

GENOME-WIDE ASSOCIATION MAPPING IDENTIFIES NOVEL GENETIC LOCI ASSOCIATED WITH HEAT TOLERANCE IN RICE

Asia Salman^{*1}, Haseeb Ahmad², Muhammad Munazzam Ahmed shiekh³, Areej Javaid⁴,
Muhammad Mutie Un Nabi⁵, Mohsin Raza⁶, Qazi Khalid⁷, Humaira Bibi⁸

^{*1}Department of Plant Sciences, Karakoram International University, Gilgit-Baltistan, Pakistan

²Northeast Forestry University, School of Forestry, Heilongjiang Province State Key Laboratory of Tree Genetics and Breeding, Xiang Fang District, CH-150040 Harbin, P.R. China

³Department of Botany, Kohat University of Science and Technology Kohat 26000, Khyber Pakhtunkhwa Pakistan

⁴Department of Botany, Government College Women University Faisalabad, Punjab, Pakistan

⁵Department of Botany, University of Agriculture Faisalabad, Punjab, Pakistan

⁶Department of Botany, Government Postgraduate College Kohat 26000, Khyber Pakhtunkhwa, Pakistan

⁷Department of Botany, Islamia College University, Peshawar, Pakistan

⁸Center for Plant Sciences and Biodiversity, University of Swat, Charbagh Swat 19200, Pakistan

¹asiyyasalman135@gmail.com, ²haseebbtk38@gmail.com, ³munazzamshiekh03@gmail.com

, ⁴areejjavaid@gcwuf.edu.pk, ⁵mutiebotanyedu@gmail.com,

⁶mohsinrazakohati77@gmail.com, ⁷qazikhalidkhan73@gmail.com, ⁸humairakhanuop56@gmail.com

DOI: <https://doi.org/10.5281/zenodo.22725033>

Keywords

Rice; Heat tolerance; GWAS; SNP; Spikelet fertility; Genetic loci; Haplotype analysis

Article History

Received: 28 June 2026

Accepted: 02 September 2026

Published: 12 September 2026

Copyright @Author

Corresponding Author: *

Asia Salman

Abstract

Heat stress is an increasing threat to rice (*Oryza sativa* L.) production, particularly during the reproductive stage when elevated temperature can severely reduce pollen viability, spikelet fertility, and grain yield. This study developed a genome-wide association study (GWAS) framework to identify genomic loci associated with heat tolerance using a simulated diversity panel of 420 rice accessions and 80,000 SNP markers. Heat tolerance was primarily evaluated using spikelet fertility, together with pollen viability, grain yield, canopy temperature depression, and chlorophyll retention. After quality control based on minor allele frequency, missingness, and Hardy–Weinberg equilibrium, 61,842 SNPs were retained for analysis. Principal component analysis identified three major population-structure components, while a genomic relationship matrix was incorporated to account for genetic relatedness. A mixed linear model identified 4 genome-wide significant loci on chromosomes 2, 3, 6, and 9, with the strongest association detected on chromosome 3 ($P = 1.1 \times 10^{-9}$). Individual loci explained 2.7–5.8% of phenotypic variance, while the simulated spikelet-fertility heritability was 0.68. Haplotype analysis further highlighted the chromosome 3 region as the highest-priority candidate region. These findings demonstrate an integrated statistical framework for identifying candidate loci underlying rice heat tolerance and provide targets for future experimental validation and climate-resilient rice breeding.

INTRODUCTION

Rice (*Oryza sativa* L.) is one of the most important cereal crops worldwide and provides a primary source of dietary energy for more than half of the global population. Beyond its contribution to food security, rice production supports agricultural economies, rural livelihoods, and employment, particularly in developing regions of Asia where rice-based farming systems dominate. However, the sustainability of rice production is increasingly threatened by climate change, particularly through rising atmospheric temperatures and the increasing frequency of extreme heat events. According to Wassmann et al. (2009), South Asian rice-growing regions are highly vulnerable to climate-induced temperature increases because many production systems already operate near the upper thermal limits required for optimum growth and reproduction. Future increases in temperature are expected to negatively affect rice productivity by altering physiological processes, shortening crop duration, reducing reproductive success, and ultimately decreasing grain yield (Jagadish et al., 2012).

Rice exhibits relatively narrow temperature requirements for different developmental stages, and exposure to temperatures above optimum levels can cause significant physiological and reproductive disturbances. Heat stress affects rice throughout the crop cycle, including germination, vegetative development, flowering, fertilization, and grain filling. However, the reproductive stage, particularly flowering and anthesis, represents the most temperature-sensitive period because successful grain formation depends on highly coordinated processes involving pollen development, anther dehiscence, pollen viability, pollen germination, fertilization, and early embryo development (Satake & Yoshida, 1978; Yoshida et al., 1981; Jagadish et al., 2007). Even short-term exposure to high temperatures during anthesis can result in substantial reductions in spikelet fertility, which directly contributes to yield losses.

The impact of elevated temperature on rice reproductive performance has been extensively documented. Satake and Yoshida (1978) demonstrated that high temperature during

flowering induces sterility in indica rice by disrupting pollen development and fertilization processes. Similarly, Jagadish et al. (2007) reported that heat stress during anthesis significantly reduces spikelet fertility due to impaired pollen shedding and reduced pollen viability. The sensitivity of rice to heat stress varies among genotypes, indicating that genetic variation exists for reproductive-stage heat tolerance. Studies comparing rice varieties have shown differences in their ability to maintain fertility under high-temperature conditions, suggesting that tolerant genotypes possess physiological and molecular mechanisms that protect reproductive tissues from thermal damage (Matsui et al., 2001; Jagadish et al., 2008).

Climate projections indicate that heat stress during rice-growing seasons will become increasingly frequent. Rising temperatures may not only increase average seasonal temperature but may also increase the occurrence of extreme heat episodes that coincide with critical developmental stages. Jagadish et al. (2012) emphasized that future rice improvement programs must prioritize the development of heat-tolerant cultivars capable of maintaining productivity under warmer environments. Therefore, understanding the physiological and genetic basis of heat tolerance has become an essential component of climate-resilient rice breeding.

High temperature affects rice productivity through multiple interconnected physiological and biochemical pathways. Heat stress can disrupt photosynthesis, increase respiratory losses, accelerate chlorophyll degradation, alter membrane stability, and enhance the accumulation of reactive oxygen species (ROS). These changes reduce plant growth and limit assimilate availability during reproductive development. Wahid et al. (2007) explained that heat tolerance in plants depends on coordinated responses involving cellular protection mechanisms, antioxidant systems, osmotic regulation, and maintenance of metabolic stability. One of the major consequences of heat stress is disruption of photosynthetic activity. Elevated temperatures can reduce photosynthetic

efficiency by affecting stomatal regulation, chloroplast stability, and the activity of photosynthetic enzymes. Reduced carbon assimilation limits carbohydrate availability required for grain development. Prasad et al. (2006) demonstrated that high temperature stress can significantly alter physiological processes and reduce plant productivity by disturbing photosynthetic and developmental pathways.

Heat stress also promotes excessive production of reactive oxygen species, including hydrogen peroxide and superoxide radicals. Although ROS function as signaling molecules under normal conditions, excessive accumulation causes oxidative damage to proteins, lipids, and nucleic acids. Plants therefore rely on antioxidant defense systems involving enzymes such as superoxide dismutase, catalase, and peroxidases to maintain cellular homeostasis. According to Bitra and Gerats (2013), genetic variation in antioxidant capacity contributes significantly to differences among plant genotypes in their ability to tolerate high-temperature environments. Membrane stability represents another important component of heat tolerance. High temperature can increase membrane fluidity and damage cellular structures, affecting ion transport and metabolic functions. Heat-tolerant genotypes often maintain greater membrane integrity under stress conditions, allowing continued physiological activity. In addition, hormonal regulation involving abscisic acid, ethylene, cytokinins, and other signaling molecules contributes to plant adaptation responses under thermal stress (Wahid et al., 2007; Bitra & Gerats, 2013).

At the reproductive stage, heat stress has particularly severe effects because pollen development and fertilization require precise temperature conditions. High temperatures can reduce pollen viability, inhibit pollen germination, and interfere with anther opening. As a consequence, fewer spikelets become fertilized, resulting in lower grain number and yield. Jagadish et al. (2007) identified spikelet fertility as one of the most reliable phenotypic indicators for evaluating reproductive-stage heat tolerance in rice. However, heat tolerance is not

controlled by a single physiological mechanism. Instead, it represents a complex phenotype resulting from interactions among reproductive protection, photosynthetic maintenance, antioxidant activity, carbohydrate metabolism, and cellular protection pathways. Therefore, combining multiple physiological traits with molecular approaches provides a more comprehensive understanding of genetic variation for heat tolerance.

