

Mining of eIF4G^{tsv1} diversity reveals new Rice tungro spherical virus resistance alleles and their evolutionary enrichment

Israel Dave Ambita

International Rice Research Institute

Ian Paul Navea

International Rice Research Institute

Gilda Jonson

International Rice Research Institute

Samantha Izabelle Alforja

International Rice Research Institute

Frances Nikki Borja

International Rice Research Institute

Dmytro Chebotarov

International Rice Research Institute

Ramil Mauleon

International Rice Research Institute

Maria Genaleen Diaz

University of the Philippines Los Baños

John Damien Platten

International Rice Research Institute

Il-Ryong Choi

International Rice Research Institute

Van Schepler-Luu

v.scheplerluu@cgiar.org

International Rice Research Institute

Research Article

Keywords: rice tungro, allele mining, RTSV, eIF4G, natural variation, SNPs, InDels

Posted Date: September 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10857362/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Rice tungro is a viral disease of rice in Asia caused by a co-infection with rice tungro bacilliform virus (RTBV) and rice tungro spherical virus (RTSV) and transmitted by green leafhoppers (GLH). Host resistance to RTBV, RTSV or GLH remains the most effective management strategy. The eukaryotic translation initiation factor 4 gamma (*eIF4G*, *LOC_Os7g36940*) acts as a susceptibility gene to RTSV. Previous studies identified two RTSV resistance-associated variants within a 30-bp region in exon 9 of *eIF4G* (*eIF4G^{tsv1}*) in Utri Merah (UR) and TKM6 (TKM). Here, we mined *eIF4G^{tsv1}* polymorphisms in 3,024 accessions from the 3K Rice Genome Panel and evaluated RTSV resistance in 345 accessions to identify additional resistance-associated alleles.

Results

Approximately 7.2% of accessions (219/3,024) carried nonsynonymous SNPs or 3-12 bp deletions within the 30-bp *eIF4G^{tsv1}* region. In addition to UR and TKM, we identified eight alternative alleles: ARC, BON, AS, KG, KK, KM, KM-UR, and KD. All accessions carrying UR and 83% of accessions carrying TKM showed lower RTSV titers than the susceptible control TN1. Similarly, 94% of accessions carrying alternative alleles had lower RTSV titers than TN1, supporting their association with RTSV resistance. Most accessions carrying resistance-associated alleles belonged to *aromatic* and *temperate japonica* groups and originated from Asia regions where tungro is prevalent, suggesting possible positive selection for RTSV resistance.

Analysis of 1,076 elite *indica* breeding lines revealed low frequency and limited diversity of *eIF4G^{tsv1}* resistance-associated alleles. TKM accounted for 95% of the resistance-allele carriers, whereas ARC, UR, and KM were rare, and the remaining six alleles were absent. To support marker-assisted selection, PACE (PCR allele competitive extension) markers were developed for KM, ARC, TKM, and UR. The contribution of TKM was further validated in IRRI 154, where removal of this allele through backcrossing increased tungro severity.

Conclusions

Natural variation in exon 9 of *eIF4G^{tsv1}* is associated with RTSV resistance in rice. The newly identified alleles and PACE markers provide genetic resources and selection tools to accelerate tungro resistance breeding in rice.

1 Introduction

Rice tungro disease (RTD) is a viral disease that impedes rice production in different countries, particularly in South Asia and Southeast Asia. This disease is caused by the co-infection of two distinct viruses, *Rice tungro bacilliform virus* (RTBV) and *Rice tungro spherical virus* (RTSV). RTSV has spherical

particles containing a single-stranded 12 kb RNA genome encoding a large polyprotein and possibly one or two other smaller proteins [1], whereas RTBV has distinct bacilliform-shaped virions [2]. Both viruses are transmitted in a semi-persistent manner by green leafhoppers (GLH, *Nephotettix virescens*) [2]. GLH can readily transmit RTSV into the host plant and is independent of RTBV transmission. In contrast, RTBV transmission requires an RTSV-derived helper factor, likely composed of virus-associated protein(s) produced in infected plant tissues, which mediates the interaction between RTBV particles and the GLH vector, thereby enabling transmission [3].

The development and utilization of rice cultivars resistant to RTSV, RTBV or GLH are key approaches for controlling RTD. Resistance against GLH helps control RTD by decreasing the probability of viral transmission through insect vectors. The use of this approach, however, has some caveats. For example, GLH resistance has shown limited durability for continued use in rice fields, as insect vectors can adapt and overcome the resistance barrier present in insect-resistant rice varieties. The emergence of new insect vectors could also be a risk associated with the continued use of GLH-resistant varieties due to selective pressure. In recent years, the focus on controlling RTD has shifted towards the use of virus-resistant varieties [4]. Molecular tools, including genomic mapping and cloning, have improved our understanding of RTD resistance and virus-host interaction [4–6].

RTSV is a positive-sense single-stranded RNA virus whose genome functions directly as mRNA and therefore depends on the host translation machinery for the polyprotein synthesis required for replication [1]. In rice, resistance to RTSV is mapped to a recessive allele of *eIF4G* (LOC_Os07g36940, *eIF4G^{tsv1}*), which encodes a core scaffold component of the eIF4F protein complex that recruits ribosomes for cap-dependent translation [7]. Reducing *eIF4G^{tsv1}* expression through artificial microRNAs or introducing targeted mutations via CRISPR/Cas9 in susceptible varieties decreases RTSV accumulation and confers resistance [8]. The molecular basis of resistance is consistent with impaired viral translation: positive-sense RNA viruses hijack host eIF4F components, including *eIF4G*, to initiate viral RNA translation [5]. Loss-of-function or altered *eIF4G^{tsv1}* alleles likely disrupt viral RNA recruitment to the translation initiation complex, thereby limiting polypeptide synthesis, replication complex formation, and genome amplification. Because resistance results from reduced host factor availability rather than immune activation, *eIF4G^{tsv1}*-mediated resistance represents a recessive loss-of-susceptibility mechanism [7, 8].

Genetic databases are rich resources of novel alleles for valuable traits, such as disease resistance. In rice, the publication of the first reference genome of Nipponbare by the International Rice Genome Sequencing Project (IRGSP) has initiated efforts to systematically explore and use rice diversity for genetic research and breeding [9]. The Nipponbare genome was used for reference-based assemblies. The completed 3000 Rice Genomes Project (3K RGP) has generated a large dataset of genomic variation in rice [10]. Along with the public database SNP-Seek, the 3K RGP has provided genomic data resources for the identification of useful rice accessions [9].

Allele mining is an approach used to dissect naturally occurring variations of genes that potentially modulate traits for crop improvement [11]. Allele mining aids in tracing the evolution of alleles, identifying new haplotypes, and developing allele-specific molecular markers that can be used for marker-assisted selection. The integration of rice genetic databases into allele mining workflows has been applied to identify novel resistance alleles against various rice pathogens [9, 12, 13]. The allele mining workflow includes the following steps: 1) comparative analysis between the sequenced genomes and known alleles or haplotypes; 2) identification of accessions harboring superior alleles or haplotypes and conducting phenotypic evaluation; 3) surveying for the presence of novel alleles in breeding lines; 4) introgression to diverse genetic backgrounds; and 5) phenotypic validation in introgression lines [9].

To date, two $\text{eIF4G}^{\text{tsv1}}$ resistance alleles, derived from Utri Merah (UR) and TKM6 (TKM), have been described [7]. These alleles carry mutations within a 30-bp region in exon 9, specifically at nucleotide positions 3163–3192 of $\text{eIF4G}^{\text{tsv1}}$. These mutations resulted in amino acid (aa) substitutions and deletions relative to the canonical susceptible $\text{eIF4G}^{\text{tsv1}}$ allele represented by Nipponbare and Taichung Native 1 (TN1). TKM carried a Val deletion at the aa position 6, whereas UR carried a Tyr-to-Cys substitution at aa position 4 and a Val-to-Ala substitution at aa position 6. The frequency of TKM and UR alleles in diverse rice panels such as the 3K RGP remains unknown. Moreover, natural variants beyond these two alleles may exist. We hypothesized that variants differing from the susceptible allele (TN1/Nipponbare) within this 30-bp region could confer resistance to RTSV. In this study, we mined $\text{eIF4G}^{\text{tsv1}}$ allelic variation in the 3K RGP to determine the distribution of known resistance alleles and to identify novel candidate variants. Accessions carrying putative resistance alleles were then selected for RTSV phenotyping by artificial inoculation to assess their contribution to resistance. The most frequent resistance-associated alleles were used to develop PCR allele competitive extension (PACE) markers for marker-assisted selection. We validated the function of the TKM allele in IRRI 154, an elite variety carrying the TKM allele, together with its near-isogenic line in which the TKM allele had been removed through backcrossing.

2 Results

2.1 Eight new alternative alleles of $\text{eIF4G}^{\text{tsv1}}$ identified in the 3K Rice Genomes Panel

To assess $\text{eIF4G}^{\text{tsv1}}$ variation in the 3K Rice Genomes Panel (RGP), the SNP matrix from the SNP-Seek database (<https://snp-seek.irri.org/>) was used. We focused on a 30-bp (nt positions 3163–3192) region in exon 9 of $\text{eIF4G}^{\text{tsv1}}$, as variation in this region contributes to *tsv1*-mediated resistance [7, 8]. In this 30-bp region, we identified ten alternative alleles relative to the canonical $\text{eIF4G}^{\text{tsv1}}$ susceptible allele represented by Nipponbare and TN1 reference (REF) allele (Fig. 1). The REF allele encoded the 10-amino-acid sequence Ala-Gly-Lys-Ser-Tyr-Val-Val-Asp-His-Pro.

