Article

Natural variation of an E3 ubiquitin ligase encoding gene *Chalk9* regulates grain chalkiness in rice

Zhi Hu[1,2,3], Hongchun Liu[1,3], Min Guo[1,3], Xiang Han[1], Youguang Li[1], Rujia Chen [1], Yifan Guo[1], Yihao Yang[1], Shengyuan Sun[1], Yong Zhou[1], Minghong Gu[1] & Changjie Yan [1,2] ✉

Grain chalkiness is an undesirable trait affecting rice quality, concerning both consumers and breeders. However, the genetic mechanisms underlying rice chalkiness remain largely elusive. Here, we identify *Chalk9* as a major gene associated with grain chalkiness in a natural population. *Chalk9* encodes an E3 ubiquitin ligase that targets OsEBP89 for its ubiquitination and degradation. Low expression of *Chalk9* results in excessive accumulation of OsEBP89, disrupting the homeostasis of storage components and leading to the chalkiness phenotype. A 64-bp insertion/deletion in the *Chalk9* promoter contributes to its differential transcriptional levels, thus causing chalkiness variation among rice varieties. Moreover, the elite allele *Chalk9*-L reduces grain chalkiness, without compromising yield. *Chalk9*-L is strongly selected in *japonica* but exhibits a complex evolutionary trajectory in *indica*. Our findings reveal the molecular and genetic mechanisms underlying chalkiness and provide a potential strategy for breeding rice varieties with improved quality.

Rice (*Oryza sativa* L.) is a staple food for over half of the global population, but enhancing its grain quality remains a significant challenge as living standards rise[1,2]. Chalkiness, a major determinant of rice quality, severely reduces the appearance quality of rice and negatively affects milling and cooking quality, thereby diminishing its commercial value[3,4]. Chalkiness is an undesirable trait for consumers and marketing[4]. Preventing grain chalkiness formation is thus a critical goal in rice breeding.

Crop breeding is a dynamic and continuous process that strongly reflects human preferences[5]. Over the past century, rice breeding efforts have primarily focused on enhancing rice productivity by developing high-yield varieties[6]. However, these increased yields often come at the cost of poor quality, particularly high chalkiness[2,7]. Seed storage proteins (SSPs) and starch, the predominant components in rice grains, determine both yield and quality. The negative correlation between yield and quality likely

arises from the disruption of their coordinated synthesis[8,9]. Breaking this trade-off between yield and quality represents a breakthrough opportunity for rice breeders.

Chalkiness, which refers to opaque regions in the endosperm, is a complex quantitative trait influenced by polygenes and environmental factors, such as high temperature and nutrient availability[10–12]. Extensive efforts have been made to dissect the genetic basis of chalkiness in rice, and numerous quantitative trait loci (QTLs) related to chalkiness have been identified on all 12 rice chromosomes using biparental mapping and natural populations[13–19]. Several genes have been functionally cloned and characterized. For example, *Chalk5* influences rice grain chalkiness by regulating pH homeostasis in developing seeds[20]. Natural variation in *WCR1* regulates redox homeostasis in rice endosperm to affect grain chalkiness[21]. Recent studies also identified *WBR7* and *LCG1* as regulators of rice chalkiness through their effects on the accumulation of grain storage components[22,23]. Despite these

[1]Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding/Zhongshan Biological Breeding Laboratory/Key Laboratory of Plant Functional Genomics of the Ministry of Education, College of Agriculture, Yangzhou University, Yangzhou 225009 Jiangsu, China. [2]Co-Innovation Center for Modern Production Technology of Grain Crops of Jiangsu Province, Yangzhou University, Yangzhou, Jiangsu, China. [3]These authors contributed equally: Zhi Hu, Hongchun Liu, Min Guo. ✉e-mail: cjyan@yzu.edu.cn

advancements, the genetic and molecular mechanisms underlying rice grain chalkiness remain unclear.

E3 ligases are critical components of the ubiquitin-proteasome system determining the substrate specificity of the cascade by covalent attachment of ubiquitin to target proteins[24,25]. RING-finger proteins, a major family of E3 ligases characterized by a 40-60 residue RING domain, confer substrate specificity through direct interaction with target proteins[26]. The RING domain, stabilized by zinc ions coordinated by cysteine and histidine residues, is essential for E3 activity. Mutations in these zinc-binding residues can disrupt the domain structure and abolish ligase activity[26]. E3 ligases play significant roles in plant growth, stress resistance, and signaling[27–29]; however, their role in regulating grain chalkiness remains unknown.

In this work, we identify *Chalk9* as a major gene controlling chalkiness variation through genome-wide association studies (GWAS) in *indica* rice germplasm and elucidate the molecular mechanism of *Chalk9*-mediated chalkiness regulation. For breeding applications, we identify the elite haplotype *Chalk9*-L, which improves rice appearance quality without yield penalty. Our findings provide insight into the molecular mechanisms underlying rice chalkiness and offer promising strategies for breeding rice varieties with high quality.

Results

Chalk9 is a major locus associated with grain chalkiness in *indica* rice

To investigate the genetic basis of grain chalkiness, we collected 175 *indica* rice varieties from a global population with high phenotypic diversity in chalky grain rate (CGR) and degree of chalkiness (DC) (Supplementary Fig. 1a–d and Supplementary Data 1). Whole-genome sequencing of these varieties generated a final set of 2,290,145 high-quality single-nucleotide polymorphisms (SNPs) after filtering. Principal component analysis showed that the score plot of principal components had continuous distribution without any distinct clusters (Supplementary Fig. 1e), indicating that these *indica* varieties did not represent a highly structured population. In addition, the average decay of linkage disequilibrium (LD) distance was estimated at about 180 kb in this population ($r^2 = 0.1$) (Supplementary Fig. 1f), consistent with the previous estimation in cultivated rice[30].

Using a linear mixed model, we identified a major locus on chromosome 9, *Chalk9*, associated with CGR and DC in the 2-year trials through GWAS (Supplementary Fig. 2). This locus explained ~28% of the total phenotypic variation (Supplementary Table 1). In the overlapped peak, the top two SNPs associated with CGR were located at 19,506,938 bp ($P = 8.12 \times 10^{-10}$) and 19,536,079 bp ($P = 7.25 \times 10^{-12}$), while the top two SNPs associated with DC were located at 19,586,699 bp ($P = 2.65 \times 10^{-10}$) and 19,536,079 bp ($P = 4.39 \times 10^{-11}$) (Fig. 1a). LD analysis delimited the candidate region within an approximately 200-kb block from 19.43 to 19.63 Mb (Fig. 1b). Interestingly, *Chalk9* was located within the previously reported chalkiness-associated QTL regions, such as *qWBR9-1* and *qCR9-1*[17].

Using a relatively strict *P* value threshold ($P < 1 \times 10^{-6}$), we identified 76 SNPs that were significantly associated with chalkiness (Supplementary Data 2). Of these, 11 caused missense mutations and 1 caused a synonymous mutation (both in gene coding regions), while 20 were in regulatory regions. These 32 SNPs, likely affecting gene function or expression, were assigned to 15 genes (Supplementary Data 3). The remaining SNPs were in the intergenic regions (14 SNPs) or gene introns (30 SNPs). For these 15 genes, three genes were annotated as either transposon-related or expressed proteins, the remaining 12 candidate genes were annotated as putative functional proteins (Supplementary Data 4).

LOC_Os09g32730 is the candidate of *Chalk9*

To identify the candidate gene for *Chalk9*, we evaluated the estimated effect of missense SNPs causing amino acid substitutions among 12 candidate genes. We found that these missense SNPs were only in 6 of the 12 genes (Supplementary Data 5). Moreover, only one SNP affected a functional domain (Supplementary Fig. 3a), but it was not conserved across plant species (Supplementary Fig. 3b), suggesting the missense SNP was unlikely to affect protein function[31]. We then randomly selected eight lines each from both high-chalky and low-chalky varieties to measure the expression levels of these 12 genes in endosperms and leaves by quantitative RT-PCR (qRT-PCR). Of the 12 genes, 11 showed no significant differences in the expression levels in the endosperms between the high-chalky and low-chalky varieties (Fig. 1c). Only gene III (*LOC_Os09g32730*) showed significantly higher expression in low-chalky varieties compared to high-chalky varieties (Fig. 1c, d). In contrast, these 12 genes exhibited similar expression levels in leaves between high-chalky and low-chalky varieties (Fig. 1e). Notably, *LOC_Os09g32730* was preferentially expressed in the developing endosperm, compared to the other candidate genes (Supplementary Fig. 3c). Since grain chalkiness is closely associated with endosperm development, *LOC_Os09g32730* was identified as a potential candidate gene for the *Chalk9* locus. Hence, we designated this gene as *Chalk9*.

To further validate *LOC_Os09g32730* as the functional gene, we generated transgenic lines that either overexpressed *Chalk9* (OE) using the constitutive CaMV 35S promoter or interfered *Chalk9* using RNA interference (RNAi) in the Zhonghua11 (ZH11) background. Two *Chalk9*-overexpression lines (OE1 and OE2) displayed decreased chalkiness with lower CGR and DC values, whereas two *Chalk9*-RNAi lines exhibited increased chalkiness with higher CGR and DC values (Fig. 2a–d). The RNAi lines developed in the variety Nipponbare (Nip) or Yangdao 6 (YD6) also displayed increased chalkiness (Supplementary Fig. 4a–h). Additionally, the CRISPR/Cas9 system was used to specifically disrupt the *Chalk9* gene in Nip. Two mutant alleles, *chalk9-1* and *chalk9-2*, were generated, each with a 1 bp deletion and a 1 bp insertion in the exon of *Chalk9*, resulting in premature stop codons at the 115th and 92nd codons, respectively (Supplementary Fig. 4i–k). Two knockout lines exhibited increased chalkiness, with higher CGR and DC values compared to Nip (Fig. 2e–g). Collectively, these results strongly suggest that *LOC_Os09g32730* is the candidate of *Chalk9*, acting as a negative regulator of grain chalkiness in rice.

An indel in the *Chalk9* promoter confers grain chalkiness variation

To address potential limitations in identifying DNA sequence variations in *Chalk9* from low-coverage genome sequencing, we re-sequenced *Chalk9* of 149 varieties selected from the GWAS panel, representing a wide range of grain chalkiness. We then conducted an association analysis with the identified variants (Supplementary Data 6). Two indels (−1331, 64-bp and −791, 1-bp; referred to as v5 and v12) and four SNPs (−1355 G > A, −817 A > G, −749 G > A, and −634 G > C; referred to as v4, v10, v14, and v15) in the promoter region of *Chalk9* exhibited stronger associations with grain chalkiness than the other SNPs (Supplementary Fig. 5a, Supplementary Data 6). However, a missense SNP in the coding region was not significantly associated with grain chalkiness (Supplementary Fig. 5a).

Based on the identified variants, we classified *Chalk9* variations into two haplotypes: one associated with high-chalky varieties [haplotype H (*Chalk9*-H)] and one with low-chalky varieties [haplotype L (*Chalk9*-L)] (Fig. 3a, b, Supplementary Table 2 and Supplementary Data 7). We randomly selected 24 accessions from each haplotype (L and H) to measure *Chalk9* expression levels. *Chalk9*-L accessions showed significantly higher *Chalk9* expression in the endosperm compared to *Chalk9*-H accessions (Fig. 3c). We further developed a near-isogenic line (NIL) carrying the *Chalk9*-H allele from the *indica* variety Kasalath in the *japonica* variety Nip, which had the *Chalk9*-L allele based on the known reference genome (Supplementary Fig. 5b, c). Compared to Nip, NIL$^{Chalk9\text{-}H}$ plants exhibited significantly higher grain chalkiness with reduced *Chalk9* expression (Fig. 3d–g). These
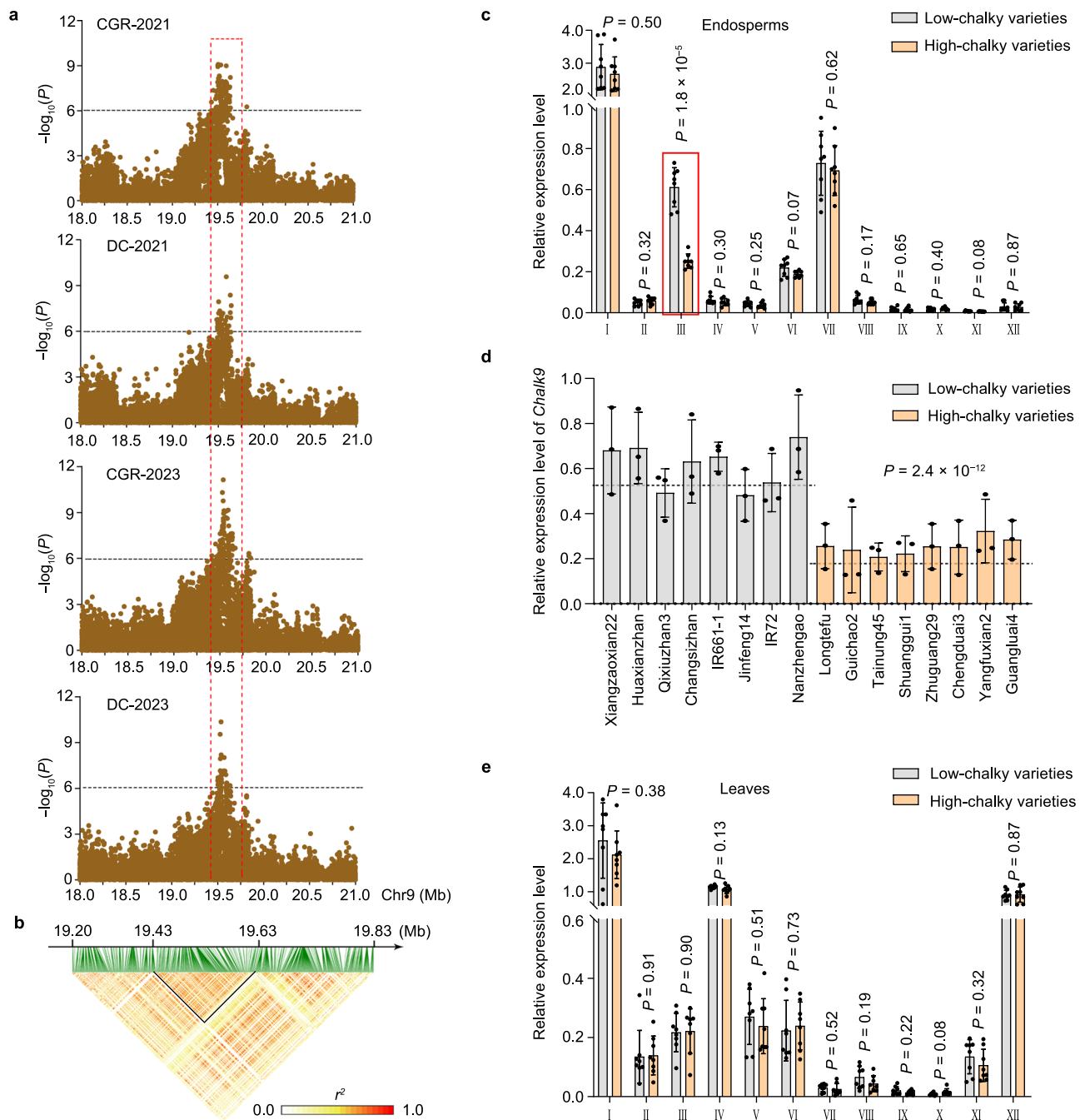
Fig. 1 | GWAS and fine mapping of the major locus that underlies grain chalkiness variation. a The genome-wide association signals for chalky grain rate (CGR) and degree of chalkiness (DC) in the region at 18–21 Mb on chromosome 9 (*x*-axis) across two years. Negative $\log_{10}$-transformed *P* values from the linear mixed model are plotted on the *y*-axis. The horizontal dashed line indicates the genome-wide significance threshold ($P = 1 \times 10^{-6}$). *P* values were determined using a two-sided Wald test and assessed after Bonferroni correction for multiple comparisons. **b** Linkage disequilibrium (LD) heatmap of the *Chalk9* locus region. Pairwise linkage disequilibrium was determined by calculating r^2 (the square of the correlation coefficient between SNPs). **c** Relative expression level of the 12 candidate genes in the endosperm of eight high-chalky and eight low-chalky varieties at 20 days after