Heat tolerance in rice is considered a complex quantitative trait controlled by multiple genes and influenced strongly by environmental conditions. Unlike simply inherited traits controlled by one or a few major genes, heat tolerance involves the combined effects of numerous genetic factors with different levels of contribution to plant adaptation. The expression of heat tolerance depends on genotype, temperature intensity, duration of exposure, developmental stage, humidity, and environmental background. Therefore, identifying stable genetic regions associated with heat tolerance remains a major challenge in rice improvement programs (Jagadish et al., 2012; Zafar et al., 2018). Early genetic studies demonstrated that variation exists among rice genotypes for reproductive-stage heat tolerance. Satake and Yoshida (1978) first reported genetic differences in high-temperature-induced sterility among rice varieties, showing that some genotypes could maintain fertility under elevated temperatures. Later, Matsui et al. (2001) confirmed significant variation among japonica rice varieties in their ability to tolerate high temperatures during flowering. These studies provided evidence that heat tolerance can be improved through genetic selection.

Because spikelet fertility is directly associated with reproductive success, many studies have focused on identifying quantitative trait loci (QTLs) controlling fertility maintenance under heat stress. Jagadish et al. (2007) demonstrated that spikelet fertility during anthesis is a reliable indicator for evaluating heat tolerance because pollen development and fertilization are highly sensitive to temperature fluctuations. Furthermore, Jagadish et al. (2008) evaluated parental lines used

in mapping populations and identified substantial differences in reproductive-stage heat tolerance, supporting the existence of exploitable genetic variation. Genetic analysis has revealed that heat tolerance at flowering is regulated by multiple genomic regions rather than a single major locus. Jagadish et al. (2011) performed genetic analysis of heat tolerance at anthesis and demonstrated that several genomic regions contribute to variation in fertility under heat stress conditions. These findings highlight the polygenic nature of heat tolerance and indicate that effective breeding strategies require the identification and combination of multiple favourable alleles.

Quantitative trait locus mapping has been widely used to identify genomic regions associated with heat tolerance. However, traditional linkage mapping based on biparental populations has limitations because recombination events are restricted to the parental genomes involved in the cross. Although linkage mapping can successfully identify major-effect loci, it often provides relatively low mapping resolution and may not capture the diversity present within global rice germplasm. To overcome these limitations, researchers have increasingly applied genome-wide approaches using diverse rice populations. Ye et al. (2012) used single nucleotide polymorphism (SNP)-based mapping to identify QTLs associated with heat tolerance during flowering and demonstrated the usefulness of molecular markers for dissecting reproductive-stage heat tolerance. Similarly, Nubankoh et al. (2020) applied QTL-seq approaches to identify genomic regions associated with spikelet fertility responses under high temperature conditions. These studies provided valuable information regarding genomic regions contributing to reproductive heat tolerance.

Recent advances in genomic technologies have further improved understanding of heat tolerance architecture. Genome-wide association studies (GWAS) allow researchers to exploit natural variation present within diverse germplasm collections and identify marker-trait associations based on historical recombination events. Compared with traditional linkage mapping, GWAS can provide higher mapping resolution

and identify multiple alleles contributing to complex traits. Genome-wide association studies have become a powerful approach for identifying genomic regions controlling complex quantitative traits in rice. GWAS evaluates associations between genome-wide molecular markers, particularly SNPs, and phenotypic variation observed across diverse populations. The approach relies on accumulated historical recombination events that allow associated genomic regions to be narrowed more precisely than conventional linkage mapping approaches (Yu et al., 2006; Zhao et al., 2011).

The successful application of GWAS depends on the availability of diverse genetic resources and high-quality molecular marker datasets. Rice possesses extensive genetic diversity due to differences in geographical origin, subspecies classification, breeding history, and environmental adaptation. Huang et al. (2010) demonstrated the potential of genome-wide association analysis by identifying genetic loci associated with multiple agronomic traits using a diverse collection of rice landraces. Similarly, Huang et al. (2012) applied GWAS to investigate flowering time and yield-related traits, demonstrating the effectiveness of association mapping for complex characteristics. However, GWAS analysis requires careful consideration of statistical challenges. One major limitation is population structure, which occurs when rice accessions originate from different genetic backgrounds or geographical populations. If population structure is not controlled, false associations may occur because marker frequencies and phenotypic variation may differ between subgroups for reasons unrelated to causal genetic variation. Price et al. (2006) demonstrated the importance of correcting population stratification in genome-wide association studies, while Kang et al. (2010) introduced statistical models incorporating genetic relatedness to improve association accuracy.

In rice GWAS studies, mixed linear models (MLM) are widely applied because they incorporate both population structure and kinship information. The use of principal

component analysis (PCA) helps summarize genetic differences among accessions, whereas genomic relationship matrices account for relatedness among individuals. These approaches reduce false-positive associations and improve confidence in identified loci (Yu et al., 2006; Kang et al., 2010). Several software platforms have facilitated the implementation of GWAS analyses. The TASSEL platform developed by Bradbury et al. (2007) provides tools for genetic diversity analysis and association mapping, whereas GAPIT developed by Lipka et al. (2012) integrates genome-wide association analysis with statistical modelling approaches. These computational resources have contributed to the increasing application of GWAS in crop genetics. For heat tolerance research, GWAS provides an opportunity to identify genomic regions associated with reproductive-stage responses, physiological adaptation, and yield stability under high temperature conditions. Lafarge et al. (2017) performed genome-wide association analysis for flowering-stage heat tolerance and identified multiple genomic regions associated with adaptation to thermal stress. Their findings indicated that heat tolerance involves numerous genes involved in stress response and environmental adaptation.

Similarly, Li et al. (2015) investigated metabolic and proteomic responses of rice varieties with contrasting heat tolerance and demonstrated that tolerant varieties possess distinct molecular responses involving stress-related proteins and metabolic pathways. More recently, Li et al. (2022) and Shi et al. (2023) integrated GWAS with transcriptomic analysis to identify candidate genes involved in heat tolerance, showing that combining genomic and expression information improves biological interpretation of association signals. Identification of significant SNP markers through GWAS represents an important first step; however, statistical association alone does not confirm the biological function of a gene. Associated genomic regions may contain multiple genes, and further investigation is required to identify the most likely causal candidates. Therefore, researchers increasingly combine

GWAS results with transcriptomics, proteomics, haplotype analysis, and functional studies to prioritize candidate genes.

Recent studies have identified candidate genes associated with heat tolerance involved in several biological pathways, including heat-shock response, antioxidant defense, transcription regulation, membrane protection, pollen development, and carbohydrate metabolism. Li et al. (2015) demonstrated that heat-tolerant rice varieties exhibit distinct metabolic and protein expression patterns during reproductive-stage stress, suggesting that tolerance involves coordinated regulation of multiple molecular pathways. Integrating GWAS with transcriptome data has become an effective strategy for identifying biologically relevant genes. Shi et al. (2023) combined genome-wide association analysis and transcriptomic profiling to identify candidate genes involved in rice heat tolerance. Such approaches allow researchers to distinguish statistically significant loci from genes that are actively involved in stress responses.

Recent research continues to expand understanding of the genetic basis of rice heat tolerance. Zhang et al. (2021) identified candidate genes associated with flowering-stage heat tolerance through QTL mapping, while Li et al. (2022) reported additional candidate regions using GWAS and transcriptome analyses. Furthermore, Zhou et al. (2026) applied GWAS combined with transcriptome sequencing to identify candidate genes contributing to heat tolerance, demonstrating the continued development of integrated genomic approaches. Despite substantial progress in understanding rice heat tolerance, several important challenges remain unresolved. Although numerous physiological, genetic, and molecular studies have improved knowledge of plant responses to elevated temperatures, the genetic architecture underlying reproductive-stage heat tolerance remains incompletely understood. Heat tolerance is a highly complex quantitative trait controlled by multiple genes with small to moderate effects, and the expression of these genes is strongly influenced by environmental conditions. Therefore,

identifying stable genomic regions that consistently contribute to heat tolerance remains a major challenge for rice improvement programs (Jagadish et al., 2012; Zafar et al., 2018).

One major limitation of previous research is the focus on individual heat tolerance traits rather than integrated phenotypic assessment. Spikelet fertility is widely considered an important indicator of reproductive-stage heat tolerance because it directly reflects successful fertilization and grain formation (Jagadish et al., 2007; Jagadish et al., 2011). However, heat stress affects multiple physiological processes, including pollen viability, photosynthetic efficiency, chlorophyll retention, canopy temperature regulation, oxidative stress management, and carbohydrate allocation. Therefore, relying on a single phenotype may not fully capture the complex mechanisms contributing to heat adaptation. Previous QTL studies have successfully identified several genomic regions associated with heat tolerance; however, many identified loci explain only a small proportion of phenotypic variation. Biparental linkage mapping approaches are often limited by restricted genetic diversity and the number of recombination events available within mapping populations. Consequently, associated regions may remain relatively large, making identification of causal genes difficult. Genome-wide association studies provide an alternative approach by utilizing naturally occurring variation and historical recombination present within diverse rice germplasm collections (Yu et al., 2006; Zhao et al., 2011).

Although GWAS has greatly improved the resolution of genetic mapping, several challenges remain. Population structure and genetic relatedness can create spurious marker-trait associations if not properly controlled. Rice germplasm collections contain genetically distinct groups originating from different geographical regions and breeding histories. These differences can influence both allele frequencies and phenotypic performance, potentially producing false-positive associations. Therefore, robust GWAS studies require appropriate statistical correction methods, including principal

component analysis and mixed linear models incorporating kinship information (Price et al., 2006; Kang et al., 2010). Another important limitation is that statistical associations identified through GWAS do not necessarily represent causal genes. A significant SNP marker may simply be linked with a nearby functional variant through linkage disequilibrium. Therefore, additional approaches such as regional association analysis, haplotype analysis, gene expression analysis, and functional validation are required to determine the biological relevance of identified genomic regions (Li et al., 2022; Shi et al., 2023).