Among the ten alternative alleles detected, two alleles (TKM and UR) were previously reported and eight additional alleles were newly detected, including ARC (ARC13829-type), AS (Adday Sel-type), BON (Bongeza-type), KG (Ketan Gundil-type), KK (Kalar Kar-type), KM (Kinmaze-type), KM-UR (Kinmaze-Utri Merah-type), and KD (Kala Digha-type). Seven of the 10 alleles, including TKM, UR, ARC, AS, KD, KM, and KM-UR, carried missense mutations affecting amino acid positions 2, 4, 5, or 6 within the 10-amino acid region. TKM carried a deletion at aa position 6. UR carried a Ser-to-Cys substitution at aa position 4 and a Val-to-Ala substitution at aa position 6. ARC carried a Tyr-to-Cys substitution at aa position 5, AS carried a Tyr-to-Asn substitution at the same position, KD carried a Ser-to-Cys substitution at aa position 4, KM carried a Gly-to-Val substitution at aa position 2, and KM-UR combined the KM and UR mutations with substitutions at aa positions 2 and 6. The remaining three alleles displayed distinct mutation patterns. BON contained an in-frame deletion spanning aa positions 5–8 corresponding to Tyr-Val-Val-Asp, whereas KG carried a Val-to-Gly substitution at aa position 6. KK shared the Tyr-to-Asn substitution at aa position 5 observed in AS and additionally carried a His-to-Tyr substitution at aa position 9. Overall, the identified alleles consisted primarily of missense mutations and one to four aa in-frame deletions concentrated within aa positions 2–8 of the resistance-associated region of eIF4G^{tsv1}, suggesting that these variants may represent resistance alleles.

2.2 The eIF4G^{tsv1} alternative alleles are highly associated with RTSV resistance

To evaluate whether eIF4G^{tsv1} alternative alleles contribute to RTSV resistance, we selected 345 accessions from the 3K RGP including all accessions carrying the eIF4G^{tsv1} alternative alleles ($n = 219$) and a random set of 119 accessions carrying the REF allele for phenotypic assessment. Among the 219 accessions harboring the alternative eIF4G^{tsv1} alleles, 30 and 16 carried the known resistance alleles TKM and UR, respectively, whereas 173 accessions carried the eight novel alternative alleles: KM ($n = 91$), ARC ($n = 65$), KD ($n = 7$), BON ($n = 4$), AS ($n = 3$), KG ($n = 1$), KK ($n = 1$), and KM-UR ($n = 1$) (Table S1). These 345 accessions were evaluated for resistance against RTSV by artificial inoculation with RTSV-viruliferous green leafhoppers (GLH), followed by quantification of RTSV titer using Double Antibody Sandwich Enzyme-Linked Immunosorbent Assay (DAS-ELISA). Accessions with RTSV titers lower than the susceptible check TN1 and comparable to the resistant check TW16 were classified as resistant, whereas accessions with RTSV titers comparable to TN1 and higher than TW16 were classified as susceptible (threshold, $p = 0.001$).

Differences in RTSV titers were observed among eIF4G^{tsv1} allele groups (Fig. 2; Table S1). Of the accessions carrying the eIF4G^{tsv1} alternative alleles, 94% (205/219) exhibited resistant phenotypes with significantly lower RTSV titers than TN1 (refer to p -values in Table S1), demonstrating the association of these alleles with RTSV resistance. Accessions carrying the two previously reported eIF4G^{tsv1} resistance alleles, TKM and UR, were 83% (25/30) and 100% (16/16) resistant to RTSV, respectively. Among the eight novel alternative alleles, KM was the most frequent ($n = 91$), with 89 accessions showing resistance (98) %. The ARC was represented by 65 accessions, of which 60 were resistant (92%). KD

showed 86% (6/7) resistant reactions while BON showed 75% (3/4) resistant reactions, although both were represented by smaller sample sizes ($n = 7$ and $n = 4$, respectively). Complete resistance was observed in AS ($n = 3$), KG ($n = 1$), KK ($n = 1$), and KM-UR ($n = 1$); however, the limited number of accessions carrying these alleles restricts broader inference on their phenotypic consistency.

Among the 126 accessions carrying the REF allele, 10% (13/126) had significantly lower RTSV titers than TN1 (refer to p -values in Table S1) and were classified as resistant. Conversely, 14 of the 219 (8%) accessions carrying the $\text{eIF4G}^{\text{tsv1}}$ alternative alleles were susceptible (Table S1). This incomplete concordance with our prediction indicated that $\text{eIF4G}^{\text{tsv1}}$ variation in the 30-bp region in exon 9 alone does not fully explain the response and perhaps additional loci might be involved in modulating the RTSV response. We therefore conducted a genome-wide association study (GWAS) and used three tests: the Wald test, log-likelihood ratio test, and score test to detect SNPs associated with resistance or susceptibility to RTSV. All three analyses revealed a strong signal on chromosome 7 corresponding to the $\text{eIF4G}^{\text{tsv1}}$ locus, and no additional significant loci was detected (Figure S1). This result indicates that $\text{eIF4G}^{\text{tsv1}}$ is a major RTSV resistance locus.

2.3 The $\text{eIF4G}^{\text{tsv1}}$ resistance-associated alleles are rare in the 3K RGP but enriched in aromatic and temperate japonica rice

To evaluate the distribution of $\text{eIF4G}^{\text{tsv1}}$ alleles in a diverse rice panel, we surveyed all accessions in the 3K RGP database. Among the 3,024 accessions, 2,785 (92%) carried the REF allele, whereas only 219 (7.2%) carried the alternative resistance-associated alleles. Twenty accessions (0.7%) were either heterozygous or unclassified because of ambiguous SNP calls (Table S2). These results indicate that the two known resistance alleles (TKM and UR) and the putative resistance alleles identified in this study (ARC, BON, AS, KG, KK, KM, KM-UR, and KD) were generally rare in the 3K panel.

Resistance-associated alleles were unevenly distributed across rice subpopulations in the 3K RGP and were most frequent in the *aromatic* and *temperate japonica* groups (Fig. 3). The *aromatic* group showed the highest proportion of resistance-associated alleles, with 48 accessions carrying alternative alleles (66.7%) compared with 24 carrying the REF allele. The *temperate japonica* group also showed substantial enrichment, with 67 accessions carrying resistance-associated alleles (23.7%) and 216 REF accessions. In contrast, resistance-associated alleles were less common in *tropical japonica* (24/347, 6.5%) and *indica*-related subpopulations, including *indicaX* (31/583, 5.0%), *indica1B* (18/187, 8.8%), *indica2* (4/278, 1.4%), *indica3* (3/472, 0.6%), and *indica1A* (1/208, 0.5%). Intermediate frequencies were observed in *admix* (11/88; 11.1%) and *japonicax* (7/74; 8.6%), whereas *aus* had few resistance-associated alleles (5/196; 2.5%). No resistance-associated alleles were detected in the *subtropical japonica* group. Because *aromatic* and *temperate japonica* rice originated from Asian regions [14] that overlap with the historical distribution of tungro disease [15], the enrichment of $\text{eIF4G}^{\text{tsv1}}$ resistance-associated alleles in these groups suggests a possible pathogen-driven selection for RTSV resistance in these groups.

Resistance-associated alleles showed subpopulation-specific distributions across the 3K RGP, with both previously characterized resistance alleles and multiple putative novel alleles displaying distinct patterns of enrichment (Fig. 3). Among the resistance-associated alleles, previously characterized TKM and UR were detected in 30 and 16 accessions, respectively. In contrast, the novel resistance-associated alleles identified in this study showed broader overall representation, with KM occurring in 91 accessions, ARC in 65, and KD in 7, whereas BON ($n = 4$), AS ($n = 3$), KG ($n = 1$), KK ($n = 1$), and KM-UR ($n = 1$) were rare. Distinct subpopulation-specific patterns were also observed. KM was strongly associated with *temperate japonica*, where all 67 accessions carried this allele, although additional KM accessions were detected in *indicax* ($n = 14$), *japonicax* ($n = 6$), *tropical japonica* ($n = 1$), *indica1B* ($n = 2$), and *indica1A* ($n = 1$). ARC was enriched in *aromatic* ($n = 45$) and *admix* ($n = 9$), with smaller numbers detected in *indicax* ($n = 3$), *tropical japonica* ($n = 6$), *japonicax* ($n = 1$), and *indica2* ($n = 1$). TKM was predominantly associated with *indica1B* ($n = 16$) and *indicax* ($n = 11$), whereas UR was largely restricted to *tropical japonica* ($n = 15$). KD occurred mainly in *indica*-related groups, including *indicax* ($n = 3$), *indica2* ($n = 2$), and *indica3* ($n = 2$). The remaining rare alleles showed narrow distributions, including BON in *aus* ($n = 4$), AS in *aromatic* ($n = 2$) and *admix* ($n = 1$), KG in *tropical japonica* ($n = 1$), KK in *aus* ($n = 1$), and KM-UR in *tropical japonica* ($n = 1$). These results indicate the subpopulation-specific distribution of alternative eIF4G^{tsv1} alleles.

2.4 The frequency and diversity of eIF4G^{tsv1} resistance-associated alleles are low in elite *indica* breeding lines

To estimate the frequency and diversity of eIF4G^{tsv1} alleles in elite *indica* breeding materials, we analyzed SNP data from 1,076 elite *indica* breeding lines. A total of 235 lines (22%) carried resistance-associated alleles, whereas most lines (841, 78.2%) carried the REF allele (Fig. 4a). The frequency of resistance-associated alleles varied across breeding groups (Fig. 4b), with a higher representation in IRRI-early-maturity lines (99/241, 41%). There were progressively lower frequencies of resistance-associated alleles in IRRI-medium maturity (63/248, 25%), IRRI-late maturity (41/226, 18%), IRRI-hybrid (22/281, 8%), and non-IRRI elite lines (10/80, 13%). Despite these differences, the TKM allele was consistently predominant across all groups ($n = 224$), whereas non-TKM alleles occurred sporadically and at a low frequency (ARC, $n = 5$; UR, $n = 5$; KM, $n = 1$). These findings showed that elite *indica* breeding lines are dominated by the TKM allele and that the diversity of eIF4G^{tsv1} alternative alleles is low, suggesting the need for diversification.

