Unintended effects of transgenic rice on grain yield and quality traits determined by quantitative proteomics

Yue Sun
Ministry of Education /College of Agronomy, JAU
Huizhen Chen
Ministry of Education /College of Agronomy, JAU
Zhongkai Chen
Ministry of Education /College of Agronomy, JAU
Chunlei Wang
Ministry of Education /College of Agronomy, JAU
Lu Qin
Ministry of Education /College of Agronomy, JAU
Xiaoli Lin
Ministry of Education /College of Agronomy, JAU
Yicong Cai
Ministry of Education /College of Agronomy, JAU
Dahu Zhou
Ministry of Education /College of Agronomy, JAU
Linjuan Ouyang
Ministry of Education /College of Agronomy, JAU
Changlan Zhu
Ministry of Education /College of Agronomy, JAU
Xiaosong Peng (pxs63@163.com)
Ministry of Education /College of Agronomy, JAU
Haohua He
Ministry of Education /College of Agronomy, JAU

Research Article

Keywords: Proteomics, transgenic rice, Chilo suppressalis, genetic background, unintended effect

Posted Date: May 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1660011/v1

Abstract

Omics techniques provide effective detection tools for assessing the potential impact of plant composition at the DNA, RNA, and protein levels. Among these, protein is the executor of gene function and the embodiment of biological traits, so that organisms show various genetic characteristics. Proteomics can be used to assess whether genetic engineering will lead to changes in plant traits beyond those introduced by conventional plant breeding. Here, we compare the extent of the proteome occurring in the leaves of three transgenic rice restore lines expressing *CRY1C* and *CRY2A* genes developed by genetic engineering and their corresponding recurrent parents developed by conventional breeding. *CRY1C* and *CRY2A* genes were inserted into chromosomes 11 and 12, respectively, which significantly improved the resistance of restore lines to *Chilo suppressalis*. Although differentially expressed proteins could be distinguished between transgenic rice and its recurrent parents, these differences were not sufficient to cause unintended effects on grain yield and quality traits of transgenic rice. In contrast, differences in phenotypic traits are more because of differences in genetic background. Functional cluster analysis showed that the differentially expressed proteins caused by the insertion of exogenous genes did not involve harmful metabolic pathways. The study successfully used 4D label-free quantitative proteomics technology to assess the unexpected changes in new rice varieties, and the results showed that transgenic rice did not cause unintended effects.

Introduction

Rice is one of the major food crops, and nearly half of the world population consume rice as a staple food (Zeigler & Barclay, 2008). Rice production is constrained by insect pests, among which lepidopterans such as the rice striped stem borer Chilo *suppressalis*, rice gall midges, and rice thrips can cause substantial yield losses (Lou et al., 2013;Li et al., 2020). *C. suppressalis* is a chewing insect pest that causes significant physical damage to rice plants (Hua et al., 2007). *C. suppressalis* larvae bore into the stem of rice plants and feed inside, resulting in “deadheart” at the tillering stage and “whiteheads” at the heading stage (Yue et al., 2018). Transgenic technology is often widely used in crops to introduce traits such as insect resistance, disease resistance, herbicide resistance, and heavy metal tolerance (ISAAA, 2018). Among these, Bt protein produced by *Bacillus thuringiensis* has a positive lethal effect on borers, and has been widely used as a pest-resistance gene in rice, corn, cotton, and other crops worldwide (Lou et al., 2013). To date, research on Bt rice has mainly focused on breeding new lines, introducing *CRY1Ab/Ac*, *CRY1C*, and *CRY2A* genes into indica rice CMS restore lines widely used in China by transgenic technology (Chen et al., 2005; Tang et al., 2006; Tang et al., 2011). Genetically modified (GM) crops have obvious yield advantages compared with traditional crops under the serious infestation of target pests (Jiang et al., 2013).

The environmental and food safety of GM crops has been a major concern and an important factor in delaying the commercialization of GM crops in many countries, including China (Baker et al., 2010), because of concerns about the horizontal transfer of GM ingredients in the human body after consumption of GM crops (Li et al., 2020). Potential sources of damage caused by transgenic technology can be divided into two types-intended effects and unintended effects (Ladics et al., 2015). The intended effect of transgenic plants refers to the introduction of foreign genes into the plant genome to enhance or improve plant functions, and these foreign genes can be stably expressed in the recipient plants with the desired traits (Kuiper et al., 2001). With the large-scale production of transgenic crops, many technical means have been developed for the detection of target traits. Among these, molecular methods can be used to detect specific nucleic acid sequences and the expression of foreign gene proteins (Stewart et al., 1993). Real-time polymerase chain reaction (PCR) is a qualitatively effective technique commonly used for molecular detection of nucleic acid sequences (Chou & Huang, 2010), whereas enzyme-linked immunosorbent assay (ELISA) is widely used to detect proteins expressed by foreign genes (Dong et al., 2017). The unintended effect of transgenic organisms refers to the change in the original genetic traits caused by the insertion of foreign genes (Cellini et al., 2004). Unintended effects can hardly be anticipated and are difficult to detect, raising caution when assessing the risk of transgenic technology (Wang

et al., 2018). Therefore, unintended effects need to be deeply investigated by comprehensive analyses to compare GM crops to their counterparts from parental or near-isogenic lines (Strauss et al., 2016; Fu et al., 2019).

Molecular biological detection techniques can be used to assess whether GM crops have unintended effects (Liu et al., 2020). Researchers have studied the unintended effects of transgenic technology at the molecular level using the new generation omics technologies such as transcriptomics, proteomics, and metabolomics (Steiner et al., 2013). Many researchers have evaluated the unintended effects at the proteome level in transgenic corn, rice, soybean, and other materials, and achieved positive results (Arthur et al., 2004; Wang et al., 2012; Li et al., 2013; Agapito-Tenfen et al., 2014; Chandni et al., 2015; Cho et al., 2020). These results usually reveal some differences in the transcriptome, proteome, and metabolome of plants under study (Ricroch et al., 2011; Gong and Wang, 2013; Wang et al., 2018b). However, the original genetic traits of the organism are changed because of the insertion of exogenous genes. Such unexpected effects related to transformation, location, recombination, insertion, and induction are produced by micro-effect reactions, and accurate quantification of differentiators is required to a certain extent (Miki et al., 2009; Schnell et al., 2015). Therefore, the precise identification of proteins is necessary for the study of unintended effects of transgenic crops, and an accurate proteome bioinformation system is a key prerequisite for the development of all mass spectrometry (MS) technologies. 4D label-free proteomics based on trapped ion mobility spectrometry-time-of-flight MS combined with parallel accumulation serial fragmentation can repeatedly measure the collision cross-section of all detected ions, greatly improving the coverage rate of proteomics with improved proteome accuracy (Ankney et al., 2018). On the basis of 3D proteomics, 4D proteomics adds the fourth dimension, ion mobility, which separates ions mainly according to their shape and cross-section, and can distinguish peptides with small M/Z difference, enabling the detection of low abundance protein signals (Ludwig et al., 2018). Parallel reaction monitoring (PRM) analysis can also be performed on proteins associated with phenotypic traits to determine whether the target gene insertion has an effect at the proteome level, thus changing the plant phenotypic traits (NASEM, 2016).

By serving as timely protection against the rapid evolution of resistance to Bt proteins by *C. suppressalis*, *CRY1C* and *CRY2A* can be widely used in transgenic rice breeding in China. In the present study, we compared the biological variation of protein expression levels in three transgenic rice restore lines and their corresponding recurrent parent rice varieties, and analyzed these datasets using 4D label-free quantitative proteomics technology. Based on these results, combined with genetic background analysis, the possible unintended effects of target gene insertion on rice yield and quality traits were comprehensively evaluated.

Results

Target gene insertion site analysis

We transformed the insect-resistance genes *CRY1C* and *CRY2A* into rice cultivar Minghui63 (MH63), forming three transgenic lines MH63(1C), MH63(2A), and MH63(1C+2A). We then created three transgenic restore lines in the BC_4F_5 generation, CH891(1C), CH891(2A), and CH891(1C+2A), by backcrossing with the hybrid traditional rice variety CH891. Thus, we performed two independent transgenic events using the same vectors harboring different *CRY1C* and *CRY2A* genes to elucidate the random insertion effects of targeted genomic regions (Fig. 1). The *CRY1C* flanking sequence in CH891(1C) and CH891(1C+2A) was 1276 bp, with 394 bp derived from the T-DNA region and the remaining 832 bp located on rice chromosome 11 in a stretch of the noncoding sequence region (by BLAST alignment). The *CRY2A* flanking sequence of CH891(2A) and CH891(1C+2A) was 948 bp, of which 356 bp was derived from the T-DNA region, and the remaining 543 bp of the flanking sequence was located on rice chromosome 12 by BLAST alignment. T-DNA was inserted in the 5'-noncoding region of a gene encoding an unknown functional protein (Fig. 2a).

Intended effects analysis

qRT-PCR showed that *CRY1C* and *CRY2A* genes were significantly enriched in the three transgenic lines than recurrent parents, and the concentrations of *CRY1C* and *CRY2A* genes in the three transgenic lines were significantly higher at the heading stage than at the tillering stage (Fig. 2b-d, Table S2). The content of Cry1C and Cry2A protein in stems of transgenic rice lines at the tillering and heading stages was measured by ELISA. ELISA showed that Cry1C and Cry2A proteins were significantly enriched in the three transgenic lines. A rising trend of the Cry1C protein concentration was found between vegetative growth and reproductive growth (Fig. 2e-g, Table S3). The three transgenic lines CH891(1C), CH891(2A), and CH891(1C+2A) of the rice variety BC_5F_3, with two transgenic events (*CRY1C* and *CRY2A*), exerted stronger resistance to insects than CH891 at the tillering and heading stages (Fig. 2h, j). The rate of deadheart stems was significantly less in the three transgenic lines than in CH891 at the tillering stage, and the whitehead panicles of the three transgenic lines were significantly lower than those of CH891 at the heading stage; thus, the insect-resistance genes *CRY1C* and *CRY2A* were responsible for the death of *C. suppressalis* larvae. Finally, we found that the polymeric strain CH891(1C+2A) had the best expected effect and insect resistance (Fig. 2i, k, Table S4).

Genetic background analysis

To identify the genetic background response rate, we conducted a genetic background response rate analysis using three BC_4F_5 populations. Genotyping was performed using 512 molecular markers, and the linkage map of SSR fingerprint markers was created (Fig. 3, Table S5). There were 22 polymorphic markers between CH891(1C) and its parents, and the percentage of polymorphic markers was 4.29%. There were 15 polymorphic markers between CH891(2A) and its parents, and the percentage of polymorphic markers was 2.93%. were 19 polymorphic markers between CH891(1C+2A) and its parents, and the percentage of polymorphic markers was 3.71%. The low rate of polymorphic markers indicates that the three transgenic recovered lines had a genetic background similar to their recurrent parents. The response rate of the actual genetic background of CH891(1C), CH891(2A), and CH891(1C+2A) selected was 97.85%, 98.54%, and 98.14%, respectively (Table S6). The actual genetic background response rate was higher than the theoretical genetic background response rate.

