



Genome-wide association analysis for drought tolerance and component traits in groundnut gene pool

Seltene Abady · Hussein Shimelis ·
Pasupuleti Janila · Ankush Wankhade ·
Vivek P. Chimote

Received: 10 November 2023 / Accepted: 3 March 2024
© The Author(s) 2024

Abstract The potential production and productivity of groundnuts are limited due to severe drought stress associated with climate change. The current study aimed to identify genomic regions and candidate genes associated with drought tolerance and component traits for gene introgression and to guide marker-assisted breeding of groundnut varieties. Ninety-nine genetically diverse groundnut genotypes were phenotyped under drought-stressed and non-stressed field conditions in 2018/19 and 2019/20, and using the LeasyScan platform under non-stressed conditions in 2019/20 at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)/India. The samples were genotyped using 48 K single nucleotide polymorphisms (SNPs) markers at the University of

Georgia/USA. Phenotypic data was collected on 17 agronomic traits and subjected to statistical analyses. The SNP data were computed, and population structure was inferred using a Bayesian clustering method in Structure version 2.3.4, while linkage disequilibrium was calculated using the GAPIT program in R software. Marker-trait associations were deduced using Tassel 5.2.86. Significant phenotypic variations were recorded for drought tolerance and the assessed agronomic traits. GWAS analysis using PCA + K and Q + K models identified significant SNPs associated with leaf area (1 SNP), leaf area Index (1 SNP), specific leaf area (1 SNP), leaf relative water content (43 SNPs), number of primary branches (1 SNP) and hundred seed weight (1 SNP). Forty-seven and one marker-trait associations were detected under drought-stressed and non-stressed conditions, respectively. The candidate genes and markers identified in the current study are useful for accelerated groundnut breeding targeting drought tolerance and market-preferred traits.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10681-024-03324-3>.

S. Abady (✉) · H. Shimelis
African Centre for Crop Improvement (ACCI), School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Private Bag X01, Scottsville, Pietermaritzburg 3209, South Africa
e-mail: TesfamariamS@ukzn.ac.za

P. Janila · A. Wankhade
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India

V. P. Chimote
State Level Biotechnology Centre, Mahatma Phule Krishi Vidyapeeth, Rahuri 413722, India

Keywords *Arachis hypogaea* · Candidate genes · Drought-stress · GWAS · Single nucleotide polymorphism

Introduction

Groundnut is cultivated on 32.72 million ha, with an annual production of 53.93 million tons worldwide

(FAOSTAT 2023). It is cultivated primarily as a rain-fed crop in the semi-arid tropics and sub-tropical regions where recurrent drought is widespread. The potential production and productivity of groundnut (*Arachis hypogaea* L.; $2n=4x=40$) is limited due to recurrent and severe drought stress associated with climate change. Developing and deploying groundnut varieties with drought tolerance and desirable product profiles is vital for food security and global trade. Groundnuts significantly improve the nutritional status of humankind. Groundnut seeds are rich sources of carbohydrates, protein, lipids, vitamins, minerals and fiber. The seed contains all the essential amino acids making them a critical component of the human diet, especially in communities where animal-derived protein sources are not readily available (Mupunga et al. 2017).

Groundnut is relatively tolerant to drought stress than other traditional leguminous species (Wan et al. 2014). The quantity and quality of grain, seed oil and fodder values of groundnut are affected by drought stress (Abady et al. 2021). Severe drought stress occurring during the reproductive growth stage can lead to a yield loss of up to 33% (Pereira et al. 2016; Carvalho et al. 2017). Drought during grain filling and maturity reduced groundnut's total oil and linoleic acid contents (Dwived et al. 1996). Additionally, drought affects the inherent symbiotic nitrogen fixation capacity of crops, limiting grain yield and quality and affecting the feed quality of the haulm such as its nitrogen content, digestibility and metabolizable energy (Blümmel et al. 2012). Therefore, there is a need to develop and deploy drought-tolerant and locally adapted groundnut varieties to mitigate the effects of drought on groundnut yield and product profiles.

Reportedly, groundnut has marked genetic variability for drought tolerance and component traits for genetic improvement programs (Azevedo et al. 2010; Abady et al. 2021). Several candidate groundnut varieties with improved drought tolerance and agronomic traits have been developed through conventional breeding by the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) and through national breeding programs (Janila et al. 2016). However, the globally pace of new variety design, release and adoption of drought-tolerant groundnut is slow for several reasons. Drought tolerance is a polygenic trait conditioned by multiple

genes with minor genetic effects and is subjected to genotype by the environmental interaction dragging selection gains (Ravi et al. 2011). Furthermore, the genetic base of the cultivated groundnut is narrow, and the introgression of genes from the wild species has limited success due to the ploidy differences providing unpredicted progeny segregations and selection response (Foncéka et al. 2009; Janila et al. 2016). Drought-tolerant and high-yielding varieties have yet to be developed and deployed globally. This is dependent on the identification of agronomic traits associated with drought tolerance and transferring of the genes underlying the target traits to locally adapted genotypes (Edae et al. 2014).

Advanced breeding and genetic innovations such as marker-assisted selection (MAS), genomic selection (GS), and targeted gene editing would accelerate groundnut variety design, and commercialization (Hasan et al. 2021; Pandey et al. 2014). These tools have been valuable in breeding programs to facilitate the identification of drought-resistance genes from germplasm including landraces and wild relatives. The candidate genes can be transferred, pyramided, and fast-tracked in advanced breeding lines (Salgotra and Stewart 2020). Genetic and genomic tools are valuable resources for precision and speed breeding to release drought-resilient and market-preferred varieties.

Identification of genetic markers associated with drought tolerance and economic traits including high kernel yield, oil and oleic acid contents, and haulm yields and quality attributes, disease and insect resistance is crucial for the development new variety with essential traits (Shaibu et al. 2020; Devate et al. 2022). To date, limited genes have been reported based on groundnut genome-wide association studies for drought tolerance (Bertioli et al. 2016). Zhou et al. (2021) identified SNP markers significantly associated with pod and hundred seed weights in groundnut. Shaibu et al. (2020) identified SNP markers for drought surrogate traits using the soil plant analysis development (SPAD) chlorophyll meter reading and leaf area index in groundnut. Zou et al. (2022) identified five SNP markers significantly associated with chlorophyll content in the groundnut. Pandey et al. (2014) identified one SSR marker associated with rust resistance and higher yield in groundnut. There is a limited knowledge on the number of genetic markers and marker-traits

association for drought tolerance and economic traits based on diverse genetic pool of groundnut. Knowledge on marker-traits associations is crucial for marker-assisted selection, trait integration and precision breeding in groundnut. Thus, the objective of the current study was to identify genomic regions and candidate genes associated with drought tolerance and component traits for gene introgression, and to guide marker-assisted breeding of drought-tolerant groundnut varieties.

Material and methods

Plant material

Ninety-nine genetically diverse groundnut genotypes acquired from ICRISAT in Patancheru, India were used for the study. The genotypes were selected based on desirable traits, including drought tolerance, resistance to foliar diseases such as late leaf spot and rust, high oil and oleic acid contents, and early-to-medium maturation. This study used a high-yielding groundnut cultivar ICGV98412 released in India, Ghana and Ethiopia as a comparative control. The details of the genotypes are described in Supplementary Table 1. The genotypes were evaluated under drought-stressed (DS) and non-stressed (NS) conditions at ICRISAT (latitude, 17.51°N, longitude, 78.27°E, and altitude 545 m) during the 2018/2019 and 2019/2020 post-rainy cropping seasons using a 10×10 alpha lattice design with two replications. The plants were phenotyped in five environments, including four experiments [drought-stressed and non-stressed conditions in two seasons (2018/19 and 2019/20)] under field conditions and using the leasyScan platform under non-stressed conditions with four replications.

Phenotypic evaluation and data analysis

Phenotypic data were collected on days to 50% flowering (DF), chlorophyll meter reading (SCMR), specific leaf area ($\text{cm}^2 \text{g}^{-1}$), leaf relative water content (LRWC), plant height (PH, expressed in cm), number of primary branches (PB), pod yield per plant (PY, expressed in g plant^{-1}), shelling percentage (SHP, expressed in %), seed yield per plant (SY, expressed in g plant^{-1}), total biomass per plant (TBM, expressed

in g plant^{-1}) and harvest index (HI) (%). From the LeasyScan experiment, leaf area (LA), projected leaf area (PLA), leaf area index (LAI), and light penetration depth (LPD), digital plant height (DPH) and digital biomass (DBM) data were collected. The phenotypic data were subjected to analysis of variance. The homogeneity of error variances was tested using the Bartlett test before pooled analysis of variance. The means of the treatments were separated using the least significant difference (LSD) procedure at the 5% significant level.

Genotyping

The 99 groundnut genotypes were grown under field conditions at ICRISAT, Hyderabad, India. Genomic DNA was extracted from the leaves of three weeks old seedlings at the Center of Excellence in Genomics and Systems Biology at ICRISAT. The DNA was extracted using the modified cetyl trimethyl ammonium bromide (CTAB) method (Mace et al. 2003). DNA was mixed with a loading dye and quantified by loading 1 μl DNA on the 0.8% (w/v) agarose gel containing 10 μl ethidium bromide (10 mg/ml) and run at 80 V for 30–45 min. Subsequently, the DNA was visualized under a UV transilluminator (Bio-Rad Universal Hood II Gel Doc System). DNA quality and concentration were estimated using NanoDrop Spectrometry (UV 160 A, Japan). A DNA sample of 47 ng/ μl per genotype was submitted for genotyping. The DNA samples were genotyped with a 48 K Affymetrix SNP array ('Axiom_Arachis') (Wankhade et al. 2023).

SNP data were analyzed using the Axiom analysis suite (Thermo 2018). SNP markers with more than 20% of missing data and minor allele frequencies lower than 0.05 were eliminated. This resulted in 15,575 SNP markers, which were used for further analysis. Ninety-nine genotypes were used after the data imputation. The genotype data filtering was performed using TASSEL version 5.2.86 software.

