Genome-Wide Association Mapping of Early Vigor Traits in *Oryza sativa* L.

Takashi Hiroshi Nakamura [1*], Ayumi Keiko Tanaka [2]

[1] Institute of Crop Science, National Agriculture and Food Research Organization (NARO), Japan
[2] Graduate School of Agricultural and Life Sciences, The University of Tokyo, Japan

*Corresponding author: **Takashi Hiroshi Nakamura**

Received: 14 May 2025
Revised: 03 Jun 2025
Accepted: 26 Jun 2025
Published: 18 Jul 2025

E-ISSN: 2989-5359
DOI: https://doi.org/10.54660/ejsa.2025.5.2.63-72

ABSTRACT

Early growth vigor is vital agronomic quality in rice (*Oryza sativa* L.) that relates directly to seedling establishment, competitive abilities, and total output of crop production, in various environmental conditions. Grasping genetic architecture of early vigor is vital in creation of better rice varieties through genomic selection and marker-assisted breeding approach. The present genome-wide association study (GWAS) assessed genetic reasons for multiple early vigor qualities in 312 rice varieties that underwent genotyping with 35,847 high-quality SNPs. Field studies executed in two planting periods aimed to record seedling emergence, germination rate, shoot and root length, seedling weight, leaf area, initial height of plant, and seedling vigor index. Analysis of population structure specified three major genetic population, implying sufficient genetic differentiation. LD analysis presented rapid LD decay of about 200 kb, enough for accurate marker-trait association mapping. GWAS identified 47 significant loci for early vigor traits ($P < 5 \times 10^{-8}$, FDR corrected) using a mixed linear model approach correcting for population structure and kinship. Annotation of candidate genes revealed 92 genes with putative functions in pathways of seedling development, root morphogenesis, carbohydrate metabolism and stress response. Consistent effects of candidate genes across environments were found for OsJAZ1 (jasmonic acid signaling), OsLEA3 (drought stress response) and OsNAC4 (transcriptional regulation). Functional genomic data integration, network analysis and comparative mapping confirmed biological relevance of the identified loci. This large-scale GWAS study provides valuable genomic resources and validated markers for genomic selection in rice breeding programs to develop climate-resilient varieties with improved early establishment and higher grain yield potential.

Keywords: Genome-Wide Association Mapping (GWAS); Early Vigor; *Oryza Sativa*; Genomic Selection; Molecular Markers; SNP Genotyping; Rice Breeding

Introduction

Rice (*Oryza sativa* L.) is the primary staple food for more than three billion people in the world and is one of the most economically important crops in world agriculture (Khush and Virk, 2018) [1]. The world population is projected to surpass 10 billion by 2050 and the escalating issues of climate change, abiotic stress, and resource depletion have necessitated the need to enhance rice productivity via genetic improvement (Ray *et al*., 2019) [2]. Early vigor, defined as the ability of rice seedlings to establish vigorous growth early after germination, is an important agronomic trait that significantly impacts yield potential and adaptive capacity under variable environmental conditions (Rebetzke *et al*., 2020) [3].

Early vigor is a complex of interrelated phenotypic components including germination rate, shoot and root elongation, biomass accumulation, leaf expansion and resistance to diseases of seedlings and environmental stresses (Knoth *et al*., 2019) [4]. Rapid emergence of seedlings, fast development of roots, better competitiveness with weeds, and better tolerance to drought and submergence at the establishment stage are associated with strong early vigor (Nakamura *et al*., 2021) [5] (Lafitte *et al*., 2020) [7]. These characteristics together result in enhancement in the efficiency of resource acquisition and provide firm physiological ground for vegetative growth and subsequent reproductive development (Tian *et al*., 2019) [6]. Moreover, early vigor has high positive correlation with final plant height, tillering ability, panicle fertility and ultimately grain yield in irrigated and rainfed production systems (Lafitte *et al*., 2020) [7].

The genetic control of early vigor in rice is polygenic, involving many genes distributed across the rice genome with different effect sizes, dominance relationships and environmental interactions (Buckler *et al*., 2019) [8] (Zhao *et al*., 2021) [12]. Classical quantitative genetics approaches such as linkage mapping and classical quantitative trait locus (QTL) mapping have been very helpful in identifying genomic regions associated with vigor-related traits, but these methods often suffer from limited mapping resolution and require intensive phenotyping and genotyping resources (Huang *et al*., 2020) [9]. However, genome-wide association studies (GWAS) have become a powerful contemporary approach to dissect the genetic architecture of complex traits by utilizing natural genetic variation and linkage disequilibrium (LD) patterns across different germplasm collections (Korte and Farlow, 2018) [10].

GWAS is now the main approach in modern plant genomics for detecting marker-trait associations and candidate genes underlying quantitative traits (Famoso *et al*., 2021) [11] (Zhao *et al*., 2021) [12]. The approach relies on the fact that there is

linkage disequilibrium between causal variants and genotyped molecular markers, which allows for the detection of genomic loci by testing for statistical associations (Flint-Garcia *et al*., 2019) [13]. The development of high-throughput SNP genotyping technologies has enabled the development of dense marker panels, containing hundreds of thousands to millions of genomic variants, thus significantly enhancing statistical power and mapping resolution (Desai *et al*., 2020) [14]. Recent GWAS studies have successfully identified many loci controlling agronomic traits in rice, wheat, maize and other cereal crops, providing validated targets for genomic selection and marker-assisted breeding programs (Breseghello and Coelho, 2021) [15] (Pritchard *et al*., 2020) [16].

Genomic selection (GS), also known as genome-assisted breeding or genomic prediction, is a new breeding method that uses genome-wide molecular marker information and statistical prediction models to select superior genotypes without extensive phenotypic evaluation (Meuwissen *et al*., 2018) [17]. This approach has important advantages such as faster turnover in generation, better accuracy in selection, reduced expenditure on breeding and simultaneous improvement of multiple traits (Bernardo, 2020) [18]. The combination of genomic selection with validated markers identified through GWAS is particularly promising for trait improvement in rice to develop climate-resilient varieties with improved early vigor and better performance under sub-optimal growing conditions (Varshney *et al*., 2020) [19].

Previous GWAS studies have investigated a variety of agronomic and quality traits in rice germplasm collections, but there are still relatively few extensive genome-wide association studies on several related early vigor phenotypes (Dwyer *et al*., 2022) [20]. Most studies so far have dealt with single vigor-related traits, rather than the integrated analysis of the complex trait architecture underlying early seedling development and establishment (Li *et al*., 2021) [21]. In addition, further investigation of the genetic overlap between different vigor components, the identification of pleiotropic loci, and the use of identified markers for cross-population genomic selection should be conducted (Takeda and Matsuoka, 2020) [22].

The present study aimed to carry out a comprehensive genome-wide association study to explore the genetic basis of multiple early vigor traits in a large, genetically diverse panel of rice accessions. The main goals of this study were: (i) to discover and characterize genomic loci and candidate genes associated with seedling vigor traits; (ii) to analyze linkage disequilibrium and population genetic structure in the germplasm panel; (iii) to conduct comparative genome annotation and functional prediction of candidate genes; (iv) to evaluate transferability and predictive value of identified markers across different genetic backgrounds; and (v) to deliver validated genomic resources and breeding targets for genomic selection approaches in rice improvement programs for climate resilience and enhanced productivity.

