Journal of Innovative Agriculture
Research Article

Genome-wide association study for composite performance index in rice (*Oryza sativa* L.)

Raveendran Muthurajan*, Williams Mohanavel, Ameena Premnath, Bharathi Ayyenar, Veera Ranjani Rajagopalan, Sudha Manickam

Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore - 641003, India.

***Correspondence**
Raveendran Muthurajan
raveendrantnau@gmail.com

Volume: 12, Issue: 1, Pages: 18-27
DOI: https://doi.org/10.37446/jinagri/rsa/12.1.2025.18-27
Received: 20 November 2024 / Accepted: 14 February 2025 / Published: 31 March 2025

Background: A critical step to maximize the usefulness of genome-wide association studies (GWAS) in plant breeding is the identification and validation of candidate genes underlying genetic associations. Once strong candidates are identified, further validation helps confirm whether the gene truly influences the trait. This process strengthens confidence in marker-trait relationships and enables the development of more precise molecular markers or genomic prediction models for crop improvement.

Methods: Genome-wide association study (GWAS) was conducted on a panel of 100 genetically diverse rice genotypes to dissect the genetic architecture of the Composite Performance Index (CPI), a multivariate principal component score integrating all major agronomic traits.

Results: Full annotation of all nine significant SNPs revealed perfect convergence on four biologically coherent functional modules. Four reproducible quantitative loci (CPI-1, CPI-4, CPI-8 and CPI-11) were identified, explaining the major gradients in field performance. These loci encompass biologically coherent modules linked to energy metabolism, growth regulation, cell-wall integrity and dehydration response. Favourable alleles were located on chromosomes 8 and 11 contributing +1.12 and +1.38 CPI units, respectively. Unfavourable alleles on chromosomes 1 and 4 were associated with stress response and growth-defense trade-offs. The combined fixation of favourable haplotypes from CPI-8 and CPI-11 while purging unfavourable alleles from CPI-1 and CPI-4 predicted a +2.5 CPI gain representing the top 1% ideotype.

Conclusion: The two tightly linked SNPs on chromosome 8 (separated by only 1,069 bp) tag the identical haplotype and are therefore merged into a single locus (CPI-8). No additional loci reached genome-wide significance, confirming that CPI is controlled by a compact, high-impact genetic architecture amenable to rapid marker-assisted pyramiding. The compact genetic basis uncovered in this study provides a practical foundation for marker-assisted selection and genomic designing of high-performing rice varieties.

Keywords: *rice, GWAS, COMPOSITE Performance Index (CPI), SNPs, genomic designing*

Introduction

Rice (*Oryza sativa* L.) is the staple food for more than half of the world's population and a model crop for understanding complex quantitative traits in plants. Ensuring sustainable rice productivity under the dual pressures of increasing global demand and climatic variability remains a central challenge for crop improvement programs (Greenwood et al., 2024). Traditional breeding approaches, though successful in developing high-yielding and stress-tolerant varieties, are often constrained by the polygenic nature of most agronomic traits and the strong influence of genotype × environment interactions. To accelerate genetic gain, it is crucial to identify the genomic regions and candidate genes controlling multi-trait performance under field conditions (Jadamba et al., 2024). Genome-wide association studies (GWAS) have emerged as a powerful strategy for dissecting the genetic basis of complex quantitative traits by leveraging the natural allelic diversity present in diverse germplasm panels (Ashfaq et al., 2023). Unlike traditional linkage mapping which relies on bi-parental populations and offers limited resolution, GWAS exploits historical recombination events accumulated over generations, thereby achieving finer mapping resolution and a broader allelic spectrum. GWAS has been successfully applied in rice to identify loci for yield components, flowering time, stress tolerance and grain quality (Yano et al., 2019). Despite these advances, most GWAS studies focus on single trait, which can fail to capture the interdependence among multiple physiological and developmental parameters that together determine field performance.

Field performance in rice is a product of complex interactions among yield components, resource-use efficiency, stress adaptation and phenological synchrony (Kadam et al., 2019). Thus, dissecting these interconnected traits individually can obscure underlying genetic correlations. A holistic approach that integrates multiple performance-related traits into a single composite index provides a more realistic representation of overall plant productivity (Li et al., 2018). In this context, principal component analysis (PCA) offers a robust framework for deriving a Composite Performance Index (CPI) which is a multivariate score summarizing the major sources of phenotypic variation across genotypes. Such integrative indices have been demonstrated to capture the underlying physiological coherence of high-performing ideotypes more effectively than single-trait metrics. This approach provides unique insights into the genetic coordination between growth, development, metabolism and stress response, which collectively define crop fitness in variable field environments (Padmashree et al., 2023). Moreover, identifying a small number of major loci with pleiotropic effects can simplify marker-assisted selection (MAS) strategies and accelerate the pyramiding of favourable alleles into elite breeding lines (Varshney et al., 2014). The present study aimed to dissect the genetic architecture of the Composite Performance Index (CPI) using genome-wide association analysis in a diverse panel of 100 rice genotypes. By integrating agronomic and physiological data into a single multivariate index, reproducible loci associated with overall field performance were identified. The objectives of the present study were to identify significant SNPs and corresponding candidate genes associated with CPI; to annotate the functional modules represented by these loci and to interpret their biological relevance in the context of field adaptability and breeding implications.