Population structure and principal component analysis

The population structure pattern and admixture detection were inferred using a Bayesian model-based clustering algorithm implemented in STRU CTURE version 2.3.4 (Pritchard et al. 2000). The

length of the burn-in period and Markov Chain Monte Carlo (MCMC) were set at 10,000 iterations (Evanno et al. 2005). The K value was set between 1 and 10 to generate the number of subpopulations in the genotypes. Twenty runs were performed for each K-value to accurately estimate the number of populations. Delta K values were calculated, and the appropriate K value was determined by the Evanno et al. (2005) method using the STRUCTURE Harvester program (Earl et al. 2012). SNP marker-based PCA and kinship analysis were subsequently conducted with GAPIT (Lipka et al. 2012).

Genome-wide association analysis (GWAS)

GWAS was performed with Tassel 5.2.86. Six models were evaluated in the marker-trait association analysis, including the naïve, Q, K, PCA, PCA+K, and Q+K. Association signals were observed on PCA+K and Q+K models using a mixed linear model (MLM). Quantile–quantile (QQ) plots were presented with $-\log_{10}(P)$ of each SNP and expected P value, and the Manhattan plots were generated using TASSEL 5.2.86. Marker-trait association with or above 20% phenotypic variance explained (PVE) was considered to be a major association. Candidate genes covering major SNPs within a 50 kb region upstream or downstream of peak SNPs were selected from the PeanutBase website tool (<https://www.peanutbase.org>).

Linkage disequilibrium (LD) and decay

The LD between polymorphic SNPs retained after filtering at a cutoff of MAF 0.05, 0.1, and 0.2 was calculated in the form of r^2 using TASSEL 5.2.86. LD decay plots were generated using the R script written by Remington et al. (2001) using R Studio (2021.09.0 Build 351© 2009–2021 R Studio, PBC).

Results

Phenotypic variation

Significant genetic variation were recorded for yield and yield components among the tested groundnut genotypes evaluated under drought-stressed and non-stressed conditions (Abady et al. 2021). Analysis of

variance for canopy-related traits phenotyped using LeasyScan platform showed highly significant ($P < 0.001$) genotype differences (Table 1). Mean performance of the groundnut genotypes for 13 phenotypic traits under drought-stressed and non-stressed conditions, and six canopy-related traits under non-stressed conditions are presented in Supplementary Tables 2, 3, and 4 in that order. Wide phenotypic variations existed for all the assessed traits.

Population structure, principal component analysis (PCA) and linkage disequilibrium (LD)

Population structure analysis of the 99 groundnut genotypes resolved three sub-populations with 32% admixture genotypes (Abady et al. 2021). Allocation into clusters was done at 70% ancestry. Twenty-four, 22 and 21% of the genotypes were assigned to sub-populations 1, 2, and 3, respectively. The PCA based on SNP marker data also confirmed the presence of three subgroups, corresponding with the population structure results (Fig. 1). The first three principal components accounted for 32% of the total variation (Fig. 1a) and revealed three distinct clusters in the population (Fig. 1b).

Three different threshold cutoff levels of MAF, i.e., 0.05, 0.1 and 0.2 were used to explore the effect of minor alleles on the nature and decay of genome-wide LD and resented in Fig. 2a, b and c in that order. LD was found to be decreasing with increasing bin distance. LD declined to half of its original value in three different threshold cutoff levels of MAF, i.e., 0.05, 0.1 and 0.2 was 3.98, 6.33 and 14.48 Mb, respectively.

Marker trait association

Forty-seven and 13 SNP markers were significantly associated with DLA, LAI, SLA, LRWC, PB and HSW, were identified using PCA+K and Q+K models, respectively (Table 2). Among the significantly associated SNP markers, nine were identified by both models. Forty-five SNP markers were significantly associated with LRWC and seven SNP markers were associated with one or two traits. Thus, in this study, 50 SNP markers were identified (Table 2 and Fig. 3) Graphical representation of significant SNPs

Table 1 Descriptive statistics, mean squares and significant tests for six phenotypic traits among 99 groundnut genotypes evaluated under LeasyScan non-stressed conditions during the 2019 rainy season at ICRISAT, Patancheru, India

Traits	Min	Max	Range	Mean	SEM	Skewness	Kurtosis	Genotype(df=98)	Rep (df=3)	Error (df=284)	CV (%)
DLA	11,413	47,438	36,025	29,469.04	2401	-0.11	-0.35	1.41E+08**	4.67E+08**	5,769,052	8.15
PLA	7378	30,171	22,793	19,274.07	1503.4	-0.17	-0.32	60,110,363**	1.79E+08**	2,260,234	7.8
LAI	0.04	0.17	0.13	0.1	0.01	-0.01	-0.23	0.001844**	0.00853**	0.00012567	10.83
LPD	26.7	61.3	34.6	44	1.73	0.14	0.08	129,2217**	238.11**	2.99	3.93
DPH	41	83	42	61.45	2.12	0.1	-0.01	192,6547*8	321.1559**	4.51	3.45
DBM	657,610	3,469,740	2,812,130	1,974,771.5	171,118.9	0.12	-0.43	1.08E+12**	4.09E+12**	29,305,653,273	8.66

*Significant at the 0.01 probability level; ns, non-significant

DLA Digital leaf area (mm²), PLA Projected leaf area (mm²), LAI Leaf area index, LPD Light penetration depth (mm), DPH Digital plant height (mm), DBM Digital biomass (mm³), Min. Minimum, Max. Maximum, SEM Standard error of the mean, Rep Replications, df Degree of freedoms

identified for the assessed traits were depicted with a Manhattan map along with QQ plots (Fig. 3). The QQ plots showed that the deviation between observed and expected P values was very small, suggesting a true positive association between the SNPs and the traits.

Traits

The GWAS output identified one SNP marker with significant association with both DLA and LAI under non-stressed (NS) conditions using the Q+K model (Fig. 3a and c). The phenotypic variance of these traits explained by the marker was 21 and 20%, respectively. PCA+K and Q+K models detected one SNP marker significantly associated with SLA under drought-stressed (DS) conditions (Fig. 3e and f). The phenotypic variance of SLA explained by the significant marker ranged from 22 to 23%. Similarly, PCA+K and Q+K models detected two SNPs markers with significant association with PB under DS (Fig. 3i and j) and NS (Fig. 3k and l) conditions. The phenotypic variance of the trait explained by the markers ranged from 20 to 23%. The study identified 43 SNPs with significant association with LRWC under DS conditions through either the PCA+K or Q+K or, both models (Table 2, Fig. 3g and h). The phenotypic variance of the trait explained by the significant SNPs ranged from 20 to 31%. Further, the GWAS analysis detected one SNP significant association with HSW under DS conditions. Both PCA+K and Q+K models were identified for this marker (Fig. 3m and n). The phenotypic variance of the trait explained by the significant SNP marker ranged from 28 to 31%.

Discussion

Phenotypic variability

Drought is the leading abiotic stress, which limits groundnut production and productivity globally. Significant progress were reported on groundnut pre-breeding for drought tolerance through conventional breeding methods (Janila et al., 2016). However, the pace of drought tolerance breeding and variety release has been slow due to the complex nature of gene action and the genotype by environment by management interaction effect (Ravi et al. 2011). Deploying

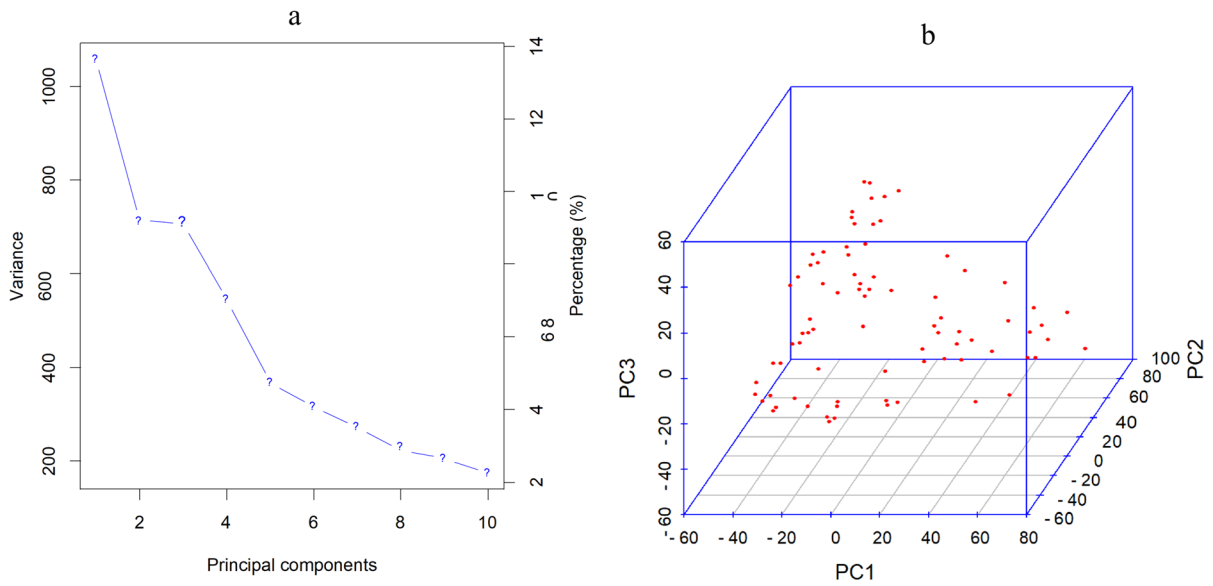


Fig. 1 Principal component analysis of the 99 groundnut genotypes based on 15,575 high-quality SNPs with $MAF > 0.05$ using the first three principal components. The first three prin-

incipal components indicated 32% of the variation as indicated on the scree plot (a). The genotypes were grouped into three distinct clusters (b)