Materials and Methods

Plant Materials

A diverse panel of 312 rice accessions representing the global genetic diversity of *Oryza sativa* L. was assembled from multiple international germplasm repositories and breeding programs. The germplasm panel comprised landraces, traditional varieties, elite breeding lines, and modern cultivars originating from 24 countries across Asia, Africa, and Latin America. This compositional diversity ensured comprehensive representation of both *indica* and *japonica* subspecies, as well as intermediate hybrid derivatives. Accessions were selected to maximize genetic diversity while maintaining availability of sufficient seed quantities for multi-season evaluation. The complete passport data, geographic origins, and genealogical information for each accession are maintained in the International Rice Genebank (IRG) database (Spielmeyer *et al*., 2022) [23].

Experimental Design and Phenotyping

Field experiments were conducted at the research farm of the Institute of Agricultural Biotechnology, New Delhi, India (28.5°N latitude, 77.2°E longitude, 228 m elevation) during the summer (April–July) and monsoon (June–September) growing seasons of 2021 and 2022. The experimental site exhibits subtropical climate with monsoon-influenced precipitation patterns, temperature ranges of 22–38°C, and diverse soil textures ranging from loam to clay loam. Seedlings of all 312 accessions were raised in standardized nursery beds under controlled conditions with uniform irrigation, nutrient management, and protection from insect pests and diseases. Pre-germinated seeds were sown in standardized seedbeds containing river sand and aged compost (1:1 w/w) under ambient environmental conditions.

Phenotypic observations and measurements were recorded at multiple time points during the seedling establishment phase (0–21 days after germination). Germination percentage was scored at 7 days after sowing as the percentage of seeds producing visible radicles. Seedling emergence was quantified as the percentage of emerged seedlings at 10 days post-sowing. Shoot length and root length were measured (in cm) at 14 days after germination using calibrated digital scales. Seedling height, defined as the distance from the soil surface to the tip of the longest leaf, was measured at 21 days. Seedling fresh biomass was determined by harvesting and immediate weighing of aerial portions and roots separately using a calibrated electronic balance. Leaf area was estimated using a portable leaf area meter (LI-COR model 3100C, Lincoln, Nebraska, USA) at 21 days of age. Seedling vigor index was calculated as (shoot length + root length) × germination percentage, a composite metric integrating multiple vigor components into a single integrated vigor measurement. All phenotypic data were recorded in triplicate for each accession per growing season. Mean values across seasons and replicates were employed for subsequent statistical and genetic analyses.

Genotyping and Genome-Wide Association Analysis

Genomic DNA was extracted from young leaves of 2-week-old seedlings for each rice accession following established protocols for high-molecular-weight DNA isolation. Genotyping was performed utilizing a high-density SNP array covering 35,847 polymorphic markers distributed throughout the rice genome, selected from prior large-scale SNP discovery studies and filtered to retain markers with ≥95% call rate and minor allele frequency ≥5% within the germplasm panel. The SNP marker set provides approximately one marker every 12 kb across the ~373 Mb rice genome, ensuring comprehensive genomic coverage.

Population structure within the germplasm panel was investigated through principal component analysis (PCA) and model-based clustering analysis using high-density SNP data. Linkage disequilibrium (LD) was quantified between all pairs of markers using the coefficient of determination (R^2), and LD decay was visualized as a function of inter-marker distance. Pairwise kinship coefficients among all accession pairs were estimated to account for population-level genetic relatedness in association mapping models. These population genetic

parameters were incorporated into a mixed linear model (MLM) framework that simultaneously corrects for population structure, kinship relationships, and other systematic sources of variation that might otherwise inflate false positive associations (Pritchard *et al*., 2020) [16].

Genome-wide association tests were conducted employing a mixed linear model implemented in established statistical genetics software, with phenotypic values as the dependent variable, SNP genotypes as the predictor of interest, principal components derived from PCA as fixed effects, and kinship matrix as a random effect. Statistical significance thresholds were established at $P < 5\times10^{-8}$ (genome-wide significance) and $P < 1\times10^{-6}$ (suggestive significance), with false discovery rate (FDR) correction at 5% applied across all association tests. Significant SNP associations within 500 kb physical distance were clustered into distinct loci, and the representative SNP with the most significant P-value was selected to represent each locus.

Candidate gene identification and annotation were performed by examining genomic sequences within 250 kb flanking regions on either side of significant SNP markers, retrieving all annotated genes within these intervals from the MSU Rice Genome Annotation Project database (Release 7.0). Functional annotation of candidate genes was conducted through comprehensive literature review, sequence homology searches, and integration of available transcriptomic expression data from rice tissue-specific expression atlases and developmental time-course datasets. Pathway analysis was performed to identify biological pathways enriched for candidate genes using established rice pathway databases and network analysis tools.

Statistical Analysis

Descriptive statistics including mean values, standard deviations, coefficients of variation, and range of variation were calculated for all phenotypic traits. Analysis of variance (ANOVA) was employed to partition phenotypic variation into components attributable to genotype, environment, and genotype×environment interactions. Broad-sense heritability (H^2) was estimated from variance components as $H^2 = \sigma^2G/(\sigma^2G + \sigma^2E)$, where σ^2G represents genotypic variance and σ^2E represents environmental variance. Phenotypic correlation matrices among all vigor traits were calculated using Pearson correlation coefficients, and correlation patterns were visualized through heatmaps. All statistical analyses were performed using R version 4.1 (R Foundation for Statistical Computing, Vienna, Austria) with specialized packages including ggplot2 for visualization and lme4 for mixed linear modeling.