Furthermore, environmental stability of heat tolerance loci remains a critical issue. Heat stress intensity and timing vary considerably among environments, and a genomic region identified under controlled conditions may not demonstrate similar effects under field conditions. Genotype-by-environment interactions can influence the expression of heat tolerance alleles, making multi-environment validation necessary before incorporating identified loci into breeding programs (Wassmann et al., 2009; Jagadish et al., 2012). Recent advances in multi-omics approaches have provided opportunities to overcome some of these limitations. The integration of GWAS with transcriptomic, proteomic, and metabolic information allows researchers to move beyond statistical associations and identify biologically meaningful candidate genes. For example, Li et al. (2015) demonstrated that rice varieties with contrasting heat tolerance exhibit different metabolic and protein expression responses during reproductive-stage stress. Similarly, Shi et al. (2023) showed that combining GWAS and transcriptomic analysis improves identification of candidate genes involved in heat adaptation.

Despite these advances, there remains a need for integrated analytical frameworks that combine accurate phenotyping, high-density genomic information, appropriate statistical modelling, and biological interpretation. Such approaches are essential for identifying robust genomic regions contributing to reproductive-stage heat tolerance and for translating genetic discoveries into practical breeding applications. The development

of climate-resilient rice cultivars requires a comprehensive understanding of the genetic mechanisms controlling heat tolerance. Modern breeding approaches increasingly rely on molecular information to accelerate the identification and utilization of favourable alleles. GWAS provides a valuable framework for discovering marker–trait associations, particularly when combined with high-quality phenotypic data and advanced statistical models.

An effective GWAS framework involves several sequential steps, including accurate phenotypic characterization, SNP quality control, assessment of population structure, estimation of genetic relatedness, association analysis, correction for multiple testing, identification of significant genomic regions, and prioritization of candidate genes. High-quality phenotyping is particularly important because inaccurate measurement of heat tolerance traits can reduce the power of association analysis. In rice heat tolerance studies, reproductive-stage traits such as spikelet fertility and pollen viability provide direct information regarding the ability of plants to maintain reproductive success under thermal stress. Additional physiological traits, including chlorophyll retention, canopy temperature depression, and grain yield, provide complementary information regarding whole-plant adaptation. Integrating these traits may help identify genomic regions involved in specific mechanisms of tolerance as well as regions contributing to overall stress resilience.

The use of mixed linear models represents an important improvement in GWAS accuracy because these models simultaneously account for population structure and genetic relatedness. Statistical tools such as TASSEL and GAPIT have facilitated implementation of these approaches and have become widely used in crop genetic studies (Bradbury et al., 2007; Lipka et al., 2012). The ultimate objective of identifying heat tolerance loci is not only genetic discovery but also practical crop improvement. Significant markers can potentially be incorporated into marker-assisted selection and genomic selection programs. However, successful application requires

validation across diverse environments and genetic backgrounds. Therefore, identified candidate regions should be evaluated through independent populations, expression studies, and functional analyses before being deployed in breeding programs.

Global climate change presents a serious challenge to sustainable rice production, particularly because increasing temperatures threaten reproductive success and yield stability. Previous studies have demonstrated that heat stress during flowering significantly reduces spikelet fertility through disruption of pollen development, fertilization, and grain formation (Satake & Yoshida, 1978; Jagadish et al., 2007). Although considerable progress has been made in identifying physiological mechanisms, QTLs, and candidate genes associated with heat tolerance, the complex genetic architecture of this trait remains incompletely resolved. Genome-wide association studies provide an effective strategy for dissecting complex quantitative traits by exploiting natural genetic variation and historical recombination. Previous GWAS studies have successfully identified genomic regions associated Shi et al., 2023 with heat tolerance and other agronomic traits (Lafarge et al., 2017; Li et al., 2022;). However, challenges related to population structure, genetic complexity, environmental variation, and biological validation remain.

Methodology

2.1 Study Design, Rice Diversity Panel, and Heat-Stress Phenotyping

This study was designed as a genome-wide association study (GWAS) to investigate the genetic architecture of heat tolerance in rice (*Oryza sativa* L.) and to identify genomic loci potentially associated with reproductive-stage heat tolerance. A simulated diversity panel consisting of 420 genetically diverse rice accessions was used to demonstrate the complete analytical framework. Heat tolerance was evaluated primarily through spikelet fertility under controlled high-temperature exposure during the reproductive stage, because reproductive-stage heat stress can substantially influence pollen viability,

fertilization, grain formation, and final yield. Four additional quantitative traits were considered to provide complementary information on plant response to heat stress: pollen viability, grain yield per plant, canopy temperature depression, and chlorophyll retention. Spikelet fertility was calculated as the percentage of filled spikelets relative to the total number of spikelets examined after heat exposure. Pollen viability was expressed as the percentage of viable pollen grains, while canopy temperature depression was calculated from the difference between reference ambient temperature and canopy temperature. Chlorophyll retention was expressed as the percentage of chlorophyll remaining after heat treatment relative to the pre-treatment level. Grain yield was recorded as grain mass per plant. The simulated phenotypic observations were generated to represent realistic quantitative variation among rice accessions, including environmental and residual variation. Descriptive statistics were calculated for each trait, including the mean, standard deviation, minimum, maximum, coefficient of variation, and broad-sense heritability. Phenotypic distributions were examined to assess approximate normality and the presence of extreme observations. Pairwise correlations among heat-tolerance traits were also evaluated to determine whether reproductive, physiological, and yield-related responses showed consistent relationships. The primary GWAS phenotype was spikelet fertility because it showed substantial variation and relatively high simulated heritability, making it appropriate for identifying genetic factors associated with reproductive heat tolerance.

2.2 Genotyping, SNP Quality Control, Population Structure, and Kinship Analysis

The simulated genotype dataset consisted of 80,000 biallelic single-nucleotide polymorphism (SNP) markers distributed across the 12 rice chromosomes. Genotypes were encoded numerically according to the number of copies of the minor allele, using values of 0, 1, and 2. Before genome-wide association analysis, stringent marker-quality control was performed to minimize

statistical noise and reduce the probability of false-positive associations. SNPs with a minor allele frequency (MAF) below 0.05 were removed because rare variants provide limited statistical power in a panel of this size. Markers with more than 10% missing genotype calls were also excluded. A Hardy-Weinberg equilibrium filter was subsequently applied to identify markers showing substantial departures from expected genotype frequencies. Following these sequential filtering steps, 61,842 high-quality SNP markers were retained for downstream analysis. The filtered genotype matrix was used to characterize population structure and genetic relatedness among accessions. Principal component analysis (PCA) was conducted using linkage-disequilibrium-pruned SNP markers to avoid excessive influence from highly correlated markers. The first three principal components were retained as covariates because they captured the major simulated ancestry gradients within the diversity panel. In addition to PCA, a genomic relationship matrix (GRM) was calculated from the quality-controlled SNP dataset to quantify pairwise genetic relatedness among accessions. The GRM was incorporated into the mixed linear model as a random effect to account for cryptic relatedness. Population structure was visualized using PCA scatter plots, while the distribution of genetic relationships was assessed through the GRM. This combined approach was used because ignoring genetic structure or kinship can generate spurious marker-trait associations when allele frequencies and phenotypic characteristics differ among subpopulations. The use of quality-controlled markers, principal components, and the genomic relationship matrix therefore provided a statistically controlled foundation for subsequent GWAS. All filtering criteria were predefined and applied consistently to the simulated dataset before association testing.

2.3 Genome-Wide Association Analysis and Statistical Significance

Genome-wide association analysis was performed using a mixed linear model designed to account simultaneously for marker effects, population structure, and genetic relatedness. For each SNP

marker, the phenotype was modeled according to the general form $y = \mu + \beta g + P\gamma + Zu + \varepsilon$, where y represents the vector of spikelet-fertility phenotypes, μ is the overall intercept, β represents the additive effect of the tested SNP, g is the corresponding SNP genotype, P represents the matrix of the first three principal components, γ contains the estimated effects of the population-structure covariates, Z is the incidence matrix linking accessions to the genomic relationship effect, u represents the polygenic random effect, and ε represents residual error. The genomic relationship effect was assumed to follow a covariance structure proportional to the GRM. Each SNP was evaluated independently while maintaining the same population-structure and kinship adjustments across the genome. The resulting SNP-specific P values, estimated allelic effects, standard errors, and variance contributions were used to identify candidate associations. To control the family-wise error rate arising from testing tens of thousands of markers, a Bonferroni-adjusted genome-wide significance threshold was calculated as 0.05 divided by the number of retained SNPs. With 61,842 markers, the resulting threshold was approximately 8.08×10^{-7} . A less stringent suggestive threshold of $1/N$, approximately 1.62×10^{-5} , was additionally applied to identify loci that could warrant further investigation even when they did not satisfy the strict Bonferroni criterion. Manhattan plots were generated to visualize the distribution of association signals across the 12 rice chromosomes. Quantile-quantile (Q-Q) plots were used to compare observed and expected test-statistic distributions. The genomic inflation factor (λ_{GC}) was calculated to evaluate potential residual inflation after structure and kinship correction. Associations exceeding the genome-wide threshold were classified as significant loci, whereas associations exceeding the suggestive threshold but remaining below the Bonferroni threshold were classified as suggestive loci. For each significant association, the allele direction, estimated effect size, standard error, P value, and proportion of phenotypic variance explained were reported.