flowering (DAF). The 12 predicted genes in the *Chalk9* locus region are labeled by I to XII. Data show means ± SD (*n* = 8 varieties). *P* values were calculated for comparisons between high-chalky and low-chalky groups, with each group comprising 8 varieties. **d** Relative expression level of the candidate gene III (*Chalk9*) in the endosperm from the selected varieties at 20 DAF. The *P* value was calculated for the comparison between high-chalky and low-chalky groups, with each group comprising 8 varieties. Data show means ± SD (*n* = 3 biological replicates). **e** Relative expression level of the 12 candidate genes in the leaves of eight high-chalky and eight low-chalky varieties. Data show means ± SD (*n* = 8 varieties). In **c**–**e**, statistical analysis between high-chalky and low-chalky groups was performed by two-tailed Student's *t*-test. Source data are provided as a Source Data file.

results suggest that the two *Chalk9* haplotypes confer different expression levels and variations in grain chalkiness in rice.

To investigate whether the functional differences between the two *Chalk9* haplotypes arise from the variants in the promoter or coding regions, we created three transgenic constructs (*pChalk9-*

L::Chalk9-L, *pChalk9-H::Chalk9-H*, and *pChalk9-L::Chalk9-H*), and used them to generate transgenic plants (Supplementary Fig. 5d) (see Methods). As shown in Fig. 3a–c and Supplementary Data 7, the *indica* variety Guichao2 carried the *Chalk9-H* allele, which was associated with low *Chalk9* expression and high chalkiness. Compared to wild-type
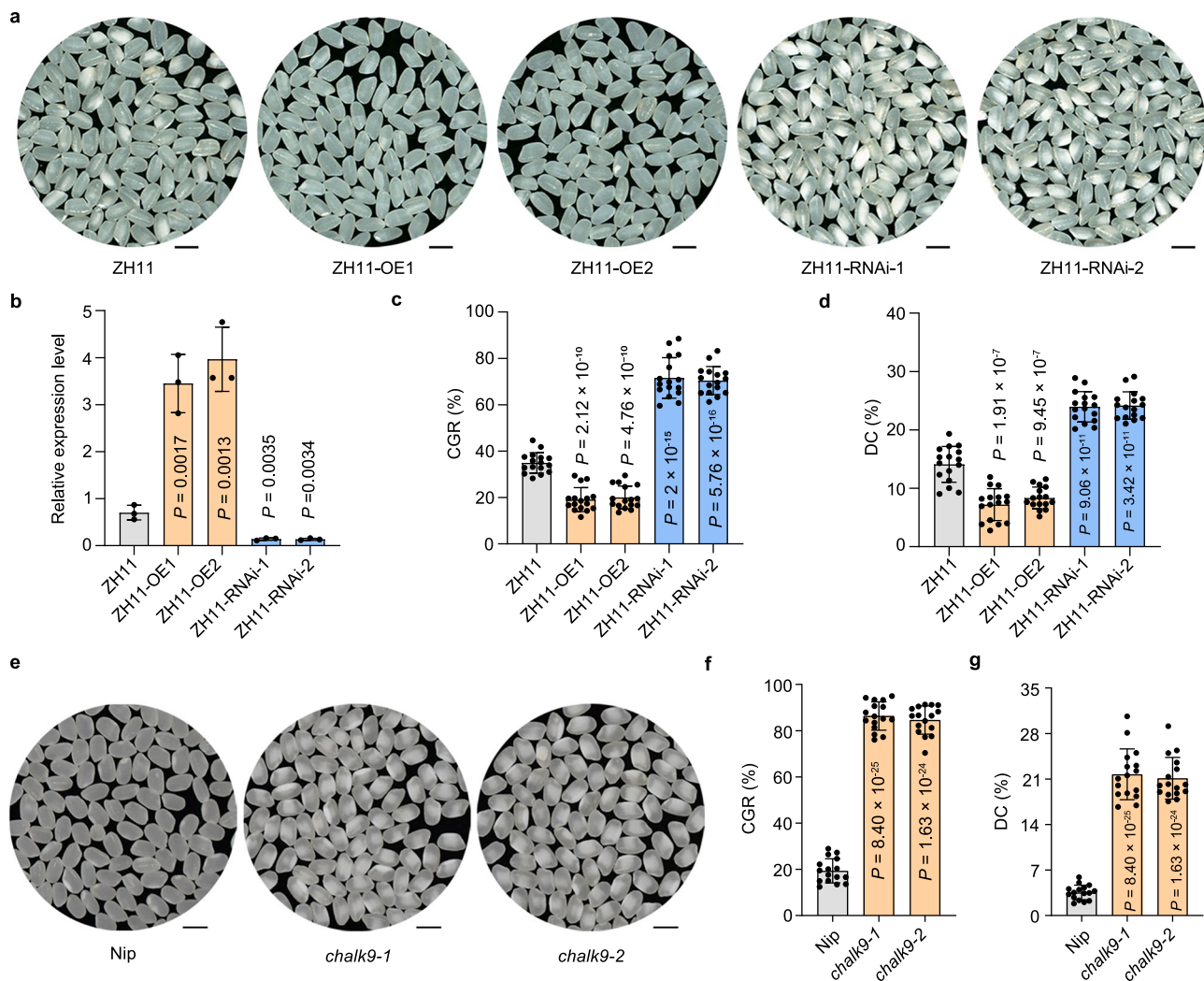
Fig. 2 | Chalk9 negatively regulates grain chalkiness in rice. a Grain chalkiness in ZH11, ZH11-OE1, ZH11-OE2, ZH11-RNAi-1, and ZH11-RNAi-2 plants. Scale bar: 5 mm. **b** Expression analysis of *Chalk9* in ZH11 and transgenic plants. Data show means ± SD ($n = 3$ biological replicates). **c, d** Chalky grain rate (**c**) and degree of chalkiness (**d**) in ZH11 and transgenic plants. Data show means ± SD ($n = 16$ plants).

e Grain chalkiness in Nip, *chalk9-1*, and *chalk9-2* plants. Scale bar: 5 mm. **f, g** Chalky grain rate (**f**) and degree of chalkiness (**g**) in Nip, *chalk9-1*, and *chalk9-2* plants. Data show means ± SD ($n = 16$ plants). In **b–d**, **f**, and **g**, statistical analysis was performed by two-tailed Student's *t*-test. Source data are provided as a Source Data file.

Guichao2 plants (*Chalk9*-H type), *pChalk9-L::Chalk9-L* and *pChalk9-L::Chalk9-H* transgenic lines showed significantly reduced grain chalkiness, with approximately 20% and 40% decreases in CGR and DC values, respectively (Fig. 3h–j). Importantly, the chalkiness-reducing effects were comparable regardless of whether the *Chalk9*-L or *Chalk9*-H coding sequence was expressed under the *Chalk9*-L promoter (Fig. 3h–j), suggesting that the amino acid changes do not impair *Chalk9*'s function in chalkiness regulation. However, *pChalk9-H::Chalk9-H* transgenic plants showed no significant difference in grain chalkiness relative to wild-type Guichao2 (Fig. 3h–j). Consistent with the phenotypes of these transgenic lines, *pChalk9-L::Chalk9-L* and *pChalk9-L::Chalk9-H* transgenic lines showed higher *Chalk9* transcript levels than wild-type Guichao2 and *pChalk9-H::Chalk9-H* transgenic plants (Fig. 3k), highlighting the importance of promoter-driven expression differences in determining haplotype-specific phenotypic effects. These findings indicate that the variants in the *Chalk9* promoter are responsible for the differences in grain chalkiness.

To further pinpoint the functional variations, we mutated the *Chalk9*-L promoter by introducing each of six variations individually (v4, v5, v10, v12, v14, and v15) from the *Chalk9*-H promoter. Transient expression assays showed that the activity of the

Chalk9-L promoter was significantly reduced by deleting the 64-bp indel, to a level that was comparable to that of the *Chalk9*-H promoter (Fig. 3l, Supplementary Fig. 5e). By contrast, none of the other five mutations affected the activity of the *Chalk9*-L promoter (Fig. 3l, Supplementary Fig. 5e). We also generated gene-edited plants with a deletion in this 64-bp indel region in Nip (*Chalk9*-L type) (Supplementary Fig. 5f). The *Chalk9*-L gene-edited (*D52*) plants exhibited reduced *Chalk9* expression and increased chalkiness (Fig. 3m–p), further confirming that the 64-bp indel in the promoter as the causal variant.