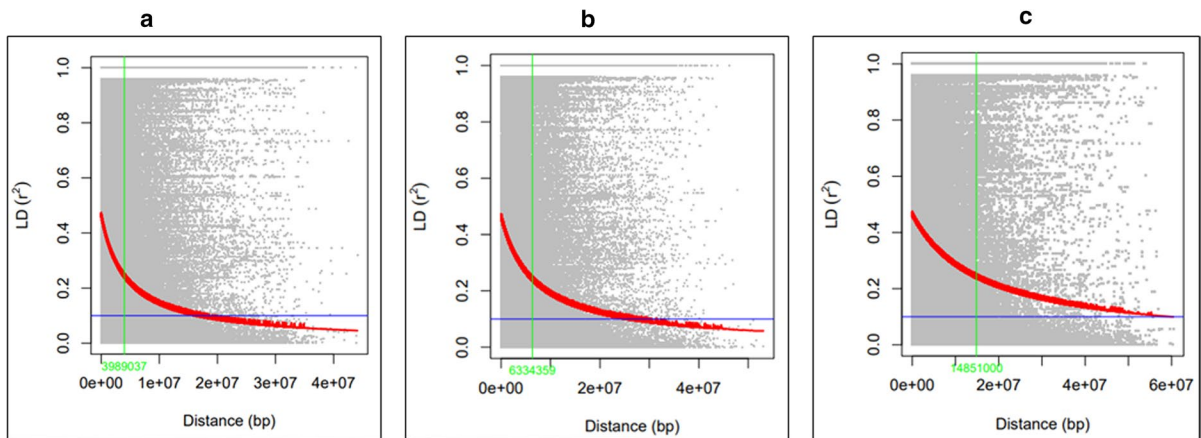


Fig. 2 Effect of three MAFs, 0.05 (a), 0.1 (b) and 0.2 (c), on the nature of LDs and their decay in advanced breeding lines

drought surrogate traits is critical for effective drought tolerance breeding in crop genetic resources, including groundnut. Furthermore, understanding the genetic base of physiological and yield-related traits in groundnut could provide an opportunity to develop drought-tolerant cultivars (Pereira et al. 2016). Wankhade et al. (2022) proposed an integrated

phenotyping approach for screening groundnut genotypes for drought tolerance. The authors reported early generation selection gains using the LeasyScan method with complementary drought stress indices under managed stress environment.

The present study revealed wide genetic variability for the assessed physiological and yield-related traits

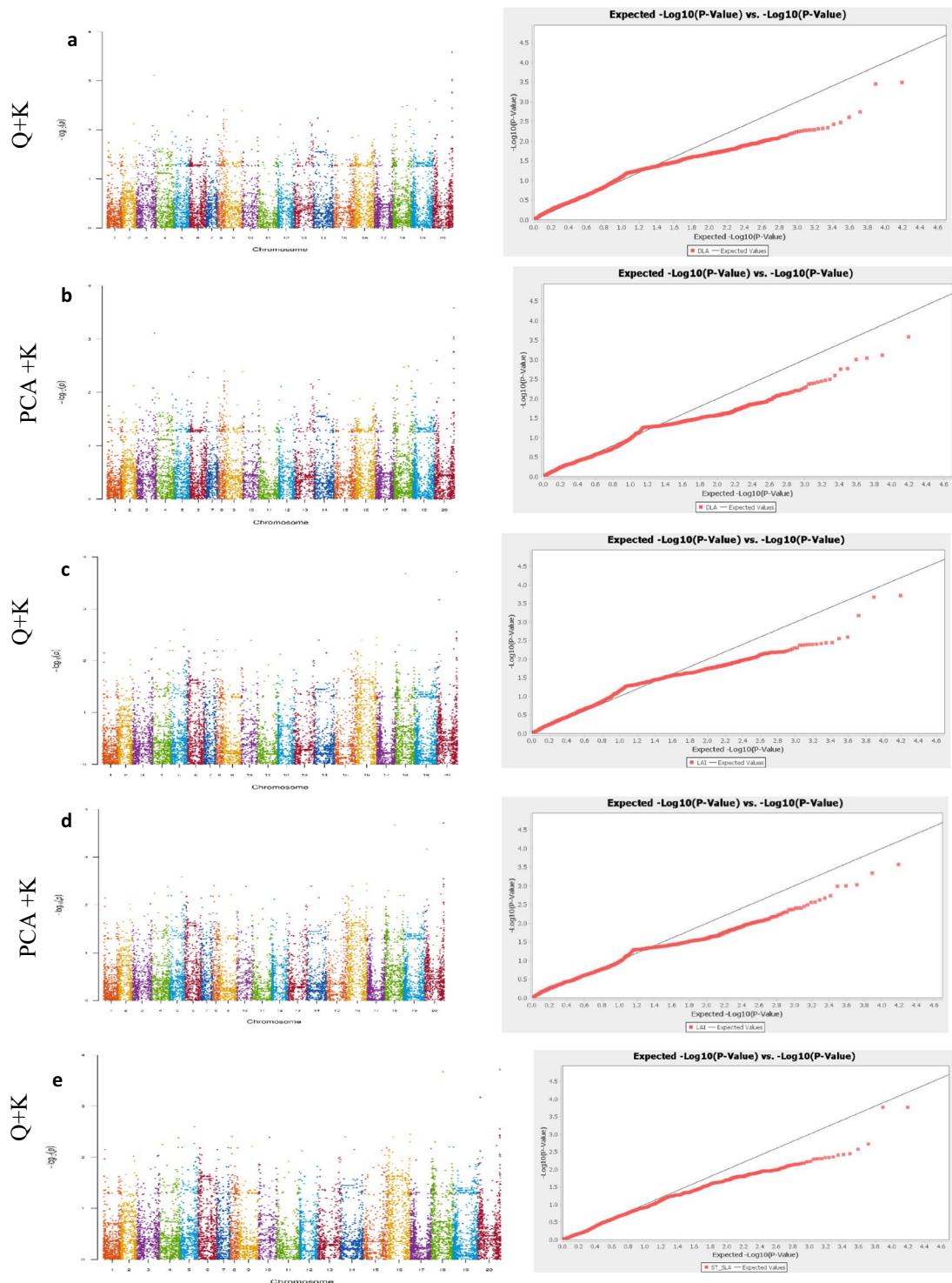


Fig. 3 Manhattan map and QQ plots showing SNP markers associated with different agronomic traits among 99 groundnut genotypes based on the PCA + K and Q + K models. Note: **a**) and **b**) denote digital leaf area under non-stressed (NS) conditions; **c**) and **d**) digital leaf area index under NS conditions; **e**)

and **f**) specific leaf area under NS conditions; **g**) and **h**) leaf relative water content under drought-stressed (DS) conditions; **i**) and **j**) number of primary branches under DS conditions; **k**) and **l**) number of primary branches under NS conditions; **m**) and **n**) hundred seed weight under DS conditions

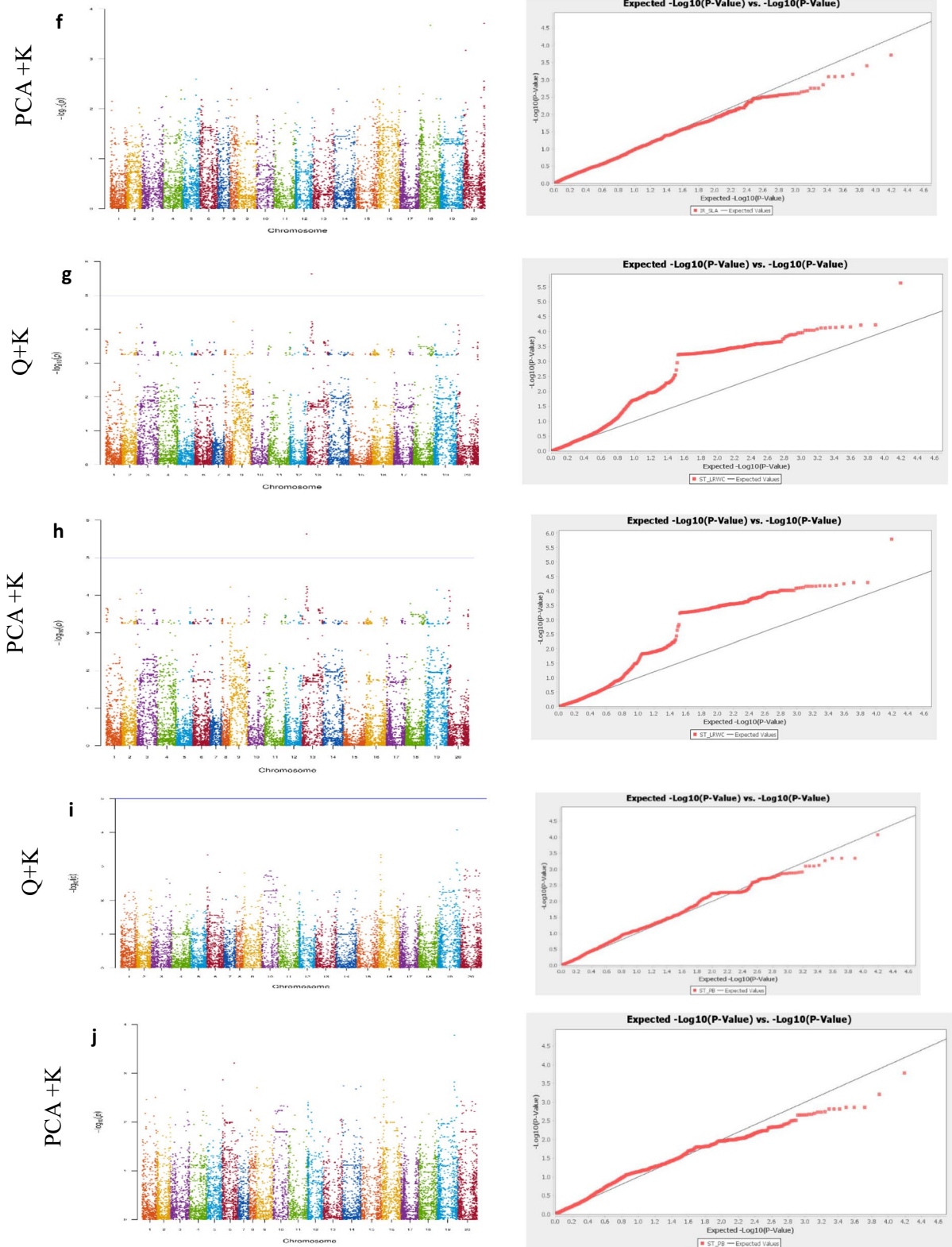


Fig. 3 (continued)