Materials and Methods

Plant material and phenotyping

A panel of 100 diverse rice genotypes, representing both *indica* and *japonica* backgrounds, was evaluated at Wetlands Farm, Tamil Nadu Agricultural University, Coimbatore. Sowing was taken up under raised nursery bed and twenty-one-day seedlings were transplanted to the main field with the spacing of 20 x 20 cm with $2m^2$ plot size by following unreplicated single-plot field evaluation (observational design). Standard agronomic management was applied to ensure uniform crop establishment. Each genotype was assessed for major agronomic and physiological traits *viz.,* days to first flowering, photosynthetic rate, stomatal conductance, transpiration rate, number of veins, plant height, number of tillers, number of productive tillers, leaf area, panicle length, spikelet number per panicle, spikelet fertility, L/B ratio of grains and single plant yield were recorded and used for analysis.

Composite Performance Index (CPI)

Trait data were normalized and subjected to principal component analysis (PCA). The first principal component (PC1) capturing the largest variance proportion (over 40%) was designated as the Composite Performance Index (CPI). CPI effectively integrates both yield potential and physiological efficiency under field conditions.

Genotyping and GWAS

Genome-wide SNP genotyping was performed using high-density rice SNP arrays. After quality filtering ($MAF \geq 0.05$, missing data $\leq 10\%$), GWAS was conducted using a mixed linear model incorporating population structure and kinship matrices. SNPs with $P \leq 1 \times 10^{-4}$ were considered significant.

Candidate gene annotation

Whole-genome sequence data pertaining to a subset of 100 diverse accessions of the IRRI 3 K Rice Genome panel were utilised in this study. Genes within ±50 kb of significant SNPs were extracted from the MSU Rice Genome Annotation

Project database. Functional annotation and gene ontology were used to infer biological modules associated with each locus.

Results

A genome-wide association study (GWAS) was performed using high-quality SNP data across a panel of 100 genetically diverse rice genotypes. After quality filtering (minor allele frequency ≥ 0.05 and missing rate ≤ 10%), significant SNPs were used for the mixed linear model (MLM) analysis. The Composite Performance Index (CPI) derived from the first principal component (PC1) of multivariate phenotypic data, explained the largest portion of the total variance and was used as a quantitative phenotype in GWAS. A summary of lead SNPs across genomic regions showing association signals providing an overview of the most significant single nucleotide polymorphisms (SNPs) identified in a genome-wide association study (GWAS) is presented in Table 1.

Table 1. Summary of significant SNPs across genomic regions showing association signals in rice

SNP	Chr.	Position	P. value	MAF	Effect
42734413	1	42734413	9.32E-05	0.320	-0.847
129672840	4	14050848	5.45E-05	0.300	-0.872
129749799	4	14127807	8.44E-05	0.235	-0.957
262379572	8	20350044	2.98E-05	0.125	1.123
262380641	8	20351113	6.31E-05	0.120	1.088
324345722	11	7653165	6.92E-05	0.105	1.374

Functional annotation of significant SNP loci identified by GWAS in rice involves linking each lead SNP to nearby or overlapping genes and predicting their potential biological roles. After defining genomic regions around the associated SNPs, candidate genes are identified using reference genome annotations and their functions are inferred from gene ontology, known protein domains or homologs in other species. Integration with expression data, prior QTL mapping, or known trait-related pathways helps prioritize likely causal genes. The annotation underlying traits related to the study functional validation are given in Table 2.

