BMC Genomics

RESEARCH Open Access

Genome wide association study uncovers significant SNPs and candidate genes for kernel grades in groundnut (*Arachis hypogaea* L.)

Ashutosh Purohit[1,2], Anitha Raman[1], Devraj Lenka[2], Tharanya Murugesan[1], Safinaaz Kounain[1], Sunita Choudhary[1] and Janila Pasupuleti[1*]

Abstract

Background Kernel grade is a key market trait that significantly influences the market price of groundnut and is directly proportional to pod yield. A set of 574 multi-parent advanced generation inter-cross (MAGIC) lines including parents and checks was assessed for kernel grade, yield, and component traits in a partially replicated (p-rep) design for two seasons (rainy and post-rainy; 2021-22). A genome-wide association study (GWAS) was conducted to identify marker-trait associations (MTAs) and potential candidate genes for kernel grades and yield.

Results MAGIC lines ICGR 171238 (79.45%) and ICGR 171206 (85.65%) showed highest percent net recovery of grade-I kernel (PNR_I) in rainy and post-rainy seasons respectively. Twenty-four high confidence SNPs were associated with kernel-grades and yield traits across 11 chromosomes. SNPs AX_147226917 (A07) and AX_177643480 (B08) showed consistent association with PNR_I and counts per ounce of grade-I kernel (CPO_I) across seasons. Key candidate genes for kernel grades include *Aradu.6Z78F* (*RING-H2 finger protein*), *Aradu.993Q7* (*ascorbate peroxidase 1*), *Araip.MKV8R* (*protein FAR1-RELATED SEQUENCE 3-like isoform X1*) and *Aradu.S3AS8* (*Vacuolar protein-sorting protein BRO1*). For yield traits, *Aradu.Y7AIG* (*cytochrome P450*), *Aradu.BD60N* (*Glucose-1-phosphate adenylyltransferase*) and *Aradu.TW8M6* (*LEA protein*) were identified.

Conclusion Predominantly these genes are known to regulate growth and development, control cell functions, confer disease resistance and stress tolerance, and influence pod size in groundnuts. The validation studies of the identified MTAs would facilitate the development of molecular markers for use in marker-assisted selection (MAS), enabling efficient selection of progenies with higher percent recovery of grade-I kernels in the segregating populations.

Keywords Kernel grades, Yield, Groundnut, GWAS, MTAs

*Correspondence:
Janila Pasupuleti
janila.pasupuleti@icrisat.org
[1]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad 502324, India
[2]Department of Plant Breeding and Genetics, College of Agriculture, Odisha University of Agriculture and Technology (OUAT), Odisha 751003 Bhubaneswar, India

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Background

Groundnut (*Arachis hypogaea* L.), a self-pollinated annual legume crop widely grown in the arid and semi-arid tropics of Asia and Africa, is highly valued for its nutritional composition, comprising 44–56% oil, 22–30% protein, and 10–25% carbohydrates [1, 2]. Groundnut cultivation spans a global area of 30.53 million hectares, yielding a total production of 54.23 million tons with an average productivity of 1776.2 kg/ha. India, China, Nigeria, and Sudan have the largest cultivated areas (5.70, 4.45, 3.40, and 3 Mn ha, respectively) (FAOSTAT, 2023). Africa produced 17.36 Mn Tn of groundnut, whereas Asia produced 31.70 Mn tns. Together, Asia and Africa contribute to 90% of global groundnut production. Groundnut kernel is composed of high-quality protein, which is increasingly used as a concentrated protein ingredient in food formulations, meat analogues, bakery products, protein bars and animal feed [3]. Considering the increasing demand for plant-based protein-rich foods, it will provide opportunities of higher exports for major groundnut producing countries.

There is a shift in the demand for groundnuts in the confectionery industry, particularly in developed nations, compared to the historical demand for oil extraction due to increased usage of roasted nuts, peanut butter, and different groundnut-based snack products [4]. This is evident from a growth of the confectionery market from $206.97 billion in 2023 to a projected $278.36 billion by 2032 (www.cgiar.org). The characteristics that are desirable for confectionery purposes include a higher percentage of sound mature kernels (>80% SMK), a creamy smooth texture, a rich nutty flavor, a 100 seed weight of more than 55 g, sugar content above 5%, protein content exceeding 30%, blanchability over 60%, and an oil content below 45% [5, 6]. The mass of one hundred seeds is a crucial factor in determining confectionery quality. Studies at ICRISAT have demonstrated that large-seeded groundnut genotypes, such as ICGV 90212 and ICGV 97051, are particularly suitable for confectionery use because of their high kernel yield and superior seed mass [7, 8].

According to the UNECE (United Nations Economic Commission for Europe) standards, peanut kernels are graded by size using the counts per ounce method, which measures the number of kernels in 28.35 g (1 oz). Grade-I kernel in peanuts, with 40–60 kernels per ounce in the global market, are preferred by the confectionery industry owing to their superior physical and sensory qualities (www.unece.org). These kernels are large, uniform in size, free of defects, and possess desirable textures and flavor profiles. These physical and chemical parameters make them ideal for value-added products such as roasted nuts and peanut butter as they evolve during roasting [9]. In addition, international markets place a premium on high-quality grade-I kernels because of their application to snacks and products. Groundnut exports for confectionery often follow strict grading standards, emphasizing the need for well-sorted visually appealing kernels (https://www.cbi.eu). Considering the increasing importance of grade-I kernel in the industry, they have been incorporated into breeding program in market segment II at ICRISAT, which focuses on developing product profiles for the confectionary industry. The kernel grade is a complex quantitative trait that is affected by the environment. To cater to the market demand for confectionery groundnuts and considering the importance of large-seeded grade-I kernels, it is important to combine yield with a high proportion of grade-I kernel recovery to develop suitable market-driven groundnut cultivars.

Grade-I kernel with widths greater than 7.5 mm [10] are in high demand in the confectionary industry. Yield improvement is always a key breeding objective in most crop improvement programs, including groundnuts. To develop suitable market-preferred groundnut cultivars, it is important that the cultivars have high yields with superior kernel grades and a crucial understanding of the molecular mechanisms governing this trait to realize their maximum potential. Since the inception of the idea of generating and utilizing multi-parent populations, such as the multi-parent advanced generation inter-cross (MAGIC) in crop improvement, several MAGIC populations have been generated in different crops, including groundnuts, for genomic dissection of complex traits [11–13]. Groundnut MAGIC populations have been used to dissect and characterize complex traits, such as drought tolerance [14], late leaf spot [15], pod-size related traits [16], pod weight, seed weight, shelling percentage, pod constriction, and pod reticulation [17]. A larger number of recombination events in MAGIC populations provide an opportunity to map genomic regions with a higher resolution.

Recent advancements in groundnut genomics and cost-effective sequencing have enhanced our understanding of complex traits. Reference genomes for cultivated tetraploid groundnut are available [18–20]. New sequencing methods like genotyping-by-sequencing [21] and "*Axiom_Arachis*" SNP arrays [22] have reduced costs, enabling detailed genetic mapping [23]. Association studies in multiparent populations help dissect complex traits with high resolution. Studies have identified genomic regions associated with pod/kernel yield and traits like hundred-seed mass and shelling percentage. Thirty significant markers explaining 11.22–32.30% phenotypic variation were found to be associated with seed-related traits through an association analysis by utilizing 104 peanut accessions and SSR markers [24]. A nested association mapping (NAM) population genotyped with 58 K SNP array revealed genomic regions linked to seed and

pod weights [25]. SSR and SNP array-based maps identified major regions on chromosome B06 and A07/B07 for pod and seed related traits [26]. QTL on chromosome A05 showed major effects on seed size in US mini-core collection [27]. Two major loci on chromosome A06 and A02 [28] and another two on chromosome A08 and B06 [29] have been identified for yield-related traits.

However, no QTL mapping or association mapping studies have identified genomic regions associated with the high recovery of grade-I kernels in groundnuts. Considering the need to develop confectionary purposes high yielding groundnut cultivars with high recovery of grade I kernels and the unavailability of genomic resources to augment the breeding pipeline, we have conducted a genome-wide association studies (GWAS) utilizing 574 MAGIC lines and a high density "48k SNP array" to identify markers linked to yield attributing traits and kernel grades.

Methods

Genotypes and experimental design

A MAGIC population consisting of five hundred and seventy-four lines that includes eight founder parents and 12 checks, was evaluated [30] (Supplementary Table S1). The founder parents viz., used to develop MAGIC population were ICGV 91114, ICGV 06040, ICGV 00440, ICGV 00308, ICGV 05155, ICGV 88145, GPBD 4 and 55–437. Among these, ICGV 00440 is a large-seeded, high-yielding variety with a hundred seed weight of 75 g. The trials were laid out in a partially replicated design (p-rep) consisting of 722 plots indexed by 19 rows and 38 columns ordered as columns within rows. 30% of the test lines (148) and checks were replicated twice. While reducing the total number of experimental plots and optimizing resources, the replicated plots increased. Each plot consisted of four rows of 4 m each, with a row-to-row spacing of 30 cm and a plant-to-plant spacing of 10 cm, grown on a broad bed. The experiments were conducted during two seasons: rainy 2021 and post rainy 2021-22 in alfisols at ICRISAT, Patancheru ($17^0$53'N, $78^0$27'E, and 545 m asl), India. The recommended agronomic management practices such as timely irrigation immediately after planting and as required thereafter, the application of gypsum at peak flowering, measures to protect against insect pests and diseases were implemented to ensure the cultivation of a healthy crop.

Traits measured

Yield traits included pod weight per plot (PW), kernel weight per plot (KW), hundred kernel weight (HKW), and shelling percentage (SHP). Kernel grade traits comprised percent net recovery of grade-I kernel (PNR-I), hundred kernel weights of grade-I kernel (HKW-I), and counts per ounce of grade-I kernel (CPO-I). The PW from each plot was recorded after drying the pods using a measuring balance. The SHP and HKW values were predicted using CT scans [31] and the details on the procedure and calibration were described later in this section. KW was determined using the PW and SHP data. The PNR-I, HKW-I, and CPO-I were manually determined.

Kernel grading

Kernel grading is a replica of industrial seed grading, in which seeds are sorted based on commercially defined screens. 500 gms of sound mature kernels (SMK) were passed through 3 sieves that were of 8.5 mm, 7.5 mm and 6.5 mm consecutively (Fig. 1). This method divided the entire material into three grades. Kernel weight, hundred kernel weight, and counts per ounce for each grade were measured using a weighing balance. The percent net recovery of each grade was calculated using the formula: $\%NR = \frac{Kernel\ weight\ of\ individual\ grade}{Total\ weight\ (500gm)} \times 100$

Digital trait value prediction of hundred kernel weight and shelling percentage using computed tomography

The methodology for a non-destructive analysis of seed traits that determines the quality of groundnut traits has been comprehensively described in a previous study [31]. The main traits of interest, hundred kernel weight and shelling percentage, were predicted as part of a pipeline that first determined kernel weight and shell weight in preliminary steps using X-ray image transformation (XRT). A 50 g peanut pod sample, representative of each plot harvest (total 722 plots, each with a 4 m×4 m plot size), was scanned to extract 2D X-ray image features, which were then processed by the XRT model. This model demonstrated high predictive accuracy, achieving a kernel weight coefficient of determination (R^2) of 0.93 for kernel weight and 0.78 for shell weight, with mean absolute errors of 0.17 and 0.08, respectively. These predictions were then used to calculate the kernel size and shelling percentage. These predicted values were subsequently used to calculate the kernel weight and the shelling percentage. This robust, statistically validated methodology provides a non-destructive, high-throughput solution for crop evaluation and quality control, making it highly effective for breeding programs and postharvest assessments.

Single trial analysis

The model for the partially replicated design is described as follows.

$$y_{ijk} = \mu + \alpha_i + r_j + c_k + e_{ijk}$$

Where.

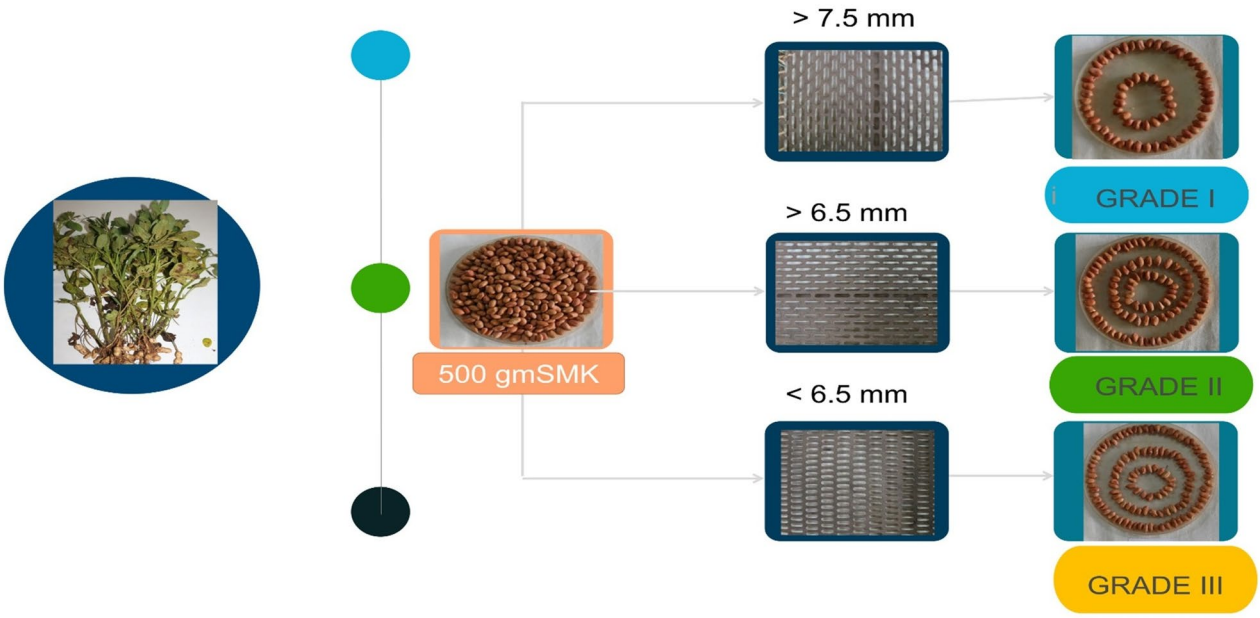


Fig. 1 Schematic diagram illustrating the methodology followed for assessing kernel grading in groundnut

α_i is the variety main effect considered "random," r_j and c_k (global trend) are the design factors, which are random terms, for rows and columns.

e_{ijk} (local + extraneous) is the residual corresponding to the observational units, which are the plots.