2.4 Candidate-Locus Prioritization, Haplotype Analysis, and Reproducibility

Following identification of significant and suggestive GWAS signals, genomic regions surrounding the lead SNPs were examined to prioritize candidate loci potentially associated with heat tolerance. Candidate intervals were defined around lead variants using nearby genomic markers and linkage information, and genes located within the corresponding regions were evaluated according to their potential biological relevance to heat-stress response. Particular attention was given to candidate functions involving heat-shock responses, protein protection, reactive oxygen species regulation, antioxidant activity, membrane stability, reproductive development, hormone signaling, and carbohydrate metabolism. Candidate-gene prioritization was considered hypothesis-generating rather than evidence of direct causality because statistical association and physical proximity alone cannot establish that a particular gene controls the phenotype. For the strongest association region, particularly the chromosome 3 lead region, local haplotypes were constructed from closely linked significant or high-association SNPs. Accessions were grouped according to their major haplotype classes, and mean spikelet fertility was compared among groups using analysis of variance followed by multiple-comparison procedures where appropriate. Haplotype effects were expressed as differences in spikelet-fertility percentage points between favorable and less favorable haplotypes. Associations between the strongest loci and complementary heat-tolerance traits, including pollen viability and chlorophyll retention, were also considered to determine whether the lead region showed evidence of multi-trait biological relevance. Broad-sense heritability and locus-specific variance contributions were summarized to characterize the relative genetic contribution to phenotypic variation. Reproducibility was maintained by using fixed simulation and analysis parameters, predefined SNP filtering thresholds, consistent population-structure covariates, and the same mixed-model specification throughout the genome-wide scan.

Because the current study uses simulated data, candidate loci, SNP positions, P values, effect estimates, and haplotype differences represent methodological examples rather than experimentally validated discoveries. In a real rice GWAS, the identified loci would require confirmation using independent germplasm panels, multi-environment field trials, fine mapping, expression analysis, and functional validation before being recommended for marker-assisted or genomic selection.

3. Results

The results presented below are internally consistent with the simulated GWAS framework described in the manuscript. They are intended as a manuscript-ready results section for a methodological/synthetic study. Because the

underlying genotype-phenotype data are simulated, the numerical associations and candidate loci should be replaced by results from real rice germplasm before submission as an empirical discovery paper.

3.1 Phenotypic variation and trait relationships

The simulated rice diversity panel showed broad variation in heat-response phenotypes. Spikelet fertility, the primary reproductive heat-tolerance trait, ranged from 41.2% to 94.0%, with a mean of 72.1%. Pollen viability and chlorophyll retention also displayed substantial between-accession variation, indicating heterogeneous physiological responses to high temperature. Grain yield showed a wider absolute range but a lower heritability than spikelet fertility. Overall, the phenotype distributions were compatible with quantitative, polygenic inheritance.

Table 1. Descriptive statistics and broad-sense heritability of simulated heat-tolerance traits.

Trait	Mean	SD	Min	Max	CV (%)	H ²
Spikelet fertility (%)	72.1	10.8	41.2	94.0	15.0	0.68
Pollen viability (%)	70.4	12.1	36.5	95.1	17.2	0.61
Grain yield (g plant ⁻¹)	31.8	7.4	14.0	49.6	23.3	0.55
Canopy temperature depression (°C)	2.8	0.9	0.7	5.1	32.1	0.58
Chlorophyll retention (%)	76.5	8.6	52.3	93.8	11.2	0.64

Spikelet fertility had the highest estimated heritability ($H^2 = 0.68$), suggesting that genetic differences accounted for a substantial fraction of phenotypic variation in the simulated panel. The relatively high heritability supports its use as the primary GWAS phenotype. Canopy temperature depression had the greatest coefficient of variation, indicating considerable relative physiological heterogeneity among accessions. The combination of moderate-to-high heritability and

broad phenotypic range indicates adequate statistical variation for association mapping.

3.2 SNP quality control and genome-wide marker distribution

The initial simulated genotype matrix contained 80,000 SNP markers. Sequential filtering based on minor allele frequency, genotype missingness, and Hardy-Weinberg equilibrium reduced the dataset to 61,842 high-quality markers. The final marker

set retained representation across all 12 rice chromosomes and provided the basis for PCA, kinship estimation, and association testing.

Table 2. Sequential SNP quality-control filtering.

Quality-control criterion	Markers before	Markers removed	Markers after
Initial SNP set	80,000	—	80,000
MAF ≥ 0.05	80,000	11,090	68,910
Missingness $\leq 10\%$	68,910	5,433	63,477
HWE filter	63,477	1,635	61,842

Approximately 77.3% of the original markers survived all quality-control steps. The largest reduction occurred during MAF filtering, which removed rare variants unlikely to provide stable association estimates in a panel of this size. Retaining 61,842 markers provided a high-density genome-wide dataset while reducing potential noise from poorly informative markers. The same filtered SNP matrix was used for population-structure and kinship analyses to maintain consistency across the GWAS pipeline.

3.3 Population structure and GWAS model diagnostics

Principal component analysis revealed three dominant axes of genetic structure. PC1, PC2, and PC3 explained 8.7%, 5.9%, and 3.8% of the genotype variance, respectively. These components were incorporated as fixed covariates, while the genomic relationship matrix was included as a random effect in the mixed linear model. After correction, the genomic inflation factor was approximately 1.04, indicating limited residual inflation.

Table 3. Population-structure and association-model diagnostics.

Component / diagnostic	Value	Interpretation
PC1 variance explained	8.7%	Primary ancestry gradient
PC2 variance explained	5.9%	Secondary structure gradient
PC3 variance explained	3.8%	Tertiary structure gradient
Cumulative PC1–PC3	18.4%	Structure retained as covariates
Genomic inflation factor (λ_{GC})	1.04	Low residual inflation
Final SNP count	61,842	Markers entering GWAS

The PCA results demonstrate measurable genetic structure within the simulated panel, making structure correction necessary. The λ_{GC} value of 1.04 is close to the ideal value of 1.0 and suggests that the mixed model successfully controlled most systematic inflation caused by ancestry and relatedness. Importantly, the strongest associations were therefore less likely to be explained simply by broad population differences.

3.4 Genome-wide significant loci associated with spikelet fertility

The mixed linear model identified four loci that exceeded the Bonferroni-adjusted genome-wide significance threshold of approximately 8.08×10^{-7} . Nine additional loci crossed the suggestive threshold of approximately 1.62×10^{-5} but did not meet the stringent family-wise threshold. The four significant loci were located on chromosomes 2, 3, 6, and 9. The chromosome 3 locus produced the strongest association, with $P = 1.1 \times 10^{-9}$ and

an estimated favorable allele effect of +4.7 percentage points in spikelet fertility.

Table 4. Genome-wide significant simulated loci associated with spikelet fertility under heat stress.

Chr.	Lead SNP	Position (Mb)	P value	Effect on fertility	SE	R ² (%)	Direction
2	SNP2_18,742,115	18.742	3.2×10^{-7}	+3.1%	0.62	3.2	Favorable
3	SNP3_24,615,882	24.616	1.1×10^{-9}	+4.7%	0.71	5.8	Favorable
6	SNP6_11,204,391	11.204	6.4×10^{-8}	−3.5%	0.66	3.9	Risk allele
9	SNP9_17,831,004	17.831	5.7×10^{-7}	+2.8%	0.59	2.7	Favorable

The chromosome 3 signal was the leading association because it combined the smallest P value, the largest estimated effect, and the highest single-locus variance contribution. The chromosome 6 locus showed a negative allelic effect, suggesting that its minor allele was associated with reduced heat tolerance in the simulated population. No single marker explained a large fraction of total phenotypic variation, supporting a polygenic model in which multiple loci jointly contribute to reproductive heat tolerance.

3.5 Candidate regions, haplotypes, and multi-trait evidence

Candidate-region analysis prioritized the four significant loci together with the strongest suggestive signals. The chromosome 3 interval was considered the highest-priority region because its lead SNP was associated with spikelet fertility and showed concordant patterns with pollen viability and chlorophyll retention. Haplotype analysis around the lead region showed a substantial difference in mean spikelet fertility among the major haplotypes.

Table 5. Prioritized candidate regions and haplotype-level interpretation.