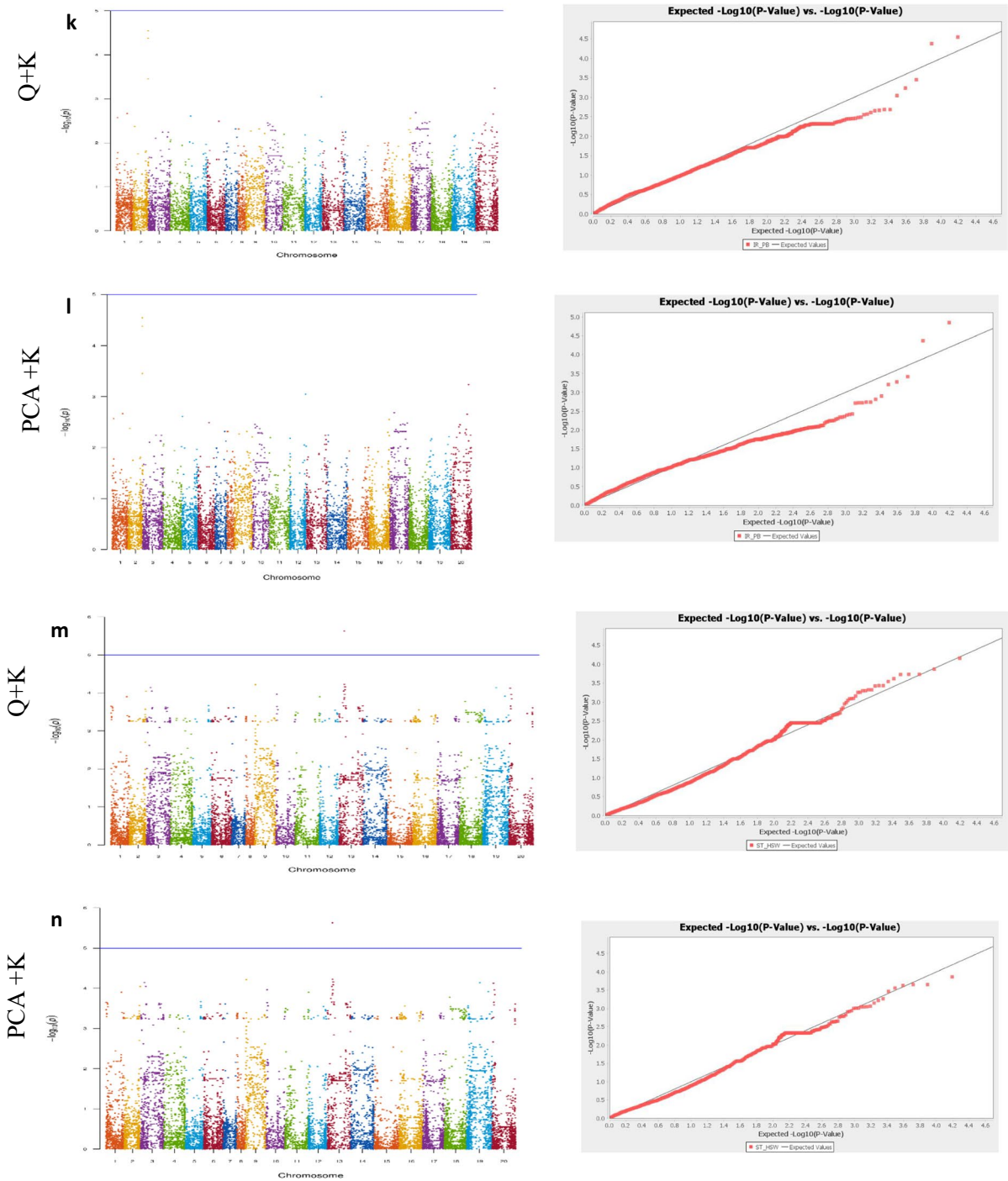


Fig. 3 (continued)

among the tested groundnut genotypes which were evaluated under drought-stressed and non-stressed conditions (Abady et al. 2021). The analysis of variance

revealed highly significant genotypic differences for canopy-related traits, including digital leaf area (LA), digital leaf area index (LAI), specific leaf area (SLA),

Table 2 Association mapping with PCA + K and Q + K models identified marker-trait associations and predicted genes within 50 kb of marker positions

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
1	PCA + k	ST_LRWC	AX-176794990	B03	29,320,209	5.792303	31.68378	45	Araip.9NG64	Araip. B03:29,297,132..29298526	RNA-binding protein 24-like [Glycine max]; IPR012677 (Nucleotide-binding, alpha-beta plat); GO:0000166 (nucleotide binding), GO:0003676 (nucleic acid binding)
2	PCA + k	ST_HSW	AX-177639302	B09	26,718,833	3.628884	28.39494	4.05	Araip.88U3C	Araip. B09:26,715,247..26716988	cleavage and polyadenylation specificity factor 5; IPR016706 (Cleavage/ polyadenylation specificity factor subunit 5); GO:0003729 (mRNA binding), GO:0005849 (mRNA cleavage factor complex), GO:0006378 (mRNA polyadenylation)
3	PCA + k	IR_PB	AX-176817979	A02	92,194,198	4.854479	25.89139	7.3	Aradu.ML3P3	Aradu. A02:92,187,308..92190172	P-type ATPase of Arabidopsis 2
4	PCA + k	IR_PB	AX-176797883	A02	92,259,164	4.369099	23.96569	7.2	Aradu.WRQ70	Aradu. A02:92,249,930..92255793	uncharacterized protein LOC102661917 isoform X2 [Glycine max]; IPR012337 (Ribonuclease H-like domain); GO:0003676 (nucleic acid binding)
5	PCA + k	IR_SLA	AX-176803565	A01	103,143,552	3.718735	23.7956	1	Aradu.YV19K	Aradu. A01:103,138,366..103141551	uncharacterized protein LOC100789274 [Glycine max]; IPR010341 (Protein of unknown function DUF936, plant)

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
6	PCA+k	ST_LRWC	AX-147244306	B03	29,871,778	4.24613	22.73632	1	Araip.4J8RL	Araip. B03:29,866,672..29871890	polynucleotide phosphatase/kinase; IPR001510 (Zinc finger, PARG-type), IPR013954 (Polynucleotide kinase 3 phosphatase), IPR023214 (HAD-like domain); GO:0003677 (DNA binding), GO:0008270 (zinc ion binding)
7	PCA+k	ST_LRWC	AX-176816874	B03	32,547,071	4.289175	22.54966	20	Araip.57P4D	Araip. B03:32,555,495..32561902	Glycosyl hydrolase family protein with chitinase insertion domain; IPR017853 (Glycoside hydrolase, superfamily); GO:0004568 (chitinase activity), GO:0005975 (carbohydrate metabolic process), GO:0006032 (chitin catabolic process)
8	PCA+k	ST_LRWC	AX-147244415	B03	34,773,499	4.289175	22.54966	1	Araip.SVH5H	Araip. B03:34,772,161..34773791	white-brown-complex ABC transporter family protein
9	PCA+k	ST_LRWC	AX-176809832	A03	22,947,365	4.194627	21.99801	9	Aradu.56ICC	Aradu. A03:22,937,241..22942399	electron transfer flavoprotein subunit alpha; IPR001308 (Electron transfer flavoprotein, alpha subunit); GO:0009055 (electron carrier activity), GO:0050660 (flavin adenine dinucleotide binding)

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
10	PCA+k	ST_LRWC	AX-177641299	B09	12,571,709	4.163747	21.93233	45	Araip.4.A1AY	Araip. B09:12,539,449..12540921	glucosylmannan 4-beta-mannosyltransferase 2-like [Glycine max]
11	PCA+k	ST_LRWC	AX-177637534	B09	11,016,521	4.175475	21.88659	10	Araip.XKV67	Araip. B09:11,009,359..11010991	transcription factor PER1ANTHIA-like [Glycine max]; IPR004827 (Basic-leucine zipper domain); GO:0003700 (sequence-specific DNA binding transcription factor activity); GO:0043565 (sequence-specific DNA binding)
12	PCA+k	ST_LRWC	AX-177643146	B09	12,518,764	4.175475	21.88659	17.82	Araip.P5TWT	Araip. B09:12,508,402..12510210	protein N-lysine methyltransferase METTL21A-like isoform XI [Glycine max]
13	PCA+k	ST_LRWC	AX-176794442	B09	13,151,882	4.175475	21.88659	1.6	Araip.N6QX9	Araip. B09:13,152,433..13156413	E3 ubiquitin-protein ligase CHIP-like [Glycine max]; IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
14	PCA+k	ST_LRWC	AX-147216599	A03	30,565,512	4.16011	21.79728	1	Aradu.35J09	Aradu. A03:30,563,780..30566010	myo-inositol oxygenase 2; IPR007828 (Inositol oxygenase); GO:0005506 (iron ion binding), GO:0005737 (cytoplasm), GO:0019310 (inositol catabolic process), GO:0050113 (inositol oxygenase activity), GO:0055114 (oxidation-reduction process) pentatricopeptide repeat-containing protein At4g04790, mitochondrial-like isoform X3 [Glycine max
15	PCA+k	ST_LRWC	AX-147244175	B03	25,355,107	4.16011	21.79728	1	Araip.Y9RQ5	Araip. B03:25,354,426..25359616	ethylene-responsive transcription factor 4 [Glycine max]; IPR008942 (ENTH/HS), IPR016177 (DNA-binding domain); GO:0003677 (DNA binding), GO:0003700 (sequence-specific DNA binding transcription factor activity)
16	PCA+k	ST_LRWC	AX-147244210	B03	26,215,510	4.16011	21.79728	1	Araip.ZS7TB	Araip. B03:26,213,405..26215873	