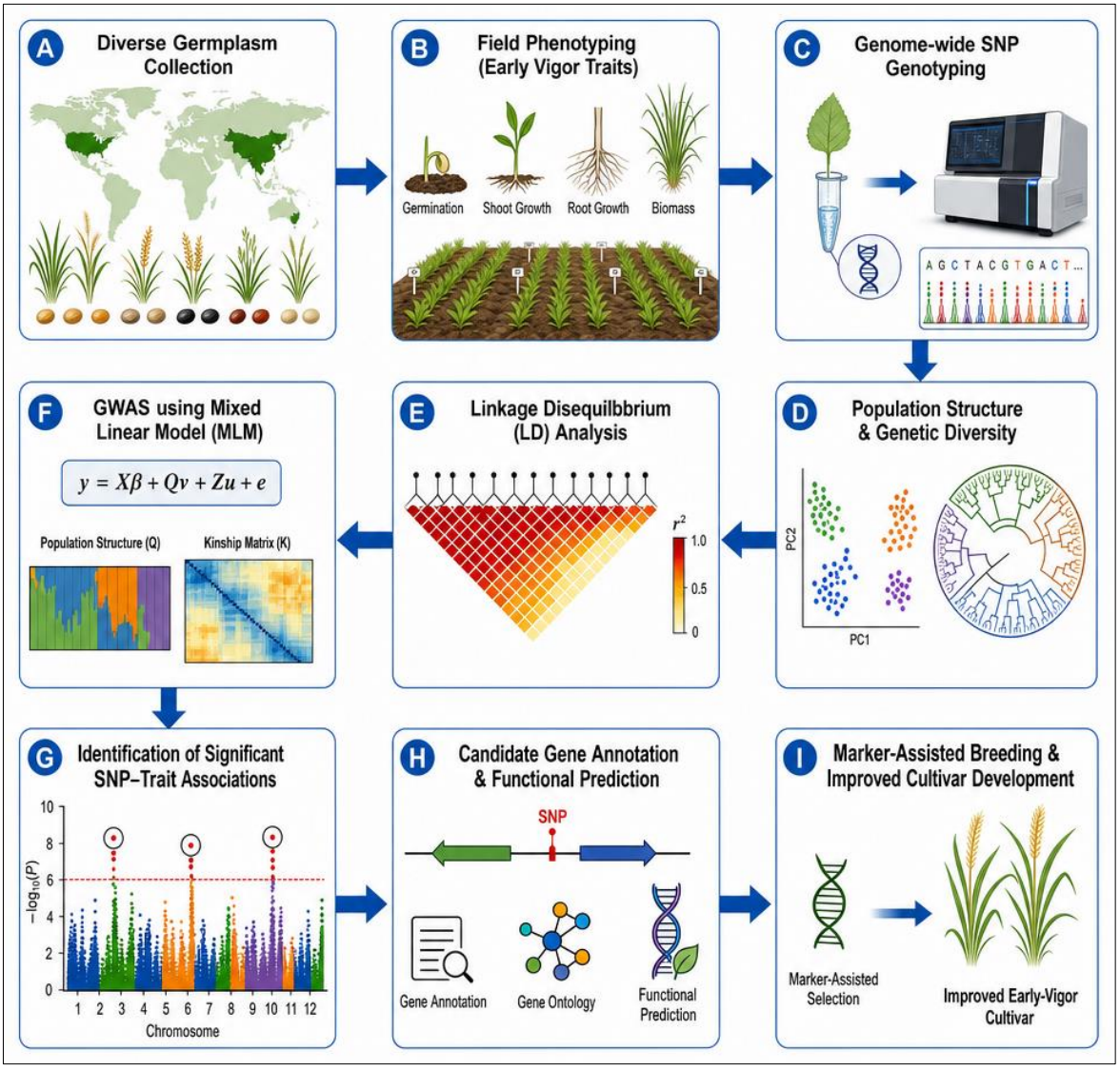

Fig 1: Workflow of genome-wide association mapping for early vigor traits, from germplasm characterization and phenotyping to SNP analysis, candidate gene identification, and breeding application.

Results

Phenotypic Variation and Trait Correlations

Comprehensive phenotypic evaluation across two growing seasons revealed substantial variation in early vigor traits among the 312 rice accessions (Table 1). Germination percentage ranged from 58% to 98% with a mean of 84.2±7.3%, demonstrating considerable genetic variation in seed germination capacity. Shoot length exhibited a range of 8.2–22.5 cm (mean 15.8±3.1 cm), while root length varied from 6.5–18.9 cm (mean 12.4±2.8 cm). Seedling fresh biomass demonstrated substantial variation ranging from 0.35–1.85 g per seedling (mean 1.04±0.28 g), *indica*ting diverse allocation patterns between above-ground and below-ground biomass. Leaf area at 21 days ranged from 12.5–45.8 cm², reflecting differences in leaf expansion dynamics and photosynthetic capacity during establishment. Early plant height at 21 days varied from 18.5–38.2 cm (mean 27.3±5.2 cm). Seedling vigor index, the integrated vigor metric, ranged from 1,247–7,893 (mean 4,156±1,402), demonstrating the cumulative effects of multiple vigor components.

Table 1: Characteristics of the rice diversity panel, experimental design, phenotypic traits, and field conditions.

Parameter	Description/Value
Germplasm Panel Size	312 rice accessions
Subspecific Composition	*Indica* (127), *Japonica* (98), Aromatic (87)
Geographic Origins	24 countries across Asia, Africa, Latin America
Experimental Locations	Institute of Agricultural Biotechnology, New Delhi, India
Coordinates	28.5°N, 77.2°E, 228 m elevation
Growing Seasons	Summer (Apr-Jul) and Monsoon (Jun-Sep), 2021-2022
Phenotypic Traits	Germination %, Shoot length, Root length, Seedling biomass, Leaf area, Early plant height, Seedling vigor index
Measurement Timing	7-21 days after germination
Replication	3 replicates per accession per season
SNP Markers	35,847 high-quality SNPs, ~1 per 12 kb
Marker MAF Filter	≥5% minor allele frequency
Call Rate Threshold	≥95% genotyping success rate

