



OPEN Whole genome sequencing reveals transcriptional and translational elements potentially regulating biotic and abiotic stress responses in cowpea

Dhanasekar Punniyamoorthy✉ & Souframanien Jegadeesan

Cowpea (*Vigna unguiculata* (L.) Walp.) is a highly versatile and resilient crop, globally ranking as the third most pivotal grain legume. However, various biotic, abiotic, and physiological challenges, often hinder its productivity. Cowpea exhibits complex environmental adaptive responses regulated at the transcriptional and translational levels through mechanisms such as resistance genes (R-genes), transcription-associated proteins (TAPs), and protein kinases (PKs). A comprehensive study was conducted based on a whole-genome hybrid assembly (Illumina and Nanopore) in cowpea, revealing the identification of 2188 R-genes (29 classes), 5573 TAPs (118 families) and 1135 PKs (22 groups, 122 families). Among the R-genes, Kinases (KIN) and transmembrane proteins (RLKs and RLPs) were prominent, while CCHC (Zn), C2H2, MYB-HB-like, WD40-like, bHLH, and ERF families were notable among TAPs. The largest kinome group, RLK-Pelle, encompassed over three-fifths of the cowpea PKs (VuPKs), followed by CAMK and CMGC groups. Two and three novel families in TAPs (ABTB and CW-ZN-B3_VAL) and PKs (RLK-Pelle-URK-1, RLK-Pelle-URK-2, TKL-Cr-3), respectively, were identified along with two novel PK groups (NAK and TLK). Dispersed and tandem duplication events under purifying selection mainly contributed to kinome expansion, with chromosome 'Vu03' anchoring the maximum PKs. This investigation delves into the biological intricacies with manipulative potential to enhance cowpeas' resilience to environmental challenges without compromising yield.

Keywords R-genes, TAPs, Kinome, Climate resilience, Duplication, Purifying selection

Identification of climate resilient crops assumes prominence among the major research priorities for ensuring global food and nutritional security. Legumes are known to evince acclimatizing adaptation to an extensive range of ecological conditions from arid to temperate climates. Among the hardy legumes, cowpea (*Vigna unguiculata* (L.) Walp.) is a multifarious crop endowed with climate smart attributes and laden with inherent potential to mitigate the vagaries of climate change^{1,2}.

Globally, cowpea is the third most important grain legume crop in terms of area and production, next only to dry beans and chickpea. Cultivated over an estimated area of 15.19 Mha and production of 9.77 Mt³, the productivity of cowpea (643.5 kg/ha) is abysmally low compared to other legumes like chickpea and dry beans. Various biotic and abiotic stresses predominantly hinder cowpea productivity and augmentation of resistance against these factors is a research prerogative of eminence. Even though cowpeas are biologically resilient, the low productivity is essentially attributed to its cultivation under subsistence or marginal conditions with minimal inputs, limited access to improved varieties and largely grown as a mixed intercrop rather than sole crop⁴. From the genomics perspective, until recently, cowpea genetic improvement was on the backfoot unlike its counterparts like soybean, chickpea, and common bean. Cowpea is a diploid fabid with a chromosome number $2n = 22$ and an estimated genome size of about 641 Mbp⁵. With next-generation sequencing becoming increasingly cost-effective and the various genomic tools and whole genome sequence of cowpea⁵ becoming readily accessible, the genetic improvement of cowpea has gathered momentum in recent times.

Like other plants, cowpea, when confronted with challenges imposed by various biotic and abiotic stresses, has a myriad of complex adaptive response mechanisms to thwart or minimize the negative impacts stemming

Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Trombay, Mumbai, Maharashtra 400 085, India. ✉email: sekar@barc.gov.in

out of such situations. The rapid and efficient responses are moderated at the transcriptional or translational levels primarily through resistance genes (R-genes), transcription-associated proteins (TAPs) and protein kinases (PKs) among others. R-genes are largely involved in the response against biotic stresses, especially the disease-causing pathogens⁶. The disease resistance reaction of plants is the consequence of the interactions between the R-genes and pathogen specific effector molecules called avirulence proteins⁷. Upon recognizing the invading pathogen, the R-genes elicit defence response signalling cascades against it. These R-genes consist of well conserved domains and motifs such as the N-terminal transmembrane (TM), serine/threonine and tyrosine kinase (STTK), lysin motif (LysM), coiled-coil (CC) and Toll/Interleukin-1 receptor (TIR), central nucleotide binding site (NB-ARC), and the C-terminal leucine rich repeats (LRR)^{8,9}. These conserved structural features are utilized by various bioinformatic tools for mining the R-genes from the whole genome sequences. In plants, among the R-genes, the nucleotide-binding leucine-rich repeat (NLR) class of genes is most widely prevalent¹⁰. A collection of 481 NLRs have been validated through experimentation in 31 plant genera¹¹. These R-genes are known to play a pivotal role in disease resistance breeding and therefore, it is imperative to envisage the R-genes prevalent across the genome.

The developmental and resilient adaptive responses of plants to environmental variations are dynamically moderated primarily through the reprogramming of gene expressions by various transcriptional regulatory factors, apart from epigenetic and translational control. In plants, intricate systems of TAPs regulate the transcription of protein encoding genes^{12,13}. TAPs are primarily comprised of: (1) Transcription factors (TFs) that bind in a sequence-specific manner to cis-acting non-coding DNA regulatory regions; (2) Transcriptional regulators (TRs) that bring out regulation by non-specific DNA binding, protein–protein interactions or chromatin remodelling; and (3) Putative TAPs (PTs) with hitherto unknown roles¹⁴. Apart from these factors, a variety of other TAPs like RNA polymerases, mediator complexes, polyadenylation factors, transcription elongation and termination factors also work in a coordinated manner to regulate the transcription of genes, ensuring the accurate and controlled synthesis of RNA molecules in response to various cellular signals and environmental cues. As per an estimate, at least 5–7% of the genome encoded proteins in plants are tasked with transcriptional regulation^{15–17}. The fraction of TFs in the expressed genome is in positive consonance with the complexity of the organism¹⁸.

The third major family of proteins coordinating the signalling pathways in tandem with other regulatory factors are the PKs. Kinomes, the whole set of kinases present in the genome, constituting about 1–2% of functional proteome, usually present a conserved catalytic domain comprising 250–300 amino-acids¹⁹. The PKs primarily regulate the adaptive and cellular functions through reversible phosphorylation and post-translational modification of downstream target proteins influencing their activities, localization, inter-protein interactions, and other features²⁰. Activation of the proteins by appending phosphate moieties conjures a cascade of signal transductions eventually modulating the adaptive, developmental, metabolic, and stress-responsive cellular processes. PKs are documented as key elements in plant responses to biotic and abiotic stresses viz., water scarcity, high salinity, low temperature, and pathogen attack²⁰.

In cowpea, the meagre reports on R-genes, TAPs, and PKs encouraged the present study to comprehend these regulatory factors using the hybrid assembly (Illumina and Nanopore) of a whole genome sequence of cowpea cultivar ‘CPD103’. The genome-wide identification and characterization of these regulatory factors will help not only understand the regulatory intricacies controlling the biological processes in legumes but also will pave the way for manipulating the processes to the relative advantage of mankind.

Materials and methods

Materials

The cowpea cultivar ‘CPD103’ is being used at our institute in cowpea mutation breeding programme for induction of resistance against cowpea aphid borne mosaic virus. To generate genomic resources, ‘CPD103’, also known as ‘CDS’, was subjected to de novo whole genome sequencing using Illumina and Nanopore sequencing techniques for hybrid assembly.

DNA extraction and quality control

Genomic DNA was extracted in duplicate from young leaves using Qiagen DNeasy Plant Mini kit as per the instructions contained in the kit and previously described²¹. Finally, DNA was eluted with 50 µl of 10 mM Tris–Cl (pH 8.0). The eluted genomic DNA was quantified and evaluated for quality using Nanodrop 2000 (Thermo Scientific, USA), Qubit (Thermo Scientific, USA), and agarose gel electrophoresis (samples with A260/A280 ratio 1.8–2.0, A260/A230 ratio > 1.8, Qubit concentration > 10–20 ng/µl for Illumina and > 50 ng/µl for Nanopore and those not showing smearing, degradation, RNA contamination or faint bands in gel electrophoresis were only used for sequencing).

Illumina and nanopore library preparation and sequencing

Library construction and sequencing was carried out at M/s Genotypic Technology, Bengaluru, India, as detailed previously²¹. Briefly, library preparation involved NEXTFLEX Rapid DNA-seq kit (BIOO Scientific, Inc. U.S.A.) for Illumina platform as per the manufacturer’s directives. The Qubit quantified DNA (500 ng) was fragmented (200–250 bp) by sonication (Covaris S220, USA), purified, ligated to multiplex-barcoded adaptors, and prepared the sequencing library by PCR-amplification for 4 cycles using kit provided primers. The library was thereafter, purified, checked for quality, Qubit quantified, and analyzed for fragment size distribution (Agilent 2200 Tape Station). The paired-end sequencing of equimolar-normalized library post multiplexing was carried out on a HiSeq X Ten Illumina sequencer as per the instruction of the manufacturer (150 cycles). For Nanopore library preparation, the end-repaired (NEBNext ultra II end repair kit, New England Biolabs, MA, USA) and purified DNA (1.5 µg) were ligated with adapter (AMX) at room temperature (20 °C) for 20 min using NEB Quick T4

DNA Ligase (New England Biolabs, MA, USA). The ligation sequencing kit (SQK-LSK109) provided elution buffer (15 µl) was used for eluting the purified reaction mixture and constituted the sequencing library. Long read sequencing was accomplished on a GridION X5 (Oxford Nanopore Technologies, Oxford, UK) sequencer with a SpotON flow cell R9.4 (FLO-MIN106) following a 48-h sequencing-protocol (20× depth) and base calling was performed on the raw reads ('fast5' format) by Guppy basecaller v2.3.4 tool (<https://nanoporetech.com/documentation/Guppy-protocol#windows-guppy>). The sequencing samples in both the cases consisted of two biological replicates and two technical replicates.

Bioinformatic analysis

The hybrid (MaSuRCA v3.4.2)²² assembled genome was processed for repeat region masking using RepeatModeler v2.0²³ and RepeatMasker v4.0.6²⁴. The completeness of drafted genome assembly was validated (BUSCO v3.0.2²⁵) by retrieving the annotation percentage of complete genes. The draft genome of 'CDS' sample along with transcript data and reference protein data were used for gene prediction in BRAKER v2.1.4 tool²⁶. The hybrid draft cowpea genome/proteome sequence of 'CDS' (Supplementary material SM) was used further for mining R-genes, TAPs, and PKs.

Prediction of R-genes

Putative R-genes in the draft cowpea genome assembly were predicted and annotated by employing Disease Resistance Analysis and Gene Orthology (DRAGO 3)²⁷ pipeline available online through web interface (<http://prgdb.org/prgdb4/drago3>) using the proteome generated through the BRAKER tool. The DRAGO 3 pipeline makes use of the Pathogen Recognition Genes database (PRGdb v4)²⁷ repository for prediction of putative genes based on the candidate recognition genes. DRAGO 3 performed multiple sequence alignment for each PRG class using MEGA X²⁸ (MUSCLE algorithm with default parameters) prior to creating hidden Markov model (HMM) using HMMER v3 package²⁹. An in-house PERL script filtered the best alignments with a minimum BLOSUM62 score of +1 and peptides with at least 10 AAs were only considered. The HMM modules created by the PERL script were used to detect LRR, Kinase, NBS and TIR domains, while CC domains and TM domains were detected by DRAGO 3 using COILS v2.2³⁰ and TMHMM v2.0c³¹ programs. In addition, DRAGO 3 also detects LYK and LYP proteins containing LYSM (Lysin motif) in the place of LRR domains and also LECRK proteins containing lectin-like motifs (LECM).

Prediction of transcription associated proteins (TAPs) and protein kinases (PKs)

The TAPs containing the TFs and TRs present in the draft cowpea genome assembly were predicted using three different identification pipelines: the PlantTFcat pipeline³², the iTAK v1.6 pipeline³³ and PlantTFDB v5.0³⁴ module in PlantRegMap³⁴. The PlantTFcat pipeline utilizes InterProScan v5.59–91.0³⁵ to systematically search proteins for TFs/TRs/chromatin remodelling (CR)-related domain signatures. The iTAK pipeline based on PFAM domain models and consensus rules summarized from different pipelines, was used to identify and classify TFs, TRs and PKs from protein or nucleotide sequences into different gene families. PlantTFDB prediction tool adopts an integrative strategy by combining sequence-based prediction (InterProScan), orthologous-based projection, and collection of annotation in canonical sources [The Arabidopsis Information Resource (TAIR³⁶) and UniProt³⁷] to identify TFs. The non-redundant families identified by all three pipelines were used to predict the maximum number of TFs and TAP families in the proteome.

The cowpea PKs (VuPKs) were predicted using iTAK that is based on significant hit to protein kinase domains (PF00069, PF07714, or PF00481) in the Pfam database³⁸ that were classified into gene families by comparing their sequences to a set of HMMs¹⁹. The sub-cellular localizations of these VuPK genes were predicted using CELLO v2.5³⁹ and LOCALIZER v1.0.4 tools⁴⁰. The VuPK protein sequences were submitted to ProtParam⁴¹ (<https://web.expasy.org/protparam>) to determine the molecular weights and theoretical isoelectric points (pIs).

For comparative analysis of TAPs and PKs, the reference genome of cowpea in NCBI (assembly ASM411807v2) was also analysed using iTAK and the results were compared with the predictions involving our genome.

Expansion mechanisms of VuPK genes

The duplication mechanisms leading to the origin of the VuPK genes were discerned using the Multiple Collinearity Scan toolkit vX⁴² (MCScanX) software package. MCScanX identified PK homologs along the *V. unguiculata* genome and categorized the duplication events into tandem and segmental duplications. The PK genes devoid of any duplicates were classified as "singletons", while those with gene ranks less than 20 (gene ranks were assigned based on the order of chromosomal location) were considered "proximal duplicates". Adjacent PK gene pairs with unit gene rank differences were classified as "tandem duplicates" and those with BLASTp hits exceeding 20 gene ranks were christened "dispersed duplicates". The anchor genes in collinear blocks across chromosomes were regarded as "WGD/segmental duplicates". Genes with multiple BLASTp hits were uniquely assigned one of the above classes in accordance with their precedence order (segmental, followed by tandem, proximal, and dispersed). The coding sequences of the tandemly duplicated VuPK genes, post alignment using Clustal Omega⁴³ (EMBL-EBI Job Dispatcher sequence analysis tools framework⁴⁴), were analysed through MEGA v11.0.13 for determining Ka (non-synonymous substitution)/Ks (Synonymous substitution) ratios. The substitution ratios resolved using standard genetic code following Nei-Gojobori method (Jukes-Cantor model) served as indicators of the selection nature these VuPK genes were subjected to. The duplicated gene pairs with Ka/Ks ratio of less than "1" could be construed to be under purifying selection (negative selection) resulting in conserved amino acid sequences, while those with more than "1" were deemed to have undergone positive or Darwinian selection leading to altered peptides. The duplicating genes with Ka/Ks ratios equal to one were profoundly uninfluenced by neutral selection, negating changes in amino acid sequences⁴⁵.