Differentially expressed proteome analysis

We analyzed the proteome of the three transgenic lines using LC–MS. Protein were extracted from isolated leaf tissue at heading for LC–MS analysis, where insect resistance is the most significant. In total, 6650 proteins were detected and identified, and 6324 of the proteins were quantified. Correlation cluster heatmap showed positive repeatability of protein data during the overall analysis of the three transgenic lines and recurrent parents (Fig. 4a). Based on our PCA results, the three transgenic lines of two transgenic events showed similar clusters of differentially abundant proteins, whereas recurrent parents were similar to the three transgenic lines. The PCA of CH891(1C) was slightly different from that of the other three lines, which might be because of the relatively low recovery rate of the genetic background (Fig. 4b). The relative standard deviation of protein quantitative values between repeated samples was relatively small, which was consistent with the correlation analysis and PCA results (Fig. 4c). Proteome heatmaps showed distinct patterns of protein abundance in four lines. All quantified proteins were compared between transgenic lines and recurrent parents, and the threshold for differential expression was padj<0.05 and $|\log_2$ fold-change$|>0$. Four clusters corresponding to the four leaf tissue samples were identified for the proteome (Fig. 4d). Overall, the number of differentially expressed proteins (DEPs) between the three transgenic lines and their recurrent parents were within the normal range of gene expression changes (Fig. 4e). Among DEPs, proteins were differentially expressed between CH891(1C) and CH891, of which 239 were upregulated and 119 were downregulated in CH891(1C) (Fig. 4f). Proteins were differentially expressed between CH891(2A) and CH891, of which 168 were upregulated and 131 were downregulated in CH891(2A) (Fig. 4g). We also found that 198 and 198 DEPs were downregulated and upregulated between CH891(1C+2A) and CH891,

respectively (Fig. 4h). This study focused on the concentration of different genes in the upper left and upper right corners of the volcano plot, and functional classification and enrichment analyses were performed for these genes.

Cluster analysis of identified proteins

We next performed Clusters of Orthologous Groups (COGs) pathway enrichment analysis of DEPs. We mainly focused on COGs enriched by significant DEPs in the three groups. The vast majority of proteins were positively correlated with transcripts, but some showed partial negative correlation. The COG pathways were mainly enriched in carbohydrate transport and metabolism, coenzyme transport and metabolism, post-translational modification, and signal transduction mechanisms. We also considered some unknown pathways (Fig. 5a, Table S7). Some genes enriched in the COG pathways were also significantly enriched in the KEGG pathways. The KEGG enrichment pathways mainly included glycine, serine, and threonine metabolism and cyanoamino acid metabolism (Fig. 5b, Table S8). Among the KEGG pathways significantly enriched in upregulation and downregulation of DEPs were the glycine, serine, and threonine metabolism and cyanoamino acid metabolism pathways, and both showed enrichment of the three transgenic lines and recurrent parents (Fig. 5c). DEPs in the glycine, serine, and threonine metabolism pathway were significantly upregulated in CH891(1C), CH891(2A), and CH891, and significantly downregulated in CH891(1C+2A) and CH891 (Fig. 5d). Most DEPs in the glycine, serine, and threonine metabolism pathway were significantly upregulated in three clusters (Fig. 5e). We identified several enzymes among the DEPs associated with the conversion reactions of glycine, serine, and threonine. Most of these enzymes were downregulated. Information on these key enzymes include enzyme symbol, EC number, and fold change. They included zinc finger matrin-type protein (ZMAT, 2.06, 1.59, 1.32), ubiquitin family protein (UBE, 4.12, 2.33, 4.18), EF-hand domain-containing protein (EFHC, 1.53, 1.56, 1.95), mitogen-activated protein kinase (MAPK, EC:2.7.11.25, 3.28, 2.40, 3.73), rhomboid-like protein (RHBD, 1.76, 2.11, 1.63), glutathione synthetase (GSS, EC:6.3.2.3, 2.45, 0.71, 2.75), C-factor (predicted) (CF, 3.69, 2.98, 1.47), and UDP-glucose 6-dehydrogenase (UG6D, EC:1.1.1.22, 3.61, 0.96, 4.99). Also upregulated were enzymes of the cyanoamino acid metabolism pathway, such as UBX domain-containing protein (PUX, 0.64, 0.75, 0.91), protease Do-like 5, chloroplast precursor (PDI, 0.64, 0.66, 0.76), plant intracellular Ras-group-related LRR protein (PIRL, 0.62, 1.01, 0.60), zeta-carotene desaturase (ZDS, 0.63, 0.85, 0.87), tyrosine-protein phosphatase (PTP, EC:3.1.3.48, 0.49, 0.58, 0.60), serine hydroxymethyltransferase (SHMT, EC:2.1.2.1, 0.54, 0.59, 0.65), nicotinate phosphoribosyltransferase (NAPRT, EC2.4.2.11, 0.51, 0.69, 0.42), and hydroxy acid dehydrogenase (PGM, EC:5.4.2.2, 0.51, 0.99, 0.74). These DEPs were significantly enriched in some basic biologically conserved COG and KEGG pathways, and belonged to constitutively expressed proteins (Table 1). Under the two insertion methods of *CRY1C* and *CRY2A genes*, the upregulation/downregulation trends of these DEPs were the same between the three transgenic lines and recurrent parents, and the changes at the proteome level were universal and representative, and could be used for subsequent evaluation of unexpected effects.

PRM validation

For proteome data validation, we selected genes encoding fifteen DEPs (MAPK, UBE, EFHC, SHMT, RHBD, GSS, CF, UG6D, PUX, PDI, PIRL, ZDS, PTP, NAPRT, and PGM) from the proteome data screened by cluster analysis. We performed PRM analysis on peptides representing these fifteen DEPs that were successfully quantified in the proteomic work (Table 5). Compared with the recurrent parents, ZMAT, UBE, EFHC, MAPK, RHBD, GSS, CF, and UG6D were significantly upregulated, and PUX, PDI, PIRL, ZDS, PTP, SHMT, NAPRT, and PGM were significantly lower in the three transgenic lines. On the other hand, except for ZDS, no significant difference was observed in DEPs among the three transgenic lines. The PRM results also correlated well with the proteomics data (Fig. 6, Table 2). Therefore, the proteome data were reasonable and reliable.

Grain yield and quality performance

CH891(1C), CH891(2A), and CH891(1C+2A) restore lines were planted in the experimental plots to evaluate agronomic traits (Table S9). Yield traits included panicles per plant, grains per panicle, 1000-grain weight, and yield per plant. Compared with the recurrent line CH891, CH891(1C), CH891(2A), and CH891(1C+2A) showed no significant difference (Fig. 7a-f). However, significant differences were found in panicle length and seed set rate between CH891(1C) and CH891. With the exception of gel consistency, significant differences were observed in other quality traits. Significant differences were found in brown rice rate, head rice rate, and chalkiness degree between CH891(1C) and CH891, while significant differences were observed in grain length to width ratio and amylose content between CH891(2A) and CH891. Significant differences were observed in amylose content between CH891(1C+2A) and CH891 (Fig. 7g-l). Previous experiments showed that no significant difference was observed in DEPs among protein groups. Therefore, the difference in grain yield and quality traits among different varieties was not caused by the change in protein composition affected by the insertion of target genes. However, the difference might be determined by the genetic background of the remaining donors in the backcross. In addition, we compared the relationship between donor chromosome fragments and agronomic traits. CH891(1C) had a polymorphism at the RM1026 locus on chromosome 9, and the *qPL9* gene related to panicle length was reported near the band locus. CH891(1C) had a polymorphism at the RM538 locus on chromosome 5, and the *PTB1* gene related to seed set rate was reported near the band locus. CH891(1C) had a polymorphism at the RM17804 locus on chromosome 5, and the *Chalk5* gene related to head rice rate and chalkiness degree was reported near the band locus. CH891(2A) had polymorphisms at RM289 and RM1364 loci on chromosomes 5 and 7, respectively, and the *GL5, GW5, GL7* genes related to grain length to width ratio was reported near the band locus. CH891(2A) and CH891(1C+2A) had a polymorphism at the RM190 locus on chromosome 6, and the *Wx* gene related to amylose content was reported near the band locus (Table 3).

Discussion

Scientific research on transgenic insect-resistant rice has been performed in China for nearly 20 years. *CRY1Ab/1Ac*, *CRY1C*, and *CRY2A* are widely used Bt insect-resistance genes. Breeding units mainly use them to obtain high-generation derivative lines and to improve better agronomic characters while obtaining insect resistance (Jiang et al., 2016). However, most studies focused on the expected and unexpected effects caused by random integration of undesirable genomic regions caused by *Agrobacterium*-mediated transformation in the transgenic process. These effects often occur in the initial stage of transgenic donor parents, without evaluation of the high-generation lines in later backcross selection, which are often the lines ultimately used (Strauss et al., 2016). Therefore, we tested whether transgenic technology in rice would have new or larger unexpected effects on conventional hybrid crops, using three transgenic lines of BC_5F_3 and one recurrent parent line.

In CH891 as the genetic background, *CRY1C* and *CRY2A* genes could be effectively and stably expressed in both transcription and translation stages to improve and maintain insect resistance. However, the expression level and resistance of different Bt proteins in the same genetic background are often different. Resistance of all Bt Bollgard cotton lines was inconsistent (Adamczyk and Sumerford. 2004). Cry1C and Cry2A proteins were also expressed differently in different rice lines. However, all transgenic lines carrying Cry1C or Cry2A exhibited nearly 100% resistance to C. suppressalis (Chen et al., 2005). The higher expression of Cry1C and Cry2A proteins in leaves than in stems can be explained by spatial differences in the plant itself, as leaves are the more susceptible part of the plant (Fearing et al., 1997). The higher expression of Cry1C and Cry2A proteins at the heading stage than at the tillering stage can be explained by differences in physiological development in the plant itself, because the heading stage was in the reproductive development stage with the most vigorous metabolism (Ye et al., 2010). Therefore, in our study, *CRY1C* and *CRY2A* genes were stably expressed in high-generation backcross lines, and effective resistance to *C. suppressalis* was generated at different insertion sites.

Genomics, transcriptomics, and proteomics have been widely used to assess the effects of transgenic technology on crop breeding at the genetic background, mRNA, and protein levels. However, the presence of residual fragments of donor genetic background in high-generation lines may also lead to unintended effects (Batista et al., 2008; Montero et al., 2011; Tan et al., 2019). Thus, more accurate methods are needed to measure omics data. 4D proteomics has the advantages of traditional 3D retention time, mass charge ratio (M/Z), ion intensity, high scanning speed, and high sensitivity, plus the collision cross-sectional area of the fourth dimension peptide – ion mobility separation, and data collection without compromising on window cycle speed. By reducing the complexity of the spectrum and increasing the ion utilization rate, proteomics has achieved a comprehensive improvement in the coverage depth, sensitivity, and flux (Ying et al., 2019). The correlation analysis and PCA of the original dataset in this study showed that the transgenic line samples had similar groups with the recurrent parents at the proteome level, whereas CH891(1C) deviates slightly. The comprehensive genetic background analysis showed that when the correlation within the group was high, differences between groups amplified in the PCA diagram, indicating the interference of a certain genetic background. This result was consistent with previous studies. There are some differences in protein components and characteristics among different transgenic lines (Jung et al., 2013; Hu et al., 2014; Zhang et al., 2019). Previous studies found no difference between transgenic and non-transgenic expression products at the transcriptome or proteome level (Kogel et al., 2010). Similar results were also found in common bean. In terms of leaf proteome, the similarity between EMBRAPA and its non-transgenic near-isogenic line was higher than that between two common bean varieties (Valentim-Neto et al., 2016). We detected many DEPs between transgenic lines and recurrent parents. COG and KEGG enrichment analysis showed that DEPs in different comparisons were enriched in some constitutive metabolic pathways, mainly including carbohydrate transport and metabolism, coenzyme transport and metabolism, post-translational modification, and signal transduction mechanisms. Previous studies used four transgenic lines with insect resistance. The crystal proteins expressed in plants are heterologous with no metabolic activity in rice plants (Fu et al., 2019). However, if the inserted genes in transgenic plants were involved in plant metabolic pathways, they may cause plant phenotypic changes (Wang et al., 2019). The results of this study were consistent with those of previous studies. Compared with traditional cross breeding, transgenic plants have no unique influence on plant metabolic pathways.