Figure 4. **Distribution of eIF4G^{tsv1} alleles in elite *indica* breeding lines.** (a) Distribution of eIF4G^{tsv1} alleles among elite *indica* rice lines. (b) Distribution of eIF4G^{tsv1} alternative alleles across the IRRI breeding groups. The cell values indicate the number of accessions carrying each allele within each breeding group. Totals per allele and breeding group are shown at the right and bottom margins. Values below each column (n) indicate the total number of elite lines in each group.

2.5 PACE markers distinguished three major eIF4G^{tsv1} alleles

To facilitate the utilization and diversification of $eIF4G^{tsv1}$ alternative alleles in breeding programs, we developed allele-specific markers for $eIF4G^{tsv1}$. PCR Allele Competitive Extension (PACE) assays were designed to target the four most frequent $eIF4G^{tsv1}$ alternative alleles (ARC, KM, TKM, and UR) in the 3K RGP and elite *indica* breeding lines. For each allele, a set of three primers was designed, consisting of two allele-specific forward primers and one common reverse primer (Table S3).

To evaluate marker performance, we estimated genotypic clustering, false positive and false negative rate of each assay. PACE assays were performed on a subset of 88 accessions representing four $eIF4G^{tsv1}$ alternative (ARC, $n = 23$; KM, $n = 3$; TKM, $n = 18$; UR, $n = 9$) and REF (susceptible, $n = 35$) alleles (Table S4). Pooled DNA samples containing both REF and $eIF4G^{tsv1}$ alternative alleles were included as a heterozygous check. PACE genotyping demonstrated a separation of the ARC, KM, and UR alternative alleles from the REF allele, with distinct clustering of homozygous REF, homozygous alternative, heterozygous/mixed DNA, and non-template control samples (Fig. 5a-c). However, the assay designed for the TKM allele did not produce clear genotype clusters (Figure S2a).

Since the PACE marker designed at the polymorphic region in exon 9 of $eIF4G^{tsv1}$ for the TKM allele did not produce clear clustering of resistant, susceptible, and heterozygous genotypes, a new linkage disequilibrium (LD) marker was designed. The marker M898 targeting the SNP at Chr 7:22,118,459, located upstream of the $eIF4G^{tsv1}$ locus, showed clear genotype clustering; hence, it can be used as an alternative marker for the TKM allele (Figure S2b).

To test whether the PACE identified alleles were correlated with the resistance phenotype, the 88 accessions selected for the PACE assay were artificially inoculated with RTSV viruliferous GLH. The RTSV titer was quantified using DAS-ELISA. Accessions carrying the ARC, KM, and UR alleles showed low RTSV titer compared to the accessions carrying the REF allele (DMRT, $p < 0.001$) (Fig. 5d, Table S4). All accessions with homozygous calls for ARC, KM, and UR alleles (FAM) were classified as resistant, whereas all accessions carrying homozygous REF allele calls (HEX) were classified as susceptible, resulting in a false positive rate and false negative rate of 0%. These results demonstrate that the ARC, KM, and UR PACE markers reliably discriminate resistant and susceptible $eIF4G^{tsv1}$ alleles and are suitable for marker-assisted selection of RTSV resistance.

2.6 Removal of the TKM allele increased tungro susceptibility in IRRI 154

As tungro disease results from co-infection with RTSV and RTBV, we evaluated the contribution of RTSV resistance-associated alleles to tungro resistance. The most abundant resistance-associated allele, TKM, was selected for the validation. IRRI 154, an elite line carrying the resistant TKM allele, was used for this purpose in the present study. A BC₆F₅-derived near-isogenic line lacking the TKM allele, IR23X1020, was developed through marker-assisted introgression from the tungro-susceptible line FL478, carrying the REF allele. IR23X1020 and IRRI 154 were artificially inoculated with viruliferous GLH

carrying RTSV and RTBV. Disease severity was assessed using a standard 1–9 evaluation scale based on plant height reduction and leaf discoloration symptoms [16].

Tungro-inoculated plants developed yellow to orange leaf discoloration and showed reduced plant height compared to control plants (Fig. 6a; Figure S3a). Across both trials, mock-inoculated IRRI 154 and IR23X1020 plants did not differ significantly in terms of plant height. In contrast, RTSV+RTBV-inoculated IR23X1020 plants were significantly shorter than the inoculated IRRI 154 plants in both trials. In the first trial, IRRI 154 showed moderate tungro symptoms, with 15 of 20 plants classified with a disease score of 5 and five plants with a score of 7, resulting in a disease index of 6 and an overall moderately resistant reaction. In contrast, IR23X1020 exhibited more severe symptoms, with one plant scored at 3, two at 5, four at 7, and 13 at 9, resulting in a disease index of 7 and a susceptible reaction (Table S5; Fig. 6b). Similarly, in the second trial, assessed 21 days after inoculation, IRRI 154 showed disease scores of 5 in five plants and 7 in 15 plants, corresponding to a disease index of 6.5, whereas IR23X1020 showed disease scores of 5 in three plants, 7 in 15 plants, and 9 in two plants, corresponding to a disease index of 7.0 (Figure S3b; Table S5). These results demonstrate that IR23X1020 was more susceptible to tungro than IRRI 154, further supporting the contribution of the TKM allele to the enhanced tungro resistance.

3 Discussion

Rice tungro disease (RTD) is a major constraint to rice production in Asia. Improving host resistance is central to the sustainable management of these diseases. Our results show that natural variation within the 30-bp (positions 3163–3192) region in exon 9 of *eIF4G^{tsv1}* is associated with RTSV resistance. The high proportion of resistant phenotypes among accessions carrying putatively resistant alleles indicates that the disruption of this gene is a reliable resistance mechanism. Notably, the eight resistance-associated alleles (ARC, BON, AS, KG, KK, KM, KM-UR, and KD) identified in this study have not been previously reported, which expands the known allelic diversity at this locus. However, their low frequency in the 3K Rice Genomes Panel (RGP) and elite *indica* breeding lines suggests that these beneficial variants remain underutilized, highlighting the opportunity to incorporate them into breeding programs through marker-assisted selection (MAS).

Recessive resistance is the predominant form of resistance in plants against viruses and can be attributed to the loss of function of susceptibility genes [17, 18]. Since the mutations are masked by the susceptible alleles, it is challenging to discern which cultivars harbor mutations. Furthermore, infection of RTSV in rice is usually asymptomatic; hence, it can only be detected through ELISA and molecular methods. Limitations in screening methods in the past, together with the recessive nature of *eIF4G^{tsv1}*-mediated resistance, might have reduced the chances of obtaining new rice accessions possessing homozygous resistance alleles.

The uneven distribution of resistance-associated alleles reflects the rice population structure. The *aromatic* and *temperate japonica* groups showed relatively high frequencies of resistance-associated alleles (Fig. 2a). From an evolutionary and pathological perspective, this may reflect the historical

overlap between rice populations and the prevalence of viral pathogens. Most *aromatic* rice accessions are cultivated in South and Southeast Asia, where RTSV is highly prevalent [19], whereas *temperate japonica* accessions are primarily associated with East Asia, including China, Japan, and Korea, where Rice waika virus, a virus closely related to RTSV and possibly representing the same or a related virus under a different name, was historically widespread [19]. However, the low frequency of resistance-associated alleles in *indica* despite its predominance in RTSV-endemic regions indicates that pathogen distribution alone does not explain the observed pattern. Population history, domestication, breeding practices, and subpopulation-specific evolutionary trajectories may also have contributed to the distribution of resistance-associated alleles [20, 21]. Together, these observations suggest that *elf4G^{tsv1}* resistance-associated alleles were preferentially retained or enriched in specific rice subpopulations that historically coexisted with RTSV or RTSV-related viral pathogens.

Interestingly, a small number of alleles (KM and ARC) were widespread, whereas others remained rare and subpopulation-specific, a pattern typical of resistance loci shaped by limited dissemination or recent origin [22]. The enrichment of resistance-associated alleles in specific rice groups may also indicate long-term natural or artificial selection imposed by persistent viral pressure. Despite the presence of multiple alleles, their overall frequency remained low across the 3K RGP, indicating that *elf4G^{tsv1}*-mediated resistance was not widely fixed. This rarity increases their value for allele mining and introgression into elite germplasms.

Analysis of the resistance-associated alleles revealed that eight consisted of nonsynonymous point mutations, whereas two contained in-frame deletions. All point mutations occurred at the first or second codon positions, suggesting the selection of amino acid changes rather than synonymous substitutions. This is consistent with previous studies showing that RTSV resistance is associated with specific amino acid substitutions in *elf4G^{tsv1}* that alter virus-host interactions [7, 8]. In contrast, the deletions observed in BON and TKM occur in multiples of three and therefore preserve the reading frame, suggesting that resistance arises from structural modification rather than from loss of function. In-frame mutations in *elf4G^{tsv1}* can confer RTSV resistance, whereas disruptive mutations are often not recovered, indicating functional constraints on this gene [7]. A similar reliance on amino acid substitutions was reported for *Pi-ta*, where single nonsynonymous changes determine resistance specificity to different rice blast (*Magnaporthe oryzae*) strains [23]. These findings suggest that *elf4G^{tsv1}* is under strong functional constraints, such that only mutations that modify key residues while preserving the overall protein integrity are retained.

Accumulation of the resistance-associated mutations from different *elf4G^{tsv1}* alleles may provide a strategy to broaden and improve the durability of RTSV resistance. However, because these mutations occur within the same gene, they cannot be pyramided using conventional marker-assisted breeding. Therefore, genome editing may offer a practical approach to combine multiple favorable mutations into a single *elf4G^{tsv1}* allele. In parallel, rapid haplotype enrichment can be used to increase the frequency of

favorable eIF4G^{tsv1} alleles in breeding populations, diversify the use of alternative resistance alleles, and reduce the risk of resistance breakdown caused by reliance on a single allele.