Validation of in silico determined R-genes, TAPs and PKs

Twenty gene sequences (CDS) each from the in silico-identified R-genes, TAPs, and PKs, were randomly selected for genic primer design using Primer3web v4.1.0⁴⁶ with default parameters. The synthesized primers were used for polymerase chain reaction (PCR) amplification in ten diverse cowpea genotypes (GC3, TC901, C-152, PL-1, ARC-1, PLM211, NBC-1, Vu-89, VBN-3, VBN-1). For each genotype, DNA was extracted from two biological replicates. Each 25 µl PCR reaction contained 75 ng of genomic DNA, 1 µM each of forward and reverse primers, 250 µM dNTPs, 1 × Taq buffer with MgCl₂, and 0.85U of Taq DNA polymerase (Qiagen). Amplifications were performed in a Nexus Eppendorf thermal cycler using the following program: initial denaturation at 94 °C for 4 min; 35 cycles of denaturation at 94 °C for 1 min, annealing at 55–60 °C (depending on primer T_m) for 30 s, and extension at 72 °C for 1 min; followed by a final extension at 72 °C for 6 min. PCR products were separated on 2% agarose gels using a 100 bp DNA ladder as a size marker and visualized using a Syngene gel documentation system (Syngene, UK). The sizes of the amplified fragments were compared with the expected amplicon lengths to validate primer specificity and amplification efficiency.

Transcriptomic analysis under biotic and abiotic stresses

RNA-seq raw data (Illumina paired-end reads) from cowpea plants subjected to biotic stress (infection with cowpea aphid-borne mosaic virus-CABMV) and abiotic stress (root dehydration) were retrieved from the NCBI Sequence Read Archive. The datasets, accessed via BioProject accessions PRJNA655993 and PRJNA605156, respectively, were then analysed for differential expression of various R-genes, TFs and PKs under these stress conditions. Details pertaining to sampling, stress application methods, sequencing, and cowpea genotypes used for transcriptome analysis have already been published⁴⁷. Briefly, for biotic stress, young trifoliate leaves of the greenhouse-grown cultivar 'IT85F-2687' were mechanically injured with carborundum before applying viral inoculum and leaf samples were collected at 60 min and 16 h post-inoculation. For abiotic stress, root dehydration in the hydroponically grown cultivar 'Pingo de Ouro' involved withdrawing the nutrient solution, with root samples taken at 25 min and 150 min after treatment. Both stress conditions were applied during the V3 development stage, and the experimental design consisted of three biological replicates and two technical replicates. Bioinformatic analyses were performed on Galaxy web platform⁴⁸. Briefly, the raw data were subjected to initial quality assessment with Falco v1.2.4⁴⁹ and the quality was further improved through trimming and filtering using Cutadapt⁵⁰ with minimum Phred score of 30 and minimum read length of 80 bp. The trimmed paired-end reads were then mapped to the *Vigna unguiculata* reference genome assembly (ASM411807v2) and gene (gtf) annotation files (downloaded from NCBI) using RNA STAR⁵¹ v 2.7.11a (with default settings excepting that the value of 200 was input as the length of genomic sequence around annotated junctions). The resulting BAM files were used for counting the number of reads per annotated gene using FeatureCounts⁵² v2.0.8 with minimum mapping quality per read of 30. The read counts were further used for analysing differential gene expression (DGE) using DESeq2⁵³ v2.11.40.8 with normalization for sequencing depth and default settings. The annotated DESeq2 files were filtered to extract genes with a significant change in gene expression (adjusted *p* value < 0.05 and |log₂FC| > 1) between treated and untreated samples. The volcano plots of DGEs were created through ggplot2⁵⁴ v3.5.2 within the Galaxy platform.

Results

Whole-genome sequencing of cowpea genotype 'CDS'

The cowpea genome ('CDS') was *de novo* assembled through a hybrid (Illumina and nanopore) whole genome sequencing approach. Illumina sequencing generated a total of ~241 million short-reads, while the nanopore sequencing produced ~7.7 million long-reads, resulting in a sequencing coverage of ~120× for Illumina data and ~20× for nanopore data. Prominent assembly features are presented in Supplementary Table S0. The hybrid genome assembly resulted in a haploid genome size of ~325 MB which covered ~87% of the haploid genome estimated by the KmerGenie program (Supplementary Fig. S1). The final draft assembly of the genome was generated post processing of the assembled genome for repeat region masking. The completeness of the assembled draft genome was validated by read utilization and identification of single copy genes. About 94% percent of read utilization and 93.4% of BUSCO (Benchmarking Universal Single-Copy Orthologs) completeness (C: 93.4% (S: 92.0%, D: 1.4%), F: 1.5%, M: 5.1%, n: 5366) confirmed good draft assembly (Supplementary Fig. S2). The proteome sequence of 'CDS' generated by the BRAKER (bacronym for Bioinformatics Re-analysis Automation of Known and Expressed Regions) tool was used further for downstream analysis.

Prediction of R-genes

A total of 65,708 protein sequences that were generated by the BRAKER tool were analysed for prediction of R-genes using DRAGO 3. R-gene related domains and motifs were predicted in 2188 proteins belonging to 28 different classes (Table 1, Supplementary Table S1a). Maximum number of proteins containing the R-genes related domains belonged to kinases (855), followed by transmembrane receptors RLKs (Kin-LLR) (258), and RLPs (Ser/Thr-LRR) (238). Eight classes (CNL, TNL, NL, CN, TN, N, CTNL, CNT) harboured nucleotide-binding site domains encompassing 392 (17.9%) R-domain proteins. The candidate R-genes in cowpea were observed to carry domains or motifs of type ranging from one to five. Most of the proteins (1063) carried two types of domains in 13 different combinations, followed by those with three types of domains (610) in 12 different combinations. Singleton domains (Kin, NBS, LYSM, TIR, TM, LRR, LECM) were observed in 319 proteins encoded by R-genes, while 183 had 4 types of domains, and only 13 presented 5 types of domains with the lone combination CC-NBS-TM-TIR-LRR. Among the R-genes, the proteins with TM-KIN motifs had the maximum representation (686), distantly followed by LRR-TM (219) and LRR-TM-KIN (277). Seven proteins with CC-NBS-TM-TIR, CC-LECM-TM, CC-TM-TIR-LRR, CC-TM-LYSM-KIN, LRR-TM-TIR, NBS-TIR, NBS-LRR-TIR motif combinations were each represented singly (Table 1). In addition, to ascertain whether

Domain/motifs	Number of R-genes	Class	Number of R-genes
CC-KIN	29	C	12
CC-TM-KIN	72	CK	101
CC-LRR-KIN-TM	3	CL	19
CC-LRR-TM	16	CLEC	1
CC-NBS-TM-LRR	84	CLECRK	3
CC-TM	12	CLK	3
CC-NBS-LRR	7	CLYSK	1
CC-NBS-TM-TIR-LRR	13	CN	33
CC-NBS-TM-TIR	1	CNL	91
CC-NBS-TM	26	CNT	1
CC-LRR	3	CT	2
CC-LECM-TM-KIN	3	CTL	1
CC-LECM-TM	1	CTNL	13
CC-TIR	2	KIN	855
CC-NBS	7	L	47
CC-TM-TIR-LRR	1	LEC	27
CC-TM-LYSM-KIN	1	LECRK	96
KIN	169	LYK	19
LECM	5	LYS	20
LECM-TM	22	N	59
LECM-TM-KIN	96	NL	77
LRR	47	RLK	258
LRR-TM	219	RLP	238
LRR-TM-KIN	277	T	33
LRR-TM-TIR	1	TL	1
NBS	14	TN	26
NBS-TM	44	TNL	92
NBS-LRR-TM	69	TRAN	59
NBS-LRR-TM-TIR	90	Total	2188
NBS-TIR	1		
NBS-LRR	10		
NBS-LRR-TIR	1		
NBS-TM-TIR	25		
LYSM	6		
TIR	19		
TM	59		
TM-KIN	686		
TM-TIR	14		
TM-LYSM	14		
TM-KIN-LYSM	19		
Total	2188		

Table 1. Prediction of R-gene domains/motifs from whole genome sequences of cowpea cultivar ‘CDS’ through DRAGO 3 pipeline of plant resistance gene database (PRGdb 4.0).

the assembled genome properly represented all the classes of R-genes, the reference genome of cowpea in NCBI (assembly ASM411807v2) was also used for R-genes prediction using the DRAGO 3 pipeline (Supplementary Table S1b). It was observed that the prediction based on the assembled genome in our study identified one class of R-genes, CLEC, that was not present in the reference genome, while LYP class found in the reference genome remained elusive in ours. Albeit the pattern of representation of different classes of R-genes remaining the same in both the genome assemblies, the preponderance of proteins within each class was higher in the reference genome excepting the L, LEC and LYS classes (Supplementary Table S1c). Each of the NBS and KIN domains were represented in 9 classes of R-genes, while LRR domains were found in 10, TIR in 8, LEC in 4, and LYS in 3 classes.

Transcription factors (TFs) and transcription-associated proteins (TAPs)

The repertoire of TFs and TAPs in cowpea were predicted using three different pipelines and all non-redundant TFs cumulatively identified by these three pipelines were anticipated to be involved in transcriptional regulation.

The PlantTFcat pipeline identified a total of 5573 TAP encoding genes with 98 families, of which CCHC (Zn) (1187), C2H2 (642), MYB-HB like (361), and WD40 like (353) were predominant (Table 2). A total of 33 families showed low representation with each coded by less than 10 genes, while families like JSW1, JMJC-ARID, LFY, MYB-related, NOZZLE, STAT, and TAZ were under-represented merely by one or two genes each. The PlantTFDB pipeline successfully uncovered 2128 genes belonging to 58 families among which 17 were under-represented with less than 10 genes each (Table 2). The prominent TF families include bHLH (195), MYB (167), ERF (149), and C2H2 (128), while NZZ/SPL, HRT-like, LFY, STAT, NF-X1, S1Fa-like, and SAP were inconspicuously under-represented with one or two genes each. The iTAK pipeline mined 2198 TFs and 504 TRs from 93 families. Some of the over-represented families in the order of preponderance include MYB (165), bHLH (163), ERF (152) and C2H2 (143). Thirty-four of the TFs were identified to be under-represented with less than 10 genes each, while HRT, LFY, MED7, SOH1, ULT, and NOZZLE were uniquely represented. Altogether, 118 non-redundant families housing the TFs and TAPs were identified using the three pipelines (Supplementary Tables S2a–c). Thus, the largest TF families predicted in cowpea include CCHC-type Zinc-finger (CCHC(Zn), Cys(2)-His(2) type (C2H2), myeloblastosis-Homo box like (MYB-HB like), Trp (W)-Asp (D) repeat proteins (WD-40-like), basic helix-loop-helix (bHLH), MYB, and ethylene response factor (ERF). On comparing the TAPs deduced from the reference and our genomes, it was observed that three families viz., RB, STAT, and ULT were exclusively found in the CDS genome. In general, the number of TAPs in the reference genome was more abundant compared to our genome. In particular, the families belonging to bZIP, C3H, FAR1, HB-BELL, LUG, MADS-MIKC, MYB-related, NAC, PHP and WRKY were predominant ($1.2\times$ to $3.5\times$). In contrast, the TFs belonging to AP2/ERF AP2, B3, C2C2 LSD and CPP were relatively more ($1.2\times$ to $2\times$) in our genome (Supplementary Table S2d).

Genome-wide identification and classification of protein kinases (kinome)

The kinome, comprising the entire set of protein kinases (PKs) encoded by the cowpea genome, was predicted in silico through iTAK (Supplementary Table S3a). A total of 1215 kinases were discerned in cowpea after the exclusion of redundant sequences (Supplementary Table S3b). The identified PKs were classified into groups and families following an approach based on Hidden Markov Models. Only 1135 of the annotated PKs could be ascertained of their families following multiple sequence alignment and clustering based on the neighbour-joining method and were used for further analysis (Supplementary Table S3c). The 1135 PKs were allocated into 22 groups, comprising of 122 families (Table 3, Supplementary Table S3a). The receptor-like kinase/Pelle (RLK-Pelle) group was the largest comprising of 56 families and housing about 68.02% (772) of the total PKs in the genome (Fig. 1). The other major groups included Ca^{2+} /calmodulin-dependent protein kinases (CAMK, 86) with 6 families, cyclin-dependent, mitogen-activated, glycogen synthase and CDK (cyclin dependent kinases)-like protein kinases (CMGC, 76) with 17 families, tyrosine kinase-like kinases (TKL, 57) with 11 families, and serine/threonine kinases (STE, 42) with 6 families. The RLK-Pelle_DLSV family was the largest with little more than one-sixth (133) of the RLK-Pelle group PKs. The other major families of the RLK-Pelle group included leucine-rich repeat-XI-1 (LRR-XI-1), leucine-rich repeat-III (LRR-III), receptor-like cytoplasmic kinase-VIIa-2 (RLCK-VIIa-2), L-type lectins (L-LEC), S domain 2b (SD-2b), LRK10-like kinase type 2 (LRK10L-2), and, *Catharanthus roseus* RLK1-like (CrRLK1L-1), each holding PKs in the range of 32–58. Some of the prominent families from other groups were calcium-dependent protein kinases (CDPK) and CAMK-like checkpoint kinase 1 (CAMKL-CHK1) of CAMK group, homologous to yeast STE11 (STE11) of STE group, plant-specific 4 (PI-4) of TKL group, ribosomal S6 kinases 2 (RSK-2) of AGC group, cyclin-dependent kinase-cdc2-related kinase 7-cyclin-dependent kinase 9 (CDK-CRK7-CDK9) of CMGC group, and nuclear receptor binding protein (NRBP) of with-no-lysine [K] kinases (WNK) group each comprising of 15–38 members. Of the 25 singleton families, only four were assigned to the largest group, RLK-Pelle. Ninety-nine families were codified by less than 15 genes; only 14 had 20 or more coding genes. The 1135 PKs were unevenly distributed across the 11 cowpea chromosomes. Chromosome 3 contained the highest number, anchoring 169 PKs (14.9%) spanning 68 families, followed by chromosome 5 with 152 PKs (13.4%) across 58 families. In contrast, chromosome 10 harboured the fewest PKs, with 69 members (6.1%) from 42 families, closely followed by chromosome 4, which housed 71 PKs (6.3%) representing 38 families (Fig. 2). One hundred fifty-two of the PK genes, associated with 24 unique families were devoid of introns, while the rest of the PKs (86.61%) have one or more introns in its genomic structure (Supplementary Table S3c). The average number of introns per family varied from zero (RLK-Pelle_LRR-VII-3) to 28 (PEK_GCN2). Twelve (~9.8%) out of 122 searched families, all belonging to RLK-Pelle group, had an average number of introns between 0 and 0.77, indicating that most of its members do not have introns. In addition, this group also housed one (RLK-Pelle_LRR-XIIIb; mean number of introns: 26.25) of the seven families with 20 or more mean number of introns, representing its structural heterogeneity. Among the major groups, STE exhibited the highest average number of introns (14.92), while the largest group RLK-Pelle displayed the least average number of introns (4.85). On comparing with the reference genome-based kinome⁴⁷, it was observed that two PK groups (NAK and TLK) and five families (TKL-Cr-3, RLK-Pelle-URK-1, RLK-Pelle-URK-2, NAK and TLK) were exclusively present in our genome. One (TKL-PI-3) of the 118 families reported in the reference genome remained elusive in our study (Supplementary Table S3d). The disparity in the total number of predicted PKs (1293 in reference vs 1135 in ours) resulted largely from a single group RLK-Pelle (908 vs 772).