PRM validation of 15 representative intra-group DEPs in the functional enrichment pathway found that most DEPs did not show significant differences among the three transgenic lines, indicating that the change in the proteome between transgenic lines would not lead to changes in grain yield and quality traits. Therefore, we attempted to analyze the relationship between polymorphic SSR markers and yield and quality traits to discuss the impact of the residue of genetic background of donor parents on unexpected effects. *qPL9* was found to affect spike length in rice. An unsubstituted fragment of MH63 was found in CH891(1C), and *qPL9* was linked to the polymorphic SSR markers RM1026 (Lian et al., 2019). *PTB1*, a ring E3 ubiquitin ligase, positively regulates the seed setting rate of rice ears by promoting pollen tube growth. An unsubstituted fragment of MH63 was found in CH891(1C + 2A), and *PTB1* was linked to the polymorphic SSR marker RM538 (Li et al., 2013). *Chalk5* encodes vacuolar H + translocation pyrophosphatase (V-PPase), and increased *Chalk5* expression increases endosperm chalkiness, possibly by disrupting pH during seed development and thereby affecting protein synthesis by the transport system, resulting in a large increase in small vesicle-like structures and chalky grain formation. An unsubstituted fragment of MH63 was found in CH891(1C), and *Chalk5* was linked to the polymorphic SSR marker RM17804 (Li et al., 2014). GW5 protein is a positive regulator of brassinolide signal transduction, which can physically interact with glycogen synthase kinase GSK2 and inhibit the activity of GSK2 kinase, regulate the expression level of brassinolide response gene, and affect grain width. An unsubstituted fragment of MH63 was found in CH891(2A), and *GW5* was linked to the polymorphic SSR marker RM289 (Liu et al., 2017). *GL7* encodes a homolog of LONGIFOLIA protein in Arabidopsis thaliana, which increases longitudinal cell division and decreases transverse cell division in grains and regulates longitudinal cell elongation. An unsubstituted fragment of MH63 was found in CH891(2A), and *GL7* was linked to the polymorphic SSR marker RM1364 (Wang et al., 2015). *Wx* encodes granule-bound starch synthesis protein. An unsubstituted fragment of MH63 was found in

CH891(1C + 2A), and *Wx* was linked to the polymorphic SSR marker RM190 (Yan et al., 2011). Yield traits (with phenotypic easy selection) were almost unchanged, whereas quality traits (with phenotypic hard observation) were changed more frequently, suggesting that the genetic background interfered with the evaluation of unintended effects of transgenes at the proteomic level in high-generation backcross lines. Our study showed that the internal difference in genetic background is much larger than the variation of plant proteome caused by the introduction of foreign genes by transgenic technology or cross breeding.

Materials And Methods

Plant materials

Three transgenic Bt rice restore lines CH891(1C), CH891(2A), and CH891(1C+2A), and their non-Bt counterpart Changhui 891 were used in this study. Among these varieties, Changhui 891 is a backcross line. The seeds for these lines were independently cultivated at the Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Jiangxi Agricultural University, Nanchang, China.

Experimental design

Field experiments were performed in May–October, 2021, in the economic and technological development zone (28˚48'10" N, 115˚49'55" E), Nanchang City, Jiangxi Province, China. The mean monthly day and night temperature during the rice growing season are shown in Table 1. The three transgenic Bt rice lines and the control line Changhui 891 were used for field evaluations. The field layout followed a randomized block design with three replications. The size of each plot was 5 m × 5 m. Twenty-day-old seedlings were transplanted at a density of 15 cm × 20 cm with one seedling per hill. The soil type of the experimental site was reddish-yellow clay-like paddy soil. Soil in the upper 15 cm at the test site had following properties at the beginning of the experiment: pH 5.01, 1.26 g kg^{-1} total N, 105.6 mg kg^{-1} available phosphorus, 125.2 mg kg^{-1} potassium, and 20.56 g kg^{-1} organic matter. The experimental field was kept flooded from transplanting until 7 days before maturity. Pests, diseases, and weeds were intensively controlled for all treatments to avoid yield losses.

Detection of target gene insertion sites and expression

DNA was extracted by the CTAB method. DNA was subjected to HindIII(Qut) and EcoII(Qut) restriction endonucleases in a 50-μL reaction system for 2 h, and the obtained fragments were cyclized in a T4 DNA ligase 50-μL reaction system. The obtained products were amplified by nested PCR, and primers used for the first and second rounds of nested PCR (Table S1). The first round of nested PCR included pre-denaturation at 94°C for 5 min; denaturation at 94°C for 1 min, denaturation at 55°C for 1 min, extension at 72°C for 2 min, 32 cycles, and final elongation at 72°C for 5 min. The second round of PCR included pre-denaturation at 94°C for 5 min, denaturation at 94°C for 1 min, annealing at 57°C for 1 min, extension at 72°C for 2 min, 30 cycles, and final extension at 72°C for 5 min, and maintained at 4°C. The target sequences were enriched and recovered by 1% agarose gel electrophoresis and sent to Tsingke Biotechnology Company, Hunan Province, for sequencing.

Total RNA was extracted from tissues at various developmental stages by grinding in TRIzol (Merck, KGaA, Germany). DNase digestion was performed to avoid contamination from genomic DNA, and the phenol–chloroform method was used to isolate total RNA. The integrity of the extracted RNA was determined by 1.5% agarose gel electrophoresis, and RNA quantity and quality were measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) based on the 260/280-nm and 260/230-nm absorbance ratios. Complementary DNA was synthesized using a PrimeScript 1st Strand cDNA Synthesis Kit (TaKaRa, 6210A, Japan) according to the instructions of the PrimeScript RT Master Mix Kit. After ten-fold dilution of the cDNA, the target genes *CRY1C* and *CRY2A* and the

reference gene Actin1 were detected by quantitative real-time PCR (qRT-PCR) according to the instructions of the SYBR Premix Taq II Kit (TaKaRa) (Table S1). The program of the 7500 Real-Time PCR system (Thermo Fisher Scientific) was incubation at 95°C for 2 min, followed by 40 cycles of 95°C for 15 s and 68°C for 30 s, and a final extension step at 68°C for 10 min (Table 3). The purity of the amplicons was confirmed in the presence of a single peak in the melting curve (Agostinetto et al., 2019). Osactin1 was used as the internal reference gene for qRT-PCR normalization, and the qRT-PCR results were analyzed by the $2^{-\Delta\Delta CT}$ method.

The amount of Cry1C and Cry2A proteins in leaves, stems, and panicles was measured at the tillering, booting, heading, filling, and maturity stages using an ELISA kit (AP003 and AP005 CRBS; EnviroLogix Inc., Portland, ME, USA). The absorbance value was measured at 450 nm using a VICTOR Nivo multimode plate reader (PerkinElmer, Waltham, MA, USA). Based on the range of the standard curve, the Cry1C and Cry2A protein extract was diluted appropriately so that its absorbance value was within the range of the standard curve (Xu et al., 2018). For ELISA, a standard curve was drawn based on the absorbance of known concentrations of Cry1C and Cry2A standard (AP003 and AP005; EnviroLogix). The concentration of each test sample was determined from the standard curve, and the Cry1C and Cry2A protein content of the sample was calculated based on its dilution ratio and the conversion formula: Cry1C and Cry2A protein content (μg g^{-1} fresh weight) = test sample concentration (ng g^{-1}) × dilution × extract volume/tissue fresh weight (mg).

Evaluation of *C. suppressalis* resistance

Thirty consecutive plants in each plot were investigated for insect resistance of the three rice lines. Deadheart stems caused by stem borers were counted at the late maximum tillering stage. Damaged leaves with visible scrapes or folds caused by leaffolders were counted within five days after peak damage appeared. Whitehead panicles were counted at the heading stage.

Genetic background detection based on SSR markers

A total of 512 SSR primers covering the whole genome of rice were used to screen the whole genome of the three transgenic lines and their corresponding recurrent parents, and the recovery rate of genetic background was analyzed. The CTAB method was used to extract DNA. SSR primers were synthesized by Shanghai Shenggong Biotechnology Company based on the sequence designed by China Rice Research Institute. The PCR system was 15 μL, including 10 × PCR buffer 1.5 μL, 100 ng•μL^{-1} DNA template 2 μL, 2.5 mmol• L^{-1} dNTPs 0.3 μL, ddH_2O 10 μL, 10 μmol• L^{-1} positive and negative primers 0.5, and 5U•μL^{-1} Taq DNA polymerase 0.2 μL. The PCR procedure was the same as above. The PCR products were observed after 8% polyacrylamide gel electrophoresis and rapid silver staining. The single plant of homozygous band type was marked as 1, the single plant of heterozygous band type was marked as 2, and the single plant of missing band type was marked as 0.

Protein extraction and liquid chromatography (LC)-MS/MS quantitative proteomics

Leaf samples from three transgenic lines and recurrent parents at the heading stage were thoroughly ground to powder in liquid nitrogen at −80°C. Phenol extraction buffer (containing 10 mM dithiothreitol, 1% protease inhibitor), four times the volume of the powder, was added to the samples and lysed by ultrasonication. An equal volume of Tris-equilibrated phenol was added, centrifuged at 5500 g for 10 min at 4°C, and five times the volume of 0.1 M ammonium acetate/methanol was added to the supernatant for overnight precipitation. The protein precipitate was washed with methanol and acetone. The final precipitation was reconstituted with 8 M urea, and the protein concentration was determined using the BCA kit (Merck, B9643, Germany) . Equal amounts of each sample protein were taken for enzymatic hydrolysis, and the volumes were adjusted to the same volume with lysis buffer. A final concentration of 20% trichloroacetic acid was slowly added, mixed by vortex, and precipitated at 4°C for 2 h, followed by centrifugation at 4500 g for 5 min. the supernatant was discarded, and the pellet was washed 2–3 times with precooled acetone. After

drying the precipitate, TEAB was added to a final concentration of 200 mM, the precipitate was ultrasonically dispersed, and trypsin was added at a ratio 1:50 (protease:protein, m/m) for overnight enzymolysis. Dithiothreitol was added to a final concentration of 5 mM and reduced at 56°C for 30 min. Iodoacetamide was added to a final concentration of 11 mM and incubated at room temperature for 15 min in the dark.