The statistical model is thus given by,

The row and column effects describe the extraneous variations that usually arise from the experimental procedures. Plot-to-plot variability can be split into global and local trends. Global trends arise from uneven soil moisture, soil depth, and other natural variations. Local trends or nuggets are small-scale spatial variations within the field, an indicator of how noisy the spatial structure is. Thus, the error structure is.

$$\mathrm{R} = \sigma^2 \left(\sum_{\mathrm{R}} \theta \sum_{\mathrm{C}}\right) + \sigma_\epsilon^2 \mathrm{I}$$

$$\sigma^2 \left[(\mathrm{SR} \,\Theta\, \mathrm{SC}) + \frac{\sigma_\epsilon^2}{\sigma^2} \mathrm{I}\right]$$

where S_R and S_C indicate the row and column correlation matrices, respectively, as the combination of row and column factors represents unique positions in the spatial grid. A separable autoregressive model of order 1 (AR1) was fitted for S_R and S_C. and σ_ϵ^2 is the nugget variance.

Genotyping

DNA was extracted from 574 MAGIC lines including eight parent plants using a NucleoSpin 96 Plant II Kit from Machery Nagel in Germany. The DNA amount was measured with a Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, MA, USA). The quality of the DNA was checked on a 0.8% agarose gel. A 48 K Affymetrix SNP array ('Axiom_Arachis 2.0') was used to genotype the extracted DNA samples. The output files (Cell Intensity File) from the Affymetrix instrument were analyzed in Axiom Analysis Suite (AAS) v 5.2 (Thermo Fisher Scientific, Inc.). The Axiom Analysis Suite integrates SNP genotyping, indel detection, multi-allele analysis, off-target variant (OTVs) calling, and copy number detection into a graphical interface. The 48 K SNP array data for the MAGIC population is provided in Supplementary Table S2.

Filtering of genotypic data and genome-wide association study (GWAS)

In Axiom Analysis Suite (AAS, Thermo Fisher Scientific) genotyping analysis was executed. AAS followed best practices workflow, where it runs genotyping algorithms, allow to view cluster graphs and export of data. Best practices workflow controls the quality with a dish QC value ≥ 0.82 and QC call rate of ≥ 97%. All markers were visually verified to inspect the quality of the cluster pattern. After filtering out 47,837 SNPs based on a missing rate exceeding 10%, a minor allele frequency (MAF) below 0.05, and heterozygosity greater than 0.3, a total of 13,937 high-quality single nucleotide polymorphisms (SNPs) were retained and used for further GWAS analysis. Previously we have calculated the LD decay for the same population and it was 2.02 Mb [15]. This information was used in the current study to support genome-wide association study (GWAS) analysis. 13,937 high-quality filtered SNPs and best linear unbiased predictor (BLUP) values of kernel grades and yield

component traits, such as PW, KW, HKW, SHP, PNR-I, HKW-I, and CPO_I were used for GWAS analysis separately for the rainy and post-rainy seasons. The BLUP values for above mentioned traits were estimated using ASREML-R package [32] in R software. GWAS analysis was performed using Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model of "GAPIT 3.0" (Genomic Association and Prediction Integrated Tool) package [33] in R v.4.1.2 software [34] because of its superiority in computing speed, high statistical power, and fewer false positives in the identification of significant MTAs. A Bonferroni-corrected threshold of 3.58757E-06 was set to reduce Type I and II errors. QQ plots, Manhattan plots, and association tables assessed the results, to detect high confidence MTAs.

Identification and gene ontology (GO) analysis of the potential candidate genes

Due to the extensive linkage disequilibrium (LD) and the computational challenges involved in identifying candidate genes within the LD region of significant SNPs, a 50 kb range both upstream and downstream of significant SNPs (100 kb window) was utilized as a confidence interval to locate candidate genes linked to the traits of interest. Data on candidate genes within SNPs' confidence interval was sourced from the Peanut Base (https://peanutbase.org/home), using the genome of the diploid ancestors of cultivated peanut, *A. duranensis* and *A. ipaensis*. Gene Ontology (GO) enrichment analysis was performed using the PlantRegMap platform (http://plantregmap.gao-lab.org) which implements a Gene Set Enrichment Analysis (GSEA)-based statistical framework. The list of candidate genes identified from PeanutBase was used as input for *Arachis duranensis* and *Arachis ipaensis*. The analysis categorized significantly enriched GO terms into three functional domains; biological process (BP), molecular function (MF), and cellular component (CC) based on hypergeometric testing and false discovery rate (FDR) correction. Enriched GO terms were further visualized and interpreted to identify the predominant biological processes and molecular functions associated with the candidate genes.

In-silico expression profiling of candidate genes and allele distribution pattern

Expression profiles of the candidate genes were analyzed *in-silico* using publicly available transcriptome datasets of *Arachis hypogaea*. The normalized expression data (FPKM >1) across 20 developmental tissues and stages were retrieved from the PeanutBase expression atlas and compiled into an expression matrix [35]. Each gene's expression pattern was compared across tissues such as cotyledon, embryo, flower, leaves, nodules, roots, seeds, and stems. Data processing was performed using Microsoft Excel and R (v4.x) for normalization and visualization. Genes with the highest mean expression values were considered *top-ranked candidates*. The expression data were visualized using heatmaps to determine tissue- or stage-specific expression trends. This computational workflow enabled the identification of genes with distinct or constitutive expression profiles without the need for experimental validation, thereby providing an efficient *in-silico* strategy for expression analysis. Fold-change was computed using Root_veg as the reference tissue, following the formula:

Fold Change = (Expression in target tissue/Expression in Root_veg).

Zero values were adjusted by adding 1 to avoid division by zero.

Favorable and unfavorable alleles were identified for significant SNPs associated with high and low yield-attributing traits and kernel grades.

Results

BLUP estimates and assessment of yield contributing traits in the MAGIC population

PW, KW, and HKW(g) were higher during the post-rainy season, whereas HKW (GRADE I) and counts per ounce (C GRADE I) had overlapping distributions (Fig. 2, Supplementary Table S3). The row and column effects were not significant for any trait. The autocorrelations were positive in both directions for yield traits HKW and SHP, indicating that trait values in plots are influenced by neighboring plots. However, PNR-I showed a significant negative autocorrelation in the row direction during the post-rainy season, indicating dissimilar neighboring plots. The MAGIC population exhibited highly significant genetic variation ($p < 0.01$) for all traits across both seasons (Tabs. 1 and 2). Among the genotypes replicated twice, ICGR 171175 and ICGR 171497 performed best for pod and kernel yields in the rainy and post-rainy seasons. ICGR 171058 and ICGR 171259 excelled in the shelling percentage in both seasons, with ICGR 171379 being the best in the post-rainy season. ICGR 171582 and ICGR 171532 had the highest HKW in the post-rainy and rainy seasons, whereas ICGR 171497 maintained good HKW across seasons (Tab. 3). Among the genotypes tested once, ICGR 171070 had the highest pod and kernel yields in the post-rainy season, whereas ICGR 171280 and ICGR 171437 had the highest pod and kernel yields in the rainy season. GG 20 performed well for both traits across all seasons. ICGR 171437 and ICGR 171005 were the best for shelling % in the rainy and post-rainy seasons. ICGV00440 ranked high for HKW in both seasons, with ICGR 171238 and ICGR 171588 performing best in the rainy and post-rainy seasons, respectively (Supplementary Table S4) .

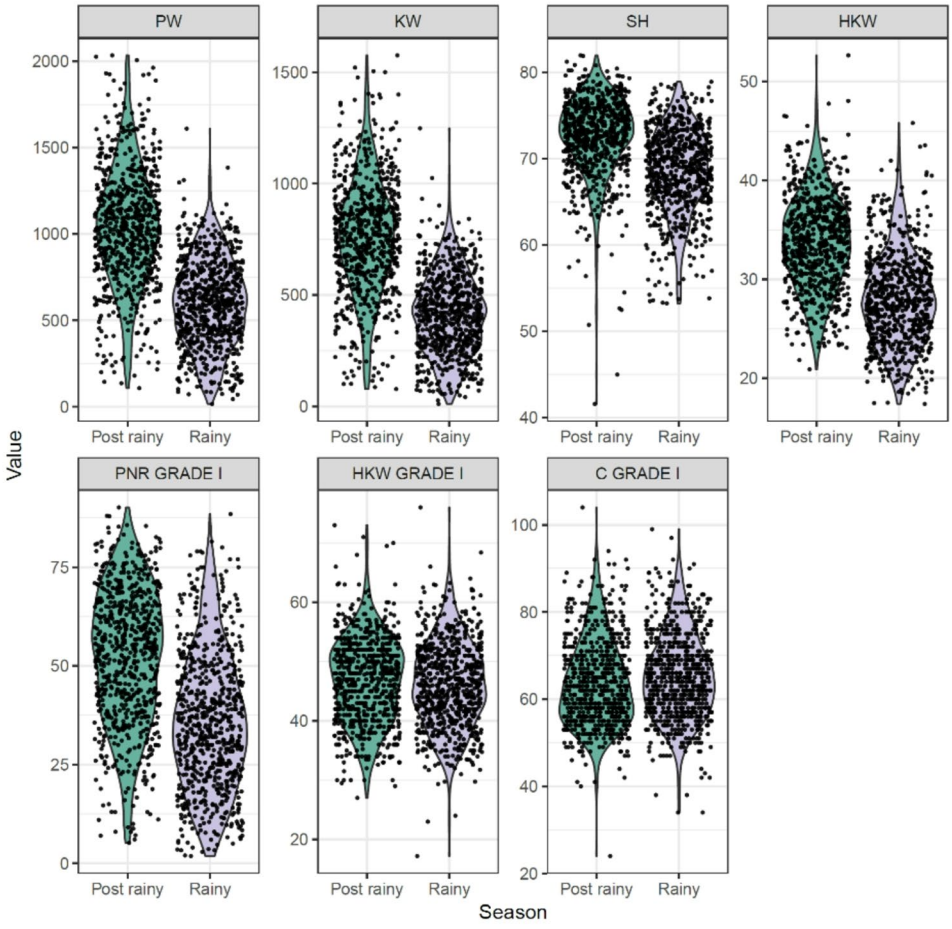


Fig. 2 Variability in pod weight (PW; kg/ha), kernel weight (KW; kg/ha), shelling percentage (SH; %), hundred-kernel weight (HKW; g), percent net recovery of grade-I kernel (PNR GRADE I; %), HKW of grade I kernels (HKW GRADE I; g), and counts per ounce of grade I kernels (C Grade I) in the MAGIC population during the rainy season of 2021 and post-rainy season of 2021–22

BLUP estimates and assessment of kernel grade contributing traits in the MAGIC population

The PNR-I kernel (kernel width > 7.5 mm) was better in the post rainy season than in the rainy season, while the HKW and CPO of grade I kernels (40–60 kernels/ounce) were comparable during both seasons. Among the replicated genotypes, ICGR 171433 (70.66%) and ICGR 171206 (85.65%) recorded the highest PNR-I in the rainy and post-rainy seasons, respectively. ICGR 171476 (57.28 gm) and M 335 (59.06 gm) recorded the highest HKW-I during the rainy and post-rainy seasons, respectively. Conversely, ICGR 171260 (51) and M 335 (48) exhibited the lowest CPO-I value in the rainy and post-rainy seasons, respectively. Among the genotypes replicated once, ICGR 171,238 (79.45) and ICGR 171,576 (84.31) recorded the maximum PNR-I kernel in the rainy and post-rainy seasons, respectively. Additionally, genotypes ICGR 171196 (64.27) and ICGR 171430 (62.34) recorded the highest HKW-I during the rainy and post-rainy seasons, respectively. Meanwhile, ICGR 171268 (41) and ICGR 171519 (41) registered the lowest CPO-I value in the rainy and post-rainy seasons, respectively.