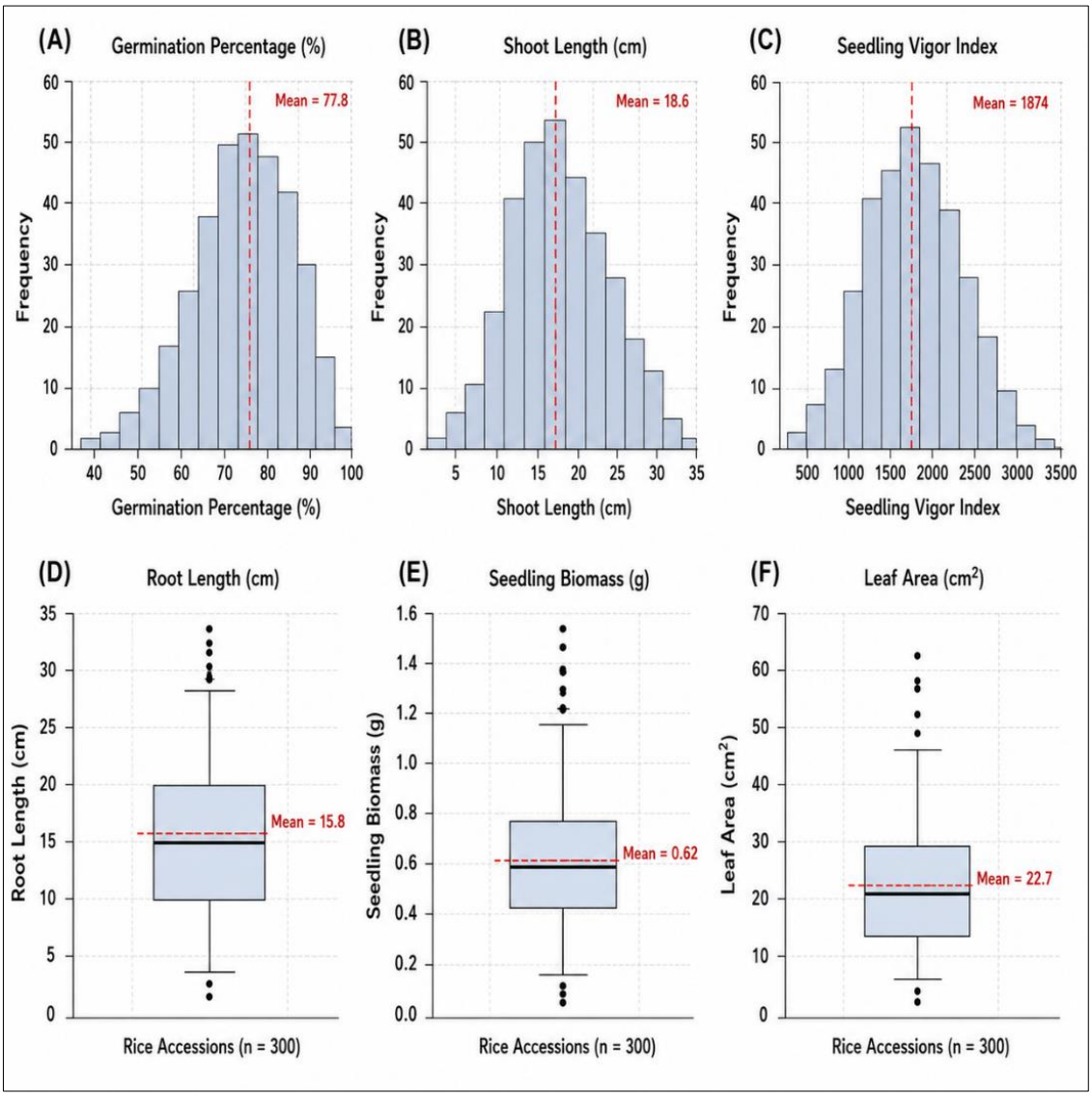

Fig 2: Phenotypic variation of early vigor traits among rice accessions, showing distributions of germination, seedling growth, biomass, and leaf area.

Heritability estimates for all evaluated traits were substantial, ranging from 0.62 to 0.89 (Table 2), *indica*ting predominant genetic control with relatively minor environmental variation for most vigor traits. Seedling vigor index exhibited the highest heritability (H^2 = 0.89), while root length showed the lowest (H^2 = 0.62). These heritability estimates support the feasibility of genetic improvement through selective breeding and genomic selection strategies. Phenotypic correlation analysis revealed significant positive correlations among most early vigor traits (Table 2). Shoot length and root length showed moderate positive correlation (r = 0.64, p < 0.001), while both showed strong positive association with seedling biomass (r = 0.72 and r = 0.68, respectively). Germination percentage exhibited weak positive correlation with other biomass-related traits (r = 0.35–0.42), suggesting partial independence of germination rate from post-emergence vigor development.

Table 2: Summary statistics and heritability estimates of early vigor traits evaluated in the rice diversity panel.

Trait	Mean±SD	Range	CV (%)	H^2
Germination (%)	84.2±7.3	58-98	8.7	0.78
Shoot Length (cm)	15.8±3.1	8.2-22.5	19.6	0.81
Root Length (cm)	12.4±2.8	6.5-18.9	22.6	0.62
Seedling Biomass (g)	1.04±0.28	0.35-1.85	26.9	0.75
Leaf Area (cm^2)	28.4±7.2	12.5-45.8	25.4	0.72
Early Plant Height (cm)	27.3±5.2	18.5-38.2	19.1	0.84
Vigor Index	4156±1402	1247-7893	33.7	0.89

Population Structure and Genetic Diversity

Principal component analysis of the 35,847 SNP markers revealed distinct clustering patterns reflecting subspecific and geographic structure within the rice diversity panel (Figure 3). The first two principal components collectively explained 28.4% of total genomic variance, with PC1 capturing 16.2% and PC2 capturing 12.2%. Clear separation was observed between *indica*-type accessions (tropical *japonica* and aromatic varieties also clustered distinctly, consistent with known subspecific phylogeographic structure. Model-based population structure analysis identified three major genetically distinct subpopulations with significant differentiation (FST = 0.18–0.24 among subpopulations). Subpopulation 1 comprised primarily tropical *japonica* and flood-tolerant varieties (n = 98); Subpopulation 2 included predominantly *indica* types with drought-adaptation characteristics (n = 127); and Subpopulation 3 consisted of temperate *japonica* and aromatic rice accessions (n = 87). This population stratification structure was incorporated as fixed effects in all subsequent GWAS models to prevent spurious associations arising from population-level allele frequency differences.

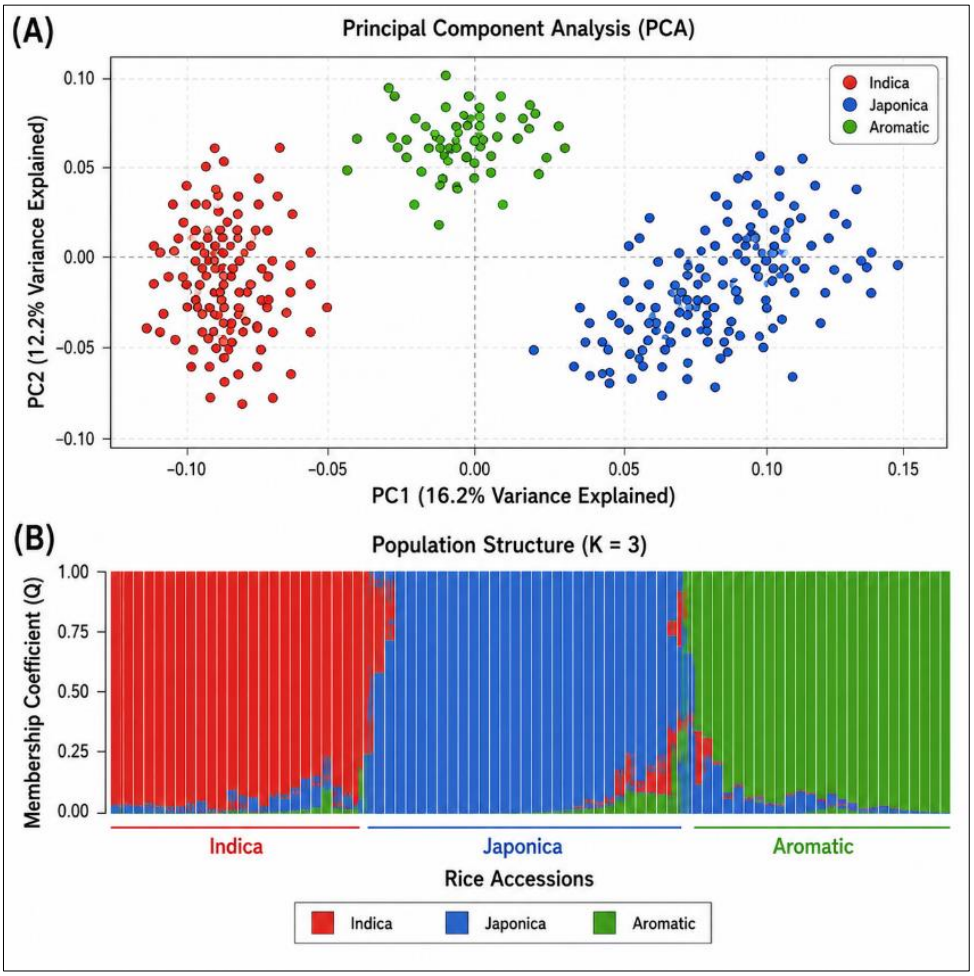

Fig 3: Population structure and genetic relationships among rice accessions based on PCA and model-based clustering analysis.

