Functional & Integrative Genomics (2025) 25:52
https://doi.org/10.1007/s10142-025-01542-z

RESEARCH

Molecular signatures that translate across omics layers and crops under high aluminium and low phosphorus stress facilitate the identification of reliable molecular targets for genotyping in lentil

Kavya Kadiyala[1] · **Noren Singh Konjengbam**[1] · **James M**[1] · **Mayank Rai**[1,2] · **Wricha Tyagi**[1,3] · **Ajay Kumar Mahato**[4]

Received: 26 November 2024 / Revised: 31 December 2024 / Accepted: 20 January 2025

Abstract

Aluminium toxicity and phosphorus deficiency are co-existing characteristics of low pH stress that significantly affect the grain yield of crops. The increasing acidity of Indian soils potentially limits the cultivable area for lentil (*Lens culinaris*), the third most widely consumed pulse. Breeding for tolerance requires an understanding of interdependent biological responses, but the molecular characterization of integrated tolerance remains elusive. Therefore, this study aimed to integrate high aluminium and low phosphorus stress responsive associations across the genomics, transcriptomics, proteomics, and metabolomics of multiple crop species. The overlapping molecular signatures were fine mapped to 23 candidates that serve multiple regulatory roles crucial for cellular homeostasis. Most of these genes have not been adequately discussed in the context of soil acidity tolerance. Thus, a multi-omics guided regulatory framework was developed to provide new insights into tolerance mechanisms. In silico genotyping of 29 lentil genotypes across 588 genes related to transomics loci yielded seven nonsynonymous and three synonymous variants likely associated with their differential response to stress. The results suggest comprehensive genotyping of multi-omics specific targets to identify potential candidates for marker-trait association studies. In conclusion, data-driven exploratory analysis of multi-omics variants highlights potential biomarkers as targets for genetically improving complex biological traits.

Keywords Aluminium toxicity · Phosphorus deficiency · Multi-omics · SNP calling · Biomarkers

Introduction

India stands as the largest consumer of lentil, yet its production significantly falls short. Annual consumption ranges from 2.6 to 2.8 MMT, with only 17.91 LMT produced according to the 2023-24 estimates (UPAg 2024). One potential cause of this substantial gap is the prevalence of acidic soils in the major consumer states, including the northeastern hill region (Majumdar et al. 2022). Soil pH is a critical factor that defines the differential solubility

---

✉ Kavya Kadiyala
kavyak.mbio@gmail.com

Noren Singh Konjengbam
norensingh27@gmail.com

James M
mailmejames.2009@gmail.com

Mayank Rai
mrai.cau@gmail.com

Wricha Tyagi
wtyagi.cau@gmail.com

Ajay Kumar Mahato
akmahato@cdfd.org.in

[1] School of Crop Improvement, College of Post Graduate Studies in Agricultural Sciences (CPGSAS), Central Agricultural University (Imphal), Umiam, Meghalaya 793103, India

[2] Post Graduate College of Agriculture, Dr. Rajendra Prasad Central Agricultural University (RPCAU), Pusa, Samastipur, Bihar 848125, India

[3] Research Program- Accelerated Crop Improvement (ACI), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana 502324, India

[4] Genome informatics, The Centre for DNA Fingerprinting and Diagnostics (CDFD), Uppal, Hyderabad, Telangana 500039, India

of aluminium (Al) and phosphorus (P) (Zhao et al. 2023). Consequently, Al toxicity and P deficiency are the most significant co-occurring problems that hinder crop productivity in low pH soils (Penn and Camberato 2019). As the solubility of aluminium increases, the mobility of phosphorus decreases. This leads to high Al bound root injuries that disrupt the root-soil interface, accompanied by limited plant growth due to low phosphorus availability under low pH conditions (Gessa et al. 2005). While liming and phosphorus fertilization may seem effective, they may not provide a sustainable solution in the long term (Holland et al. 2019; Solangi et al. 2023). Therefore, breeding improved cultivars that are inherently adaptable to unfavourable environments will pave the way for resilient and sustainable agriculture.

Associated morphological traits of tolerance primarily focus on root architectural characteristics (Singh et al. 2012; Kulkarni et al. 2021; Ramtekey et al. 2021; Aski et al. 2022). QTL analysis across different crops has identified major transporters that regulate uptake and translocation. Among the significant loci associated with aluminium tolerance in lentil, *qAlt_ma* (Singh et al. 2021) and *qAlt_fs* (Singh et al. 2018) were mapped for malate secretion and callose accumulation, respectively. *Phosphorus uptake 1 (Pup1)* (Wissuwa et al. 2002) was identified for high phosphorus use efficiency (PUE) in rice. Several transcriptional regulators have been mapped as causal genetic factors of tolerance from transcriptome assays. *Aluminum-activated malate transporter 1 (ALMT1)* (Sasaki et al. 2004), *aluminium sensitive 1 (ALS1)* (Larsen et al. 2005), *multidrug and toxic compound extrusion (MATE)* (Magalhaes et al. 2007), *sensitive to proton rhizotoxicity 1 (STOP1)* (Iuchi et al. 2007), *sensitive to Al rhizotoxicity 1 (STAR1)* (Huang et al. 2009), and *Al resistance transcription factor 1 (ART1)* (Yamaji et al. 2009) are the aluminium inducible genes reported as candidates for improving aluminium tolerance. *Phosphorus-starvation tolerance 1 (PSTOL1)* (Gamuyao et al. 2012), *phosphate transporter 1 (PHT1)* (Shin et al. 2004), *pin-formed 1 (PIN3)* (Lian et al. 2023), *growth regulating factor 10 (GRF10)* (Wang et al. 2024a, b), *phosphate starvation response 1 (PHR1)* (Rubio et al. 2001), *PHO1* (Hamburger et al. 2002), and *purple acid phosphatase (PAP)* (Tran et al. 2010) are low P inducible genes reported as targets for improving phosphorus deficiency tolerance.

Despite notable advancements in understanding molecular mechanisms, practical applications face various challenges. Crop yield, the ultimate goal of improvement under stress, is influenced by numerous constraints that operate both spatially and temporally (Soriano et al. 2021). Plants respond to these variable constraints through interdependent molecular gradients, whose relationships are similarly defined by spatial and temporal factors (Roychowdhury et al. 2023). Current molecular perspectives on Al stress are mostly centered around detoxification and sequestration (Kocjan et al. 2024), while responses to reduced phosphorus mobility are geared towards solubilization and reallocation (Ojeda et al. 2022). The exploration of overlapping genetic factors responsible for adaptive responses to both Al toxicity and P deficiency remains uncharted territory. Additionally, the efficiency of related genetic factors seems to rely heavily on genetic background (Gu et al. 2016). Failing to properly account for genetic heterogeneity can obscure the fundamental concept of stress biology. This oversight can cause our years of experimentation to yield spurious associations and conclusions relevant to limited genetic backgrounds. Therefore, identifying potential mechanisms for enhancing combined tolerance necessitates an understanding of complex genetic causation.

Recently, it has been acknowledged that broadening biological inferences to a multi-modal context enhances our understanding of the mechanisms underlying tolerance, which is instrumental in developing effective resilience (Hasin et al. 2017; Mersha et al. 2021). In this context, the present study aimed to identify promising molecular targets for genotyping in lentil. Understanding the genetic architecture of complex adaptive traits while considering biological heterogeneity demands a multimodal approach. Therefore, the objectives of this study include identifying overlapping molecular signals across multiple omics layers and crops in response to high Al and low P stress, pinpointing lentil orthologs for prioritized transomics loci, and genotyping selected genes, including their related genes across the genome, by utilizing relevant Sequence Read Archive (SRA) accessions of lentil. Consequently, we extended our search beyond the *Lens* genus for molecular information concerning high Al and low P stress. Candidate genes associated with combined tolerance were recognized through a meta-analysis of representative multi-omics data across various crop species. This investigation was further expanded to identify single nucleotide variants (SNVs) with predictable effects on the candidate loci.

Methodology

Multi-omics data retrieval, processing and filtering

A comprehensive literature mining was conducted to examine studies on low pH stress response across various crop species, focusing specifically on high aluminium and low phosphorus stress. All representative data available from the earliest records through 2023 were included in the search. Genomics, transcriptomics, proteomics, and metabolomics comprised the representative omics layers in this compilation. The genomic datasets included associated markers, QTLs (Quantitative trait loci), and causal GWAS (Genome Wide Association Study) variants. Differentially expressed genes (DEGs), differentially available proteins (DAPs), and differentially abundant metabolites (DAMs) represented the datasets for transcriptomics, proteomics, and metabolomics, respectively. Crop models from which data were curated include rice, wheat, maize, barley, *Sorghum*, switchgrass, soybean, lentil, chickpea, peanut, alfalfa, stylo, hydrangea, citrus, parrot's feather, Masson's pine, pepper, tomato, triticale, aspen, Chinese fir, rapeseed, rubber, popcorn, *Arabidopsis*, buckwheat, and common bean. Associated QTLs, markers, and GWAS variants for various traits responsive to high aluminium and low phosphorus stress were curated from studies across multiple crop species. Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) from the National Center for Biotechnology Information (NCBI) was utilized to identify the genomic coordinates of associated genomic variants/loci using the forward and reverse primer sequences of flanking/associated markers. To align these heterogeneous genomic intervals among distant crop species, the genomic data viewer (https://www.ncbi.nlm.nih.gov/gdv) from NCBI was employed to locate the annotated gene models present within or near associated signals. For crops whose genomic data were not updated in NCBI, respective crop specific genomic databases were used. These additional crop-specific databases include EMBL-EBI (https://www.ebi.ac.uk/), Phytozome (https://phytozome-next.jgi.doe.gov/), Gramene (https://www.gramene.org/), rap-db (https://rapdb.dna.affrc.go.jp/), Rice Genome Annotation Project (https://rice.uga.edu/), MaizeGDB (https://www.maizegdb.org/), BRAD (https://brassicadb.cn/), TAIR (https://www.arabidopsis.org/), SoyBase (https://www.soybase.org/), GrainGenes (https://wheat.pw.usda.gov/GG3/), KnowPulse (https://knowpulse.usask.ca/), and PulseDB (https://www.pulsedb.org/). To reduce the computational burden of large datasets, DEGs, DAPs, and DAMs with $\log 2FC \geq 2$ and $p\text{-value} < 0.01$ were designated as thresholds for meta-analysis. Table S1 includes the data that had been curated, processed, and filtered. Figure 1 illustrates the strategic workflow executed for this study.

Meta-analysis

The strategy for this multi-omics analysis was primarily inspired by the concept of ortho-MQTL as explored in a study on the dissection of meta-QTL (Khahani et al. 2021). Additionally, the issue of heterogeneity between the metabolome and the other omics datasets was addressed through pathway enrichment of genes involved in metabolite pathways, as referenced by Maan et al. (2023). To identify consensus gene models of associated genomic loci across multiple crop species, the genome of *Arabidopsis thaliana* served as the common reference. The BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) algorithm from NCBI was utilized to identify orthologous gene models between distinct crop species. Genomic intervals overlapping with *Arabidopsis* gene models were defined as meta-genomic loci. The meta-analysis of transcriptomic responses was conducted