The dispersion duplication mechanism was the main apparatus for VuPK expansion in cowpea genome, responsible for the expansion of 841 VuPKs (Supplementary Table S4a). None of the VuPK genes showed expansion through whole genome duplication (WGD) event (Fig. 3). About 10 VuPK genes did not show duplication and were considered singletons (Supplementary Table S4b). Eighty-five VuPK genes belonging to 6 groups (CAMK, CMGC, TTK, WEE, STE and RLK-Pelle) exhibited proximal duplications (Supplementary Table S4c). Seventy-three tandem duplication events covering a total of 198 genes and composing of 119 duplicated

PlantTFcat		PlantTFDBv5.0		iTAK	
Family	Frequency	Family	Frequency	Family	Frequency
A20-like	10	AP2	37	Alfin-like	10
ABTB	5	ARF	32	AP2/ERF-AP2	32
AP2-EREBP	188	ARR-B	23	AP2/ERF-ERF	152
ARF	32	B3	82	AP2/ERF-RAV	2
ARID	12	BES-1	9	ARID	17
ARID-HMG	6	bHLH	195	AUX/IAA	35
AS2-LOB	44	BRR-BPC	5	B3	86
AUX-IAA	22	bZIP	99	B3-ARF	32
B3-Domain	88	C2H2	128	BBR-BPC	5
BED-type(Zn)	64	C3H	54	BES1	9
BES/BZR	9	CAMTA	10	bHLH	163
bHLH	209	CO-Like	11	bZIP	93
Bromodomain	31	CPP	6	C2C2-CO-like	11
BTB-POZ	34	DBB	13	C2C2-Dof	43
BTB-POZ-MATH	8	Dof	43	C2C2-GATA	33
bZIP	129	E2F/DP	11	C2C2-LSD	6
C2C2-CO-like	36	EIL	7	C2C2-YABBY	11
C2C2-Dof	43	ERF	149	C2H2	143
C2C2-GATA	35	FAR1	65	C3H	72
C2C2-YABBY	11	G2-like	66	CAMTA	9
C2H2	642	GATA	33	Coactivator P15	3
C3H	88	GeBP	5	CPP	6
C3H-WRC/GRF	30	GRAS	64	CSD	4
CCHC(Zn)	1187	GRF	13	DBB	10
CG1-CAMTA	10	HB-other	12	DBP	2
CHROMO-DOMAIN	60	HB-PHD	3	DDT	9
CW-Zn	9	HD-ZIP	60	E2F-DP	10
CW-Zn-B3/VAL	6	NZZ/SPL	2	EIL	7
DDT	12	HRT-like	1	FAR1	64
DICER	0	HSF	31	GARP-ARR-B	20
E2F-DP	11	LBD	44	GARP-G2-like	66
EIL	7	LFY	1	GeBP	5
FAR	65	LSD	6	GNAT	45
FHA-SMAD	32	MIKC-MADS	43	GRAS	63
FYR	7	M-type_MADS	32	GRF	13
GAGA-Binding-like	5	MYB	167	HB-BELL	16
GARP-G2-like	23	MYB_related	98	HB-HD-ZIP	55
GeBP	5	NAC	99	HB-KNOX	19
GRAS	64	NF-YA	10	HB-other	16
GRF	104	NF-YB	22	HB-PHD	3
Hap2/NF-YA	10	NF-YC	16	HB-WOX	21
Hap3/NF-YB	83	Nin-like	13	HMG	9
HD-SAD	26	RAV	3	HRT	1
HD-ZIP	22	SBP	26	HSF	31
HMG	9	SRS	11	IWS1	12
Homeodomain-LIKE	15	STAT	2	Jumonji	33
Homeodomain-PHD	3	TALE	35	LFY	1
Homeodomain-TALE-BEL	18	TCP	26	LIM	8
Homeodomain-TALE-KNOX	24	TRIHILIX	38	LOB	44
Homobox-WOX	123	VOZ	3	LUG	9
HSA	3	WHIRLY	3	MADS-MIKC	41
HSF-type-DNA-binding	34	WOX	21	MADS-M-type	34
ISWI	2	WRKY	106	MBF1	3
JmjC	39	YABBY	11	MED6	4
JmjC-ARID	2	ZF-HD	18	MED7	1
Continued					

PlantTFCat		PlantTFDBv5.0		iTAK	
Family	Frequency	Family	Frequency	Family	Frequency
JmjN	13	NF-X1	2	MTERF	32
JUMONJI	13	S1Fa-Like	2	MYB	165
Lambda-DB	5	SAP	1	MYB-related	91
LFY	1	Total	2128	NAC	98
LIM	25	Families	58	NF-YA	10
LisH	38			NF-YB	20
MADS-MIKC	42			NF-YC	16
MADS-type1	33			OFP	19
MYB	14			Others	78
MYB/SANT	30			PHD	53
MYB-HB-like	361			PLATZ	17
MYB-related	0			Pseudo ARR-B	8
NAM	99			RB	3
Nin-like	13			Rcd1-like	2
Nozzle	2			RWP-RK	13
PAZ-Argonaute	18			SAP	1
PHD	194			SBP	26
PLATZ	17			SET	50
RAV	3			SNF2	51
RB	3			SOH1	1
RR-A-type	63			SRS	13
RR-B-type	11			STAT	2
S1Fa-like	3			SW1/SNF-BAF60b	20
SAP	12			SW1/SNF-SW13	6
SBP	26			TAZ	8
SET	44			TCP	26
SNF2	58			Tify	18
ssDNA-binding-TF	6			TRAF	21
SSXT	5			Trihelix	34
STAT	2			TUB	12
STY-LRP1	11			ULT	1
SWIB-Plus-3	6			VOZ	3
TAZ	1			Whirly	3
TCP	26			WRKY	106
Tc-PD	3			zf-HD	18
Tesmin	6			NF-X1	2
TIFY	22			NOZZLE	1
TTF-type (Zn)	3			S1Fa-like	2
TUBBY	12			Total	2702
WD40-like	353			Families	93
WRKY	107				
YEATS	3				
ZF-HD	18				
Znf-B	45				
Znf-LSD	7				
Total	5573				
Families	98				

Table 2. Whole genome prediction of transcription factors and transcription regulators through PlantTFCAT, PlantTFDB v5.0 and iTAK in cowpea.

gene pairs were identified. The tandem duplicated genes were observed primarily in 4 groups, viz., CAMK, CMGC, TKL, and RLK-Pelle. About 95% (187) of the tandemly duplicated genes belonged to the RLK-Pelle group (Supplementary Table S4d). The number of tandem duplication events in each chromosome varied from 3 to 14 with chromosome 5 (35 genes) and chromosome 7 (31 genes) housing the maximum number of tandemly

Group/Family	No. of families	No. of PKs
<i>Group AGC</i>		
AGC_RSK-2, AGC_MAST, AGC_PDK1, AGC_NDR, AGC_PKA-PKG, AGC-PI	6	37
<i>Group CAMK</i>		
CAMK_CDPK, CAMK_OST1L, CAMK_CAMKL-CHK1, CAMK_CAMKL-LKB, CAMK_AMPK, CAMK_CAMK1-DCAMKL	6	86
<i>Group CK1</i>		
CK1_CK1, CK1_CK1-PI	2	14
<i>Group CMGC</i>		
CMGC_GSK, CMGC_CDK-CCRK, CMGC_CDK-PI, CMGC_DYRK-PRP4, CMGC_MAPK, CMGC_CDK-CRK7-CDK9, CMGC_DYRK-YAK, CMGC_CDK-CDK8, CMGC_CDK-CDK7, CMGC_CDKL-Cr, CMGC_CK2, CMGC_SRPK, CMGC_CLK, CMGC_RCK, CMGC_CDK-PITSLR, CMGC_GSKL, CMGC_PI-Tthe	17	76
<i>Group Others</i>		
IRE1, TTK, NEK, NAK, WNK_NRB, WEE, TLK, SCY1_SCYL2, Aur, BUB, PEK_GCN2, PEK_PPK, SCY1_SCYL1, ULK_Fused, ULK_ULK4	15	44
<i>Group-PI-3</i>	1	3
<i>Group-PI-4</i>	1	3
<i>Group-PI-2</i>	1	1
<i>Group RLK-Pelle</i>		
RLK-Pelle_LRR-I-2, RLK-Pelle_C-LEC, RLK-Pelle_LRR-Xb-2, RLK-Pelle_RLCK-XV, RLK-Pelle_LRR-III, RLK-Pelle_LRK10L-2, RLK-Pelle_RLCK-VI, RLK-Pelle_RLCK-VIIa-1, RLK-Pelle_RLCK-VIIa-2,		
RLK-Pelle_RLCK-X, RLK-Pelle_LRR-VII-1, RLK-Pelle_RLCK-V, RLK-Pelle_LRR-XI-1, RLK-Pelle_LRR-V, RLK-Pelle_DLSV, RLK-Pelle_LRR-IX, RLK-Pelle_RLCK-IXb, RLK-Pelle_CR4L, RLK-Pelle_LRR-I-1, RLK-Pelle_RLCK-IV, RLK-Pelle_RLCK-Os, RLK-Pelle_Singleton, RLK-Pelle_L-LEC, RLK-Pelle_WAK_LRK10L-1, RLK-Pelle_LRR-VIII-1, RLK-Pelle_CrRLK1L-1, RLK-Pelle_RLCK-XII-1, RLK-Pelle_PERK-2, RLK-Pelle_LRR-XII-1, RLK-Pelle_LRR-XIIIa, RLK-Pelle_PERK-1, RLK-Pelle_LRR-VI-2, RLK-Pelle_RLCK-XI, RLK-Pelle_LRR-IV, RLK-Pelle_LysM, RLK-Pelle_RLCK-II, RLK-Pelle_Extensin, RLK-Pelle_RLCK-VIII, RLK-Pelle_LRR-VII-2, RLK-Pelle_SD-2b, RLK-Pelle_LRR-II, RLK-Pelle_LRR-VI-1, RLK-Pelle_LRR-Xb-1, RLK-Pelle_WAK, RLK-Pelle_RLCK-XVI, RLK-Pelle_LRR-Xa, RLK-Pelle_RKF3, RLK-Pelle_RLCK-IXa, RLK-Pelle_URK-2, RLK-Pelle_LRR-XV, RLK-Pelle_LRR-XI-2, RLK-Pelle_LRR-XIIIb, RLK-Pelle_URK-1, RLK-Pelle_LRR-XIV, RLK-Pelle_LRR-VII-3, RLK-Pelle_RLCK-VIIb	56	772
<i>Group STE</i>		
STE_STE-PI, STE_STE11, STE_STE7, STE_STE20-YSK, STE_STE20-Fray, STE_STE20-PI	6	42
<i>Group TKL</i>		
TKL_CTR1-DRK-2, TKL-PI-6, TKL-PI-4, TKL-PI-5, TKL_CTR1-DRK-1, TKL-PI-1, TKL-PI-2, TKL-PI-8, TKL-PI-7, TKL-Gdt, TKL-Cr-3		57
<i>Total</i>	122	1135

Table 3. Classification of genome-wide protein kinases predicted in cowpea by iTAK.

arrayed VuPK genes. The number of VuPK genes within a tandem in each chromosome varied from 2 to 7, with chromosome 6 and chromosome 9 carrying the maximum genes per tandem event (Supplementary Table S4d).

The coding sequences of the VuPK genes undergo nucleotide substitutions that act as the driving force for natural selections to act upon. The ratio of non-synonymous to synonymous substitution rates (K_a/K_s) is often construed as an informative parameter of gene evolution under selection. The pairwise comparisons between the tandem duplicated genes showed that the K_s/K_a values varied between 0.08 and 7.75 (Supplementary Table S5) and the mean ratio of the tandem pairs was 0.67. Eighty-five percent of these gene pairs had less than a unit K_a/K_s ratio, suggesting their influence under purifying selection. About 15% of the gene pairs displayed more than a unit K_a/K_s ratio implicating the pertinent role of positive selection in driving their evolution.

The subcellular localizations of the VuPKs were also predicted through CELLO and LOCALIZER. Most of the PKs (419, 36.92%) were found localized to the plasma membrane followed by the nucleus (27.67%), cytoplasm (19.91%), chloroplast (6.43%), mitochondria (5.90%), extracellular (3.08%), and only one of the PKs (0.09%) was found localized to endoplasmic reticulum (Fig. 4). About 98.3% of the VuPKs localized to plasma membrane belonged to the RLK-Pelle group (Supplementary Table S6a). On the contrary, the LOCALIZER resolved 86 (6%), 46 (4.1%), 502 (44.2%) and 65 (5.7%) VuPKs to be localized to chloroplasts, mitochondria, nuclear with no transit peptides and nuclear with transit peptides, respectively (Supplementary Table S6b).

The isoelectric points (pIs) of the predicted VuPKs varied from 4.2 to 11.08, with MWs ranging from 9070 to 194,290 Da. The pIs and MWs of VuPKs varied widely within the groups exhibiting both extremes of values. CK1 was the only group displaying narrow intra-pI values (8.67–10.22) (Supplementary Table S7).

Validation of in silico determined R-genes, TAPs and PKs

All twenty genic primers designed from gene sequences of R-genes, TAPs, and PKs successfully amplified the target regions across the ten cowpea genotypes. While many primers exhibited monomorphic amplification patterns, a subset revealed presence/absence variations among the genotypes. The sizes of the amplified products matched the expected amplicon lengths precisely. The primer sequences along with their expected amplicon lengths are listed in Supplementary Tables S8a–c. Representative amplification profiles for two primers from each gene regulatory class are shown in Figs. 5, 6 and 7 (The original uncropped images of the gels are provided in Supplementary Fig. S3).

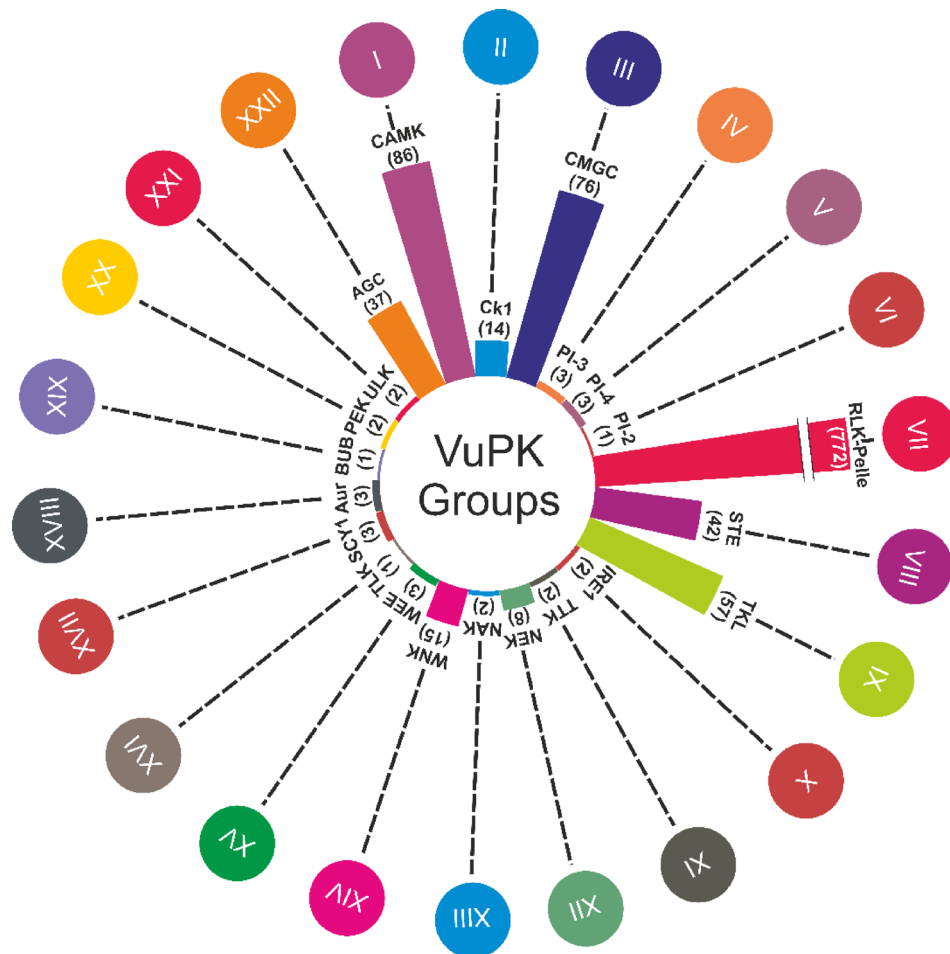


Fig. 1. Distribution of VuPKs among 22 kinase groups in the cowpea genome. The circular layout shows each VuPK group (outer ring, in Roman numerals), with bar length and associated Arabic numerals representing the number of genes per group. The variation in bar heights reflects the relative abundance of each group.