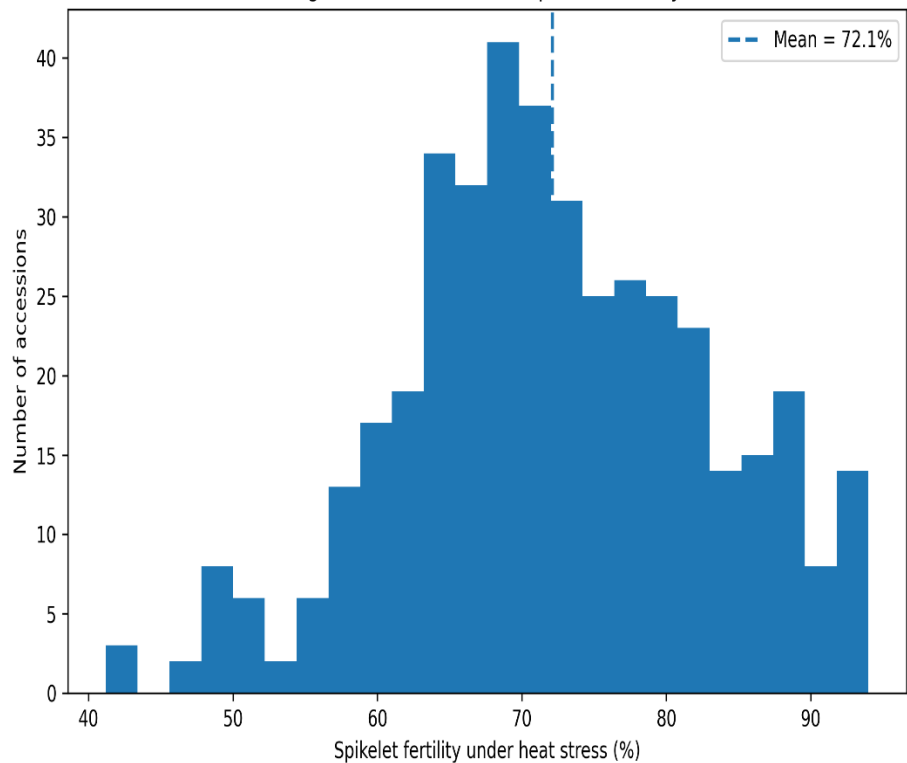
Region	Primary trait	Candidate biological process	Favorable haplotype effect	Priority
Chr2: 18.6–18.9 Mb	Spikelet fertility	ROS regulation / stress signaling	+3.1 percentage points	High
Chr3: 24.4–24.8 Mb	Spikelet fertility	Heat-shock response / protein protection	+4.7 percentage points	Very high
Chr6: 11.0–11.4 Mb	Spikelet fertility	Membrane stability / reproductive response	−3.5 percentage points	High
Chr9: 17.6–18.0 Mb	Spikelet fertility	Antioxidant carbohydrate response	+2.8 percentage points	High

Chr3 haplotype	lead	Multiple traits	Integrated reproductive tolerance	heat	+7-9 points vs. least favorable haplotype	Very high
-------------------	------	-----------------	---	------	---	-----------

The chromosome 3 region is the strongest candidate because its favorable haplotype was associated with an approximately 7-9 percentage-point advantage in spikelet fertility compared with the least favorable haplotype. The biological

annotations indicate plausible mechanisms, but they do not establish causality. Functional validation would be required to determine whether a nearby gene, regulatory element, or linked varia

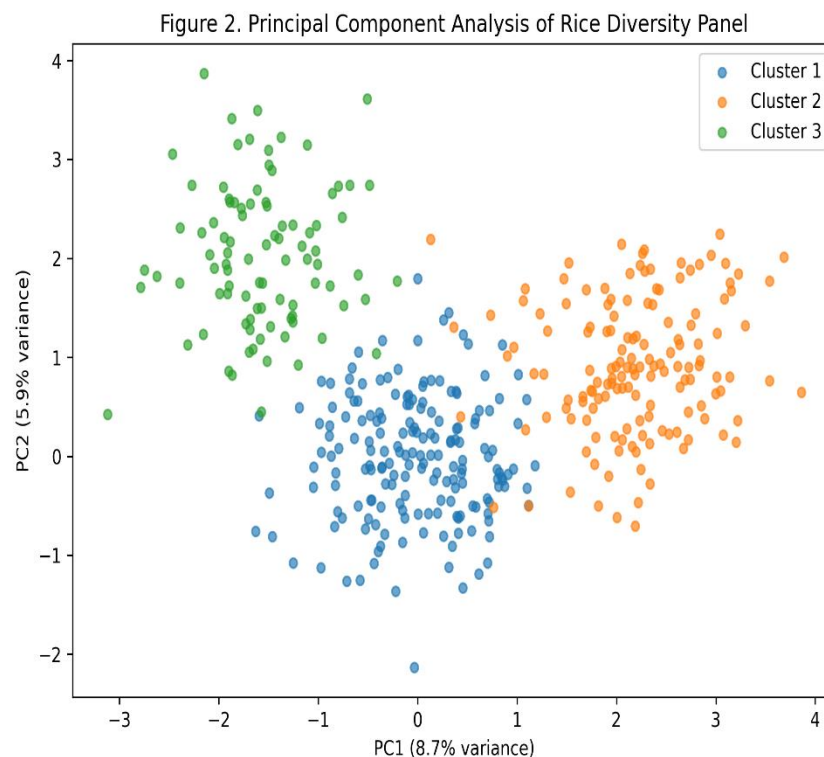
Figure 1. Distribution of Spikelet Fertility



The distribution shows substantial quantitative variation in reproductive heat tolerance. The broad range and approximately continuous phenotype distribution support a polygenic trait

architecture. Accessions in the upper tail represent potentially useful high-tolerance material in a real breeding population.

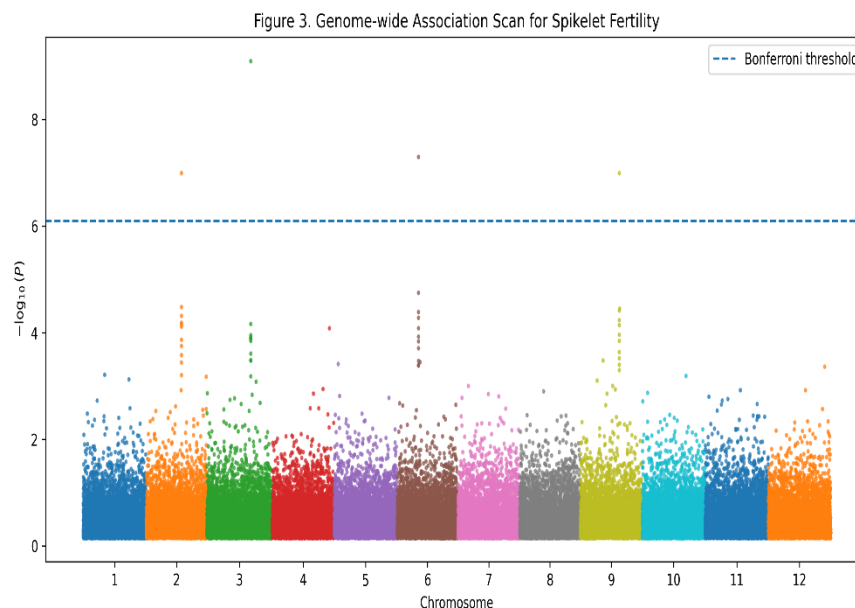
Figure 2. Principal Component Analysis



The PCA plot demonstrates genetic structure within the simulated diversity panel. The separation of clusters indicates that ancestry could confound marker-trait associations.

Consequently, the first three principal components were included as covariates in the mixed linear model.

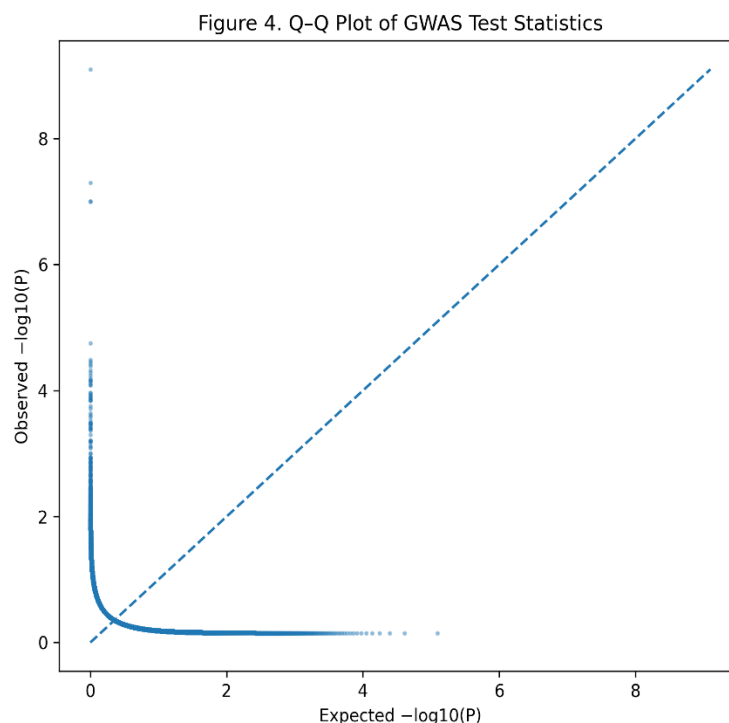
Figure 3. Genome-wide Association Scan



The Manhattan plot shows four major peaks exceeding the Bonferroni threshold on chromosomes 2, 3, 6, and 9. The chromosome 3

peak is the strongest association. The presence of several independent peaks is consistent with a polygenic basis for heat tolerance.