To understand why the 64-bp indel resulted in different expressions, we further analyzed its sequence and identified binding sites for some conserved transcription factors, including AT-Hook, TCR, B3, and ZF-HD families (Supplementary Data 8). Among these, a rice B3 domain transcription factor (OsB3), highly expressed in endosperms and homologous to ABI3 (essential for seed maturation in *Arabidopsis*[32]), was found (Supplementary Fig. 6a, b). We found that OsB3 protein activated the promoter of *Chalk9*-L (Supplementary Fig. 6c, d). In the absence of the 64-bp sequence of the *Chalk9*-L promoter, the activation of OsB3 protein was significantly reduced (Supplementary Fig. 6c, d). These results demonstrate that the 64-bp
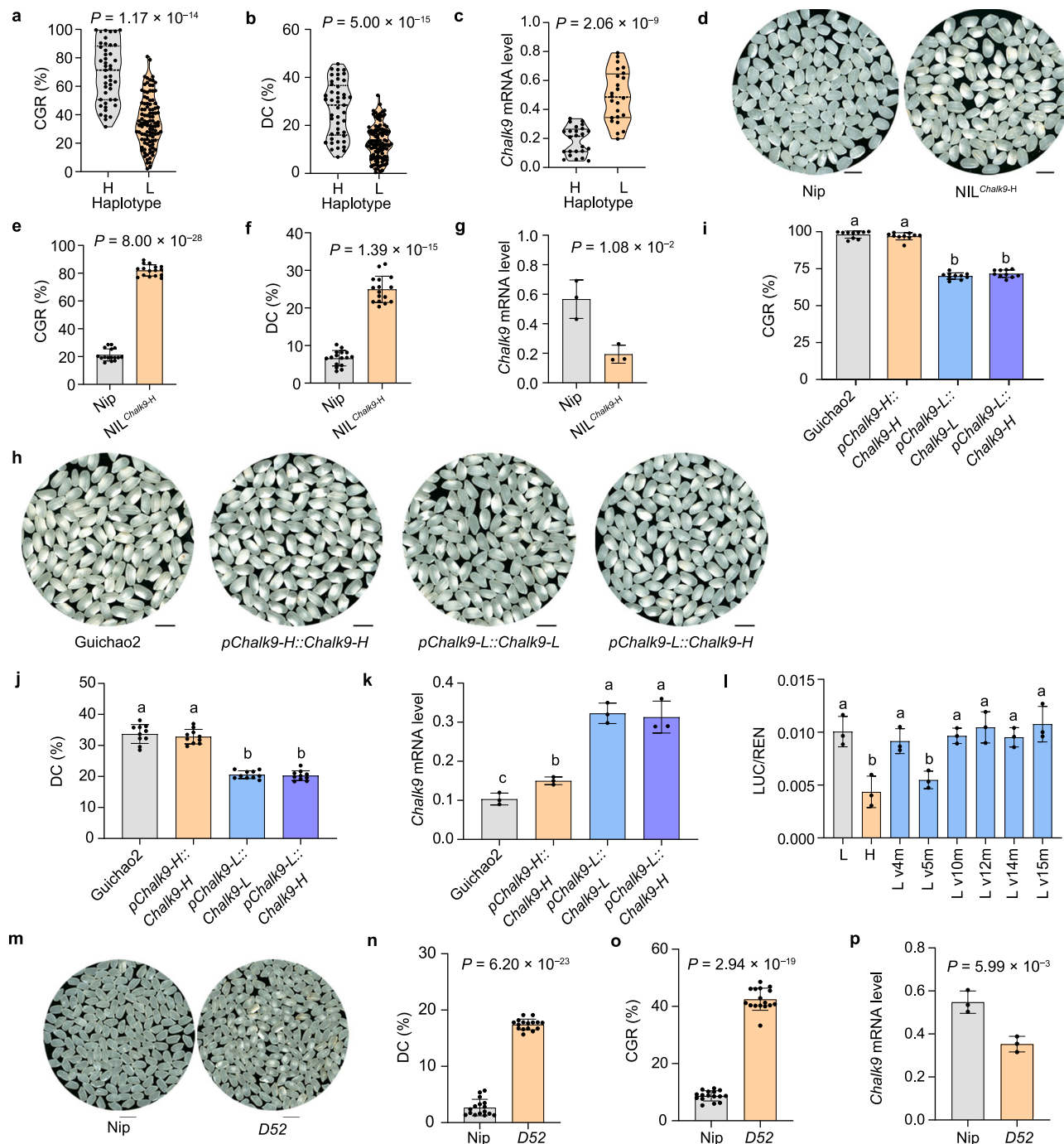
Fig. 3 | A 64-bp indel in the *Chalk9* promoter confers different grain chalkiness in rice. a, b The distribution of chalky grain rate (**a**) and degree of chalkiness (**b**) in haplotype H (*n* = 45 accessions) and haplotype L (*n* = 104 accessions).
c Expression analysis of *Chalk9* in haplotype H (*n* = 24 accessions) and haplotype L (*n* = 24 accessions) in endosperms. **d** Grain chalkiness of Nip and NIL$^{Chalk9\text{-}H}$ plants. Scale bar: 5 mm. **e, f** Chalky grain rate (**e**) and degree of chalkiness (**f**) of Nip and NIL$^{Chalk9\text{-}H}$ plants. **g** Relative *Chalk9* expression levels of Nip and NIL$^{Chalk9\text{-}H}$ plants in endosperms. **h** Grain chalkiness of Guichao2, *pChalk9-H::Chalk9-H*, *pChalk9-L::Chalk9-L*, and *pChalk9-L::Chalk9-H* plants. Scale bar: 6 mm. **i, j** Chalky grain rate (**i**) and degree of chalkiness (**j**) of Guichao2, *pChalk9-H::Chalk9-H*, *pChalk9-L::Chalk9-L*, and *pChalk9-L::Chalk9-H* plants. **k** Relative *Chalk9* expression levels of Guichao2, *pChalk9-H::Chalk9-H*, *pChalk9-L::Chalk9-L*, and *pChalk9-L::Chalk9-H* plants in endosperms. **l** Transient expression assays of the effect of different variations on gene expression, shown by firefly luciferase/Renilla luciferase activity ratio (LUC/REN). **m** Grain chalkiness of Nip and *D52* plants. Scale bar: 5 mm. **n, o** Degree of chalkiness (**n**) and chalky grain rate (**o**) in Nip and *D52* plants. **p** Relative *Chalk9* expression levels of Nip and *D52* plants in endosperms. Data show means ± SD (*n* = 16 plants in **e, f n**, and **o**; *n* = 3 biological replicates in **g, k, l**, and **p**; *n* = 10 plants in **i** and **j**). In **a–c**, the bars within violin plots represent the 25th percentile, median, and 75th percentile. In **a–c**, **e–g**, and **n–p**, statistical analysis was performed by two-tailed Student's *t*-test. In **i–l**, different letters indicate significant differences (*P* < 0.05, one-way ANOVA with Tukey's multiple comparison test); for *P* values, see Source Data. Source data are provided as a Source Data file.
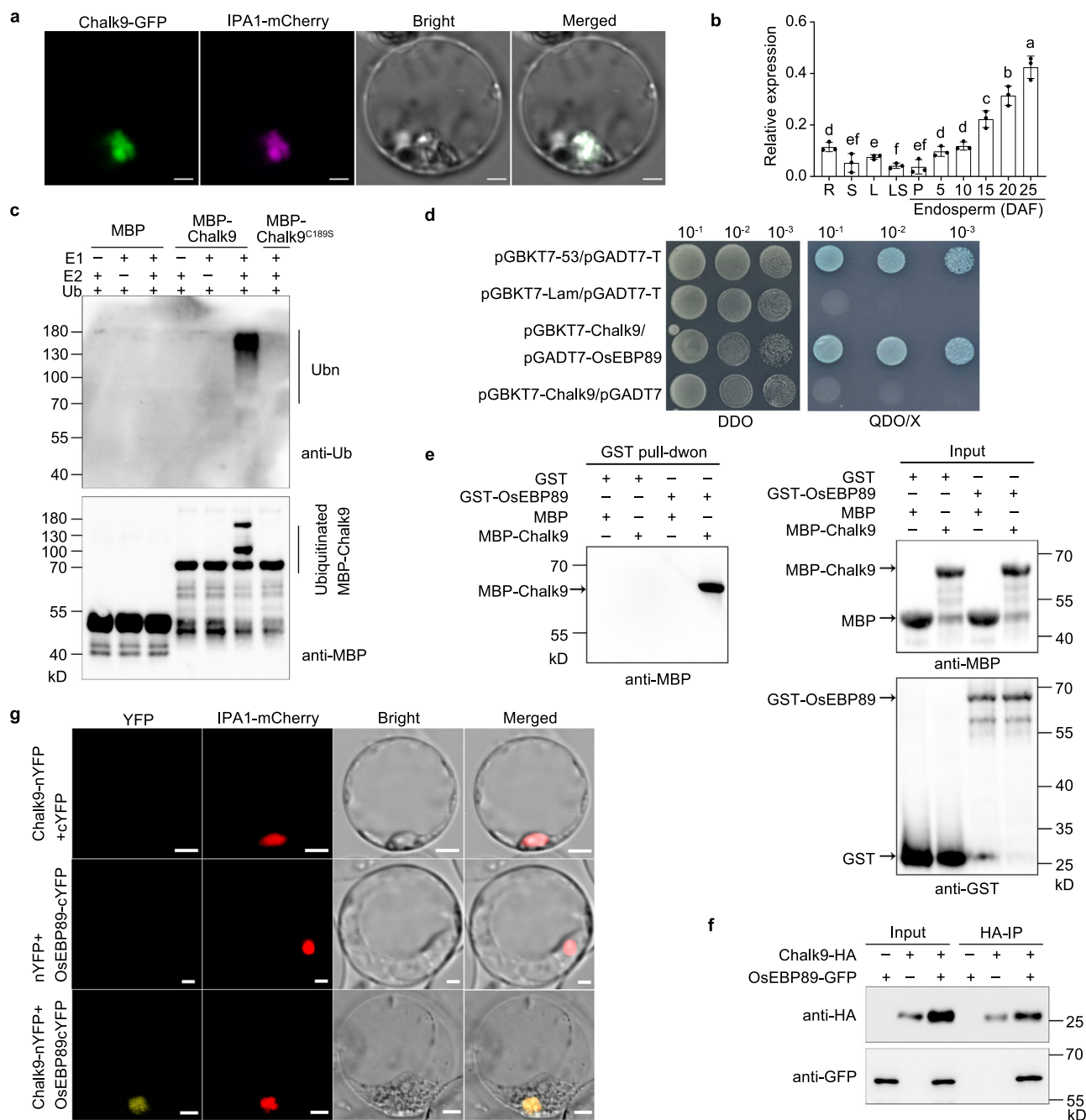
Fig. 4 | Chalk9 is an E3 ubiquitin ligase that interacts with OsEBP89.
a Subcellular localization of Chalk9-GFP fusion protein in rice protoplasts. IPA1-mCherry was used as a nuclear marker. Scale bars: 5 μm. **b** Quantitative PCR with reverse transcription (qRT–PCR)-based transcript abundance analysis of *Chalk9* in various tissues. R, root; S, stem; L, leaf; LS, leaf sheath; P, panicle; DAF, days after flowering. Data show means ± SD (*n* = 3 biological replicates). *OsActin* was used as a control. Different letters indicate significant differences (*P* < 0.05, one-way ANOVA with Tukey's multiple comparison test). **c** Ubiquitin ligase activity of Chalk9. MBP-Chalk9 was expressed in *E. coli* strain BL21, and ubiquitinated proteins were detected using both anti-MBP and anti-ubiquitin (Ub) antibodies. **d** Yeast two-hybrid (Y2H) assay showing the interaction between Chalk9 and OsEBP89. DDO and QDO/X represent SD/−Trp–Leu and SD/−Trp–Leu–His–Ade + X-α-Gal selection medium, respectively. **e** Pull-down assay. GST-OsEBP89 was used as bait, and the pull-down MBP-Chalk9 was detected by the anti-MBP antibody. **f** Co-immunoprecipitation (Co-IP) assay of rice protoplasts co-expressing Chalk9-HA and OsEBP89-GFP. The immunoprecipitants were probed with antibodies against HA and GFP. IP, immunoprecipitation. **g** Interaction between Chalk9 and OsEBP89 demonstrated by bimolecular fluorescence complementation (BiFC) assays in rice protoplasts. IPA1-mCherry was used as a nuclear control. Scale bars: 5 μm. In **a**, **c**, and **e**–**g**, the experiments have three independent experiments with similar results. Source data are provided as a Source Data file.

sequence in the *Chalk9*-L promoter contains the DNA binding elements by the OsB3 protein in rice.

Chalk9 exhibits E3 ubiquitin ligase activity

To investigate the molecular function of Chalk9, we analyzed its localization and expression pattern. The results showed that Chalk9

was localized in the nucleus and highly expressed in the developing endosperm with gradually increasing during grain filling (Fig. 4a, b). Similar results were observed through GUS staining (Supplementary Fig. 7a). Chalk9 is predicted to be a RING-C3HC4 type E3 ubiquitin ligase. To confirm its E3 ligase activity, we produced recombinant MBP-Chalk9 protein in *Escherichia coli* (*E. coli*) for in vitro

ubiquitination assays. When ubiquitin, ubiquitin-activating enzyme (E1), and ubiquitin-conjugating enzyme (E2) were present, Chalk9 underwent auto-ubiquitination, whereas no ubiquitination was detected when E1, E2, or MBP-Chalk9 was absent (Fig. 4c). A key amino acid residue site was identified through sequence alignment (Supplementary Fig. 7b) and structural analysis based on the different RING-finger domains in plants[29]. We mutated the conserved cysteine at position 189 to serine, creating the MBP-Chalk9^{CI89S} mutant (Supplementary Fig. 7b). The self-ubiquitination was abolished by the substitution in the RING finger domain (Fig. 4c), confirming that Chalk9 is a functional RING finger E3 ligase.

Additionally, both Chalk9-L and Chalk9-H proteins showed the same protein subcellular localizations (Supplementary Fig. 8a). Moreover, both Chalk9-L and Chalk9-H proteins could be auto-ubiquitinated, and the E3 ubiquitin ligase activity of Chalk9-L was similar to that of Chalk9-H (Supplementary Fig. 8b). These results indicate that the polymorphisms in the coding regions of *Chalk9*-L and *Chalk9*-H do not affect the function of Chalk9 as an E3 ubiquitin ligase.

Using yeast two-hybrid assays to screen the substrate of Chalk9, we successfully identified OsEBP89, a transcription factor involved in amylose biosynthesis[33–35], that interacts with Chalk9 (Fig. 4d, Supplementary Fig. 9a). The C-terminal domain of OsEBP89 was found to be critical for this interaction (Supplementary Fig. 9b). This interaction was further validated by in vitro pull-down (Fig. 4e, Supplementary Fig. 9c) and coimmunoprecipitation (CoIP) assays in rice protoplasts (Fig. 4f). Co-localization and bimolecular fluorescence complementation (BiFC) assays confirmed that the interaction between Chalk9 and OsEBP89 occurred in the nucleus (Fig. 4g, Supplementary Fig. 9d, e).

Given that Chalk9 functions as an active RING finger E3 ligase and interacts with OsEBP89 (Fig. 4), we hypothesized that OsEBP89 is the direct substrate of Chalk9. In vitro ubiquitination assay was performed using MBP-Chalk9 and GST-OsEBP89. In the presence of E1, E2, ubiquitin, GST-OsEBP89 was ubiquitinated by MBP-Chalk9 (Fig. 5a). In contrast, no polyubiquitination was observed in the absence of E1, E2, ubiquitin or MBP-Chalk9 (Fig. 5a). Furthermore, replacing MBP-Chalk9 with the MBP-Chalk9^{CI89S} mutant failed to ubiquitinate GST-OsEBP89 (Fig. 5a), confirming that Chalk9 targeted OsEBP89 for ubiquitination.

Since ubiquitination often leads to 26S proteasome-dependent degradation of target proteins, we tested whether Chalk9 influences the protein stability of OsEBP89 in a rice cell-free system. GST-OsEBP89 protein was expressed in *E. coli*, and purified protein was incubated in cell-free extracts from Nip and *chalk9-1* seedlings. The GST-OsEBP89 protein was found to be more stable in the *chalk9-1* mutant extract compared to Nip (Fig. 5b, c). The addition of MG132 significantly inhibited GST-OsEBP89 degradation in both Nip and *chalk9-1* extracts (Fig. 5b), indicating that Chalk9 mediated the stability of OsEBP89 in vivo through the 26S proteasome system.

To further determine OsEBP89 abundance in seeds from Nip and *chalk9* mutants, we generated a specific antibody against OsEBP89 (Supplementary Fig. 9f). OsEBP89 was more abundant in *chalk9* mutants than in Nip, although the transcript levels of *OsEBP89* remained unchanged (Fig. 5d–f). This suggests that the loss of Chalk9 function leads to the accumulation of OsEBP89 protein in rice. We also compared OsEBP89 protein levels between Nip (*Chalk9*-L type) and NIL$^{Chalk9-H}$ plants. The OsEBP89 protein level in seeds was higher in NIL$^{Chalk9-H}$ plants than in Nip (Fig. 5g, h), while *OsEBP89* expression was similar (Fig. 5i), suggesting that *Chalk9*-L promotes more degradation of OsEBP89 than *Chalk9*-H.

Chalk9–OsEBP89 module regulates grain chalkiness through regulation of the storage components in the endosperm

We observed that the *chalk9* mutants produced white-belly endosperms (Supplementary Fig. 10a). Scanning electron microscopy revealed that the chalky endosperm of the *chalk9* mutants contained loosely packed spherical starch granules interspersed with large air spaces, whereas the non-chalky endosperm of Nip consisted of densely and regularly packed polyhedral starch granules (Fig. 6a), which is consistent with previous studies[36,37]. Although the total starch content remained unchanged, the *chalk9* mutants exhibited significantly higher amylose content (Fig. 6b, c). Transmission electron microscopy further showed that chalky endosperm cells of the *chalk9* mutants contained increased numbers and larger mean areas of spherical protein body I (PBI) and irregularly shaped PBII compared to Nip (Fig. 6d–f). This observation aligned with the greatly increased levels of seed storage proteins in *chalk9* mutants, including glutelin, prolamin, and albumin (Fig. 6g–k). We further performed a transcriptome deep sequencing (RNA-seq) analysis on seeds from Nip and *chalk9-1* (Supplementary Fig. 10b). A total of 2,658 differentially expressed genes were identified in *chalk9-1* compared to Nip (Supplementary Fig. 10c, Supplementary Data 9). We found that the *Waxy* (*Wx*) gene for amylose synthesis and some genes for seed storage protein (SSP) exhibited significantly higher expression levels in *chalk9-1* compared to Nip (Supplementary Fig. 10d, Supplementary Data 10). These findings were further validated by qRT-PCR analysis, which confirmed the increased expression of related genes in *chalk9-1* (Supplementary Fig. 11).