After separation by ultra-high performance liquid phase system, the peptide was injected into the ion source for ionization using the Orbitrap Exploris™ 480 mass spectrometer (Thermo Fisher Scientific). The ion source voltage was set to 2.3 kV, and the FAIMS compensation voltage was set to −45 V, −65 V. The peptide parent ions and their secondary fragments were detected and analyzed by high-resolution Orbitrap. The scanning range of primary MS was set to 400–1200 m/Z and the scanning resolution was set to 60000. The fixed start point of the secondary mass spectrum scanning range was 110 m/Z, the secondary scanning resolution was set to 15000, and TurboTMT was turned off. According to the sequence of HCD-1.0 ion fragmentation, the secondary ion fragmentation mode was used for analysis. To improve the effective utilization of MS, the automatic gain control (AGC) was set to 100%, the signal threshold was set to 5E4 ions/s, the maximum implantation time was set to auto, and the dynamic exclusion time of tandem MS (MS/MS) scanning was set to 20 s to avoid repeated scanning of parent ions. The MS proteomics data are available at the ProteomeXchange Consortium via the PRIDE partner repository (https://www.ebi.ac.uk/pride/) with the dataset identifier PXD033443.

PRM analysis

Protein extraction was performed in line with the 4D label-free proteome method. The tryptic peptides were dissolved in 0.1% formic acid (solvent A), directly loaded onto a customized reversed-phase analytical column. The gradient increased from 6% to 23% solvent B (0.1% formic acid in 98% acetonitrile) over 38 min, from 23% to 35% in 14 min, climbed to 80% in 4 min, and stayed at 80% for the last 4 min, all at a constant flow rate of 700 nL/min on an EASY-nLC 1000 ultra-performance LC (UPLC) system (Thermo Fisher Scientific). The peptides were subjected to NSI source, followed by MS/MS in Q ExactiveTM Plus (Thermo Fisher Scientific) with online UPLC. The electrospray voltage applied was 2.0 kV. The m/z scan range was 350 to 1000 for full scan, and intact peptides were detected in the Orbitrap at 35,000 resolution. Peptides were then selected for MS/MS using NCE setting as 27, and the fragments were detected in the Orbitrap at 17,500 resolution. A data-independent procedure that alternated between one MS scan followed by 20 MS/MS scans was followed. AGC was set at 3E6 for MS and 1E5 for MS/MS. The maximum IT was set at 20 MS for full MS and auto for MS/MS. The isolation window for MS/MS was set at 2.0 m/z. The resulting MS data were processed using Skyline (v.3.6). Peptide settings: enzyme was set as trypsin [KR/P], maximum missed cleavage was set as 2, peptide length was set as 8–25, variable modification was set as carbamidomethyl on Cys and oxidation on Met, and maximum variable modification was set as 3. Transition settings: precursor charges were set at 2, 3, ion charges were set at 1, 2, ion types were set at b, y, p. The product ions were set from ion 3 to last ion, the ion match tolerance was set as 0.02 Da. After normalizing the quantitative information by the heavy isotope-labeled peptide, a relative quantitative analysis (three biological replicates) was performed on the target peptides.

Assays for grain yield and quality traits

Transgenic rice lines were planted in paddy fields at the Transgenic Experimental Plots of Jiangxi Agricultural University (Nanchang, Jiangxi, China) to evaluate the agronomic performance. Non-transgenic control line CH891 was planted in paddy fields adjacent to the transgenic lines. Six blocks of 6 m2 (2 m × 3 m) were randomly chosen for the evaluation. Each block contained about 100 plants, and each plant had 10–15 tillers. Five yield traits and seven quality traits were measured, including panicles per plant, grains per panicle, 100-grain weight, seed set rate, and yield per plant. After maturity, all plants from each plot were harvested for measurements of brown rice rate, head rice rate, grain length to width ratio, chalky kernel percentage, chalkiness degree, gel consistency, and amylose content.

Data analysis

Vector NTI Suite 8 software (Invitrogen Corp., Carlsbad, CA, USA) was used to compare the flanking sequence with the sequence on the transformed Vector PPZP201-rubisk-BT to determine whether the obtained sequence was the target fragment and remove the part identical with the Vector sequence. Sequencing results were analyzed for similarity using the NCBI database (http://last.ncbi.nlm.nih.gov). The GRAMENE library (http://www.gramene.org) was used to retrieve and analyze the rice genome sequence. After PCR band type sorting, data was inputted into a Microsoft Excel 2007 spreadsheet (Microsoft Corp., Redmond, WA, USA) for data processing, and CASS2.1 software was used to draw linkage maps. The regression rate of the recurrent parent background was calculated by the formula $E[G(g)] = 1 - (½)^{g+1}$. The common formula $G(g) = [L + X(g)] / 2L$ was used to calculate the response rate of the genetic background in actual analysis, where G(g) represented the response rate of the genetic background in g generation. g represents the number of generations used for backcrossing; L represents the number of molecular markers involved in analysis; X(g) represents the number of band markers in the backcross g generation, which were the same as those of recurrent parents.

The protein data with quantitative values in all samples were selected for dimensionality reduction. Data were first transformed by $\log_2$, and the mean value was subtracted. Then the PRCOMP function in R was used for principal component analysis (PCA). Protein annotation mainly included Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation and subcellular localization. KEGG database was used to annotate the protein pathways. First, BLAST comparison (blastp, evalue $\leq 1e^{-4}$) was performed on the identified proteins. Based on the BLAST comparison results, sequences with the highest scores were selected for annotation. Hierarchical clustering was then performed based on differentially expressed protein functional classification (e.g., GO, Domain, Pathway, Complex). All enriched categories along with their *P*-values were first collated, then those categories that were enriched in at least one of the clusters with $P < 0.05$ were filtered. This filtered *P*-value matrix was transformed using the function $x = -\log_{10}(P\text{-value})$. KEGG online service tools KAAS (https://www.genome.jp/tools/kaas/) was used to annotate the KEGG database description of proteins. Finally, the annotation results of the KEGG database pathways were mapped using KEGG online service tools KEGG mapper. These pathways were classified into hierarchical categories using the KEGG website. In the experimental design, more than two unique peptides were used for quantitative analysis of each protein (only one unique peptide was suitable for PRM verification for some proteins), and only one peptide was identified for some proteins due to sensitivity and other reasons. After normalizing the quantitative information by the heavy isotope-labeled peptide, a relative quantitative analysis (three biological replicates) was performed on the target proteins. The protein relative expression levels were processed and analyzed using Excel 2007 and SPSS 16.00 (IBM Corp., Armonk, NY, USA). Agronomic traits of transgenic plants were compared with the recurrent parents using one-way analysis of variance (ANOVA). Values were presented as means (±SD). The borer mortality rate data were processed and analyzed using Excel 2007 and SPSS 16.00. Data were analyzed by one-way ANOVA, and treatment means were compared using the least significant difference test at $P = 0.05$. Figures were constructed using Origin 2017 (OriginLab Corp., Northampton, MA, USA).

Conclusion

We successfully used 4D label-free quantitative proteomics technology to investigate the changes in protein expression by functional clustering in three transgenic lines and recurrent parents. The results show that the inserted *CRY1C* and *CRY2A* genes were inherited stably in higher generations, and the newly bred transgenic restored lines showed high insect resistance and superior agronomic traits, which can lead to changes in the proteome. Moreover, no negative or unintended effects of proteomic changes were observed on grain yield and quality traits of these transgenic lines. Thus, this new omics technology can provide an effective detection method for identifying the unintended effects of transgenic varieties. We also found that the residual fragments of donor genetic background during backcross selection

may have more influence on the agronomic traits of transgenic varieties. Therefore, it is necessary to establish a comprehensive evaluation system of unintended effects at the multiomics analysis level of commercial variety selection before the analysis of unintended effects considered in this study.

Declarations

Acknowledgments

We gratefully acknowledge Jingjie PTM BioLab Co. Ltd. (Hangzhou, China) for help with the mass spectrometry analysis. We would like to thank TopEdit (www.topeditsci.com) for linguistic assistance during the preparation of this manuscript.

Funding

This research was supported by the Major Project supported by the Major Project of Jiangxi Provincial Department of Science and Technology (S2016NYZPF0256), the Major Project supported by Breeding New Varieties of the National Transgene of China (2016ZX08001-001-005), Jiangxi Provincial Natural Science Foundation of China (20202BAB215001).

Authors' Contribution

HH and XP designed the experiments and drafted the manuscript. YS and HC wrote the manuscript. ZC, CW, LQ, XL, YC and DZ participated in phenotype measurement. YS, ZC and LO performed the data analyses. YS, CZ, XP and HH participated in the revision process. All authors read and approved the final article.

Availability of supporting data and materials

The original contributions presented in the study are publicly available. The mass spectrometry proteomics data are available at the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD033443. The data sets supporting the results of this article are included within the article and its additional files.

Declarations

Ethics approval and consent to participate

(Not applicable)

Consent for publication

All authors reviewed the manuscript and agreed to publish it.

Competing interests

All authors declare that they have no confict of interest.

Authors' information

[1] Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education /College of Agronomy, JAU, Nanchang 330045, Jiangxi, China;

[2] Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China.

References

1. Adamczyk, J. J. , & Meredith, W. R. 2004 Breeding and genetics genetic basis for variability of Cry1Ac expression among commercial transgenic bacillus thuringiensis (Bt) cotton cultivars in the united states. Journal of Cotton ence 8(1)
2. Agapito-Tenfen, S. , Vilperte, V. , Benevenuto, R. , Rover, C. , Traavik, T. , & Nodari, R (2014) Effect of stacking insecticidal cry and herbicide tolerance epsps transgenes on transgenic maize proteome. Bmc Plant Biology 14(1):346
3. Ankney, J.A., Muneer, A., and Chen, X (2018) Relative and absolute quantitation in mass spectrometry-based proteomics. Annual Review of Analytical Chemistry 11(1):49–77
4. Arthur, J. W. , & Wilkins, M. R (2004) Using proteomics to mine genome sequences. Journal of Proteome Research 3(3):393–402
5. Baker, J. M. , Hawkins, N. D. , Ward, J. L. , Lovegrove, A. , Napier, J. A. , & Shewry, P. R (2010) A metabolomics study of substantial equivalence of feld-grown genetically modifed wheat. Plant Biotechnology Journal 4(4):381–392
6. Batista, R., Saibo, N., Lourenco, T. and Oliveira, M.M (2008) Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. Proc. Natl. Acad. Sci. USA 105:3640–3645
7. Cellini, F. , A C hesson, Colquhoun, I. , A C onstable, Davies, H. V. , & Engel, K. H (2004) Unintended effects and their detection in genetically modifed crops. Food Chem Toxicology 42(7):1089–1125.
8. Chandni, M. , Kathuria, P. C. , Pushpa, D. , Singh, A. B. , & Manoj, P (2015) Lack of detectable allergenicity in genetically modified maize containing "cry" proteins as compared to native maize based on in silico & in vitro analysis. Plos One 10(2):e0117340
9. Chen, H., Tang, W., Xu, C., Li, X., Lin, Y., Zhang, Q. 2005 Transgenic indica rice plants harboring a synthetic cry2A* gene of Bacillus thuringiensis exhibit enhanced resistance against lepidopteran rice pests. Theoretical and Applied Genetics 111(7):1330–1337
10. Cho, J. I. , Park, S. H. , Lee, G. S. , Kim, S. M. , & Park, S. C (2020) Current status of gm crop development and commercialization. Korean Journal of Breeding Science 52(Special Issue):40–48
11. Chou, J., & Huang, Y 2010 Differential expression of thaumatin-like proteins in sorghum infested with greenbugs. Zeitschrift Für Naturforschung C Journal of Biosciences 65(3–4):271–276
12. Dong, S., Zhang, X., Liu, Y., Zhang, C., Xie, Y., & Zhong, J., Xu, C., & Liu, X (2017) Establishment of a sandwich enzyme-linked immunosorbent assay for specific detection of Bacillus thuringiensis (Bt) cry1ab toxin utilizing a monoclonal antibody produced with a novel hapten designed with molecular model. Analytical & Bioanalytical Chemistry 409(8):1985–1994
13. Fearing, P.L., Brown, D., Vlachos, D., Meghji, M., Privalle, L. 1997 Quantitative analysis of CryIA (b) expression in Bt maize plants, tissues, and silage and stability of expression over successive generations. Mol. Breed. 3(3):169–176
14. Fu, W., Wang, C., Xu, W. & Zhu, P (2019) Unintended effects of transgenic rice revealed by transcriptome and metabolism. GM Crops Food 10, 20–34
15. Glen, A., Evans, C. A., Gan, C. S., Cross, S. S., Hamdy, F. C., Gibbins, J (2010) Eight-Plex iTRAQ analysis of variant metastatic human prostate cancer cells identifies candidate biomarkers of progression: an exploratory study. Prostate 70:1313–1332.
16. Gong, C.Y. and Wang, T (2013) Proteomic evaluation of genetically modified crops: current status and challenges. Frontiers in Plant Science 4:41
17. Hua, H., Lu, Q., Cai, M., Xu, C., Zhou, D. X., & Li, X., & Zhang, Q (2007) Analysis of rice genes induced by striped stem borer (Chilo suppressalis) attack identified a promoter fragment highly specifically responsive to insect feeding.