A relatively large number of RTSV resistance sources were identified. Among the 219 accessions carrying resistance-associated alleles, 94% of them were RTSV-resistant. This number represents a substantial germplasm resource for RTSV resistance breeding. In comparison, only 34 rice accessions in the 3K RGP with resistance to the rice yellow mottle virus have been reported [24]. Besides, among the 126 accessions carrying the canonical susceptible eIF4G^{tsv1} allele (REF), 13 accessions were also resistant to RTSV. Conversely, 14 of the 219 accessions carrying alternative alleles were classified as susceptible (Fig. 2; Table S1). This discrepancy may partly reflect environmental effects, as some accessions showed reduced growth under Philippine conditions, which could influence their response to RTSV infection. Because the panel included accessions of diverse geographical origins, variations in local adaptation may have affected phenotypic expression. In addition, the allele classification used in this study was based on 10 amino acid residues (30-bp) in exon 9, associated with previously reported eIF4G^{tsv1} alleles. Accessions grouped under the REF allele may still differ from one another at other causal residues or genomic regions that we did not evaluate in this study. In attempt to identify additional genomic loci controlling RTSV resistance that could explain the discordance between genotype-based predictions and the observed phenotype, we conducted a genome-wide association study (GWAS) using the accessions evaluated by DAS-ELISA. However, the GWAS did not identify any additional loci significantly associated with RTSV resistance (Figure S1). This may reflect the inherent limitations of GWAS in detecting rare or low-frequency alleles, as statistical detection power depends strongly on minor allele frequency thresholds and adequate allele representation within the study population [25, 26]. In addition, the selected accessions were enriched for specific eIF4G^{tsv1} alleles and therefore may not fully represent the broader genetic diversity of rice, potentially reducing mapping resolution. The complex genetic architecture of RTSV resistance, including possible epistatic interactions or loci with small phenotypic effects, may have further constrained GWAS resolution.

We demonstrated that PACE assays can discriminate natural alleles at the eIF4G^{tsv1} locus, particularly for the ARC, KM, and UR alleles. The clustering of S and R accessions and the intermediate positioning of mixed samples indicate reliable allele calling and heterozygote detection. The failure to obtain clear discrimination against the TKM allele is likely due to the nature of its mutation. The TKM allele contains a 3-bp deletion within a tandem GTT repeat. Tandem repeat sequences can reduce primer specificity, particularly when the polymorphism involves a small indel [27]. Hence, alternative linkage disequilibrium markers (M898) located upstream of the 30bp region was a practical replacement. The PACE markers, together with the potential donor accessions (Table S4), provide a framework for marker-assisted introgression of RTSV resistance alleles. Nevertheless, the performance of these PACE markers should be further tested across different breeding populations to confirm their robustness and reliability.

Because tungro disease results from the co-infection of RTSV and RTBV, evaluating the effects of different eIF4G^{tsv1} alleles under mixed infection conditions is essential. We showed that the near-

isogenic line IR23X1020, which lacks the TKM allele, displayed higher disease severity, stronger leaf discoloration, and greater plant height reduction than IRRI 154. Notably, IRRI 154 carried only the TKM allele and no known resistance gene against RTBV, indicating that the TKM allele alone was sufficient to reduce tungro symptoms under mixed infection conditions. However, IRRI 154 remained moderately resistant rather than fully resistant, likely because RTBV infection still contributed to symptom development. Interestingly, TW16 variety that harbor UR, used as a resistant check, was more resistant to tungro than IRRI154, indicating that this UR allele might be more effective than the widely used TKM allele. However, the effect of the TW16 genetic background on this UR allele was not tested. Since only the TKM allele was evaluated in our experiment, the resistance effects of UR and the other novel $eIF4G^{tsv1}$ alleles require further validation in elite breeding backgrounds. In addition, the effect of TKM, UR and other alleles also need to be tested against different RTSV and RTBV strains to better inform the deployment strategy of these alleles in different rice-growing countries or regions.

Allele utility depends on its relevance within target breeding populations rather than on its mere association with a trait. Markers linked to favorable alleles are most useful when those alleles are rare or absent in breeding materials, as they enable efficient identification and introgression of beneficial variants into elite germplasm [28]. Therefore, assessing allele frequency within breeding populations is a critical step before marker development. In our study, almost 80% of IRRI's breeding materials carry the susceptible allele. Among lines carrying resistance-associated alleles, the TKM allele was predominant, while other alleles occurred at low frequency or were absent (Fig. 4b). The predominance of the TKM allele also indicates that it is already widely available in elite *indica* breeding germplasm and therefore represents the most readily deployable source of RTSV resistance for breeding programs. At the same time, the limited representation of additional resistance-associated alleles highlights the potential value of diversifying $eIF4G^{tsv1}$ variation in elite germplasm. Expanding the allelic diversity of $eIF4G^{tsv1}$ may provide opportunities to strengthen resistance durability and broaden the genetic resources available for future RTSV breeding efforts.

4 Conclusions

We demonstrated that publicly available genetic databases, such as SNP Seek, can be used to identify naturally occurring alleles associated with Rice tungro spherical virus (RTSV) resistance. Ten natural variants within a 30-bp region of exon 9 predicted RTSV response with approximately 93% accuracy. The distribution of resistance-associated $eIF4G^{tsv1}$ alleles across rice subpopulations suggests a possible relationship between allele frequency and the historical prevalence of RTSV or related viruses, providing insight into the evolutionary and pathological dynamics shaping resistance variation in rice populations.

The predominance of the TKM allele in elite *indica* breeding germplasm indicates that it is already widely available and represents the most readily deployable source of RTSV resistance for breeding programs. In contrast, the limited representation of other resistance-associated $eIF4G^{tsv1}$ highlights the importance of expanding allelic diversity to broaden the genetic base and improve the durability of RTSV resistance in future breeding efforts.

5 Materials and methods

5.1 eIF4G^{tsv1} allele mining

Since mutations within a 30-bp region of exon 9 in the canonical eIF4G^{tsv1} susceptible allele (REF, Chr. 7: 22,119,334 – 22,119,363) were previously associated with Rice tungro spherical virus (RTSV) resistance [7], this region was examined using the SNP matrix retrieved from the SNPSeek Database. Variants within the target region were inspected and filtered to retain only SNPs and InDels located within the 30-bp interval relative to the REF allele represented by TN1 and Nipponbare. Alleles identical to TN1 and Nipponbare were designated as REF allele, whereas all sequence variants differing from the REF allele were classified as alternative alleles. Rice accessions were subsequently grouped according to sequence variation within this 30-bp region.

Whole-genome resequencing data from 1,076 IRRI elite *indica* breeding lines were retrieved from an in-house variant database (data unpublished) generated by alignment against the IRGSP-1.0 Nipponbare reference genome. Variant call format files were used to extract SNPs within the 30-bp region of exon 9 of eIF4G^{tsv1} to estimate allele frequencies in the elite breeding lines.

5.2 RTSV resistance phenotyping

To evaluate RTSV response, 126 accessions carrying the canonical eIF4G^{tsv1} susceptible allele (REF) and 219 accessions carrying the alternative alleles were selected for phenotyping. TN1 and TW16 were included as susceptible and resistant checks, respectively. Ten plants from each accession were inoculated with RTSV-viruliferous green leafhoppers (GLH) carrying RTSV strain A, one of two biological RTSV variants reported in the Philippines, together with Vt6, which are distinguished using TKM6 as a differential host, where strain A is avirulent to TKM6, whereas Vt6 is avirulent [29]. Three RTSV-viruliferous GLH were allowed to feed on ten-day-old seedlings for 24 h at approximately 25–30°C under shaded conditions. Each plant was enclosed in a tubular mylar cage to confine the insects to the target plant. The cages and insects were removed after the 24 h inoculation period.

Because single RTSV infection produces only mild or no visible symptoms in rice plants, resistance was evaluated based on RTSV titer rather than symptom severity. RTSV titer was estimated at 14 days post inoculation by double-antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) [30]. The second youngest leaf was collected from each inoculated plant and homogenized in phosphate-buffered saline containing Tween 20 (PBS-T; 0.015 M NaCl, 0.735 mM KH₂PO₄, 0.01 M Na₂HPO₄, 0.0015 mM NaN₃, and 0.5 mL L⁻¹ Tween 20) at a ratio of 1:10 (w/v) using a mechanical homogenizer (Wenig & Koch Elektromotoren GMBH, Hannover, Germany). RTSV particles were captured by coat protein-specific antibodies immobilized on ELISA plates and detected using enzyme-conjugated RTSV antibodies. RTSV titer was estimated from optical density measurements at 405 nm (OD₄₀₅) after substrate conversion using an 800 TS Microplate Reader (BioTek Instruments, Inc., USA), where higher absorbance values indicated greater RTSV accumulation. The OD₄₀₅ values were used to classify each accession by comparison with the resistant (TW16) and susceptible (TN1) checks using a *t*-test (*p* = 0.001). Extracts

from healthy and RTSV-inoculated TN1 and TW16 plants, together with PBS-T, were included as controls in the DAS-ELISA assay.

5.3 Genome-wide association for RTSV resistance

Using the 3K RG 29 million biallelic SNP dataset together with SNP data from the selected 345 rice accessions, genome-wide association analysis (GWAS) was performed to identify genomic regions associated with RTSV resistance. SNP datasets were processed in PLINK for genotype filtering and variant call format (VCF) file generation. The resulting VCF files were filtered using VCFtools [31] and BCFtools [32]. A minor allele frequency (MAF) of ≥ 0.05 and missing genotype frequency of ≥ 0.2 thresholds were set to filter out SNPs used for GWAS resulting in 3,692,651 SNPs.

Phenotypic data consisted of DAS-ELISA OD₄₀₅ obtained from ten plants per accession. The trait distribution in the panel was evaluated for normality using the Shapiro-Wilk test implemented in R [33] prior to GWAS.