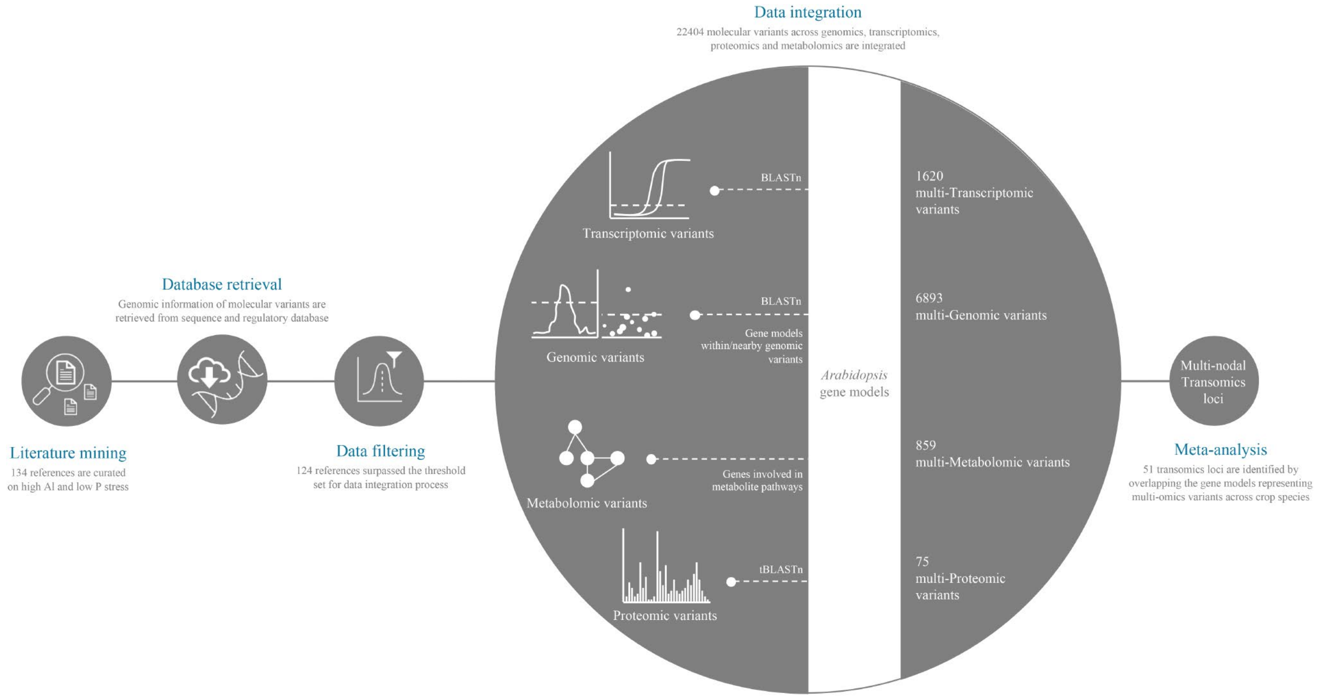


Fig. 1 Graphic summary of enrichment strategy to identify multi-nodal transomics patterns

by curating transcriptomic experiments from the past two decades across various crop species under high aluminum and low phosphorus stress. The transcript sequences of filtered transcripts were retrieved from crop-specific databases and further analyzed for collinearity among distant crop models using the BLASTn algorithm from NCBI. The transcriptome of *Arabidopsis thaliana* was considered a common platform to align these distinct transcripts that are commonly accumulated across diverse crop species in response to low pH stress. A similar workflow was employed to identify proteins that respond to low pH stress across plant taxa, with *Arabidopsis thaliana* as the reference proteomic platform. *The* BLASTp tool was used to align the retrieved protein sequences of DAPs. The meta-analysis of metabolomic datasets was conducted differently from other omics layers. Given that metabolomic identifiers are often too distinct to overlap across studies, integration focused on the genes involved in the biochemical pathways of DAMs. BioCyc (https://biocyc.org/), MetaCyc (https://metacyc.org/), AraCyc (https://www.plantcyc.org/databases/aracyc/15.0), PlantCyc (https://www.plantcyc.org/), and PubChem (https://pubchem.ncbi.nlm.nih.gov/) served as reference databases for retrieving information on biochemical pathways involving DAMs. AraCyc was the primary platform for identifying overlapping *Arabidopsis* genes associated with the biochemical pathways of DAMs. The identifiers of the scrutinized meta-genomic, meta-transcriptomic, meta-proteomic, and meta-metabolomic signals from each omics layer were converted into common gene identifiers using the ID conversion tool of the DAVID Knowledgebase (https://davidbioinformatics.nih.gov/). The processed data at each step of integration was tabulated in Table S2.

The common biological signals from individual omics layers converge through multiple nodes into single multi-omics loci. The various nodes of this multi-modal framework include:

1. Multiple representations across studies: The associated molecular variants must be frequently represented in various studies.
2. Inducibility to high Al and low P: The associated signals must be responsive to both high aluminium and low phosphorus stress.
3. Multi-omics translation: The overlapping molecular signals must be translated across the biological layers with at least three omics layers involved.
4. Cross-species validation: The causal factors should be primarily observed in legumes, followed by other crop species.
5. Pattern congruence with model species: The associated molecular patterns must align with those observed in low pH sensitive or tolerant crop models. Several findings have suggested tolerant model crops (Zhu et al. 2002; Gallon et al. 2004; Shen et al. 2006; Du et al. 2009; Guo et al. 2009; Jessy et al. 2009; Ma et al. 2023) and sensitive model crops (Wang et al. 2006; Langer et al. 2009) for Al toxicity and P deficiency under low soil pH stress. Accordingly, buckwheat, rubber, stylo, and parrot's feather have been considered tolerant model species, While barley and alfalfa are regarded as sensitive models.

Based on these five modalities (Fig. 2), primary determinants were prioritized from aligned loci across multiple omics gradients. These causal factors were considered MTOCL (Multi-nodal Trans-Omics based Candidate Loci).

Determination of gene orthologs and summarization of GO annotations

Gene orthologs of identified transomics loci in *Lens culinaris* were identified using BLASTn (https://knowpulse.usask.ca/blast/nucleotide/nucleotide) of KnowPulse database. The parameters of choice for alignment were set to a maximum threshold for e-value (1 × of 1×10^{-6}) and a word size of 28 to ensure maximum coverage, as recommended by Moreno-Hagelsieb and Latimer (2008). The top hits were regarded as the respective gene orthologs in lentil. Gene ontology (GO) annotations for these orthologs were collected from Phytozome (https://phytozome-next.jgi.doe.gov/) and Uniprot (https://www.uniprot.org/) using the gene identifiers provided in the gene attributes track (https://knowpulse.usask.ca/jbrowse/Lens-culinaris/2) of the KnowPulse database. Concurrently, these orthologs were assigned to gene families based on information archived in Uniprot. The BLAST-based method (Zhang et al. 2020) was utilized to identify all other genes belonging to the family of the transomics-based candidate genes. Alignment was performed using the default parameters of BLASTn in the KnowPulse database. BLAST hits with bit-scores > 80 were considered members of the same gene family (Table S3), which was stringent beyond the ideal threshold of 50 (Pearson 2013).

In silico genotyping of identified biomarkers

The investigation was further extended to identify single nucleotide variants that predictably impact the variable expressivity of multi-omics variants. Precoz, L4076, ILL7537, WBL77, BM4, L830, ILWL118, IPL220, and ILL7663 were selected as susceptible genotypes based on previous studies characterizing their response to Al toxicity and P deficiency. Accordingly, Northfield, Boomer, Cassab, PBA Ace, ILL6002, ILL6788, Indian Head, L4147, DPL62, PDL1, L4602, Digger, ILL358, ILL7979, ILL7978, JL3,

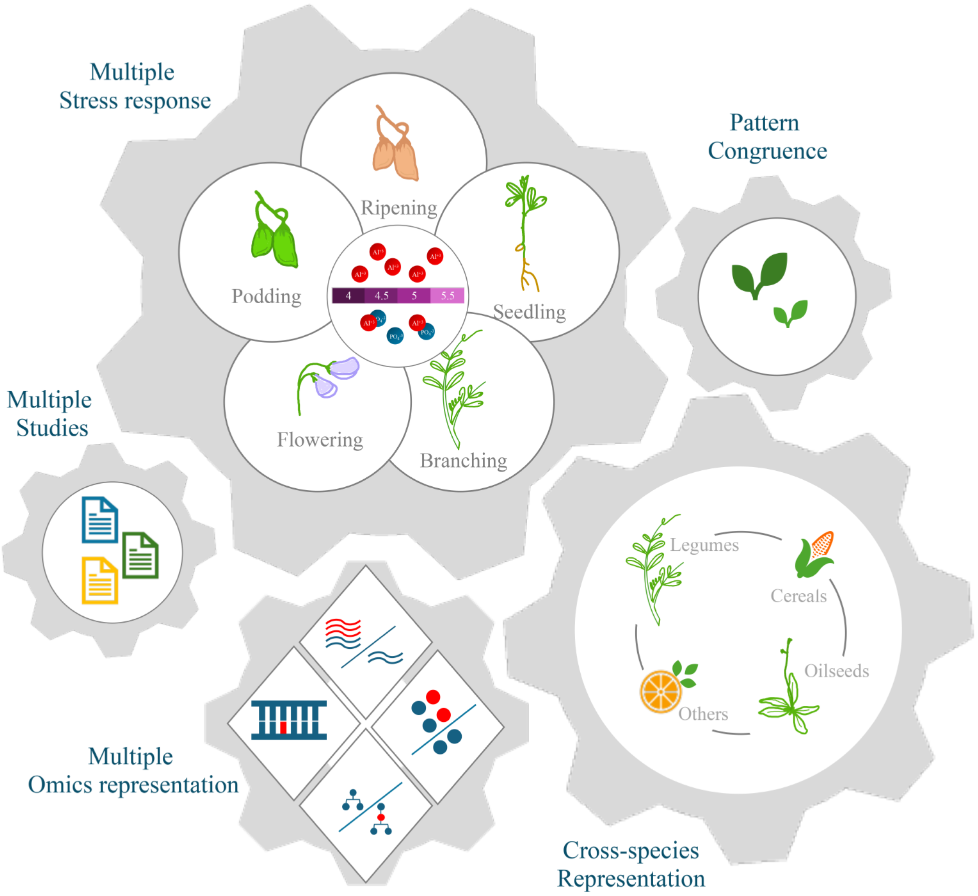


Fig. 2 Five decisive nodes for prioritizing transomics based candidate loci

PDL2, ILL4605, ILL5722, and ILL5883 were identified as tolerant genotypes. The genotypic information for DPL62, ILL358, ILL4605, ILL5722, ILL5883, ILL7663, ILL7978, ILL7979, ILWL118, IPL220, and BM4 was curated using the genomic coordinates of genes from the genotype matrix browser (https://knowpulse.usask.ca/chado/genotype/Lens) of the KnowPulse database. The genotypic information for the remaining 18 genotypes was imputed by assembling their raw sequence datasets archived in the SRA of NCBI.

Paired-end FASTQ files of SRA accessions were retrieved using the SRA toolkit, configured from the terminal (https://github.com/ncbi/sra-tools/wiki/02.-Installing-SRA-Toolkit). The FASTQ format was converted to FASTA for faster alignment using SED commands (https://bioinformaticsworkbook.org/dataWrangling/fastaq-manipulations/converting-fastq-format-to-fasta.html#gsc.tab=0). Raw sequence data was then aligned to the reference genome of lentil, *Lens culinaris*: CDC Redberry Genome Assembly v2.0 (https://knowpulse.usask.ca/bio_data/2690904). Reference-based paired-end sequence alignments were performed by executing BWA commands on a LINUX terminal (https://hcc.unl.edu/docs/applications/app_specific/bioinformatics_tools/alignment_tools/bwa/running_bwa_commands/). Variant calling was carried out using SAMtools (https://github.com/samtools/samtools/issues/298), and the identified SNPs were filtered with the default threshold settings of bcftools (https://samtools.github.io/bcftools/howtos/variant-calling.html). Genotypic imputations at non-variant sites were retained during filtering (https://samtools.github.io/bcftools/bcftools.html#expressions) to compare allelic configurations between the individuals.

A genotypic matrix was compiled for all 29 genotypes using imputed data, including curated data. The allelic frequencies of the reference allele (A) and SNP (B) were calculated independently for tolerant and susceptible genotypes using the following formula:

$$\text{Frequency of A allele} = \frac{\text{Number of genotypes having A allele}}{\text{Total number of genotypes}}$$

$$\text{Frequency of B allele} = \frac{\text{Number of genotypes having B allele}}{\text{Total number of genotypes}}$$

A chi-square test to determine significant associations was not performed due to the limited number of genotypes sampled. Allelic frequencies exceeding the Mendelian expectation of 0.5 (Hardy 1908) were considered as deviating allelic frequencies, as recommended by Feller et al. (2024). SNPs co-segregating with more than 50% of the individuals were therefore regarded as probable associations with Al toxicity and a low P stress response. The impacts of SNPs on gene translation were predicted by translating alternate reading frames using the translate tool in the Sequence Manipulation Suite (https://www.bioinformatics.org/sms2/). The standard genetic code was chosen to translate on the direct template. Table S4 contains curated and called SNP data.

Results

Genomic and pathway collinearity based cross-species coordination model for meta-analysis of multi-omics datasets

The current exploratory data analysis involved a systematic search for evidence on high Al and low P stress across various crop models. A total of 134 references were curated from the earliest records through 2023, of which only 124 were included in the data integration process. The difficulty in retrieving genomic information and the absence of overlapping patterns with other crop models led to the exclusion of 10 studies. Information on filtered variants from independent studies is tabulated in Table S1. In total, 23,779 molecular variants across distinct omics platforms were sampled, of which only 18,759 variants were retained after filtering. The sampled variants were from 24 different crop models, with only four models (rice, soybean, maize, and wheat) represented across all the omics layers. Therefore, to balance these biased datasets, variants with multiple representations within each omics layer were considered for the final mapping of multi-omics loci.