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
17	PCA+k	ST_LRWC	AX-176814526	B03	31,644,259	4.097391	21.78212	10	Araip.Z05PC	Araip. B03:31,633,649..31639039	Protein phosphatase 2C family protein; IPR001932 (Protein phosphatase 2C (PP2C)-like domain), IPR015655 (Protein phosphatase 2C); GO:0003824 (catalytic activity)
18	PCA+k	ST_LRWC	AX-176808224	A02	91,075,199	4.113854	21.52883	6.6	Aradu.6K8RX	Aradu. A02:91,078,341..91080561	Pentatricopeptide repeat (PPR) superfamily protein; IPR002885 (Pentatricopeptide repeat), IPR011990 (Tetratricopeptide-like helical); GO:0005515 (protein binding)
19	PCA+k	ST_LRWC	AX-176808762	B03	38,111,010	4.108408	21.49726	25.78	Araip.XZL7Y	Araip. B03:38,093,901..38096941	probable pectinesterase/pectinesterase inhibitor 6-like [Glycine max]; IPR006501 (Pectinesterase inhibitor domain), IPR011050 (Pectin lyase fold/vinulence factor); GO:0004857 (enzyme inhibitor activity), GO:0005618 (cell wall), GO:0030599 (pectinesterase activity), GO:0042545 (cell wall modification)

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
20	PCA+k	ST_LRWC	AX-176823645	B03	33,575,120	4.091025	21.39657	13.12	Araip.VHP1S	Araip. B03:33,564,599..33569421	hexokinase I; IPR001312 (Hexokinase); GO:0005524 (ATP binding), GO:0005975 (carbohydrate metabolic process)
21	PCA+k	ST_LRWC	AX-177640271	B08	127,304,285	3.951915	21.07514	5.84	Araip.NS4X8	Araip. B08:127,296,958..127302290	MACPF domain- containing protein A14g24290-like isoform X3 [Glycine max]; IPR020864 (Membrane attack complex component/perforin (MACPF) domain)
22	PCA+k	ST_LRWC	AX-147244325	B03	30,517,838	3.948692	20.99685	1	Araip.KTN2B	Araip. B03:30,514,008..30524471	zinc finger, C3HC4 type (RING finger) protein; IPR013083 (Zinc finger, RING/ FYVE/PHD-type); GO:0005515 (protein binding), GO:0008270 (zinc ion binding)
23	PCA+k	ST_LRWC	AX-176795390	A03	32,267,174	4.015988	20.96291	36.45	Aradu.GA1C4	Aradu. A03:32,283,542..32284090	FAR1 DNA-binding domain protein; IPR004330 (FAR1 DNA binding domain)
24	PCA+k	ST_LRWC	AX-176804539	A03	106,338,595	4.015988	20.96291	49.5	Aradu.98BQH	Aradu. A03:106,303,427..106310412	DNA binding; ATP binding; IPR003594 (Histidine kinase- like ATPase, ATP- binding domain); GO:0005524 (ATP binding)
25	PCA+k	ST_LRWC	AX-176801328	A03	110,853,974	4.015988	20.96291	1	Araip.J235V	Araip. B03:17,440,104..17443010	APO RNA-binding protein; IPR023342 (APO domain); GO:0003723 (RNA binding)

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
26	PCA+k	ST_LRWC	AX-176822255	B03	39,225,929	4.015988	20.96291	36.45	Araip.WEZIZ	Araip. B03:39,241,101..39242598	WRKY family transcription factor; IPR003657 (DNA-binding WRKY), IPR018872 (Zn-cluster domain); GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0043565 (sequence-specific DNA binding)
27	PCA+k	ST_LRWC	AX-176824188	B03	107,874,365	4.015988	20.96291	10.17	Araip.U28W7	Araip. B03:107,861,685..107868998	DNA binding; ATP binding; IPR003594 (Histidine kinase-like ATPase, ATP-binding domain); GO:0005524 (ATP binding)
28	PCA+k	ST_LRWC	AX-176819162	B03	110,012,359	4.015988	20.96291	14.43	Araip.PN6I4	Araip. B03:110,019,908..110024050	unknown protein; Has 84 Blast hits to 81 proteins in 31 species: Archae—0; Bacteria—0; Metazoa—42; Fungi—0; Plants—41; Viruses—0; Other Eukaryotes—1 (source: NCBI BLINK)

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
29	PCA+k	ST_LRWC	AX-176821311	B03	111,532,428	4.015988	20.96291	40.5	Araip.Y13UG	Araip. B03:111,548,638..111554664	diacylglycerol kinase 2; IPR000756 (Diacylglycerol kinase, accessory domain), IPR001206 (Diacylglycerol kinase, catalytic domain), IPR002219 (Protein kinase C-like, phorbol ester/ diacylglycerol binding), IPR016064 (ATP- NAD kinase- like domain); GO:0003951 (NAD+ kinase activity), GO:0004143 (diacylglycerol kinase activity), GO:0007205 (protein kinase C-activating G-protein coupled receptor signaling pathway), GO:0008152 (metabolic process), GO:0035556 (intracellular signal transduction)

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
30	PCA+k	ST_LRWC	AX-147245090	B03	115,102,239	4.015988	20.96291	1	Araip.S31HP	Araip. B03:115,102,056..115108203	cellulose synthase 6; IPR005150 (Cellulose synthase), IPR013083 (Zinc finger, RING/ FYVE/PHD-type); GO:0016020 (membrane), GO:0016760 (cellulose synthase (UDP-forming) activity), GO:0030244 (cellulose biosynthetic process)
31	PCA+k	ST_LRWC	AX-176823852	B03	117,059,487	4.015988	20.96291	64.95	Araip.R6RNX	Araip. B03:117,087,747..117096168	regulation of nuclear pre-mRNA domain-containing protein 1B-like isoform X4 [Glycine max]; IPR006569 (CID domain), IPR008942 (ENTH/VHS)
32	PCA+k	ST_LRWC	AX-176823026	B03	106,858,732	3.991245	20.94642	18	Araip.I1331	Araip. B03:106,848,944..106850432	serine/threonine-protein phosphatase 7 long form homolog [Glycine max]; IPR019557 (Aminotransferase-like, plant mobile domain)
33	PCA+k	ST_LRWC	AX-177642688	B09	60,912,926	3.506374	20.91062	35.07	Araip.MQ358	Araip. B09:60,930,794..60931914	calcium-dependent protein kinase 16; IPR011992 (EF-hand domain pair); GO:0005509 (calcium ion binding)
34	PCA+k	ST_LRWC	AX-176816985	B03	114,953,493	3.963163	20.65861	7.29	Araip.X4M23	Araip. B03:114,948,098..114950666	Unknown protein
35	PCA+k	ST_LRWC	AX-176800334	B03	116,725,590	3.963163	20.65861	32.21	Araip.XQ9YQ	Araip. B03:116,704,526..116709229	unknown protein

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
36	PCA+k	ST_LRWC	AX-176810103	B03	20,375,224	3.962957	20.65743	6.56	Araip.MK8EH	Araip. B03:20,377,934..20382008	Core-2/1-branching beta-1,6-N-acetylglucosaminyltransferase family protein; IPR003406 (Glycosyl transferase, family 14); GO:0008375 (acetylglucosaminyltransferase activity), GO:0016020 (membrane)
37	PCA+k	ST_LRWC	AX-176795936	A03	34,012,289	3.962161	20.65284	20	Aradu.JFP11	Aradu. A03:34,022,664..34023983	Transketolase; IPR009014 (Transketolase, C-terminal/Pyruvate-ferredoxin oxidoreductase, domain II); GO:0003824 (catalytic activity), GO:0008152 (metabolic process) unknown protein
38	PCA+k	ST_LRWC	AX-176819367	B03	36,814,313	3.962161	20.65284	3.21	Araip.NL7RX	Araip. B03:36,815,404..36816054	stress enhanced protein 2, chloroplastic-like [Glycine max]
39	PCA+k	ST_LRWC	AX-176814346	A03	114,447,106	3.917219	20.55669	20	Aradu.F7E59	Aradu. A03:114,433,849..114434528	APO RNA-binding protein; IPR023342 (APO domain); GO:0003723 (RNA binding)
40	PCA+k	ST_LRWC	AX-147217283	A03	107,970,504	3.937764	20.51259	1	Araip.J235V	Araip. B03:17,440,104..17443010	

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
41	PCA+k	ST_LRWC	AX-176820569	B03	110,279,421	3.937764	20.51259	5	Araip.G1168	Araip. B03:110,271,121..1110276675	tyrosyl-DNA phosphodiesterase- related; IPR008984 (SMAD/ FHA domain), IPR010347 (Tyrosyl-DNA phosphodiesterase 1), IPR014905 (HIP116, Rad5p N-terminal), IPR027415 (Tyrosyl-DNA phosphodiesterase C-terminal domain); GO:0003676 (nucleic acid binding), GO:0005515 (protein binding), GO:0005634 (nucleus), GO:0006281 (DNA repair), GO:0008081 (phosphoric diester hydrolase activity), GO:0008270 (zinc ion binding)
42	PCA+k	ST_LRWC	AX-176820403	B03	111,032,086	3.937764	20.51259	6.05	Araip.A10SM	Araip. B03:111,035,157..1111036796	serine/threonine- protein phosphatase 7 long form homolog [Glycine max]; IPR019557 (Aminotransferase- like, plant mobile domain)

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
43	PCA+k	ST_LRWC	AX-176816848	B03	24,967,473	3.457547	20.50669	2.95	Araip.KLX3M	Araip. B03:24,961,776..24966652	RNA polymerase sigma factor; IPR014284 (RNA polymerase sigma-70 like domain); GO:0003677 (DNA binding), GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0016987 (sigma factor activity)
44	PCA+k	ST_LRWC	AX-176821705	B03	28,587,190	3.887135	20.32597	20	Araip.23WVH	Araip. B03:28,595,377..28596251	UDP-Glycosyltransferase superfamily protein; IPR002213 (UDP-glucuronosyl/UDP-glucosyltransferase); GO:0008152 (metabolic process)
45	PCA+k	ST_PB	AX-177640364	B09	123,045,228	3.782608	20.28576	45	Araip.S2FNK	Araip. B09:123,014,823..123017842	ribosome-binding factor, A family protein; IPR000238 (Ribosome-binding factor A), IPR015946 (K homology domain-like, alpha/beta); GO:0006364 (rRNA processing)
46	PCA+k	ST_LRWC	AX-147235264	A10	5,693,627	3.894073	20.26184	1	Aradu.PKW10	Aradu.A10:5,690,513..5695190	CD2 antigen cytoplasmic tail-binding-like protein
47	PCA+k	ST_LRWC	AX-147217633	A03	118,847,253	3.870841	20.12873	1	Aradu.LN5U5	Aradu. A03:118,846,024..118848320	epoxide hydrolase; IPR000073 (Alpha/beta hydrolase fold-1), IPR000639 (Epoxide hydrolase-like); GO:0003824 (catalytic activity)