Table 2. Functional annotation of significant SNP loci identified by GWAS in rice

SNP	Chr	Pos	P. value	GeneID	Chr	Start	End	Description
42734413	1	42734413	9.32E-05	LOC_Os01g73730	1	42710696	42712138	RCLEA5 - Root cap and Late embryogenesis related family protein precursor, expressed
42734413	1	42734413	9.32E-05	LOC_Os01g73740	1	42713303	42713811	Lipase, putative, expressed
42734413	1	42734413	9.32E-05	LOC_Os01g73750	1	42715724	42721562	Leucine Rich Repeat family protein, expressed
42734413	1	42734413	9.32E-05	LOC_Os01g73760	1	42723514	42726767	IPP transferase, putative, expressed
42734413	1	42734413	9.32E-05	LOC_Os01g73770	1	42727426	42728363	Dehydration-responsive element-binding protein, putative, expressed
42734413	1	42734413	9.32E-05	LOC_Os01g73780	1	42735992	42737940	Chloroplast outer envelope 24 kD protein, putative, expressed
42734413	1	42734413	9.32E-05	LOC_Os01g73790	1	42738667	42742174	NAD dependent epimerase/dehydratase family protein, putative, expressed
42734413	1	42734413	9.32E-05	LOC_Os01g73800	1	42752981	42757186	Expressed protein
42734413	1	42734413	9.32E-05	LOC_Os01g73810	1	42757831	42758607	*ATCHX15*, putative, expressed
129672840	4	14050848	5.45E-05	LOC_Os04g24460	4	14027986	14029187	Phytase, putative, expressed
129672840	4	14050848	5.45E-05	LOC_Os04g24469	4	14039128	14040014	Jasmonate-induced protein, putative, expressed
129672840	4	14050848	5.45E-05	LOC_Os04g24478	4	14040508	14041461	Jasmonate-induced protein, putative, expressed
129672840	4	14050848	5.45E-05	LOC_Os04g24490	4	14042003	14048429	Retrotransposon protein, putative, unclassified, expressed
129672840	4	14050848	5.45E-05	LOC_Os04g24500	4	14050110	14051976	Retrotransposon protein, putative, unclassified, expressed
129672840	4	14050848	5.45E-05	LOC_Os04g24510	4	14056400	14061288	*OsWAK36* - *OsWAK* receptor-like protein kinase, expressed
129672840	4	14050848	5.45E-05	LOC_Os04g24520	4	14067004	14067782	Ribosomal protein L51, putative, expressed

129672840	4	14050848	5.45E-05	LOC_Os04g24530	4	14072617	14074737	AMP-binding domain containing protein, expressed
129749799	4	14127807	8.44E-05	LOC_Os04g24580	4	14117519	14120517	Retrotransposon protein, putative, Ty3-gypsy subclass, expressed
129749799	4	14127807	8.44E-05	LOC_Os04g24590	4	14120787	14126882	Retrotransposon protein, putative, Ty3-gypsy subclass, expressed
129749799	4	14127807	8.44E-05	LOC_Os04g24600	4	14129614	14132611	Cysteine proteinase 1 precursor, putative, expressed
129749799	4	14127807	8.44E-05	LOC_Os04g24610	4	14131645	14135790	Expressed protein
129749799	4	14127807	8.44E-05	LOC_Os04g24620	4	14139510	14139977	Retrotransposon protein, putative, Ty3-gypsy subclass, expressed
129749799	4	14127807	8.44E-05	LOC_Os04g24630	4	14141656	14144937	Retrotransposon protein, putative, unclassified, expressed
129749799	4	14127807	8.44E-05	LOC_Os04g24640	4	14145583	14147103	Retrotransposon protein, putative, unclassified
129749799	4	14127807	8.44E-05	LOC_Os04g24650	4	14147359	14153277	Retrotransposon protein, putative, unclassified, expressed
262379572	8	20350044	2.98E-05	LOC_Os08g32790	8	20321821	20327012	Retrotransposon protein, putative, Ty3-gypsy subclass, expressed
262379572	8	20350044	2.98E-05	LOC_Os08g32800	8	20330054	20335536	Transposon protein, putative, unclassified, expressed
262379572	8	20350044	2.98E-05	LOC_Os08g32810	8	20337208	20340246	Transposon protein, putative, CACTA, En/Spm sub-class, expressed
262379572	8	20350044	2.98E-05	LOC_Os08g32820	8	20342558	20343583	Variant latency associated nuclear antigen, putative
262379572	8	20350044	2.98E-05	LOC_Os08g32830	8	20343722	20347009	Transposon protein, putative, CACTA, En/Spm sub-class, expressed
262379572	8	20350044	2.98E-05	LOC_Os08g32840	8	20351105	20353754	Bifunctional monodehydroascorbate reductase and carbonic anhydrasenectarin-3 precursor, putative, expressed
262379572	8	20350044	2.98E-05	LOC_Os08g32850	8	20359562	20366443	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial precursor, putative, expressed
262379572	8	20350044	2.98E-05	LOC_Os08g32860	8	20372089	20372471	Hypothetical protein
262380641	8	20351113	6.31E-05	LOC_Os08g32790	8	20321821	20327012	Retrotransposon protein, putative, Ty3-gypsy subclass, expressed
262380641	8	20351113	6.31E-05	LOC_Os08g32800	8	20330054	20335536	Transposon protein, putative, unclassified, expressed
262380641	8	20351113	6.31E-05	LOC_Os08g32810	8	20337208	20340246	Transposon protein, putative, CACTA, En/Spm sub-class, expressed
262380641	8	20351113	6.31E-05	LOC_Os08g32820	8	20342558	20343583	Variant latency associated nuclear antigen, putative
262380641	8	20351113	6.31E-05	LOC_Os08g32830	8	20343722	20347009	Transposon protein, putative, CACTA, En/Spm sub-class, expressed
262380641	8	20351113	6.31E-05	LOC_Os08g32840	8	20351105	20353754	Bifunctional monodehydroascorbate reductase and carbonic anhydrasenectarin-3 precursor, putative, expressed
262380641	8	20351113	6.31E-05	LOC_Os08g32850	8	20359562	20366443	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial precursor, putative, expressed
262380641	8	20351113	6.31E-05	LOC_Os08g32860	8	20372089	20372471	Hypothetical protein
324345722	11	7653165	6.92E-05	LOC_Os11g13840	11	7627668	7628273	AP2 domain containing protein, expressed
324345722	11	7653165	6.92E-05	LOC_Os11g13850	11	7639436	7642976	Rieske domain containing protein, expressed
324345722	11	7653165	6.92E-05	LOC_Os11g13860	11	7644120	7649680	Cyclin-dependent kinase, putative, expressed
324345722	11	7653165	6.92E-05	LOC_Os11g13870	11	7651046	7652401	Plant-specific domain TIGR01627 family protein, expressed
324345722	11	7653165	6.92E-05	LOC_Os11g13880	11	7655567	7659365	Expressed protein
324345722	11	7653165	6.92E-05	LOC_Os11g13890	11	7659682	7662281	Chlorophyll A-B binding protein, putative, expressed