Common molecular patterns identified within each omics scale included 6,893, 1,620, 75, and 903 from genomics, transcriptomics, proteomics, and metabolomics, respectively. The heterogeneity resulting from cross species representation was resolved through the genomic collinearity of each variant with the transcriptome of *Arabidopsis thaliana.* The overlapping top *Arabidopsis* hits for each variant at a single omics level were ultimately mapped onto 51 transomics loci. The heterogeneity caused by molecular gradients across omics layers was addressed through pathway collinearity of each variant with their associated gene models at different biological scales. Consensus maps corresponding to the respective data integration steps are illustrated in Fig. 3. Finally, a total of 23 candidate genes were prioritized over the identified transomics loci based on their validity across five decisive nodes (multiple representations across studies, inducibility to high Al and low P, variability translating through omics gradient, cross-species representation, and pattern congruence with crops considered as tolerant and susceptible models). Figure 4 presents multi-nodal transomics-based candidate loci (MTOCL) that were selected for validation in the lentil model. Table S2 includes the processed data at every step of the meta-analysis.

Lentil orthologs of multi-modular patterns may serve as potential modulators of cellular homeostasis

BLAST (Basic Local Alignment Search Tool) alignment of selected MTOCL with the *Lens culinaris* genome determined the corresponding lentil orthologs, showing a similarity percentage ranging from 43 to 85% (Fig. 5). The observed lower identity may indicate greater sequence divergence between *Arabidopsis thaliana* and *Lens culinaris* during evolution. Referring to information curated in the UniProt database, these 23 lentil genes were assigned to 19 distinct gene families. Typically, the identified transomics loci are key regulators of cellular homeostasis. *O-methyltransferase (OMT)*, *Glycerol-3-phosphate 2-O-acyltransferase (GPAT)*, *Phosphopyruvate hydratase (PPH)*, *Glycerolipase (GL)*, *Cytochrome P450 (CYP450)*, *Heat shock protein 70 (HSP70)*, *Transmembrane and coiled-coil protein (TMCO)*, *Tyrosine aminotransferase (TAT)*, *Gibberellin receptor (GID1)*, *Thiazole biosynthetic enzyme (THI)*, *Flavin-containing monooxygenase (FMO)*, *and Cinnamyl alcohol dehydrogenase-like protein (CAD)* are genes that have not been previously discussed as key regulators of soil acidity tolerance, providing novel insights into high Al and low P tolerance mechanisms. Meanwhile, *Peroxidase (POD)*, *Phenylalanine ammonia lyase (PAL)*, *UDP-glycosyltransferase (UGT)*, *Glutathione S-transferase (GST)*, *and Purple acid phosphatase (PAP)* are the genes frequently mentioned among the identified multi-omics signatures. *The* roles of TMCO have yet to be characterized for their biological relevance in this study.

Figure 6 illustrates the annotated functions and related pathways of these genes, with foundational information accessed from the Gene Ontology (GO) annotations listed in UniProt and pathways curated in the PANTHER database. Charting the underlying signaling networks, lipid signaling appears to be primarily involved, followed by phytohormonal regulatory networks. The immediate biological implications of lipid signaling in abiotic stress may encompass plasma membrane integrity and secondary cell wall maturation. Plasma membrane integrity can be maintained either directly through lipid remodeling mechanisms or indirectly through ROS detoxification. Secondary cell wall maturation may be regulated by altering its composition through cutinization and suberization. Genes associated with phytohormonal activities regulate root and leaf growth by influencing the cell cycle and energy signaling. Moreover, optimal growth in the face of nutrient imbalance is likely supported by inducing membrane transporters responsible for nutrient mobilization. Given the various

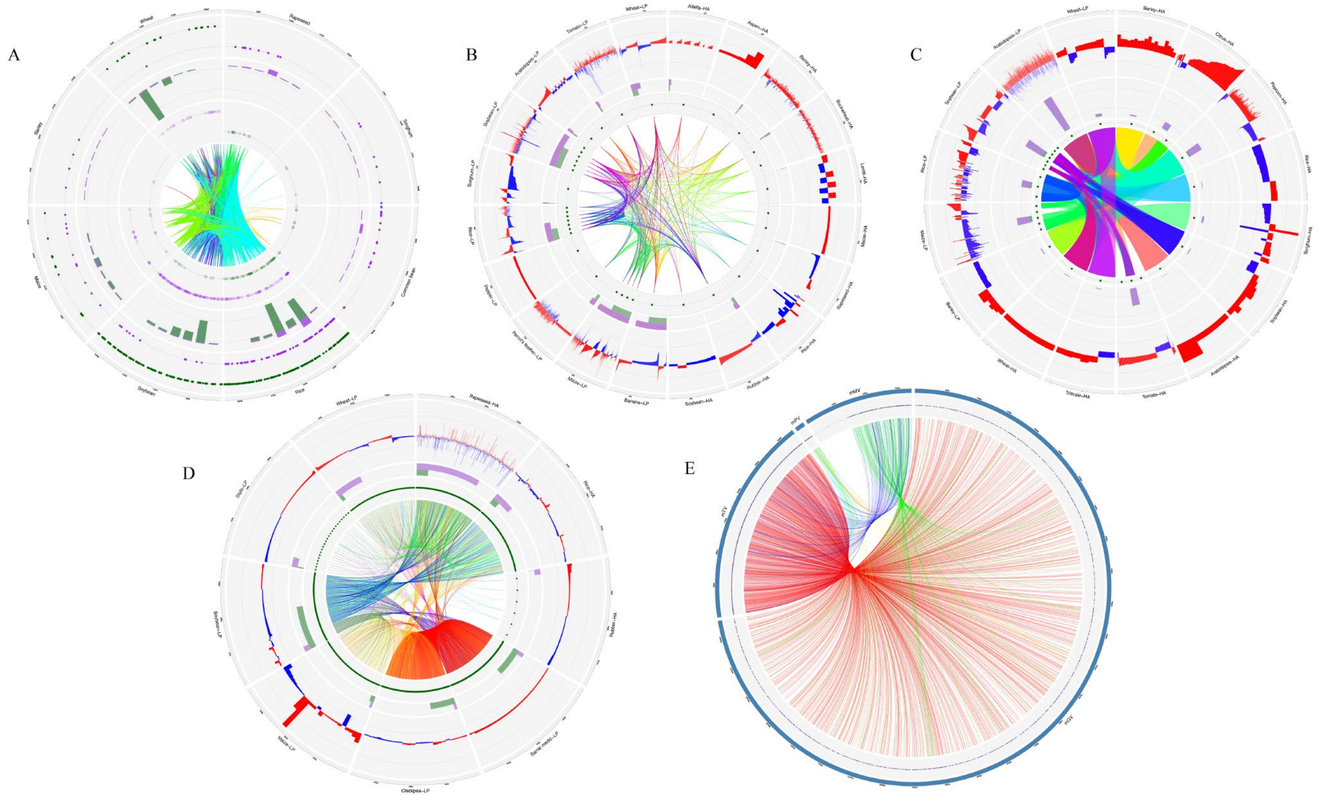


Fig. 3 Circos plots depicting the meta-analysis of multi-omics variants. **A**: Multi-genomics map illustrating high Al and low P associations. The outer track indicates the positions of QTLs, associated markers, and GWAS variants corresponding to the crop genomes. The second track summarizes the number of studies, associated traits, variants, genes, multi-genomic variants, and overlaps for each crop in a clockwise direction. The inner track displays only the overlapping multi-genomic variants across studies. **B**: Multi-transcriptomics map of high Al and low P stress responses. Bidirectional bars in the outer track represent log2FC values of upregulated and downregulated genes in red and blue, respectively. Horizontally clustered columns in the second track indicate the number of multi-transcriptomic variants in each crop. The inner track represents studies with overlapping transcriptomic responses. **C**: Multi-proteomics map of DAPs under Al toxicity and P deficiency. Bidirectional bars in the outer track express log2FC values of upregulated and downregulated proteins in red and blue, respectively. The second track presents bar graphs for the number of multi-proteomic variants identified for each crop. The inner track denotes studies with overlapping proteomic responses. **D**: Multi-metabolomics map of high Al and low P responsive metabolites. Bidirectional bars in the outer track illustrate log2FC values of upregulated and downregulated metabolites in red and blue, respectively. Horizontally clustered columns in the second track signify the number of multi-transcriptomic variants in each crop. The inner track displays each crop's overlapping genes involved in metabolite-dependent pathways. In all maps, links connect recurring patterns between studies. Purple and green colours distinguish high Al studies from low P studies

contributions of these genes to plastic phenotypes, members of the corresponding gene families were targeted to capture the underlying functional diversity. By considering bit scores above 80 as ideal BLAST hits for inferring homology, all related genes within their respective gene families were identified, resulting in a total of 588 genes (Table S3).

Assembling low pH stress responsive SRA datasets calls putative SNPs likely associated with differential responses to stress

Paired-end FASTQ files of 17 SRA accessions with known responses to high Al and low P were retrieved from the National Centre for Biotechnology Information (NCBI) using the SRA toolkit configured in the terminal. Genotypic information was then imputed by aligning with the *Lens culinaris*: CDC Redberry genome v2 using BWA MEM codes, followed by SAMtools and bcftools. In addition to these filtered variants from assembling SRA accessions, genotypic data for the remaining 12 genotypes were obtained from the genotype matrix browser of the KnowPulse database. Details of accession numbers and SNPs identified and collected for each accession are tabulated in Table S4. Genotypic information at non-variant sites was also retained for comparisons between individuals. Precoz, L4076, ILL7537, WBL77, BM4, L830, ILWL118, IPL220, and ILL7663 have frequently been characterized as susceptible in studies on high Al and low P stress. In contrast, Northfield, Boomer, Cassab, PBA Ace, ILL6002, ILL6788, Indian Head, L4147, DPL62, PDL1, L4602, Digger, ILL358, ILL7979, ILL7978,

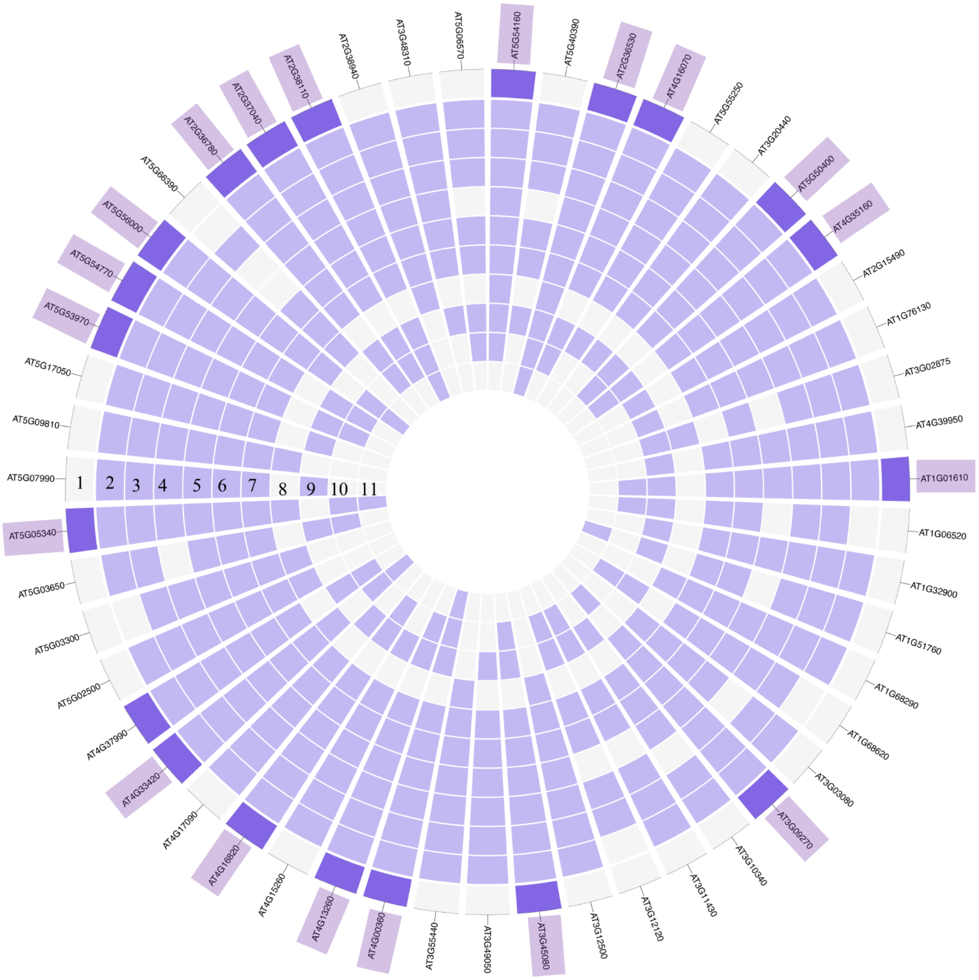


Fig. 4 Circos plot illustrating the multi-modality representation of transomics loci. Track 1, featuring dark purple cells, represents prioritized transomics loci with highlighted gene labels. Light purple cells from tracks 2 to 11 signify the existence of representations under high Al stress, low P stress, genomics, transcriptomics, proteomics, metabolomics, tolerant models, sensitive models, legume species, and other distant species, respectively

JL3, PDL2, ILL4605, ILL5722, and ILL5883 were identified as tolerant. References supporting this categorization are tabulated in Table 1. After comparing the alleles among individuals, 109 comparable sites were found, of which only 22 sites displayed alternate alleles (reference allele/SNP) between the tolerant and susceptible groups. The genotype matrix of those variant sites is shown in Fig. 7.