Table 3 and Supplementary Table S4 provide detailed information on the performance of the top twenty MLs, replicated twice and once, including checks for pod yield and net recovery of grade-I kernels, respectively during rainy and post rainy season. ICGR 171410, a Spanish bunch type ML, shows excellent yield potential with 2033.88 kg/ha during Rainy (R) and 1296.91 kg/ha during post rainy (PR), and high recovery of grade-I kernels at 66.84% PR and 63.96% R during both seasons. ICGR 171044 demonstrated good yield with 2003 kg/ha PR and 1291.27 kg/ha R, and a good shelling percentage of 78.63% during PR and 72.05% during R. ICGR 171376 recorded 2345.18 kg/ha during PR and 1283.60 kg/ha R, with bold seed size and high HKW of 40.85 gm PR and 32.96 gm R across seasons. ICGR 171260 shows high recovery of grade-I kernels at 70.68% PR and 52.18% R, with HKW of 37.96 gm PR and 31.83 gm R. ICGR 171576 records a grade-I kernel recovery of 84.31% PR

Table 1 Likelihood ratio test (LRT) for yield and kernel grade component traits during the rainy season of 2021

Trait-Season	Variance components	Row	Column	Nugget	Genotype	Row autocorrelation	Column autocorrelation	Error
PW-R	Estimate	0.0037	899.37	0.059	23387.36	−0.059	0.052	36970.24
	p-LRT		0.1269		0.000000069***	0.5558	0.30038	
	Model Action	Boundary	Dropped	Absent	Retained	Swapped	Swapped	
KW-R	Estimate	0.0013	524.14	0.030	12032.41	−0.038	0.064	18791.46
	p-LRT		0.1067		0.000000071***	0.870	0.220	
	Model Action	Boundary	Dropped	Absent	Retained	Swapped	Swapped	
HKW-R	Estimate	0.0000006	0.0000003	6.929	10.196	0.840	0.840	4.411
	p-LRT				0.00000000002***	0.00000001***	0.000000000002***	
	Model Action	Boundary	Boundary	Absent	Retained	Unswapped	Unswapped	
SHP-R	Estimate	0.244	0.0000026	9.494	10.204	0.773	0.719	3.401
	p-LRT	0.1935			0.0000000004***	0.00006**	0.0000044***	
	Model Action	Dropped	Boundary	Absent	Retained	Unswapped	Unswapped	
PNR-I-R	Estimate	0.0000049	0.0000018	32.612	240.176	0.575	0.729	17.133
	p-LRT				< 2.2e-16 ***	0.0294*	0.0005**	
	Model Action	Boundary	boundary	Absent	Retained	Unswapped	Unswapped	
HKW-I-R	Estimate	0.0000003	0.858	17.81	31.811	0.945	0.980	3.261
	p-LRT		0.0975		0.0000000000000001***	0.0004**		
	Model Action	Boundary	Dropped	Absent	Retained	Unswapped	Fixed, Unswapped	
CPO-I-R	Estimate	0.0000001	0.0053	0.501	1.281	0.973	0.980	0.10
	p-LRT		0.3974		0***	0.0015**		
	Model Action	Boundary	Dropped	Absent	Retained	Unswapped	Fixed, Unswapped	

Retained: model term retained; Absent: model term excluded; Fixed: autocorrelation fixed at boundary (≤ 1); Unswapped: residual structure unchanged due to significant autocorrelation; Swapped: residual structure modified to exclude non-significant autocorrelation

Table 2 Likelihood ratio test (LRT) for yield and kernel grade component traits during the post- rainy season of 2021-22

Trait-Season	Variance components	Row	Column	Nugget	Genotype	Row correlation	Column correlation	Error
PW-PR	Estimate	165.0393	4073.81	27937.22	62913.82	0.935	0.980	36101.41
	p-LRT	0.4698817	0.0083		0.00000000	0.0000		
	Action	Dropped	Retained	Absent	Retained	Unswapped	Fixed, Unswapped	
KW-PR	Estimate	142.5337	1922.652	14176.460	37783.93	0.917	0.964	15514.01
	p-LRT	0.453	0.018*		0***	0***	0***	
	Model Action	Dropped	Retained	Absent	Retained	Unswapped	Unswapped	
HKW-PR	Estimate	0.0000006	0.0000010	4.961	15.727	0.961	0.898	2.396
	p-LRT				< 2.2e-16***	0.0000005***	0.000002***	
	Model Action	Boundary	Boundary	Absent	Retained	Unswapped	Unswapped	
SHP-PR	Estimate	0.020	0.0753634	3.379	15.874	0.874	0.952	3.892
	p-LRT	0.469	0.355		< 2.2e-16***	0.00000003***	0.0000000000007***	
	Model Action	Dropped	Dropped	Absent	Retained	Unswapped	Unswapped	
PNR-I-PR	Estimate	0.0000007	4.954	32.125	250.233	−0.782	0.634	8.896
	p-LRT		0.0095		< 2.2e-16 ***	0.035*	0.3864	
	Model Action	Boundary	Retained	Absent	Retained	Unswapped	Swapped	
HKW-I-PR	Estimate	0.0000025	0.236	0.000009	29.872	−0.070	0.129	20.778
	p-LRT		0.344		0.000000000000006***	0.4166	0.09648	
	Model Action	Boundary	Dropped	Absent	Retained	Swapped		
CPO-I-PR	Estimate	0.0000000	0.0000	0.047	0.258	0.022	0.385	0.072
	p-LRT				0.000	0.8685	0.0431	
	Model Action	Boundary	Boundary	Absent	Retained	Unswapped	Swapped	

Retained: model term retained; Absent: model term excluded; Fixed: autocorrelation fixed at boundary (≤ 1); Unswapped: residual structure unchanged due to significant autocorrelation; Swapped: residual structure modified to exclude non-significant autocorrelation

Table 3 Top twenty groundnut MAGIC lines/checks based on pod weight (PW) and percent net recovery of grade I kernel (PNR_I) tested in two replications

PW (Kg/ha)				PNR_I (%)			
Genotype	**Rainy**	**Genotype**	**post Rainy**	**Genotype**	**Rainy**	**Genotype**	**Post rainy**
ICGR 171586	1381.66	ICGR 171497	2815.24	ICGR 171433	70.66	ICGR 171206	85.65
ICGR 171175	1374.12	ICGV03043	2523.59	ICGR 171073	69.08	ICGR 171223	77.69
ICGV03043	1334.01	ICGV02266	2494.15	M 335	68.35	ICGR 171188	74.81
ICGR 171172	1297.39	ICGR 171138	2398.62	ICGR 171018	65.73	ICGR 171277	74.38
ICGR 171410	1296.91	ICGR 171276	2364.39	ICGR 171172	64.87	ICGR 171476	73.41
ICGR 171228	1291.95	ICGV6040	2359.77	ICGR 171033	64.39	ICGR 171433	71.35
ICGR 171044	1291.27	ICGR 171376	2345.18	ICGR 171410	63.96	ICGR 171254	70.99
ICGR 171376	1283.60	ICGR 171528	2311.79	ICGR 171376	63.53	ICGR 171545	70.90
ICGR 171315	1281.23	ICGR 171477	2194.13	ICGR 171442	62.52	ICGR 171260	70.68
ICGR 171101	1257.87	ICGR 171282	2173.94	ICGV 88145	54.03	ICGR 171278	69.75
ICGR 171087	1243.46	ICGR 171228	2162.06	ICGR 171383	53.67	ICGR 171118	69.51
ICGR 171580	1238.28	ICGR 171582	2148.42	ICGR 171191	53.66	ICGR 171305	69.39
ICGR 171491	1220.37	ICGR 171318	2144.43	ICGR 171068	52.90	ICGR 171563	69.37
ICGR 171427	1220.25	ICGR 171251	2068.10	ICGR 171254	52.72	ICGR 171076	68.56
ICGR 171157	1217.72	ICGR 171540	2061.03	ICGR 171079	52.61	ICGR 171105	68.35
ICGR 171260	1207.54	ICGR 171580	2042.21	ICGR 171260	52.18	ICGR 171379	67.84
ICGR 171379	1188.08	ICGR 171410	2033.88	ICGR 171223	51.84	ICGR 171073	67.76
ICGR 171532	1184.22	ICGR 171044	2003.00	ICGR 171499	48.44	ICGR 171383	67.39
ICGR 171600	1181.65	ICGR 171324	1942.08	ICGR 171315	47.98	ICGR 171018	67.07
ICGR 171276	1176.87	ICGR 171349	1936.25	ICGR 171427	46.50	ICGR 171410	66.84

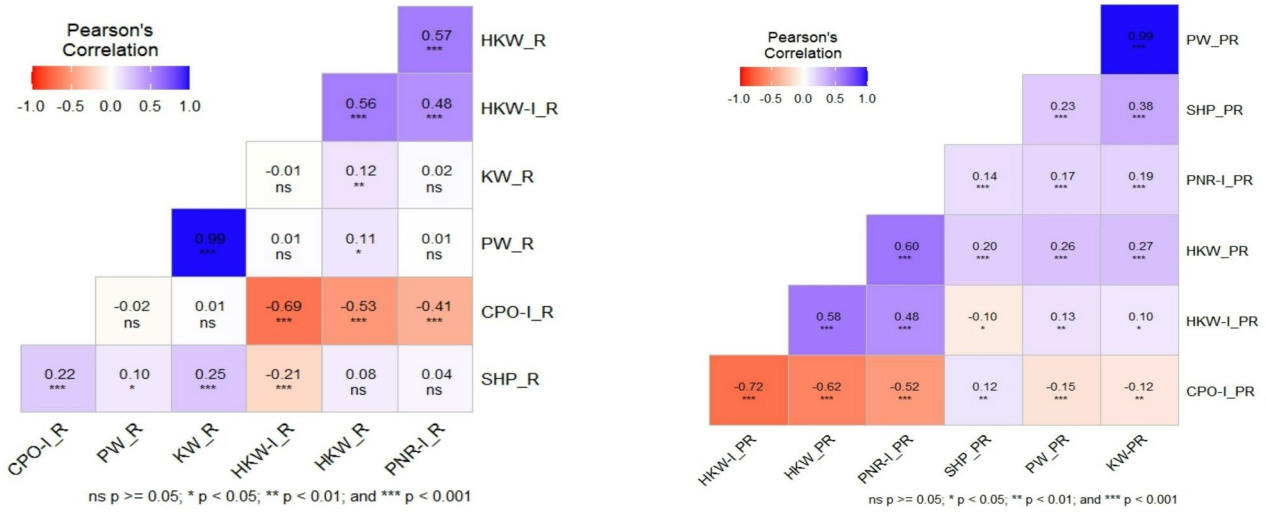


Fig. 3 Pearson correlation matrix showing relationships among yield-attributing traits and kernel grade components in the MAGIC population across seasons

and 72.68% R, with an SHP of 77.20% PR and 72.89% R in both seasons.

Correlation between yield and kernel grade contributing traits

A correlation study was conducted to examine the relationship between yield-contributing traits and kernel-grade traits during the rainy and post-rainy seasons. The percent net recovery of grade-I kernels was positively and significantly associated with HKW and HKW-I during both seasons, whereas the counts per ounce of grade-I kernels (CPO-I) were negatively and significantly associated with PNR-I. PW and KW showed a significant positive association with PNR-I during the post-rainy season but a non-significant association during the rainy season. (Fig. 3).

Genomic regions associated with yield contributing traits

PW, KW, HKW, and SHP are the yield-contributing traits. A total of fifteen SNPs identified to be associated

with yield-contributing traits; out of which twelve were unique, significant SNPs above the Bonferroni corrected threshold of 3.58757E-06. Three SNPs (AX_176822892, AX_176805020, and AX_147234427) were identified for pod yield on chromosomes A01, A04, and A09 during the rainy season, with PVE of 3.37–4.97%, SNP AX_176803444 was detected on A03 during the post-rainy season, with a PVE of 10.88%. Three SNPs associated with kernel yield were identified: AX_176805020 and AX_147234427 on chromosomes A04 and A09 during the rainy season, explaining 7.42 and 7.11% of phenotypic variance, respectively, and AX_176803444 on A03 during the post-rainy season, explaining 12.48% of phenotypic variance. Three SNPs, namely AX_176805020 and AX_147234427, identified during the rainy season, and AX_176803444, identified during the post-rainy season, were found to be common loci associated with both pod and kernel yield traits. Four SNPs were associated with HKW during the rainy season on chromosomes A03, A05, B07, and B08, with a PVE of 2.13–7.90%. Three SNPs were identified for HKW during the post-rainy season on A07 and B08, with PVE of 2.83–12.94%. One significant SNP was identified on B03 for SHP during the rainy season, with a PVE of 22.52% (Fig. 4, Supplementary Table S5, Supplementary Table S6, Supplementary Figure S1).

Genomic regions associated with grade-I kernel

Three component traits, namely PNR-I, HKW-I, and CPO-I, constitute grade-I kernel traits. A total of fifteen SNPs were identified to be associated with the kernel-grade component traits. Out of which, nine were unique significant SNPs above the Bonferroni corrected threshold. Three SNPs (AX_147226917, AX_177642221, and AX_176803178) were identified on chromosomes A07 and B09 for PNR-I, which explained a PVE of 5.32 to 11.90%. SNP AX_147226917 was consistently associated with PNR-I on chromosome A07 during both seasons. Five SNPs (AX_147226949, AX_177637658, AX_176812240, AX_176820983, and AX_177643480) were associated with HKW-I on chromosomes A07, B09, A04, B05, and B08, respectively, during the rainy season with a PVE range of 1.68–7.58%, while one SNP, AX_177638905, on chromosome B07 was associated with HKW-I during the post-rainy season, which explained a PVE of 30.04%. Four SNPs, AX_147226949,