Differential gene expression under biotic and abiotic stresses

Biotic stress (cowpea aphid borne mosaic virus): The RNA-seq data from infected and non-infected cowpea plants showed significant upregulation of nine R-genes (Supplementary Table S9a). The R-genes belonged to typical NBS (1), TNL (1), RLK (1), RLP (1), LECRK (1) and KIN (4) classes with log₂FC in the range of 2.11 (KIN) to 3.11 (NBS). None of the R-genes were downregulated in the resistant cowpea genotype IT85F-2687 post infection with the virus. Twenty-four TFs belonging to AP2/ERF-ERF (8), WRKY (4), TIFY (3), bHLH (2), MYB (2), and one each of GRAS, Jumonji, TCP, SBP and C2H2 were significantly upregulated (log₂FC: 1.63–2.8). Alternately, 11 TFs, four of AP2/ERF-ERF, three of NACs, two of WRKYs and one each of MYB and PLATZ were significantly under expressed consequent to infection (log₂FC: – 2.82 to – 1.23). Six PKs (RLK-Pelle_LRR-IV, CMGC_CDK-CRK7-CDK9, STE_STE11, CAMK_CAMKL-CHK1, RLK-Pelle_DLSV and CMGC_CDK-CRK7-CDK9) were found upregulated after infection with log₂FC values in the range of 2.1–2.56, while three PKs (CAMK_CAMKL-CHK1, RLK-Pelle_LRR-I-1, RLK-Pelle_CR4L) were significantly downregulated (log₂FC of – 2.99 to – 1.48) owing to infection. The volcano plot of differentially expressed genes of cowpea plants infected with CABMV and the control plants is depicted in Fig. 8.

Abiotic stress (root dehydration): When cowpea plants were subjected to dehydration, 11 R-genes were over-expressed, while 18 genes were under-expressed (Supplementary Table S9b). R-genes belonging to the classes CTNL (2), TNL (2), TN, CK, NL, RLK, CNL, CLK and KIN (1 each) with NBS and or LRR and Kinase domains were upregulated (log₂FC: 1.15–2.45). Likewise, R-genes belonging to RLKs, KINs, TNs, TNLs, LECRK, NLs, CTNLs, CNs, CNLs, and CKs were downregulated (log₂FC: – 3.69 to – 1.11). Dehydration resulted in the enhanced expression of RLK-Pelle_DLSV and RLK-Pelle_LRR-XI-1 PKs (log₂FC: 1.26–2.45), while other classes of RLK-Pelle group (LRR-XI-2, RLCK-Os, SD-2b, DLSV, RLCK-VIII-2, LRR-XI-1, LRR10L-2) and AGC group (RSK-2) were under-expressed (log₂FC: – 2.47 to – 1.11). Incidentally, different isoforms of RLK-Pelle_DLSV and RLK-Pelle_LRR-XI-1 were both up- and down-regulated due to dehydration. The TFs MADS-MIKK, LOB, and HSFs were over-expressed (log₂FC: 1.08–1.75), while a good number of AP2/ERF (7) TFs were down-regulated along with others such as C2H2, WRKY, MYB, NAC, and GRAS (log₂FC: – 1.70 to – 1.04). The

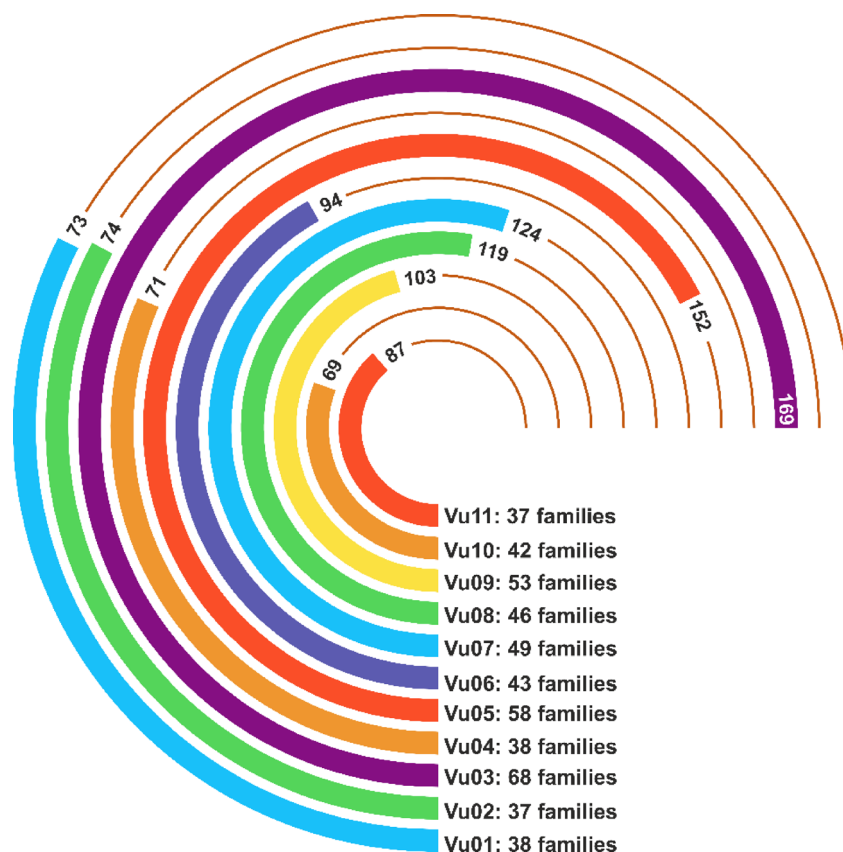


Fig. 2. Chromosomal distribution of VuPKs in the cowpea genome. Each arc represents one of the 11 chromosomes (Vu01–Vu11), with the numbers indicating the total count of VuPKs per chromosome. The length of each solid arc is proportional to the number of encoded VuPKs, reflecting their relative abundance. The number of distinct protein kinase families present on each chromosome is listed alongside.

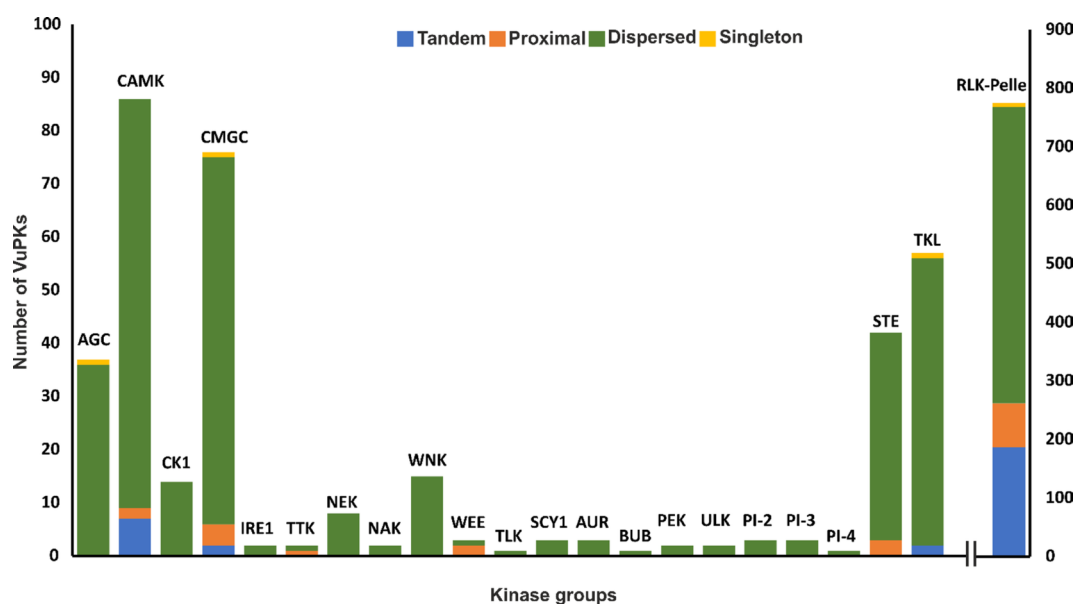


Fig. 3. Expansion mechanisms of protein kinases (VuPKs) across 22 kinase groups in the cowpea genome. The bar plot illustrates the number of VuPKs in each group, with color segments representing different duplication modes: tandem (blue), proximal (orange), dispersed (green), and singleton (yellow). Dispersed duplication is the most prevalent mechanism contributing to kinase group expansion.

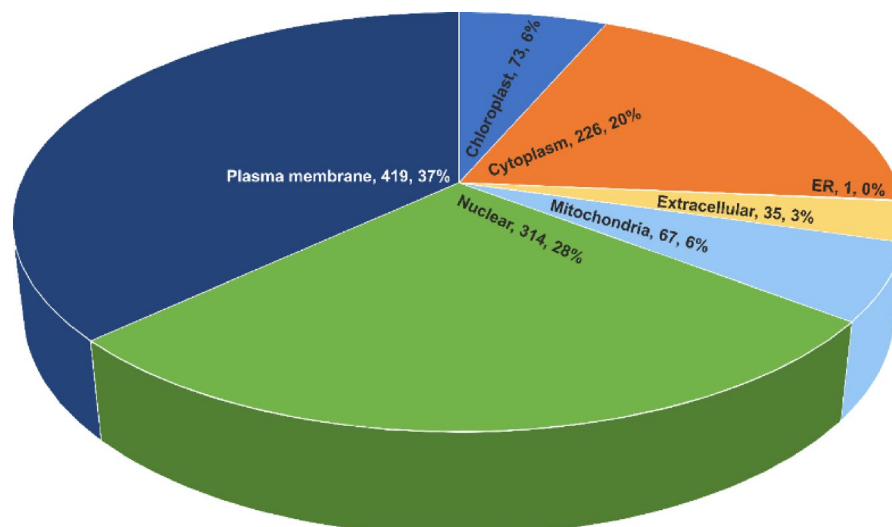


Fig. 4. Predicted subcellular localization of cowpea protein kinases (VuPKs). The pie chart illustrates the distribution of VuPKs across major cellular compartments. The majority localize to the plasma membrane (37%), followed by the nucleus (28%) and cytoplasm (20%).

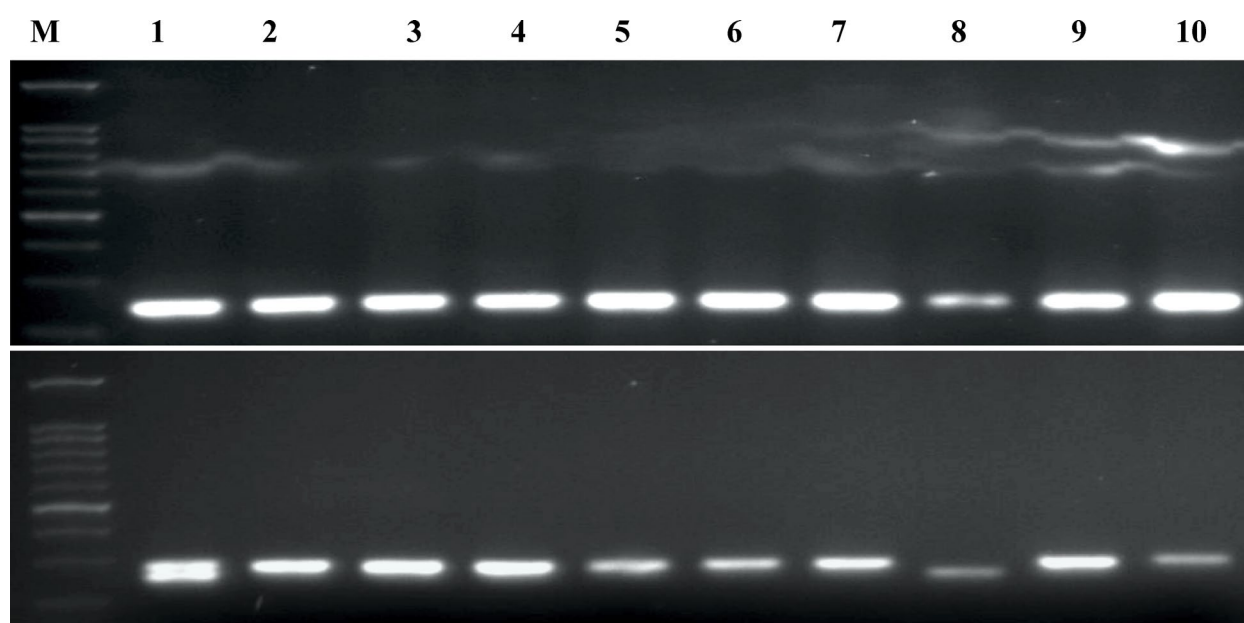


Fig. 5. PCR amplification profiles of R-gene-specific primers VuRGENE8 (top panel) and VuRGENE11 (bottom panel) across ten cowpea genotypes. Lane M: 100 bp DNA ladder; Lanes 1–10: GC3, TC901, C-152, PL-1, ARC-1, PLM211, NBC-1, Vu-89, VBN-3, and VBN-1, respectively.

volcano plot of differentially expressed genes of cowpea plants subjected to root dehydration and the control plants is depicted in Fig. 9.

Discussion

Cowpea, like other legumes, has evolved intricate molecular networks to mitigate diverse biotic and abiotic stresses. Whole-genome sequencing enables the comprehensive identification and characterization of molecular moderators, including R-genes, TAPs and PKs, facilitating the discovery of key regulators involved in plant stress responses. Our hybrid genome assembly of cowpea, leveraging Illumina and nanopore sequencing data, produced a high-quality draft. Although $\sim 30\times$ coverage is ideal for de novo nanopore assemblies, this study used $\sim 20\times$ nanopore data supplemented with $\sim 120\times$ Illumina reads, balancing cost and computational efficiency without compromising assembly quality. The assembly (~ 325 Mbp) attained $>93\%$ completeness (BUSCO), validating its robustness for functional annotation. The smaller genome size estimate (compared to a

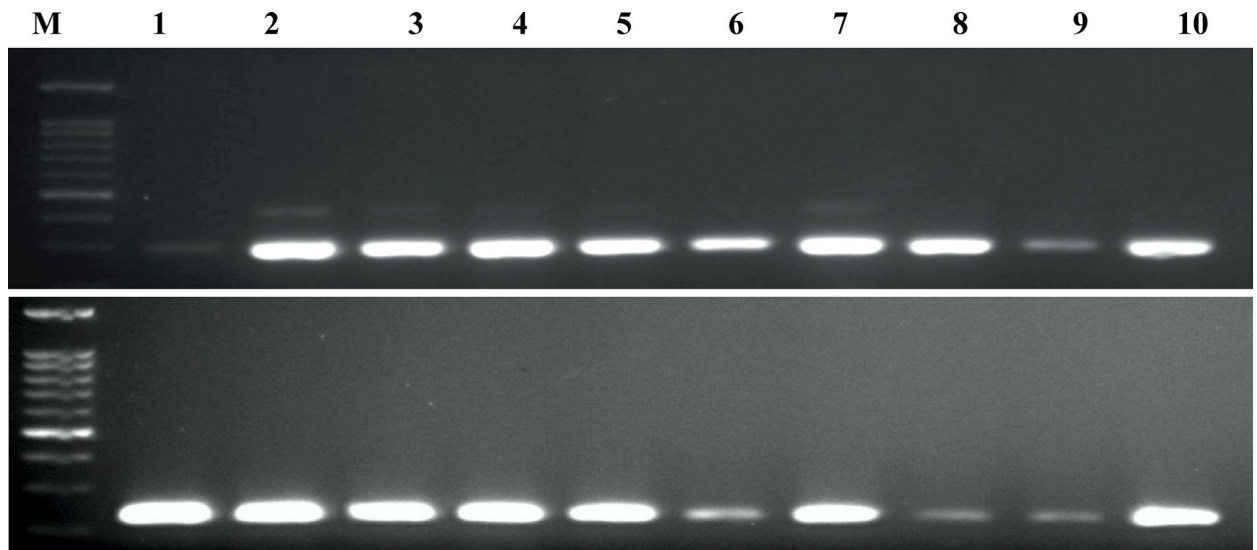


Fig. 6. PCR amplification profiles of TAP-gene specific primers VuTAP2 (top panel) and VuTAP11 (bottom panel) across ten cowpea genotypes. Lane M: 100 bp DNA ladder; Lanes 1–10: GC3, TC901, C-152, PL-1, ARC-1, PLM211, NBC-1, Vu-89, VBN-3, and VBN-1, respectively.