Based on the differences in allelic frequencies between tolerant and susceptible genotypes, ten SNPs are identified as likely candidates for follow-up marker-trait association studies. The allelic frequencies of these SNPs co-segregate with more than 50% of individuals in both the tolerant and susceptible groups (Fig. 8). However, the statistical significance of their association has not been established due to the limited number of sampled genotypes. These ten SNPs are still valuable, as they are exonic, including seven non-synonymous and three synonymous SNPs, when their respective reading frames are translated using a bioinformatics tool called the Sequence Manipulation Suite. Table 2 displays supporting impact predictions of SNPs on gene products. Given the wide range of possibilities, targeting multi-omics signatures in genotyping populations with variable phenotypic responses to stress can help identify markers associated with trait variation. The identified markers ultimately enhance the utility of marker-assisted selection in improving cultivars for tolerance.

Discussion

Maximizing the potential utility of legumes (Ferguson et al. 2013) in acidic soils faces significant limitations, particularly in lentil, which is highly sensitive (Ryan 2018) to low pH. Improving resilience to low pH stress relies on interdependent regulatory mechanisms (Liang et al. 2013) manifested against high Al and low P problems (Gessa et al. 2005) of low soil pH. Unlike salinity and drought tolerance, tolerance to acidic stress has not been adequately addressed. Furthermore, lentil is the least studied crop among those sensitive to low pH. The proposed method of utilizing a multi-omics design to identify reliable targets is inherently challenging because, unlike human models, there are no established datasets for multi-omics analysis in crops. Furthermore, the datasets are highly heterogeneous, complicating data integration with existing

Fig. 5 Lollipop plot illustrating BLASTn similarity indices of lentil orthologs

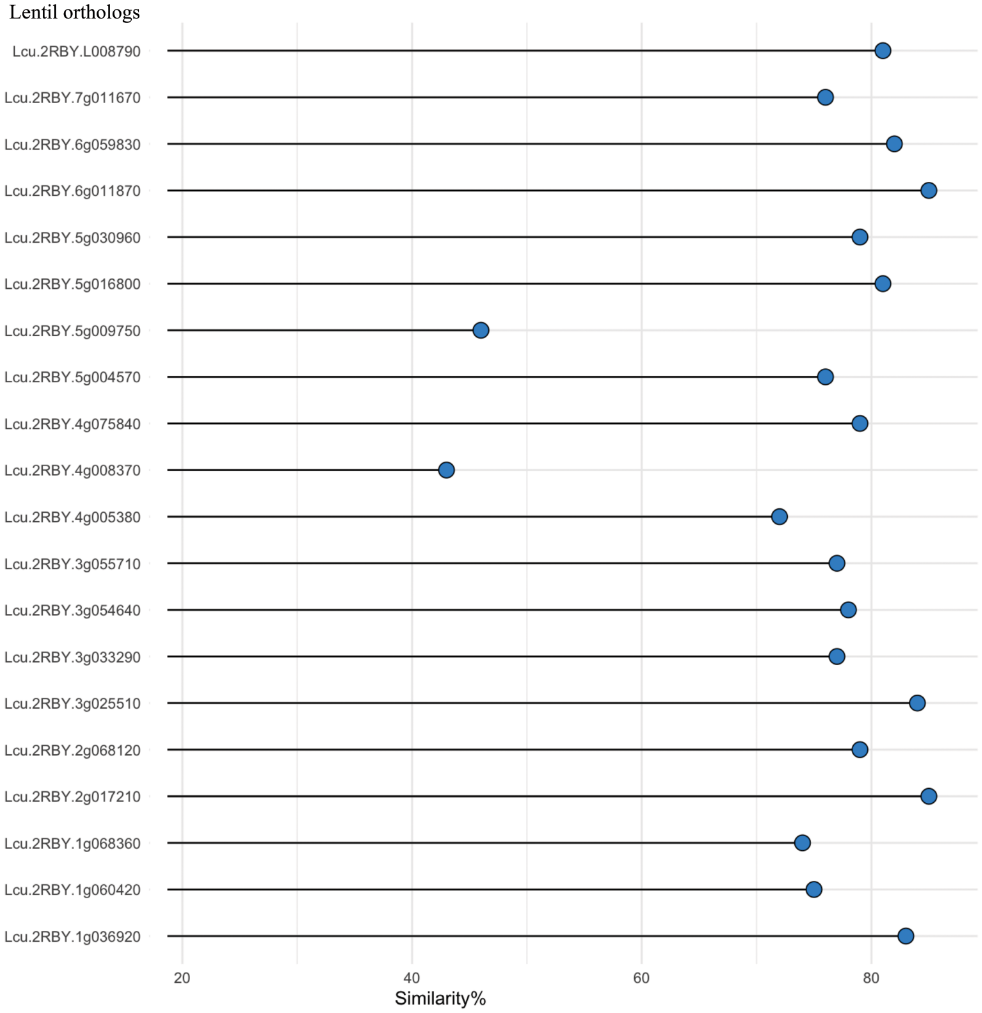
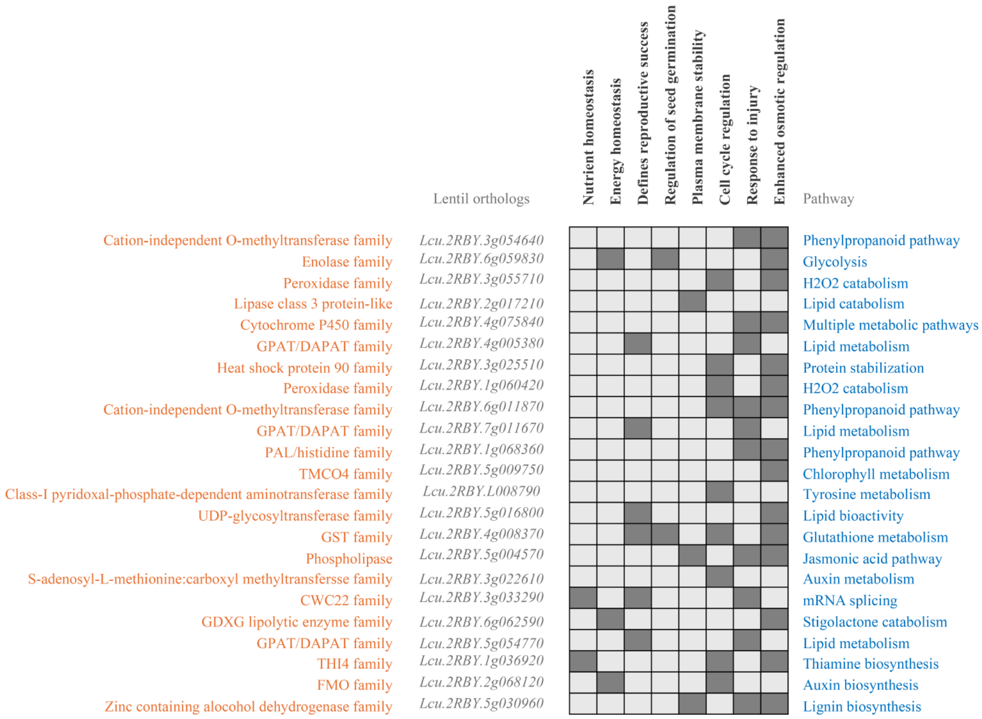


Fig. 6 Gene ontology plot of prioritized transomics loci. Dark gray cells indicate biological functions associated with genes



	Lentil orthologs	Nutrient homeostasis	Energy homeostasis	Defines reproductive success	Regulation of seed germination	Plasma membrane stability	Cell cycle regulation	Response to injury	Enhanced osmotic regulation	Pathway
Cation-independent O-methyltransferase family	*Lcu.2RBY.3g054640*							▨	▨	Phenylpropanoid pathway
Enolase family	*Lcu.2RBY.6g059830*		▨		▨				▨	Glycolysis
Peroxidase family	*Lcu.2RBY.3g055710*						▨		▨	H2O2 catabolism
Lipase class 3 protein-like	*Lcu.2RBY.2g017210*					▨				Lipid catabolism
Cytochrome P450 family	*Lcu.2RBY.4g075840*							▨	▨	Multiple metabolic pathways
GPAT/DAPAT family	*Lcu.2RBY.4g005380*			▨				▨		Lipid metabolism
Heat shock protein 90 family	*Lcu.2RBY.3g025510*						▨		▨	Protein stabilization
Peroxidase family	*Lcu.2RBY.1g060420*						▨		▨	H2O2 catabolism
Cation-independent O-methyltransferase family	*Lcu.2RBY.6g011870*						▨	▨	▨	Phenylpropanoid pathway
GPAT/DAPAT family	*Lcu.2RBY.7g011670*			▨				▨		Lipid metabolism
PAL/histidine family	*Lcu.2RBY.1g068360*							▨	▨	Phenylpropanoid pathway
TMCO4 family	*Lcu.2RBY.5g009750*								▨	Chlorophyll metabolism
Class-I pyridoxal-phosphate-dependent aminotransferase family	*Lcu.2RBY.L008790*						▨			Tyrosine metabolism
UDP-glycosyltransferase family	*Lcu.2RBY.5g016800*			▨					▨	Lipid bioactivity
GST family	*Lcu.2RBY.4g008370*			▨	▨		▨		▨	Glutathione metabolism
Phospholipase	*Lcu.2RBY.5g004570*					▨		▨	▨	Jasmonic acid pathway
S-adenosyl-L-methionine:carboxyl methyltransfersse family	*Lcu.2RBY.3g022610*						▨			Auxin metabolism
CWC22 family	*Lcu.2RBY.3g033290*	▨		▨				▨		mRNA splicing
GDXG lipolytic enzyme family	*Lcu.2RBY.6g062590*		▨						▨	Stigolactone catabolism
GPAT/DAPAT family	*Lcu.2RBY.5g054770*			▨				▨		Lipid metabolism
THI4 family	*Lcu.2RBY.1g036920*	▨					▨		▨	Thiamine biosynthesis
FMO family	*Lcu.2RBY.2g068120*		▨				▨			Auxin biosynthesis
Zinc containing alcohol dehydrogenase family	*Lcu.2RBY.5g030960*					▨		▨	▨	Lignin biosynthesis

Table 1 Phenotypic classification of genotypes with supporting references

Genotypes	Response to stress		References	
	High Al	Low *P*	High Al	Low *P*
Precoz	-	Susceptible	Devi et al. (2021)	Ramtekey et al. (2021) Ganguly et al. (2021)
L4076	Susceptible	Susceptible	Singh et al. (2016) Singh et al. (2012)	Ramtekey et al. (2021) Ganguly et al. (2021)
ILL7537	Susceptible	-	Kulkarni et al. (2021) Singh et al. (2012)	
WBL77	Susceptible	Susceptible	Devi et al. (2021) Singh et al. (2016)	Ramtekey et al. (2021) Ganguly et al. (2021)
BM4	Susceptible	Susceptible	Devi et al. (2021) Singh et al. (2016) Singh et al. (2012)	Ramtekey et al. (2021)
L830	Susceptible	Susceptible	Singh et al. (2012)	Ramtekey et al. (2021)
ILWL118	-	Susceptible		Ramtekey et al. (2021)
IPL220	-	Susceptible		Ramtekey et al. (2021)
ILL7663	Susceptible	-	Singh et al. (2012)	
DPL62	Tolerant	Tolerant	Devi et al. (2021) Singh et al. (2016) Singh et al. (2012)	Ramtekey et al. (2021)
ILL358	Tolerant	-	Singh et al. (2016)	
ILL4605	Tolerant	-	Singh et al. (2016)	
ILL4605	Tolerant	-	Singh et al. (2016)	
ILL5722	Tolerant	-	Singh et al. (2016)	
ILL5883	Tolerant	-	Singh et al. (2016)	
ILL7979	Tolerant	-	Singh et al. (2016)	
PDL2	Tolerant	-	Singh et al. (2016)	
PDL1	Tolerant	-	Devi et al. (2021) Singh et al. (2016)	
PBA Ace	Tolerant	-	Kulkarni et al. (2021)	
Northfield	Tolerant	-	Kulkarni et al. (2021)	
L4602	Tolerant	Tolerant	Devi et al. (2021) Singh et al. (2016) Singh et al. (2012)	Ramtekey et al. (2021)
L4147	Tolerant	Tolerant	Devi et al. (2021) Singh et al. (2016) Singh et al. (2012)	Ramtekey et al. (2021)
JL3	Tolerant	-	Singh et al. (2016) Singh et al. (2012)	
Indian head	Tolerant	-	Kulkarni et al. (2021)	
ILL6788	Tolerant	-	Kulkarni et al. (2021)	
ILL6002	Tolerant	Tolerant	Singh et al. (2016) Singh et al. (2012)	Ganguly et al. (2021)
Digger	Tolerant	-	Kulkarni et al. (2021)	
Cassab	Tolerant	-	Kulkarni et al. (2021)	
Boomer	Tolerant	-	Kulkarni et al. (2021)	

correlation and Bayesian approaches. This study elucidated the molecular networks that regulate combined tolerance through a multi-omics model based on genomic and pathway collinearity for a crop with very few molecular references. The 23 transomics loci identified as a result are key regulators of cellular homeostasis. The majority of these genes are rarely discussed as critical determinants of low pH tolerance, providing novel insights into tolerance, as discussed further.