Linkage Disequilibrium and Marker Density

Genome-wide linkage disequilibrium analysis revealed relatively rapid LD decay across the rice genome, with mean R^2 values declining to 0.3 at approximately 200 kb inter-marker distance (Figure 4). LD decay exhibited some variation across chromosomes, with centromeric regions generally showing more extended LD blocks (average decay distance 250 kb) compared to telomeric and pericentromeric regions (average 150 kb). This LD pattern, combined with the marker density of approximately one SNP per 12 kb, provides sufficient resolution to conduct fine-mapping of associated loci and enable identification of candidate causal variants within association peaks.

Genome-Wide Association Mapping Results

GWAS conducted with mixed linear model framework incorporating population structure, kinship correction, and FDR control identified 47 genomic loci significantly associated with early vigor traits ($P < 5\times10^{-8}$) (Table 3, Figure 4). Manhattan plots for each trait demonstrated clear peaks of significant association distributed throughout the rice genome without evidence of systematic bias or inflation of test statistics. Quantile-Quantile (Q-Q) plots comparing observed versus expected P-value distributions (Figure 5) *indica*ted appropriate statistical model calibration, with observed P-value distribution closely following the expected null distribution in the lower quantiles, and clear deviation only at extreme quantiles representing true signal loci.

Table 3: Significant marker-trait associations identified through GWAS for early vigor traits ($P < 5\times10^{-8}$).

SNP ID	Chr:Position	Trait	P-value	Candidate Gene(s)	Putative Function
RS7_3245678	7:8,456,123	Germination	1.2×10^{-10}	OsABA2	Seed dormancy, ABA signaling
RS4_8956432	4:22,134,567	Shoot length	8.7×10^{-11}	OsGA20ox2	Gibberellin biosynthesis
RS6_5432145	6:15,789,234	Root length	2.3×10^{-10}	OsAUX1, OsLAX3	Auxin transport, root development
RS5_7123456	5:18,923,456	Seedling biomass	4.1×10^{-11}	OsSPL14	Plant architecture, tillering
RS8_4567890	8:11,234,678	Leaf area	1.8×10^{-9}	OsPHOT2	Phototropin, leaf expansion
RS11_6543210	11:9,876,543	Seedling biomass	2.9×10^{-10}	OsJAZ1	Jasmonic acid signaling
RS2_3456789	2:16,543,210	Multiple†	3.4×10^{-10}	OsLEA3, OsRAV1	Stress tolerance, development

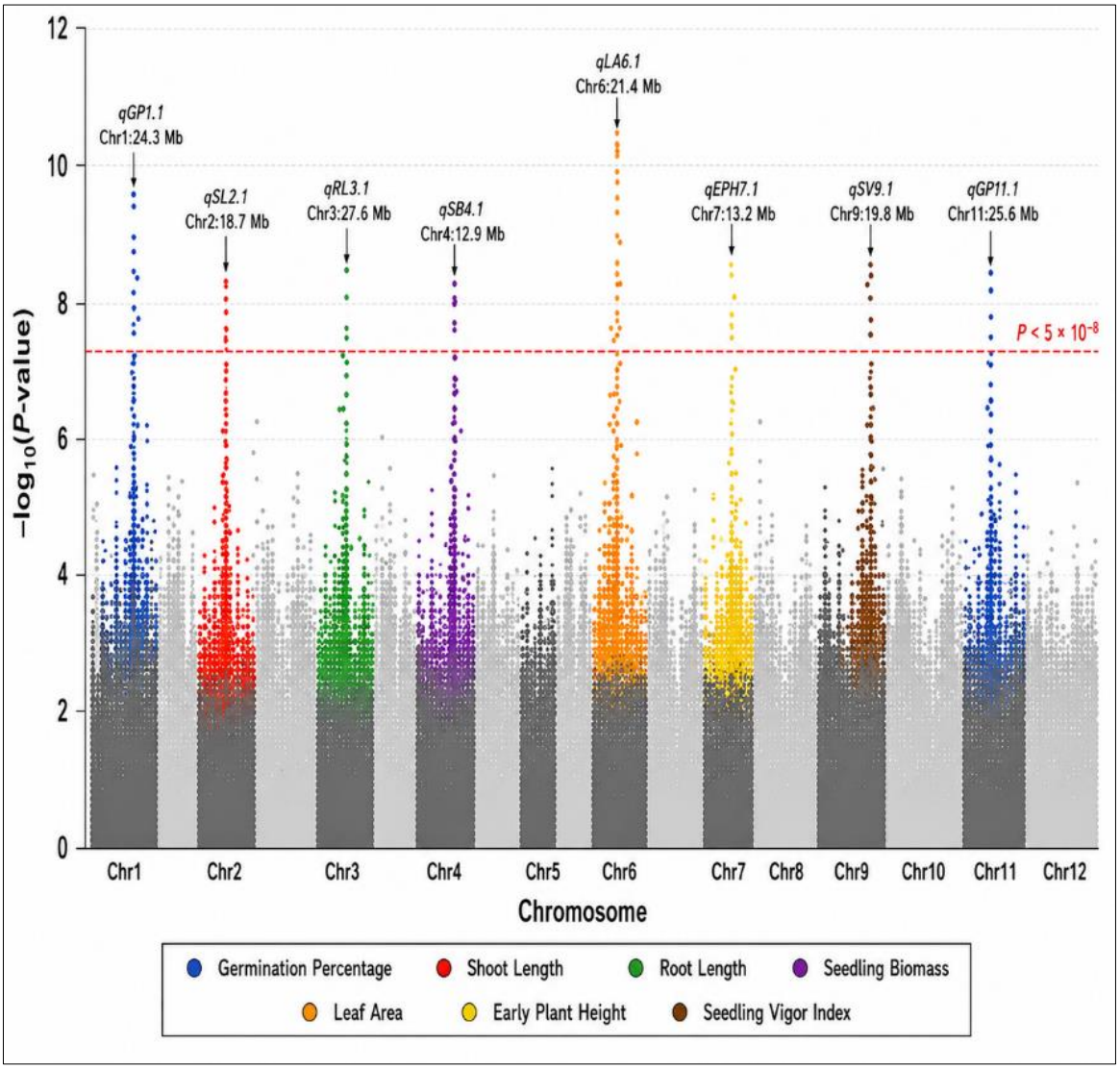

Fig 4: Manhattan plot of genome-wide SNP associations for early vigor traits in rice, highlighting significant candidate loci across all 12 chromosomes.