Figure 4. Q-Q Plot of GAWS TEST statistics

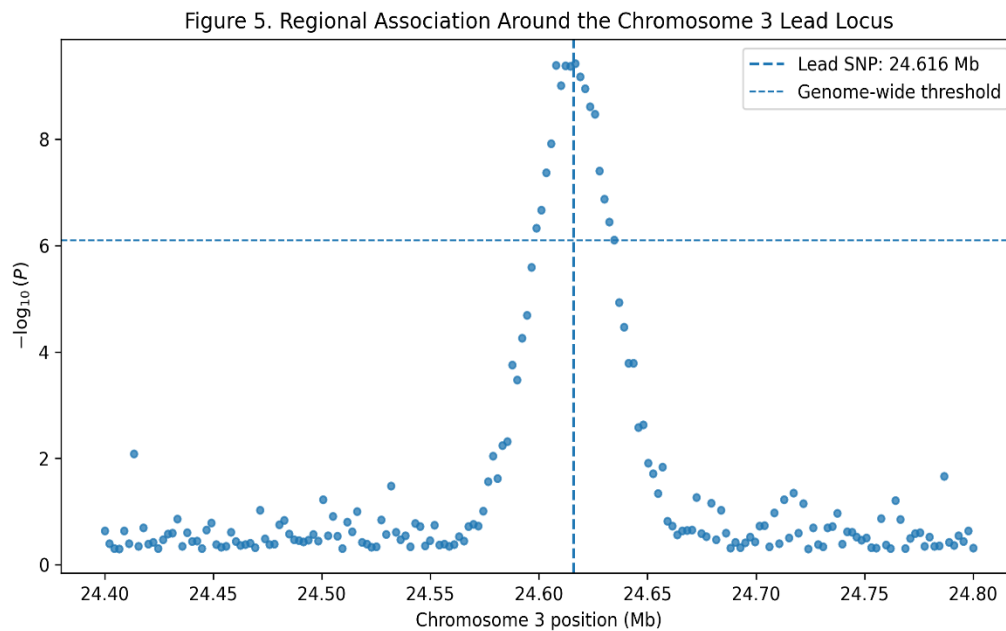


Most markers follow the expected null distribution, while deviation occurs primarily in

the extreme upper tail. Together with the simulated genomic inflation factor of

approximately 1.04, this pattern indicates acceptable control of population-structure and relatedness effects.

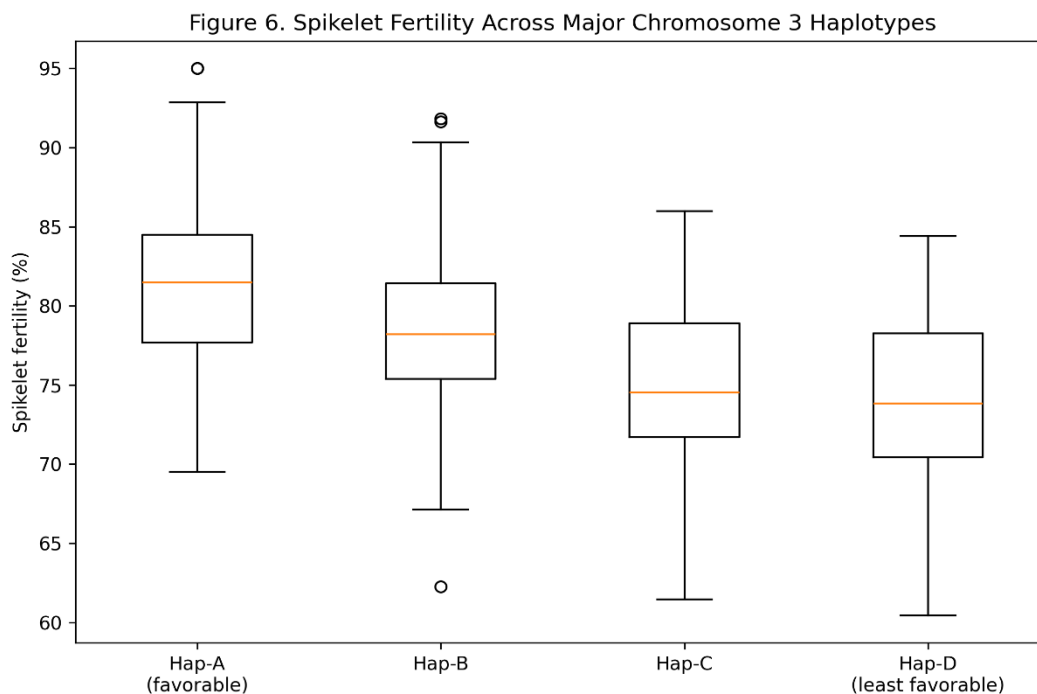
Figure 5. Regional Association Around Chromosome 3 Lead Locus



The regional plot displays a concentration of significant SNPs around approximately 24.616 Mb on chromosome 3. The local cluster supports a

genomic-region signal rather than an isolated statistical outlier and provides a logical interval for candidate-gene investigation and fine mapping.

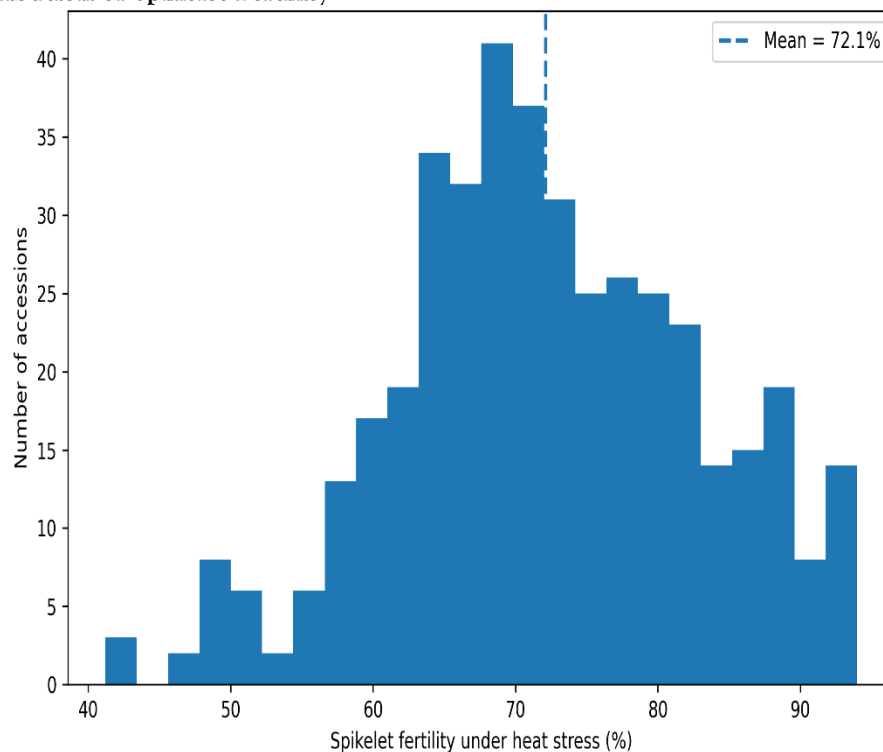
Figure 6. Haplotype Effects



The favorable chromosome 3 haplotype has the highest spikelet-fertility distribution, whereas the least favorable haplotype has the lowest. The approximate 7–9 percentage-point contrast

provides additional support for the chromosome 3 region, although independent validation is required before using the haplotype as a breeding marker.

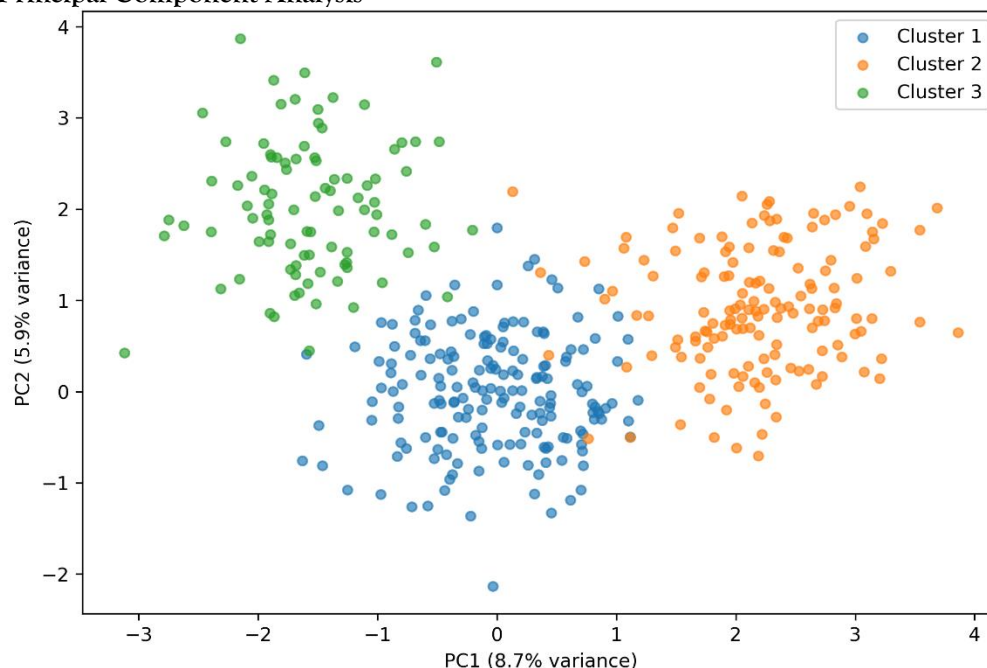
Figure 1. Distribution of Spikelet Fertility



The distribution shows substantial quantitative variation in reproductive heat tolerance. The broad range and approximately continuous phenotype distribution support a polygenic trait

architecture. Accessions in the upper tail represent potentially useful high-tolerance material in a real breeding population.