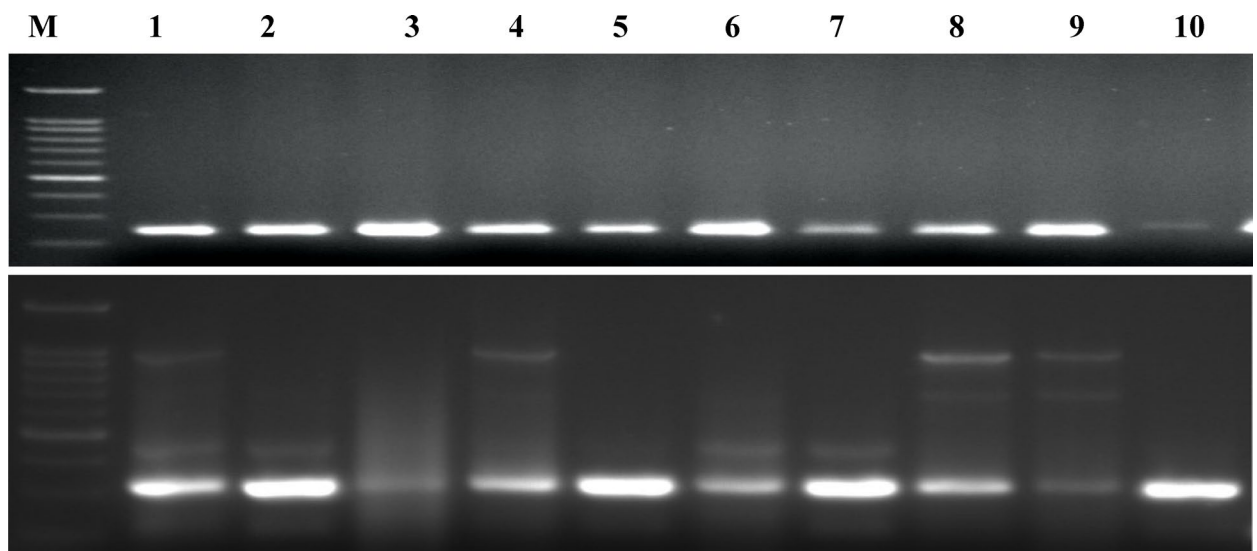


Fig. 7. PCR amplification profiles of PK-gene specific primers VuPK12 (top panel) and VuPK16 (bottom panel) across ten cowpea genotypes. Lane M: 100 bp DNA ladder; Lanes 1–10: GC3, TC901, C-152, PL-1, ARC-1, PLM211, NBC-1, Vu-89, VBN-3, and VBN-1, respectively.

previous estimate of 519 Mbp⁵) is likely due to the limitation of hybrid assemblers in collapsing repeat regions⁵⁵. Optical map-based PacBio sequencing is a more reliable estimator of genome size, in repeat-rich genomes like cowpea⁵. However, this limitation did not compromise our ability to identify key gene families. Functional annotation of biomolecules through computational prediction tools such as Hidden Markov Models (HMM)⁵⁶ utilizes conserved domains and structural features, offering a cost- and time-effective alternative to experimental methods⁵⁷.

R-genes

R-genes play a central role in plant immunity by encoding proteins that recognize pathogens and trigger defence responses. Widely used in resistance breeding, the predominant class (NB-LRR or NLR genes⁵⁸) features a conserved nucleotide-binding domain and a variable leucine-rich repeat domain that determines pathogen specificity. R-genes confer resistance to a wide range of pathogens—including bacteria, fungi, viruses, and nematodes—despite encoding a limited set of proteins with conserved domains^{58–60}. Through their modular structure, R-proteins can both recognize pathogen effectors (AVR proteins) and modulate defence signalling.

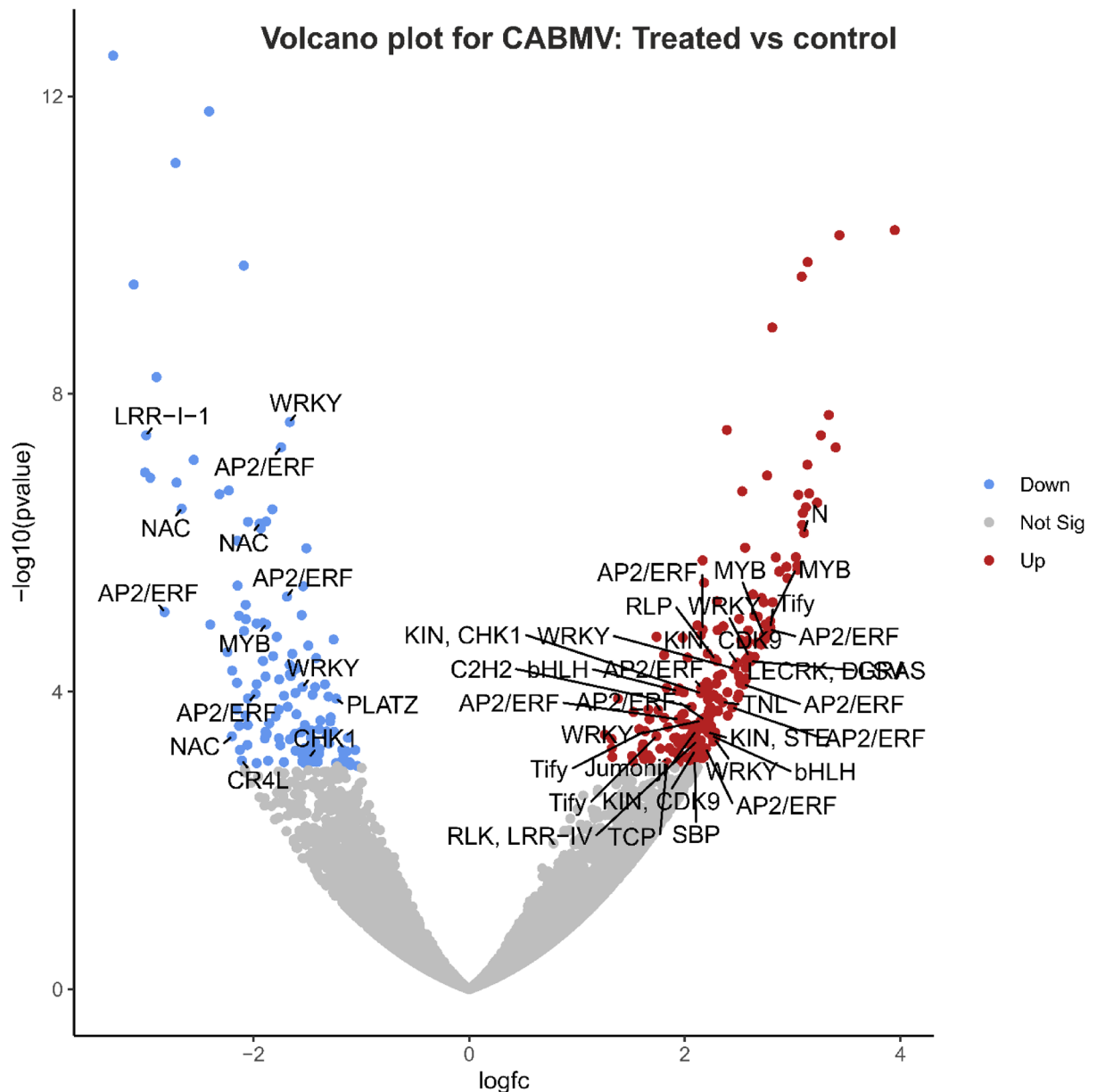


Fig. 8. Volcano plot depicting differential expression of R-genes, transcriptionally active proteins (TAPs), and protein kinases (PKs) in cowpea plants infected with cowpea aphid-borne mosaic virus (CABMV) compared to control plants. Red dots indicate significantly upregulated genes, blue dots represent significantly downregulated genes, and grey dots denote non-significant changes. Selected gene families with significant differential expression are annotated on the plot.

Recognition follows one or more of four models: direct interaction (elicitor-receptor^{61,62}), indirect sensing via modified host targets (guard model⁵⁹), detection of altered decoy proteins (decoy model⁶³), or incorporation of decoy-like domains within NLRs (integrated decoy model^{64,65}).

With the increasing accessibility of whole-genome sequences, comprehensive analyses of R-genes have been undertaken across several crops, including blackgram, mungbean, chickpea, rice, tomato, *Medicago*, and *Arabidopsis*⁶⁶. In dicots, R-genes typically represent 0.18% (papaya)⁶⁷ to 5.3% (*Arabidopsis*)⁶⁰ of the total gene content. In our study, R-genes comprised 3.3% of the cowpea genome, positioning it well within the reported range. While the proportion is notably higher than *Medicago*⁵⁸ (1.2%), it is comparable to blackgram⁶⁶ (3.9%), highlighting cowpea's relatively rich repertoire of immune-related genes among legumes. Strikingly, NBS-domain containing genes accounted for 17.9% of total R-genes in cowpea, more than double the proportion observed in blackgram (8.6%)⁶⁶. The 392 NBS-domain genes identified in this study is in line with a previous report in cowpea (402)⁵. It also matches the counts reported in other crops such as sorghum (346)⁶⁸, soybean (319)⁶⁹, common bean (325)⁷⁰ and *Arachis duranensis* (393)⁷¹, reinforcing the evolutionary conservation and functional

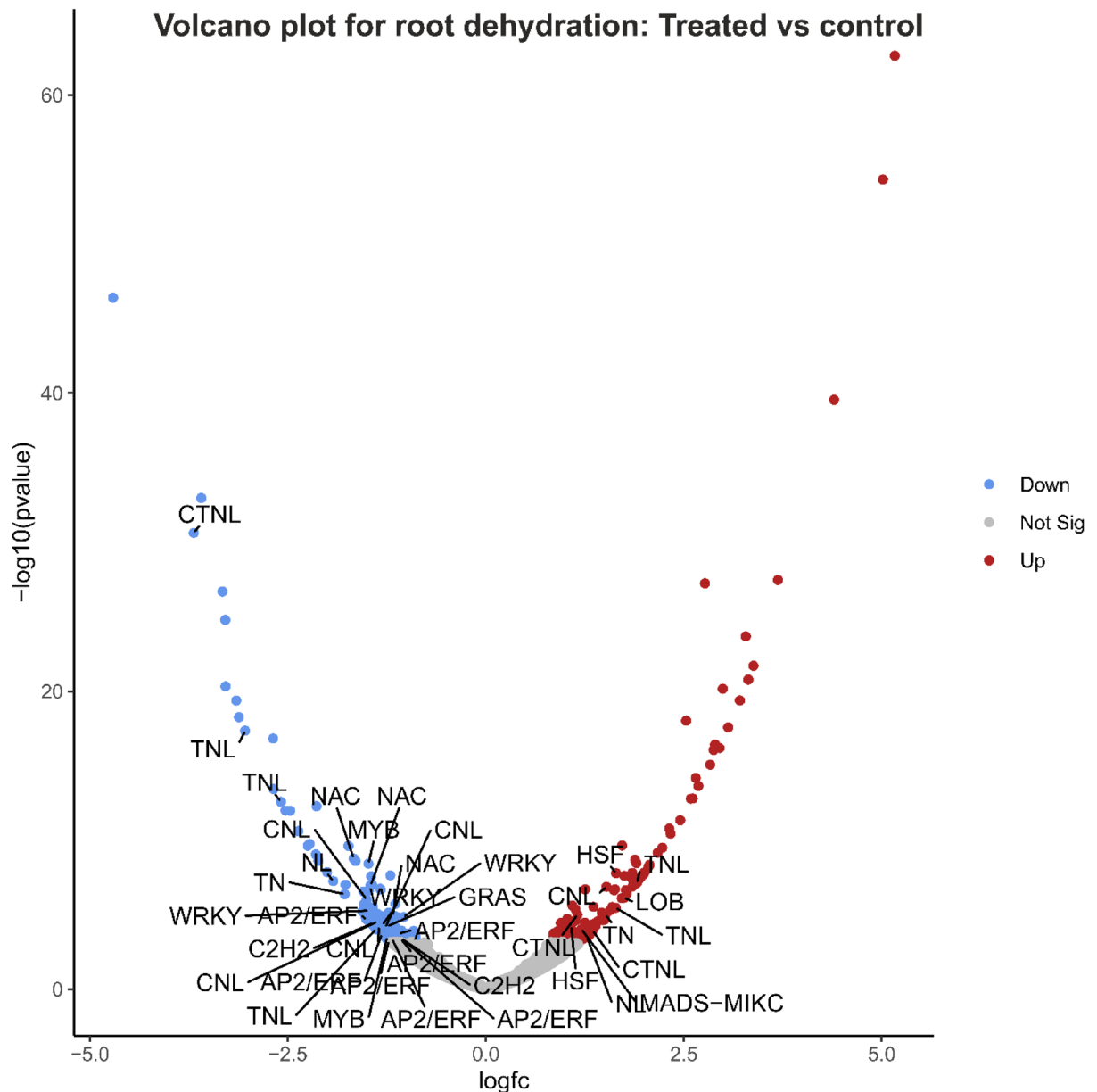


Fig. 9. Volcano plot showing differential expression of R-genes, transcriptionally active proteins (TAPs), and protein kinases (PKs) in cowpea plants subjected to root dehydration stress compared to control plants. Red dots represent significantly upregulated genes, blue dots indicate significantly downregulated genes, and grey dots correspond to non-significant changes. Notable gene families exhibiting significant differential expression are annotated.

importance of this class. NBS-LRR proteins, including TIR-NBS-LRR (TNL) and coiled coil-NBS-LRR (CNL), are central to effector triggered immunity (ETI), a critical defence response against pathogen effectors⁷². Their broad involvement in resistance against fungal (wheat stripe/stem rust^{73,74}, barley powdery mildew⁷⁵, flax rust⁷⁶, downy mildew in *Arabidopsis*⁷⁷), viral (tobacco mosaic virus in tobacco⁷⁸) and bacterial (rice blight⁷⁹, *Arabidopsis* bacterial wilt⁸⁰) diseases across crops is well documented, yet their functional relevance in cowpea remains underexplored. Additionally, receptor like kinases (RLK) and receptor like proteins (RLP), which mediate pathogen associated molecular pattern triggered immunity (PTI)⁷², comprised 23.3% of total R-genes in cowpea (Table 1), aligning closely with estimates in mungbean (25.7%)⁶⁶. Significant differences were observed across the R-gene classes between reference and our genomes. CLEC class was detected exclusively in ours, whereas the LYP class was unique to the reference genome. This presence-absence variation suggested genotypic divergence likely shaped by selective pressures or breeding history⁸¹. Such variation also reflects the dynamic nature of the cowpea pan-genome, where non-core genes (genes present in some individuals but not all), often involved in stress responses, contribute disproportionately to genetic diversity⁸¹. In addition, our genome also exhibited

an enrichment of lectin-domain (L, LEC, CLEC) and lysin-motif (LYS) containing R-genes (Supplementary Table S1c), which are key sensors of pathogen-associated molecular patterns like chitin and peptidoglycans, the hallmarks of fungal and bacterial pathogens⁸². In contrast, the reference genome harboured a higher proportion of canonical R-genes belonging to CN, CNL, CTNL, RLK, and TNL classes. Thus, our genome is primarily augmented with PTI-related R-genes, while the reference genome shows relative abundance of ETI-associated R-proteins. This contrasting distribution reveals divergent evolutionary trajectories of immune gene families and suggests potential specialization in pathogen defence across cowpea genotypes. Collectively, these findings not only highlight the richness and diversity of R-genes in cowpea but also underscore the importance of genome-level exploration in uncovering genotype-specific resistance mechanisms. They also provide a strong foundation for functional studies and targeted breeding efforts aimed at enhancing disease resistance in this climate-resilient legume.