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
48	Q+K	ST_HSW	AX-177639302	B09	26,718,833	4.214745	31.26693	4.05	Araip.88U3C	Araip. B09:26,715,247..267,16988	cleavage and polyadenylation specificity factor 5; IPR016706 (Cleavage/ polyadenylation specificity factor subunit 5); GO:0003729 (mRNA binding), GO:0005849 (mRNA cleavage factor complex), GO:0006378 (mRNA polyadenylation)
49	Q+K	ST_LRWC	AX-176794990	B03	29,320,209	5.661677	30.56852	45	Araip.9NG64	Araip. B03:29,297,132..292,98526	RNA-binding protein 24-like [Glycine max]; IPR012677 (Nucleotide-binding, alpha-beta plait); GO:000166 (nucleotide binding), GO:0003676 (nucleic acid binding)
50	Q+K	ST_PB	AX-177640364	B09	123,045,228	4.298727	23.37434	45	Araip.S2FNK	Araip. B09:123,014,823..12,301,7842	ribosome-binding factor A family protein; IPR000238 (Ribosome-binding factor A), IPR015946 (K homology domain-like, alpha/beta); GO:0006364 (rRNA processing)
51	Q+K	IR_SLA	AX-176803565	A01	103,143,552	3.757097	22.75878	1	Aradu.YV19K	Aradu. A01:103,138,366..10314,1551	uncharacterized protein LOC100789274 [Glycine max]; IPR010341 (Protein of unknown function DUF936, plant)

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
52	Q+K	IR_PB	AX-176797883	A02	92,259,164	4.198361	22.74301	7.2	Aradu.WRQ70	Aradu. A02:92,249,930..92255793	uncharacterized protein LOC102661917 isoform X2 [Glycine max], IPR012337 (Ribonuclease H-like domain); GO:0003676 (nucleic acid binding)
53	Q+K	ST_LRWC	AX-176816848	B03	24,967,473	3.684408	21.81545	2.95	Araip.KLX3M	Araip. B03:24,961,776..24966652	RNA polymerase sigma factor; IPR014284 (RNA polymerase sigma-70 like domain); GO:0003677 (DNA binding), GO:0003700 (sequence-specific DNA binding transcription factor activity), GO:0016987 (sigma factor activity)
54	Q+K	DLA	AX-177643135	B10	14,792,775	3.590692	21.13553	42.87	Araip.6RA6D	Araip. B10:14,761,107..14763635	Expressed protein n = 2 Tax = Oryza sativa subsp. japonica ReplID = QR6F5_ ORYSJ
55	Q+K	LAI	AX-177641314	B08	67,546,583	3.51811	21.02522				P-type ATPase of Arabidopsis 2
56	Q+K	IR_PB	AX-176817979	A02	92,194,198	3.980584	20.61059	7.3	Aradu.ML3P3	Aradu. A02:92,187,308..92190172	zinc finger MYM-type protein 1-like [Glycine max]
57	Q+K	ST_LRWC	AX-176798839	A05	91,532,176	3.942134	20.49478	17.29	Aradu.VIU0I	Aradu. A05:91,521,255..91524561	
58	Q+K	LAI	AX-177643135	B10	14,792,775	3.464072	20.05719	42.87	Araip.6RA6D	Araip. B10:14,761,107..14763635	Expressed protein n = 2 Tax = Oryza sativa subsp. japonica ReplID = QR6F5_ ORYSJ
59	Q+K	ST_LRWC	AX-147235264	A10	5,693,627	3.881946	20.05087	1	Aradu.PKW10	Aradu.A10:5,690,513..5695190	CD2 antigen cytoplasmic tail-binding-like protein

Table 2 (continued)

Sr. No	Model	Trait	Marker	Chromosome	Position (bp)	-10Log P	PVE%	Within (kb)	Gene Model	Gene Position (bp)	Annotation
60	Q + K	ST_LRWC	AX-147244306	B03	29,871,778	3.814704	20.04224	1	Araip_4J8RL	Araip. B03:29,866,672..29871890	polynucleotide phosphatase/kinase; IPR001510 (Zinc finger, PARP-type), IPR013954 (Polynucleotide kinase 3 phosphatase), IPR023214 (HAD-like domain); GO:0003677 (DNA binding), GO:0008270 (zinc ion binding)

Q, population structure; K, kinship matrix; PCA, principal component analysis; DLA, digital leaf area under non-stressed (NS) conditions; LAI, digital leaf area index under NS conditions; IR_SLA, specific leaf area under NS conditions; ST_LRWC leaf relative water content under drought-stressed (DS) conditions; ST_PB, number of primary branches under DS conditions; IR_PB, number of primary branches under NS conditions; ST_HSW, hundred seed weight under DS conditions

leaf relative water content (LRWC), digital plant height (PH), digital biomass (DBM) (Table 1). Traits related to canopy development are tightly associated with plant water use (Vadez et al. 2015). The leaf area influences the rate of transpiration as the wider the leaf area, the greater the rate of transpiration because broad leaves tend to have more stomata (Maylani et al. 2020). Thus, selecting genotypes with small leaf area could enhance groundnut productivity under DS conditions. Diffuse light penetrates deeper into a plant canopy, and increases photosynthesis and crop production (Zhang et al. 2022). Reportedly, there is a strong positive association between biomass production and transpiration efficiency under drought-stressed (DS) conditions due to the genotypes' root system to mobilize water from the soil for stem elongation and biomass accumulation (Vadez et al. 2016). Previous findings indicated a positive association between reduced SLA and increased leaf thickness under DS conditions. This correlation results in thicker cell wall, which helps to prevent water loss by evaporation and achieve higher water use efficiency (Zhou et al. 2020). LRWC is the most useful parameter to measure plant water status in terms of the physiological consequence of cellular water deficit (Barr and Weatherley 1962). This parameter represents the balance between the water supply to the leaf tissue and the transpiration rate (Lugojan and Ciulca 2011). Thus, maintenance of higher LRWC under -stressed conditions could be a good indication of drought tolerance. The observed genetic variability in this study could be utilized in groundnut breeding programs to develop drought-tolerant and high-yielding varieties.

Population structure and PCA

In the present study, the population structure of the 99 groundnut genotypes revealed the presence of three sub-populations (Abady et al. 2021). Similarly, the PCA results displayed the presence of three sub-groups (Fig. 1b). The low number of sub-groups indicates low genetic differentiation, given that most genotypes were India collections. Combining information generated from the genetic population structure and PCA is useful for the selection of various parents in breeding programs and the mapping of marker-trait associations.

Linkage disequilibrium in groundnut

LD is the non-random co-occurrence of two or more alleles (Lewontin and Kojima 1960). Determination of LD and its decay with the genetic distance helps to assess the resolution of association mapping and desirable numbers of SNPs on arrays (Vos et al. 2017). LD decay depends on cultivation patterns, breeding methods, breeding history, and evolutionary history (Devate et al. 2022). Higher LD decay was observed in the present study than in the previous findings (Pandey et al. 2014; Otyama et al. 2019). This could be attributed to possible intercrosses among the advanced breeding lines. This suggests that the utilization of more SNP markers and population size could enhance the power and efficiency of MTAs in the groundnut breeding programs.

Candidate genes associated with SNPs

Identifying genomic regions using genome-wide SNP markers is a vital approach for developing climate-resilient varieties. In this study, a pleiotropic gene effect was detected between digital leaf area and leaf area index using the marker AX-177643135, indicating collinearity between leaf canopy traits (Table 2, Fig. 3a and c).

Significant SNPs for LRWC were identified on chromosomes A02, A03, A05, A10, B03, B08 and B09. In addition, the following five SNP markers with major effect were identified: AX-176804539, AX-176794990, AX-177641299, AX-176795390 and AX-176822255 located on chromosome A03 at 49.5 kb (PVE=20%), B03 at 45 kb (PVE=31%) and B09 at 45 kb (PVE=21%), A03 at 36.5 kb (PVE=20%) and B03 at 36.5 kb (PVE=20%), in that order (Table 2, Fig. 3g and h). These markers showed strong association signals for LRWC under drought-stressed conditions.

For leaf relative water content (LRWC) under drought-stressed conditions, the SNP AX-176794990 [chromosome B03; $-\log_{10}(P)$ value of 5.68 to 5.79] was located within 21.68 kb of the *Araip.9NG64* gene, which encodes for an RNA-binding protein (RBP) 24-like protein family, involved in RNA processing, export and stability. Muthuswamy et al. (2021) and Yan et al. (2022) reviewed the role of RBPs in abiotic stress response and proposed that the proteins regulate stress response

through RNA metabolism. Based on BLAST analysis of the gene sequence of *Araip.9NG64*, a match was found with UBPI-associated protein 2C (UBP2c) with two RNA recognition motifs of 85 amino acids each, reportedly playing a crucial role in leaf senescence (Na et al. 2015). Li et al. (2002) reported that ABA Activated Protein Kinase (AAPK), which is present in guard cells, interacts with the AAPK Interacting protein (AAPKIP 1), which is a RBP that interacts with mRNA of dehydriin, a protein implicated in drought stress.