To investigate whether *OsEBP89* is involved in chalkiness regulation, we generated *OsEBP89* knockout plants (*osebp89-1* and *osebp89-2*) using CRISPR/Cas9 (Supplementary Fig. 12a), and *OsEBP89* overexpression lines (*OsEBP89-OE1* and *OsEBP89-OE2*) driven by the constitutive CaMV 35S promoter (Supplementary Fig. 12b). Compared to the wild-type Nip, the *OsEBP89* knockout plants showed a slight yet significant reduction in both CGR and DC values, whereas the *OsEBP89*-overexpression lines exhibited markedly increased chalkiness with higher CGR and DC values (Supplementary Fig. 12c–e). These results indicate that *OsEBP89* positively regulates chalkiness in rice. Notably, a significant decrease in *Wx* expression was detected in *OsEBP89* knockout mutants, whereas its expression increased in *OsEBP89*-overexpressing plants (Supplementary Fig. 12f), which is consistent with previous studies showing that *OsEBP89* binds to the GCC box and GCC box-like sequences in the *Wx* promoter, thereby promoting its expression[33–35]. Several such binding sites were also identified in the promoters of *SSP* genes (Supplementary Data 11). The expression of *SSP* genes was greatly repressed in *OsEBP89* knockout mutants but upregulated in *OsEBP89*-overexpressing plants (Supplementary Fig. 12g–l). Furthermore, yeast one-hybrid assays demonstrated that OsEBP89 directly bound to the promoters of *SSP* genes (Supplementary Fig. 12m).

We crossed the *chalk9-1* mutant with the *osebp89-1* mutant to generate double mutant plants (*chalk9-1/osebp89-1*). While the *chalk9-1* mutant showed increased chalkiness, the *chalk9-1/osebp89-1* double mutant exhibited a reduced chalkiness, resembling the phenotype of *osebp89-1* (Fig. 7a–c). In addition, the *chalk9-1/osebp89-1* double mutant showed decreased amylose and total protein similar to *osebp89-1* mutants, while the *chalk9-1* mutant contained increased levels of both (Fig. 7d, e). Taken together, these results reveal that the Chalk9-OsEBP89 module regulated the synthesis of grain storage components by modulating the expression of genes involved in storage components, thereby influencing chalkiness formation in rice.

In addition, based on our whole-genome sequencing data, we observed that *OsEBP89* had a single major haplotype in *indica* rice (Supplementary Data 12). Extending this analysis to 4,726 accessions of cultivated rice[38–40], this major haplotype occupied 97.5% of *indica* (Supplementary Data 13), indicating the strong genetic conservation and unlikely contribution to chalkiness variation in *indica* varieties. Moreover, *OsEBP89* knockout plants did not show significant differences from Nip in major agronomic traits, including plant height, tiller number, grain length, grain width, 1000-grain weight, number of grains per panicle, number of primary branches, and seed-setting rate (Supplementary Fig. 13). These results suggest that loss-of-function of *OsEBP89* does not compromise key yield-related traits.
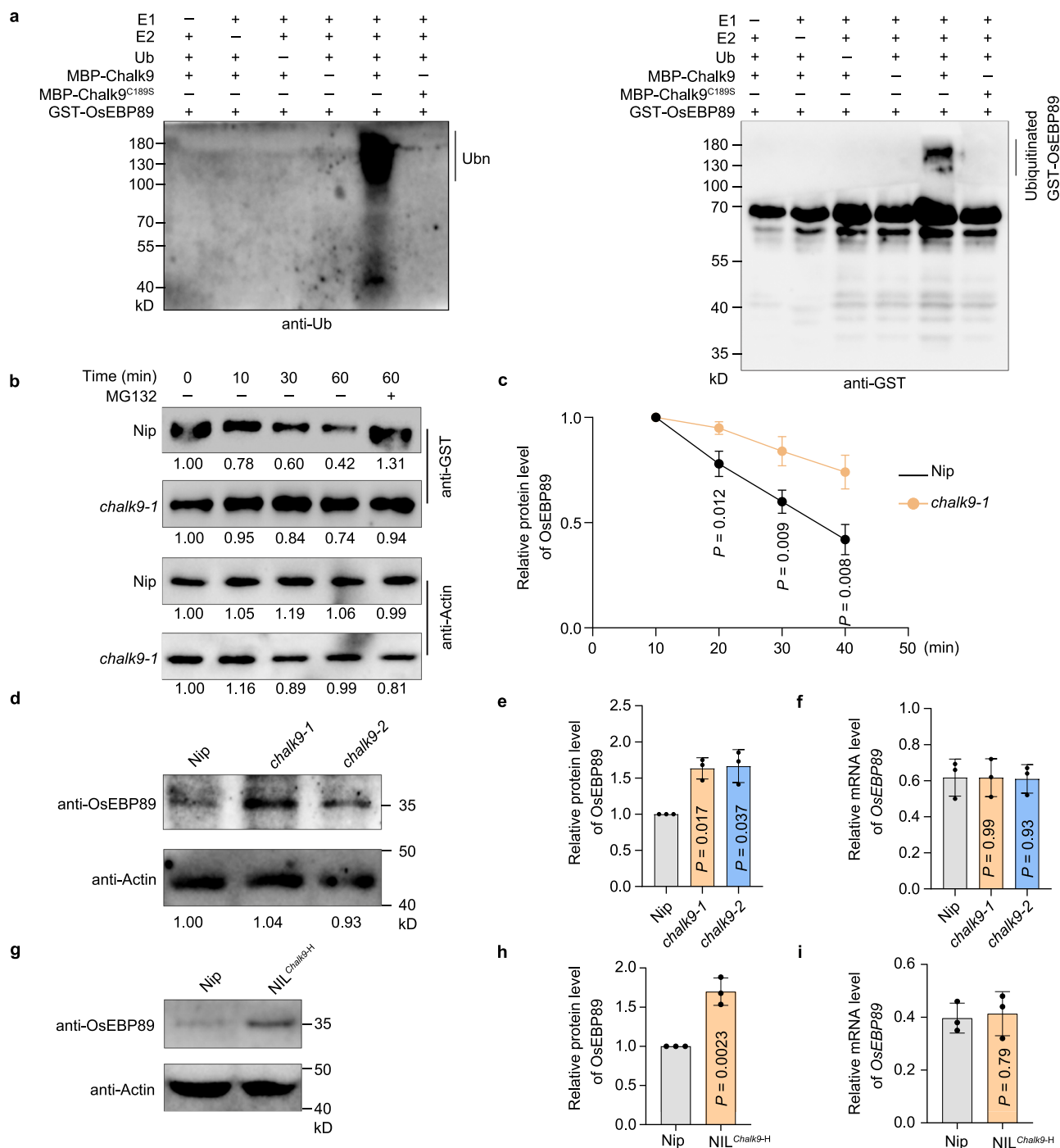
Fig. 5 | Chalk9 ubiquitinates OsEBP89 and regulates its stability. a In vitro ubiquitination of OsEBP89 by Chalk9. Ubiquitinated proteins were detected using anti-GST and anti-Ub antibodies. **b** Cell-free degradation of GST-OsEBP89 in the protein extracts from Nip and *chalk9-1* seedlings. Protein levels of GST-OsEBP89 were detected using an anti-GST antibody, and Actin was used as a loading control for total protein extraction. Relative fold changes of GST-OsEBP89 to Actin were quantified by ImageJ. The protein level at time point 0 min was marked as 1. **c** GST-OsEBP89 degradation rate in Nip and *chalk9-1* seedlings. **d** Detection of OsEBP89 protein abundance in Nip and *chalk9-1* plants. OsEBP89 protein abundance was determined by immunoblotting. **e** Relative quantification of OsEBP89 protein abundance in Nip, *chalk9-1*, and *chalk9-2* plants. The protein level of OsEBP89 in Nip was normalized to 1 within each biological replicate, and the relative protein levels

of OsEBP89 in *chalk9-1* and *chalk9-2* mutants were calculated relative to Nip. **f** *OsEBP89* mRNA level in Nip, *chalk9-1*, and *chalk9-2* plants. **g** Detection of OsEBP89 protein abundance in Nip and NIL$^{Chalk9\text{-}H}$ plants. Total proteins were extracted from seeds at 15 DAF. OsEBP89 protein abundance was determined by immunoblotting. **h** Relative quantification of OsEBP89 protein abundance in the Nip and NIL$^{Chalk9\text{-}H}$ plants. The protein level of OsEBP89 in Nip was normalized to 1 within each biological replicate, and the relative protein level of OsEBP89 in NIL$^{Chalk9\text{-}H}$ was calculated relative to Nip. **i** *OsEBP89* mRNA level in Nip and NIL$^{Chalk9\text{-}H}$ plants. In **c**, **e**, **f**, **h**, and **i**, data show means ± SD ($n = 3$ biological replicates); statistical analysis was performed by two-tailed Student's *t*-test. In **a** and **b**, the experiments have three independent experiments with similar results. Source data are provided as a Source Data file.
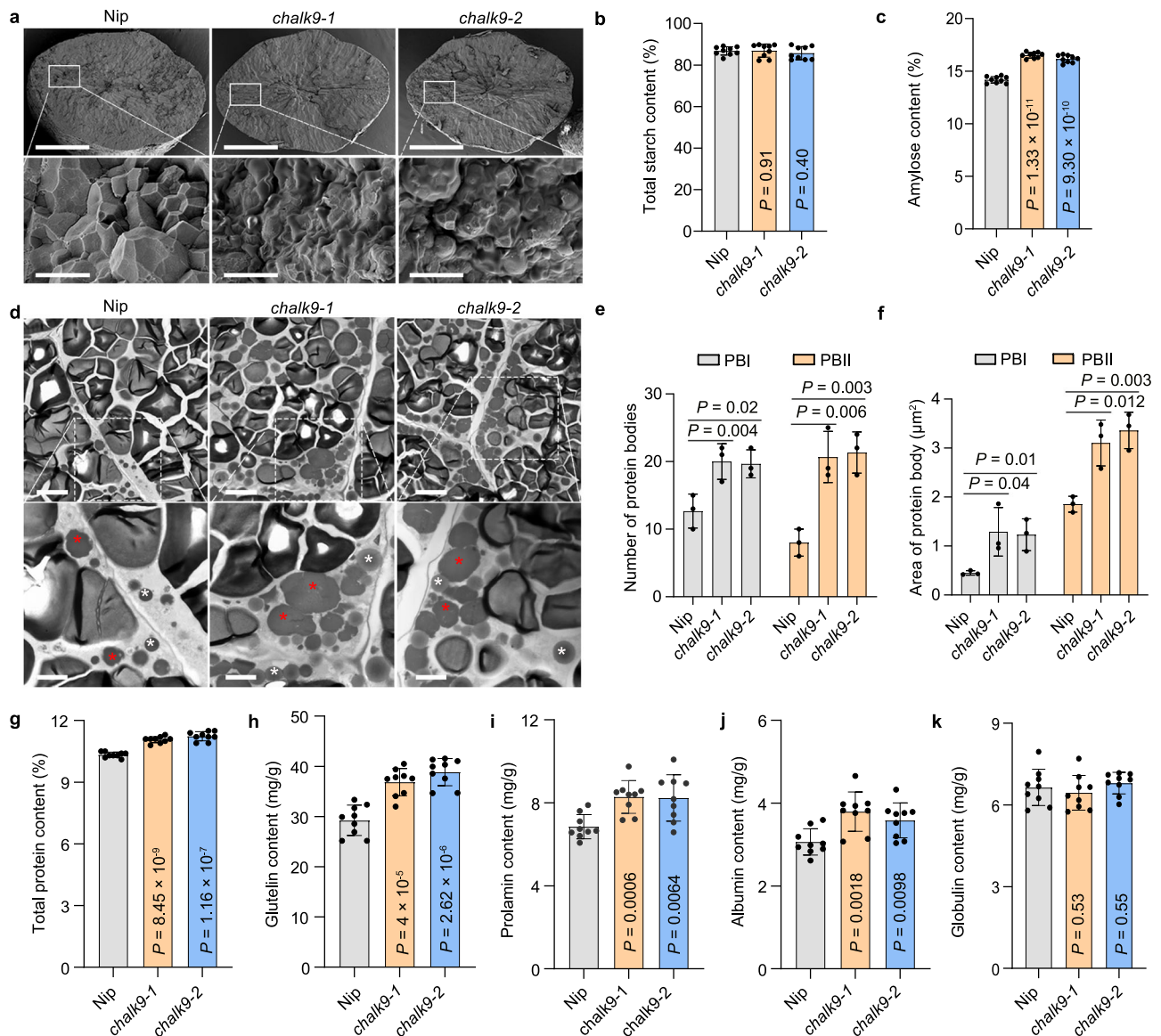
Fig. 6 | *Chalk9* regulates rice grain chalkiness by influencing seed storage substance biosynthesis. a The scanning electron microscopy observation of transverse sections of mature seeds from Nip, *chalk9-1*, and *chalk9-2* plants. Scale bars: 0.8 mm (upper), 5 μm (down). **b**, **c** Starch (**b**) and amylose (**c**) contents of Nip, *chalk9-1*, and *chalk9-2* plants. **d** Transmission electron microscopy analysis of the endosperm cells from Nip, *chalk9-1*, and *chalk9-2* plants at 18 DAF. Scale bars: 5 μm (upper), 2 μm (down). The white asterisk indicates PBI; the red asterisk indicates PBII. **e**, **f** Number (per 400 μm²) of protein bodies (**e**) and mean area of protein bodies (**f**) in the endosperms from Nip, *chalk9-1*, and *chalk9-2* plants. **g**–**k** Total protein (**g**), glutelin (**h**), prolamin (**i**), albumin (**j**), and globulin (**k**) contents of Nip, *chalk9-1*, and *chalk9-2* plants. Data show means ± SD (*n* = 9 plants in **b**, **c**, and **g**–**k**; *n* = 3 biological replicates in **e** and **f**). In **b**, **c**, **e**, **f**, and **g**–**k**, statistical analysis was performed by two-tailed Student's *t*-test. Source data are provided as a Source Data file.