Because the phenotype distribution was strongly left-skewed and approximated non-normality (Figure S3), OD₄₀₅ values were transformed prior to analysis using rank-based inverse normal transformation to approximate normality and reduce the influence of extreme values [34]. The transformed phenotypic data together with the filtered SNP dataset were used for GWAS.

GWAS was conducted using vcf2gwas [35], a Python-based pipeline that implements Genome-wide Efficient Mixed Model Association (GEMMA) for mixed-model association analysis. Manhattan and quantile-quantile (QQ) plots were generated in R using the package qqman [36]. Genomic regions with p -values $\leq 5 \times 10^{-8}$ were considered significantly associated with RTSV resistance.

5.4 PACE marker design

The four most frequent eIF4G^{tsv1} alternative alleles (KM, ARC, TKM, and UR) in the 3K RGP were selected for PACE (PCR allele Competitive Extension, 3CR Bioscience) marker design. Fifty base upstream and downstream sequences of the 30bp region (Chr. 7: 22,119,334 – 22,119,363) in exon 9 of LOC_Os7g36940 harboring the SNP/InDel in each unique eIF4G^{tsv1} allele were extracted from SNPseek (<http://snp-seek.irri.org/>). The list of the target SNPs/InDels was submitted to Intertek (<https://www.intertek.com/>) for the design and synthesis of PACE markers. For TKM allele, the PACE primers were designed to capture the SNP at Chr 7:22,118,459, upstream of the targeted 30bp region. The primer sequences used in this study are listed in Table S3. For each allele, three primers were designed. The first primer was specific for the eIF4G^{tsv1} REF allele and tagged with a HEX-specific tail sequence at the 5' end. The second primer was specific to eIF4G^{tsv1} alternative allele and tagged with a FAM-specific tail sequence at the 5' end. The third primer was a common reverse marker without tail sequences.

5.5 PACE assay performance evaluation

The accessions used in the assay are listed in Table S4. Young leaves were collected from individual plants, placed in glassine bags, and stored at -20°C until processing. DNA was extracted using the CTAB (N-acetyl-N,N,N-trimethylammonium bromide) method [37]. For each *eIF4G^{tsv1}* allele, a genotyping assay was prepared according to the PACE™ 2.0 Genotyping Master Mix User Guide (3CR Bioscience Limited, UK). The assay consisted of: (i) allele-specific primer 1-FAM (100 µM), (ii) allele-specific primer 2-HEX (100 µM), (iii) a common reverse primer (100 µM), and (iv) PCR-grade sterile distilled water. Each PACE PCR reaction contained 1× PACE 2.0 GMM, PACE genotyping assay mix, water, and genomic DNA in a final volume of 10 µL. Genotyping reactions were cycled according to the conditions described in Table S6 using thermal cyclers. The primer sequences used in this study is listed in Table S3. After thermocycling using a G-Storm GS1 Thermal Cycler PCR system with a standard 96-well thermal block (G-Storm, Somerset, UK). Fluorescent signals were detected using a CLARIOstar® Plus microplate reader (BMG LABTECH, Ortenberg, Germany) configured for FAM and HEX fluorescence detection. Genotype clusters were analyzed using KlusterCaller software. Genotypes were assigned based on relative HEX and FAM fluorescence intensities, corresponding to the REF and *eIF4G^{tsv1}* alternative alleles, respectively.

Marker performance was evaluated by comparing genotype calls with RTSV resistance phenotypes determined by DAS-ELISA. False-positive rate (FPR) was calculated as the proportion of REF allele-carrying accessions incorrectly classified as R, whereas false-negative rate (FNR) was calculated as the proportion of *eIF4G^{tsv1}* alternative allele-carrying accessions incorrectly classified as S.

5.6 Functional validation of the TKM allele in the IRRI 154 background

IRRI 154, an elite line carrying the TKM allele, was used to evaluate the contribution of this allele to tungro resistance. TKM6, the donor of the TKM allele, was included as a reference genotype. IR23X1020, an IRRI 154 BC₆F₅-derived near-isogenic line lacking the TKM allele, was also evaluated. IR23X1020 was developed through marker-assisted selection using a PACE assay (M898) targeting the *eIF4G^{tsv1}* locus at chromosome 7 position 22,118,459 (C/T), allowing the contribution of the TKM allele to be assessed by comparing genotypes that differ in TKM allele status. TN1 and TW16 were included as susceptible and resistant checks, respectively.

The genotypes were evaluated for tungro response under controlled inoculation conditions in two independent trials. In both trials, 9-day-old seedlings were inoculated with viruliferous GLH carrying mixed RTSV and RTBV to induce typical tungro disease symptoms. Twenty seedlings per genotype were inoculated, while five mock-inoculated seedlings served as controls. Disease response was evaluated at different time points in the two trials: 14 dpi in Trial 1 and 21 dpi in Trial 2. Disease severity was evaluated using the standard 1–9 scale [16] based on plant height reduction and leaf discoloration: 1 = no symptoms; 3 = 1–10% height reduction without distinct leaf discoloration; 5 = 11–30% height reduction without distinct leaf discoloration; 7 = 31–50% height reduction with yellow to orange leaf discoloration; and 9 = ≥ 50% height reduction with yellow to orange leaf discoloration. Disease index

values were calculated from individual plant scores and classified as resistant, moderately resistant, or susceptible.

5.7 Statistical analyses

Two-sample, two-tailed Welch's *t*-tests [38] were conducted to compare the OD₄₀₅ values of each accession with the resistant (TW16) and susceptible (TN1) controls. *P*-values were adjusted for multiple testing. Accessions were classified as resistant (R) when they differed significantly from TN1 (adjusted *p*-value ≤ 0.001) but not from TW16 (adjusted *p*-value > 0.001), and as susceptible when they did not differ significantly from TN1 (adjusted *p*-value > 0.001) but differed significantly from TW16 (adjusted *p*-value ≤ 0.001). For accessions that differed significantly from both controls, classification was assigned based on the higher adjusted *p*-value, indicating greater similarity to either the R or S control.

Differences in OD₄₀₅ values among PACE marker (allele-wise) groups were analyzed using one-way analysis of variance, followed by Duncan's multiple range test for mean separation [39].

Welch's *t*-tests were also used to evaluate the effects of mixed RTSV + RTBV inoculation on IRRI 154 and IR23X1020 by comparing mock- and inoculated plants within each genotype, as well as inoculated plants between genotypes.

Abbreviations

3K RGP

3000 Rice Genomes Project: eIF4G^{tsv1} allele codes

bp

base pair(s)

CDS

coding sequence

CRISPR

clustered regularly interspaced short palindromic repeats

CTAB

N-acetyl-N,N,N-trimethylammonium bromide

DAS-ELISA

double antibody sandwich enzyme-linked immunosorbent assay

eIF4F

eukaryotic initiation factor 4F complex

eIF4G/ eIF4G^{tsv1}

eukaryotic initiation factor 4 gamma (*tsv1* allele)

FAM

6-carboxyfluorescein

GLH

green leafhoppers (*Nephotettix virescens*)

GWAS

genome-wide association study

HEX

hexachloro-fluorescein

InDel

insertion-deletion polymorphism

IRGSP

International Rice Genome Sequencing Project

MAS

marker-assisted selection

MEGA

Molecular Evolutionary Genetics Analysis

PACE

PCR Allele Competitive Extension

PCR

polymerase chain reaction

QTL

quantitative trait locus/loci

RTBV

Rice tungro bacilliform virus

RTD

Rice tungro disease

RTSV

Rice tungro spherical virus

SNP

single nucleotide polymorphism

SNPSeek

SNP-Seek Database

TN1

Taichung Native 1

TW16

tungro-resistant line (Utri Merah × TN1)

vcf

Variant Call Format.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All the data of this study are presented in the main and supplementary tables and figures. SNP data can be retrieved from <https://snp-seek.irri.org>

Competing interests

The authors declare no conflicts of interest.

Funding

This work was supported, in whole or in part, by the Gates Foundation via the AGGRi project (INV-008226/ OPP1194925), the CGIAR Accelerated Breeding and Plant Health Initiatives. IDA received a scholarship from the Philippines Department of Science and Technology - Science Education Institute (DOST-SEI) through the Accelerated Science and Technology Human Resource Development Program (ASTHRDP).

Authors' contributions

IRC, MGQD and VSL conceived of the research. RM and MGQD served as university committees of IDA. IDA and SIA carried out phenotypic and genotypic analysis with technical support from RM, DC, GJ and FNB. IPN conducted statistical analyses, analyzed allele distributions in the 3K RGP and elite lines, performed data analysis and visualization. IPN and JDP evaluated allele/marker utility in the 3K RGP and elite breeding lines. JDP developed the IRRI 154 near-isogenic lines lacking the TKM allele. GJ and SIA conducted the evaluation of IRRI 154 near isogenic lines. IDA, IPN, and VSL interpreted the data and wrote the manuscript. All authors read and approved the manuscript.

Acknowledgments

We thank Gerald Violanta, Melencio Apostol, Celso Lantican, Esquirion Baguioso, Jun Domingo and Edgardo Coloquio for their assistance in performing the phenotyping experiments. The authors declare that GPT 5.2 was used to assist with text proofreading. The authors have subsequently reviewed, edited, and approved all the generated contents and are taking full responsibility for the contents of the publication.

Data availability

All data supporting the findings of this study are available within the paper and its Supplementary Information.