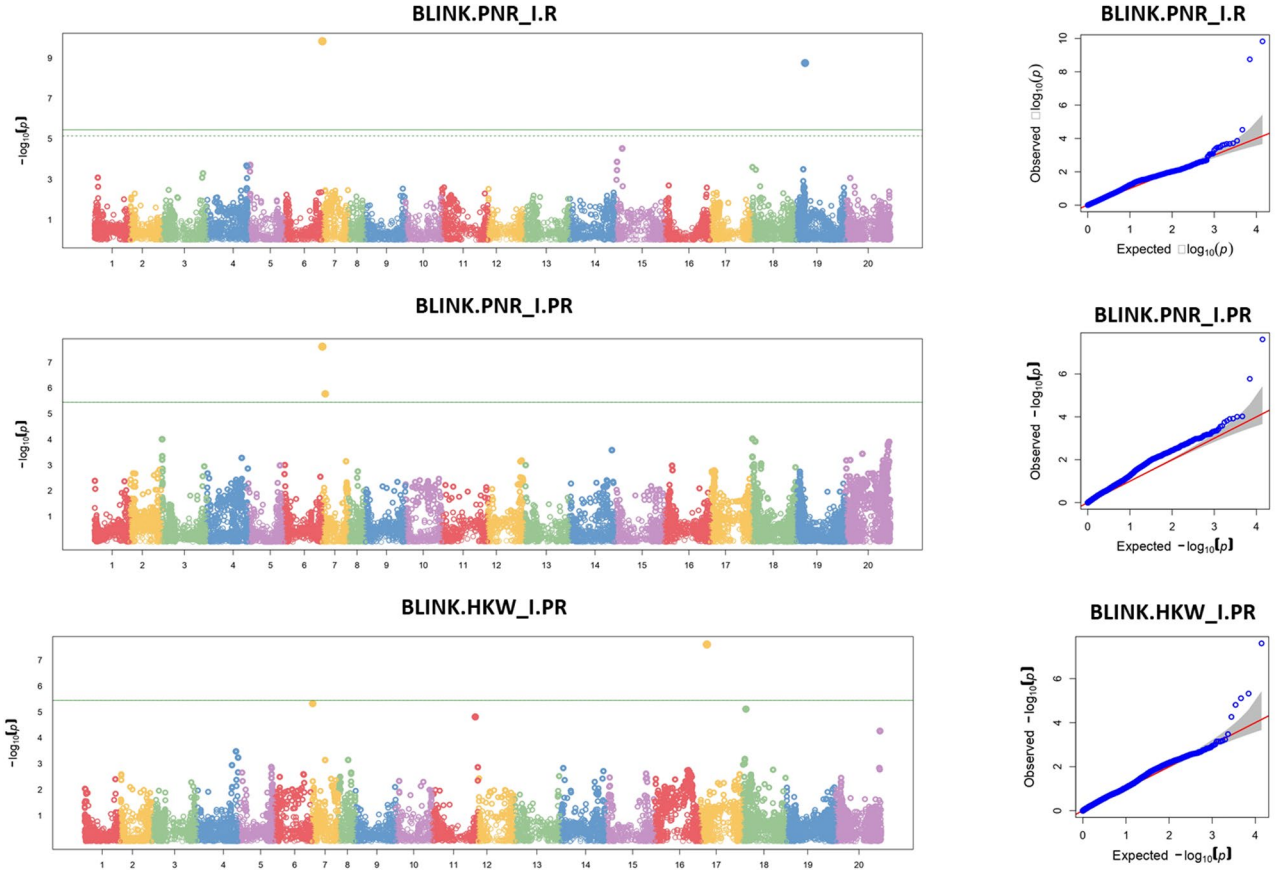


Fig. 4 Manhattan and Q-Q plot for kernel grade component traits such as **a**. PNR_I (Percent net recovery of grade-I kernel) during rainy 2021 **b**. PNR_I during post rainy 2021-22 **c**. HKW_I (Hundred kernel weight of grade-I kernel) during post rainy 2021-22 in 574 MAGIC lines. The significant associations are represented above the threshold lines

AX_176817367, AX_177643480, and AX_176818356, on chromosomes A07, B05, B08, and A03, were associated with CPO-I during the rainy season, which explained a PVE of 1.55–9.52%. Two SNPs, AX_177643480 and AX_147226917 on chromosomes B08 and A07, respectively, were associated with CPO-I during the post-rainy season, with a PVE of 16.53–32.94% (Supplementary Table S5, Supplementary Table S6).

Three SNPs, AX_147226917, AX_177643480, and AX_147226949, were pleiotropic in nature. The SNP AX_147226917 controlled HKW, PNR-I, and CPO-I, whereas SNP AX_177643480 was associated with HKW, HKW-I, and CPO-I. SNP AX_147226949 was associated with HKW-I and CPO-I (Supplementary Table S6). Out of these three SNPs, AX_147226917 on chromosome A07 was associated with PNR-I during both rainy and post rainy seasons. Similarly, AX_177643480 on chromosome B08 was associated with CPO-I during both rainy and post rainy seasons. Allele effect box plots of three important SNPs; AX_176803178, AX_147226949, and AX_177643480 were tested and found to be significant (Fig. 5). A chromosome map is given in Fig. 6 for clear visualization.

Potential candidate genes for yield and kernel grade traits

Potential candidate genes linked to yield and kernel grade traits were identified in PeanutBase (https://Peanutbase.org/) within a 100 kb window of significant SNPs (50 kb upstream and 50 kb downstream) using their physical locations and reference genome sequences. A total of fifty-six genes were identified for yield-related traits (PW, KW, HKW, and SHP) and another thirty-eight genes were identified for kernel grades (Supplementary Table S7). In order to prioritize candidate genes, GO analysis was performed for all these identified genes. The GO analysis identified significantly enriched GO terms across the three principal categories: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). Among the identified candidate genes from A-sub genome (*Arachis duranensis)*, the most prominent enrichment was observed for the oxidation–reduction process (GO:0055114; $p = 0.0021$) under the Biological Process category, involving eight genes (*Aradu.3YG82*, *Aradu.993Q7*, *Aradu.BD60N*, *Aradu.D1YZ0*, *Aradu.SFU0J*, *Aradu.T9TSZ*, *Aradu.Y7AIG*, and *Aradu.YHK80*). In the Cellular Component category, the nucleolus (GO:0005730; $p = 0.0023$) was significantly enriched, suggesting a role in ribosomal biogenesis and nuclear organization. Within the Molecular Function category, NAD binding (GO:0051287; $p = 0.0093$) was enriched, implying that several genes are involved in enzymatic redox reactions and dehydrogenase activity (Supplementary Table S9).

Similarly, the GO analysis of *Arachis ipaensis* candidate genes revealed multiple enriched GO terms. Under the Biological Process category, several ion-related processes were significantly overrepresented, including metal ion homeostasis (GO:0055065; $p = 0.0033$), cation homeostasis (GO:0055080; $p = 0.0042$), and inorganic ion homeostasis (GO:0098771; $p = 0.0052$), represented by genes such as *Araip.7Q1HT* and *Araip.U0AUG*. In the Cellular Component category, the integral component of membrane (GO:0016021; $p = 0.0031$) and intrinsic component of membrane (GO:0031224; $p = 0.0038$) were enriched, highlighting that several genes encode membrane-associated proteins possibly involved in transport and signal transduction. In the Molecular Function category, enrichment of oxidoreductase activity (GO:0016705; $p = 0.0016$), iron ion binding (GO:0005506; $p = 0.0021$), and heme binding (GO:0020037; $p = 0.0028$) was observed, represented by genes including *Araip.C78LH*, *Araip.CLW9Z*, *Araip.F3W88*, and *Araip.L96AH*. These categories reflect a strong association with redox regulation and electron transfer. Furthermore, general molecular function such as metal ion binding (GO:0046872; $p = 0.006$) and (GO:0005488; $p = 0.0094$)

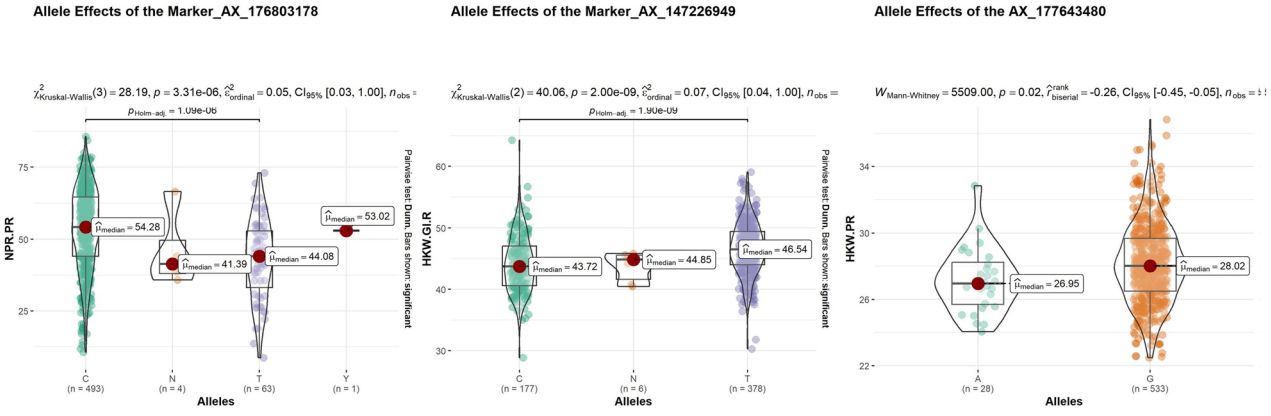


Fig. 5 Allelic effect box plots for SNP markers associated with kernel grade traits: **a** AX_176803178, **b** AX_147226949, and (**c**) AX_177643480

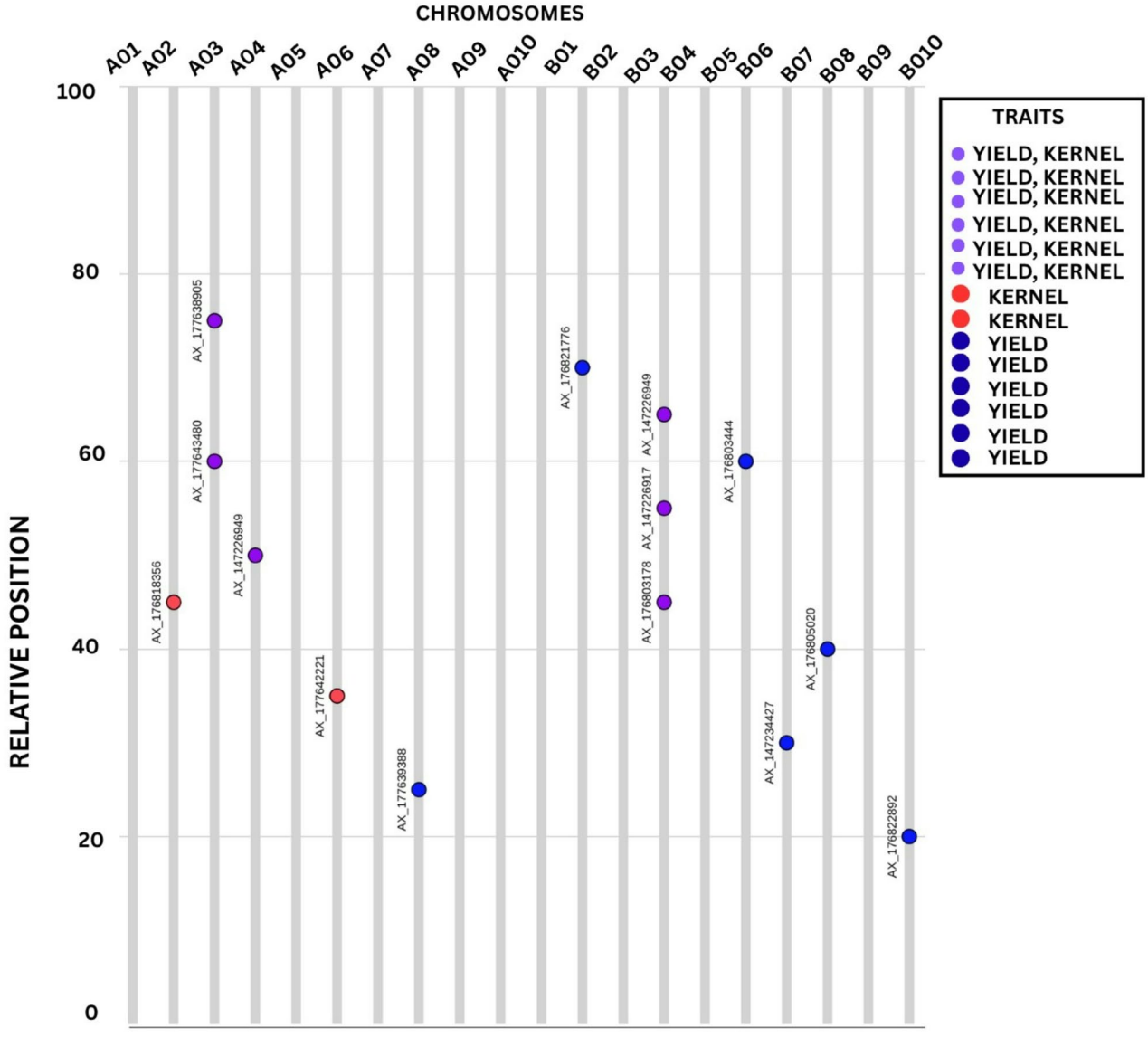


Fig. 6 distribution of yield (YIELD)and kernel grades (KERNEL) SNPs identified through GWAS in groundnut MAGIC population

was also significantly enriched, suggesting the involvement of these genes in a variety of enzymatic and regulatory activities (Supplementary Table S9).

Sixteen genes for yield and seven genes for kernel grades were prioritized based on literature review, gene ontology study, and in silico expression analysis for their roles in biological processes and yield regulation (Tab. 4). The loci *Aradu.6262P* (*CHLG*), *Aradu.Y7AIG* (*CYP*), and *Aradu.Z67WQ* (*CCCH-ZFP*) were identified on chromosome A01. The locus *Aradu.V14167* (*PP7*) was located on chromosome A04. Additionally, the loci *Aradu.7P3Q2* (*B3-DBP*) and *Aradu.D1YZ0* (*ZEP*) were situated on chromosome A09, while *Aradu.TW8M6* (*LEA*) was found on chromosome A03.These loci were associated with pod yield. SNPs and candidate genes associated with kernel yield were also linked to pod yield. Four loci around SNP AX_176815442 on chromosome A03 include *Aradu.45HCQ* (*DnaJ*), *Aradu.A28JW* (*DNAH*), *Aradu.F9ZRP* (*PP7*), and *Aradu.HP9LA* (*PPP*). Three loci on chromosome B07 were *Araip.2TN3Y* (*COX11*), *Araip.8E63N* (*KIN*), and *Araip.C78LH* (*CYP*). Loci on chromosome A07 include *Aradu.188J4* (*DNAH*), *Aradu.8G9XJ* (*ENO*), and *Aradu.BD60N* (*GPT*). Additionally, potential candidate genes *Araip.CLW9Z* (*CYP*) and *Araip.UT46I* (*bHLH10*) for shelling percentage were associated with SNP AX_147246094 on chromosome B03. A total of thirty-eight genes were identified for kernel grades (Supplementary Table S8). Of these, only seven were prioritized as potential candidate genes based on a literature search. Key genes regulating kernel grades

Table 4 Gene description of SNPs associated with yield attributing and kernel grade component traits