OMTs are a highly represented molecular group across multiple studies from various omics and crop species. Enzymes in this group are key regulators of flavonoid (Zhao et al. 2024) and lignin (Cai et al. 2023) biosynthesis pathways. They transfer methyl groups from S-adenosyl-L-methionine (*SAM*) to organic compounds found in flavonoids and lignins. The adaptive response may primarily include the lignification of cell walls and vascular tissues. Additionally, there are changes in levels of secondary metabolites like melatonin, which has multiple positive effects (Khan et al. 2024) such as increased chlorophyll and carotenoid levels, along with reduced damage from oxidative stress markers. Therefore,

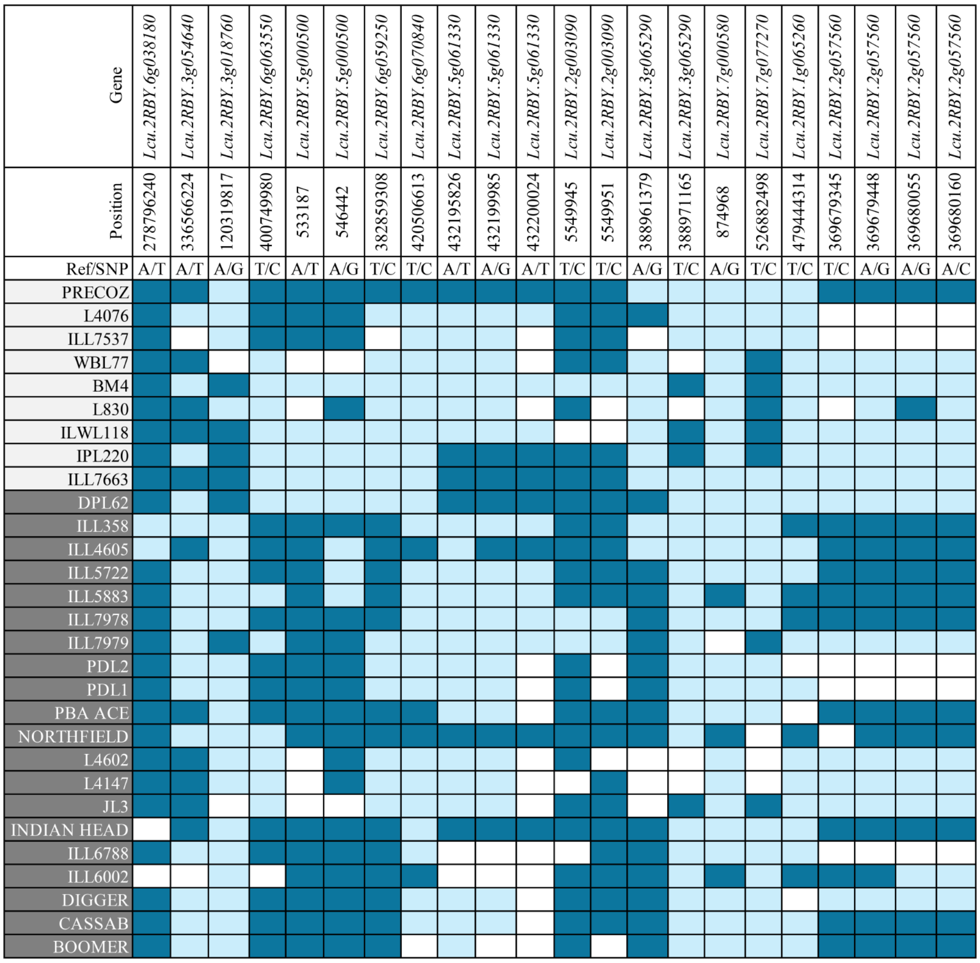


Fig. 7 Genotype matrix showing 22 variant sites between tolerant and susceptible accessions. Light and dark gray genotypes correspond to the susceptible and tolerant groups, respectively. Dark and light blue cells indicate genotypes that are homozygous for the SNP and reference allele. White cells represent heterozygous genotypes

Gene	*Lcu.2RBY.6g038180*	*Lcu.2RBY.3g054640*	*Lcu.2RBY.3g018760*	*Lcu.2RBY.6g063550*	*Lcu.2RBY.5g000500*	*Lcu.2RBY.5g000500*	*Lcu.2RBY.6g059250*	*Lcu.2RBY.6g070840*	*Lcu.2RBY.5g061330*	*Lcu.2RBY.5g061330*	*Lcu.2RBY.5g061330*	*Lcu.2RBY.2g003090*	*Lcu.2RBY.2g003090*	*Lcu.2RBY.3g065290*	*Lcu.2RBY.3g065290*	*Lcu.2RBY.7g000580*	*Lcu.2RBY.7g077270*	*Lcu.2RBY.1g065260*	*Lcu.2RBY.2g057560*	*Lcu.2RBY.2g057560*	*Lcu.2RBY.2g057560*	*Lcu.2RBY.2g057560*
Position	278796240	336566224	120319817	400749980	533187	546442	382859308	420506613	432195826	432199985	432200024	5549945	5549951	388961379	388971165	874968	526882498	479444314	369679345	369679448	369680055	369680160
Ref/SNP	A/T	A/T	A/G	T/C	A/T	A/G	T/C	T/C	A/T	A/G	A/T	T/C	T/C	A/G	T/C	A/G	T/C	T/C	T/C	A/G	A/G	A/C
PRECOZ	■	■	▨	■	■	■	■	■	■	■	■	■	■	▨	▨	▨	▨	▨	■	■	■	■
L4076	■	▨	▨	■	■	■	▨	▨	▨	▨	▨	■	■	■	▨	▨	▨	▨	□	□	□	□
ILL7537	■	□	▨	■	■	■	□	▨	▨	▨	□	■	■	□	▨	▨	▨	▨	□	□	□	□
WBL77	■	■	□	▨	□	□	▨	▨	▨	▨	□	■	■	▨	□	▨	■	▨	▨	▨	▨	▨
BM4	■	▨	■	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	■	▨	■	▨	▨	▨	▨	▨
L830	■	■	▨	▨	□	■	▨	▨	▨	▨	□	■	□	▨	□	▨	■	▨	□	▨	■	▨
ILWL118	■	■	■	▨	▨	▨	▨	▨	▨	▨	▨	□	□	▨	■	▨	■	▨	▨	▨	▨	▨
IPL220	■	▨	■	▨	▨	▨	▨	▨	■	■	■	■	■	▨	■	▨	■	▨	▨	▨	▨	▨
ILL7663	■	■	■	▨	▨	▨	▨	▨	■	■	■	■	■	▨	▨	▨	▨	▨	▨	▨	▨	▨
DPL62	■	▨	■	▨	▨	▨	▨	▨	■	■	■	■	■	■	▨	▨	▨	▨	▨	▨	▨	▨
ILL358	▨	▨	▨	■	■	■	■	▨	▨	▨	▨	■	■	▨	▨	▨	▨	■	■	■	■	■
ILL4605	▨	■	▨	■	■	▨	■	■	▨	■	■	■	■	▨	▨	▨	▨	▨	■	■	■	■
ILL5722	■	▨	▨	■	■	▨	■	▨	▨	▨	▨	■	■	■	▨	▨	▨	▨	■	■	■	■
ILL5883	■	▨	▨	▨	■	▨	■	▨	▨	▨	▨	■	■	■	▨	■	▨	■	■	■	■	■
ILL7978	■	▨	▨	■	■	■	■	▨	▨	▨	▨	▨	▨	■	▨	▨	▨	■	■	■	■	■
ILL7979	■	▨	■	▨	■	■	▨	▨	▨	▨	▨	▨	▨	■	▨	□	■	▨	▨	▨	▨	▨
PDL2	■	▨	▨	■	■	■	▨	▨	▨	▨	□	■	□	■	▨	▨	▨	□	□	□	□	□
PDL1	■	▨	▨	■	■	■	▨	▨	▨	▨	□	■	□	■	▨	▨	▨	▨	□	□	□	□
PBA ACE	■	■	▨	■	■	■	■	■	▨	▨	□	■	■	■	▨	▨	▨	□	■	■	■	■
NORTHFIELD	■	▨	▨	▨	■	■	■	■	■	■	■	■	■	■	▨	■	□	■	□	■	■	■
L4602	■	■	▨	▨	□	■	▨	▨	▨	▨	□	■	□	□	□	▨	□	▨	▨	▨	▨	▨
L4147	■	■	▨	▨	□	■	▨	▨	▨	▨	□	□	■	□	□	▨	□	▨	▨	▨	▨	▨
JL3	■	■	□	▨	□	□	▨	▨	▨	▨	□	■	■	□	■	▨	■	▨	▨	▨	▨	▨
INDIAN HEAD	□	■	▨	■	■	■	■	▨	■	■	■	■	■	■	▨	▨	▨	▨	■	■	■	■
ILL6788	■	▨	▨	■	■	■	■	▨	□	□	□	□	■	■	▨	▨	▨	▨	□	□	□	□
ILL6002	□	□	▨	□	■	■	■	■	□	□	□	■	■	■	▨	■	▨	■	■	■	▨	▨
DIGGER	■	▨	▨	■	■	■	■	▨	▨	▨	□	■	■	■	▨	▨	▨	□	▨	▨	▨	▨
CASSAB	■	▨	▨	■	■	■	■	▨	▨	▨	□	■	■	■	▨	▨	▨	▨	■	■	■	■
BOOMER	■	▨	▨	■	■	■	■	□	▨	□	□	■	□	■	▨	▨	▨	▨	■	■	■	■

with supporting evidence, it is clear that *OMTs help* mitigate damage from physical and physiological injuries due to cell wall and membrane degradation, respectively.

The genes of *GPAT* are also frequently represented but not sufficiently featured in studies of low pH stress. Enzymes transcribed from this gene family are known to catalyze the synthesis of glycerol lipids (Yang et al. 2024). They are functionally characterized by their ability to regulate the plasma membrane fluidity of cells and cell organelles by altering the fatty acid composition of phosphatidylglycerol. This mechanism may be relevant in maintaining plasma membrane integrity against aluminium ion injuries. Additionally, the chemical reactions catalyzed by these enzymes contribute to root suberization (Gully et al. 2024). Suberin serves as an apoplastic barrier, regulating water and nutrient uptake under stressed osmotic environments due to ionic imbalance (Franke et al. 2007). The differential expression of phosphopyruvate hydratase prepares the cells under nutrient-deprived conditions with glycolytic activity to generate energy (Khan et al. 2022). Glycerolipase remodels the lipid composition (Wang et al. 2019) of membranes in a stress-induced manner, preventing cellular damage caused by toxic elements. Photosynthetic and respiratory efficiencies are regulated as a result of their role in maintaining the membrane structure of cell organelles (Wang et al. 2022). Several findings (Zangar et al. 2004; Pandian et al. 2020; Waseem et al. 2021) suggest that *CYPs* are pleiotropic regulators in the biosynthesis of secondary metabolites, antioxidants, and phytohormones. The versatile reactions they catalyze may eventually help recover cells from the osmotic imbalance caused by aluminium toxicity and phosphorus starvation.