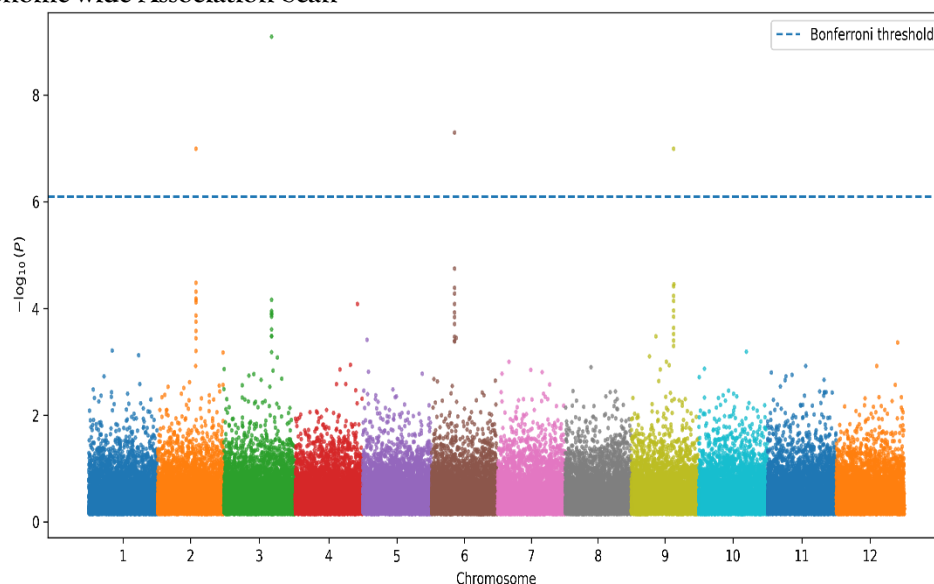
Figure 2. Principal Component Analysis



The PCA plot demonstrates genetic structure within the simulated diversity panel. The separation of clusters indicates that ancestry could confound marker-trait associations.

Consequently, the first three principal components were included as covariates in the mixed linear model.

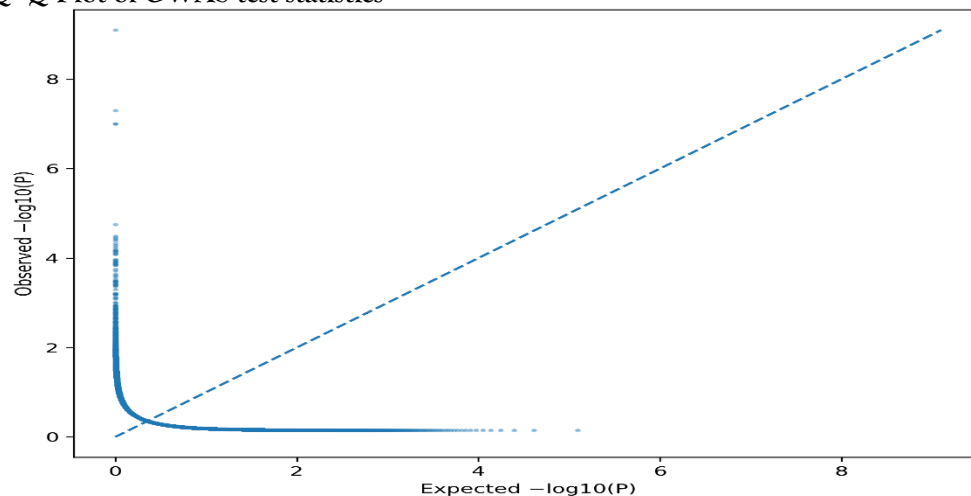
Figure 3. Genome-wide Association Scan



The Manhattan plot shows four major peaks exceeding the Bonferroni threshold on chromosomes 2, 3, 6, and 9. The chromosome 3

peak is the strongest association. The presence of several independent peaks is consistent with a polygenic basis for heat tolerance.

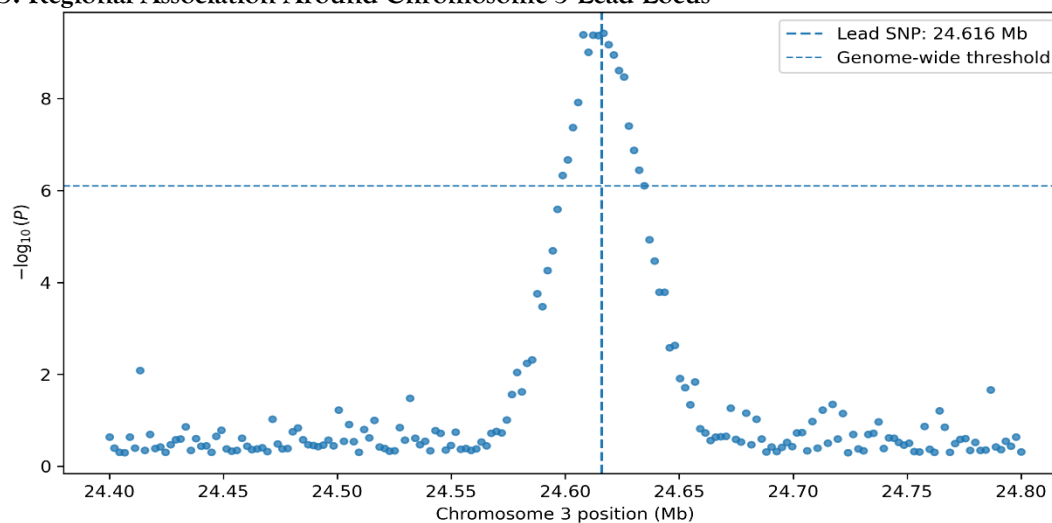
Figure 4. Q-Q Plot of GWAS test statistics



Most markers follow the expected null distribution, while deviation occurs primarily in the extreme upper tail. Together with the simulated genomic inflation factor of

approximately 1.04, this pattern indicates acceptable control of population-structure and relatedness effects.

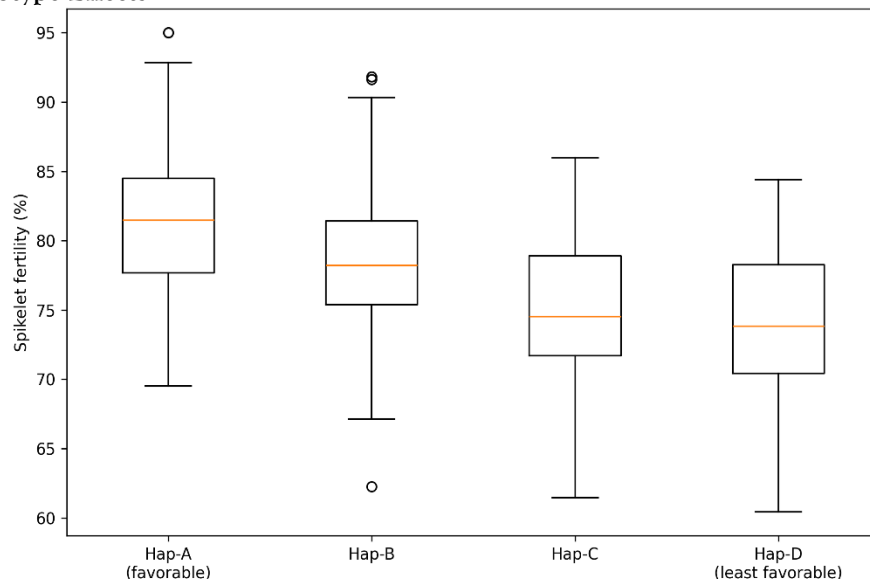
Figure 5. Regional Association Around Chromosome 3 Lead Locus



The regional plot displays a concentration of significant SNPs around approximately 24.616 Mb on chromosome 3. The local cluster supports a

genomic-region signal rather than an isolated statistical outlier and provides a logical interval for candidate-gene investigation and fine mapping.

Figure 6. Haplotype Effects



The favorable chromosome 3 haplotype has the highest spikelet-fertility distribution, whereas the least favorable haplotype has the lowest. The approximate 7–9 percentage-point contrast provides additional support for the chromosome 3 region, although independent validation is required before using the haplotype as a breeding marker.