Trait	SNP ID	Chr	Diploid Gene ID	Tetraploid Gene ID	Sart	End	Length	Gene description
PW_R	AX_176822892	A01	*Aradu.6262P*	*Arahy.44A4XL*	3,398,298	3,399,687	1389	*chlorophyll synthase*
			Aradu.7L5GB	*Arahy.00G2U6*	3,377,240	3,383,903	6663	*ATP-binding ABC transporter*
			Aradu.Y7AIG	*Arahy.08738Y*	3,416,192	3,417,702	1510	*Cytochrome P450 superfamily protein*
	AX_176805020	A04	*Aradu.V14167*	*Arahy.6VZ2KG*	97,104,284	97,106,562	2278	*serine/threonine-protein phosphatase 7 long form homolog*
	AX_147234427	A09	*Aradu.7P3Q2*	*Arahy.108MS0*	115,055,440	115,077,534	22,094	*plant-specific B3-DNA-binding domain protein*
			Aradu.D1YZ0	*Arahy.MS9EFZ*	115,143,020	115,147,880	4860	*zeaxanthin epoxidase*
PW_PR	AX_176803444	A03	*Aradu.TW8M6*	*Arahy.FE1PUB*	25,790,840	25,792,591	1751	*Late embryogenesis abundant (LEA) protein-related (Root cap)*
KW_R	AX_147234427	A09	*Aradu.7P3Q2*	*Arahy.108MS0*	115,055,440	115,077,534	22,094	*plant-specific B3-DNA-binding domain protein*
HKW_R	AX_176815442	A03	*Aradu.45HCQ*	*Arahy.1B470G*	13,135,235	13,138,969	3734	*Chaperone DnaJ-domain superfamily protein, (DnaJ domain)*
			Aradu.A28JW	*Arahy.JS7C8X*	13,170,555	13,174,493	3938	*large subunit GTPase 1 homolog*
	AX_177638855	B07	*Araip.2TN3Y*	*Arahy.87SQNY*	114,655	117,664	3009	*cytochrome c oxidase assembly protein CtaG/Cox11 family*
			Araip.8E63N	*Arahy.07YELL*	66,887	73,309	6422	*Protein kinase superfamily protein*
			Araip.C78LH	*Arahy.05IA5I*	109,894	111,123	1229	*Cytochrome P450 superfamily protein*
	AX_176822338	B08	*Araip.U0AUG*	*Arahy.44QEU2*	26,178,326	26,183,214	4888	*sodium/calcium exchanger family protein*
	AX_147226917	A07	*Aradu.188J4*	*Arahy.TSR8I7*	1,225,601	1,228,083	2482	*ATP-dependent DNA helicase, (P-loop containing nucleoside triphosphate hydrolase)*
			Aradu.8G9XJ	*Arahy.86EPEX*	1,210,713	1,215,639	4926	*phosphopyruvate hydratase (Enolase)*
PNR_R	AX_176803178	A07	*Aradu.6Z78F*	*Arahy.08TDR6*	9,722,343	9,723,004	661	*RING-H2 finger protein 2B, (Zinc finger, RING/FYVE/PHD-type)*
			Aradu.993Q7	*Arahy.5D1G3K*	9,730,990	9,733,378	2389	*ascorbate peroxidase 1*
PNR-PR	AX_147226949	A07	*Aradu.HR82P*	*Arahy.1PBH7D*	1,472,710	1,473,414	705	*ALG-2 interacting protein X-like [Glycine max]*
			Aradu.S3AS8	*Arahy.1NQ2CM*	1,454,176	1,456,622	2447	*Vacuolar protein-sorting protein BRO1*
	AX_177643480	B08	*Araip.MKV8R*	*Arahy.29S8ZW*	8,698,068	8,699,400	1333	*protein FAR1-RELATED SEQUENCE 3-like isoform X1 [Glycine max]*
HKW_I_PR	AX_147226949	A07	*Aradu.HR82P*	*Arahy.10Y6YW*	1,472,710	1,473,414	705	*ALG-2 interacting protein X-like [Glycine max]*
			Aradu.S3AS8	*Arahy.17Q87X*	1,454,176	1,456,622	2447	*Vacuolar protein-sorting protein BRO1*

Table 5 Fold-change analysis of top-ranked candidate genes

Gene ID	Nearest Ref ID	Mean Expression (FPKM)	Highest Fold Change (vs. Root_veg)	Tissue of Maximum Expression
XLOC_075642	AH19G43450, AH19G43460	7.57×10^6	~6.4-fold	Seeds_25
XLOC_075639	AH19G43420	7.57×10^6	~6.2-fold	Flower
XLOC_075638	AH19G43410	7.57×10^6	~5.9-fold	PodWall_mature
XLOC_075637	AH19G43400	7.57×10^6	~5.7-fold	Seeds_15
XLOC_077872	AH19G43370	7.57×10^6	~5.5-fold	Seeds_25

include *Aradu.6Z78F* (*RHF2B*), *Aradu.993Q7* (*APX1*), *Aradu.HR82P* (*ALIX*), *Araip.MKV8R* (*FRS3*), and *Aradu.S3AS8* (*BRO1*) (Table 5).

In-silico gene expression analysis and allelic distribution pattern

The tissue-specific expression of the identified candidate genes was examined using the *A. hypogaea* gene expression atlas (*AhGEA*) for the *fastigiata* subspecies. Among the potential candidate genes identified for yield, only 13 genes exhibited differential expression in at least one tissue during critical developmental stages, as observed across 20 tissues in the gene expression atlas (Figs. 7 and 8). Fold-change expression analysis was carried out to compare the relative expression levels of the top-ranked genes across different tissues. Mean expression values (FPKM) were calculated for all tissues, and the top five highly expressed genes (XLOC_075642, XLOC_075639, XLOC_075638, XLOC_075637, and XLOC_077872)

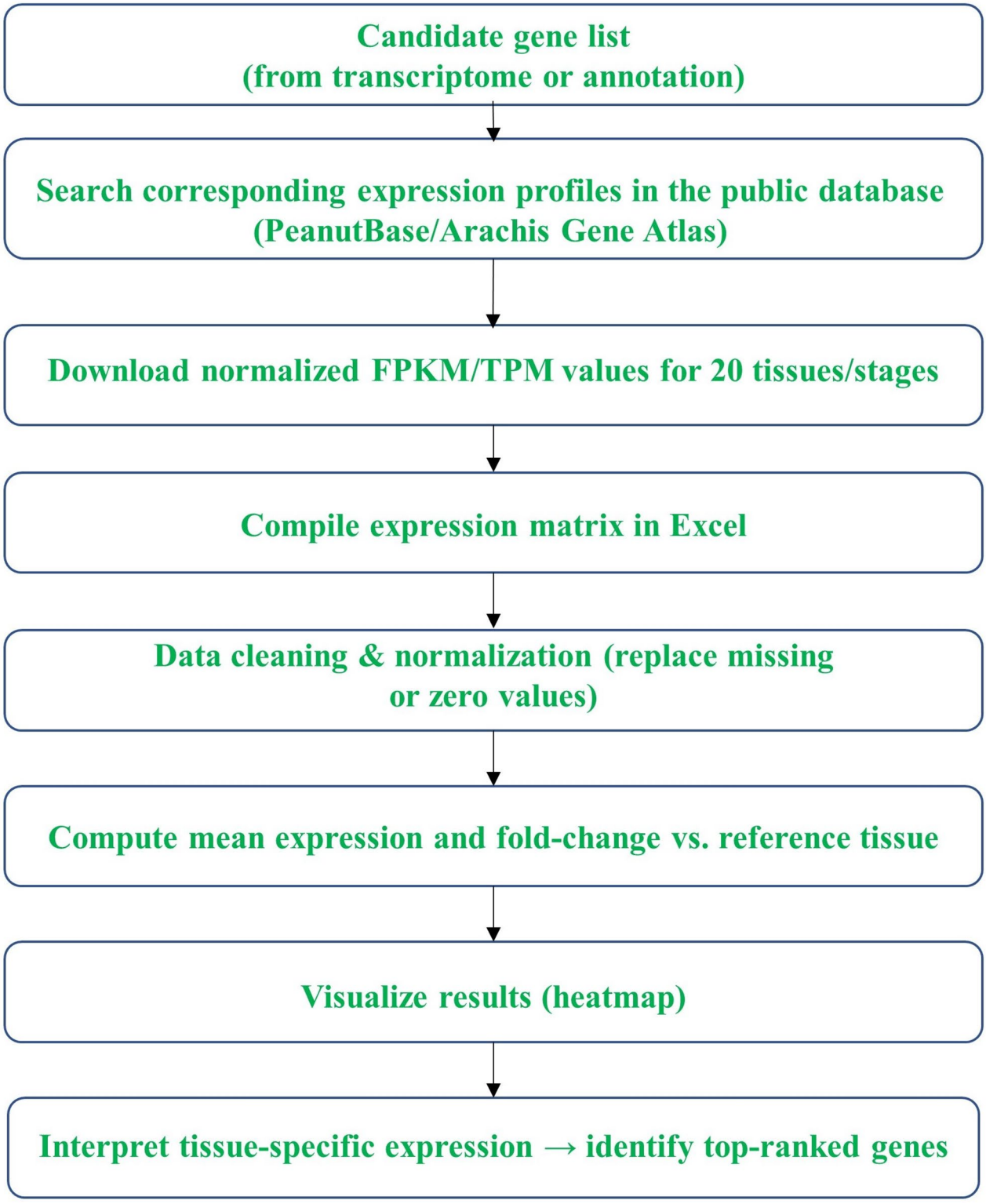


Fig. 7 Workflow for tissue-specific expression analysis of identified candidate genes in *Arachis hypogaea*. The pipeline involves retrieval of expression data from public repositories, normalization and matrix compilation, followed by statistical interpretation and visualization to identify tissue-preferentially expressed genes

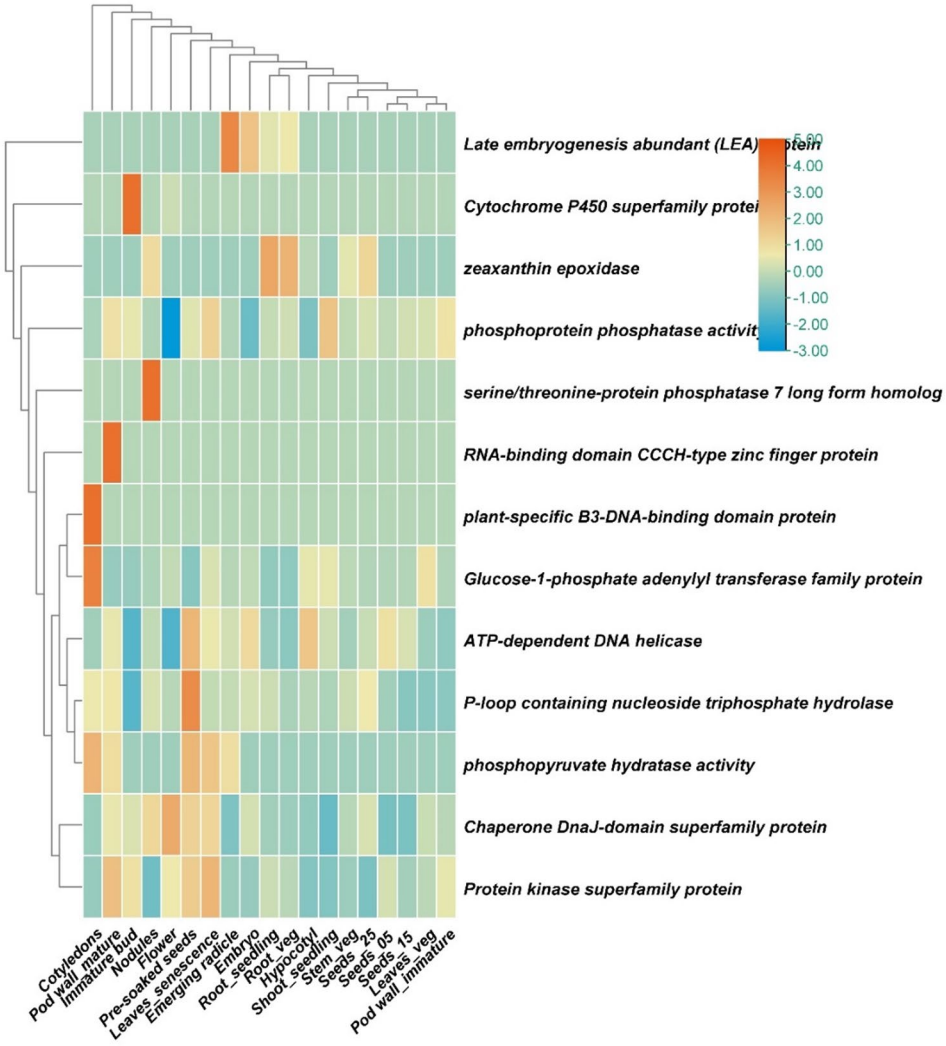


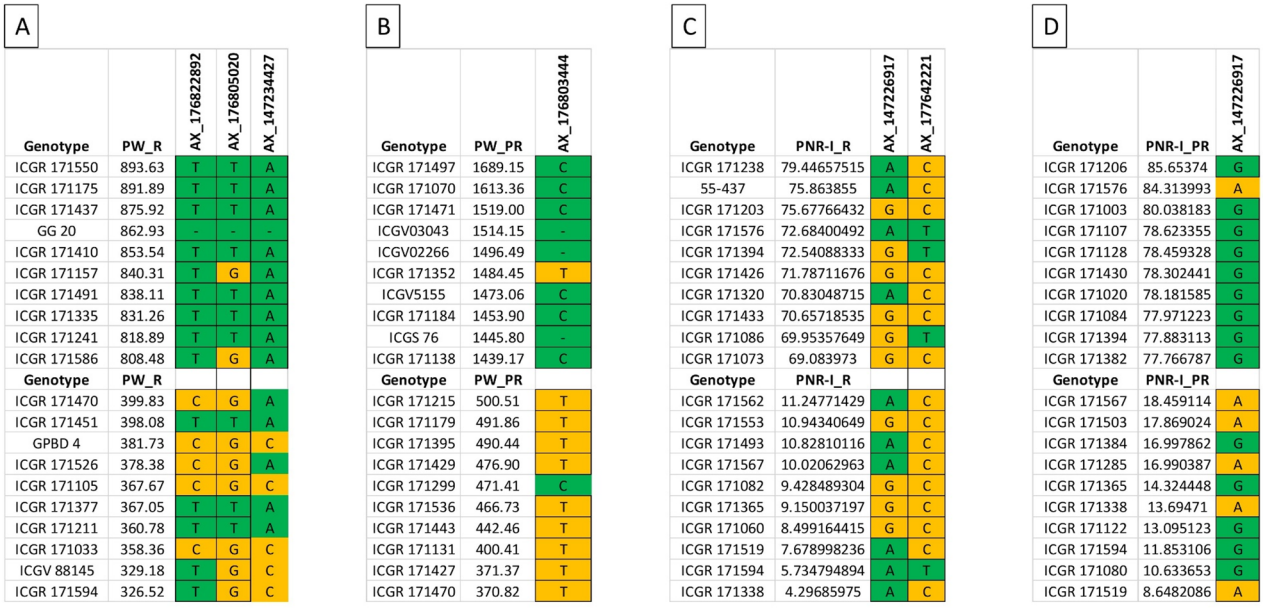
Fig. 8 *In-silico* gene expression analysis heat map for kernel grades and yield traits

were selected based on their average expression across the dataset. A list of top ranked candidate genes across different tissues is given in Tab. 5. The results revealed that these genes were highly upregulated in reproductive tissues (flower, seeds, pod wall) compared with vegetative tissues (root and stem). For example, XLOC_075642 and XLOC_075639 exhibited > 5-fold higher expression in seeds and flower compared with roots, suggesting their potential roles in reproductive development. These findings confirm strong transcriptional activity of the top-ranked candidate genes in reproductive organs, emphasizing their potential biological significance in pod and seed development. The phenotypic data of the top ten MAGIC lines (MLs) with high and low pod yields and percent net recovery of grade-I kernels from the 2021-22 rainy and post-rainy seasons were analyzed to assess SNP utility and efficiency (Fig.9). MLs with high yield and percent net recovery of grade-I kernels mostly carried favorable alleles, whereas MLs with low yield and percent net recovery of grade-I kernels predominantly carried unfavorable alleles.