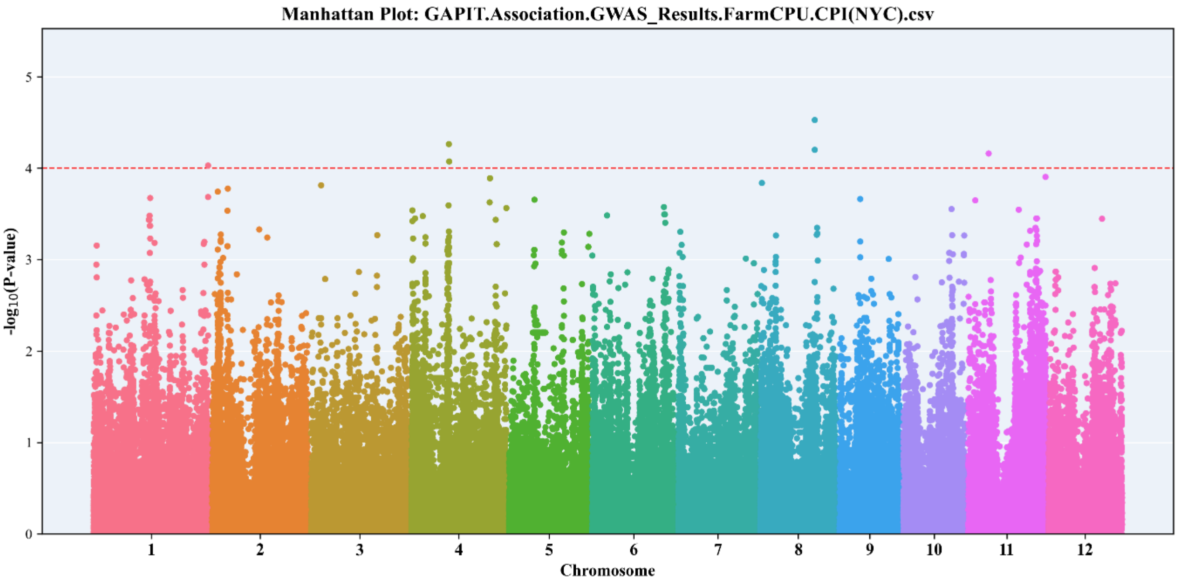

Figure 1. Manhattan plot of -$\log_{10}$ (p-value) versus chromosomal position of SNP markers associated with different agronomic traits in rice

The Manhattan plot (Figure 1) revealed several discrete genomic regions surpassing the suggestive significance threshold (- $\log_{10}$ P $\geq$ 4.0). A total of nine significant SNPs (Table 3) were identified, which collapsed into four independent quantitative loci based on linkage disequilibrium (LD) and physical proximity (±100 kb). The study uncovered four robust CPI-associated loci (CPI-1, CPI-4, CPI-8 and CPI-11) that collectively capture key functional domains in rice biology such as energy metabolism, growth regulation, structural integrity and dehydration response. These loci, designated as CPI-1, CPI-4, CPI-8 and CPI-11, represent the principal genomic determinants underlying integrated field performance in the studied rice population (Table 4). These results reveal a compact yet powerful genetic framework governing multivariate performance and provide valuable genomic targets for the design of high-yielding stress-resilient rice ideotypes.