References

1. Hull R. Molecular biology of rice tungro viruses. *Annu Rev Phytopathol.* 1996;34:275–97. <https://doi.org/10.1146/annurev.phyto.34.1.275>.
2. Khan ZR, Saxena RC. Behavior and biology of *Nephotettix virescens* (Homoptera: Cicadellidae) on tungro virus-infected rice plants: Epidemiology Implications. *Environ Entomol.* 1985;14:297–304. <https://doi.org/10.1093/ee/14.3.297>.
3. Hibino H, Cabunagan R. Rice tungro-associated viruses and their relations to host plants and vector leafhoppers. JIRCAS. 1985.
4. Khush G, Angeles E, Virk P. D. Brar. Breeding rice for resistance to tungro virus at IRRI. 2004;36:101–6.
5. Kang B-C, Yeam I, Jahn MM. Genetics of plant virus resistance. *Annu Rev Phytopathol.* 2005;43:581–621. <https://doi.org/10.1146/annurev.phyto.43.011205.141140>.
6. Fraser RSS. The Genetics of resistance to plant viruses. *Annu Rev Phytopathol.* 1990;28:179–200. <https://doi.org/10.1146/annurev.py.28.090190.001143>.
7. Lee J-H, Muhsin M, Atienza GA, Kwak D-Y, Kim S-M, De Leon TB, et al. Single nucleotide polymorphisms in a gene for translation initiation factor (*eIF4G*) of Rice (*Oryza sativa*) Associated with Resistance to *Rice tungro spherical virus*. *MPMI.* 2010;23:29–38. <https://doi.org/10.1094/MPMI-23-1-0029>.
8. Macovei A, Sevilla NR, Cantos C, Jonson GB, Slamet-Loedin I, Čermák T, et al. Novel alleles of rice *eIF4G* generated by CRISPR/Cas9-targeted mutagenesis confer resistance to *Rice tungro spherical virus*. *Plant Biotechnol J.* 2018;16:1918–27. <https://doi.org/10.1111/pbi.12927>.
9. Leung H, Raghavan C, Zhou B, Oliva R, Choi IR, Lacorte V, et al. Allele mining and enhanced genetic recombination for rice breeding. *Rice.* 2015;8:34. <https://doi.org/10.1186/s12284-015-0069-y>.
10. Mansueto L, Fuentes RR, Chebotarov D, Borja FN, Detras J, Abriol-Santos JM, et al. SNP-Seek II: A resource for allele mining and analysis of big genomic data in *Oryza sativa*. *Curr Plant Biology.* 2016;7–8:16–25. <https://doi.org/10.1016/j.cpb.2016.12.003>.
11. Sharma MK, Singh S, Sengar RS. Allele mining and its Applications. *Biot Today.* 2014;4:34. <https://doi.org/10.5958/2322-0996.2014.00016.7>.
12. Mohanavel W, Ramalingam AP, Ayyenar B, Rajagopalan VR, Mohanavel V, Subburaj S, et al. Mining of candidate novel alleles using GWAS and haplotype identification for rice blast resistance. *Plant Pathol.* 2025;74:873–83. <https://doi.org/10.1111/ppa.14059>.
13. Pidon H, Chéron S, Ghesquière A, Albar L. Allele mining unlocks the identification of RYMV resistance genes and alleles in African cultivated rice. *BMC Plant Biol.* 2020;20:222. <https://doi.org/10.1186/s12870-020-02433-0>.
14. Civián P, Ali S, Batista-Navarro R, Drosou K, Ihejieta C, Chakraborty D, et al. Origin of the *Aromatic* group of cultivated rice (*Oryza sativa* L.) traced to the Indian subcontinent. *Genome Biol Evol.* 2019;11:832–43. <https://doi.org/10.1093/gbe/evz039>.

15. Azzam O, Chancellor TCB. The biology, epidemiology, and management of rice tungro disease in Asia. *Plant Dis.* 2002;86:88–100. <https://doi.org/10.1094/PDIS.2002.86.2.88>.
16. International Rice Research Institute. Standard Evaluation System for Rice (SES). In: Standard Evaluation System for Rice (SES). 4th edition. Los Baños, Philippines. p. 11.
17. Hashimoto M, Neriya Y, Yamaji Y, Namba S. Recessive resistance to plant viruses: Potential resistance genes beyond translation initiation factors. *Front Microbiol.* 2016;7. <https://doi.org/10.3389/fmicb.2016.01695>.
18. Wang A, Krishnaswamy S. Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. *Mol Plant Pathol.* 2012;13:795–803. <https://doi.org/10.1111/j.1364-3703.2012.00791.x>.
19. Widiarta IN, Firmansyah F, Yunus M, Apriana A, Sisharmini A, Santoso TJ, et al. Rice virus disease in Indonesia: epidemiology and varietal resistance. *Phytopathol Res.* 2025;7:11. <https://doi.org/10.1186/s42483-024-00302-4>.
20. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature.* 2018;557:43–9. <https://doi.org/10.1038/s41586-018-0063-9>.
21. Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature.* 2012;490:497–501. <https://doi.org/10.1038/nature11532>.
22. Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature.* 2012;490:497–501. <https://doi.org/10.1038/nature11532>.
23. Bryan GT, Wu K-S, Farrall L, Jia Y, Hershey HP, McAdams SA, et al. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. *Plant Cell.* 2000;12:2033–45. <https://doi.org/10.1105/tpc.12.11.2033>.
24. Odongo PJ, Onaga G, Ricardo O, Natsuaki KT, Alicai T, Geuten K. Insights into natural genetic resistance to rice yellow mottle virus and implications on breeding for durable resistance. *Front Plant Sci.* 2021;12:671355. <https://doi.org/10.3389/fpls.2021.671355>.
25. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods.* 2013;9:29. <https://doi.org/10.1186/1746-4811-9-29>.
26. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47:D1005–12. <https://doi.org/10.1093/nar/gky1120>.
27. Narzisi G, Schatz MC. The challenge of small-scale repeats for Indel discovery. *Front Bioeng Biotechnol.* 2015;3. <https://doi.org/10.3389/fbioe.2015.00008>.
28. Collard BCY, Mackill DJ. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Trans Royal Soc B: Biol Sci.* 2008;363:557–72. <https://doi.org/10.1098/rstb.2007.2170>.
29. Cabauatan P. Biological variants of rice tungro viruses in the Philippines. *Phytopathology.* 1995;85:77. <https://doi.org/10.1094/Phyto-85-77>.

30. Druka A, Burns T, Zhang S, Hull R. Immunological characterization of rice tungro spherical virus coat proteins and differentiation of isolates from the Philippines and India. *J Gen Virol*. 1996;77:1975–83. <https://doi.org/10.1099/0022-1317-77-8-1975>.
31. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8. <https://doi.org/10.1093/bioinformatics/btr330>.
32. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10:giab008. <https://doi.org/10.1093/gigascience/giab008>.
33. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2025.
34. Chien L-C. A rank-based normalization method with the fully adjusted full-stage procedure in genetic association studies. *PLoS ONE*. 2020;15:e0233847. <https://doi.org/10.1371/journal.pone.0233847>.
35. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. 2012;44:821–4. <https://doi.org/10.1038/ng.2310>.
36. Turner D. qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. *JOSS*. 2018;3:731. <https://doi.org/10.21105/joss.00731>.
37. Dellaporta SL, Wood J, Hicks JB. A plant DNA miniprep: Version II. *Plant Mol Biol Rep*. 1983;1:19–21. <https://doi.org/10.1007/BF02712670>.
38. Welch BL. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*. 1947;34:28–35. <https://doi.org/10.1093/biomet/34.1-2.28>.
39. Duncan DB. Multiple range and multiple F tests. *Biometrics*. 1955;11:1. <https://doi.org/10.2307/3001478>.

Figures

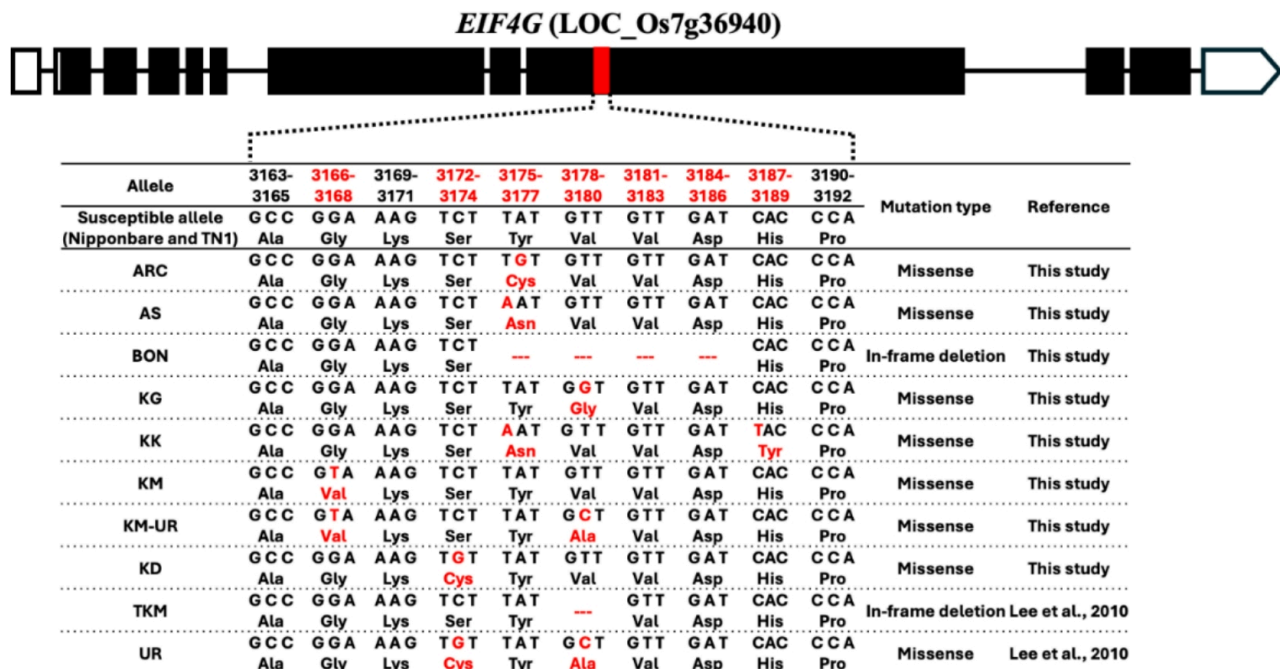


Figure 1

Allelic variation in the *EIF4G*(LOC_Os07g36940) coding region. Schematic representation of the *EIF4G* gene showing the position of the analyzed polymorphic region (highlighted in red). The table summarizes nucleotide substitutions across alleles relative to Nipponbare and TN1 (canonical susceptible allele), along with the corresponding amino acid changes. Variant positions are indicated by genomic coordinates, and nonsynonymous substitutions and deletions are highlighted in red.