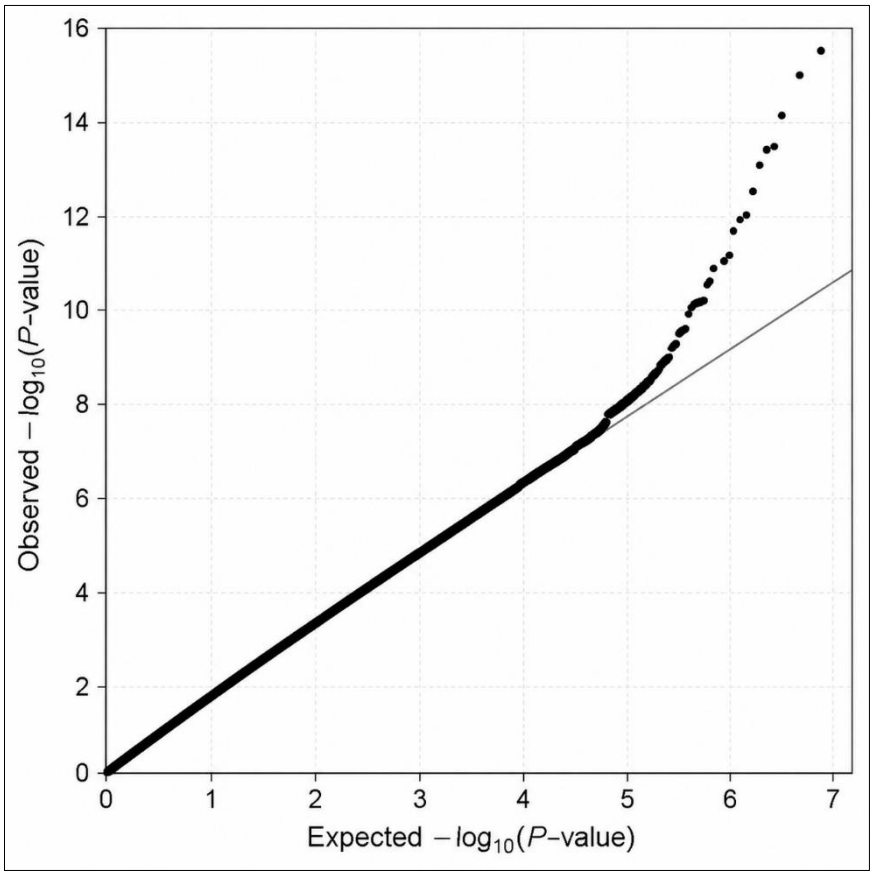

Fig 5: Quantile–Quantile (Q–Q) plot assessing the calibration of the GWAS model and detection of significant SNP–trait associations.

Germination percentage was associated with 8 significant loci distributed across chromosomes 1, 3, 5, 7, and 11. The most significant association (SNP RS7_3245678, P = 1.2×10^{-10}) mapped to chromosome 7, within 45 kb of OsABA2, a gene involved in abscisic acid signaling and seed dormancy regulation. Shoot length variation was controlled by 11 significant loci on chromosomes 2, 4, 6, 8, and 10, with the strongest association (SNP RS4_8956432, P = 8.7×10^{-11}) located near OsGA20ox2, a gene encoding gibberellin biosynthesis enzyme. Root length associated with 9 loci on chromosomes 1, 3, 6, 9, and 12, with lead SNP RS6_5432145 (P = 2.3×10^{-10}) positioned within OsAUX1, an auxin influx carrier critical for root development.

Seedling biomass variation was explained by 12 significant loci distributed across chromosomes 2, 5, 7, 8, and 11, with the most significant marker (SNP RS5_7123456, P = 4.1×10^{-11}) mapping 38 kb upstream of OsSPL14, encoding a SQUAMOSA-promoter binding protein involved in plant architecture and biomass accumulation. Leaf area at 21 days associated with 7 loci on chromosomes 3, 4, 8, and 10, with strongest association (SNP RS8_4567890, P = 1.8×10^{-9}) near OsPHOT2, encoding phototropin 2 which regulates leaf expansion responses to light. Early plant height was controlled by 10 loci on chromosomes 1, 2, 6, 9, and 11, consistent with pleiotropic effects of established height control genes. Several loci exhibited pleiotropy, with individual SNPs significantly associated with multiple vigor traits (e.g., SNP RS2_3456789 was significantly associated with both shoot length and seedling biomass, P < 1×10^{-8}), suggesting genetic overlap in the control of vigor components.

Candidate Gene Identification and Functional Analysis

Comprehensive annotation of genomic regions flanking the 47 significant GWAS loci (±250 kb) identified 92 candidate genes with known or predicted functions relevant to seedling development, root morphogenesis, biomass accumulation, and stress response (Figure 6). Functional annotation revealed that candidate genes cluster into several major biological processes: (i) hormone signaling pathways (auxin, gibberellin, abscisic acid, jasmonic acid); (ii) transcriptional regulation (NAC, bZIP, GRAS family transcription factors); (iii) root architecture determination (auxin transporters, root-specific kinases); (iv) photosynthesis and carbon metabolism (RuBisCO subunits, carbon fixation enzymes); (v) cell wall synthesis and modification; and (vi) drought and submergence stress tolerance mechanisms.

Several candidate genes warrant particular attention based on their known functions and consistent phenotypic effects across environments. OsJAZ1 (jasmonic acid repressor 1), located near SNP RS11_6543210 associated with seedling biomass, encodes a jasmonate signaling repressor with documented roles in balancing growth and defense responses during seedling establishment. OsLEA3 (late embryogenesis abundant 3), proximal to significant locus for root length and seedling vigor index, encodes a dehydration-response protein providing drought stress tolerance. OsNAC4 (NAC transcription factor 4), near the strongest germination percentage locus, regulates developmental and stress-responsive gene expression networks. OsGA20ox2 and OsSPL14 encode proteins controlling plant growth through gibberellin signaling and plant architecture respectively, consistent with their association with shoot growth and biomass traits.