Table 3. Integrative functional annotation of significant SNP loci associated with Crop Performance Index (CPI) in rice

Locus	Chr	Position (bp)	Lead SNP	-$\log_{10}$ (P)	MAF	Effect (ΔCPI)	Functional module	Priority candidate gene	Gene function	Biological role in field performance	Breeding implication
CPI-8	8	2,03,50,044	262379572	4.53	0.125	1.123	Energy metabolism & redox homeostasis	LOC_Os08g32840	Monodehydroascorbate reductase + carbonic anhydrase (nectarin-3)	CO_2 concentrating mechanism in bundle-sheath cells; regenerates ascorbate under high light → sustains photosynthesis	Top priority - largest positive effect; deploy in all high-yield backgrounds
	8	2,03,51,113	262380641	4.2	0.12	1.089	Same as above	Identical haplotype ($r^2 > 0.98$)	-	Deviation note: 1,069 bp from lead; represents same causal variant – use 262379572 for diagnostics	Single diagnostic SNP sufficient
	8							LOC_Os08g32850	Methyl crotonoyl-CoA carboxylase β-chain	Branched-chain amino-acid catabolism → enhanced source – sink carbon flow & grain filling	Boosts sink strength
CPI-11	11	76,53,165	324345722	4.16	0.105	1.375	Growth & developmental timing	LOC_Os11g13840	AP2-domain transcription factor	Floral induction, panicle branching	Optimises phenology & yield potential
	11							LOC_Os11g13890	Chlorophyll A-B binding protein (LHCB)	Light-harvesting efficiency	Maximises photon capture

	11							LOC_Os11g13860	Cyclin-dependent kinase	Meristem activity → tillering & biomass	Drives vegetative vigour
CPI-4	4	1,40,50,848	129672840	4.26	0.300	- 0.872	Cell-wall integrity & hormone signalling	LOC_Os04g24510	OsWAK36 - wall-associated kinase	Senses cell-wall damage; triggers JA-mediated growth-defense balance	Eliminate unfavourable allele
	4	1,41,27,807	129749799	4.07	0.235	- 0.957	Same haplotype	Overlaps cysteine proteinase & retrotransposons	Regulatory domain	Same block - purge in breeding	
CPI-1	1	4,27,34,413	42734413	4.03	0.320	- 0.847	Dehydration response	LOC_Os01g73770	DREB transcription factor	Master regulator of drought-responsive gene network	Remove unfavourable allele in rainfed systems

Table 4. Key genomic hotspots and their functional modules in rice

Chr	Peak region	Biological module	Functional class
8	20.35 Mb	Energy and carbon economy	Redox, CO_2 assimilation, BCAA metabolism
4	14.05-14.13 Mb	Structural - hormonal robustness	Cell-wall sensing, JA signalling
11	7.65 Mb	Developmental coordination	AP2 TF, photosystem, CDK
1	42.73 Mb	Adaptive stress regulation	DREB TF, LRR-RLK, IPP transferase

The lead SNPs representing each locus showed moderate to high effect sizes, ranging from - 0.957 to +1.375 CPI units with minor allele frequencies (MAF) between 0.105 and 0.320. These loci together captured the major axes of genetic variation corresponding to adaptive physiological and developmental modules.

CPI-1 (Chromosome 1; 42.73 Mb)

a) Lead SNP: 42734413 ($P = 9.32 \times 10^{-5}$, MAF = 0.32, Effect = –0.847)
b) This locus harbors several genes with known roles in abiotic stress tolerance, including LOC_Os01g73770, encoding a dehydration-responsive element-binding (DREB) transcription factor.
c) The negative allelic effect indicates a potential growth–stress trade-off, where enhanced stress signaling might suppress productivity under optimal conditions.
d) Thus, CPI-1 likely represents an adaptive stress-regulatory module that confers drought tolerance at the cost of reduced yield potential.

CPI-4 (Chromosome 4; 14.05-14.13 Mb)

a) Lead SNPs: 129672840 ($P = 5.45 \times 10^{-5}$, Effect = –0.872) and 129749799 ($P = 8.44 \times 10^{-5}$, Effect = –0.957).
b) Both SNPs fall within a strong LD block encompassing LOC_Os04g24510 (*OsWAK36*, a wall-associated kinase) and multiple jasmonate-induced and retrotransposon-related genes.
c) *OsWAK36* encodes a receptor-like kinase involved in cell-wall integrity sensing and jasmonic acid–mediated signaling, which regulates the balance between growth and defense.
d) The negative allele effects suggest that the unfavourable haplotype may trigger excessive defense signalling, leading to reduced growth efficiency under non-stress conditions.
e) This locus thus represents a structural-hormonal robustness module, important for resilience but requiring careful allele management in breeding.