Discussion

Uniqueness of kernel grades in groundnut and need of marker development for this trait

The kernel grade is highly influenced by environment. With increasing confectionary groundnut demand versus oil markets, there is demand for developing high-yielding cultivars with better grade-I kernel recovery. Studies show bold seed size contributes to grade-I kernel recovery percentage. Kernel grading ensures product quality and market value, benefiting processing and exports. Well-graded groundnuts meet standards for consumer acceptance and safety. Quality kernels with uniform sizes enhance their use in food products and confectionery [4]. Premium grades command higher prices for their quality. In exports, adherence to international standards like the International Organization for Standardization



A

Genotype	PW_R	AX_176822892	AX_176805020	AX_147234427
ICGR 171550	893.63	T	T	A
ICGR 171175	891.89	T	T	A
ICGR 171437	875.92	T	T	A
GG 20	862.93	-	-	-
ICGR 171410	853.54	T	T	A
ICGR 171157	840.31	T	G	A
ICGR 171491	838.11	T	T	A
ICGR 171335	831.26	T	T	A
ICGR 171241	818.89	T	T	A
ICGR 171586	808.48	T	G	A
Genotype	**PW_R**			
ICGR 171470	399.83	C	G	A
ICGR 171451	398.08	T	T	A
GPBD 4	381.73	C	G	C
ICGR 171526	378.38	C	G	A
ICGR 171105	367.67	C	G	C
ICGR 171377	367.05	T	T	A
ICGR 171211	360.78	T	T	A
ICGR 171033	358.36	C	G	C
ICGV 88145	329.18	T	G	C
ICGR 171594	326.52	T	G	C

B

Genotype	PW_PR	AX_176803444
ICGR 171497	1689.15	C
ICGR 171070	1613.36	C
ICGR 171471	1519.00	C
ICGV03043	1514.15	-
ICGV02266	1496.49	-
ICGR 171352	1484.45	T
ICGV5155	1473.06	C
ICGR 171184	1453.90	C
ICGS 76	1445.80	-
ICGR 171138	1439.17	C
Genotype	**PW_PR**	
ICGR 171215	500.51	T
ICGR 171179	491.86	T
ICGR 171395	490.44	T
ICGR 171429	476.90	T
ICGR 171299	471.41	C
ICGR 171536	466.73	T
ICGR 171443	442.46	T
ICGR 171131	400.41	T
ICGR 171427	371.37	T
ICGR 171470	370.82	T

C

Genotype	PNR-I_R	AX_147226917	AX_177642221
ICGR 171238	79.44657515	A	C
55-437	75.863855	A	C
ICGR 171203	75.67766432	G	C
ICGR 171576	72.68400492	A	T
ICGR 171394	72.54088333	G	T
ICGR 171426	71.78711676	G	C
ICGR 171320	70.83048715	A	C
ICGR 171433	70.65718535	G	C
ICGR 171086	69.95357649	G	T
ICGR 171073	69.083973	G	C
Genotype	**PNR-I_R**		
ICGR 171562	11.24771429	A	C
ICGR 171553	10.94340649	G	C
ICGR 171493	10.82810116	A	C
ICGR 171567	10.02062963	A	C
ICGR 171082	9.428489304	G	C
ICGR 171365	9.150037197	G	C
ICGR 171060	8.499164415	G	C
ICGR 171519	7.678998236	A	C
ICGR 171594	5.734794894	A	T
ICGR 171338	4.29685975	A	C

D

Genotype	PNR-I_PR	AX_147226917
ICGR 171206	85.65374	G
ICGR 171576	84.313993	A
ICGR 171003	80.038183	G
ICGR 171107	78.623355	G
ICGR 171128	78.459328	G
ICGR 171430	78.302441	G
ICGR 171020	78.181585	G
ICGR 171084	77.971223	G
ICGR 171394	77.883113	G
ICGR 171382	77.766787	G
Genotype	**PNR-I_PR**	
ICGR 171567	18.459114	A
ICGR 171503	17.869024	A
ICGR 171384	16.997862	G
ICGR 171285	16.990387	A
ICGR 171365	14.324448	G
ICGR 171338	13.69471	A
ICGR 171122	13.095123	G
ICGR 171594	11.853106	G
ICGR 171080	10.633653	G
ICGR 171519	8.6482086	A

Fig. 9 Allele distribution pattern of significant Markers associated with **A**. Pod weight during rainy **B**. Pod weight during post rainy **C**. Percent net recovery of grade-I kernels during rainy **D**. Percent net recovery of grade-I kernels during post-rainy seasons

(ISO) or the United States Department of Agriculture (USDA) is essential. Importing countries require kernels free from contaminants like aflatoxins [36]. Processing industries benefit from graded kernels as uniform sizes simplify mechanical shelling. Larger kernels suit confectionery while smaller ones are used for oil extraction. Grading prevents contaminated kernels from entering the food chain [37]. This practice enhances quality and marketability.

BLUP means, variance components and correlation

The high genotypic variance in PW, KW, and PNR-I across seasons indicate wide variability in the MAGIC population. MAGIC lines performed better during post-rainy season than rainy season for yield and kernel gradess, due to lower disease pressure and better pod filling. During the rainy season, MLs; ICGR 171586 (1381.66 kg/ha) and ICGR 171175 (1374.12 kg/ha) recorded higher yields than check variety ICGV 03043 (1334.01 kg/ha). This demonstrate their adaptability under rainfed conditions. In contrast, under post-rainy conditions, where temperature and moisture regimes were more favorable and irrigation was controlled, ICGR 171497 (2815.24 kg/ha) and ICGV 03043 (2523.59 kg/ha), exhibited superior performance compared to check ICGV 02266 (2494.15 kg/ha). During post-rainy season, shelling percentage varied from 49% in ICGR 171387 due to immature pods, to 81% in ICGR 171379. High shelling percentage upto 78.9% has been reported [38] in breeding lines. A very high shelling percentage is possible for certain improved genotypes under optimal agronomy, with correct harvest timing, low moisture and controlled lab measurement. Spanish bunch types of groundnut have higher shelling percentage (up to 78%) due to thinner shells than Virginia types [1]. A very high shelling percentage in ICGR 171379 may be a result of good production practices mentioned above. However, owing to CT scan model prediction error it may vary ± 2%. ML, ICGR 171576 records a grade-I kernel recovery of 84.31% PR and 72.68% R, with an SHP of 77.20% PR and 72.89% R in both seasons. These consistent performing line with better PNR-I kernel and SHP can be used as parent in breeding programs to simultaneously improve multiple traits across seasons.

Kernel grade as a trait per se has received little attention in groundnut breeding programs. A previous study assessed two cultivars, GG-20 and TG-37 A for their kernel width and reported that 71% kernel were having a width of >8.19 mm in GG-20 whereas in TG-37A, 29% kernels were having width of >8.62 mm [39]. Out of the large set of ML assessed for kernel grade, only seven MLs (ICGR 171238, ICGR 171203, ICGR 171576, ICGR 171394, ICGR 171426, ICGR 171320 and ICGR 171433) had a PNR-I value more than 70% during rainy season. During post rainy season, three MLs (ICGR 171206, ICGR 171576 and ICGR 171003) had PNR-I kernel value >80% and forty-five MLs had PNR-I kernel value >70%. Higher recovery of grade-I kernel during post rainy can be attributed to cooler night temperatures, wider diurnal variation, reduced disease and waterlogging, and controlled irrigation that together favor kernel development and better yield [40]. Positive correlations between PNR-I and HKW suggest bold-seeded varieties contribute more to grade-I kernel PNR. PW and KW positively

correlated with PNR-I in post-rainy season only, likely due to extended maturity allowing better pod filling and kernel development.

Suitability of the MAGIC population for this study

MAGIC populations enable studying genomic architecture and discovering genomic regions governing complex traits with precision by integrating genetic diversity and high recombination rates [41]. MAGIC population used in the current study showed wide variability in yield, disease, kernel grades, and quality traits due to recombination events creating diverse allele combinations. MAGIC populations derived from multiple founders showed increased allelic diversity [42]. Enhanced genetic diversity augments GWAS power by providing more genetic variation for association testing. MAGIC populations undergo numerous recombination events during multiple generations of intercrossing, creating smaller linkage disequilibrium (LD) blocks [43, 44]. Self-pollinating crops, such as groundnut, typically exhibit larger LD blocks and demonstrate slower LD decay due to limited recombination events [45]. Numerous studies have reported extensive LD in various groundnut breeding populations [46, 47]. For instance, in a previous GWAS experiment, an LD decay of up to 4.8 MB, more than the LD reported in the current study has been estimated [48]. Although a high LD block was estimated for the current MAGIC population due to the tetraploid nature and genome complexity of groundnut, still significant associations were identified through the GWAS with a high p-value. Manhattan plots showed distinct peaks for various yield and kernel grade component traits, and Q-Q Plots exhibited inflation only at the tail of the distribution, reflecting true associations. All of these factors validated the suitability of the MAGIC population for association studies.

Genome-wide association study

Association analysis was conducted separately for rainy and post rainy seasons, as these two are different seasons where one has protected irrigation. Rainy and post-rainy season adaptation is needed in India, as commodity production occurs largely in the rainy season, but seed production occurs in the post-rainy season. A comprehensive literature search was carried out for a comparative analysis of yield and kernel-grade associated genomic regions in previous and the current study. Yield traits in groundnut are influenced by multiple genes and key QTLs were mapped on chromosome A03 for breeding purposes [49]. QTLs for seed size were reported on chromosomes A05 and A07 through QTL mapping studies using RIL populations [50]. Through GWAS, we have also identified genomic regions associated with hundred kernel weight (HKW) on chromosome A07. Several genomic regions associated with hundred seed and pod weight were identified utilizing a NAM population and SNP array through GWAS on chromosome A03, A05, A06, A07, A07, A09, B06, B07, B08 and B09 [25]. In this previous study, significant MTAs were co-localized for pod weight and hundred kernel weight on chromosomes A05, A06, B05, and B06. This study supports results from our study, where we have also identified significant SNPs for HKW on chromosomes A03, A05, A07, B07 and B08. QTLs for hundred kernel weight (HKW) on chromosomes A02 and A06, at loci A02-86439145 and A06-108577126 were identified through QTL mapping in another study [28]. Overlapping QTLs for shelling percentage and HSW were identified on chromosomes A05, A08, B10, B06 and A08 [51]. But in the current study, we have found associations with shelling percentage only on chromosome B03. In the current study, all SNPs associated with yield traits like pod and kernel weight are mapped on different chromosomes of A genome, suggesting a significant role of A genome compared to B genome. In our study, three component traits of grade-I kernel revealed nine significant SNPs. SNP AX_147226917 on chromosome A07 was detected for PNR-I across two seasons, indicating a stable genetic determinant. PVE by the detected SNPs ranged broadly from 5.3% up to 32.9%, which suggests a mixture of major and minor effect loci underpinning kernel-grade traits. SNP AX_177638905 on B07 accounted for more than 30% PVE for HKW-I in post-rainy season, which is a major effect SNP. Overlapping SNPs associated with HKW-I and CPO-I suggest pleiotropy or linked loci controlling kernel grade trait. These findings support polygenic architecture for kernel grade traits. Though several GWAS were conducted previously for seed size, the current study is the first attempt, specifically to identify genomic regions for kernel grades in groundnut.

Phenotypic variance explained by the associated SNPs for PW (3.37–10.89%), KW (7.43–12.48%), HKW (3.20–12.94.20.94%), SHP (22.52%) appeared to be modest in the current study. Previous linkage mapping studies for yield related traits have also identified QTLs for pod yield (6.27–6.87%), hundred seed weight (5.89–13.87%) and shelling percentage (10.98–11.65%) with less PVEs [52]. Similar studies have attributed less PVE to the complex and quantitative nature of the yield attributing traits and G×E interactions, which reduces the detectable PVE per marker [29]. In the current study, sufficient number of SNPs (13,937) from a high-density SNP array and a statistically powerful model (BLINK) were used for GWAS analysis to identify causal loci with smaller effects.