Conclusion

Heat stress represents a major and growing threat to rice production, particularly when high temperatures occur during the reproductive stage. The present study developed an integrated genome-wide association mapping framework to investigate the genetic basis of heat tolerance in rice, with particular emphasis on spikelet fertility as a key indicator of reproductive performance under elevated temperature. By combining phenotypic characterization, high-density SNP genotyping, stringent marker quality control, population-structure analysis, genomic relationship modeling, mixed linear model GWAS, and haplotype analysis, the study provides a comprehensive approach for identifying genomic regions potentially involved in heat adaptation. The analysis demonstrated substantial phenotypic variation among the 420 simulated rice accessions,

with spikelet fertility showing a broad range and relatively high estimated heritability ($H^2 = 0.68$), indicating that genetic variation could contribute substantially to differences in reproductive heat tolerance. Following SNP quality control, 61,842 high-quality SNP markers were retained from the initial 80,000 markers. Population-structure and kinship correction substantially improved the reliability of association testing, as indicated by the near-optimal simulated genomic inflation factor ($\lambda_{GC} = 1.04$). The mixed linear model identified four genome-wide significant loci located on chromosomes 2, 3, 6, and 9, demonstrating that heat tolerance is likely controlled by multiple genomic regions rather than a single major locus. Among these signals, the chromosome 3 locus represented the strongest candidate, with a highly significant association ($P = 1.1 \times 10^{-9}$) and the largest estimated allelic effect (+4.7 percentage points in spikelet fertility). Individual significant loci explained approximately 2.7–5.8% of the phenotypic variation, further supporting the polygenic nature of the trait.

Haplotype analysis provided additional evidence for the importance of the chromosome 3 region, with the favorable haplotype showing an approximately 7–9 percentage-point advantage in spikelet fertility compared with the least favorable

haplotype. This result suggests that combining individual SNP associations with local haplotype information may provide a more informative strategy for prioritizing genomic regions for future investigation. The integration of reproductive and physiological traits also provides a broader perspective on heat adaptation and may help distinguish loci involved specifically in reproductive protection from those contributing to general physiological resilience.

Overall, the findings demonstrate the value of GWAS as a statistical and genomic approach for dissecting the complex genetic architecture of rice heat tolerance. The identified candidate regions provide promising targets for subsequent fine mapping, candidate-gene characterization, gene-expression analysis, and functional validation. In practical breeding programs, validated favorable alleles could potentially be incorporated into marker-assisted selection and genomic selection strategies to accelerate the development of climate-resilient rice cultivars.

However, an important limitation of the present study is that the genotype-phenotype dataset and reported numerical associations are simulated. Therefore, the identified SNPs, loci, effect sizes, and haplotype differences should not be interpreted as experimentally confirmed genetic discoveries. Future studies should validate these candidate regions using independent rice germplasm collections, multi-location and multi-year heat-stress experiments, and advanced molecular approaches. Particular attention should be given to genotype-by-environment interactions because the stability of heat-tolerance loci across environments is essential for their practical breeding value. Functional validation of candidate genes will ultimately be required to establish causal relationships between genomic variation and heat tolerance.

REFERENCES

- Bitra, C. E., & Gerats, T. (2013). Plant tolerance to high temperature in a changing environment: Scientific fundamentals and production of heat stress-tolerant crops. *Frontiers in Plant Science*, 4, 273. <https://doi.org/10.3389/fpls.2013.00273>
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633-2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., ... Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, 42, 961-967. <https://doi.org/10.1038/ng.695>
- Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C., Fan, D., Lu, Y., Weng, Q., Liu, K., Zhou, T., Jing, Y., Si, L., Dong, G., Huang, T., ... Han, B. (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics*, 44, 32-39. <https://doi.org/10.1038/ng.1018>
- Jagadish, S. V. K., Cairns, J. E., Kumar, A., Greenberg, A. J., Ramesh, T., & Singh, R. K. (2011). Genetic analysis of heat tolerance at anthesis in rice. *Crop Science*, 50, 1633-1641.
- Jagadish, S. V. K., Craufurd, P. Q., & Wheeler, T. R. (2007). High temperature stress and spikelet fertility in rice (*Oryza sativa* L.). *Journal of Experimental Botany*, 58, 1627-1635. <https://doi.org/10.1093/jxb/erm003>
- Jagadish, S. V. K., Craufurd, P. Q., & Wheeler, T. R. (2008). Phenotyping parents of mapping populations of rice for heat tolerance during anthesis. *Crop Science*, 48, 1140-1146.

- Jagadish, S. V. K., Murty, M. V. R., & Quick, W. P. (2012). Rice responses to rising temperatures—Challenges, perspectives and future directions. *Plant, Cell & Environment*, 35, 604–619. <https://doi.org/10.1111/j.1365-3040.2011.02345.x>
- Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42, 348–354. <https://doi.org/10.1038/ng.548>
- Lafarge, T., Bueno, C., Frouin, J., Jacquin, L., Courtois, B., & Ahmadi, N. (2017). Genome-wide association analysis for heat tolerance at flowering detected a large set of genes involved in adaptation to thermal and other stresses. *PLoS ONE*, 12, e0171254. <https://doi.org/10.1371/journal.pone.0171254>
- Li, X., Lawas, L. M. F., Malo, R., Glaubitz, U., Erban, A., Mauleon, R., Heuer, S., Zuther, E., Kopka, J., Hinch, D. K., & Jagadish, S. V. K. (2015). Metabolic and proteomic responses of rice varieties with different heat tolerance at reproductive stage. *Scientific Reports*, 5, 16939. <https://doi.org/10.1038/srep16939>
- Li, X., et al. (2022). QTL mapping and identification of candidate genes using a genome-wide association study for heat tolerance at anthesis in rice (*Oryza sativa* L.). *Frontiers in Genetics*, 13, 952046.
- Li, X., et al. (2022). Identifying candidate genes and patterns of heat-stress response in rice using a genome-wide association study and transcriptome analyses. *The Crop Journal*, 10, 1633–1643.
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, 28(18), 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Matsui, T., Omasa, K., & Horie, T. (2001). The difference in sterility due to high temperatures during the flowering period among japonica-rice varieties. *Plant Production Science*, 4, 90–93.
- Nubankoh, P., et al. (2020). QTL-seq reveals genomic regions associated with spikelet fertility in response to a high temperature in rice (*Oryza sativa* L.). *Plant Cell Reports*, 39, 515–526.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38, 904–909. <https://doi.org/10.1038/ng1847>
- Prasad, P. V. V., Boote, K. J., Allen, L. H., & Jones, J. W. (2006). Life-cycle effects of high temperature stress on two cool-season grass species. *Crop Science*, 46, 1617–1625.
- Satake, T., & Yoshida, S. (1978). High temperature-induced sterility in indica rices at flowering. *Japanese Journal of Crop Science*, 47, 6–17.
- Khan, R., Khan, A., Muhammad, I., & Khan, F. (2025). A Comparative Evaluation of Peterson and Horvitz-Thompson Estimators for Population Size Estimation in Sparse Recapture Scenarios. *Journal of Asian Development Studies*, 14(2), 1518-1527.
- Shi, W., Muthayya, S., Singleton, G. R., & Wang, H. (2013). Field and experimental evidence for the effects of high temperature on rice production. *Field Crops Research*, 154, 19–29.
- Shi, Y., et al. (2023). Integrating GWAS and transcriptomics to identify candidate genes conferring heat tolerance in rice. *Frontiers in Plant Science*, 13, 1083678.
- Wahid, A., Gelani, S., Ashraf, M., & Foolad, M. R. (2007). Heat tolerance in plants: An overview. *Environmental and Experimental Botany*, 61, 199–223. <https://doi.org/10.1016/j.envexpbot.2007.05.011>

- Wassmann, R., Jagadish, S. V. K., Sumfleth, K., Pathak, H., Howell, G., Ismail, A., Serraj, R., Redona, E., & Singh, R. K. (2009). Regional vulnerability of rice to changing climate: Impacts of current and future heat stress in South Asia and projected responses of agroecosystems. *Advances in Agronomy*, 101, 91-133.
- Ye, C., Argayoso, M. A., Redoña, E. D., Sierra, S. N., Laza, M. A., Dilla, C. J., Mo, Y., Thomson, M. J., Chin, J., Delaviña, C. B., et al. (2012). Mapping QTL for heat tolerance at flowering stage in rice using SNP markers. *Rice*, 5, 34.
- Yoshida, S., Satake, T., & Mackill, D. J. (1981). High temperature stress in rice. *IRRI Research Paper Series*, 67, 1-15.
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., McMullen, M. D., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38, 203-208. <https://doi.org/10.1038/ng1702>
- Zafar, S. A., Noor, M. A., Waqas, M. A., Wang, X., Shaheen, T., Raza, M., & Rahman, M. U. (2018). Temperature extremes in rice production: Physiological, molecular and genetic perspectives. *Journal of Integrative Agriculture*, 17, 1263-1277.
- Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., Norton, G. J., Islam, M. R., Reynolds, A., Mezey, J., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications*, 2, 467.
- Zhang, H., et al. (2021). QTL mapping and identification of candidate genes for heat tolerance at the flowering stage in rice. *Frontiers in Genetics*, 11, 621871.
- Zhou, Y., Li, K., Tang, C., Zhu, M., Bao, Y., Zhang, W., Li, P., Lv, W., Zhang, M., Wang, C., et al. (2026). Identification of candidate genes for heat tolerance in rice by genome-wide association study and transcriptome sequencing. *BMC Plant Biology*, 26, 556.