In response to an environmental cue, *HSP70* proteins are known to restore the structural homeostasis of proteins, thereby enabling them to fulfill their intended biological roles (Anaraki et al. 2018). Several physiological parameters related to protein content can optimize plant growth and survival under stress. *TAT* significantly contributes to the biosynthesis of tocopherols, which are crucial for mitigating the effects of ROS (Riewe et al. 2012) and lipid radicals (Hollander-Czytko et al. 2005). *GID1* facilitates the GA signaling pathway by degrading its repressor, *SLR1* (Ueguchi-Tanaka et al. 2005). GAs produced in this manner regulate plant growth and development. Stress-induced thiamine thiazole synthase accumulates thiamine differentially (Li et al. 2016). The accumulated thiamine provides oxidative protection (Tunc-Ozdemir et al. 2009). *FMOs* modulate

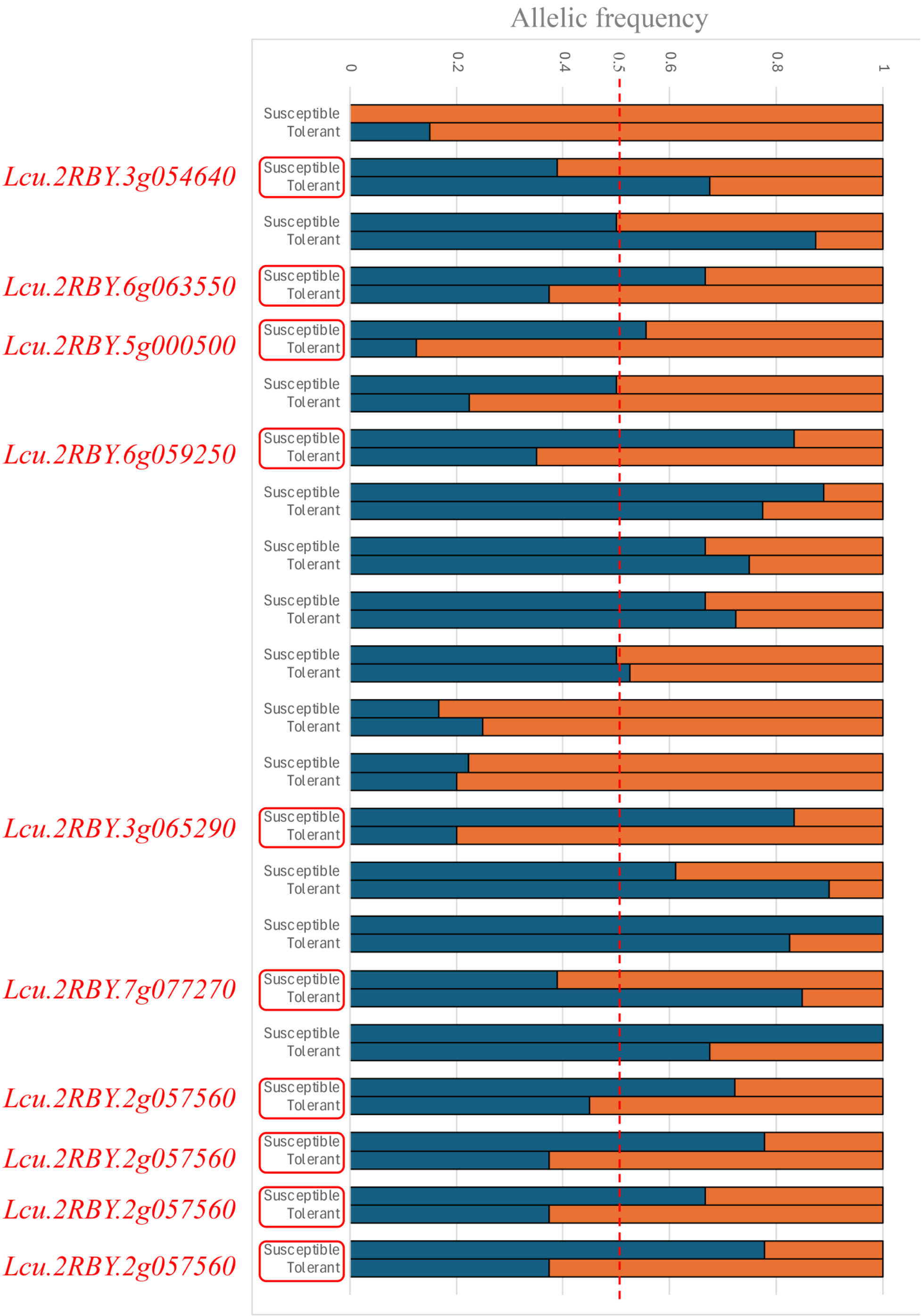


Fig. 8 Allelic frequency plot of the reference allele (**A**) and SNP (**B**) across 29 differentially responsive lentil genotypes. Blue and orange bars indicate the frequency of the A and B alleles, respectively. The red dotted line represents the Mendelian expectation of 0.5. Red boxes highlight SNPs that segregate with tolerant and susceptible individuals, showing deviating allelic frequencies from the Mendelian expectation

oxidative homeostasis by adding molecular oxygen to lipophilic compounds and ROS (Eswaramoorthy et al. 2006). Yeast two-hybrid screening revealed its physical interaction with *CAT2* as a positive regulator of ROS scavenging (Wang et al. 2023). Enzymes of the cinnamyl alcohol dehydrogenase (*CAD*) family regulate cell wall lignification by catalyzing the reduction of cinnamaldehydes to alcohol (Preisner et al. 2018). This secondary cell wall modification strengthens the root cell wall against degradation due to metal ion injuries (Bachir et al. 2022).

Peroxidases are recognized for their role in ROS detoxification. Furthermore, they are involved in the synthesis of lignin and suberin in roots (Quiroga et al. 2000). The genes of *PAL* are functionally significant in responding to phosphorus deficiency, as they catalyze the first step in the lignin and flavonoid biosynthetic pathway (Gho et al. 2020). Regarding Al tolerance, the accumulation of lignins and flavonoids enhances ROS scavenging ability (Yin et al. 2024). The broad spectrum glucosyltransferase activity of UDP glucosyltransferase

Table 2 Predicted SNP impacts

Gene	Ref/SNP	Sequence position	Gene Position	Variant type	Amino acid substitution
Lcu.2RBY.3g054640	A/T	336,566,224	Exon 4	Non-synonymous	E/V
Lcu.2RBY.6g063550	T/C	400,749,980	Exon 1	Non-synonymous	Q/R
Lcu.2RBY.5g000500	A/G	533,187	Exon 1	Non-synonymous	A/V
Lcu.2RBY.6g059250	T/C	382,859,308	Exon 7	Synonymous	D/D
Lcu.2RBY.3g065290	A/G	388,961,379	Exon 1	Synonymous	G/G
Lcu.2RBY.7g077270	T/C	526,882,498	Exon 1	Non-synonymous	I/V
Lcu.2RBY.2g057560	T/C	369,679,345	Exon 2	Synonymous	P/P
	A/G	369,679,448	Exon 2	Non-synonymous	S/P
	A/G	369,680,055	Exon 1	Non-synonymous	M/T
	A/C	369,680,160	Exon 1	Non-synonymous	K/N

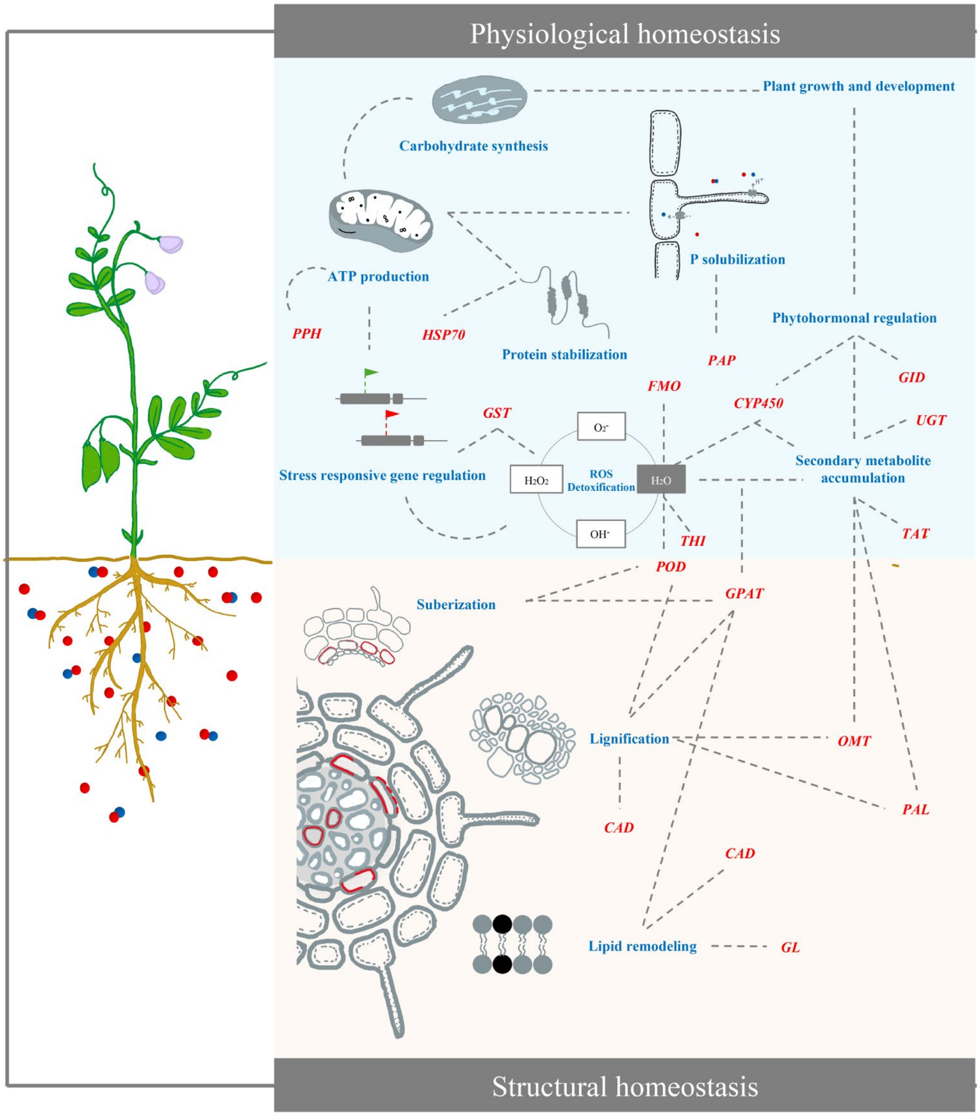


Fig. 9 Mechanistic insights on integrated tolerance to high aluminium and low phosphorus stress. Red and blue circles represent aluminium and phosphate ions, respectively. Causal genes are highlighted in red, while their mechanistic associations are denoted with dotted lines

regulates flavonoid-specific auxin levels, which may influence cell proliferation and expansion (Dong et al. 2020). Despite their general induction by multiple stimuli, *GSTs* play a crucial role in protecting plants against oxidative damage by modifying sensitivity through complex interactions with numerous genes (Sappl et al. 2009). *PAPs* are essential enzymes in solubilizing phosphorus from organophosphates in the rhizosphere (Wang et al. 2012). In summary, the causal mechanisms of tolerance are categorized into structural and physiological homeostasis, as illustrated in Fig. 9.

The identified SNPs across these prioritized loci, which serve as targets for genotyping, including their related genes throughout the genome, are valuable. Supporting this assertion, all ten variants were located in exonic regions, indicating their potential to influence downstream functions. Although seven of them are non-synonymous, they may have an indirect regulatory effect on gene expression, which needs confirmation in follow-up experiments. The primary structure of proteins contributes to their stability and functional activity (Deller et al. 2016). The amino acid substitutions predicted by this study suggest that the identified SNPs may play a role in regulating protein function at the downstream level. The utility of these markers may enhance the likelihood of associations with tolerance-related traits. Additionally, this indicates that genotyping costs can be reduced by concentrating on the most reliable markers rather than a large number of random markers. The results of this study emphasize the flexibility of the experimental approach adopted to address technical and informational limitations. This bioinformatics approach optimally utilizes the available datasets to identify ideal targets and call variants for marker selection in subsequent association studies. However, this study has limitations. The sampled genotypes are insufficient, given the lack of extensive studies on molecular and morphological characterization. Increased experimentation over time can adequately validate these findings in follow-up studies.