Both positive and negative allelic effects were found for identified SNPs, with alleles increasing or decreasing trait values. A positive effect suggests that the reference allele enhances the phenotypic value, whereas a negative effect denotes a reduction. For PW, SNP AX_176822892 on chromosome A01 showed a positive effect, while three

SNPs had negative effects. All SNPs associated with KW showed negative effects. For HKW, five SNPs had positive effects and two had negative effects. SNP AX_147246094 on B03 showed a positive effect on SHP. For PNR-I kernel, three SNPs had positive effects and one had negative effect. HKW-I showed both positive and negative effects, while all CPO-I SNPs had negative effects.

Potential candidate genes for yield and kernel grades

The candidate genes are categorized into various functional groups. Gene Ontology (GO) enrichment analysis, literature search and expression analysis indicate that these genes are implicated in oxidoreductase activity (including *Cytochrome P450 superfamily protein*, *ascorbate peroxidase 1*, *cytochrome c oxidase assembly protein*), photosynthesis and pigment biosynthesis (such as *chlorophyll synthase* and *zeaxanthin epoxidase*), metabolism (*enolase*, *glucose-1-phosphate adenylyltransferase*), ion transport (*ATP-binding ABC transporters*, *sodium/calcium exchanger*, *vacuolar protein-sorting protein BRO1*), signal transduction (*serine/threonine-protein phosphatase 7*, *protein kinase superfamily protein*), regulatory proteins (*B3-DNA-binding domain protein*, *FAR1-RELATED SEQUENCE 3-like isoform X1*), and stress-related proteins (*LEA protein*, *DnaJ-domain protein*).

Oxidoreductase activity, mediated by *Cytochrome P450 protein*, regulates yield by influencing the biosynthesis of gibberellic acid and brassinosteroids [53], which are critical for cell division, seed development, and nutrient transport. The *cytochrome c oxidase assembly protein* is essential for pollen development and growth in *Arabidopsis* [54]. *Chlorophyll synthase* catalyzes the biosynthesis of chlorophyll a, thereby enhancing photosynthetic efficiency [55–57]. Increased chlorophyll content is associated with improved photosynthetic rates, leading to a higher number of pods and increased kernel weight [58]. *Zeaxanthin epoxidase* regulates yield by mediating abscisic acid (ABA) biosynthesis, which is crucial for stress tolerance and development. In groundnuts, *ZEP* activity maintains ABA levels, facilitating seed maturation and germination, thereby contributing to higher yields [59]. *Phosphopyruvate hydratase* (*enolase*) is a pivotal glycolytic enzyme that supports carbon and energy supply for seed development. Mutations in this enzyme can restrict carbohydrate flux, resulting in smaller seeds in *Arabidopsis* [60]. *Glucose-1-phosphate adenyltransferase proteins* (*AGPase*) are involved in starch biosynthesis, and their downregulation leads to reduced starch content in groundnut leaves [61].

Membrane transport proteins are crucial for nutrient partitioning towards developing seeds, promoting higher seed-filling rates and kernel mass [62]. The vacuolar protein-sorting protein *BRO1* confers bacterial wilt resistance in groundnuts [63]. *Serine/threonine-protein phosphatases* regulate cell division in *Arabidopsis* [64] and influence yield-related traits in groundnuts [65, 66]. *Phosphatases* balance growth and defense mechanisms, while *protein kinases* regulate cell growth through protein phosphorylation. *PSW1* regulates pod size, with *PSW1 HapII* enhancing seed size [67]. *B3-DNA-binding proteins* modulate auxin-responsive genes affecting yield [59]. *FAR1* transcription factors and *RING-type E3 ligases* control hormonal signaling in developing kernels [68]. *Late embryogenesis abundant* (*LEA*) protein is involved in abiotic stress responses in peanuts, including drought and low temperature [69]. Chaperone *DnaJ-domain superfamily proteins* are critical for drought tolerance and seed weight in groundnuts [65, 66, 70]. The transcriptional regulator *STERILE APETALA-like* influences pod and seed size variation during peanut evolution [71].

Conclusion

Considering the increasing demand of groundnut for confectionary industry, it is high time to include traits like kernel grades in the breeding programs. ML ICGR 171238 (79.45%) and ICGR 171206 (85.65%) with highest percent net recovery of grade-I kernel (PNR-I) in rainy and post-rainy seasons respectively can serve as parents in breeding programs. Another ML, ICGR 171576 with a grade-I kernel recovery of 84.31% PR and 72.68% R, with an SHP of 77.20% PR and 72.89% R can also be used in crossing programs to improve these traits simultaneously. Consistent SNP AX_147226917 (A07) and AX_177643480 (B08) associated with PNR_I and counts per ounce of grade-I kernel (CPO-I) across seasons, and SNP AX_177638905 on B07 with highest PVE associated with HKW-I can be further validated and developed in to KASP assay that can augment rapid selection of progenies in early generations to develop groundnut cultivars with high recovery of grade-I kernels. Candidate genes identified for kernel grades in the current study include *Aradu.6Z78F* (*RING-H2 finger protein*), *Aradu.993Q7* (*ascorbate peroxidase 1*), *Araip.MKV8R* (*protein FAR1-RELATED SEQUENCE 3-like isoform X1*) and *Aradu.S3AS8* (*Vacuolar protein-sorting protein BRO1*). For yield traits, *Aradu.Y7AIG* (*cytochrome P450*), *Aradu.BD60N* (*Glucose-1-phosphate adenylyltransferase*) and *Aradu.TW8M6* (*LEA protein*). These are potential/putative in nature and can be targeted for improving kernel grades once functionally validated.

Abbreviations

ALIX	ALG-2 interacting protein X-like [Glycine max]
APX1	Ascorbate peroxidase 1
B3 or B3-DBP	Plant-specific B3-DNA-binding domain protein
bHLH10	Basic helix-loop-helix 10
BRO1	Vacuolar protein-sorting protein BRO1
CHLG	Chlorophyll synthase
CCCH-ZFP	CCCH- zing finger protein

CYP	Cytochrome P450 superfamily protein
DNAH	ATP-dependent DNA helicase
DnaJ or HSP40	Chaperone DnaJ-domain superfamily protein (DnaJ domain)
ENO	Phosphopyruvate hydratase (Enolase)
FRS3	Protein FAR1-RELATED SEQUENCE 3-like isoform X1 [Glycine max]
GPT	Glucose-6-phosphate
LEA	Late embryogenesis abundant (LEA) protein
PK or KIN	Protein kinase superfamily protein
PPP	Phosphoprotein Phosphatase
PP7	Serine/threonine-protein phosphatase 7 long form homolog
NCX	Sodium/calcium exchanger family protein
RHF2B or RING	RING-H2 finger protein 2B
ZEP	Zeaxanthin epoxidase

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12332-z.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Supplementary Material 6.

Supplementary Material 7.

Supplementary Material 8.

Supplementary Material 9.

Acknowledgements
The authors are thankful to Gopi Potupureddi, Nallapu Harshavardhan, Siri Muddada, and Latha for their support in conducting experiments.

Authors' contributions
AP: ** Investigation, Writing-original draft, Data curation, Validation, Visualization, Formal analysis, software, Writing-review and editing **AR: ** Formal analysis, software, Writing-original draft, Writing-review and editing **DL: ** Conceptualization, Writing-review and editing, Supervision **TM: ** Investigation **, Methodology, Writing-original draft Supervision **SK: ** Data curation, Writing-review and editing**SC: **Methodology, Writing-review and editing **JP: ** **Conceptualization, Methodology, Project administration, Funding acquisition, Resources, Supervision, Writing-review and editing.**.

Funding
Not available.

Data availability
The datasets generated during the current study are provided in the supplementary materials.

Declarations

Ethics approval and consent to participate
No specific permits were required for this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 31 July 2025 / Accepted: 12 November 2025
Published online: 29 November 2025

References

1. Janila P, Variath MT, Pandey MK, Desmae H, Motagi BN, Okori P, Manohar SS, Rathnakumar AL, Radhakrishnan T, Liao B, Varshney RK. Genomic tools in groundnut breeding program: status and perspectives. Front Plant Sci. 2016;7:289. https://doi.org/10.3389/fpls.2016.00289.
2. Shasidhar Y, Vishwakarma MK, Pandey MK, Janila P, Variath MT, Manohar SS, Nigam SN, Guo B, Varshney RK. Molecular mapping of oil content and fatty acids using dense genetic maps in groundnut (Arachis Hypogaea L). Front Plant Sci. 2017;8:794. https://doi.org/10.3389/fpls.2017.00794.
3. Zhao T, Ying P, Zhang Y, Chen H, Yang X. Research advances in the high-value utilization of peanut meal resources and its hydrolysates: A review. Molecules. 2023;28(19):6862. https://doi.org/10.3390/molecules28196862.
4. Janila P, Nigam SN, Pandey MK, Nagesh P, Varshney RK. Groundnut improvement: use of genetic and genomic tools. Front Plant Sci. 2013;4:23. https://doi.org/10.3389/fpls.2013.00023.
5. Kona P, Mahatma MK, Gangadhara K, Ajay BC, Kumar N, Reddy KK, Dagla T, Kumar TL, Gor HK. Evaluation and identification of promising advanced breeding lines for quality and yield traits in groundnut (Arachis Hypogaea L). J Agric Sci. 2021;17(2):280–6.
6. Kona P, Ajay BC, Gangadhara K, Kumar N, Choudhary RR, Mahatma MK, Singh S, Reddy KK, Bera SK, Sangh C, Rani K. AMMI and GGE biplot analysis of genotype by environment interaction for yield and yield contributing traits in confectionery groundnut. Sci Rep. 2024;14(1):2943.
7. Hariprasanna K, Lal C, Radhakrishnan T, Gor HK, Chikani BM. Performance of confectionery groundnut cultures of ICRISAT at Junagadh, India. Legume Research-An Int J. 2013;36:280–3.
8. Kona P, Sangh C, Ravikiran KT, Ajay BC, Kumar N. Genetic and genomic resource to augment breeding strategies for biotic stresses in groundnut. In genomics-aided breeding strategies for biotic stress in grain legumes. Singapore: Springer Nature Singapore; 2024. pp. 359–403. https://doi.org/10.1007/978-981-97-3917-2_11.
9. Weissburg J, Dean LL, Hendrix K. Changes in runner peanut quality parameters as a function of roast times. Peanut Sci. 2024;51(1):45–58.
10. Schirack AV, Sanders TH, Sandeep KP. Effect of processing parameters on the temperature and moisture content of microwave-blanched peanuts. J Food Process Eng. 2007;30:225–40. https://doi.org/10.1111/j.1745-4530.2007.00110.x.
11. Novakazi F, Krusell L, Jensen JD, Orabi J, Jahoor A, Bengtsson T. You had me at MAGIC! Four barley MAGIC populations reveal novel resistance QTL for powdery mildew. Genes. 2020;11(12):1512.
12. Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, Gosman N, Greenland AJ, Horsnell R, Howells R, O'Sullivan DM, Rose GA. An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. G3: Genes, Genomes, Genetics. 2014;4(9):1603–1610.
13. Desai SP. Development of Magic Population and Tagging of Resistance to Major Diseases in Maize (Zea Mays L.). 2022. Doctoral dissertation, University of Agricultural Sciences, GKVK, Bangalore.
14. Sharma V, Mahadevaiah SS, Latha P, Gowda SA, Manohar SS, Jadhav K, Bajaj P, Joshi P, Anitha T, Jadhav MP, Sharma S. Dissecting genomic regions and underlying candidate genes in groundnut MAGIC population for drought tolerance. BMC plant biol. 2024;24;(1):1044. https://doi.org/10.1186/s12870-024-05749-3.
15. Wankhade AP, Chimote VP, Viswanatha KP, Yadaru S, Deshmukh DB, Gattu S, Sudini HK, Deshmukh MP, Shinde VS, Vemula AK, Pasupuleti J. Genome-wide association mapping for LLS resistance in a MAGIC population of groundnut (Arachis Hypogaea L). Theor Appl Genet. 2023;136:43. https://doi.org/10.1007/s00122-023-04256-7.
16. Sun Z, Zheng Z, Qi F, Wang J, Wang M, Zhao R, Liu H, Xu J, Qin L, Dong W, Huang B. Development and evaluation of the utility of genobaits peanut 40K for a peanut MAGIC population. Mol Breeding. 2023;43(10):72. https://doi.org/10.1007/s11032-023-01417-w.
17. Thompson E, Wang H, Korani W, Fountain JC, Culbreath AK, Holbrook CC, Clevenger JP, Guo B. Genetic and genomic characterization of a multiparent advanced generation intercross (MAGIC) population of peanut (Arachis Hypogaea L). Crop Sci. 2024;65:21402.

18. Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, Leal-Bertioli SC, Ren L, Farmer AD, Pandey MK, Samoluk SS. The genome sequence of segmental allotetraploid peanut Arachis Hypogaea. Nat Genet. 2019;51:877–84. https://doi.org/10.1038/s41588-019-0405-z.
19. Chen X, Lu Q, Liu H, Zhang J, Hong Y, Lan H, Li H, Wang J, Liu H, Li S, Pandey MK. Sequencing of cultivated peanut, Arachis hypogaea, yields insights into genome evolution and oil improvement. Mol Plant. 2019;12:920–34. https://doi.org/10.1016/j.molp.2019.03.005.
20. Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang WC, Zhang L, Zhang X, Tang R, Garg V. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet. 2019;51:865–76. https://doi.org/10.1038/s41588-019-0402-2.
21. Wang C, Liu R, Liu Y, Hou W, Wang X, Miao Y, He Y, Ma Y, Li G, Wang D, Ji Y. Development and application of the Faba_bean_130K targeted next-generation sequencing SNP genotyping platform based on transcriptome sequencing. Theor Appl Genet. 2021;134:3195–207. https://doi.org/10.1007/s00122-021-03885-0.
22. Pandey MK, Agarwal G, Kale SM, Clevenger J, Nayak SN, Sriswathi M, Chitikineni A, Chavarro C, Chen X, Upadhyaya HD, Vishwakarma MK. Development and evaluation of a high-density genotyping 'Axiom_ arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. Sci Rep. 2017;7:40577. https://doi.org/10.1038/srep40577.
23. Pandey MK, Pandey AK, Kumar R, Nwosu CV, Guo B, Wright GC, Bhat RS, Chen X, Bera SK, Yuan M, Jiang H. Translational genomics for achieving higher genetic gains in groundnut. Theor Appl Genet. 2020;133:1679–702. https://doi.org/10.1007/s00122-020-03592-2.
24. Zhao J, Huang L, Ren X, Pandey MK, Wu B, Chen Y, Zhou X, Chen W, Xia Y, Li Z, Luo H. Genetic variation and association mapping of seed-related traits in cultivated peanut (Arachis Hypogaea L.) using single-locus simple sequence repeat markers. Front Plant Sci. 2017;8:2105. https://doi.org/10.3389/fpls.2017.02105.
25. Gangurde SS, Wang H, Yaduru S, Pandey MK, Fountain JC, Chu Y, Isleib T, Holbrook CC, Xavier A, Culbreath AK, Ozias-Akins P. Nested-association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (Arachis hypogaea). Plant Biotechnol J. 2020;18:1457–71. https://doi.org/10.1111/pbi.13311.
26. Chavarro C, Chu Y, Holbrook C, Isleib T, Bertioli D, Hovav R, Butts C, Lamb M, Sorensen RA, Jackson S, Ozias-Akins P. Pod and seed trait QTL identification to assist breeding for peanut market preferences. G3: genes. Genomes Genet. 2020;10:2297–315. https://doi.org/10.1534/g3.120.401147.
27. Chu Y, Chee P, Isleib TG, Holbrook CC, Ozias-Akins P. Major seed size QTL on chromosome A05 of peanut (Arachis hypogaea) is conserved in the US mini core germplasm collection. Mol Breeding. 2020;40:1–16. https://doi.org/10.1007/s11032-019-1082-4.
28. Zhou X, Guo J, Pandey MK, Varshney RK, Huang L, Luo H, Liu N, Chen W, Lei Y, Liao B, Jiang H. Dissection of the genetic basis of yield-related traits in the Chinese peanut mini-core collection through genome-wide association studies. Front Plant Sci. 2021;12:637284. https://doi.org/10.3389/fpls.2021.637284.
29. Guo M, Deng L, Gu J, Miao J, Yin J, Li Y, Fang Y, Huang B, Sun Z, Qi F, Dong W. Genome-wide association study and development of molecular markers for yield and quality traits in peanut (ArachishypogaeaL). BMC Plant Biol. 2024;24:244. https://doi.org/10.1186/s12870-024-04937-5.
30. Wankhade AP, Kadirimangalam SR, Viswanatha KP, Deshmukh MP, Shinde VS, Deshmukh DB, Pasupuleti J. Variability and trait association studies for late leaf spot resistance in a groundnut MAGIC population. Agronomy. 2021;11(11):2193.
31. Domhoefer M, Chakraborty D, Hufnagel E, Claußen J, Wörlein N, Voorhaar M, Anbazhagan K, Choudhary S, Pasupuleti J, Baddam R, Kholova J. X-ray driven peanut trait estimation: computer vision aided agri-system transformation. Plant Methods. 2022;18:76.
32. Butler DG, Cullis BR, Gilmour AR, Gogel BJ, Thompson RA. ASReml-R reference manual version 4. VSN International Ltd. Hemel Hempstead. HP1 1ES UK.2017.
33. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012;28:2397–9. https://doi.org/10.1371/journal.pgen.1005767.
34. Team RC, Team MRC, Suggests MASS, Matrix S. Package stats. R Stats Package. 2018;2018:1–3.
35. Sinha P, Bajaj P, Pazhamala LT, Nayak SN, Pandey MK, Chitikineni A, Huai D, Khan AW, Desai A, Jiang H, Zhuang W. Arachis Hypogaea gene expression atlas for fastigiata subspecies of cultivated groundnut to accelerate functional and translational genomics applications. Plant Biotechnol J. 2020;18:2187–200. https://doi.org/10.1111/pbi.13374.
36. Florkowski WJ, Kolavalli S. Strategies to control aflatoxin in groundnut value chains. Volume 1369. Intl Food Policy Res Inst; 2014.
37. Kachapulula PW, Akello J, Bandyopadhyay R, Cotty PJ. Aflatoxin contamination of groundnut and maize in zambia: observed and potential concentrations. J Appl Microbiol. 2017;122:1471–82. https://doi.org/10.1111/jam.13448.
38. Deshmukh DB, Marathi B, Sudini HK, Variath MT, Chaudhari S, Manohar SS, Rani CVD, Pandey MK, Pasupuleti J. Combining high oleic acid trait and resistance to late leaf spot and rust diseases in groundnut (Arachis Hypogaea L). Front Genet. 2020;11:514. https://doi.org/10.3389/fgene.2020.00514.
39. Gojiya DK, Dobariya UD, Pandya PA, Gojiya KM. Studies on physical properties of peanut seed. Acta Sci Agric. 2020;4(3):1–5.
40. Padmalatha Y, Reddy SR, Reddy TY. The relationship between weather parameters during developmental phase and fruit attributes and yield of peanut. Peanut Sci. 2006;33(2):118–24.
41. Shah R, Huang BE, Whan A, Fradgley NS, Newberry M, Verbyla K, Morell MK, Cavanagh CR. Recombination and structural variation in a large 8-founder wheat MAGIC population. G3: Genes, Genomes, Genetics. 2025;15(4):jkaf037. https://doi.org/10.1093/g3journal/jkaf037.
42. Arrones A, Vilanova S, Plazas M, Mangino G, Pascual L, Díez MJ, Prohens J, Gramazio P. The dawn of the age of multi-parent MAGIC populations in plant breeding: novel powerful next-generation resources for genetic analysis and selection of Recombinant elite material. Biology. 2020;9(8):229.
43. Scott MF, Ladejobi O, Amer S. Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity. 2020;125:396–416. https://doi.org/10.1038/s41437-020-0336-6.
44. Yuan G, Sun K, Yu W, Jiang Z, Jiang C, Liu D, Wen L, Si H, Wu F, Meng H, Cheng L, Yang A, Wang Y. Development of a MAGIC population and high-resolution quantitative trait mapping for nicotine content in tobacco. Front Plant Sci. 2023;13:1086950.
45. Vos PG, Paulo MJ, Voorrips RE, van Visser RG. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. Theor Appl Genet. 2017;130(1):123–35. https://doi.org/10.1007/s00122-016-2798-8.
46. Zhang H, Chu Y, Dang P, Tang Y, Jiang T, Clevenger JP, Ozias-Akins P, Holbrook C, Wang ML, Campbell H, Hagan A, Chen C. Identification of QTLs for resistance to leaf spots in cultivated peanut (Arachis Hypogaea L.) through GWAS analysis. Theor Appl Genet. 2020;133(7):2051–61. https://doi.org/10.1007/s00122-020-03576-2.
47. Zhou Q, Zhou C, Zheng W, Mason AS, Fan S, Wu C, Fu D, Huang Y. Genome-Wide SNP markers based on SLAF-Seq uncover breeding traces in rapeseed (*Brassica Napus* L). Front Plant Sci. 2017;8:648. https://doi.org/10.3389/fpls.2017.00648.
48. Otyama PI, Wilkey A, Kulkarni R, Assefa T, Chu Y, Clevenger J, O'Connor DJ, Wright GC, Dezern SW, MacDonald GE, Anglin NL. Evaluation of linkage disequilibrium, population structure, and genetic diversity in the US peanut mini core collection. BMC Genomics. 2019;20:1–17. https://doi.org/10.1186/s12864-019-5824-9.
49. Zhang X, Larson SR, Gao L, Teh SL, DeHaan LR, Fraser M, Sallam A, Kantarski T, Frels K, Poland J, Wyse D. Uncovering the genetic architecture of seed weight and size in intermediate wheatgrass through linkage and association mapping. Plant Genome. 2017;10:2017–03. https://doi.org/10.3835/plantgenome2017.03.0022.
50. Luo H, Guo J, Ren X, Chen W, Huang L, Zhou X, Chen Y, Liu N, Xiong F, Lei Y, Liao B. Chromosomes A07 and A05 associated with stable and major QTLs for pod weight and size in cultivated peanut (Arachis Hypogaea L). Theor Appl Genet. 2018;131:267–82. https://doi.org/10.1007/s00122-017-3000-7.
51. Gangurde SS, Pasupuleti J, Parmar S, Variath MT, Bomireddy D, Manohar SS, Varshney RK, Singam P, Guo B, Pandey MK. Genetic mapping identifies genomic regions and candidate genes for seed weight and shelling percentage in groundnut. Front Genet. 2013;14:1128182. https://doi.org/10.3389/fgene.2023.1128182.
52. Joshi P, Soni P, Sharma V, Manohar SS, Kumar S, Sharma S, Pasupuleti J, Vadez V, Varshney RK, Pandey MK, Puppala N. Genome-wide mapping of quantitative trait loci for yield-attributing traits of peanut. Genes. 2024;15(2):140.
53. Li QF, Wang JD, Xiong M, Wei K, Zhou P, Huang LC, Zhang CQ, Fan XL, Liu QQ. iTRAQ-based analysis of proteins co-regulated by brassinosteroids and gibberellins in rice embryos during seed germination. International Journal of Molecular Sciences, 2018;19:3460.https://www.mdpi.com/1422-0067/19/11/3460#.

54. Radin I, Mansilla N, Rödel G, Steinebrunner I. The Arabidopsis COX11 homolog is essential for cytochrome c oxidase activity. Front Plant Sci. 2015;6:1091. https://doi.org/10.3389/fpls.2015.01091.
55. Li R, He Y, Chen J, Zheng S, Zhuang C. Research progress in improving photosynthetic efficiency. Int J Mol Sci. 2023;24(11):9286.
56. Kumar A, Pandey SS, Kumar D, Tripathi BN. Genetic manipulation of photosynthesis to enhance crop productivity under changing environmental conditions. Photosynth Res. 2023;155(1):1–21.
57. Gao F, Guo J, Shen Y. Advances from chlorophyll biosynthesis to photosynthetic adaptation, evolution and signaling. Plant Stress. 2024;12(100470):10–1016.
58. Bennett EJ, Roberts JA, Wagstaff C. The role of the pod in seed development: strategies for manipulating yield. New Phytol. 2021;190(4):838–53.
59. Okushima Y, Overvoorde PJ, Arima K, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Lui A, Nguyen D, Onodera C. Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in Arabidopsis thaliana: unique and overlapping functions of ARF7 and ARF19. Plant Cell. 2005;17:444–63. https://doi.org/10.1105/tpc.104.028316.
60. Liu Z. ENO2 affects the seed size and weight by adjusting cytokinin content and forming ENO2-bZIP75 complex in Arabidopsis Thaliana. Front Plant Sci. 2020;11:574316.
61. Zeng R, Chen T, Wang X, Cao J, Li X, Xu X, Chen L, Xia Q, Dong Y, Huang L, Wang L. Physiological and expressional regulation on photosynthesis, starch and sucrose metabolism response to waterlogging stress in peanut. Front Plant Sci. 2021;12:601771.
62. Do THT, Martinoia E, Lee Y, Hwang JU. Update on ATP-binding cassette (ABC) transporters: how they Meet the needs of plants. Plant Physiol. 2021;187(4):1876–92.
63. Luo H, Pandey MK, Khan AW, Wu B, Guo J, Ren X, Zhou X, Chen Y, Chen W, Huang L, Liu N. Next-generation sequencing identified genomic region and diagnostic markers for resistance to bacterial wilt on chromosome B02 in peanut (Arachis Hypogaea L). Plant Biotechnol J. 2019;17(12):2356–69.
64. Ühlken C, Horvath B, Stadler R, Sauer N, Weingartner M. MAIN-LIKE 1 is a crucial factor for correct cell division and differentiation in Arabidopsis Thaliana. Plant J. 2014;78:107–20. https://doi.org/10.1111/tpj.12455.
65. Zhang X, Zhu L, Ren M, Xiang C, Tang X, Xia Y, Song D, Li F. Genome-Wide association studies revealed the genetic loci and candidate genes of Pod-Related traits in peanut (Arachis Hypogaea L). Agronomy. 2023;13:1863. https://doi.org/10.3390/agronomy13071863.
66. Dang P, Patel J, Sorensen R, Lamb M, Chen CY. Genome-wide association analysis identified quantitative trait loci (QTLs) underlying drought-related traits in cultivated peanut (Arachis Hypogaea L). Genes. 2024;15:868. https://www.mdpi.com/2073-4425/15/7/868#.
67. Zhao K, Wang L, Qiu D, Cao Z, Wang K, Li Z, Wang X, Wang J, Ma Q, Cao D, Qi Y. PSW1, an LRR receptor kinase, regulates pod size in peanut. Plant Biotechnol J. 2023;21:2113–24. https://doi.org/10.1111/pbi.14117.
68. Fang Y, Tai Z, Hu K, Luo L, Yang S, Liu M, Xie, X. Comprehensive Review on Plant Cytochrome P450 Evolution: Copy Number, Diversity, and Motif Analysis From Chlorophyta to Dicotyledoneae. Genome biology and evolution.2024;16(11), p.evae240.
69. Huang R, Xiao D, Wang X, Zhan J, Wang A, He L. Genome-wide identification, evolutionary and expression analyses of LEA gene family in peanut (Arachis Hypogaea L). BMC Plant Biol. 2022;22(1):155.
70. Wang Z, Yan L, Chen Y, Wang X, Huai D, Kang Y, Jiang H, Liu K, Lei Y, Liao B. Detection of a major QTL and development of KASP markers for seed weight by combining QTL-seq, QTL-mapping and RNA-seq in peanut. Theor Appl Genet. 2022;135(5):1779–95.
71. Alyr MH, Pallu J, Sambou A, Nguepjop JR, Seye M, Tossim HA, Djiboune YR, Sane D, Rami JF, Fonceka D. Fine-mapping of a wild genomic region involved in pod and seed size reduction on chromosome A07 in peanut (Arachis hypogaea L.). Genes.2020;11(12):1402.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.