Transcription factors (TFs) and transcription associated proteins (TAPs)

TAPs, including TFs, TRs and putative proteins, orchestrate complex gene expression networks that enable plants to respond to developmental cues and environmental stimuli. While TFs directly bind cis-regulatory elements, TRs often function as coactivators/corepressors or as chromatin remodellers. In the pursuit of developing climate-resilient pulse crop varieties, the TFs which form the key regulators for stress and developmental responses, are of paramount importance⁸³.

While several curated databases exist for TF and TAP identification, no single pipeline captures their full diversity¹³. To address this, we employed a combinatorial approach using PlantTFcat, PlantTFDB, and iTAK, leading to the identification of 6464 TAP-encoding domains from 5226 transcripts-accounting for 9.8% of the genome. This result aligns well with previous reports in cowpea (~7.26% of the transcriptome)¹³. Although the number of TF families (118) identified in this study was lower than that of Misra et al.¹³ (136 families), two families (ABTB and CW-Zn-B3_VAL) were uniquely revealed in our assembly, potentially reflecting the increased sensitivity of our hybrid assembly and BRAKER-based annotation strategy. However, these two families were discovered in other legumes such as common bean¹³. Pipeline-specific differences were also observed while comparing previous annotations by Misra et al.¹³. They identified five TAP families (CW-Zn-B3_VAL, Dicer, JmjC-ARID, Rel, and RF-X) exclusively from raw cowpea genome using MAKER⁸⁴ and AUGUSTUS⁸⁵ gene prediction tools. Contrarily, we were able to discern first three of the five TAP families even from our transcripts, reinforcing the superiority of BRAKER-based gene prediction. Our study also strengthens the conservation of stress-regulatory TFs such as NAC and WRKY in cowpea. We identified 99 NAC and 106 WRKY genes that were consistent with earlier reports (vs 90 NAC⁸³ and 92 WRKY⁸⁶). Low-copy TF families (Table 2; HRT, LFY, MED7, SOH1, ULT with single copy and others with few copies) although less represented, play important roles with their specialized, tightly regulated and often conserved functions (Table 2). Together with minimal functional redundancy (have few or no paralogs), they serve as strategic targets for precision crop improvement through gene editing or transgenic approaches. For instance, in soybean, editing *GmE1* (a B3-domain low copy TF) led to early flowering under long day conditions⁸⁷. Sadhukhan et al.⁸⁸ identified a *DREB2* ortholog in cowpea (*VuDREB2A*) with implications for imparting drought tolerance and confirmed the role of this potential candidate gene in conferring water stress tolerance through a transgenic approach. Comparative analysis with the reference genome revealed differential enrichment of TAP families. While all the TAP families in the reference genome were discoverable in ours, three families, RB, ULT, and STAT, were exclusive to our CDS genome. The reference genotype was richer in TFs associated with abiotic/biotic stress responses (e.g., bZIP, NAC, WRKY, MYB-related, C3H)⁸⁹, and floral and meristem identity (e.g., MADS-MIKC, FAR1, HB-BELL, LUG, PHP)⁹⁰. This suggests its adaptive advantage under environmental stresses such as drought, salinity, or pathogen attack. PHP may offer additional epigenetic regulation of flowering time. Conversely, CDS harboured greater abundance of TFs involved in developmental regulation⁹⁰, including AP2/ERF AP2, B3, C2C2-LSD, and CPP. AP2/ERF and B3 TFs are known to regulate seed development and hormone signalling⁹¹, while C2C2-LSD is implicated in fine-tuning programmed cell death to limit pathogen spread⁹². Thus, CDS may exhibit enhanced developmental plasticity, earlier flowering, or higher seed yield under non-stress or moderate stress conditions. These findings underscore functional divergence between the genotypes. While the reference genotype appears better adapted for stress-prone environments, CDS may be optimized for reproductive success and yield stability. Crossbreeding strategies incorporating both could yield cultivars with synergistic improvements in stress resilience and productivity.

Protein kinases (PKs)

Protein kinases form one of the most expansive and functionally diverse gene families in plants, orchestrating complex signalling networks essential for development, environmental sensing, and stress adaptation. In the present study, we identified 1135 VuPKs in cowpea, accounting for 3.6% of predicted proteins, consistent with proportions observed in common bean⁹³ (3.3%), *Arabidopsis*⁹⁴ (3.4%), cucumber⁹⁵ (3.69%), grapevine⁹⁶ (3.7%), pineapple⁹⁷ (2.8%), but lower than in soybean⁹⁸ (4.7%). This reflects the evolutionary conservation of this regulatory machinery across angiosperms. A slightly higher proportion (4%) and a larger gene count (1298) in a previous cowpea kinome⁴⁷, is likely due to differences in the genotypes and sequencing depth. We identified 22 PK groups, including two additional ones (NAK and TLK) unreported in the reference genome, and 122 families, of which five were novel (TKL-Cr-3, RLK-Pelle-URK-1, RLK-Pelle-URK-2, NAK and TLK). One (TKL-PI-3) of the 118 families in the reference genome-based kinome⁴⁷ remained elusive in our study. These newly resolved families, largely underexplored in plant systems^{94,99}, expand the known functional repertoire of PKs and highlight the potential for discovering genotype-specific signalling components. Their exclusive detection in our genome suggests lineage-specific expansions or adaptive retention, offering valuable leads for functional validation and targeted crop improvement. Incorporating findings from both studies bring the

total known cowpea PK repertoire to at least 123 families across 22 groups. The RLK-Pelle group dominated the cowpea kinome (~68%), mirroring trends in other crops (63–75% in common bean⁹³, *Arabidopsis*⁹⁴, pineapple⁹⁷, soybean⁹⁸). This was followed by CAMK, CMGC, TKL, STE and AGC; together forming 94% of the kinome (Table 3, Fig. 1). All groups in the reference genome and ours showed similar abundance excepting RLK-Pelle group. In the former, we observed a relative enrichment of families like DLSV, LRR10L-2, LRR-XI-1, LRR-XII-1, and WAK within the RLK-Pelle group. This possibly reflects an evolved and diversified receptor system, likely enhancing the plants' ability to sense and respond to broad range of environmental cues and pathogens¹⁰⁰. The predominance of RLK-Pelle_DLSV family (~77% of RLK-Pelle group and 11.7% of total VuPKs) and the hierarchy of abundance of different PK groups (RLK-Pelle > CAMK > CMGC > TKL > STE) align with observations in different crop species including common bean⁹³ (Fig. 1). The low representing atypical PK groups with minimal functional redundancy (PI02, TLK, BUB, IRE1, TTK, NAK, PEK, ULK, PI-3, PI-4, SCY1, Aur, and WEE) showed congruency with a previous study⁴⁷ and may serve unique regulatory roles, making them promising candidates for gene function studies. Spatially, VuPKs were unevenly distributed across chromosomes, with Vu3 and Vu5 exhibiting the greatest abundance and diversity, while Vu10 carried the least VuPKs (Fig. 2). This finding corroborated with a previous study in cowpea⁴⁷ and also mirrored syntenic patterns seen in common bean⁹³, where chromosomes Pv8 and Pv10 (syntenic with Vu5 and Vu10)⁵ showed similar trends. The predominance of intron-containing PKs (86.6%) suggests evolutionary selection for structural complexity, potentially enhancing regulatory versatility¹⁰¹. The extent of intron-less PKs observed (13.4%) was similar to the previous reports in cowpea (13.6%)⁴⁷ and common bean (13.5%)⁹³, well within the range reported in other crops (9.5%–16.6% in grapevine⁹⁶, pineapple⁹⁷, and wheat¹⁰²). The maximum introns per family (28) observed in the study is the same as that in other Fabids including common bean⁹³ and soybean⁹⁸.

Gene duplication is a pivotal mechanism driving genome evolution and functional diversification responsible for the vast expanse of PKs in plants⁹⁹. Importantly, dispersed duplication emerged as the primary mechanism driving VuPK expansion (74.2%), followed by tandem (17.4%) and proximal (7.5%) duplications (Fig. 3). This pattern contrasts with legumes like common bean⁹³ and soybean⁹⁸, where whole-genome or segmental duplications predominate. Lack of recent polyploidy events and transposon-rich genome⁵ facilitated dispersed and tandem duplications in cowpea, responsible for the expansion of ~82% of VuPKs. While all three non-WGD mechanisms were distinct in RLK-Pelle, CAMK, and CMGC groups (Fig. 3), dispersed duplication exclusively was responsible for expansion in 14 specific groups (Fig. 3), notably CK1, NEK and WNK. Copies emanating through dispersed duplication might be the outcome of different transposition events (replicative, non-replicative or conservative) occurring in different plant genomes^{42,103}. Tandem duplication was the second largest event forcing the expansion of 17.4% of VuPKs as previously deduced in cowpea⁴⁷ and common bean⁹³. PKs expanded through tandem duplication often play roles in biotic stress responses⁹⁹, and over 85% of tandem duplicated gene pairs exhibited $K_a/K_s < 1$, suggesting purifying selection and potential functional redundancy, buffering against gene loss during diversification. Subcellular localization analysis showed a striking 98.3% of VuPKs targeted to the plasma membrane, belonged entirely to the RLK-Pelle group, consistent with their roles as transmembrane receptors in pathogen detection and hormonal signalling¹⁰⁴. Other VuPKs localized to diverse compartments, including the nucleus, cytoplasm, chloroplast, mitochondria, extracellular space and endoplasmic reticulum, reflecting their functional breadth across signalling axes (Fig. 4, Supplementary Table S6a). The biochemical parameters of VuPKs like pI and MW varied extremely even within the groups like in common bean⁹³ but contrasted to that in other crops like grapevine⁹⁶, wherein the values generally remained similar within a group.

PCR validation of in silico determined genes

All designed primers successfully amplified the expected targets, validating the utility of the genomic data. Due to strong purifying (negative) selection as discussed above, short genic regions (~200–300 bp) within exons typically exhibit low polymorphism¹⁰⁵. Nevertheless, some primers captured presence-absence variations (Fig. 5), a common feature in regulatory gene families¹⁰⁶.

Interplay of R-genes, TFs and PKs under biotic and abiotic stresses

The expression dynamics of R-genes, TAPs, and PKs revealed distinct stress-specific regulatory patterns in cowpea. In the present study, nine R-genes were specifically induced in response to cowpea aphid borne mosaic virus (CABMV) infection. These gene classes included four kinases, one each of a TNL, RLK, RLP and LECRK, in addition to a canonical NLR, aligning with their established roles in pathogen perception and signal activation^{107,108}. Though hardly reported in cowpea, such activation, mirrors findings in other legumes. For instance, *Co-1* to *Co-10* confer resistance to anthracnose (*Colletotrichum lindemuthianum*) in common bean¹⁰⁹, while *Phg-1* to *Phg-5* and *I* genes provide resistance against angular leaf spot and bean common mosaic virus, respectively¹⁰⁹. In soybean, the *Rps1-Rps8*, *rhg1-Rhg4*, *Rsv1-Rsv4* and *Rpp3* (a TNL) mediate resistance to *Phytophthora sojae*¹¹⁰, soybean cyst nematode¹¹¹, soybean mosaic virus¹¹², and *Phakopsora*¹¹³ (rust), respectively. In chickpea, the AB4.1 QTL associated with *Ascochyta* blight encompassed 12 predicted genes including those annotated as NBS-LRR RLK, WAK, zinc finger protein, and STPK¹¹⁴. In mungbean, several NLRs (VrNBS) showed significant activation response to mungbean yellow mosaic India virus (MYMIV)¹¹⁵. Interestingly, emerging evidence suggests that R-genes, especially those encoding NBS-LRR proteins, may also contribute to abiotic stress responses. In this study, RNA-seq data revealed differential regulation of 29 R-genes under root dehydration stress, with 11 upregulated and 18 downregulated genes, many belonging to the NBS-LRR class. Comparable patterns have been reported in other legumes. In grass pea, nine LsNBS genes (including *LsNBS-D18*, *LsNBS-D204*, and *LsNBS-D180*) exhibited significant stress-dependent expressions (both up- and down-regulation) under salt stress¹¹⁶. In *Arabidopsis*, overexpression of *ADRI*, an NLR gene, enhanced drought tolerance¹¹⁷. Such findings point to a broader functional scope of R-genes, suggesting their involvement in both biotic and abiotic stress signalling, potentially mediated through crosstalk with hormone-regulated pathways.

TFs demonstrated a complex and context-specific response. CABMV infection upregulated families like TIFY, GRAS, bHLH, TCP, C2H2, SBP, and Jumonji, whereas NAC and PLATZ were exclusively downregulated. TFs like WRKY, MYB, and AP2/ERF showed mixed response. Most of the upregulated TFs like AP2/ERF, MYB, bHLH are majorly intricately involved in regulating and synthesizing secondary metabolites like phenols, lignin, flavonoids, tannins etc. under biotic stress in various crops^{118,119}. Simultaneously, many of these TFs are also involved in growth and developmental processes¹¹⁸. Therefore, under a given stress, isoforms of these TFs could show contrasting response within the same genotype as evident in this study. TFs are largely implicated in abiotic stress tolerance. In cowpea, two NAC genes, *VuNAC1* and *VuNAC2*, isolated from a drought-hardy genotype imparted tolerance to multiple abiotic stresses such as drought, salinity and oxidative stresses^{7,8}. The soybean NAC (*GmNAC109*) and WRKY (*GmWRKY13*, *GmWRKY21*, and *GmWRKY54*) genes enhance lateral root growth and contribute to drought and salt stress alleviation^{120,121}. A chickpea MYB (1R-MYB) has been reported to co-regulate drought tolerance¹²². TFs like bZIP play crucial roles in ABA-mediated signalling pathways and are involved in modulating responses to abiotic stresses like drought, salinity and temperature extremes, such as *OsZIP62* in rice¹²³. AP2/ERF and DREB TFs are integral to regulating gene expressions in response to abiotic stresses such as drought, salinity and cold as in cowpea¹³ and mungbean¹²⁴. Such TFs typically associated with abiotic stresses including drought were predominantly downregulated in this study. This atypical response may reflect severe stress adaptation (roots completely exposed to air), genotype-specific repression, or shift toward growth arrest and resource allocation^{125,126}. In contrast, MADS-MIKC, LOB, and HSF TFs were upregulated under dehydration, suggesting alternative pathways that contribute to root development and protective responses^{127,128}.