The formation of cuticular layers with increased wax and cutin content on leaf surfaces is closely related to drought tolerance. Identification of drought tolerance-associated wax components and cutin monomers and the genes responsible for their biosynthesis is essential for understanding the physiological and genetic mechanisms underlying drought tolerance and improving crop drought resistance (Yang et al. 2022). SNP, AX-147235264 located on chromosome B10 ($-\log_{10}(P)$ value of 3.88) accounted for 20% of the variance in LRWC under drought-stressed conditions (ST_LRWC). It was present near the *Aradu.PKW10* gene, which encodes for a CD2 antigen cytoplasmic tail-binding-like protein. CD2 cytoplasmic tail binding protein 2 is a component of the U5 snRNP complex involved in RNA splicing. PSTPIP1, encodes CD2 antigen-binding protein 1 (CD2BP1), also known as proline/serine/threonine phosphatase-interacting protein 1 (PSTPIP1). SNP, AX-147235264 was reported by Otyama et al. (2022) to be responsible for linolenic acid accumulation and wax formation under drought-stress conditions (Yang et al., 2022).

The current association analyses identified a SNP, AX- AX-176816874 (chromosome B09; $-\log_{10}(P)$ value of 4.29) affecting ST_LRWC. It was present in the *Araip.57P4D* gene, which encodes for a Chitinase (Class V). Some chitinases are expressed in response to abiotic stress (Hamid 2013; Zhou et al. 2020). Lv et al. (2022) found that drought stress treatment induces significant upregulation of Class V chitinases.

SNP, AX-147244306 (chromosome B03; $-\log_{10}(P)$ value of 4.30) was identified as associated with ST_LRWC. It was present in the *Araip.4J8RL* gene, which encodes for a polynucleotide phosphatase/kinase type with Intracellular protein interaction domains with a role in abiotic stress tolerance (Dasuni

and Nailwal 2020). It could assist in the selection for drought tolerance in groundnut.

SNP, AX-147244415 specific to ST_LRWC (chromosome B03; $-\log_{10}$ (P value) of 4.29) was present within the *Araip.SVH5H* gene encodes for a G family of Abscisic acid (ABA) transporter protein of the half-size transporters. The SNPs are expressed in the vascular tissues and are mainly involved in the translocation of ABA across the plasma membrane and tonoplast (Jarzyniak and Jasiński 2014). Kuromori et al. (2011) reported that a mutant version of this gene results in increased transpiration losses and drought susceptibility.

SNP AX-176798839 (chromosome A05; $-\log_{10}$ (P value) of 3.94) associated with ST_LRWC was present in the *Aradu.VIU0I* gene encoding for a zinc finger MYM-type 1 like protein. It is responsible for signalling and regulation under abiotic stress. Zinc finger proteins enhance plant drought resistance by increasing the levels of osmotic adjustment substances (Han et al. 2020).

SNP AX-176817979 [(chromosome A02; $+\log_{10}$ (P value of 4.85 and 3.98)] was identified as associated with the number of primary branches under non-stressed conditions (Table 2, Fig. 3i and j). It was present within 4.0 kb of the *Aradu.ML3P3* gene encodes for a P-type ATPase of the Arabidopsis 2 protein. P-type ATPase type 2 belongs to the haloacid dehalogenase (HAD) superfamily and is split into four groups, i.e., Na^+ , K^+ , H^+ Ca^{2+} , Mg^{2+} and phospholipids (Thever and Saier 2009). Animals and fungi have Na^+/K^+ -ATPases (P2C ATPases) and Na^+ -ATPases (P2D ATPases), respectively, that carry Na^+ exclusion (Axelsen and Palmgren, 2001). The Na^+/K^+ -ATPase helps to maintain low Na^+ and high K^+ concentrations within the cells.

In addition, for SN AX-177639302, which is on chromosome B09 at 4.05 kb (PVE=28%) shows a strong association signal for seed weight under drought-stressed conditions (Table 2, Fig. 3m and n). Similarly, Gangurde et al. (2020 and 2022) reported a seed weight-associated genomic region on chromosome B09 in groundnut. This marker could provide an opportunity for seed size improvement in groundnut.

Conclusions

The study identified SNP-traits associations through association mapping in *Arachis hypogaea*. Forty-eight significant associated regions were detected for important physiological and yield-related traits using the PCA + K and Q + K models. Forty-seven SNPs significantly associated with leaf area, leaf area index, specific leaf area, leaf relative water content, number of primary branches and hundred seed weight under drought-stressed conditions were identified. The identified MTAs and candidate genes in this study could be used to understand the genetic basis of genomic regions of important physiological and yield-related traits and to accelerate the development of drought-tolerant and high yielding groundnut cultivars. Furthermore, the markers could be validated and deployed in groundnut breeding programs for gene pyramiding and trait integration.

Acknowledgements The authors are very thankful to the Groundnut Breeding Program and Center of Excellence in Genomics and Systems Biology (CEG) at ICRISAT, India for the technical assistance during DNA extraction and field experimentation. The authors also acknowledge the University of Georgia, Tifton, United States for providing technical assistance during genotyping. ICRISAT is thanked for providing the groundnut germplasm used in the study.

Author contributions Conceptualization: SA HS, PJ, AW, VC, Data curation: SA, Formal analysis: SA, AW, Funding acquisition: HS, Methodology: SA, HS, Project administration: PJ, Resources: HS, PJ, AW, VC, Supervision: HS, PJ, Validation: HS, Writing – original draft: SA, HS. All the authors read and approved the manuscript.

Funding Open access funding provided by University of KwaZulu-Natal. This work was financially supported by the Organization of the Petroleum Exporting Countries (OPEC) Fund for International Development (OFID), the International Foundation for Science (IFS), and the University of KwaZulu-Natal and conducted under CGIAR Research Program on Grain Legume and Dry Land Cereals (CRP-GLDC).

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have declared no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

References

- Abady S, Shimelis H, Janila P, Yaduru S, Shayanowako AIT, Deshmukh D, Chaudhari S, Manhor SS (2021) Assessment of the genetic diversity and population structure of groundnut germplasm collections using phenotypic traits and SNP markers: implications for drought tolerance breeding. *PLoS ONE* 16(11):e0259883. <https://doi.org/10.1371/journal.pone.0259883>
- Axelsen KB, Palmgren MG (2001) Inventory of the superfamily of P-type ion pumps in Arabidopsis. *Plant Physiol* 126:696–706. <https://www.jstor.org/stable/4279931>
- Azevedo NAD, Nogueira RJMC, Melo FPA, Santos R (2010) Physiological and biochemical responses of peanut genotypes to water deficit. *J Plant Interact* 5:1–10
- Barr HD, Weatherley PE (1962) A re-examination of the relative turgidity technique for estimating water deficit in leaves. *Aust J Biol Sci* 15:413–428
- Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet* 48:438–446. <https://doi.org/10.1038/ng.3517>
- Blummel M, Ratnakumar P, Vadez V (2012) Opportunities for exploiting variations in haulm fodder traits of intermittent drought tolerant lines in a reference collection of groundnut (*Arachis hypogaea* L.). *Field Crops Res* 126:200–206. <https://doi.org/10.1016/j.fcr.2011.10.004>
- Carvalho MJ, Vorasoot N, Puppala N, Muiitia A, Jogloy S (2017) Effects of terminal drought on growth, yield and yield components in Valencia peanut genotypes. *SABRAO J Breed Genet* 49:270–279
- Dasrauni K, Nailwal TK (2020) Zinc finger proteins: Novel sources of genes for abiotic stress tolerance in plants. *Transcription Factors for Abiotic Stress Tolerance in Plants*. Academic Press, Cambridge, pp 29–45. <https://doi.org/10.1016/B978-0-12-819334-1.00003-4>
- Devate NB, Krishna H, Parmeshwarappa SKV, Manjunath KK, Chauhan D, Singh S, Singh JB, Kumar M, Patil R, Khan H, Jain N, Singh GP, Singh PK (2022) Genome-wide association mapping for component traits of drought and heat tolerance in wheat. *Front Plant Sci* 13:943033. <https://doi.org/10.3389/fpls.2022.943033>
- Dwivedi SL, Nigam SN, Nageswara Rao RC, Singh U, Rao KVS (1996) Effect of drought on oil, fatty acids and protein contents of groundnut (*Arachis hypogaea* L.) seeds. *Field Crops Res* 48:125–133
- Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Eadae EA, Byrne PF, Haley SD, Lopes MS, Reynolds MP (2014) Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theor Appl Genet* 127:791–807. <https://doi.org/10.1007/s00122-013-2257-8>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- FAOSTAT (2023) “Food and Agriculture Organization of the United Nations Database of Agricultural Production.” FAO Statistical Databases. <http://www.fao.org/faostat/>. Accessed 14 January 2023
- Gangurde SS, Wang H, Yaduru S, Pandey MK, Fountain JC, Chu Y, Isleib T, Holbrook CC, Xavier A, Culbreath AK, Ozias-Akins P, Varshney RK, Guo B (2020) Nested-association mapping (NAM) based genetic dissection uncovers candidate genes for seed and pod weights in peanut (*Arachis hypogaea*). *Plant Biotechnol J* 18:1457–1471
- Gangurde SS, Khan AW, Janila P, Variath MT, Manohar SS, Singam P, Chitkineni A, Varshney RK, Pandey MK (2022) Whole-genome sequencing based discovery of candidate genes and diagnostic markers for seed weight in groundnut. *Plant Genome* 16:e20265. <https://doi.org/10.1002/tpg2.20265>
- Hamid R, Khan MA, Ahmad M, Ahmad MM, Abidin MZ, Musarrat J, Javed S (2013) Chitinases: an update. *J Pharm and Bioallied Sci* 5:21. <https://doi.org/10.4103/0975-7406.106559>
- Han G, Lu C, Guo J, Qiao Z, Sui N, Qiu N, Wang B (2020) C₂H₂ zinc finger proteins: master regulators of abiotic stress responses in plants. *Front Plant Sci* 11:115. <https://doi.org/10.3389/fpls.2020.00115>
- Hasan N, Choudhary S, Naaz N, Sharma N, Laskar RA (2021) Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *J Genet Eng and Biotechnol* 19:128. <https://doi.org/10.1186/s43141-021-00231-1>
- Janila P, Nigam SN, Pandey MK, Nagesh P, Varshney RK (2016) Groundnut improvement: use of genetic and genomic tools. *Front Plant Sci* 4:1–16. <https://doi.org/10.3389/fpls.2013.00023>
- Jarzyniak KM, Jasiński M (2014) Membrane transporters and drought resistance—a complex issue. *Front Plant Sci* 5:687. <https://doi.org/10.3389/fpls.2014.00687>
- Kuromori T, Sugimoto E, Shinozaki K (2011) Arabidopsis mutants of AtABCG22, an ABC transporter gene, increase water transpiration and drought susceptibility. *Plant J* 67:885–894. <https://doi.org/10.1111/j.1365-313X.2011.04641.x>
- Lewontin RC, Kojima K (1960) The evolutionary dynamics of complex polymorphisms. *Evolution* 14:458–472
- Li J, Kinoshita T, Pandey S, Ng CKY, Gygi SP, Shimazaki KI, Assmann SM (2002) Modulation of an RNA-binding