Conclusion

A meta-analysis of multi-omics signatures reveals a multidimensional perspective on complex tolerance responses. Furthermore, lentil is the least studied crop under high aluminum and low phosphorus stress. With this motivation, this study represents the first effort to integrate highly heterogeneous datasets, despite the absence of trained bioinformatics models for crops. The transomics loci identified suggest that a major survival strategy may involve balancing physical and physiological disorders caused by injuries and ROS, respectively. Candidate genes with cross-species and multi-omics relevance are mainly involved in the lignification and suberization of the cell wall, lipid remodeling of the plasma membrane, ROS detoxification, and the regulation of secondary metabolites and phytohormones, including phosphorus solubilization. These mechanisms ultimately help to optimize plant growth and development, enhancing photosynthesis and respiration amid osmotic and ionic imbalances due to stress. The candidate genes identified in this study serve as potential biomarkers for follow-up studies. Polymorphic variants across these biomarkers, with altered allele frequencies between contrasting genotypes, may primarily be targeted in marker-trait association studies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10142-025-01542-z.

Acknowledgements The research work was funded by CPGSAS, CAU (Imphal), Umiam. Financial support was provided by CAU (Imphal).

Author contributions Kadiyala Kavya: Conceptualization, Methodology, Data curation, Software, Validation, Illustrations, Writing - Original draft, Writing – Review and Editing. Noren Singh Konjengbam: Supervision. M James: Methodology, Writing – Review and Editing. Mayank Rai: Methodology, Resources. Wricha Tyagi: Writing – Review and Editing, Resources. Ajay Kumar Mahato: Writing – Review and Editing.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Anaraki ZE, Tafreshi SAH, Shariati M (2018) Transient silencing of heat shock proteins showed remarkable roles for HSP70 during adaptation to stress in plants. Environ Exp Bot 155:142–157. https://doi.org/10.1016/j.envexpbot.2018.06.031

Aski M, Mehra R, Mishra GP, Singh D, Yadav P, Rai N, Reddy VRP, Kumar A, Pandey R, Singh MP, Bansal R (2022) Genotypic variation in root architectural traits under contrasting phosphorus levels in Mediterranean and Indian origin lentil genotypes. PeerJ 10:12766. https://doi.org/10.7717/peerj.12766

Bachir IH, Ployet R, Teulieres C, Cassan-Wang H, Mounet F, Grima-Pettenati J (2022) Regulation of secondary cell wall lignification by abiotic and biotic constraints. Adv Bot Res 104:363–392. https://doi.org/10.1016/bs.abr.2022.03.008

Cai T, Sharif Y, Zhuang Y, Yang Q, Chen X, Chen K, Chen Y, Gao M, Dang H, Pan Y, Raza A (2023) In-silico identification and characterization of O-methyltransferase gene family in peanut (*Arachis hypogaea* L.) reveals their putative roles in development and stress tolerance. Front Plant Sci 14:1145624. https://doi.org/10.3389/fpls.2023.1145624

Deller MC, Kong L, Rupp B (2016) Protein stability: a crystallographer's perspective. Acta Crystallogr Sect F: Struct Biology Commun 72:72–95. https://doi.org/10.1107/S2053230X15024619

Dong NQ, Sun Y, Guo T, Shi CL, Zhang YM, Kan Y, Xiang YH, Zhang H, Yang YB, Li YC, Zhao HY (2020) UDP-glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. Nat Commun 11:2629. https://doi.org/10.1038/s41467-020-16403-5

Du YM, Tian J, Liao H, Bai CJ, Yan XL, Liu GD (2009) Aluminium tolerance and high phosphorus efficiency helps *Stylosanthes* better adapt to low-P acid soils. Ann Botany 103:1239–1247. https://doi.org/10.1093/aob/mcp074

Eswaramoorthy S, Bonanno JB, Burley SK, Swaminathan S (2006) Mechanism of action of a flavin-containing monooxygenase. Proceedings of the National Academy of Sciences 103:9832–9837. https://doi.org/10.1073/pnas.0602398103

Feller AF, Peichel CL, Seehausen O (2024) Testing for a role of postzygotic incompatibilities in rapidly speciated Lake Victoria cichlids. Evolution 78:652–664. https://doi.org/10.1093/evolut/qpae007

Ferguson B, Lin MH, Gresshoff PM (2013) Regulation of legume nodulation by acidic growth conditions. Plant Signal Behav 8:23426. https://doi.org/10.4161/psb.23426

Franke R, Schreiber L (2007) Suberin—a biopolyester forming apoplastic plant interfaces. Curr Opin Plant Biol 10:252–259. https://doi.org/10.1016/j.pbi.2007.04.004

Gallon C, Munger C, Premont S, Campbell PG (2004) Hydroponic study of aluminium accumulation by aquatic plants: effects of fluoride and pH. Water Air Soil Pollut 153:135–155. https://doi.org/10.1023/B:WATE.0000019943.67578.ed

Gamuyao R, Chin JH, Pariasca-Tanaka J, Pesaresi P, Catausan S, Dalid C, Slamet-Loedin I, Tecson-Mendoza EM, Wissuwa M, Heuer S (2012) The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. Nature 488:535–539. https://doi.org/10.1038/nature11346

Gessa CE, Mimmo T, Deiana S, Marzadori C (2005) Effect of aluminium and pH on the mobility of phosphate through a soil-root interface model. Plant Soil 272:301–311. https://doi.org/10.1007/s11104-004-5693-z

Gho YS, Kim SJ, Jung KH (2020) Phenylalanine ammonia-lyase family is closely associated with response to phosphate deficiency in rice. Genes Genomics 42:67–76. https://doi.org/10.1007/s13258-019-00879-7

Gu M, Chen A, Sun S, Xu G (2016) Complex regulation of plant phosphate transporters and the gap between molecular mechanisms and practical application: what is missing? Mol Plant 9:396–416. https://doi.org/10.1016/j.molp.2015.12.012

Gully K, Berhin A, De Bellis D, Herrfurth C, Feussner I, Nawrath C (2024) The GPAT4/6/8 clade functions in Arabidopsis root suberization nonredundantly with the GPAT5/7 clade required for suberin lamellae. Proc Natl Acad Sci 121:2314570121. https://doi.org/10.1073/pnas.2314570121

Guo JX, Xu QJ, Jin XC, Hu XW, Li H (2009) Effect of phosphorus concentration on growth of Myriophyllum spicatum and Hydrilla verticillata. Acta Sci Circum 29:118–123

Hamburger D, Rezzonico E, MacDonald-Comber Peteetot J, Somerville C, Poirier Y (2002) Identification and characterization of the *Arabidopsis* PHO1 gene involved in phosphate loading to the xylem. Plant Cell 14:889–902. https://doi.org/10.1105/tpc.000745

Hardy GH (1908) Mendelian proportions in a mixed population. Science 28:49–50. https://doi.org/10.1126/science.28.706.49

Hasin Y, Seldin M, Lusis A (2017) Multi-omics approaches to disease. Genome Biol 18:1–15. https://doi.org/10.1186/s13059-017-1215-1

Holland JE, White PJ, Glendining MJ, Goulding KWT, McGrath SP (2019) Yield responses of arable crops to liming–An evaluation of relationships between yields and soil pH from a long-term liming experiment. Eur J Agron 105:176–188. https://doi.org/10.1016/j.eja.2019.02.016

Hollander-Czytko H, Grabowski J, Sandorf I, Weckermann K, Weiler EW (2005) Tocopherol content and activities of tyrosine aminotransferase and cystine lyase in *Arabidopsis* under stress conditions. J Plant Physiol 162:767–770. https://doi.org/10.1016/j.jplph.2005.04.019

Huang CF, Yamaji N, Mitani N, Yano M, Nagamura Y, Ma JF (2009) A bacterial-type ABC transporter is involved in aluminium tolerance in rice. Plant Cell 21:655–667. https://doi.org/10.1105/tpc.108.064543

Iuchi S, Koyama H, Iuchi A, Kobayashi Y, Kitabayashi S, Kobayashi Y, Ikka T, Hirayama T, Shinozaki K, Kobayashi M (2007) Zinc finger protein STOP1 is critical for proton tolerance in *Arabidopsis* and coregulates a key gene in aluminum tolerance. Proceedings of the National Academy of Sciences 104:9900–9905. https://doi.org/10.1073/pnas.0700117104

Jessy MD, Nair AS, Bai MM, Rajendran P, Punnoose KI (2009) Self–sustainability of phosphorous cycle in rubber (*Hevea brasiliensis*) plantations: annual recycling through litter and removal through latex. J Plantation Crops 37:177–184

Khahani B, Tavakol E, Shariati V, Rossini L (2021) Meta-QTL and ortho-MQTL analyses identified genomic regions controlling rice yield, yield-related traits and root architecture under water deficit conditions. Sci Rep 11:6942. https://doi.org/10.1038/s41598-021-86259-2

Khan MN, Ahmed I, Ud Din I, Noureldeen A, Darwish H, Khan M (2022) Proteomic insight into soybean response to flooding stress reveals changes in energy metabolism and cell wall modifications. PLoS ONE 17:0264453. https://doi.org/10.1371/journal.pone.0264453

Khan Z, Jan R, Asif S, Farooq M, Jang YH, Kim EG, Kim N, Kim KM (2024) Exogenous melatonin induces salt and drought stress tolerance in rice by promoting plant growth and defense system. Sci Rep 14:1214. https://doi.org/10.1038/s41598-024-51369-0

Kocjan A, Kwasniewska J, Szurman-Zubrzycka M (2024) Understanding plant tolerance to aluminum: exploring mechanisms and perspectives. Plant Soil 8:1–25. https://doi.org/10.1007/s11104-024-06745-0

Kulkarni V, Sawbridge T, Kaur S, Hayden M, Slater AT, Norton SL (2021) New sources of lentil germplasm for aluminium toxicity tolerance identified by high throughput hydroponic screening. Physiol Mol Biology Plants 27:563–576. https://doi.org/10.1007/s12298-021-00954-y

Langer H, Cea M, Curaqueo G, Borie F (2009) Influence of aluminum on the growth and organic acid exudation in alfalfa cultivars grown in nutrient solution. J Plant Nutr 32:618–628. https://doi.org/10.1080/01904160802715430

Larsen PB, Geisler MJ, Jones CA, Williams KM, Cancel JD (2005) ALS3 encodes a phloem-localized ABC transporter-like protein that is required for aluminum tolerance in *Arabidopsis*. Plant J 41:353–363. https://doi.org/10.1111/j.1365-313X.2004.02306.x

Li CL, Wang M, Wu XM, Chen DH, Lv HJ, Shen JL, Qiao Z, Zhang W (2016) THI1, a thiamine thiazole synthase, interacts with Ca2+-dependent protein kinase CPK33 and modulates the S-type anion channels and stomatal closure in *Arabidopsis*. Plant Physiol 170:1090–1104. https://doi.org/10.1104/pp.15.01649

Lian W, Yu Q, Zhang P, Cui Y, Yin Z, Cui H, Chen L, Jia H (2023) NtPIN3 positively regulates low phosphorus tolerance by changing root elongation, pi concentration and antioxidant capacity in tobacco. Environ Exp Bot 208:105257. https://doi.org/10.1016/j.envexpbot.2023.105257

Liang C, Pineros MA, Tian J, Yao Z, Sun L, Liu J, Shaff J, Coluccio A, Kochian LV, Liao H (2013) Low pH, aluminum, and phosphorus coordinately regulate malate exudation through GmALMT1 to improve soybean adaptation to acid soils. Plant Physiol 161:1347–1361. https://doi.org/10.1104/pp.112.208934

Ma X, Cheng L, Peng W, Xie G, Liu Z, Yang Z, Wang Y, An F (2023) Comparative transcriptome and metabolome analysis of rubber trees (*Hevea brasiliensis* Muell. Arg.) Response to aluminum stress. Forests 14:568. https://doi.org/10.3390/f14030568

Maan K, Baghel R, Dhariwal S, Sharma A, Bakhshi R, Rana P (2023) Metabolomics and transcriptomics based multi-omics integration reveals radiation-induced altered pathway networking and underlying mechanism. NPJ Syst Biology Appl 9:42. https://doi.org/10.1038/s41540-023-00305-5