A similar trend was observed in PKs. Biotic stress induced RLK-Pelle, CAMK, and CMGC kinase groups—consistent with their roles in early signal transduction and immune response¹⁹. PKs belonging to RLK-Pelle, CMGC, STE and CAMK were also found upregulated in cowpea subjected to CABMV and CPSMV viral infections⁴⁷. Similarly, significant involvement of RLK-Pelle and CAMK families in response to various stressors were elucidated in sunflower¹²⁹. Interestingly, different isoforms of the PKs belonging to same family of CAMK group were up- and downregulated under CABMV infection. While many PKs are involved in stress response, other isoforms have roles in development and may be downregulated because of the need for resource allocation upon stress treatment⁹⁹. However, several PKs were suppressed under dehydration, possibly reflecting a stress-phase-specific metabolic adjustment. Many families belonging to RLK-Pelle group (LRR-XI-2, RLCK-Os, SD-2b, RLCK-VIIa-2, LRR-XI-1, LRK10L-2) and AGC group (RSK-2) were downregulated under root dehydration stress. Likewise, downregulation of RLCK-VIIa-2 was also observed in wheat under waterlogging conditions¹³⁰. Like in biotic stress, different isoforms within the same family (DLSV and LRR-XI-1) were contrastingly expressed under root dehydration. This is congruent to similar observations in wheat¹³⁰. Interestingly, RLK-Pelle_DLSV was upregulated under both stresses, underscoring its potential role as a convergent signalling hub, similar to reports in cowpea, wheat and sunflower^{47,129,130}.

Thus, the R-genes are primarily involved in signal perception triggering immunity against invading pathogens, while the PKs are implicated in relaying the signal from the membrane to the nucleus. The signal transduction through their cascading effect phosphorylates or dephosphorylates the TFs, which modulate expression of stress-responsive genes by binding on to the promoters or cis-elements. The interplay between these three groups occurs at various levels and their feedback ensures a dynamic, context specific response that balances defence and growth of the plants.

Conclusion

Cowpea is a hardy legume of high agricultural value, particularly in the context of climate change. As it frequently encounters various stresses, identifying and understanding the roles of key regulatory elements, such as R-genes, TAPs and PKs in enduring these stresses is essential. The present study provides valuable insights into the repertoire, structural diversity, functional profiles, and genomic organization of these regulatory elements. The highly diversified and structurally complex regulatory units, enriched with novel and under characterized gene families may hold the key to unlocking stress tolerance and signalling specificity. The genotype-specific presence of unique gene groups and classes of the regulatory units, underscores cowpea's evolutionary innovation in signal transduction. This presents rich opportunities for molecular breeding and translational research towards developing climate-smart cowpeas.

Data availability

The NGS genomic datasets generated during the current study are available in the NCBI SRA repository [under the accession number PRJNA858559]. All other data generated or analysed during this study are included in this published article (and its Supplementary Information files).

Received: 11 November 2024; Accepted: 1 September 2025

Published online: 25 September 2025

References

1. Ayalew, T. & Yoseph, T. Cowpea (*Vigna unguiculata* L. Walp.): A choice crop for sustainability during the climate change periods. *J. Appl. Biol. Biotechnol.* **10**, 154–162. <https://doi.org/10.7324/JABB.2022.100320> (2022).
2. Gonçalves, A., Ribeiro, T., Silva, L. R. & Ferreira, I. C. F. R. Cowpea (*Vigna unguiculata* L. Walp.), a renewed multipurpose crop for a more sustainable agri-food system: Nutritional advantages and constraints. *J. Sci. Food Agric.* **96**, 2941–2951. <https://doi.org/10.1002/jsfa.7644> (2016).
3. FAOSTAT. Food and agriculture data. <https://www.fao.org/faostat> (2022).

4. Mohammed, S. B., Shehu, M. Y. & Singh, B. B. Appraisal of cowpea cropping systems and farmers' perceptions of production constraints and preferences in the dry savannah areas of Nigeria. *CABI Agric. Biosci.* **2**, 25. <https://doi.org/10.1186/s43170-021-00046-7> (2021).
5. Lonardi, S. et al. The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J.* **98**, 767–782. <https://doi.org/10.1111/tjp.14349> (2019).
6. Gururani, M. A., Venkatesh, J. & Tran, L.-S.P. Plant disease resistance genes: Current status and future directions. *Physiol. Mol. Plant Pathol.* **78**, 51–65. <https://doi.org/10.1016/j.pmpp.2012.01.002> (2012).
7. McDowell, J. M. & Woffenden, B. J. Plant disease resistance genes: Recent insights and potential applications. *Trends Biotechnol.* **21**, 178–183. [https://doi.org/10.1016/S0167-7799\(03\)00053-2](https://doi.org/10.1016/S0167-7799(03)00053-2) (2003).
8. van Ooijen, G., van den Burg, H. A., Cornelissen, B. J. C. & Takken, F. L. W. Structure and function of resistance proteins in solanaceous plants. *Annu. Rev. Phytopathol.* **45**, 43–72. <https://doi.org/10.1146/annurev.phyto.45.062806.094430> (2007).
9. Li, P. et al. RGAugury: A pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics* **17**, 852. <https://doi.org/10.1186/s12864-016-3197-x> (2016).
10. Jones, J. D. G., Vance, R. E. & Dangl, J. L. Intracellular innate immune surveillance devices in plants and animals. *Science* **354**, aaf6395. <https://doi.org/10.1126/science.aaf6395> (2016).
11. Kourtelis, J., Sakai, T., Adachi, H. & Kamoun, S. RefPlantNLR is a comprehensive collection of experimentally validated plant disease resistance proteins from the NLR family. *PLoS Biol.* **19**, e3001124. <https://doi.org/10.1371/journal.pbio.3001124> (2021).
12. Liu, L., White, M. J. & Singer, C. E. Transcription factors and their genes in higher plants. *Eur. J. Biochem.* **262**, 247–257. <https://doi.org/10.1046/j.1432-1327.1999.00349.x> (1999).
13. Misra, V. A., Wang, Y. & Timko, M. P. A compendium of transcription factor and transcriptionally active protein coding gene families in cowpea (*Vigna unguiculata* L.). *BMC Genomics* **18**, 898. <https://doi.org/10.1186/s12864-017-4287-7> (2017).
14. Filiz, E., Vatansever, R. & Ozyigit, I. I. Bioinformatics database resources for plant transcription factors. In *Plant Bioinformatics* (eds Hakeem, K. R. et al.) 97–116 (Springer, 2017). https://doi.org/10.1007/978-3-319-67156-7_5.
15. Riechmann, J. L. et al. *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* **290**, 2105–2110. <https://doi.org/10.1126/science.290.5499.2105> (2000).
16. Qu, L. J. & Zhu, Y. X. Transcription factor families in *Arabidopsis*: Major progress and outstanding issues for future research. *Curr. Opin. Plant Biol.* **9**, 544–549. <https://doi.org/10.1016/j.pbi.2006.07.005> (2006).
17. Richardt, S., Lang, D., Reski, R., Frank, W. & Rensing, S. A. PlantAPDB, a phylogeny-based resource of plant transcription-associated proteins. *Plant Physiol.* **143**, 1452–1466. <https://doi.org/10.1104/pp.106.091264> (2007).
18. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147–151. <https://doi.org/10.1038/nature01763> (2003).
19. Lehti-Shiu, M. D. & Shiu, S.-H. Diversity, classification and function of the plant protein kinase superfamily. *Philos. Trans. R. Soc. B* **367**, 2619–2639. <https://doi.org/10.1098/rstb.2012.0003> (2012).
20. Wang, P. et al. Mapping proteome-wide targets of protein kinases in plant stress responses. *Proc. Natl. Acad. Sci. USA* **117**, 3270–3280. <https://doi.org/10.1073/pnas.1919901117> (2020).
21. Dhanasekar, P. & Souframanien, J. Gamma-rays induced genome wide stable mutations in cowpea deciphered through whole genome sequencing. *Int. J. Radiat. Biol.* **100**, 1072–1084. <https://doi.org/10.1080/09553002.2024.2345087> (2024).
22. Zimin, A. V. et al. The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677. <https://doi.org/10.1093/bioinformatics/btt476> (2013).
23. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **117**, 9451–9457. <https://doi.org/10.1073/pnas.1921046117> (2020).
24. Smit, A.F.A., Hubley, R. & Green, P. RepeatMasker Open-4.0. <http://www.repeatmasker.org> (2015).
25. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351> (2015).
26. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5 (2019).
27. García, J. C. et al. PRGdb 4.0: An updated database dedicated to genes involved in plant disease resistance process. *Nucleic Acids Res.* **50**, D1483–D1490. <https://doi.org/10.1093/nar/gkab1087> (2022).
28. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. <https://doi.org/10.1093/molbev/msy096> (2018).
29. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37. <https://doi.org/10.1093/nar/gkr367> (2011).
30. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164. <https://doi.org/10.1126/science.252.5009.1162> (1991).
31. Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182. <https://doi.org/10.1186/1471-2105-14-321> (1998).
32. Dai, X., Sinharoy, S., Udvardi, M. & Zhao, P. X. PlantTFCat: An online plant transcription factor and transcriptional regulator categorization and analysis tool. *BMC Bioinform.* **14**, 321. <https://doi.org/10.1186/1471-2105-14-321> (2013).
33. Zheng, Y. et al. iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* **9**, 1667–1670. <https://doi.org/10.1016/j.molp.2016.09.014> (2016).
34. Tian, F., Yang, D. C., Meng, Y. Q., Jin, J. & Gao, G. PlantRegMap: Charting functional regulatory maps in plants. *Nucleic Acids Res.* **48**, D1104–D1113. <https://doi.org/10.1093/nar/gkz1020> (2020).
35. Mulder, N. J. & Apweiler, R. InterPro and InterProScan: Tools for protein sequence classification and comparison. *Methods Mol. Biol.* **396**, 59–70. <https://doi.org/10.1385/1-59745-513-9:59> (2007).
36. Rhee, S. Y. et al. The Arabidopsis information resource (TAIR): A model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* **31**, 224–228. <https://doi.org/10.1093/nar/gkg076> (2003).
37. The UniProt Consortium. UniProt: The universal protein knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617. <https://doi.org/10.1093/nar/gkad066> (2025).
38. Mistry, J. et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419. <https://doi.org/10.1093/nar/gkaa913> (2021).
39. Yu, C.-S., Chen, Y.-C., Lu, C.-H. & Hwang, J.-K. Prediction of protein subcellular localization. *Proteins* **64**, 643–651. <https://doi.org/10.1002/prot.21018> (2006).
40. Sperschneider, J. et al. Localizer: Subcellular localization prediction of both plant and effector proteins in the plant cell. *Sci. Rep.* **7**, 44598. <https://doi.org/10.1038/srep44598> (2017).
41. Gasteiger, E. et al. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* 571–607 (Humana press, Totowa, NJ, 2005). <https://doi.org/10.1385/1-59259-890-0:571>.
42. Wang, Y. et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49. <https://doi.org/10.1093/nar/gkr1293> (2012).
43. Sievers, F. & Higgins, D. G. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* **1079**, 105–116. https://doi.org/10.1007/978-1-62703-646-7_6 (2014).

44. Madeira, F. et al. The EMBL-EBI job dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* **52**, W521–W525. <https://doi.org/10.1093/nar/gkae241> (2024).
45. Nekrutenko, A., Makova, K. D. & Li, W.-H. The KA/KS ratio test for assessing the protein-coding potential of genomic regions: An empirical and simulation study. *Genome Res.* **12**, 198–202. <https://doi.org/10.1101/gr.217302> (2002).
46. Untergasser, A. et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115. <https://doi.org/10.1093/nar/gks596> (2012).
47. Ferreira-Neto, J. R. C. et al. The cowpea kinome: Genomic and transcriptomic analysis under biotic and abiotic stresses. *Front. Plant Sci.* **12**, 667013. <https://doi.org/10.3389/fpls.2021.667013> (2021).
48. The Galaxy Community. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res.* **52**, W83–W94. <https://doi.org/10.1093/nar/gkae410> (2024).
49. de Sena Brandine, G. & Smith, A. D. Falco: High-speed FastQC emulation for quality control of sequencing data. *F1000Res* **8**, 1874. <https://doi.org/10.12688/f1000research.21142.2> (2021).
50. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10. <https://doi.org/10.14806/ej.17.1.200> (2011).
51. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. <https://doi.org/10.1093/bioinformatics/bts635> (2013).
52. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. <https://doi.org/10.1093/bioinformatics/btt656> (2013).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
54. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016). <https://doi.org/10.1007/978-3-319-24277-4>.
55. Kong, W., Wang, Y., Zhang, S., Yu, J. & Zhang, X. Recent advances in assembly of complex plant genomes. *Genom. Proteom. Bioinform.* **21**, 427–439. <https://doi.org/10.1016/j.gpb.2023.04.004> (2023).
56. Yoon, B. J. Hidden Markov models and their applications in biological sequence analysis. *Curr. Genomics* **10**, 402–415. <https://doi.org/10.2174/138920209789177575> (2009).
57. Peng, W. Improving protein function prediction using domain and protein complexes in PPI networks. *BMC Syst. Biol.* **8**, 35. <https://doi.org/10.1186/1752-0509-8-35> (2014).
58. Ameline-Torregrosa, C. et al. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* **146**, 5–21. <https://doi.org/10.1104/pp.107.110041> (2008).
59. Dangl, J. L. & Jones, J. D. Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–833. <https://doi.org/10.1038/35081161> (2001).
60. Meyers, B. C. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**, 809–834. <https://doi.org/10.1105/tpc.009308> (2003).
61. Catanzariti, A. M. et al. The AvrM effector from flax rust has a structured C-terminal domain and interacts directly with the M resistance protein. *Mol. Plant Microbe Interact.* **23**, 49–57. <https://doi.org/10.1094/MPMI-23-1-0049> (2010).
62. Steinbrenner, A. D., Goritschnig, S. & Staskawicz, B. J. Recognition and activation domains contribute to allele-specific responses of an *Arabidopsis* NLR receptor to an oomycete effector protein. *PLoS Pathog.* **11**, e1004665. <https://doi.org/10.1371/journal.ppat.1004665> (2015).
63. van der Hoorn, R. A. & Kamoun, S. From guard to decoy: A new model for perception of plant pathogen effectors. *Plant Cell* **20**, 2009–2017. <https://doi.org/10.1105/tpc.108.060194> (2008).
64. Le Roux, C. et al. A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* **161**, 1074–1088. <https://doi.org/10.1016/j.cell.2015.04.025> (2015).
65. Sarris, P. F. et al. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* **161**, 1089–1100. <https://doi.org/10.1016/j.cell.2015.04.024> (2015).
66. Souframien, J., Raizada, A., Dhanasekar, P. & Suprasanna, P. Draft genome sequence of the pulse crop blackgram (*Vigna mungo* (L.) Hepper) reveals potential R-genes. *Sci. Rep.* **11**, 11247. <https://doi.org/10.1038/s41598-021-90683-9> (2021).
67. Ming, R. et al. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **452**, 991–996. <https://doi.org/10.1038/nature06856> (2008).
68. Mace, E. et al. The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. *BMC Plant Biol.* **14**, 253. <https://doi.org/10.1186/s12870-014-0253-z> (2014).
69. Kang, Y. J. et al. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC Plant Biol.* **12**, 139. <https://doi.org/10.1186/1471-2229-12-139> (2012).
70. Wu, J., Zhu, J., Wang, L. & Wang, S. Genome-wide association study identifies NBS-LRR-encoding genes related with anthracnose and common bacterial blight in the common bean. *Front. Plant Sci.* **8**, 1398. <https://doi.org/10.3389/fpls.2017.01398> (2017).
71. Song, H. et al. Comparative analysis of NBS-LRR genes and their response to *Aspergillus flavus* in *Arachis*. *PLoS ONE* **12**, e0171181. <https://doi.org/10.1371/journal.pone.0171181> (2017).
72. Chisholm, S. T., Coaker, G., Day, B. & Staskawicz, B. J. Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell* **124**, 803–814. <https://doi.org/10.1016/j.cell.2006.02.008> (2006).
73. Liu, W. et al. The stripe rust resistance gene Yr10 encodes an evolutionarily conserved and unique CC-NBS-LRR sequence in wheat. *Mol. Plant* **7**, 1740–1755. <https://doi.org/10.1093/mp/ssu112> (2014).
74. Periyannan, S. et al. The gene Sr33, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. *Science* **341**, 786–788. <https://doi.org/10.1126/science.1239028> (2013).
75. Zhou, F. et al. Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a Rar1-independent signalling pathway. *Plant Cell* **13**, 337–350. <https://doi.org/10.1105/tpc.13.2.337> (2001).
76. Anderson, P. A. et al. Inactivation of the flax rust resistance gene M associated with loss of a repeated unit within the leucine-rich repeat coding region. *Plant Cell* **9**, 641–651. <https://doi.org/10.1105/tpc.9.4.641> (1997).
77. Noel, L. et al. Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus of *Arabidopsis*. *Plant Cell* **11**, 2099–2112. <https://doi.org/10.1105/tpc.11.11.2099> (1999).
78. Whitham, S. et al. The product of the tobacco mosaic virus resistance gene N: Similarity to toll and the interleukin-1 receptor. *Cell* **78**, 1101–1115. [https://doi.org/10.1016/0092-8674\(94\)90283-6](https://doi.org/10.1016/0092-8674(94)90283-6) (1994).
79. Iyer, A. S. & McCouch, S. R. The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. *Mol. Plant Microbe Interact.* **17**, 1348–1354. <https://doi.org/10.1094/MPMI.2004.17.12.1348> (2004).
80. Deslandes, L. et al. Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive *RRS1-R* gene, a member of a novel family of resistance genes. *Proc. Natl. Acad. Sci. USA* **99**, 2404–2409. <https://doi.org/10.1073/pnas.032485099> (2002).
81. Liang, Q. et al. A view of the pan-genome of domesticated Cowpea (*Vigna unguiculata* [L.] Walp.). *Plant Genome* **17**, e20319. <https://doi.org/10.1002/tpg2.20319> (2024).
82. Naithani, S., Komath, S. S., Nonomura, A. & Govindjee, G. Plant lectins and their many roles: Carbohydrate-binding and beyond. *J. Plant Physiol.* **266**, 153531. <https://doi.org/10.1016/j.jplph.2021.153531> (2021).
83. Srivastava, R. & Sahoo, L. Genome-wide analysis of cowpea NAC family elucidating the genetic and molecular relationships that interface stress and growth regulatory signals. *Plant Gene* **31**, 100363. <https://doi.org/10.1016/j.plgene.2022.100363> (2022).