Chalk9-L is artificially selected in cultivated rice during domestication and breeding

We performed a geographic distribution analysis of haplotypes in 1,424 cultivated varieties from the 3 K Rice Genomes Project[38]. The distribution of rice accessions carrying either *Chalk9*-L or *Chalk9*-H was variable across Asia regions relative to other areas (Supplementary Fig. 14a). The frequency of *Chalk9*-L was nearly 100% in Southeast Asia (e.g., Myanmar, Philippines, Laos, and Thailand), but it was relatively lower in China (71.1%) and South Asia, including Bangladesh (62%), Nepal (68.1%), Pakistan (70%), and India (76.2%) (Supplementary Fig. 14a). We further performed haplotype analysis in 4,726 accessions of cultivated rice[39–41]. Eight out of 9 unique high-confidence haplotypes belonged to the *Chalk9*-L group, while only one belonged to the *Chalk9*-H group (Supplementary Data 14). *Chalk9*-L was present in 12.3% of Aus, 85.3% of aromatic, 99.9% of *japonica*, and 80.1% of *indica*

varieties (Supplementary Table 3). Within the *indica* subgroups, its frequency was 40.9% in *indica* I, 96.6% in *indica* II, 94% in *indica* III, and 84% in *indica* intermediate (Supplementary Table 3). In 445 accessions of the wild ancestor *Oryza rufipogon* (*O. rufipogon*)[39], *O. rufipogon* had a high frequency of *Chalk9*-L (89.4%) (Supplementary Table 4). These results suggest that the allele distribution of *Chalk9* in different rice subgroups may be correlated to their evolution and selection.

A selective sweep surrounding the *Chalk9* locus was observed between *japonica* and wild rice, with significantly reduced nucleotide diversity in *japonica* compared to wild rice (Fig. 8a), indicating a strong artificial selection in *Chalk9* locus of *japonica*. Tajima's *D* values in the *Chalk9* locus were significantly negative in *japonica* (Fig. 8b), reflecting directional selection across this region. In contrast, no obvious selection was detected in *indica* because the relative ratio of nucleotide diversity in *indica* to wild rice was higher than that in *japonica* to wild
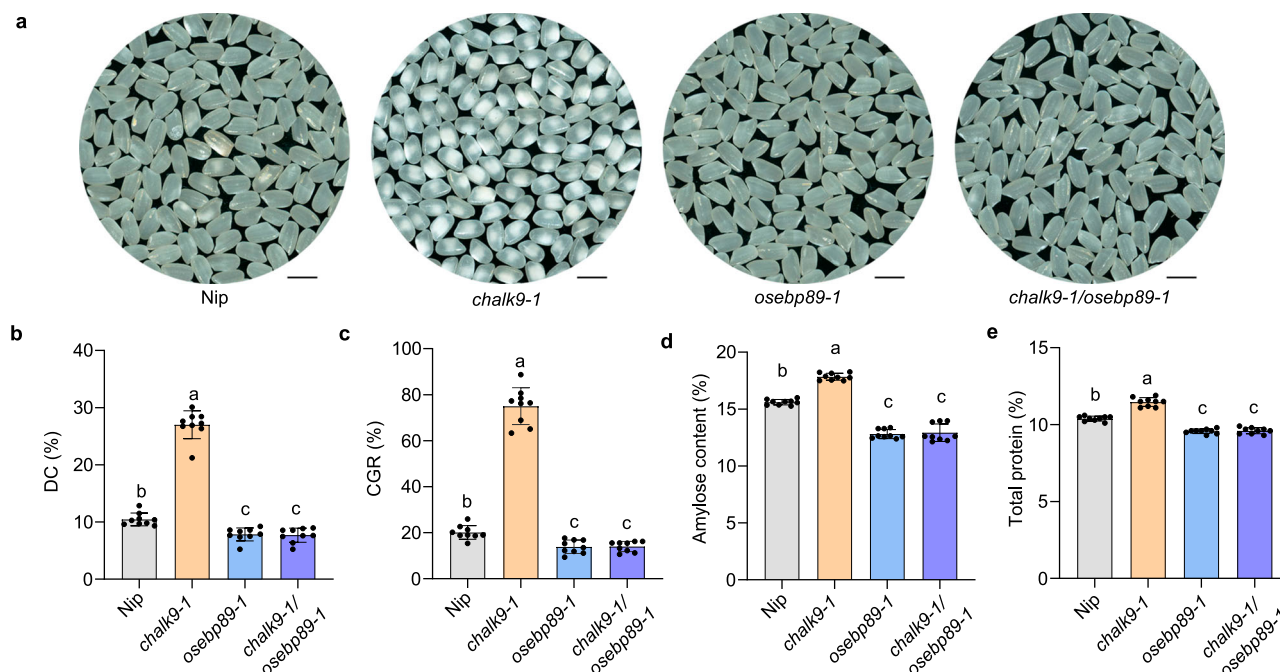
Fig. 7 | Chalk9-OsEBP89 module regulates rice grain chalkiness. a Grain chalkiness of Nip, *chalk9-1*, *osebp89-1* and *chalk9-1/osebp89-1* plants. Scale bar: 5 mm. **b**–**e** Degree of chalkiness (**b**), chalky grain rate (**c**), amylose (**d**), and total protein (**e**) in Nip, *chalk9-1*, *osebp89-1*, and *chalk9-1/osebp89-1* plants. In **b**–**e**, data show means ± SD (*n* = 9 plants); different letters indicate significant differences (*P* < 0.05, one-way ANOVA with Tukey's multiple comparison test); for *P* values, see Source Data. Source data are provided as a Source Data file.

rice in *Chalk9* locus (Fig. 8a). Further phylogenetic analysis showed that the *Chalk9*-L haplotype in *japonica* rice formed a tight cluster, while in *indica* rice, *Chalk9*-L was more widely distributed and genetically diverse (Fig. 8c). Haplotype network also showed that *Chalk9*-L in *japonica* was closely related to *O. rufipogon*, with few mutational differences, whereas *Chalk9*-L in *indica* exhibited more complex connections and mutational steps (Fig. 8d), suggesting that *Chalk9*-L in *japonica* evolved from *O. rufipogon* through a single lineage, while *Chalk9*-L in *indica* had a more complex evolution history with multiple origins.

To trace the selection of *Chalk9*-L during *indica* rice breeding, we developed a 64-bp InDel marker in the *Chalk9* promoter and genotyped *Chalk9* in 127 *indica* varieties from the 1950s to the 2000s. The frequency of *Chalk9*-L in varieties prior to 1990 was relatively low, but it increased significantly thereafter (Supplementary Fig. 14b). This trend aligns with the significant reduction of chalkiness observed in *indica* varieties post-1990 (Supplementary Fig. 14c, d), indicating that *Chalk9*-L has been artificially selected in modern *indica* rice breeding programs. All 123 *japonica* varieties carried *Chalk9*-L (Supplementary Fig. 14b), consistent with the lower chalkiness observed (Supplementary Fig. 14e, f). These findings suggest that *Chalk9*-L might have been under artificial selection to reduce chalkiness. In addition, we investigated the distribution of *Chalk9*-L in 105 elite *indica* varieties released over the past ten years in China and found that 81% of varieties harbor the elite allele (Supplementary Fig. 14g). This indicates that significant progress has been made in *indica* rice breeding for improved grain quality, particularly in reducing chalkiness, in China over the past decade.

Chalk9-L holds the potential for breeding low-chalky rice cultivars without yield penalty

We further investigated the effect of *Chalk9* on yield. *Chalk9* knockout plants displayed no significant differences from Nip in major agronomic traits, including heading date, tiller number, plant height, grain size, and weight, as well as yield per plant and yield per plot

(Supplementary Fig. 15). These results suggest that *Chalk9* has no impact on rice yield. NIL[Chalk9-H] plants showed no significant differences in grain weight or yield per plant compared to Nip (*Chalk9*-L type) (Supplementary Fig. 16a, b). Furthermore, introducing the *Chalk9*-L transgene into the high-yield variety Guichao2 significantly reduced chalkiness without affecting other agronomic traits, particularly yield per plant (Supplementary Fig. 16c–h), demonstrating the potential of *Chalk9*-L to reduce chalkiness in high-yield rice cultivars without compromising productivity.

To investigate the effect of *Chalk9* on grain filling, we analyzed the grain filling rate in wild-type Nip and *chalk9-1* mutant plants. Compared to Nip, *chalk9-1* showed no significant differences in the grain filling rate from 5 to 35 DAF (Supplementary Fig. 17a), suggesting that *Chalk9* does not affect the grain filling process. We further measured sucrose content in the endosperms of Nip and *chalk9-1* plants at 5 and 10 DAF. No significant differences were observed between the two genotypes (Supplementary Fig. 17b, c), indicating that *Chalk9* does not disrupt photosynthetic assimilate supply for the synthesis of endosperm storage components.

Discussion

To date, little progress has been made in understanding the genetic and molecular mechanisms underlying natural variation associated with chalkiness in rice. Here, we reported that *Chalk9* is the major gene controlling chalkiness variation in *indica* rice. A 64-bp indel variant in *Chalk9* promoter leads to differing expression levels, conferring chalkiness variation among rice varieties. Moreover, we deciphered a Chalk9-OsEBP89 regulatory module that mediates chalkiness variation (Fig. 8e). These findings deepen our understanding of the genetic and molecular mechanisms underlying grain chalkiness variation in rice.

Developing high-yielding rice with superior quality is challenging for rice breeding due to the trade-off between these traits[2]. One notable reason is that many QTLs associated with chalkiness are closely linked to yield-associated genes[12,20]. Fortunately, *Chalk9* does not exhibit such a linkage drag, as the yield in its near-isogenic lines shows
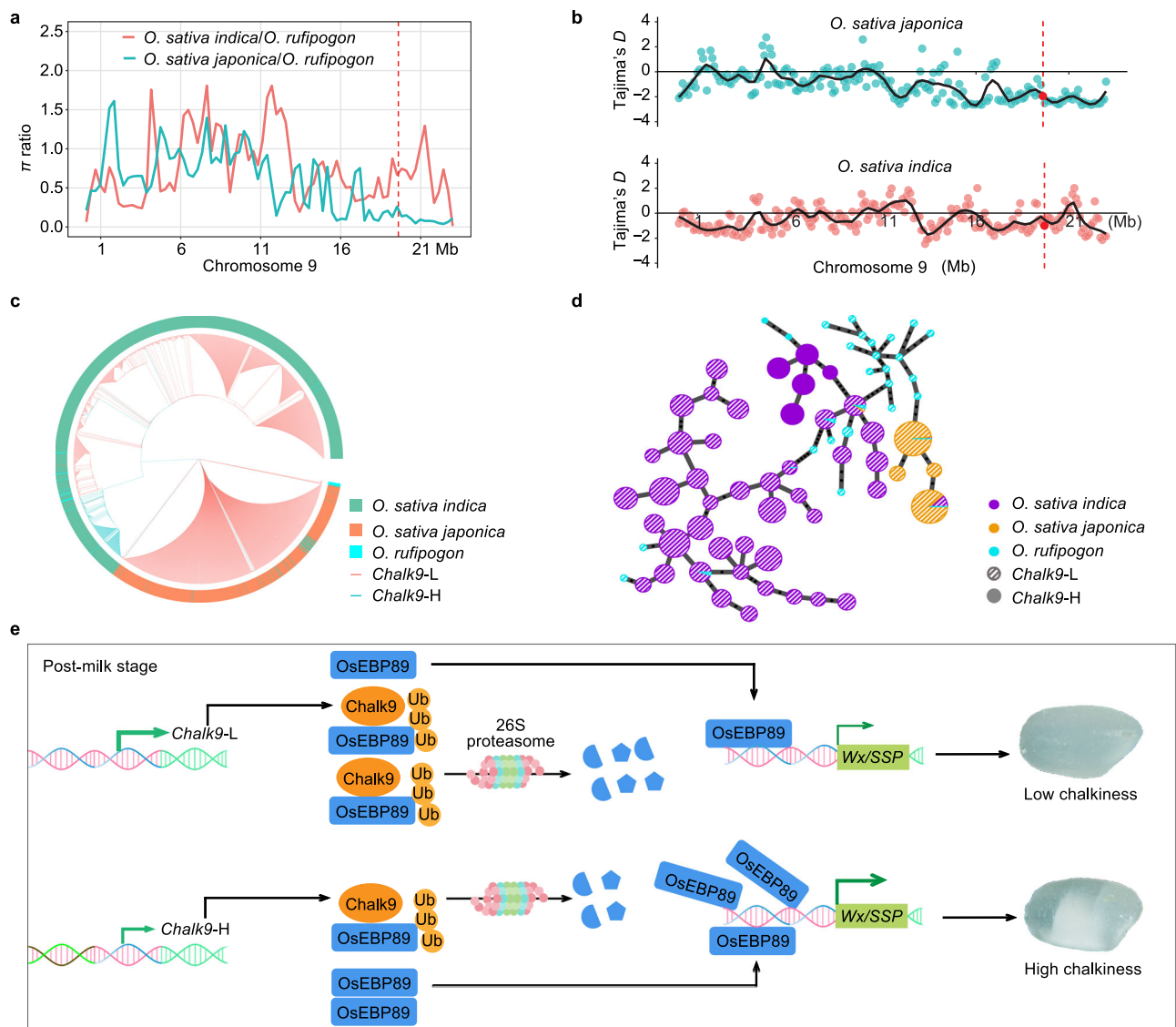
Fig. 8 | Geographical distribution, genomic differentiation, and genomic selection of *Chalk9* between *japonica* and *indica* subspecies. a, b The relative ratio of nucleotide diversity (**a**) and Tajima's *D* (**b**) analyses in the whole chromosome 9 of cultivated and wild rice. The red dashed line indicates the *Chalk9* locus. **c, d** Phylogeny (**c**) and haplotype network (**d**) generated from the genomic sequences of *Chalk9* in both cultivated and wild rice varieties. The outer circle of the tree indicates various rice populations. The circle size of the network is proportional to the number of samples for each haplotype. Black spots on the lines indicate mutational steps between two haplotypes. **e** A proposed model for the Chalk9−OsEBP89 module in the regulation of grain chalkiness. In rice varieties with the *Chalk9*-H allele, *Chalk9* expression in the endosperm is relatively lower, which reduces the degradation of OsEBP89. This accumulation of OsEBP89 leads to the upregulation of *Wx* and *SSP* genes, resulting in increased levels of amylose and storage protein in the endosperm. This elevated synthesis of storage compound during the post-milk stage contributes to the formation of chalky grains. Conversely, in rice varieties with the *Chalk9*-L allele, *Chalk9* is highly expressed, which accelerates OsEBP89 degradation. The reduction in OsEBP89 levels leads to the downregulation of *Wx* and *SSP* genes in the endosperm, resulting in decreased synthesis of storage substances during the post-milk stage and the formation of translucent grains.

no significant difference compared to the wild type (Supplementary Fig. 16a, b). *Chalk9*-L as an elite haplotype showed increased *Chalk9* expression, conferring reduced chalkiness (Fig. 3a−g). By introducing this favorable allele into a well-known high-yielding *indica* variety but with high chalkiness, the chalkiness in the lines was significantly decreased but did not compromise yield (Fig. 3h−j, Supplementary Fig. 16g, h). This is primarily attributed to the fact that *Chalk9* does not affect grain filling or assimilate supply (Supplementary Fig. 17).