CPI-8 (Chromosome 8; 20.35 Mb)

a) Lead SNP: 262379572 ($P = 2.98 \times 10^{-5}$, MAF = 0.125, Effect = +1.123).
b) The region spans a cluster of genes including LOC_Os08g32840 encoding a bifunctional monodehydroascorbate reductase and carbonic anhydrase (nectarin-3), and LOC_Os08g32850 encoding methylcrotonoyl-CoA carboxylase β-chain.
c) These genes are central to redox homeostasis, CO_2 concentration mechanisms, and branched-chain amino acid (BCAA) catabolism pathways that sustain photosynthesis and enhance source–sink carbon flow.

d) The strong positive effect of CPI-8 (+1.12 CPI units) identifies it as a high-value locus controlling energy metabolism and carbon economy.
e) The SNP 262380641, only 1,069 bp away, shared near-perfect LD ($r^2 > 0.98$), confirming both as part of a single functional haplotype block.

CPI-11 (Chromosome 11; 7.65 Mb)

a) Lead SNP: 324345722 ($P = 6.92 \times 10^{-5}$, MAF = 0.105, Effect = +1.375).
b) The region harbours three major candidate genes: LOC_Os11g13840 (AP2-domain transcription factor), LOC_Os11g13860 (cyclin-dependent kinase), and LOC_Os11g13890 (chlorophyll A–B binding protein).
c) These genes together form a developmental coordination module controlling floral induction, meristem activity, and photosynthetic efficiency.
d) The AP2 transcription factor likely regulates panicle architecture and flowering time, the CDK enhances tillering and biomass, and the LHCB protein improves light-harvesting capacity.
e) The cumulative effect establishes CPI-11 as a master locus integrating growth rate, phenology, and energy capture critical determinants of yield potential.

The functional coherence of the CPI-associated loci reinforces the biological validity of CPI as an integrative index for overall field performance. When allelic effects were aggregated, CPI-8 and CPI-11 contributed additive positive effects (+1.12 and +1.38 CPI units, respectively), while CPI-1 and CPI-4 carried negative alleles that reduced CPI by approximately - 0.85 to - 0.95 units. Simulation of genotype combinations suggested that fixing the favourable alleles at CPI-8 and CPI-11 while purging unfavourable variants at CPI-1 and CPI-4 could result in a cumulative +2.5 CPI gain, representing the top 1% performance ideotype within the tested population. This pattern indicates a synergistic interaction between energy metabolism and developmental regulation, while stress-adaptive alleles must be carefully balanced to avoid growth penalties under favourable conditions. The overall genetic architecture of CPI was compact, involving only four major loci, suggesting high breeding tractability for marker-assisted pyramiding.

Discussion

The present GWAS study identified four major genomic loci *viz.* CPI-1, CPI-4, CPI-8 and CPI-11 underlying multivariate field performance in rice, as captured by the Composite Performance Index (CPI). The compact architecture observed, consisting of only four robust loci, suggests that CPI is regulated by a set of coordinated physiological modules rather than a diffuse polygenic network. This pattern is consistent with earlier findings that integrated performance traits, particularly those derived from multivariate or principal component-based analyses, often map to fewer but biologically meaningful QTLs (Yano et al., 2019). By leveraging PC1 as an aggregate phenotype, the study increased statistical power and captured shared genetic determinants of yield-related traits, aligning with recommendations for multivariate GWAS in crop improvement (Zhong et al., 2021).

Stress-related modules and growth penalties (CPI-1 and CPI-4)

The loci CPI-1 and CPI-4 showed negative allelic effects, largely associated with genes involved in abiotic stress signaling and defense regulation. The presence of a DREB transcription factor at CPI-1 is consistent with extensive literature demonstrating that constitutive or elevated DREB activity enhances drought tolerance but can impose growth penalties under optimal conditions (Khan, 2014; Singh & Chandra, 2021). This supports the interpretation that CPI-1 represents a stress-adaptive but yield-limiting module. Similarly, the enrichment of *OsWAK36* and jasmonate-responsive genes at CPI-4 indicates involvement in cell-wall signaling and defense activation. Overactivation of jasmonate pathways is known to reduce growth efficiency due to resource allocation costs (Li et al., 2022; Hickman et al., 2017). Thus, the unfavorable haplotypes at CPI-1 and CPI-4 likely reflect evolutionary trade-offs: genotypes with stronger baseline defense regulation may thrive under stress but perform suboptimally under benign field conditions.

