	17	5	19	6	14	2	12	1	4	11	40
Arahy.18	1496	910	136	126	49	38	14	21	5	22	4	24	0	17	4	12	0	13	6	30
Arahy.19	1586	790	150	112	59	58	16	31	5	23	1	29	2	21	0	21	3	13	10	42
Arahy.20	1356	715	126	116	39	40	27	23	7	18	4	16	2	15	2	10	3	19	5	39
Genome	30184	20084	3012	3471	1053	1024	347	529	126	412	89	401	59	299	56	266	33	270	181	832

*DEL: deletion, INS: insertion

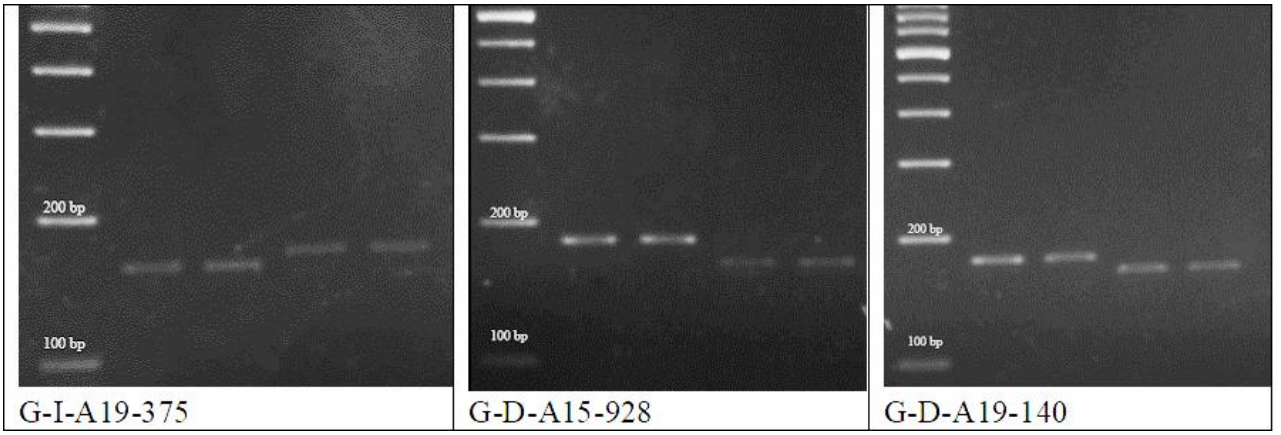

Figure S1. Amplification of peanut DNAs with use of selected markers (Ladder 100 bp).