Plant Molecular Biology 65:519–530

18. ISAAA. (2018) Brief 54: global status of commercialized biotech/GM crops:2018. ISAAA Briefs, 54, Ithaca, NY: ISAAA
19. Jiang, Y. , Ling, L. , Zhang, L. , Wang, K. , Cai, M. , Zhan, M. , Li, C. , Wang, J. , Chen, X. , Lin, Y , Cao, C (2016) Transgenic bt (Cry1ab/ac) rice lines with different genetic backgrounds exhibit superior field performance under pesticide-free environment. Field Crops Research 6:1–6
20. Jung, K.H., Gho, H.J., Giong, H.K (2013) Genome-wide identification and analysis of Japonica and Indica cultivar-preferred transcripts in rice using 983 Affymetrix array data. Rice 6:19
21. Kogel, K.H., Voll, L.M., Schafer, P (2010) Transcriptome and metabolome profiling of field-grown transgenic barley lack induced differences but show cultivar-specific variances. Proc. Natl. Acad. Sci USA 107:6198–6203
22. Kuiper, H. A., Kleter, G. A., Notenorn, H. O. & Kok, E. J (2001) Assessment of the food safety issues related to genetically modifed foods. Plant Journal 27(6):503–528
23. Ladics, G.S., Bartholomaeus, A., Bregitzer, P. (2015) Genetic basis and detection of unintended effects in genetically modified crop plants. Transgenic Research 24:587–603
24. Li, H. , Olson, M. , Lin, G. , Hey, T. , Tan, S. Y. , & Narva, K. E (2013) Bacillus thuringiensis Cry34Ab1/Cry35Ab1 interactions with western corn rootworm midgut membrane binding sites. Plos One 8(1):e53079
25. Li S. , Li W. , Huang B. , Cao X. , Zhou X. , Ye S. , Li C. , Gao F. , Zou T. , Xie K. , Ren Y. , Ai P. , Tang Y. , Li X. , Deng Q. , Wang S. , Zheng A. , Zhu J. , Liu H. , Wang L. & Ping Li (2013) Natural variation in PTB1 regulates rice seed setting rate by controlling pollen tube growth. Nature Communications 4:2793
26. Li, Y. , Fan, C. , Xing, Y. , Yun, P. , Luo, L. , Yan, B. , Peng B. , Xie W. , Wang G. , Li X. , Xiao J. , Xu C. & He Y 2014 Chalk5 encodes a vacuolar h+-translocating pyrophosphatase influencing grain chalkiness in rice. Nature Genetics 46(4):398-404
27. Li, Y., Hallerman, E.M., Wu, K. and Peng, Y (2020) Insect-resistant genetically engineered crops in China: development, application, and prospects for use. Annual Review of Entomology 65(1):273–292
28. Lian Guangqian, Zhai Xue, Shang Fei, Wang Jianfei. (2019) Fine mapping and candidate genes analysis of a major QTL, qPL9, for panicle length in rice(Oryza sativa L.). Journal of Nanjing Agricultural University 42(3):398-405
29. Liu, J. , Chen, J. , Zheng, X. , Wu, F. , Lin, Q. , Heng, Y. , Tian, P. , Cheng, Z. , Yu, X. , Zhou, K. , Zhang, X. , Guo, X. , Wang, J. , Wang, H. , Wan, J (2017) Gw5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nature Plants 3:17043
30. Liu, Q. , Yang, X. , Tzin, V. , Peng, Y. , & Li, Y (2020) Plant breeding involving genetic engineering does not result in unacceptable unintended effects in rice relative to conventional cross–breeding. The Plant Journal 103:2236–2249
31. Lou, Y. G., Zhang, G. R., Zhang, W. Q., Hu Y., Zhang J (2013) Biological control of rice insect pests in China. Biological Control 6(7):8–20
32. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R (2018) Data-independent acquisition-based SWATH-MS for quantitative proteomics: A tutorial. Molecular Systems Biology 14(8):e8126
33. Miki, B., Abdeen, A., Manabe, Y. and MacDonald, P (2009) Selectable marker genes and unintended changes to the plant transcriptome. Plant Biotechnology Journal 7:211–218
34. Montero, M., Coll, A., Nadal, A., Messeguer, J. and Pla, M (2011) Only half the transcriptomic differences between resistant genetically modified and conventional rice are associated with the transgene. Plant Biotechnology Journal 9:693–702
35. NASEM (National Academies of Sciences, Engineering, and Medicine) (2016) Genetically engineered crops: experiences and prospects. Washington, DC USA:The National Academies Press.

36. Ricroch, A.E., Berge, J.B. and Kuntz, M (2011) Evaluation of genetically engineered crops using transcriptomic, proteomic, and metabolomic profiling techniques. Plant Physiology 155:1752–1761
37. Schnell, J., Steele, M., Bean, J., Neuspiel, M., Girard, C., Dormann, N., Pearson, C., Savoie, A., Bourbonniere, L. and Macdonald, P (2015) A comparative analysis of insertional effects in genetically engineered plants: considerations for pre-market assessments. Transgenic Res 24:1–17
38. Steiner, H. Y. , Halpin, C. , Jez, J. M. , Kough, J. , & Hannah, L. C (2013) Editor's choice: evaluating the potential for adverse interactions within genetically engineered breeding stacks. Plant Physiology 161(4):1578–1594
39. Stewart, C., & Via, L. E (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14(5):748–50
40. Strauss, S. H. & Sax, J. K (2016) Ending event-based regulation of GMO crops. Nature Biotechnology 34(5):474–477
41. Tan, Y., Zhang, J., Sun, Y., Tong, Z., Peng, C., Chang, L., Guo, A. and Wang, X (2019) Comparative proteomics of phytase-transgenic maize seeds indicates environmental influence is more important than that of gene insertion. Science Report 9:8219
42. Valentim-Neto, P.A., Rossi, G.B., Anacleto, K.B., de Mello, C.S., Balsamo, G.M. and Arisi, A.C (2016) Leaf proteome comparison of two GM common bean varieties and their non-GM counterparts by principal component analysis. J. Sci. Food Agric 96:927–932
43. Wang, W., Mauleon, R., Hu, Z (2019) Genetic variation assessment of stacked-trait transgenic maize via conventional breeding. BMC Plant Biol 19:346
44. Wang, X.J., Zhang, X., Yang, J.T. and Wang, Z.X (2018) Effect on transcriptome and metabolome of stacked transgenic maize containing insecticidal cry and glyphosate tolerance epsps genes. The Plant Journal 93, 1007–1016
45. Wang Y, Xu W, Zhao W, Hao J, Luo Y, Tang X, Zhang Y, Huang K (2012) Comparative analysis of the proteomic and nutritional composition of transgenic rice seeds with Cry1ab/ac genes and their non-transgenic counterparts. Journal of Cereal Science 55(2):226–233
46. Wang Y., Xiong G., Hu J., Jiang L., Yu H., Xu J., Fang Y., Zeng L., Xu E., Xu J., Ye W., Meng X., Liu R., Chen H., Jing Y., Wang Y., Zhu X., Li J., Qian Q. (2015) Copy number variation at the GL7 locus contributes to grain size diversity in rice. Nature Genetics 47(8):944–948
47. Yan Su, Yuchun Rao, Shikai Hu, Yaolong Yang, Zhenyu Gao, Guanghen Zhang, Jian Liu, Jiang Hu, Meixian Yan, Guojun Dong, Li Zhu, Longbiao Guo, Qian Qian, Dali Zeng (2011) Map-based cloning proves qGC-6, a major QTL for gel consistency of japonica/indica cross, responds by Waxy in rice (Oryza sativa L.). Theoretical and Applied Genetics 123(5):859-867
48. Ye, R., Huang, H., Zhou, Y., Chen, T., & Lin, Y. 2010 Development of insect-resistant transgenic rice with cry1C*-free endosperm. Pest Management Science 65(9):1015–1020
49. Ying Jiang, Aihua Sun, Yang Zhao1, Wantao Ying, Huichuan Sun, Xinrong Yang, Baocai Xing, Wei Sun, Liangliang Ren, Bo Hu, Chaoying Li, Li Zhang, Guangrong Qin, Menghuan Zhang, Ning Chen, Manli Zhang, Yin Huang, Jinan Zhou, Yan Zhao, Mingwei Liu, Xiaodong Zhu, Yang Qiu, Yanjun Sun, Cheng Huang, Meng Yan, Mingchao Wang, Wei Liu, Fang Tian, Huali Xu, Jian Zhou, Zhenyu Wu, Tieliu Shi, Weimin Zhu, Jun Qin, Lu Xie, Chinese Human Proteome Project (CNHPP) Consortium, Jia Fan, Xiaohong Qian & Fuchu He (2019) Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. Nature 567(7747):257-261
50. Yue, W., Di, J.u, Xueqing, Y., Ma, D., & Wang, X (2018) Comparative transcriptome analysis between resistant and susceptible rice cultivars responding to striped stem borer (SSB), Chilo suppressalis (Walker) infestation. Frontiers in Physiology 9:1717
51. Zeigler, R. S., & Barclay, A (2008) The relevance of rice. Rice 1(1):3–10

52. Zhang, J., Liu, Y.X., Zhang, N (2019) NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. Nature Biotechnology 37:676–684

Tables

Table 1

Differentially expressed proteins involved in three transgenic lines and recurrent parent.