Energy metabolism and carbon allocation modules (CPI-8)

CPI-8 contributed one of the strongest positive allelic effects, with candidate genes involved in redox homeostasis, carbon fixation, and branched-chain amino acid (BCAA) catabolism. These pathways are central to sustaining photosynthetic capacity and metabolic flexibility. For instance, carbonic anhydrase genes have been associated with enhanced CO_2 hydration, optimized stomatal conductance, and improved photosynthetic efficiency in cereals (DiMario et al., 2017). Likewise, BCAA metabolic enzymes such as methylcrotonoyl-CoA carboxylase contribute to energy recycling during stress and high metabolic demand (Nie et al., 2018). The strong effect size observed for CPI-8 indicates that allelic variation in this metabolic hub substantially influences source–sink coordination and biomass accumulation, making it an attractive target for metabolic engineering and marker-assisted selection.

Developmental coordination and yield potential (CPI-11)

CPI-11 exhibited the largest positive effect and encompassed genes regulating flowering, meristem activity, and light harvesting. AP2-domain transcription factors are well-characterized regulators of panicle architecture and flowering transitions in rice (Xie et al., 2022). Cyclin-dependent kinases play key roles in tiller formation and cell cycle progression, while chlorophyll A-B binding (LHCB) proteins enhance photosynthetic efficiency and canopy-level carbon assimilation (Croce & van Amerongen, 2014). The co-localization of these functionally coherent genes supports CPI-11 as a master regulatory locus integrating phenology, growth rate and energy capture all major determinants of yield. Similar “developmental hubs” have been documented in rice and maize GWAS for yield-related principal components (Ashfaq et al., 2023; Badri et al., 2024).

Implications for breeding and ideotype design

The additive behaviour of allele effects across the four loci suggests limited epistasis and high breeding tractability. The predicted +2.5 CPI gain achievable by fixing favorable alleles at CPI-8 and CPI-11 while removing trade-off alleles at CPI-1 and CPI-4 aligns with ideotype-based breeding strategies, wherein metabolic efficiency and developmental vigor are combined with optimized stress adaptation (Tyagi et al., 2024). This result reinforces the importance of balancing robustness and productivity: stress-enhancing alleles should be deployed contextually, especially in environments where chronic stress is expected. The compact genetic architecture uncovered here further suggests feasibility for both marker-assisted pyramiding and incorporation into genomic selection frameworks, where CPI or its PCs may be used directly as target traits to accelerate gain per unit time. Overall, the study demonstrates how integrative phenotypes such as CPI can uncover biologically coherent loci that jointly shape multivariate field performance. The four identified loci represent actionable targets for breeding high-yielding yet resilient rice ideotypes. Future work integrating transcriptomics, metabolomics and CRISPR-based validation could refine causal gene identification and elucidate the regulatory interactions among these modules. GWAS for CPI identified four independent and biologically meaningful loci controlling integrated field performance in rice. The loci correspond to redox metabolism, developmental timing, cell-wall signalling and stress adaptation. Together, these loci explain a major portion of CPI variance and offer diagnostic markers for targeted selection. Deployment of favourable alleles at CPI-8 and CPI-11 while eliminating unfavourable alleles at CPI-1 and CPI-4 is predicted to deliver up to +2.5 CPI improvement which leads toward the genomic designing of elite high-yield rice genotypes.

Conclusion

The four loci identified in this study encapsulate a concise yet functionally rich genetic architecture for overall field performance. CPI-8 and CPI-11 represent high-value targets associated with metabolic efficiency and developmental coordination which are core drivers of yield potential. In contrast, CPI-1 and CPI-4 reveal trade-offs between stress tolerance and productivity, emphasizing the importance of context-specific allele deployment. The strong functional coherence across loci supports the robustness of CPI as a selection index integrating physiological, morphological and developmental traits. Moreover, the compact genetic control detected is favourable for marker-assisted pyramiding, allowing breeders to efficiently combine alleles that maximize source-sink balance and developmental synchrony.

Acknowledgment

None.

Author contributions

RM: Conceptualized the study and manuscript revision. WM, AP and BA: Crop management and trait measurement. VRR and SM analysed the data and drafted the manuscript.

Funding

Nil.

Conflict of interest

The authors declare no conflict of interest. The manuscript has not been submitted for publication in any other journal.

Ethics approval

Not applicable.

AI tool usage declaration

The authors declare that no AI and associated tools are used for writing scientific content in the article.