Interestingly, Li et al. [42]. have demonstrated that different *Wx* alleles significantly influence amylose content and chalkiness formation. Specifically, *Wx^a* is associated with high amylose content (> 24%) and increased chalkiness, whereas *Wx^b* corresponds to moderate amylose content (~16%) and reduced chalkiness. Using CRISPR/Cas9 to edit the *Wx^a* promoter region, the downregulation of *Wx* expression

reduces amylose content and effectively alleviates chalkiness formation. Similarly, a recent study revealed that *LCG1* regulates *Wx* expression through the OsBP5/OsEBP89 complex, thereby influencing chalkiness formation[22]. Lower *LCG1* expression in *indica* rice leads to higher amylose content and increased chalkiness, while higher *LCG1* expression in *japonica* rice reduces both traits. Consistently, our study observed increased amylose content in chalky endosperms (Fig. 6b, c), further supporting the involvement of *Wx* in grain chalkiness formation.

High temperature is a critical environmental factor that significantly compromises rice quality. Heat stress during the grain-filling stage disrupts endosperm development, accelerating chalkiness formation[10,12]. Although our study primarily focused on the genetic mechanisms underlying chalkiness formation in natural populations,

the potential interaction between temperature stress and *Chalk9* alleles remains to be explored. Future studies investigating whether *Chalk9*-L contributes to improved rice quality under high-temperature conditions would provide valuable insights into the environmental adaptability of *Chalk9* variants and their potential application in breeding heat-tolerant rice varieties.

The distribution of *Chalk9*-L in cultivated rice appears to have been influenced by evolution and artificial selection during domestication and breeding. Our evolutionary analysis revealed that *Chalk9* originated from wild rice but diverged significantly between *japonica* and *indica* rice (Fig. 8a–d). In *japonica* rice, *Chalk9*-L is likely derived from a single origin in *O. rufipogon*, while, in *indica* rice, *Chalk9*-L has multiple origins and exhibits greater genetic diversity. Moreover, the increasing incorporation of *Chalk9*-L in modern *indica* breeding programs has contributed to a significant reduction of chalkiness. In the light of that approximately 30% of *indica* varieties lack *Chalk9*-L and that *Chalk9* explains 28% of the variance in chalkiness phenotype, our results strongly indicate that *Chalk9*-L is a key target for improving rice appearance quality of *indica* rice.

The accumulated knowledge showed that the regulatory regions of genes involved in starch and storage protein biosynthesis usually share common motifs, which facilitates their co-regulation by common transcription factors, such as OsNAC20 and OsNAC26 in rice[43]. Similarly, OsEBP89 not only influences *Wx* expression but also regulates the expression of part of *SSP* genes, thereby coordinating the synthesis of amylose and storage proteins (Supplementary Fig. 12). In addition, Chalk9 acts as an E3 ubiquitin ligase, targeting OsEBP89 for ubiquitination and subsequent degradation via the 26S proteasome pathway (Figs. 4, 5). This discovery underscores the critical role of the 26S proteasome in maintaining OsEBP89 protein homeostasis. Notably, recent research showed that OsSK41 phosphorylates OsEBP89, thereby reducing its stability[35]. Whether this phosphorylation is involved in Chalk9-mediated degradation of OsEBP89 remains to be elucidated.

We propose that *OsEBP89* is a positive regulator of chalkiness in rice. Genetic analysis demonstrates that *Chalk9* operates in an *OsEBP89*-dependent manner to modulate the expression of genes involved in the biosynthesis of storage substances, thereby influencing chalkiness (Figs. 6, 7, Supplementary Figs. 10, 11). Notably, *OsEBP89* exhibits a single major haplotype in *indica* varieties, highlighting its high conservation in *indica* rice. Consequently, the variation in chalkiness observed in *indica* rice is largely attributed to genetic variation in *Chalk9*. Moreover, our findings suggest that OsB3 acts as a potential upstream regulator of *Chalk9*, mediating its differential expression in response to the 64-bp indel. Future studies should aim to elucidate the role of *OsB3* in regulating chalkiness and its contribution to chalkiness variation in rice. These efforts will help elucidate the *OsB3-Chalk9-OsEBP89* pathway in chalkiness regulation.

Endosperm development involves the coordinated synthesis and accumulation of storage substances, a process closely associated with chalkiness. This developmental progress begins in the pre-milk stage, peaks during mid-milk, and tapers off in the post-milk stage[44–46]. Similarly, *Wx* and *SSP* genes, which are central to this process, exhibit finely tuned temporal expression patterns that align with the synthesis of storage compounds[47,48]. This coordination is crucial for optimizing grain quality by balancing biosynthetic processes that determine grain texture and appearance. Our findings reveal that *Chalk9* expression gradually increases during endosperm development, reaching its peak in the post-milk stage (Fig. 4b), a period when the synthesis of storage substances naturally declines. At this stage, *Chalk9* functions as a regulatory "brake", limiting storage substance accumulation by promoting OsEBP89 degradation. This regulatory mechanism aligns with the natural decline in storage substance synthesis, supporting seed maturation and contributing to the formation of translucent grains. Thus, we propose a model in which the Chalk9-OsEBP89 regulatory module governs

chalkiness variation in rice (Fig. 8e). In rice varieties carrying the *Chalk9*-H allele, reduced *Chalk9* expression leads to OsEBP89 stabilization, which subsequently upregulates the expression of *Wx* and *SSP* genes. This increased synthesis of storage compounds disrupts the natural decline in their accumulation during the post-milk stage, resulting in the formation of chalky endosperm. In contrast, the *Chalk9*-L allele enhances *Chalk9* expression, promoting OsEBP89 degradation. This reduction in OsEBP89 levels downregulates the expression of *Wx* and *SSP* genes, reducing storage product synthesis during the post-milk stage, leading to translucent grains and improved grain quality.

Methods

Plant materials and genotyping
All 175 *indica* accessions, obtained from germplasm banks and breeders around the world, are listed in Supplementary Data 1. The *japonica* rice varieties (Nip and ZH11) and the *indica* rice varieties (YD6 and Guichao2) were in this study. All rice materials used in this study were cultivated simultaneously during the summer in paddy fields at the experimental station of Yangzhou University, located in Yangzhou, China. The plants were grown under standardized crop management practices.

Total genomic DNA was extracted from the samples and used to generate DNA sequencing libraries. Sequencing was performed by the Bioacme Biotechnology Co., Ltd (Wuhan, China). The resulting libraries were size-checked using an Agilent 2100 Bioanalyzer system. The libraries were ultimately sequenced on an Illumina Xten platform, producing 150 bp paired-end reads. After removing nucleotide variations with missing rates ≥ 0.25 and minor allele frequency <0.05, all nucleotide polymorphisms were categorized based on their location in the reference genome.

Measurements of grain chalkiness and storage components
Seeds harvested after full maturation were air-dried, stored at room temperature for three months. Images of 200−300 polished rice grains, randomly selected from each plant, were captured using a ScanWizard EZ scanner and analyzed with the rice quality TS-G automated analysis system (Hangzhou Shansheng Testing Technology Co., China). For chalkiness traits, the chalky grain rate (CGR) refers to the proportion of chalky grains among all rice grains, while the degree of chalkiness (DC) represents the extent of chalkiness. Total starch, amylose, total protein, and storage protein fractions were measured from rice flour[49]. In detail, the total protein content was analyzed with the kjeltec automatic nitrogen analyzer (FOSS, Hilleroad, Denmark). The content of total starch, amylose and 4 kinds of storage protein was measured using a spectrophotometer (BioTek, Winooski, Vermont, United States).

Genome-wide association study
GCR and DC were surveyed in 175 *indica* varieties over two years (2021 and 2023) and subsequently used for genome-wide association studies (GWAS). The analysis was performed using GEMMA (version 0.941), which fits a linear mixed model[50]. The P-value threshold for significance was set at 1×10^{-5} using the Bonferroni correction[51], and the leading SNP was determined to be the SNP with the minimum P-value in the associated signal. The phenotypic variance explained (PVE) was calculated using the standard formula[52] implemented in GEMMA, based on the effect size (β), standard error (SE), minor allele frequency (MAF), and sample size (N) derived from the GWAS analysis. Linkage disequilibrium (LD), evaluated as r^2, between SNPs in the 175 varieties was calculated using plink v1.9[53], and LD heatmap surrounding the peak region was constructed using LDBlockShow v1.40[54].

Constructs for genetic transformation
For the *Chalk9* RNA-interference vector, *Chalk9*-specific sequences from the coding region were amplified, and inserted in both sense and

antisense orientations into a modified pTAC303-RNAi vector. For the *Chalk9* overexpression vector, the full-length coding sequence of *Chalk9* from Nip was inserted into the pCAMBIA2300-*35S* vector to generate the pCAMBIA2300-*35S:Chalk9* construct. For the *Chalk9* knockout vectors, two single-guide RNA (sgRNA) sequences targeting the *Chalk9* coding region were cloned into the pYLCRISPR/Cas9-MH vector to generate the *Chalk9* CRISPR-Cas9 construct. Additionally, two sgRNA sequences targeting the *Chalk9* promoter surrounding the 64-bp indel were designed and inserted into the pYLCRISPR/Cas9-MH vector to generate the *Chalk9* promoter-editing construct. For the *Chalk9* promoter-GUS vector, a 2-kb genomic upstream region of *Chalk9* was amplified and cloned into the pCAMBIA1381z vector.

For the *pChalk9-L::Chalk9-L* vector, the 2,645-bp genomic region including the 2-kb upstream sequence and the 645-bp coding sequence was amplified from low-chalky variety IR72 (*Chalk9*-L type) genomic sequence and cloned into plant binary vector pCAMBIA2300. The construct *pChalk9-H::Chalk9-H* contains the 2-kb upstream sequence and the 645-bp coding sequence from high-chalky variety Guichao2 (*Chalk9*-H type). The 645-bp coding sequence from Guichao2 was driven by the 2-kb promoter sequence from IR72 to generate the *pChalk9-L::Chalk9-H* construct.

For the *OsEBP89* knockout vector, two sgRNA sequences targeting the *OsEBP89* coding region were cloned into the pYLCRISPR/Cas9-MH vector to generate the *OsEBP89* CRISPR-Cas9 construct. For the *OsEBP89* overexpression vector, the full-length coding sequence of *OsEBP89* from Nip was inserted into the pCAMBIA2300-*35S* vector to generate the pCAMBIA2300-*35S:OsEBP89* construct. Agrobacterium-mediated transformation was used to generate transgenic rice plants. Primer sequences used in this study are listed in Supplementary Data 15.

GUS analysis
To determine the spatial expression patterns of *Chalk9*, various rice tissues, including young roots, stems, leaf sheaths, leaves, spikelets, and developing seeds from *proChalk9::GUS* transgenic plants, were stained with a GUS staining kit (Coolaber Biotech, SL7160) at 37 °C in the dark for 12 h. The stained tissues were then decolorized with 100% ethanol and observed using a microscope (OLYMPUS, MVX10, Japan).

Gene expression analysis
Total RNA was extracted from rice samples for first-strand cDNA synthesis. After synthesizing first-strand cDNA, quantitative PCR was performed using ChamQ SYBR qPCR Master Mix (Vazyme Biotech, Q713). Data analysis was conducted from three replicates for each experiment, using *OsActin* (*LOC_Os10g36650*) as the internal reference. Gene-specific primers are listed in Supplementary Data 15.

Transcriptome deep sequencing (RNA-seq) analysis
Seeds harvested at 20 days after flowering (DAF) were used for total RNA extraction. RNA-seq libraries were prepared in triplicate from wild-type Nip and *chalk9-1* mutant samples. RNA-seq was performed by the Bioacme Biotechnology Co., Ltd (Wuhan, China). Paired-end sequencing (150 bp) was conducted using an Illumina HiSeq 2500. Raw reads were cleaned using FastQC and mapped to the reference genome (*Oryza sativa* L. *japonica* cv. Nipponbare) using Tophat v. 2.0.10. The counts per gene were normalized as fragments per kilobase of transcript per million mapped reads (FPKM) values before analyzing transcript abundance. Differentially expressed genes were identified using DESeq2 with a *P*-value < 0.05 and $|\log_2(\text{fold change})| > 1$. Correlation analysis, heatmap plotting, and volcano plot analysis were performed using BMKCloud (www.biocloud.net).

Transmission electron microscopy
Seeds from WT and *chalk9-1* mutant plants at 18 DAF were collected for transmission electron microscopy (TEM) analysis. The TEM samples were fixed and prepared as previously described[55], with slight modifications. In brief, developing seeds were harvested and immediately fixed in a solution containing 4% (v/v) paraformaldehyde and 0.1% (v/v) glutaraldehyde at 4 °C overnight. Fixed tissues were then dehydrated through a graded ethanol series (30%, 50%, 70%, 80%, 90%, and 100%) and embedded in resin. Micrographs of the endosperm cells were captured on 80-nm ultra-thin sections using a transmission electron microscope.

Scanning electron microscopy
For scanning electron microscopy (SEM), brown rice grains were naturally broken from the middle and then coated with gold using an E-100 ion sputter coater. The morphology of starch granules was observed using a scanning electron microscope (S570, Hitachi, Tokyo, Japan) at an accelerating voltage of 10 kV and a spot size of 30 nm. At least three biological replicates from different mature grains were analyzed.