Gene Symbol	Protein Description	CH891(1C)/CH891	CH891(2A)/CH891	CH891(1C+2A)/CH891
B8A8P2	1,4-alpha-D-glucan glucanohydrolase	0.526	0.758	0.822
A2YUR2	Tyrosine-protein phosphatase domain-containing protein	0.493	0.581	0.603
A2YCP9	Serine hydroxymethyltransferase	0.543	0.593	0.654
B8AN97	Nicotinate phosphoribosyltransferase	0.511	0.685	0.422
B8AC53	MoCF_biosynth domain-containing protein	0.591	0.665	0.659
B8BHS8	J domain-containing protein	0.490	1.200	0.593
A2ZMS2	Protease Do-like 5, chloroplast, putative, expressed	0.643	0.656	0.748
A2ZAG4	Plant intracellular Ras-group-related LRR protein 5	0.624	1.011	0.599
A2ZBX3	Calcium-dependent protein kinase 24	0.567	0.574	0.812
B8BPH4	UDP-glucose 6-dehydrogenase	3.608	0.961	4.989
B8AVF1	OSIGBa0106G07.1 protein	7.107	5.766	4.810
A2WJU9	Peptidyl-prolyl cis-trans isomerase	1.850	2.073	1.081
A2XLE8	Matrin-type domain-containing protein	2.063	1.591	1.319
B8AME3	Ubiquitin family protein, expressed	4.128	2.330	4.183
B8BCI9	Fe2OG dioxygenase domain-containing protein	3.572	2.699	2.746
B8B2Q3	Glutathione synthetase	2.445	0.709	2.754
B8APR2	Putative alcohol dehydrogenase	6.968	6.381	0.926
A2ZMK7	C-factor	3.691	2.980	1.467
B8B9E6	WD_REPEATS_REGION domain-containing protein	2.255	1.095	1.955
A2XB60	Acyl-CoA binding protein-like	1.705	1.482	1.832
B8BJ06	EF-hand domain-containing protein	1.532	1.555	1.951
B8B9C9	RHOMBOID-like protein	1.756	2.110	1.629
A2X0W6	Mitogen-activated protein kinase	3.280	2.389	3.729
B8B894	Zeta-carotene desaturase	0.630	0.851	0.866
B8BG13	Phosphoglucomutase (alpha-D-glucose-1,6-bisphosphate-dependent)	0.510	0.990	0.741

B8ARD8	UBX domain-containing protein	0.638	0.750	0.907

Table 2

Peptide sequences of fifteen candidate proteins in PRM.

Protein Symbol	Peptide Sequence	Retention Time	CH891(1C)/CH891 Ratio	CH891(2A)/CH891 Ratio	CH891(1C+2A)/CH891 Ratio
A2XLE8 (ZMAT)	CEICGNHSYWGR	12.05	1.4	1.17	1.67
B8AME3 (UBE)	ALIATAGNVHAAVER	13.37	1.28	1.53	1.73
B8BJ06 (EFHC)	AIEYDNFIECCLTVK	23.95	0.92	1.44	1.16
A2X0W6 (MAPK)	YLHSAEILHR	10.5	0.91	1.36	0.99
B8B9C9 (RHBD)	SNAIEHAHFR	7.72	1.54	0.96	1.53
B8B2Q3 (GSS)	ELAPIFNDLVDR	25.83	0.87	1.31	1.28
A2ZMK7 (CF)	TALNQLTK	11.64	1.8	1.37	1.64
B8BPH4 (UG6D)	ETPAIDVCHGLLGDK	18.25	0.71	1.61	1.17
B8ARD8 (PUX)	AFHFVQPIPR	17.24	0.7	0.98	0.83
A2ZMS2 (PDI)	LVGCDPSYDLAVLK	21.92	0.88	0.88	0.87
A2ZAG4 (PIRLs)	VFDDLIQR	18.3	0.75	0.88	0.62
B8B894 (ZDS)	ALVDPDGALQQVR	18.22	0.39	1.06	0.41
A2YUR2 (PTP)	FIAGGQWR	15.28	0.57	0.6	0.49
B8AN97 (NAPRT)	AYVVPQHVEELLK	19.49	0.39	0.48	0.5
B8BG13 (PGM)	EHWATYGR	9.44	1.1	0.83	0.81

Table 3

SSR polymorphic markers related to grain yield and quality traits.

SSR markers	Polymorphic loci lines	Chr	Gene sympol	Trait name
RM1026	CH891(1C)	9	*qPL9*	Panicle length
RM538	CH891(1C)	5	*PTB1*	Seed set rate
RM17804	CH891(1C)	5	*Chalk5*	Head rice rate
RM17804	CH891(1C)	5	*Chalk5*	Chalkiness
RM289	CH891(2A)	5	*GW5*	Grain length to width ratio
RM1364	CH891(2A)	7	*GL7*	Grain length to width ratio
RM190	CH891(2A)	6	*Wx*	Amylose content
RM190	CH891(1C+2A)	6	*Wx*	Amylose content

Figures

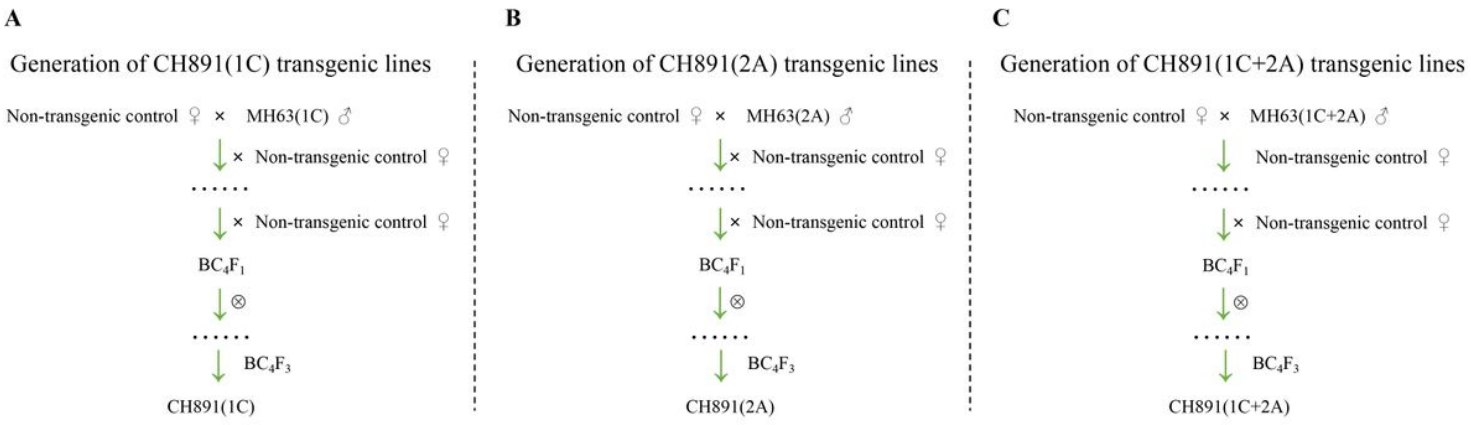

Figure 1

Genetic relations of transgenic restorer lines in this study. **a** The transgenic CH891 (1C) was produced by backcrossing imports the *CRY1C* gene. **b** The transgenic CH891 (2A) was produced by backcrossing imports the *CRY2A* gene. **c** The transgenic CH891 (1C+2A) was produced by backcrossing imports the *CRY1C* and *CRY2A*

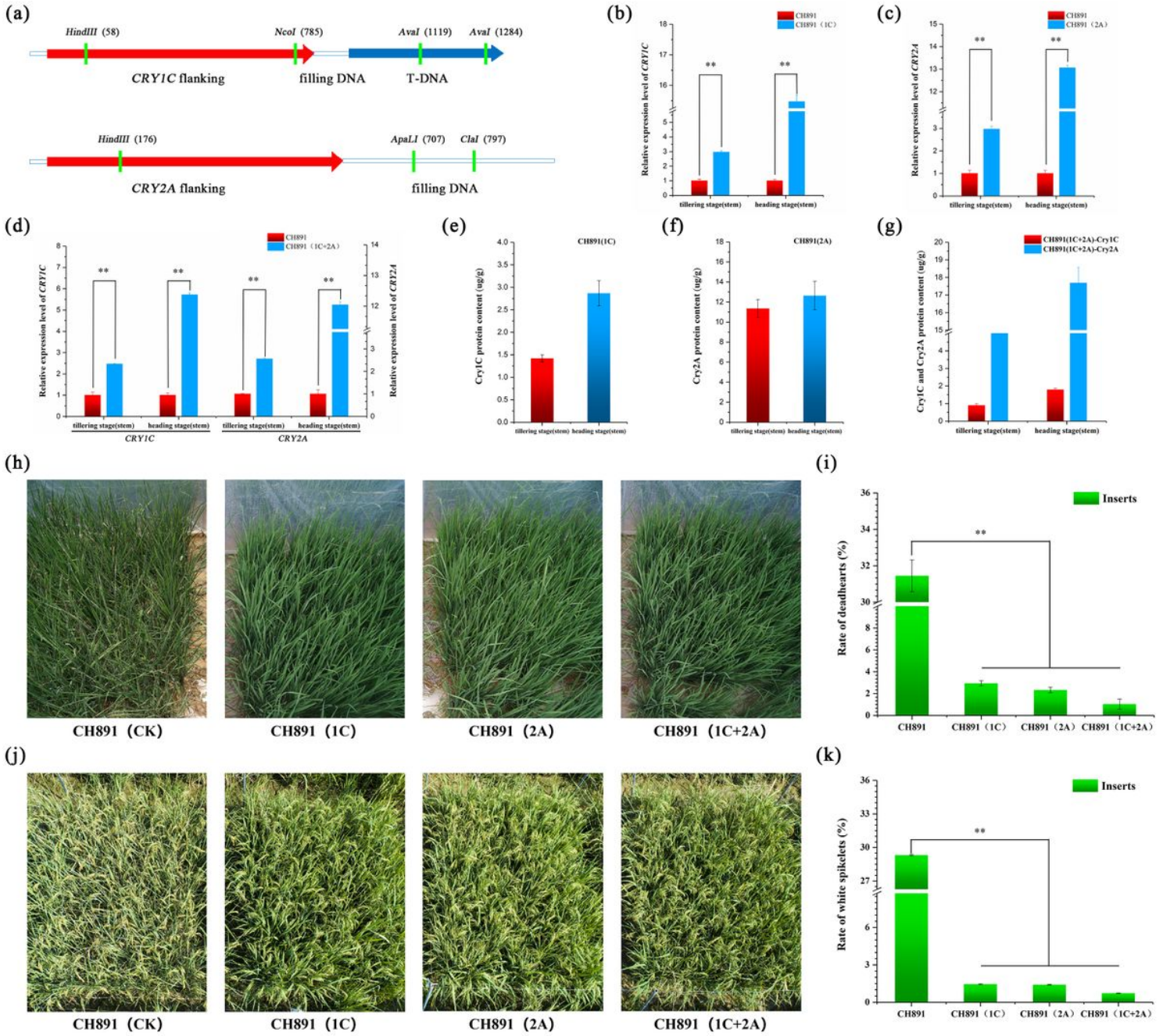

Figure 2

Field performance of insect resistance of CH891 expressing *CRY1C* and *CRY2A* gene cassette in genemodifcation CH891(1C), CH891(2A) and CH891(1C+2A) rice lines grown in Nanchang field (E115.5°, N28.5°), China. **a** Schematic map of flanking sequences of *CRY1C* and *CRY2A* insertion site. **b-d** The relative expressive level of *CRY1C* and *CRY2A* was detected by qRT-PCR. **e-g** The Cry1C and Cry2A protein content was detected by enzyme-linked immunosorbent assay (ELISA). **h** Field phenotype of transgenic restorer lines and recurrent parent exposed naturally to insect at tillering stage. **i** Comparison of infection index to insect (*Bacillus thuringiensis*) resistance at tillering stage in CH891(1C), CH891(2A) and CH891(1C+2A), compared to CH891. **j** Field phenotype of transgenic restorer lines and recurrent parent exposed naturally to insect at heading stage. **j** Comparison of infection index to insect (Bacillus thuringiensis) resistance at heading stage in CH891(1C), CH891(2A) and CH891(1C+2A), compared to CH891.