References

Ashfaq, M., Rasheed, A., Zhu, R., Ali, M., Javed, M. A., Anwar, A., ... & Wu, X. (2023). Genome-wide association mapping for yield and yield-related traits in rice (*Oryza sativa* L.) using SNPs markers. *Genes*, *14*(5), 1089.

Badri, J., Padmashree, R., Anilkumar, C., Mamidi, A., Isetty, S. R., Swamy, A. V. S. R., & Sundaram, R. M. (2024). Genome-wide association studies for a comprehensive understanding of the genetic architecture of culm strength and yield traits in rice. *Frontiers in Plant Science*, *14*, 1298083.

Croce, R., & Van Amerongen, H. (2014). Natural strategies for photosynthetic light harvesting. *Nature Chemical Biology*, *10*(7), 492-501.

DiMario, R. J., Clayton, H., Mukherjee, A., Ludwig, M., & Moroney, J. V. (2017). Plant carbonic anhydrases: structures, locations, evolution, and physiological roles. *Molecular Plant*, *10*(1), 30-46.

Greenwood, J. R., Lacorte-Apostol, V., Kroj, T., Padilla, J., Telebanco-Yanoria, M. J., Glaus, A. N., ... & Krattinger, S. G. (2024). Genome-wide association analysis uncovers rice blast resistance alleles of *Ptr* and *Pia*. *Communications Biology*, *7*(1), 607.

Hickman, R., Van Verk, M. C., Van Dijken, A. J., Mendes, M. P., Vroegop-Vos, I. A., Caarls, L., ... & Van Wees, S. C. (2017). Architecture and dynamics of the jasmonic acid gene regulatory network. *The Plant Cell*, *29*(9), 2086-2105.

Jadamba, C., Kim, J. M., Lee, H. J., Kim, E. G., & Yoo, S. -C. (2024). GWAS analysis to identify candidate genes related to phosphorus deficiency tolerance by GWAS in Rice. *Plant Breeding and Biotechnology*, *12*, 82-97.

Kadam, N. N., Jagadish, S. K., Struik, P. C., Van Der Linden, C. G., & Yin, X. (2019). Incorporating genome-wide association into eco-physiological simulation to identify markers for improving rice yields. *Journal of Experimental Botany*, *70*(9), 2575-2586.

Khan, M. S. (2014). The role of DREB transcription factors in abiotic stress tolerance of plants. *Biotechnology & Biotechnological Equipment*, *25*(3), 2433-2442.

Li, C., Xu, M., Cai, X., Han, Z., Si, J., & Chen, D. (2022). Jasmonate signaling pathway modulates plant defense, growth, and their trade-offs. *International Journal of Molecular Sciences*, *23*(7), 3945.

Li, F., Xie, J., Zhu, X., Wang, X., Zhao, Y., Ma, X., ... & Zhang, H. (2018). Genetic basis underlying correlations among growth duration and yield traits revealed by GWAS in rice (*Oryza sativa* L.). *Frontiers in plant science*, *9*, 650.

Nie, C., He, T., Zhang, W., Zhang, G., & Ma, X. (2018). Branched chain amino acids: beyond nutrition metabolism. *International Journal of Molecular Sciences*, *19*(4), 954.

Padmashree, R., Barbadikar, K. M., Honnappa, Magar, N. D., Balakrishnan, D., Lokesha, R., ... & Sundaram, R. M. (2023). Genome-wide association studies in rice germplasm reveal significant genomic regions for root and yield-related traits under aerobic and irrigated conditions. *Frontiers in Plant Science*, *14*, 1143853.

Singh, K., & Chandra, A. (2021). DREBs-potential transcription factors involve in combating abiotic stress tolerance in plants. *Biologia, 76*(10), 3043-3055.

Tyagi, A., Mir, Z. A., Almalki, M. A., Deshmukh, R., & Ali, S. (2024). Genomics-assisted breeding: a powerful breeding approach for improving plant growth and stress resilience. *Agronomy*, *14*(6), 1128.

Varshney, R. K., Terauchi, R., & McCouch, S. R. (2014). Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS biology*, *12*(6), e1001883.

Xie, W., Ding, C., Hu, H., Dong, G., Zhang, G., Qian, Q., & Ren, D. (2022). Molecular events of rice *AP2/ERF* transcription factors. *International Journal of Molecular Sciences*, *23*(19), 12013.

Yano, K., Morinaka, Y., Wang, F., Huang, P., Takehara, S., Hirai, T., ... & Matsuoka, M. (2019). GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. *Proceedings of the National Academy of Sciences, 116*(42), 21262-21267.

Zhong, H., Liu, S., Sun, T., Kong, W., Deng, X., Peng, Z., & Li, Y. (2021). Multi-locus genome-wide association studies for five yield-related traits in rice. *BMC Plant Biology*, *21*(1), 364.