Protoplast isolation and transfection
The protoplast isolation and transfection assays were carried out as described in a previous publication[56], with some modifications. In detail, rice protoplasts were isolated from 10-day-old seedlings cultured on 1/2 MS medium in the dark. Stem and sheath tissues from 40-60 seedlings were cut into 0.5 mm strips and immediately treated with 0.6 M mannitol for plasmolysis. Following the removal of mannitol, the strips were enzymatically digested for 4 h in the dark with gentle shaking. The enzyme solution consisted of 1.5% (v/v) cellulase RS, 0.75% (v/v) macerozyme R10, 0.6 M mannitol, 10 mM MES (pH 5.7), 10 mM $CaCl_2$, and 0.1% (v/v) BSA. After enzymatic digestion, an equal volume of W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, and 2 mM MES, pH 5.7) was added to terminate the enzymatic reaction. Protoplasts were collected by filtration through 40 μm nylon meshes. After washing once with W5 solution, the pellets were resuspended in MMG solution (0.4 M mannitol, 15 mM $MgCl_2$, and 4 mM MES, pH 5.7). 10 μg of plasmid DNA were mixed with protoplasts, and freshly prepared PEG solution (40% [v/v] PEG 4000, 0.2 M mannitol, and 0.1 M $CaCl_2$) was added; the mixture was incubated at room temperature for 20 min in the dark. After incubation, W5 solution was added slowly to stop the reaction, and the protoplasts were collected by centrifugation. Finally, the transfected protoplasts were cultured in W5 solution at room temperature for 16 h.

Subcellular colocalization assay
The coding regions of *IPA1* and *Chalk9* were amplified by PCR and individually cloned into the 163-mCherry and 163-GFP plasmids, respectively. Protoplasts isolated from 10-day-old Nip rice seedlings were transfected with the constructs. GFP and mCherry were excited with 488-nm and 543-nm laser lines, respectively, and fluorescence signals were detected at 500-580 nm for GFP and 565-615 nm for mCherry using confocal laser-scanning microscopy. Images presented in the figures are representative of at least five protoplasts.

Y2H assay
For the Y2H screening, developing seeds at the reproductive stage were combined to construct a two-hybrid library by Shanghai OE Biotech Company. The coding sequence of *Chalk9* was cloned into the pGBKT7 vector to generate the bait vector (pGBKT7-Chalk9), which was used to screen the library through yeast mating. To confirm interactions, plasmids from positive clones were co-transformed with pGBKT7-Chalk9 into fresh Y2HGold cells and retested. To identify which domains of OsEBP89 mediate its interaction with Chalk9, we constructed three truncated proteins of OsEBP89 (OsEBP89 [1-119], OsEBP89 [120-201], and OsEBP89 [202-326]) in the pGADT7 prey vector and tested their binding to Chalk9 (in pGBKT7) using the yeast two-hybrid system. For the reciprocal bait-prey validation, the coding

sequences of *OsEBP89* and *Chalk9* were cloned into the pGBKT7 (bait) and pGADT7 (prey) vectors, respectively. The yeast strain Y2HGold was employed for transformation. The pGBKT7-53 and pGADT7-T vectors were used as positive controls, while pGBKT7-Lam and pGADT7-T served as negative controls. Transformed yeast cells were plated onto synthetic defined medium lacking tryptophan, leucine, adenine, and histidine, and containing 40 ng/mL X-α-Gal, to verify interactions. Y2H assays were performed according to the manufacturer's instructions (Clontech). Primers used for construction are listed in Supplementary Data 15.

BiFC assay
For BiFC assays, reciprocal constructs were generated. For example, the coding regions of *OsEBP89* and *Chalk9* were amplified and cloned into the pUC-SPYCE and pUC-SPYNE vectors, respectively, and vice versa. The IPA1-mCherry vector served as a nuclear marker. The transfected protoplasts with the indicated constructs were observed using a fluorescence microscope (Leica TCS SP5), and images were analyzed with Image LAS-AF software. Yellow fluorescent protein (YFP) and mCherry were excited with 514-nm and 543-nm laser lines, respectively, and detected at 522-555 nm and 565-615 nm. Images presented in the figures are representative of at least five protoplasts. Primers used for construction are listed in Supplementary Data 15.

Co-IP assay
The coding sequences of *OsEBP89* and *Chalk9* were amplified and cloned into the 163-GFP and pUC35S-HA vectors, respectively, to generate the OsEBP89-GFP and Chalk9-HA constructs. The resulting constructs were transfected into protoplasts isolated from 10-day-old rice seedlings. Three sets of protoplast transformation experiments were established: (1) OsEBP89-GFP alone, (2) Chalk9-HA alone, and (3) co-transfection with OsEBP89-GFP and Chalk9-HA. For the Co-IP assay, total proteins were extracted from these transfected protoplasts using Co-IP buffer containing 100 mM HEPES (pH 7.5), 200 mM KCl, 2 mM EDTA, 2 mM MgCl$_2$, 0.4 mM CaCl$_2$, 0.5 mM dithiothreitol (DTT), 0.5% (v/v) Triton X-100, and 1× proteinase inhibitor cocktail (Roche). The total protein extracts were immunoprecipitated using anti-HA beads at 4 °C for 2 h. The immunoprecipitants were resuspended in an SDS loading sample buffer and boiled for 5 minutes. The supernatants were centrifuged and collected. Eluted proteins and Co-IP input samples were analyzed by immunoblotting with anti-HA (1:3000, ab9110, Abcam) and anti-GFP (1:3000, ab290, Abcam) antibodies.

In vitro pull-down assays
The coding sequences of *OsEBP89* and *Chalk9* were cloned into the pGEX-5X-1 and pMAL-c5X vectors, respectively, to produce GST-OsEBP89 and MBP-Chalk9. The resulting constructs, along with the empty vectors, were individually transformed into *E. coli* BL21 and induced with 0.2 mM IPTG for 12 h at 16 °C. The GST-tagged and MBP-tagged proteins were then purified using glutathione-sepharose resins (CW0190S; CWBIO) or amylose resins (E8021V; NEB), respectively, for Pull-down assays. Before performing the pull-down assays, we validated the presence of the target proteins in the input samples using anti-GST (CW0084M; CWBIO) or anti-MBP (HT701; Transgene) antibodies. For the GST pull-down assay, the GST-tagged proteins were coupled to glutathione-sepharose resins and incubated with MBP-tagged proteins for 3 h at 4 °C. The beads were then washed five times with PBS buffer (100 mM NaCl, 2 mM KCl, 5 mM Na$_2$HPO$_4$, 1.5 mM KH$_2$PO$_4$, pH 7.4) to remove non-specific interactions. The bound proteins were eluted by boiling in an SDS loading sample buffer for 5 minutes. The pull-down samples were analyzed by immunoblotting with an anti-MBP antibody. Similarly, for the MBP pull-down assay, the MBP-tagged proteins were coupled to amylose resins (E8021V; NEB) and incubated with GST-tagged proteins. The resulting beads were washed and eluted as described above. The relevant pull-down samples were analyzed by immunoblotting using an anti-GST antibody.

In vitro self-ubiquitination and substrate ubiquitination analyses
Recombinant MBP-Chalk9 and its single amino acid substitution mutant (MBP-Chalk9^{C189S}) were expressed in *E. coli* and purified using amylose resins (E8021V; NEB) for in vitro self-ubiquitination analyses. The ubiquitination assay was performed as previously described, with some modifications[57]. In detail, 400 μg of MBP-Chalk9, MBP-Chalk9^{C189S}, or MBP protein was incubated in a 50-μL reaction mixture containing ubiquitination buffer (50 mM Tris-HCl, pH 7.5, 5 mM MgCl$_2$, 2 mM DTT, 4 mM ATP, 15 μg ubiquitin). The reactions were carried out at 30 °C for 2 h in the presence or absence of 50 ng E1 (Beyotime, Shanghai, China) and 100 ng E2 (Beyotime). The reactions were stopped by the addition of 5×SDS loading sample buffer and heated at 95 °C for 5 minutes. The reaction products were separated by SDS-PAGE, followed by immunoblot analysis using anti-MBP (1:5000; HT701; Transgene) and anti-ubiquitin antibodies (1:1,000, RM4934; Biodragon).

For in vitro substrate ubiquitination assays, GST-OsEBP89 was used as the target substrate. 300 ng of the GST-OsEBP89 fusion protein was mixed with an equal amount of MBP-Chalk9 or MBP-Chalk9^{C189S} in the presence or absence of the following: 50 ng of E1, 100 ng of E2, and 5 μg of ubiquitin. The reaction was performed in a total volume of 50 μL containing ubiquitination buffer at 30 °C for 3 h. Ubiquitination levels of proteins were determined by Western blotting using a polyclonal anti-ubiquitin antibody (1:1,000, RM4934; Biodragon) and an anti-GST antibody (1:10,000, CW0084M; CWBIO).

Cell-free degradation assays
The leaf powder, frozen in liquid nitrogen from Nip and *chalk9-1* plants, was suspended in extraction buffer (5 mM MgCl$_2$, 40 mM Tris-HCl, pH 7.5, 5 mM NaCl, 1 mM DTT, and 10 mM ATP) and vigorously vortexed at 4 °C for 1 hour. After centrifugation at 16,000 g at 4 °C for 30 minutes, the supernatant was collected for the cell-free degradation assay. The GST-OsEBP89 recombinant protein was incubated with the supernatant at 30 °C for different periods, in the presence or absence of 50 μM MG132 (Beyotime). The reactions were terminated by adding 5× SDS loading sample buffer, and the samples were immunoblotted using anti-GST (CW0084M; CWBIO) and anti-Actin (CW0264M; CWBIO) antibodies. The protein levels were quantified using ImageJ software (http://rsb.info.nih.gov/ij/).

Yeast one-hybrid assay
Yeast one-hybrid (Y1H) assays were performed using the Matchmaker™ Gold Yeast One-Hybrid System (Clontech). The coding sequence of *OsEBP89* was fused to the activation domain of the GAL4 protein in the pGADT7 vector, generating the prey construct pGADT7-OsEBP89. 2-kb promoter sequences from *GluB1a, GluB2, GluB4, PROLM20, PROLM22,* and *PROLM23* were individually inserted into the pAbAi vector, generating the bait constructs. The pBait-AbAi vectors were transformed into the Y1HGold yeast strain, and the resulting transformants were cultured on SD/-Ura plates to select positive colonies. Subsequently, minimal inhibitory concentration of Aureobasidin A (AbA) was determined by plating the strains on SD/-Ura medium supplemented with AbA at concentrations of 100, 200, 300, and 500 ng/mL. The growth of the bait strains was observed to determine the minimal inhibitory concentration of AbA. Successful pBait-AbAi transformants were streaked onto SD/-Ura plates for selection and subsequently used for competent cell preparation. The prey construct pGADT7-OsEBP89 was then introduced into these competent pBait-AbAi yeast cells. The interaction between the empty pGADT7 and the corresponding bait plasmid served as the negative control. Positive protein-DNA interactions were verified by culturing

yeast cells on SD/-Leu culture media with or without AbA for 3–5 days at 30 °C. Primer sequences are shown in Supplementary Data 15.

Luciferase activity assay in rice protoplasts

To investigate the regulatory effect of the *Chalk9* promoter on gene expression, approximately 2-kb promoter sequences of *Chalk9* were amplified from Nip and Guichao2 and inserted into the pGreenII 0800-LUC vector to generate *proChalk9-L:LUC* and *proChalk9-H:LUC*, respectively. Six variants were generated based on *proChalk9-L:LUC* using a Fast Mutagenesis System (FM111, Transgen Biotech). All vectors were transformed into protoplasts, respectively. Afterwards, these protoplasts were incubated in W5 solution for 12 h at 28 °C. Activities of firefly luciferase (LUC) and Renilla luciferase (REN) were examined using a dual luciferase assay kit (Vazyme Biotech, Jiangsu, China). The primers used for PCR amplification and mutagenesis are listed in Supplementary Data 15.

For analyzing the transcriptional activity of the OsB3 protein on the *Chalk9* alleles, approximately 2-kb promoter sequences of *Chalk9*-L and *Chalk9*-L v5m were individually cloned into the pGreenII 0800-LUC vector to create reporter constructs, respectively. The coding sequence of *OsB3* was cloned into the pGreenII 62-SK vector to generate the effector construct. The empty pGreenII 62-SK vector was used as a negative control. Plasmid combinations were co-transformed into rice protoplasts for transcriptional activity analysis. The transformed cells were incubated in the dark at 28 °C for 12 h and then used to measure transcriptional activity using a dual luciferase assay kit (Vazyme Biotech, Jiangsu, China). The relevant primers are listed in Supplementary Data 15.

Immunoblot analysis

Developing seeds were homogenized in protein extraction buffer (2 mM EDTA, 100 mM NaCl, 20 mM Tris-HCl, pH 7.5, 0.1% [v/v] Triton X-100, 1 mM PMSF, and 1× proteinase inhibitor cocktail). Total proteins were collected after the homogenate was centrifuged at 16,000 × g at 4 °C for 20 min. Protein samples with SDS loading sample buffer were heated to 96 °C for 10 min, separated by 10% (w/v) SDS-PAGE gels, and subsequently transferred to PVDF membranes (Immobilon-P, USA). The PVDF membranes were incubated in TBST buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% Tween-20) with 5% milk powder for at least 60 min, followed by sequential incubation in TBST buffer with 1% milk powder supplemented with primary antibodies and secondary antibodies. Protein signals were detected using the eECL Western Blot Kit (CW0049S; CWBIO) and chemiluminescence detection system.

OsEBP89 polyclonal antibody preparation

To generate a specific antibody against OsEBP89, we chose a truncated sequence (residues 1–120) for recombinant protein production. The corresponding coding sequence was amplified and cloned into the pET28a vector with an N-terminal His-tag. The recombinant protein was expressed in *E. coli* strain BL21 (DE3) transformed with the resulting construct and then purified using a Ni-NTI agarose resin matrix (Qiagen). The purified recombinant protein served as the antigen to raise antibodies in two rabbits, a process conducted by GenScript. The antibody against OsEBP89 was further affinity-purified from serum using immobilized recombinant protein and specifically detect endogenous OsEBP89.

Population genetic and evolutionary analyses

The geographical information and genomic sequences of 1424 cultivated varieties were obtained from the 3 K Rice Genomes Project[38], and marked on map using *Cartopy* package v0.20.0 in Python v3.6.0 software to observe the geographic distribution of the two types of *Chalk9*. Using VCFtools v0.1.16[58], the nucleotide diversity (π) and Neutral test (Tajima's *D*) were calculated in 50-kb windows for each *japonica*, *indica*, and wild rice population. All sites in the *Chalk9*

locus with a minor allele frequency ≥0.01 were used to perform phylogenetic and haplotype network analyses[59]. In brief, the phylogenetic tree for the promoter sequences of *Chalk9* and its homologs was constructed using a maximum likelihood method with IQ-TREE v.2.1.2[60]. The haplotype network was generated using *pegas* package v1.2[61] in R v4.1.2 software and then displayed using the plotting module matplotlib v3.6.0[62] by Python v3.6.0 software.