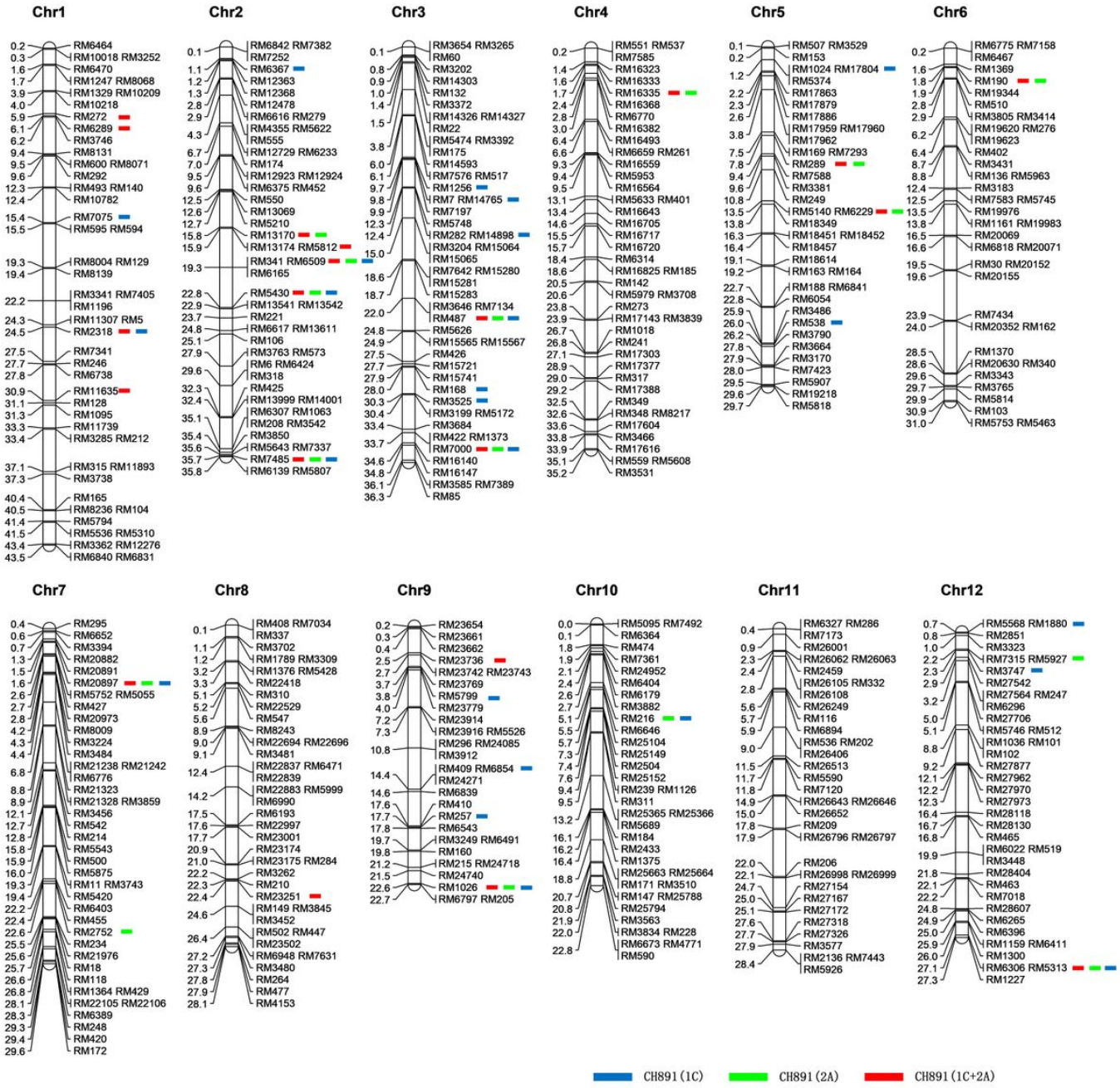

Figure 3

The linkage map of CH891(1C), CH891(2A) and CH891(1C+2A). The blue, green and red bars represent the Minghui 63 chromosomal segments in CH891(1C), CH891(2A) and CH891(1C+2A). The scale of the ruler indicates the genetic distance, the unit “cM”.

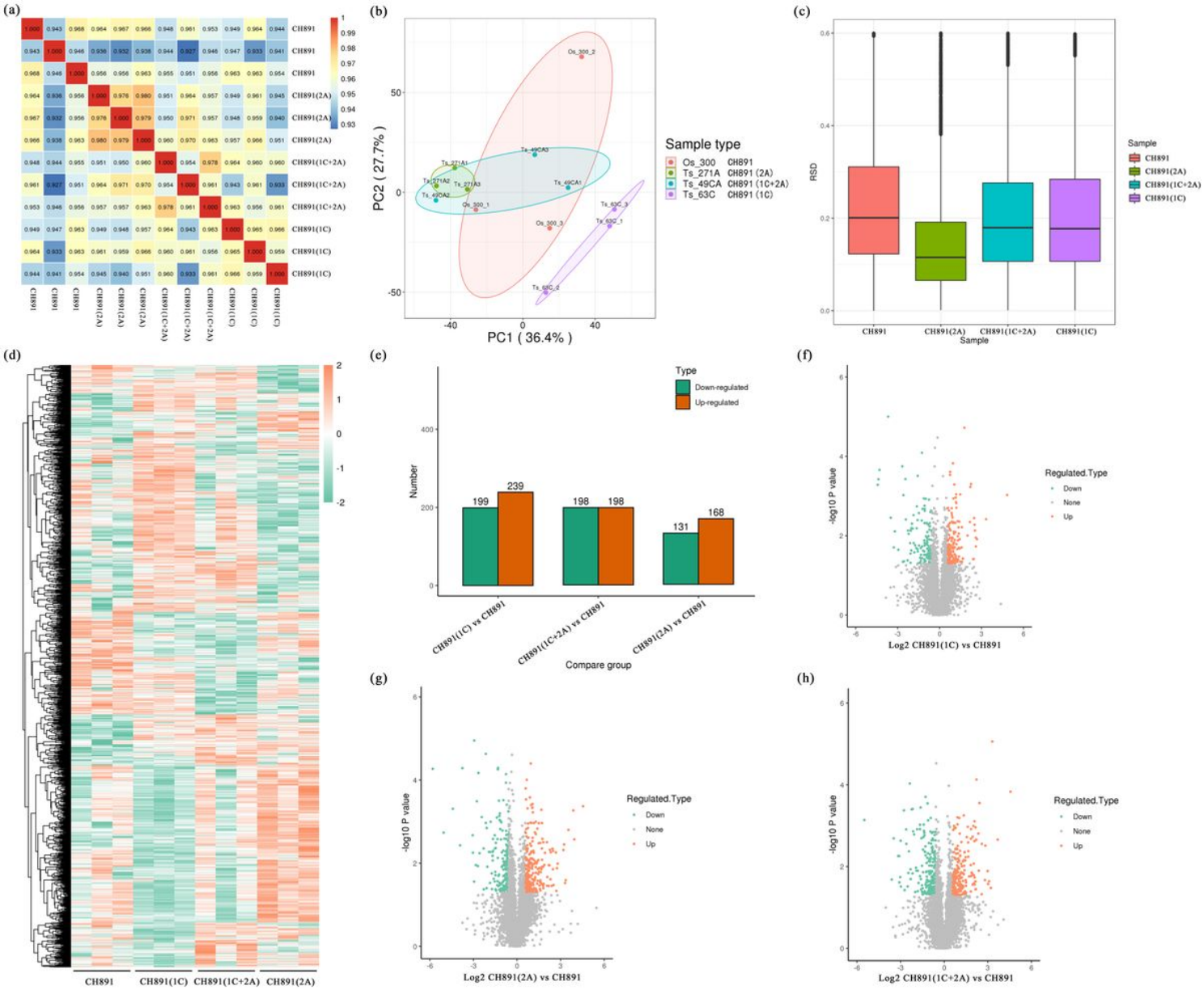

Figure 4

Overall description of proteome data. **a** All proteins in each sample were used to calculate the correlation coefficients within and between the samples, and then a heat map was drawn. **b** Principal components (PCs) analyses of protein expression levels in leaves of 4 rice lines. **c** The distribution of protein expression levels in different samples was shown in a box diagram. **d** Heatmap of fold-change differences in protein expression among samples of NIL groups. **e** Statistics of differentially expressed proteins (DEPs) among samples of NIL groups. **f** Volcano plot of DEPs between CH891(1C) and CH891. **g** Volcano plot of DEPs between CH891(2A) and CH891. h Volcano plot of DEPs between CH891(1C+2A) and CH891. The ordinate represents the multiple of difference and the ordinate represents the level of significance. The scatter in the upper left corner represents significant gene/protein expression, differentially upregulated. The scatter in the upper right represents significant gene/protein expression, differentially upregulated.