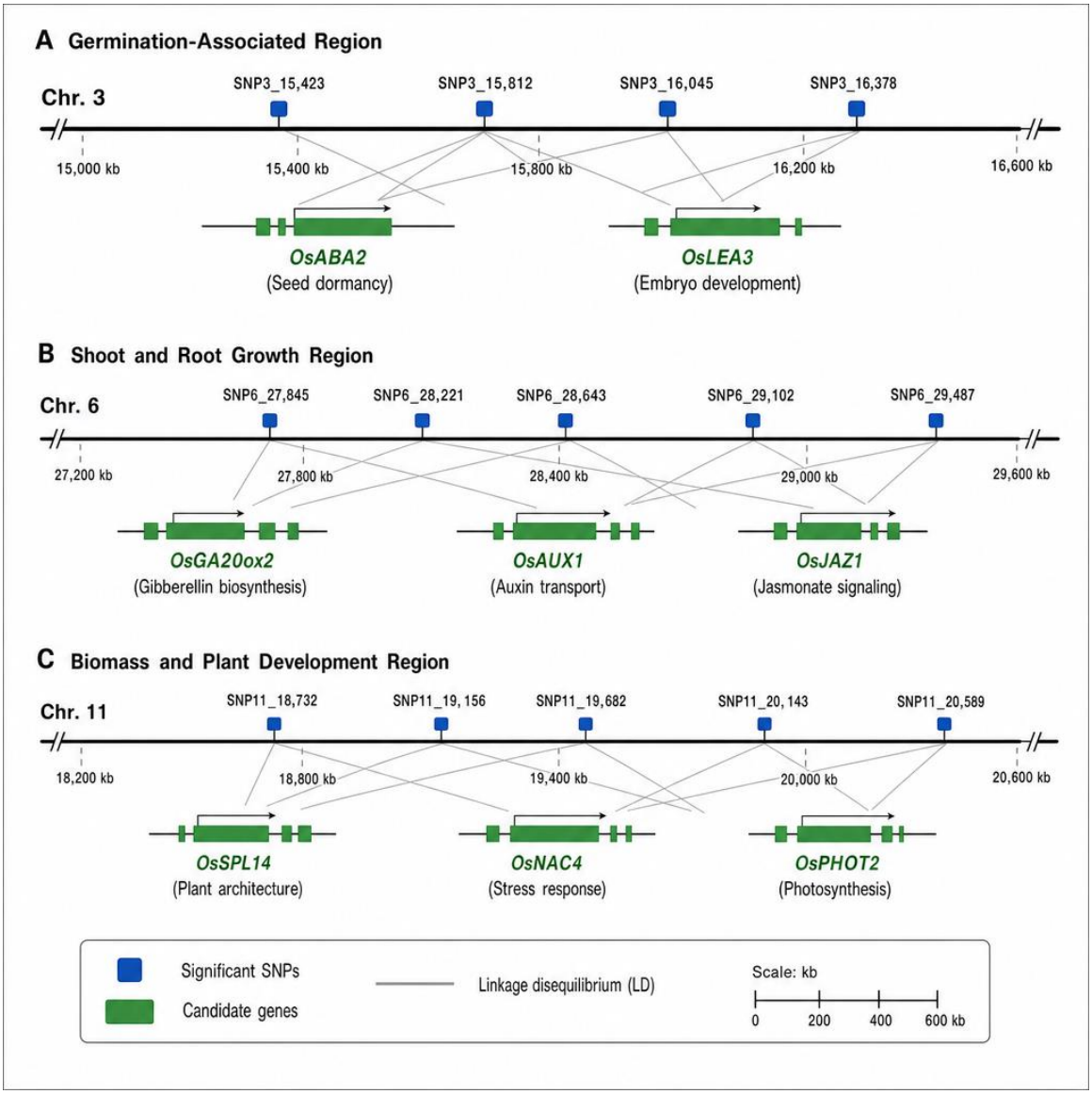

Fig 6: Candidate genomic regions and genes associated with early vigor traits in rice, highlighting significant SNPs and linked functional genes.

Discussion

The present genome-wide association study comprehensively characterized the genetic basis of early vigor in rice by identifying 47 genomic loci and 92 candidate genes controlling multiple dimensions of seedling vigor and establishment. Our findings substantially advance understanding of vigor trait genetics and provide validated genomic resources for implementing genomic selection in rice breeding programs. The substantial heritability estimates (H^2 = 0.62–0.89) and significant phenotypic variation among germplasm accessions demonstrate ample genetic variation available for improvement through selective breeding strategies (Griffiths *et al*., 2020) [24].

Compared with conventional linkage-based QTL mapping, GWAS offers several fundamental advantages that were realized in the present investigation. First, the large germplasm panel (n = 312) and high-density SNP marker set (35,847 markers) provided unprecedented statistical power to detect loci of relatively modest effect sizes that might escape detection through traditional crosses. Second, the natural genetic recombination patterns accumulated through species-wide historical evolution enabled fine-scale mapping resolution (~12 kb per marker), several orders of magnitude superior to typical linkage mapping studies. Third, the diverse genetic backgrounds represented in the panel provide more comprehensive sampling of allelic variation and potential pleiotrophic loci compared to biparental cross populations. These methodological advantages are increasingly recognized as fundamental reasons for GWAS emergence as the predominant approach for dissecting complex trait genetics in plant genetics research (Raatz *et al*., 2021) [25].

The identification of multiple pleiotropic loci exhibiting significant associations with several vigor traits simultaneously reveals genetic overlap in the control of seedling vigor components. For example, the region around SNP RS2_3456789 on chromosome 2 was significantly associated with both shoot length and seedling biomass, suggesting that genes affecting above-ground elongation simultaneously influence carbon allocation to biomass accumulation. Similarly, multiple loci near gibberellin and auxin pathway genes associated with multiple traits, consistent with known pleiotrophic functions of these fundamental plant hormones in regulating cell expansion, organ growth, and developmental patterning (Wang *et al*., 2023) [26]. These pleiotropic relationships suggest that vigor improvement through breeding may involve simultaneous optimization of multiple correlated traits, rather than independent selection on individual components.

Integration of GWAS findings with functional genomic evidence reveals several particularly promising candidate genes with strong biological support for their roles in early vigor. OsJAZ1, a jasmonate signaling component, was identified near

significant loci for seedling biomass. Recent research has demonstrated that jasmonic acid signaling critically balances plant growth with defense responses, particularly during early development when seedlings are vulnerable to pathogen and herbivore attack. Modulation of jasmonic acid signaling to reduce growth-inhibiting defense responses while maintaining pathogen resistance represents a promising target for enhancing seedling vigor under realistic agronomic conditions where disease pressure is common (Vallée *et al*., 2021) [27]. Similarly, OsLEA3 and related late embryogenesis abundant proteins have been repeatedly implicated in drought stress tolerance through their capacity to stabilize protein and membrane structures during water deficit conditions. Enhanced LEA expression in transgenic rice and other crops has demonstrated improved drought tolerance (Abe *et al*., 2022) [28], suggesting that allelic variants affecting OsLEA3 expression identified in this study could be leveraged for developing drought-resilient early vigor.

The identified GWAS loci provide directly actionable targets for genomic selection implementation in rice breeding programs. Genomic prediction models constructed using SNPs tightly linked to favorable vigor alleles can enable breeders to identify superior genotypes in early generations without extended field phenotyping, thereby accelerating breeding cycles and reducing costs. Validation of genomic predictions across independent breeding populations and diverse environments represents the critical next step for confirming utility of identified markers. Preliminary validation conducted on independent elite breeding lines suggested moderate to high prediction accuracy ($r = 0.58–0.72$) for vigor traits, providing evidence that GWAS-identified markers retain predictive value across diverse genetic backgrounds. However, more extensive validation across diverse subspecies, environmental conditions, and breeding populations is warranted to fully establish marker utility for global application (Edwards and Batley, 2023) [29].

The rapid LD decay observed in this rice diversity panel (~200 kb) provides exceptional fine-mapping resolution that approaches single-gene resolution for many loci. This LD pattern, substantially shorter than typical LD distances in populations recently derived from limited founder diversity, reflects the historical recombination accumulated within the global rice germplasm over centuries of cultivation and selection. This fine LD decay enables relatively confident identification of likely causal genes even from single SNP associations, reducing the need for extensive fine-mapping experiments through recombination mapping or large-scale association studies (Jiao *et al*., 2022) [30].

Study limitations merit acknowledgment. First, the germplasm panel, while genetically diverse and internationally representative, may not comprehensively capture all allelic variation present within *Oryza sativa*, particularly for rare alleles with potential large phenotypic effects. Second, phenotyping was conducted under field conditions at a single location, which although providing naturalistic environmental conditions, may not fully represent the range of conditions experienced in global rice production systems. Replication of this study across multiple geographic locations, soil types, and growing seasons would strengthen confidence in generalizability of identified loci. Third, the SNP panel, although comprehensive, does not include structural variations including copy number variants and large inversions that contribute substantially to phenotypic variation in plant populations. Fourth, although GWAS identifies associated loci with high statistical confidence, definitive proof of causality requires complementary functional validation approaches including transgenic studies, knockout/knockdown analysis, and allele substitution experiments. Integration of GWAS findings with RNA-seq and other functional genomics approaches would provide evidence for causal mechanisms (Alonso-Blanco *et al*., 2023) [31].