84. Cantarel, B. L. et al. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
85. Stanke, M. & Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).
86. Matos, M. K. D. S. et al. The WRKY transcription factor family in cowpea: Genomic characterization and transcriptomic profiling under root dehydration. *Gene* **823**, 146377. <https://doi.org/10.1016/j.gene.2022.146377> (2022).
87. Wan, Z. et al. CRISPR/Cas9-mediated targeted mutation of the E1 decreases photoperiod sensitivity, alters stem growth habits, and decreases branch number in soybean. *Front. Plant Sci.* **13**, 1066820. <https://doi.org/10.3389/fpls.2022.1066820> (2022).
88. Sadhukhan, A. et al. *VtDREB2A*, a novel DREB2-type transcription factor in the drought-tolerant legume cowpea, mediates DRE-dependent expression of stress-responsive genes and confers enhanced drought resistance in transgenic *Arabidopsis*. *Planta* **240**, 645–664. <https://doi.org/10.1007/s00425-014-2111-5> (2014).
89. Baillo, E. H., Kimotho, R. N., Zhang, Z. & Xu, P. Transcription factors associated with abiotic and biotic stress tolerance and their potential for crops improvement. *Genes* **10**, 771. <https://doi.org/10.3390/genes10100771> (2019).
90. Jin, J. P. et al. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **45**, D1040–D1045 (2017).
91. Yuan, H. Y., Kagale, S. & Ferrie, A. M. R. Multifaceted roles of transcription factors during plant embryogenesis. *Front. Plant Sci.* **14**, 1322728. <https://doi.org/10.3389/fpls.2023.1322728> (2024).
92. Dietrich, R. A., Richberg, M. H., Schmidt, R., Dean, C. & Dangl, J. L. A novel zinc finger protein is encoded by the *Arabidopsis* *LSD1* gene and functions as a negative regulator of plant cell death. *Cell* **88**, 685–694. [https://doi.org/10.1016/s0092-8674\(00\)81911-x](https://doi.org/10.1016/s0092-8674(00)81911-x) (1997).
93. Aono, A. H. et al. Genome-wide characterization of the common bean kinome: Catalog and insights into expression patterns and genetic organization. *Gene* **855**, 147127. <https://doi.org/10.1016/j.gene.2022.147127> (2023).
94. Zulawski, M., Schulze, G., Braginets, R., Hartmann, S. & Schulze, W. X. The *Arabidopsis* kinome: Phylogeny and evolutionary insights into functional diversification. *BMC Genomics* **15**, 1–15 (2014).
95. Costa, F. C. L. & Pereira, W. A. The *Cucumis sativus* kinome: Identification, annotation, and expression patterns in response to powdery mildew infection. *BioRxiv* 2023.03.16.532963. <https://doi.org/10.1101/2023.03.16.532963> (2024).
96. Zhu, K. et al. The grapevine kinome: Annotation, classification and expression patterns in developmental processes and stress responses. *Hortic. Res.* **5**, 19. <https://doi.org/10.1038/s41438-018-0027-0> (2018).
97. Zhu, K. et al. The kinome of pineapple: Catalog and insights into functions in crassulacean acid metabolism plants. *BMC Plant Biol.* **18**, 199. <https://doi.org/10.1186/s12870-018-1389-z> (2018).
98. Liu, J. et al. Soybean kinome: Functional classification and gene expression patterns. *J. Exp. Bot.* **66**, 1919–1934. <https://doi.org/10.1093/jxb/eru537> (2015).
99. Lehti-Shiu, M. D., Zou, C., Hanada, K. & Shiu, S. H. Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiol.* **150**, 12–26 (2009).
100. Shiu, S. H. & Blecker, A. B. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. USA* **98**, 10763–10768. <https://doi.org/10.1073/pnas.181141598> (2001).
101. Champion, A. et al. *Arabidopsis* kinome: After the casting. *Funct. Integr. Genomics* **4**, 163–187. <https://doi.org/10.1007/s10142-003-0096-4> (2004).
102. Wei, K. & Li, Y. Functional genomics of the protein kinase superfamily from wheat. *Mol. Breed.* **39**, 141. <https://doi.org/10.1007/s11032-019-1045-9> (2019).
103. Wang, Y., Ficklin, S. P., Wang, X., Feltus, F. A. & Paterson, A. H. Large-scale gene relocations following an ancient genome triplication associated with the diversification of core eudicots. *PLoS ONE* **11**, e0155637. <https://doi.org/10.1371/journal.pone.0155637> (2016).
104. Minkoff, B. B. et al. A cell-free method for expressing and reconstituting membrane proteins enables functional characterization of the plant receptor-like protein kinase FERONIA. *J. Biol. Chem.* **292**, 5932–5942. <https://doi.org/10.1074/jbc.M116.761981> (2017).
105. Li, Y. et al. Contrasting patterns of nucleotide polymorphism suggest different selective regimes within different parts of the *PgiC1* gene in *Festuca ovina* L. *Heredity* **154**, 11. <https://doi.org/10.1186/s41065-017-0032-6> (2017).
106. Zhong, Y., Liu, S., Zhang, X., Li, Y. & Yang, Q. Evolutionary pattern of the presence and absence genes in *Fragaria*. *Can. J. Plant Sci.* **102**, 427–436. <https://doi.org/10.1139/cjps-2020-0316> (2021).
107. Sun, Y., Qiao, Z., Muchero, W. & Chen, J. G. Lectin receptor-like kinases: the sensor and mediator at the plant cell surface. *Front. Plant Sci.* **11**, 596301. <https://doi.org/10.3389/fpls.2020.596301> (2020).
108. Tör, M., Lotze, M. T. & Holton, N. Receptor-mediated signalling in plants: Molecular patterns and programmes. *J. Exp. Bot.* **60**, 3645–3654. <https://doi.org/10.1093/jxb/erp233> (2009).
109. Gonçalves-Vidigal, M. C. et al. Co-segregation analysis and mapping of the anthracnose *Co-10* and angular leaf spot *Phg-ON* disease-resistance genes in the common bean cultivar Ouro Negro. *Theor. Appl. Genet.* **126**, 2245–2255. <https://doi.org/10.1007/s00122-013-2131-8> (2013).
110. McCoy, A. G. et al. A global-temporal analysis on *Phytophthora sojae* resistance-gene efficacy. *Nat. Commun.* **14**, 6043. <https://doi.org/10.1038/s41467-023-41321-7> (2023).
111. Yu, N. et al. Impact of *Rhg1* copy number, type, and interaction with *Rhg4* on resistance to *Heterodera glycines* in soybean. *Theor. Appl. Genet.* **129**, 2403–2412. <https://doi.org/10.1007/s00122-016-2779-y> (2016).
112. Hayes, A. J. et al. Molecular marker mapping of *RSV4*, a gene conferring resistance to all known strains of soybean mosaic virus. *Crop Sci.* **40**, 1434–1437. <https://doi.org/10.2135/cropsci2000.4051434x> (2000).
113. Bish, M. D. et al. The soybean *Rpp3* gene encodes a TIR-NBS-LRR protein that confers resistance to *Phakopsora pachyrhizi*. *Mol. Plant Microbe Interact.* **37**, 561–570. <https://doi.org/10.1094/MPMI-01-24-0007-R> (2024).
114. Li, Y. Genome analysis identified novel candidate genes for *Ascochyta* blight resistance in chickpea using whole genome re-sequencing data. *Front. Plant Sci.* **8**, 359. <https://doi.org/10.3389/fpls.2017.00359> (2017).
115. Purwar, S. et al. Genome-wide identification and analysis of NBS-LRR-encoding genes in mungbean (*Vigna radiata* L. Wilczek) and their expression in two wild non-progenitors reveal their role in MYMIV resistance. *J. Plant Growth Regul.* **42**, 6667–6680. <https://doi.org/10.1007/s00344-023-10948-7> (2023).
116. Alsamman, A. M. et al. Identification, characterization, and validation of NBS-encoding genes in grass pea. *Front. Genet.* **14**, 1187597. <https://doi.org/10.3389/fgene.2023.1187597> (2023).
117. Chini, A. et al. Drought tolerance established by enhanced expression of the CC-NBS-LRR gene, *ADRI*, requires salicylic acid, *EDS1* and *ABI1*. *Plant J.* **38**, 810–822. <https://doi.org/10.1111/j.1365-3113.2004.02086.x> (2004).
118. Kajla, M., Roy, A., Singh, I. K. & Singh, A. Regulation of the regulators: transcription factors controlling biosynthesis of plant secondary metabolites during biotic stresses and their regulation by miRNAs. *Front. Plant Sci.* **14**, 1126567. <https://doi.org/10.3389/fpls.2023.1126567> (2023).
119. Biswas, D., Gain, H. & Mandal, A. MYB transcription factor: a new weapon for biotic stress tolerance in plants. *Plant Stress* **10**, 100252. <https://doi.org/10.1016/j.stress.2023.100252> (2023).
120. Zheng, C. et al. Transcription factors involved in plant stress and growth and development: NAC. *Agronomy* **15**, 949. <https://doi.org/10.3390/agronomy15040949> (2025).

121. Zhou, Q.-Y. et al. Soybean WRKY-type transcription factor genes, *GmWRKY13*, *GmWRKY21*, and *GmWRKY54*, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. *Plant Biotechnol. J.* **6**, 486–503 (2008).
122. Ramalingam, A. et al. Gene expression and yeast two-hybrid studies of 1R-MYB transcription factor mediating drought stress response in chickpea (*Cicer arietinum* L.). *Front. Plant Sci.* **6**, 1117. <https://doi.org/10.3389/fpls.2015.01117> (2015).
123. Yang, S. et al. A stress-responsive bZIP transcription factor *OsbZIP62* improves drought and oxidative tolerance in rice. *BMC Plant Biol.* **19**, 260. <https://doi.org/10.1186/s12870-019-1872-1> (2019).
124. Muhammad, L. A. et al. Genome-wide identification of AP2/ERF transcription factors in mungbean (*Vigna radiata*) and expression profiling of the *VrDREB* subfamily under drought stress. *Crop Pasture Sci.* **69**, 1009–1019 (2018).
125. Shao, H.-B., Chu, L.-Y., Jaleel, C. A. & Zhao, C.-X. Water deficit stress induced anatomical changes in higher plants. *C. R. Biol.* **331**, 215–225. <https://doi.org/10.1016/j.crvi.2008.01.002> (2008).
126. Nakashima, K., Yamaguchi-Shinozaki, K. & Shinozaki, K. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front. Plant Sci.* **5**, 170. <https://doi.org/10.3389/fpls.2014.00170> (2014).
127. Reddy, A. S. N., Marquez, Y., Kalyna, M. & Barta, A. Complexity of the alternative splicing landscape in plants. *Plant Cell* **25**, 3657–3683. <https://doi.org/10.1105/tpc.113.117523> (2013).
128. Song, X., Li, Y., Cao, X. & Qi, Y. MicroRNAs and their regulatory roles in plant–environment interactions. *Annu. Rev. Plant Biol.* **70**, 489–525. <https://doi.org/10.1146/annurev-arplant-050718-100334> (2019).
129. Yan, N. et al. Genome-wide characterization of the sunflower kinase: Classification, evolutionary analysis and expression patterns under different stresses. *Front. Plant Sci.* **15**, 1450936. <https://doi.org/10.3389/fpls.2024.1450936> (2024).
130. Yan, J. et al. Phylogeny of the plant receptor-like kinase (RLK) gene family and expression analysis of wheat RLK genes in response to biotic and abiotic stresses. *BMC Genomics* **24**, 224. <https://doi.org/10.1186/s12864-023-09303-7> (2023).

Author contributions

D.P. investigated the experiment, analysed the data and results, and wrote the manuscript; S.J. supervised the investigation, assisted with data analysis and review of manuscript. All authors reviewed the manuscript.

Funding

Open access funding provided by Department of Atomic Energy.

Declarations

Competing interests

The author(s) declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-18334-x>.

Correspondence and requests for materials should be addressed to D.P.

Reprints and permissions information is available at www.nature.com/reprints.

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

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/>.

© The Author(s) 2025