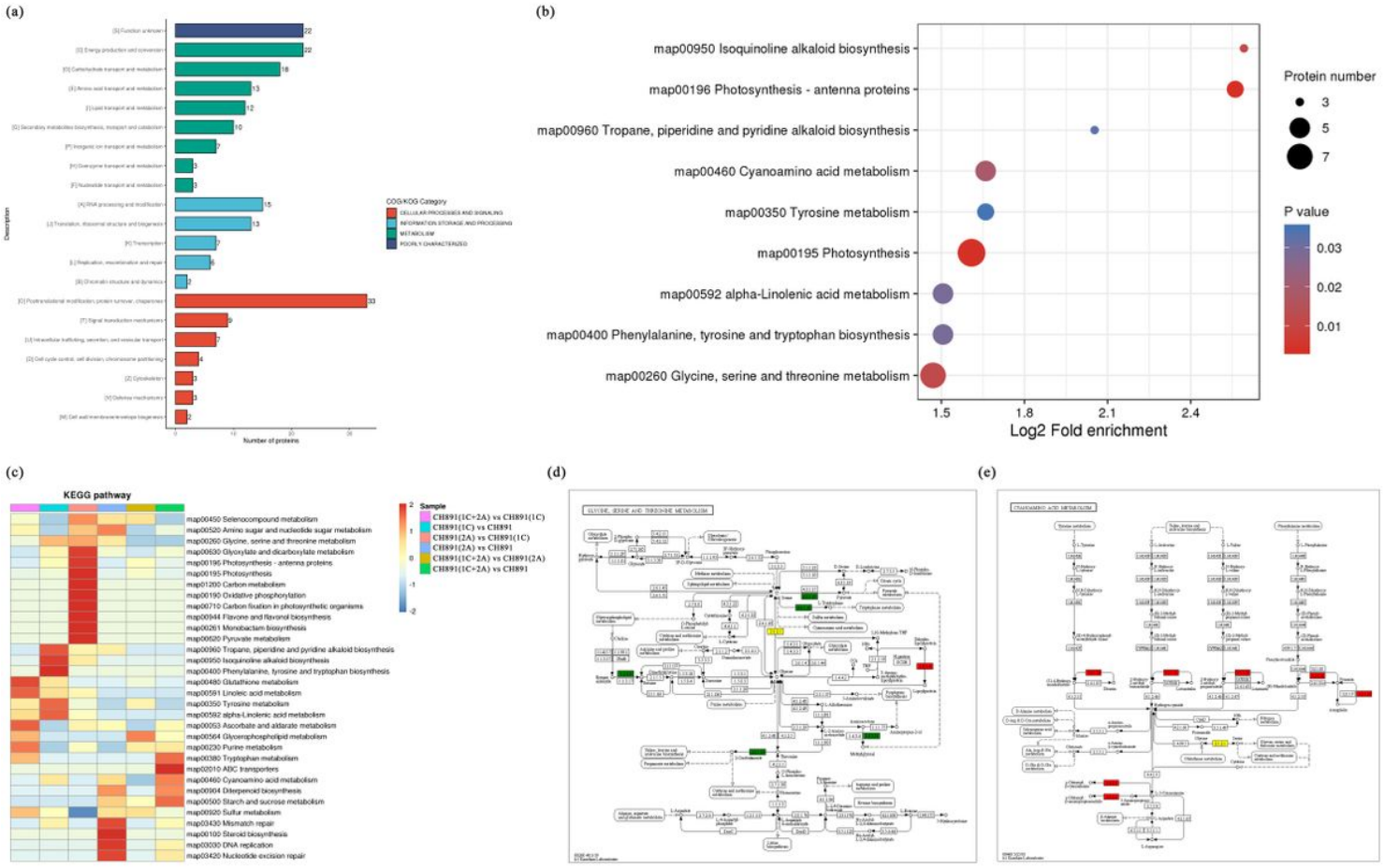

Figure 5

Enrichment and cluster analysis of differentially expressed proteins (DEPs). **a** Clusters of Orthologous Groups of proteins (COG) displayed the degree of accumulation of conserved proteins among samples of NIL groups. **b** KEGG enrichment analysis was performed for each protein class by Fisher's exact test method, and the enrichment results for different classifications were combined. Then, the top 40 ($p<0.05$) significantly enriched functional classifications were identified. The horizontal axis shows fold enrichment after Log_2 transformation, the vertical axis shows the functional classification, the bubble size represents the number of proteins, and the bubble color represents the value of p of enrichment significance. **c** Kyoto Encyclopedia of Genes and Genomes pathways significantly enriched in sets of DEPs among samples of NIL groups. **d** Differentially expressed proteins in the Glycine, Serine and Threonine metabolism pathways. **e** Differentially expressed proteins in the Cyanoamino acid metabolism pathways.

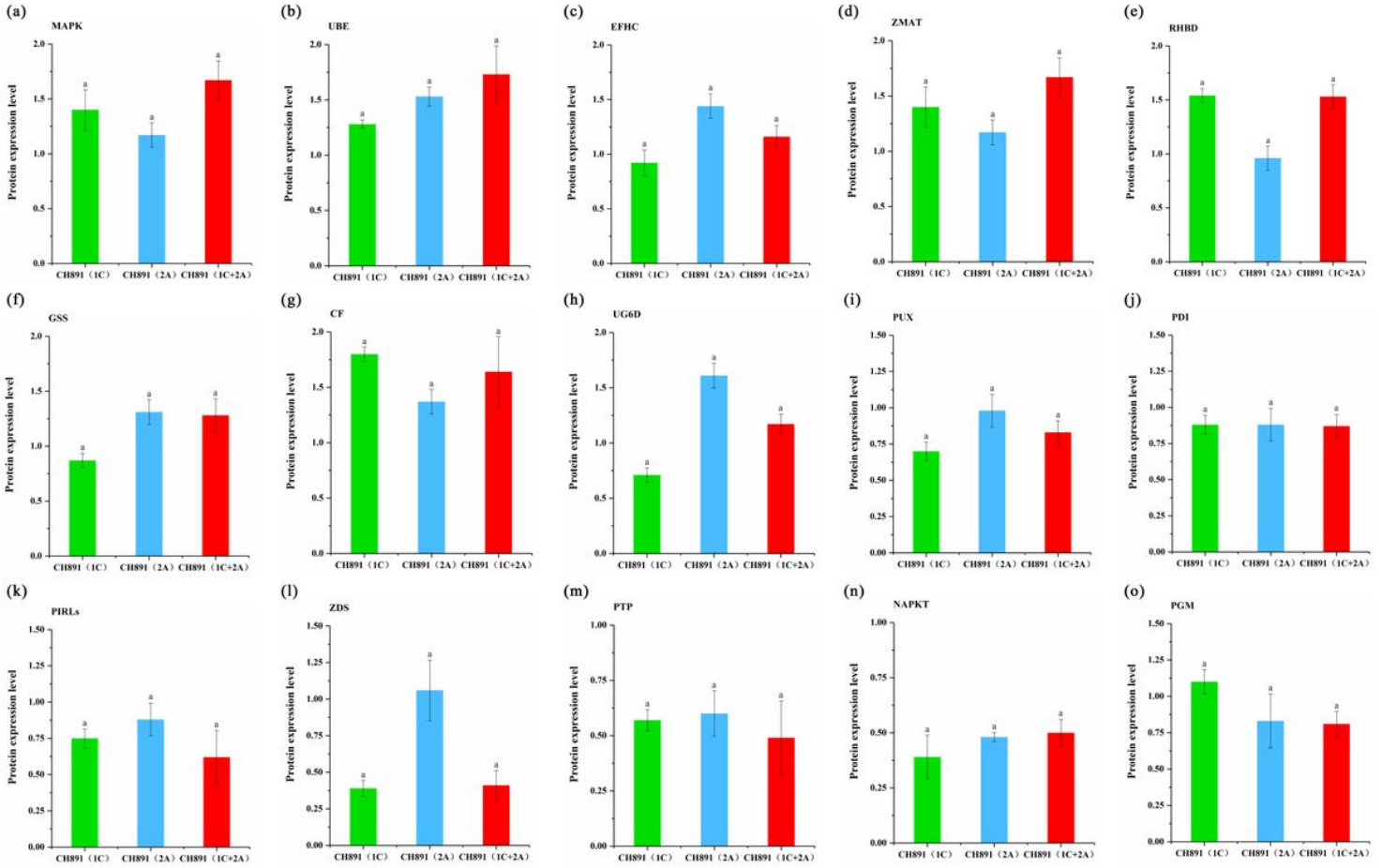

Figure 6

Parallel reaction monitoring (PRM) of proteins involved in the KEGG enrichment pathways by which three transgenic event. Samples were obtained from leaves at heading stage in CH891(1C),CH891(2A) and CH891(1C+2A). **a-o** represents PRM results of key enzymes MAPK, UBE, EFHC, SHMT, PPI, GSS, CF, UG6D, PUX, PDI, PIRLs, ZDS, PTP, NAPKT and PGM, respectively. Relative protein rates of each sample were compared to those in CH891. The data are means of three biological replicates, different letters indicate signifcant diferences within NIL groups ($p<0.05$) and bars between columns indicate signifcant diferences among proteins among NIL groups.

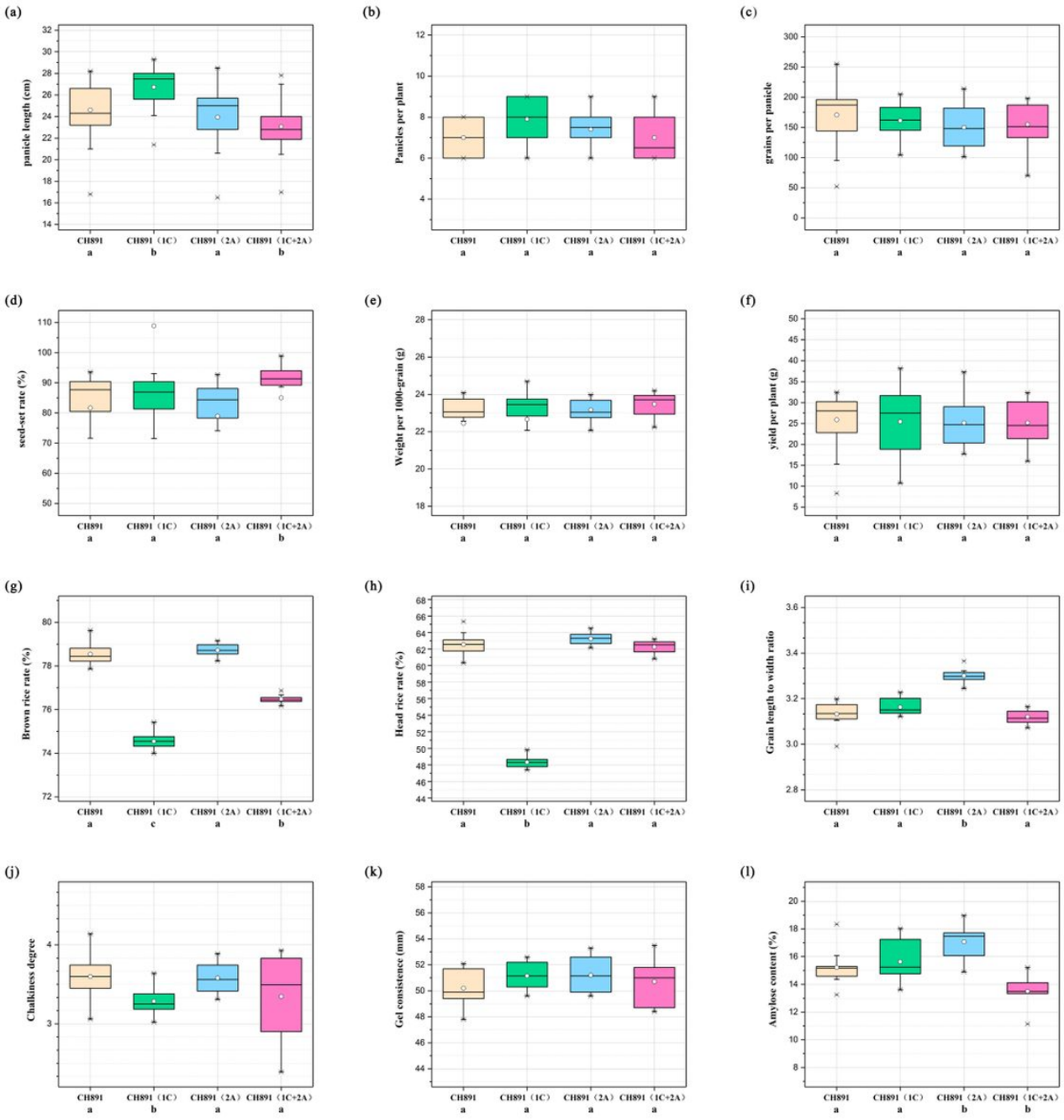

Figure 7

Boxplots of near isogenic lines (NILs) grouped by Bt gene status for the recurrent NIL groups CH891(1C), CH891(1C), CH891(1C+2A) for yield (**a-f**) and quality (**g-l**) traits. Medians are indicated by solid bold lines. Comparisons are based on least significant difference (LSD) across feld trials. Diferent letters below one-way ANOVA indicate signifcant diferences among NIL groups ($p<0.05$).

Supplementary Files

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

- SupportingInformation.docx