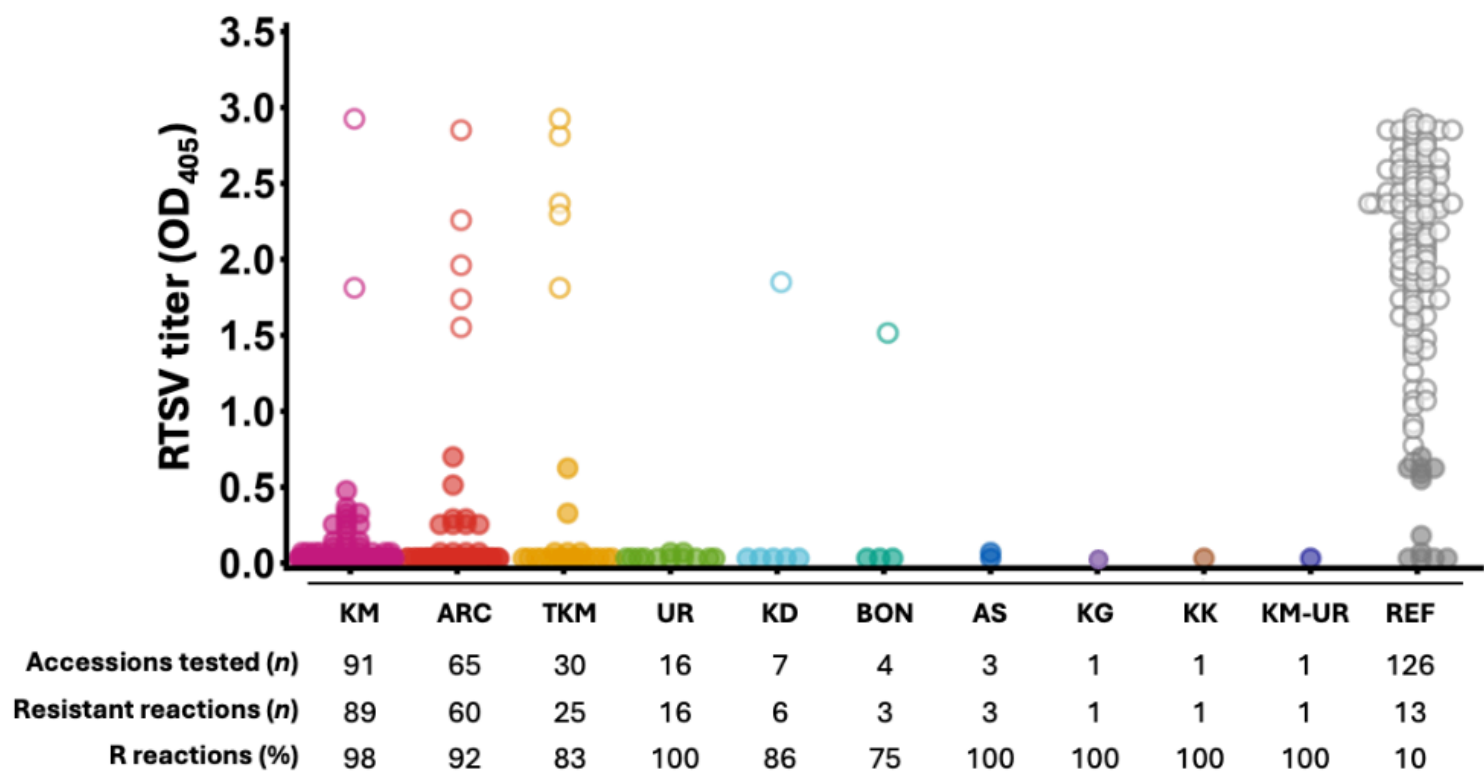


Figure 2

Distribution of RTSV titers among rice accessions carrying distinct *elf4G^{tsv1}* alleles. RTSV titer was estimated by DAS-ELISA at 14 days post inoculation and expressed as optical density at 405 nm (OD₄₀₅). Disease reaction was determined using two-tailed Welch's *t*-tests comparing each accession with both controls, followed by adjustment for multiple testing. Accessions were classified as resistant when they differed significantly from TN1 ($p \leq 0.001$) but not from TW16 ($p > 0.001$), whereas accessions were classified as susceptible when they did not differ significantly from TN1 ($p > 0.001$) but differed significantly from TW16 ($p \leq 0.001$). For accessions showing significant differences from both TN1 and TW16, classification was assigned based on the higher adjusted *p*-value, indicating greater similarity to either the resistant or susceptible control. Color-filled circles represent resistant accessions, whereas unfilled circles represent susceptible accessions. The table below the plot summarizes the number of accessions analyzed, the number classified as resistant, and the proportion of resistant reactions within each allele group.

	Alternative allele (n)	REF allele (n)	Alternative allele (%)		KM	ARC	TKM	UR	KD	BON	AS	KG	KK	KM-UR
<i>temperate japonica</i>	67	216	23.7%		67	0	0	0	0	0	0	0	0	0
<i>aromatic</i>	48	24	66.7%		0	45	1	0	0	0	2	0	0	0
<i>indicax</i>	31	583	5.0%		14	3	11	0	3	0	0	0	0	0
<i>tropical japonica</i>	24	347	6.5%		1	6	0	15	0	0	0	1	0	1
<i>indica1B</i>	18	187	8.8%		2	0	16	0	0	0	0	0	0	0
<i>admix</i>	11	88	11.1%		0	9	1	0	0	0	1	0	0	0
<i>japonicax</i>	7	74	8.6%		6	1	0	0	0	0	0	0	0	0
<i>aus</i>	5	196	2.5%		0	0	0	0	0	4	0	0	1	0
<i>indica2</i>	4	278	1.4%		0	1	1	0	2	0	0	0	0	0
<i>indica3</i>	3	472	0.6%		0	0	0	1	2	0	0	0	0	0
<i>indica1A</i>	1	208	0.5%		1	0	0	0	0	0	0	0	0	0
<i>subtropical japonica</i>	0	112	0.0%		0	0	0	0	0	0	0	0	0	0
					91	65	30	16	7	4	3	1	1	1

Figure 3

Distribution and subpopulation-specific enrichment of alternative eIF4G^{tsv1} alleles in the 3K Rice Genome Panel. Left panel indicates accessions carrying alternative and REF eIF4G^{tsv1} alleles across rice subpopulations. Values indicate the number of accessions carrying alternative alleles, the number carrying the REF allele, and the percentage of accessions with alternative alleles within each subpopulation. Right panel shows the distribution of individual alternative eIF4G^{tsv1} alleles across rice subpopulations. Numbers represent accession counts for each allele-subpopulation combination. The bottom row indicates the total number of accessions carrying each alternative allele across the 3K Rice Genome Panel.

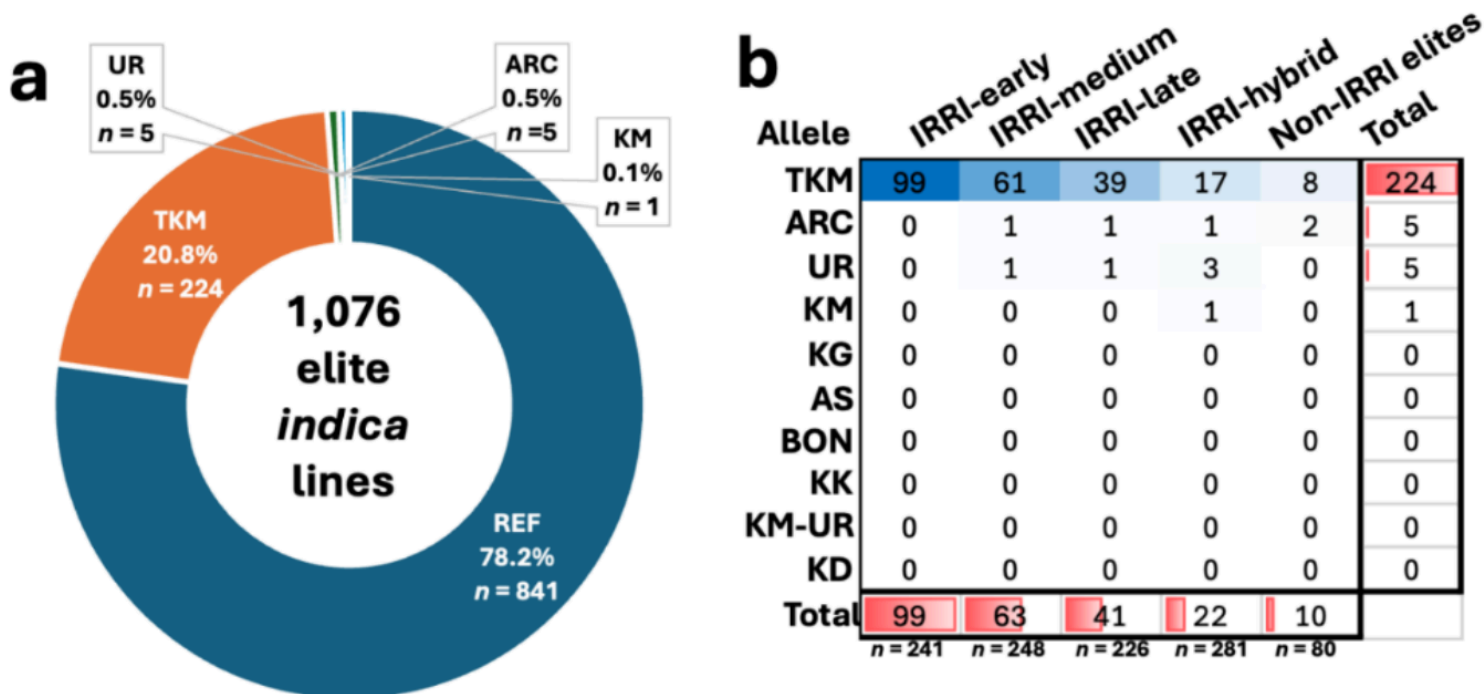


Figure 4

Distribution of $eIF4G^{tsv1}$ alleles in elite *indica* breeding lines. (a) Distribution of $eIF4G^{tsv1}$ alleles among elite indica lines. (b) Distribution of $eIF4G^{tsv1}$ alternative alleles across IRRI breeding groups. Cell values indicate the number of accessions carrying each allele within each breeding group. Totals per allele and breeding group are shown at the right and bottom margins, respectively. Values below each column (n) indicate the total number of elite lines in each breeding group.