- protein by abscisic-acid-activated protein kinase. *Nature* 418:793–797
- Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ et al (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28(18):2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Lugoan C, Ciulca S (2011) Evaluation of relative water content in winter wheat. *J Horticult, ForestBiotechnol* 15:173–177
- Lv P, Zhang C, Xie P, Yang X, El-Sheikh MA, Hefft DI, Ahmad P, Zhao T, Bhat JA (2022) Genome-wide identification and expression analyses of the chitinase gene family in response to white mold and drought stress in soybean (*Glycine max*). *Life* 12:1340. <https://doi.org/10.3390/life12091340>
- Mace ES, Buhariwalla KK, Buhariwalla HK, Crouch JH (2003) A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Mol Biol Rep* 21:459–460. <https://doi.org/10.1007/BF02772596>
- Mathew I, Shimelis H, Shayanowako AIT, Laing M, Chaplot V (2019) Genome-wide association study of drought tolerance and biomass allocation in wheat. *PLoS ONE* 14(12):e0225383. <https://doi.org/10.1371/journal.pone.0225383>
- Maylani ED, Yuniati R, Wardhana W (2020) The Effect of leaf surface character on the ability of water hyacinth, *Eichhornia crassipes* (Mart.) Solms. to transpire water. *IOP Conf Ser Mater Sci Eng* 902(1):012070. <https://doi.org/10.1088/1757-899X/902/1/012070>
- Mupunga I, Mngqawa P, Katerere DR (2017) Peanuts, aflatoxins and undernutrition in children in Sub-Saharan Africa. *Nutrients* 9:1287. <https://doi.org/10.3390/nu9121287>
- Muthusamy M, Kim JH, Kim JA, Lee SI (2021) Plant RNA binding proteins as critical modulators in drought, high salinity, heat, and cold stress responses: an updated overview. *Int J Mol Sci* 22:6731
- Na JK, Kim JK, Kim DY, Assmann SM (2015) Expression of potato RNA-binding proteins StUBA2a/b and StUBA2c induces hypersensitive-like cell death and early leaf senescence in *Arabidopsis*. *J Exp Bot* 66:4023–4033
- Otyama PI, Wilkey A, Kulkarni R, Assefa T, Chu Y, Clevenger J et al (2019) Evaluation of linkage disequilibrium, population structure, and genetic diversity in the U.S. peanut mini core collection. *BMC Genom* 20:481. <https://doi.org/10.1186/s12864-019-5824-9>
- Otyama PI, Chamberlin K, Ozias-Akins P, Graham MA, Cannon EK, Cannon SB, MacDonald GE, Anglin NL (2022) Genome-wide approaches delineate the additive, epistatic, and pleiotropic nature of variants controlling fatty acid composition in peanut (*Arachis hypogaea* L.). *Theor Appl Genet* 12:1–21. <https://doi.org/10.1093/g3journal/jkab382>
- Pandey MK, Upadhyaya HD, Rathore A, Vadez V, Sheshshaye MS, Sriswathi M, Govil M, Kumar A, Gowda MVC, Shivali S et al (2014) Genome-wide association studies for 50 agronomic traits in peanut using the reference set comprising 300 genotypes from 48 countries of semi-arid tropics of the world. *PLoS ONE* 9(8):e105228. <https://doi.org/10.1371/journal.pone.0105228>
- Pereira JW, Albuquerque MB, Melo Filho PA, Nogueira RJMC, Lima LM, Santos RC (2016) Assessment of drought tolerance of peanut cultivars based on physiological and yield traits in a semiarid environment. *Agric Water Manag* 166:70–76. <https://doi.org/10.1016/j.agwat.2015.12.010>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multi locus genotype data. *Genetics* 155:945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Ravi K, Vadez V, Isobe S, Mir RR, Guo Y, Nigam SN, Gowda MVC, Radhakrishnan T, Bertoli DJ, Knapp SJ, Varshney RK (2011) Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea* L.). *Theor Appl Genet* 122:1119–1132
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA* 98:11479–11484. <https://doi.org/10.1073/pnas.201394398>
- Salgotra RK, Stewart CN Jr (2020) Functional markers for precision plant breeding. *Int J Mol Sci* 21(13):4792
- Shaibu AS, Sneller C, Motagi BN, Chepkoech J, Chepngetich M, Miko ZL, Isa AM, Ajeigbe HA, Mohammed SG (2020) Genome-wide detection of snp markers associated with four physiological traits in groundnut (*Arachis hypogaea* L.) mini core collection. *Agronomy* 10:192. <https://doi.org/10.3390/agronomy10020192>
- Singh N, Agarwal N, Yadav HK (2019) Genome-wide SNP-based diversity analysis and association mapping in linseed (*Linum usitatissimum* L.). *Euphytica* 215:139. <https://doi.org/10.1007/s10681-019-2462-x>
- Thermo Fisher Scientific Inc (2018) *Axiom™ Analysis Suite (AxAS) v4.0 USER GUIDE*. Available at: https://download.thermofisher.com/Affymetrix_Softwares/Axiom_Analysis_Suite_AxAS_v4.0_User_Guide.pdf
- Thever MD, Saier MH (2009) Bioinformatic characterization of P-Type ATPases encoded within the fully sequenced genomes of 26 Eukaryotes. *J Membr Biol* 229:115–130. <https://doi.org/10.1007/s00232-009-9176-2>
- Vadez V, Ratnakumar P (2016) High transpiration efficiency increases pod yield under intermittent drought in dry and hot atmospheric conditions but less so under wetter and cooler conditions in groundnut (*Arachis hypogaea* (L.)). *Field Crops Res* 193:16–23
- Vadez V, Kholová J, Hummel G, Zhokhavets U, Gupta SK, Hash CT (2015) LeasyScan: a novel concept combining 3D imaging and lysimetry for high-throughput phenotyping of traits controlling plant water budget. *J Exp Bot* 66:5581–5593. <https://doi.org/10.1093/jxb/erv251>
- Vos PG, Paulo MJ, Voorrips RE, Visser RGF, van Eck HJ, van Eeuwijk FA (2017) Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theor Appl Genet* 130:123–135. <https://doi.org/10.1007/s00122-016-2798-8>
- Wan L, Wu Y, Huang J, Dai X, Lei Y, Yan L, Jiang H, Zhang J, Varshney RK, Liao B (2014) Identification of ERF genes in peanuts and functional analysis of AhERF008 and AhERF019 in abiotic stress response. *Funct Integr Genomics* 14:467–477. <https://doi.org/10.1007/s10142-014-0381-4>
- Wankhade AP, Chimote VP, Viswanatha KP, Yadaru S, Deshmukh DB, Gattu S, Sudini HK, Deshmukh MP, Shinde VS, Vemula AK, Pasupuleti J (2023)

- Genome-wide association mapping for LLS resistance in a MAGIC population of groundnut (*Arachis hypogaea* L.). *Theor Appl Genet* 136:43. <https://doi.org/10.1007/s00122-023-04256-7>
- Wankhade A, Purohit A, Janila P (2022) Step-wise selection for early canopy traits followed by stress tolerance indices as an approach for improving drought tolerance in groundnut (*Arachis hypogaea* L.). In: The 7th Congress on Plant Production In Water - Limited Environment, 28 Nov - 02 Dec 2022, King Fahd Hotel, Dakar, Senegal
- Yan Y, Gan J, Tao Y, Okita TW, Tian L (2022) RNA-Binding proteins: the key modulator in stress granule formation and abiotic stress response. *Front Plant Sci* 13:882596. <https://doi.org/10.3389/fpls.2022.882596>
- Yang F, Han Y, Zhu QH, Zhang X, Xue F, Li Y, Luo H, Qin J, Sun J, Liu F (2022) Impact of water deficiency on leaf cuticle lipids and gene expression networks in cotton (*Gossypium hirsutum* L.). *BMC Plant Biol* 22:404. <https://doi.org/10.1186/s12870-022-03788-2>
- Zhang Y, Yang J, Van Haaften M, Li L, Lu S, Wen W, Zheng X, Pan J, Qian T (2022) Interactions between diffuse light and cucumber (*Cucumis sativus* L.) canopy structure, simulations of light interception in virtual canopies. *Agronomy* 12:602. <https://doi.org/10.3390/agronomy12030602>
- Zhou N, An Y, Gui Z, Xu S, He X, Gao J, Zeng D, Gan D, Xu W (2020) Identification and expression analysis of chitinase genes in *Zizania latifolia* in response to abiotic stress. *Sci Hortic* 261:108952. <https://doi.org/10.1016/j.scienta.2019.108952>
- Zhou X, Guo J, Pandey MK, Varshney RK, Huang L, Luo H, Liu N, Chen W, Lei Y, Liao B, Jiang H (2021) Dissection of the genetic basis of yield-related traits in the chinese peanut mini-core collection through genome-wide association studies. *Front Plant Sci* 12:637284. <https://doi.org/10.3389/fpls.2021.637284>
- Zou K, Kim KS, Kang D, Kim MC, Ha J, Moon JK, Jun TH (2022) Genome-Wide association study of leaf chlorophyll content using high-density SNP array in peanuts (*Arachis hypogaea* L.). *Agronomy* 12:152. <https://doi.org/10.3390/agronomy12010152>

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