Magalhaes JV, Liu J, Guimaraes CT, Lana UG, Alves VM, Wang YH, Schaffert RE, Hoekenga OA, Pineros MA, Shaff JE, Klein PE (2007) A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. Nat Genet 39:1156–1161. https://doi.org/10.1038/ng2074

Majumdar S, Wanniang SK, Behera U (2022) Acid soil management in north eastern hill region of India. Indian Farming 72:3. https://epubs.icar.org.in/index.php/IndFarm/article/view/113461

Mersha TB, Afanador Y, Johansson E, Proper SP, Bernstein JA, Rothenberg ME, Khurana Hershey GK (2021) Resolving clinical phenotypes into endotypes in allergy: molecular and omics approaches. Clin Rev Allergy Immunol 60:200–219. https://doi.org/10.1007/s12016-020-08787-5

Moreno-Hagelsieb G, Latimer K (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics 24:319–324. https://doi.org/10.1093/bioinformatics/btm585

Ojeda-Rivera JO, Alejo-Jacuinde G, Najera-Gonzalez HR, Lopez-Arredondo D (2022) Prospects of genetics and breeding for low-phosphate tolerance: an integrated approach from soil to cell. Theor Appl Genet 135:4125–4150. https://doi.org/10.1007/s00122-022-04095-y

Pandian BA, Sathishraj R, Djanaguiraman M, Prasad PV, Jugulam M (2020) Role of cytochrome P450 enzymes in plant stress response. Antioxidants 9:454. https://doi.org/10.3390/antiox9050454

Pearson WR (2013) An introduction to sequence similarity (homology) searching. Current protocols in bioinformatics 42:3 – 1. https://doi.org/10.1002/0471250953.bi0301s42

Penn CJ, Camberato JJ (2019) A critical review on soil chemical processes that control how soil pH affects phosphorus availability to plants. Agriculture 9:120. https://doi.org/10.3390/agriculture9060120

Preisner M, Wojtasik W, Kostyn K, Boba A, Czuj T, Szopa J, Kulma A (2018) The cinnamyl alcohol dehydrogenase family in flax: differentiation during plant growth and under stress conditions. J Plant Physiol 221:132–143. https://doi.org/10.1016/j.jplph.2017.11.015

Quiroga M, Guerrero C, Botella MA, Barcelo A, Amaya I, Medina MI, Alonso FJ, de Forchetti SM, Tigier, Valpuesta V (2000) A tomato peroxidase involved in the synthesis of lignin and suberin. Plant Physiol 122:1119–1128. https://doi.org/10.1104/pp.122.4.1119

Ramtekey V, Bansal R, Aski MS, Kothar D, Singh A, Pandey R, Tripathi K, Mishra GP, Kumar S, Dikshit HK (2021) Genetic variation for traits related to phosphorus use efficiency in lens species at the seedling stage. Plants 10:2711. https://doi.org/10.3390/plants10122711

Riewe D, Koohi M, Lisec J, Pfeiffer M, Lippmann R, Schmeichel J, Willmitzer L, Altmann T (2012) A tyrosine aminotransferase involved in tocopherol synthesis in Arabidopsis. Plant J 71:850–859. https://doi.org/10.1111/j.1365-313X.2012.05035.x

Roychowdhury R, Das SP, Gupta A, Parihar P, Chandrasekhar K, Sarker U, Kumar A, Ramrao DP, Sudhakar C (2023) Multi-omics pipeline and omics-integration approach to decipher plant's abiotic stress tolerance responses. Genes 14:1281. https://doi.org/10.3390/genes14061281

Rubio V, Linhares F, Solano R, Martin AC, Iglesias J, Leyva A, Paz-Ares J (2001) A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. Genes Dev 15:2122–2133. http://www.genesdev.org/cgi/doi/ https://doi.org/10.1101/gad.204401

Ryan PR (2018) Assessing the role of genetics for improving the yield of Australia's major grain crops on acid soils. Crop Pasture Sci 69:242–264. https://doi.org/10.1071/CP17310

Sappl PG, Carroll AJ, Clifton R, Lister R, Whelan J, Harvey Millar A, Singh KB (2009) The Arabidopsis glutathione transferase gene family displays complex stress regulation and co-silencing multiple genes results in altered metabolic sensitivity to oxidative stress. Plant J 58:53–68. https://doi.org/10.1111/j.1365-313X.2008.03761.x

Sasaki T, Yamamoto Y, Ezaki B, Katsuhara M, Ahn SJ, Ryan PR, Delhaize E, Matsumoto H (2004) A wheat gene encoding an aluminum-activated malate transporter. Plant J 37:645–653. https://doi.org/10.1111/j.1365-313X.2003.01991.x

Shen RF, Che RF, Ma JF (2006) Buckwheat accumulates aluminum in leaves but not in seeds. Plant Soil 284:265–271. https://doi.org/10.1007/s11104-006-0044-x

Shin H, Shin HS, Dewbre GR, Harrison MJ (2004) Phosphate transport in Arabidopsis: Pht1; 1 and Pht1; 4 play a major role in phosphate acquisition from both low-and high-phosphate environments. Plant J 39:629–642. https://doi.org/10.1111/j.1365-313X.2004.02161.x

Singh D, Dikshit HK, Singh R (2012) Variation of aluminium tolerance in lentil (*Lens culinaris* Medik). Plant Breeding 131:751–761. https://doi.org/10.1111/j.1439-0523.2012.01999.x

Singh CK, Singh D, Tomar RSS, Karwa S, Upadhyaya KC, Pal M (2018) Molecular mapping of aluminium resistance loci based on root re-growth and Al-induced fluorescent signals (callose accumulation) in lentil (*Lens culinaris* Medikus). Mol Biol Rep 45:2103–2113. https://doi.org/10.1007/s11033-018-4368-4

Singh CK, Singh D, Sharma S, Chandra S, Tomar RSS, Kumar A, Upadhyaya KC, Pal M (2021) Mechanistic association of quantitative trait locus with malate secretion in lentil (*Lens culinaris* Medikus) seedlings under aluminium stress. Plants 10:1541. https://doi.org/10.3390/plants10081541

Solangi F, Zhu X, Khan S, Rais N, Majeed A, Sabir MA, Iqbal R, Ali S, Hafeez A, Ali B, Ercisli S (2023) The global dilemma of soil legacy phosphorus and its improvement strategies under recent changes in agro-ecosystem sustainability. ACS Omega 8:23271–23282. https://doi.org/10.1021/acsomega.3c00823

Soriano JM, Colasuonno P, Marcotuli I, Gadaleta A (2021) Meta-QTL analysis and identification of candidate genes for quality, abiotic and biotic stress in durum wheat. Sci Rep 11:11877. https://doi.org/10.1038/s41598-021-91446-2

Tran HT, Hurley BA, Plaxton WC (2010) Feeding hungry plants: the role of purple acid phosphatases in phosphate nutrition. Plant Sci 179:14–27. https://doi.org/10.1016/j.plantsci.2010.04.005

Tunc-Ozdemir M, Miller G, Song L, Kim J, Sodek A, Koussevitzky S, Misra AN, Mittler R, Shintani D (2009) Thiamin confers enhanced tolerance to oxidative stress in Arabidopsis. Plant Physiol 151:421–432. https://doi.org/10.1104/pp.109.140046

Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, Chow TY, Hsing YIC, Kitano H, Yamaguchi I, Matsuoka M (2005) GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature 437:693–698. https://doi.org/10.1038/nature04028

UPAg (2024) Production, third advance estimates, 2023–2024. Unified Portal for Agricultural Statistics. Department of Agriculture and Farmers Welfare. Ministry of Agriculture, New Delhi. https://upag.gov.in

Wang L, Liu D (2012) Arabidopsis purple acid phosphatase 10 is a component of plant adaptive mechanism to phosphate limitation. Plant Signal Behav 7:306–310. https://doi.org/10.4161/psb.19019

Wang JP, Raman H, Zhang GP, Mendham N, Zhou MX (2006) Aluminium tolerance in barley (*Hordeum vulgare* L.): physiological mechanisms, genetics and screening methods. J Zhejiang Univ Sci B 7:769–787. https://doi.org/10.1631/jzus.2006.B0769

Wang K, Durrett TP, Benning C (2019) Functional diversity of glycerolipid acylhydrolases in plant metabolism and physiology. Prog Lipid Res 75:100987. https://doi.org/10.1016/j.plipres.2019.100987

Wang L, Qian B, Zhao L, Liang MH, Zhan X, Zhu J (2022) Two triacylglycerol lipases are negative regulators of chilling stress tolerance in Arabidopsis. Int J Mol Sci 23:3380. https://doi.org/10.3390/ijms23063380

Wang L, Zhou Y, Ding Y, Chen C, Chen X, Su N, Zhang X, Pan Y, Li J (2023) Novel flavin-containing monooxygenase protein FMO1 interacts with CAT2 to negatively regulate drought tolerance through ROS homeostasis and ABA signaling pathway in tomato. Hortic Res 10:037. https://doi.org/10.1093/hr/uhad037

Wang H, Chen Z, Luo R, Lei C, Zhang M, Gao A, Pu J, Zhang H (2024a) Caffeic acid O-Methyltransferase Gene Family in Mango (*Mangifera indica* L.) with Transcriptional Analysis under biotic and abiotic stresses and the role of MiCOMT1 in Salt Tolerance. Int J Mol Sci 25:2639. https://doi.org/10.3390/ijms25052639

Wang Y, Peng X, Lian X, Yu Q, Zhang L, Li T, Luo H, Yu K, Zhang W, Zhong D, Zhao Y (2024b) Natural variation in ZmGRF10 regulates tolerance to phosphate deficiency in maize by modulating phosphorus remobilization. Crop J. https://doi.org/10.1016/j.cj.2024.08.004

Waseem M, Huang F, Wang Q, Aslam MM, Abbas F, Ahmad F, Ashraf U, Hassan W, Fiaz S, Ye X, Yu L (2021) Identification, methylation profiling, and expression analysis of stress-responsive cytochrome P450 genes in rice under abiotic and phytohormones stresses. GM Crops Food 12:551–563. https://doi.org/10.1080/21645698.2021.1908813

Wissuwa M, Wegner J, Ae N, Yano M (2002) Substitution mapping of Pup1: a major QTL increasing phosphorus uptake of rice from a phosphorus-deficient soil. Theor Appl Genet 105:890–897. https://doi.org/10.1007/s00122-002-1051-9

Yamaji N, Huang CF, Nagao S, Yano M, Sato Y, Nagamura Y, Ma JF (2009) A zinc finger transcription factor ART1 regulates multiple genes implicated in aluminum tolerance in rice. Plant Cell 21:3339–3349. https://doi.org/10.1105/tpc.109.070771

Yang C, Ma J, Qi C, Ma Y, Xiong H, Duan R (2024) Genome-wide identification, characterization, evolutionary analysis, and expression pattern of the GPAT Gene Family in Barley and functional analysis of HvGPAT18 under Abiotic Stress. Int J Mol Sci 25:6101. https://doi.org/10.3390/ijms25116101

Yin T, Xu R, Zhu L, Yang X, Zhang M, Li X, Zi Y, Wen K, Zhao K, Cai H, Liu X (2024) Comparative analysis of the PAL gene family in nine citruses provides new insights into the stress resistance mechanism of Citrus species. BMC Genomics 25:1020. https://doi.org/10.1186/s12864-024-10938-3

Zangar RC, Davydov DR, Verma S (2004) Mechanisms that regulate production of reactive oxygen species by cytochrome P450. Toxicol Appl Pharmcol 199:316–331. https://doi.org/10.1016/j.taap.2004.01.018

Zhang Y, Yin D, Song H (2020) Genome-wide identification and characterization of gene families in Arachis: methods and strategies. Front Genet 11:525. https://doi.org/10.3389/fgene.2020.00525

Zhao W, Gu C, Zhu M, Yan Y, Liu Z, Feng X, Wang X (2023) Chemical speciation of phosphorus in farmland soils and soil aggregates around mining areas. Geoderma 433:116465. https://doi.org/10.1016/j.geoderma.2023.116465

Zhao W, Huang L, Xu S, Wu J, Wang F, Li P, Li L, Tian M, Feng X, Chen Y (2024) Identification of one O-methyltransferase gene involved in methylated flavonoid biosynthesis related to the UV-B irradiation response in *Euphorbia lathyris*. Int J Mol Sci 25:782. https://doi.org/10.3390/ijms25020782

Zhu YG, He YQ, Smith SE, Smith FA (2002) Buckwheat (*Fagopyrum esculentum* Moench) has high capacity to take up phosphorus (P) from a calcium (Ca)-bound source. Plant Soil 239:1–8. https://doi.org/10.1023/A:1014958029905