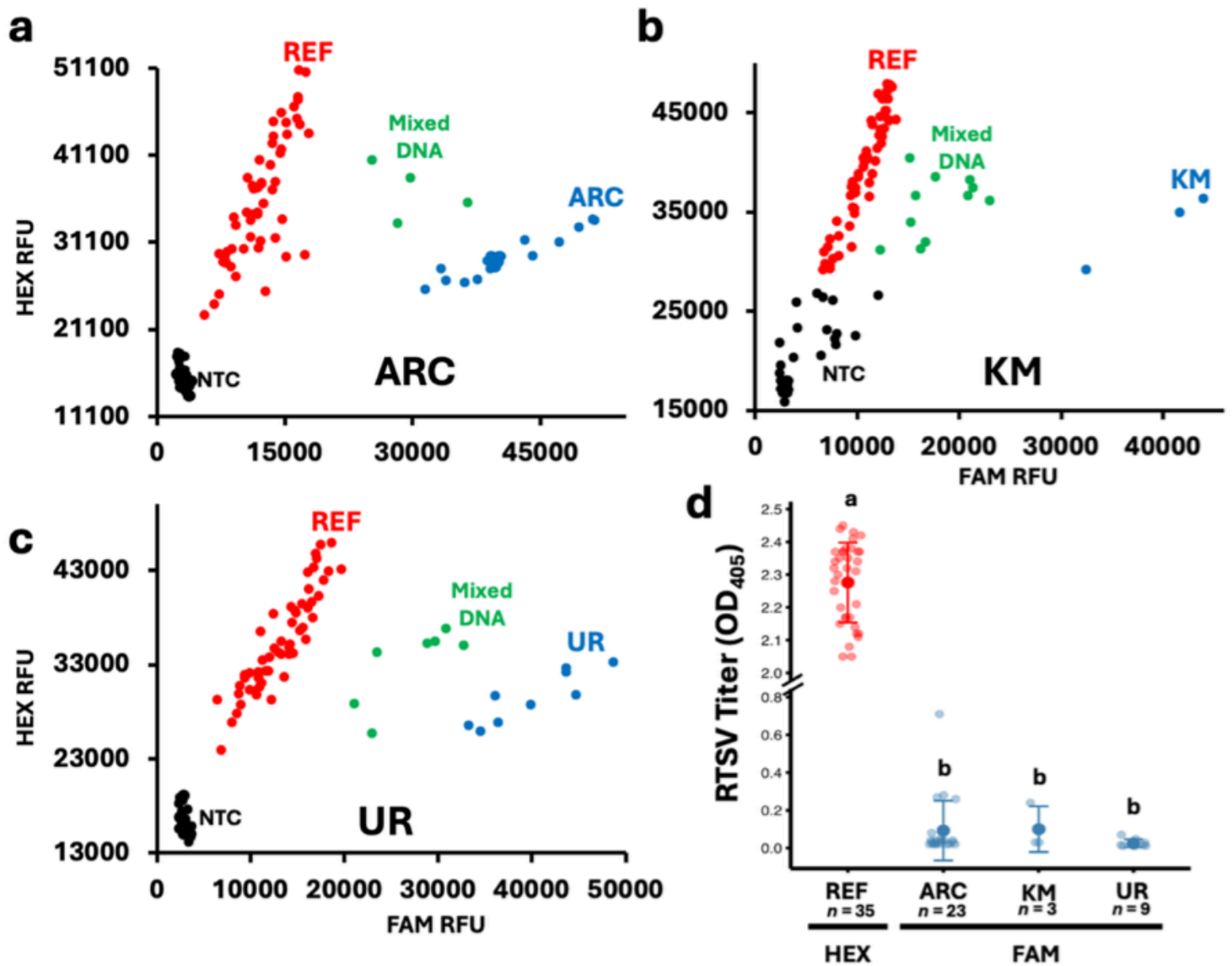


Figure 5

Genotyping using PCR allele competitive extension (PACE) and phenotyping of selected accessions carrying *eIF4G^{tsv1}* alternative alleles. (a-c) PACE clustering plots for the ARC, KM, and UR alleles. Red, blue, green, and black dots represent homozygous S allele, homozygous alternative allele, heterozygous/mixed DNA, and non-template control (NTC), respectively. (d) RTSV titer measured by DAS-ELISA in accessions carrying the REF, ARC, KM, and UR alleles. Different letters indicate significant differences among allele groups based on DMRT ($p < 0.001$). Error bars represent standard deviation. The y-axis is truncated (0.8-2.0) to improve visualization of variation among low-response groups.

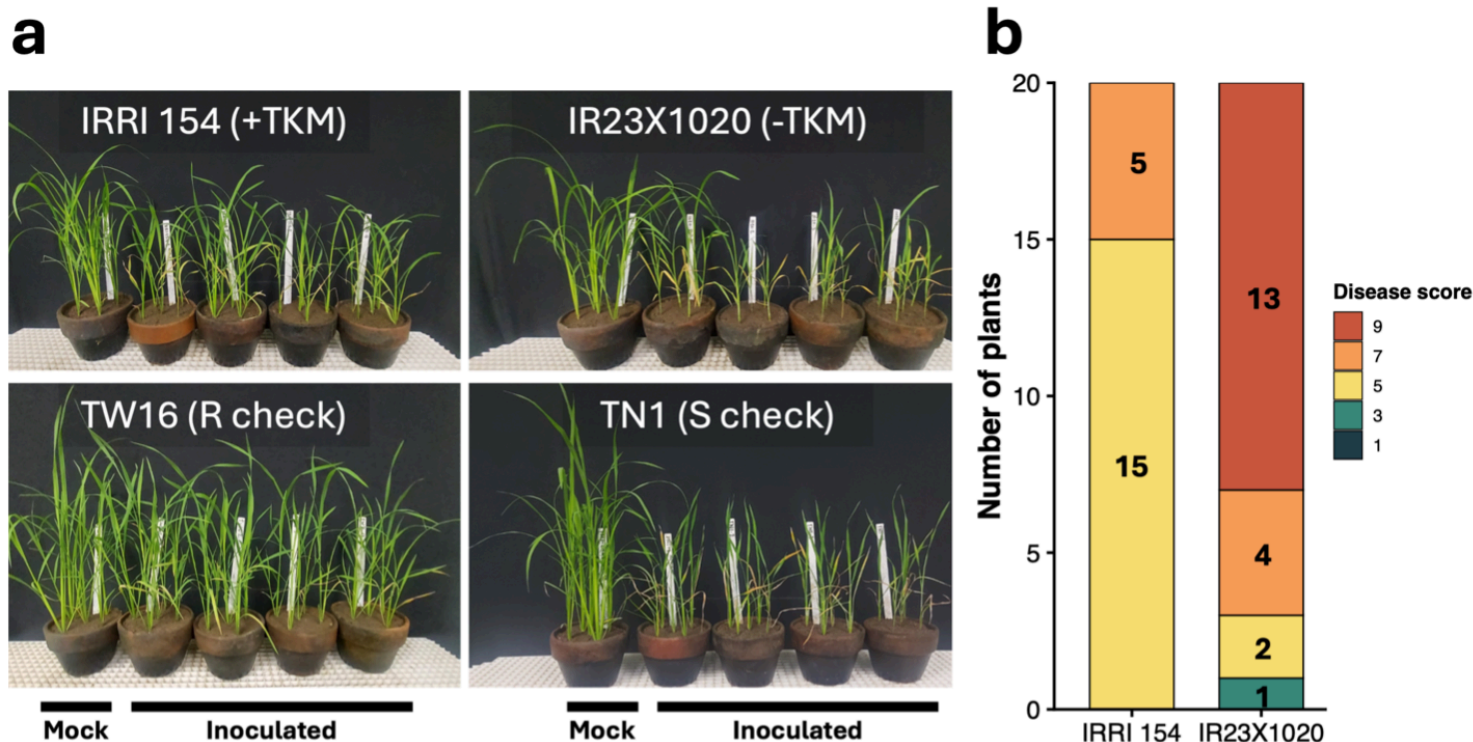


Figure 6

Phenotypic response of IRRI 154 carrying the TKM allele following tungro inoculation. (a) Representative plants of IRRI 154 (+TKM), IR23X1020 (-TKM), TW16 (resistant check), and TN1 (susceptible check) under mock and tungro-inoculated conditions at 14 day-post inoculation. **(b)** Distribution of disease severity scores among tungro-inoculated IRRI 154 (+TKM) and IR23X1020 (-TKM) plants. The numbers within each bar segment indicate the number of plants within the respective genotype that received the indicated disease severity score.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [20260829EIF4GAlleleMiningSupplementaryTablesfinal.xlsx](#)
- [SupplementalMaterials.docx](#)