Spatiotemporal gene expression and TFBS enrichment analysis

The spatio-temporal gene expression pattern was analyzed by RiceXPro[63]. Additionally, the 64-bp sequence of the *Chalk9* promoter was examined for transcription factor binding site (TFBS) enrichment analysis using PlantPan v4.0[64].

Assays of grain filling rate and sucrose content

At noon on the day, the flowering spikelets on the main panicles were marked. Developing grains were sampled at different post-flowering stages to measure the grain filling rate. The marked panicles were excised, placed in preservation bags, and transported to the laboratory. Filled grains from the marked spikelets were collected, oven-dried at 80 °C for 72 h, and then dissected to isolate the endosperm for dry weight measurement. For each line, a minimum of three biological replicates were performed, with 10 endosperms measured per replicate.

Sucrose concentrations in the endosperm were quantified at 5 and 10 DAF using a commercial sucrose assay kit (ZT-1-Y, Suzhou Comin Biotechnology Co., Ltd) following the manufacturer's standardized protocol. For each assay, a minimum of six biological replicates were analyzed to ensure statistical reliability.

Statistical analysis

Prism v.6.0 (GraphPad) software was used for all statistical tests and data visualization. Sample sizes (*n*) and *P*-values are indicated in the individual figures and figure legends. For comparisons between two groups, statistical significance was determined using two-tailed paired Student's *t*-test. For comparisons among more than two groups, statistical significance was determined using one-way analysis of variance (ANOVA) with Tukey's multiple comparisons test.

Accession numbers

Sequence data related to this article can be obtained from the Rice Database (https://www.ricedata.cn/gene) under following accession numbers *LOC_Os09g32730* for *Chalk9*, *LOC_Os03g08460* for *OsEBP89*, *LOC_Os06g04200* for *Wx*, *LOC_Os02g15178* for *GluB1a*, *LOC_Os02g15150* for *GluB2*, *LOC_Os02g16830* for *GluB4*, *LOC_Os02g16820* for *GluB5*, *LOC_Os02g14600* for *GluB7*, *LOC_Os02g25640* for *GluC*, *LOC_Os02g15090* for *GluD*, *LOC_Os05g26350* for *PROLM4*, *LOC_Os05g26460* for *PROLM11*, *LOC_Os05g26368* for *PROLM13*, *LOC_Os05g26720* for *PROLM16*, *LOC_Os07g11910* for *PROLM20*, *LOC_Os07g11920* for *PROLM22*, and *LOC_Os06g31060* for *PROLM23*.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequencing data of the GWAS generated in this study have been deposited in the NCBI SRA database under the BioProject accession PRJNA1278882 and PRJNA1277491. The RNA-seq raw data generated in this study have been deposited in the NCBI SRA database under the BioProject accession PRJNA1276623. The genomic sequences of *Chalk9* in cultivated rice varieties were obtained from the Rice Functional Genomics and Breeding Database (RFGB) (http://www.rmbreeding.cn/). The genomic sequences of *Chalk9* in wild rice accessions were downloaded from OryzaGenome (http://viewer.

shigen.info/oryzagenome2detail/index.xhtml). These genomic sequences of *Chalk9* are provided in Supplementary Data 16. Genomic variations of *OsEBP89* in cultivated rice varieties were obtained from RiceVarMap v2.0 database (http://ricevarmap.ncpgr.cn/) and are provided in Supplementary Data 17. All constructs and transgenic plants generated in this study are available from the corresponding author upon request. Source data are provided with this paper.

References

1. Tian, Z. et al. Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proc. Natl Acad. Sci. USA* **106**, 21760–21765 (2009).
2. Zeng, D. et al. Rational design of high-yield and superior-quality rice. *Nat. Plants* **3**, 17031 (2017).
3. Zhao, D., Zhang, C., Li, Q. & Liu, Q. Genetic control of grain appearance quality in rice. *Biotechnol. Adv.* **60**, 108014 (2022).
4. Fitzgerald, M. A., McCouch, S. R. & Hall, R. D. Not just a grain of rice: the quest for quality. *Trends Plant Sci.* **14**, 133–139 (2009).
5. Zhang, J. et al. Releasing a sugar brake generates sweeter tomato without yield penalty. *Nature* **635**, 647–656 (2024).
6. Sakamoto, T. & Matsuoka, M. Identifying and exploiting grain yield genes in rice. *Curr. Opin. Plant Biol.* **11**, 209–214 (2008).
7. Zhao, D. et al. *GS9* acts as a transcriptional activator to regulate rice grain shape and appearance quality. *Nat. Commun.* **9**, 1240 (2018).
8. Cao, S. et al. Orchestrating seed storage protein and starch accumulation toward overcoming yield–quality trade-off in cereal crops. *J. Integr. Plant Biol.* **66**, 468–483 (2024).
9. She, K. C. et al. A novel factor *FLOURY ENDOSPERM2* is involved in regulation of rice grain size and starch quality. *Plant Cell* **22**, 3280–3294 (2010).
10. Tabassum, R. et al. *FLOURY ENDOSPERM11-2* encodes plastid HSP70-2 involved with the temperature-dependent chalkiness of rice (*Oryza sativa* L.) grains. *Plant J.* **103**, 604–616 (2020).
11. Custodio, M. C. et al. Rice quality: how is it defined by consumers, industry, food scientists, and geneticists?. *Trends Food Sci. Tech.* **92**, 122–137 (2019).
12. Sreenivasulu, N. et al. Designing climate-resilient rice with ideal grain quality suited for high-temperature stress. *J. Exp. Bot.* **66**, 1737–1748 (2015).
13. Misra, G. et al. Genome-wide association coupled gene to gene interaction studies unveil novel epistatic targets among major effect loci impacting rice grain chalkiness. *Plant Biotechnol. J.* **19**, 910–925 (2021).
14. Qiu, X. et al. Genetic dissection of rice appearance quality and cooked rice elongation by genome-wide association study. *Crop J.* **9**, 1470–1480 (2021).
15. Gong, J. et al. Dissecting the genetic basis of grain shape and chalkiness traits in hybrid rice using multiple collaborative populations. *Mol. Plant* **10**, 1353–1356 (2017).
16. Misra, G. et al. Dissecting the genome-wide genetic variants of milling and appearance quality traits in rice. *J. Exp. Bot.* **70**, 5115–5130 (2019).
17. Peng, B. et al. Comparative mapping of chalkiness components in rice using five populations across two environments. *BMC Genet* **15**, 49 (2014).
18. Wan, X. Y. et al. Stability of QTLs for rice grain dimension and endosperm chalkiness characteristics across eight environments. *Theor. Appl. Genet.* **110**, 1334–1346 (2005).
19. Zhao, X. et al. Identification of stable QTLs causing chalk in rice grains in nine environments. *Theor. Appl. Genet.* **129**, 141–153 (2016).
20. Li, Y. et al. *Chalk5* encodes a vacuolar H+-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat. Genet.* **46**, 398–404 (2014).
21. Wu, B. et al. Natural variation in *WHITE-CORE RATE 1* regulates redox homeostasis in rice endosperm to affect grain quality. *Plant Cell* **34**, 1912–1932 (2022).
22. Tu, B. et al. The LCG1-OsBP5/OsEBP89-*Wx* module regulates the grain chalkiness and taste quality in rice. *Plant Biotechnol. J.* **23**, 36–50 (2025).
23. Shi, H. et al. Natural variation of *WBR7* confers rice high yield and quality by modulating sucrose supply in sink organs. *Plant Biotechnol. J.* **22**, 2985–2999 (2024).
24. Zheng, N. & Shabek, N. Ubiquitin ligases: structure, function, and regulation. *Annu. Rev. Biochem.* **86**, 129–157 (2017).
25. Vierstra, R. D. The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends Plant Sci.* **8**, 135–142 (2003).
26. Deshaies, R. J. & Joazeiro, C. A. RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* **78**, 399–434 (2009).
27. Lee, H. K. et al. Drought stress-induced Rma1H1, a RING membrane-anchor E3 ubiquitin ligase homolog, regulates aquaporin levels via ubiquitination in transgenic *Arabidopsis* plants. *Plant Cell* **21**, 622–641 (2009).
28. Wang, J. et al. Tissue-specific ubiquitination by IPA1 INTERACTING PROTEIN1 modulates IPA1 protein levels to regulate plant architecture in rice. *Plant Cell* **29**, 697–707 (2017).
29. Sun, J., Sun, Y., Ahmed, R. I., Ren, A. & Xie, A. M. Research progress on plant RING-finger proteins. *Genes* **10**, 973 (2019).
30. Huang, X. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967 (2010).
31. Worth, C., Gong, S. & Blundell, T. Structural and functional constraints in the evolution of protein families. *Nat. Rev. Mol. Cell Biol.* **10**, 709–720 (2009).
32. Gazzarrini, S. & Song, L. LAFL Factors in Seed Development and Phase Transitions. *Annu. Rev. Plant Biol.* **75**, 459–488 (2024).
33. Yang, H. J. et al. The *OsEBP-89* gene of rice encodes a putative EREBP transcription factor and is temporally expressed in developing endosperm and intercalary meristem. *Plant Mol. Biol.* **50**, 379–391 (2002).
34. Zhu, Y., Cai, X. L., Wang, Z. Y. & Hong, M. M. An interaction between a MYC protein and an EREBP protein is involved in transcriptional regulation of the rice *Wx* gene. *J. Biol. Chem.* **278**, 47803–47811 (2003).
35. Hu, Z. et al. The kinase OsSK41/OsGSK5 negatively regulates amylose content in rice endosperm by affecting the interaction between OsEBP89 and OsBP5. *J. Integr. Plant Biol.* **65**, 1782–1793 (2023).
36. Yan, H. et al. Rice LIKE EARLY STARVATION1 cooperates with FLOURY ENDOSPERM6 to modulate starch biosynthesis and endosperm development. *Plant Cell* **36**, 1892–1912 (2024).
37. Wang, Y. et al. OsRab5a regulates endomembrane organization and storage protein trafficking in rice endosperm cells. *Plant J.* **64**, 812–824 (2010).
38. Alexandrov, N. et al. SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res* **43**, D1023–D1027 (2015).
39. Huang, X. et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).
40. Wang, W. et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**, 43–49 (2018).
41. Zhao, H. et al. RiceVarMap: a comprehensive database of rice genomic variations. *Nucleic Acids Res* **43**, D1018–D1022 (2015).
42. Li, J. et al. Fine mapping of the grain chalkiness quantitative trait locus *qCGP6* reveals the involvement of *Wx* in grain chalkiness formation. *J. Exp. Bot.* **74**, 3544–3559 (2023).
43. Wang, J., Chen, Z., Zhang, Q., Meng, S. & Wei, C. The NAC transcription factors OsNAC20 and OsNAC26 regulate starch and storage protein synthesis. *Plant Physiol.* **184**, 1775–1791 (2020).

44. Yamagata, H., Sugimoto, T., Tanaka, K. & Kasai, Z. Biosynthesis of storage proteins in developing rice seeds. *Plant Physiol.* **70**, 1094–1100 (1982).

45. Yang, J. et al. Changes in activities of three enzymes associated with starch synthesis in rice grains during grain filling. *Acta Agron. Sin.* **27**, 157–164 (2001).

46. Zhong, L. & Cheng, F. Varietal differences in amylose accumulation and activities of major enzymes associated with starch synthesis during grain filling in rice. *Acta Agron. Sin.* **29**, 452–456 (2003).

47. Duan, M. & Sun, S. S. M. Profiling the expression of genes controlling rice grain quality. *Plant Mol. Biol.* **59**, 165–178 (2005).

48. Liu, X. et al. Transcriptome analysis of grain-filling caryopses reveals involvement of multiple regulatory pathways in chalky grain formation in rice. *BMC Genomics* **11**, 730 (2010).

49. Yang, Y. et al. Natural variation of *OsGluA2* is involved in grain protein content regulation in rice. *Nat. Commun.* **10**, 1949 (2019).

50. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).

51. Si, L. et al. *OsSPL13* controls grain size in cultivated rice. *Nat. Genet.* **48**, 447–456 (2016).

52. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).

53. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

54. Dong, S. S. et al. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief. Bioinform.* **22**, bbaa227 (2021).

55. Hu, Z. et al. Autophagy targets Hd1 for vacuolar degradation to regulate rice flowering. *Mol. Plant* **15**, 1137–1156 (2022).

56. Zhang, Y. et al. A highly efficient rice green tissue protoplast system for transient gene expression and studying light/chloroplast-related processes. *Plant Methods* **7**, 30 (2011).

57. Huang, L. et al. The LARGE2-APO1/APO2 regulatory module controls panicle size and grain number in rice. *Plant Cell* **33**, 1212–1228 (2021).

58. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

59. Chen, R. et al. A de novo evolved gene contributes to rice grain shape difference between *indica* and *japonica*. *Nat. Commun.* **14**, 5906 (2023).

60. Minh, B. Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Bio. Evol.* **37**, 1530–1534 (2020).

61. Paradis, E. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420 (2010).

62. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

63. Sato, Y. et al. RiceXPro: a platform for monitoring gene expression in *japonica* rice grown under natural field conditions. *Nucleic Acids Res* **39**, D1141–D1148 (2011).

64. Chow, C. N. et al. PlantPAN 4.0: updated database for identifying conserved non-coding sequences and exploring dynamic transcriptional regulation in plant promoters. *Nucleic Acids Res* **52**, D1569–D1578 (2024).

Acknowledgements

Author contributions

Z.H., H.L., and M.G. (Min Guo) performed experiments and analyzed the data. Z.H. and X.H. collected phenotype data of rice varieties and genetic materials in the field. Y.L. conducted the GWAS analysis. R.C. performed the evolutionary analysis. Y.G., Y.Y., S.S., Y.Z., and M.G. (Minghong Gu) participated in the experiments. Z.H. and C.Y. designed the research and wrote the manuscript. C.Y. supervised the project. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-61683-4.

Correspondence and requests for materials should be addressed to Changjie Yan.

Peer review information *Nature Communications* thanks Pinky Agarwal, who co-reviewed with Aswathi PVYuqing He, and Jianlong Xu for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.