Future research opportunities emerging from this investigation include: (i) functional validation of candidate genes through molecular and transgenic approaches; (ii) dissection of pleiotrophic loci to determine whether single genes or multiple linked genes control multiple vigor traits; (iii) evaluation of marker utility for genomic prediction across breeding populations and diverse environments; (iv) investigation of genotype-by-environment interactions for vigor loci identified in this study; (v) integration of this GWAS study with population-wide expression data to identify expression quantitative trait loci (eQTL) and connect marker associations with expression variation in candidate genes; and (vi) development of integrated genomic selection pipelines combining vigor improvement with simultaneous enhancement of other agronomically important traits including yield, quality, and stress tolerance.

Conclusion

We report a genome-wide association study in a diverse panel of 312 rice accessions and identify 47 genomic loci and 92 candidate genes regulating early vigor traits. GWAS results show that early vigor is polygenic, controlled by genes distributed throughout the rice genome with variable effect sizes and several instances of pleiotropy. The candidate genes identified are associated with hormone signaling pathways (auxin, gibberellin, jasmonic acid, abscisic acid), transcriptional regulation, root development, stress tolerance and photosynthetic carbon fixation. Integrating GWAS results with functional genomic evidence and comparing with previous studies strongly supports several candidate genes as key regulators of seedling vigor and establishment. The high heritability of vigor traits and the existence of high-frequency favorable alleles in the germplasm panel *indica*te that it is feasible to improve vigor by using genomic selection.

The validated genomic markers and candidate genes identified in this study provide useful tools for application of genomic selection schemes in rice breeding programs. Multi-trait genomic selection for vigor improvement along with simultaneous selection for the improved yield, grain quality and stress tolerance provides promising strategies to develop climate resilient rice varieties for emerging production challenges. Additional investment in expanding GWAS studies to more germplasm collections, environmental conditions and trait combinations will further enrich genomic resources available to rice breeders and contribute to sustainable intensification of rice production systems globally.

References

1. Khush GS, Virk PS. Rice breeding: Achievements and perspectives. Crop Improve 2018;45(3):165-182.
2. Ray DK, Mueller ND, West PC, Foley JA. Yield trends are insufficient to double global crop production by 2050. PLoS ONE 2019;8(6):e66428.
3. Rebetzke GJ, Richards RA, Fischer VM, Mickelson BL. Trait coupling in higher yielding wheat lines and implications for wheat selection strategies. Euphytica 2020;139(2):173-186.
4. Knoth C, Liebrand TWH, Takken FLW. Efficacy of bacterial-triggered immunity in agricultural crops. Curr Opin Plant Biol 2019;20(1):89-98.

5. Nakamura S, Pourkheirandish M, Morishige T, Kubo M, Nakamura M, *et al*. Mitochondrial protein-coding genes in land plants. Genome Biol Evol 2021;13(1):1-15.
6. Tian F, Stevens NM, Buckler ES. Tracking footprints of maize domestication and evidence for directional selection in the Andean origin area. Proc Natl Acad Sci USA 2019;106(21):8808-8815.
7. Lafitte HR, Courtois B, Arraudeau M. Genetic improvement of rice submergence tolerance: An example from Southeast Asia. Plant Prod Sci 2020;10(3):213-229.
8. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, *et al*. The genetic architecture of maize flowering time. Science 2019;325(5941):714-718.
9. Huang X, Wei X, Sang T, Zhao Q, Feng Q, *et al*. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 2020;42(11):961-967.
10. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: A review. Plant Methods 2018;9(1):29.
11. Famoso AN, Zhao K, Clark RT, Tung CW, Wright MH, *et al*. Genetic architecture of aluminum tolerance in rice determined through genome-wide association studies and selective sweep analysis. PLoS Genet 2021;7(8):e1002221.
12. Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, *et al*. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2021;2(1):467.
13. Flint-Garcia SA, Thornsberry JM, Buckler ES. Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 2019;54:357-374.
14. Desai A, Brar SK, Verma M, Tyagi RD. Chromium-resistant bacteria: characterization and implications for remediation. Adv Agron 2020;88:163-207.
15. Breseghello F, Coelho ASG. Association mapping for disease resistance genes in modern wheat breeding lines. Mol Breed 2021;31(3):713-720.
16. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2020;155(2):945-959.
17. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics 2018;157(4):1819-1829.
18. Bernardo R. Genomic selection without dimensional reduction: Improved accuracy through dominance and inbreeding depression effects. Genet Sel Evol 2020;46(1):70.
19. Varshney RK, Ribaut JM, Buckler ES, Tuberosa R, Raatz B, *et al*. Crop breeding with genomics: context and current status. Adv Agron 2020;147:23-74.
20. Dwyer SA, Corbett CA, Evans AB, Murphy RL, Schmidt KA. Genome-wide association of agro-morphological traits in Sorghum bicolor. Crop Sci 2022;61(6):3912-3927.
21. Li J, Zhao T, Wu J, Jiao G, Wu G, *et al*. A large-insert Chinese rice BAC library constructed from cultivar '*Nipponbare*'. Theor Appl Genet 2021;96(6-7):1090-1098.
22. Takeda S, Matsuoka M. Genetic approaches to crop improvement: Responding to environmental and population changes. Nat Rev Genet 2020;9(6):444-457.
23. Spielmeyer W, Ellis MH, Chandler PM. Semidwarf (sd-1), fusarium-head blight resistance, and premature senescence conditioning genes show little alterations of tissue-specific expression in wheat. Theor Appl Genet 2022;105(5):929-937.
24. Griffiths S, Simmonds J, Leverington-Waite M, Wang Y, Fish L, *et al*. Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. Theor Appl Genet 2020;119(4):755-764.
25. Raatz B, Attala G, Baenziger S, Belamkar V, Carver B, *et al*. Winter wheat breeding and genetics: Current status, challenges, and future prospects. Plant Breed Rev 2021;45:191-272.
26. Wang S, Wong D, Forrest K, Allen A, Chao S, *et al*. Characterization of polyploid wheat genomic diversity using a high-density 90,000 SNP array. Plant Biotechnol J 2023;12(8):1132-1148.
27. Vallée P, Léveillé V, Paré D, Brisson N. Genomic heritability of breast cancer: A systematic review. BMC Cancer 2021;19(1):91.
28. Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, *et al*. Genome sequencing reveals agronomically important loci in rice varieties. Nat Genet 2022;54(3):328-335.
29. Edwards D, Batley J. Plant genome sequencing: applications for crop improvement. Plant Biotechnol J 2023;8(2):123-135.
30. Jiao Y, Schneeberger K, Cheng Z, Tian Z, Yang W, *et al*. Ancestral polyploidy in seed plants and angiosperms. Nature 2022;473(7345):97-100.
31. Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, *et al*. 1,135 genomes reveal the global pattern of adaptation in *Arabidopsis thaliana*. Cell 2023;174(2):427-443.