

ORIGINAL ARTICLE

Open Access



# Both semi-dwarf and photoperiod-insensitive traits in rice were important for the Green Revolution

Cheng-Chieh Wu<sup>1,2†</sup>, Yi-Tzu Tseng<sup>1,3†</sup>, Lin-Tzu Huang<sup>1,7</sup>, Chun-Kai Liu<sup>1</sup>, Kai-Cheng Hsu<sup>4</sup>, Hao-Ching Wang<sup>5</sup>, Yue-le Caroline Hsing<sup>1</sup> and Wan-Yi Chiou<sup>6\*</sup> 

## Abstract

**Background** The semi-dwarf high-yielding rice (*Oryza sativa*) variety IR8 was the major modern line that gave rise to the “Green Revolution” in Asia approximately 60 years ago. Although the *sd1* mutation in IR8 responsible for semi-dwarfism is well known, the sequence variation, evolutionary origins, and functional impacts of *SD1* haplotypes remain incompletely characterized.

**Results** By analyzing whole-genome sequences from the 3K Rice Genomes Project, Taiwanese landraces, and Asian wild rice, we classified *SD1* into 12 haplotypes, including two newly identified loss-of-function alleles: a 1,279-bp deletion and an E100/R340 variant. Phylogenetic and nucleotide analyses traced the origins of several haplotypes, including the widely used 383-bp deletion (*DGWW* type), to the wild accession W1718 in southern China, from which they were later introgressed into landraces and subsequently brought to Taiwan. Protein structural modeling further revealed that distinct amino acid substitutions affect GA20ox-2 activity by altering the catalytic pocket or local protein stability. Screening of *SD1* and *Hd1* genotypes showed that most modern *indica* varieties carry loss-of-function alleles of both *sd1* and *hd1*, reflecting strong post–Green Revolution selection for semi-dwarfism and early maturity.

**Conclusion** Our results clarify the evolutionary history and functional consequences of *SD1* variation and highlight that the combined loss-of-function *sd1*–*hd1* genotype was central to the development of high-yielding, lodging-resistant, and photoperiod-insensitive rice varieties. These findings provide a comprehensive framework for understanding key Green Revolution genes and offer guidance for future rice improvement and genome-editing strategies.

**Keywords** Heading date, Indels, Natural variation, Rice, Semi-dwarf, SNPs, Protein structure

<sup>†</sup>Cheng-Chieh Wu and Yi-Tzu Tseng made equal contributions to this article.

\*Correspondence:

Wan-Yi Chiou  
wychiou@nchu.edu.tw

<sup>1</sup>Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

<sup>2</sup>Graduate Institute of Medical Genomics and Proteomics, National Taiwan University, Taipei, Taiwan

<sup>3</sup>Institute of Plant Biology, National Taiwan University, Taipei, Taiwan

<sup>4</sup>Graduate Institute of Cancer Biology and Drug Discovery, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

<sup>5</sup>Graduate Institute of Translational Medicine, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

<sup>6</sup>Department of Agronomy, National Chung-Hsing University, Taichung, Taiwan

<sup>7</sup>Present address: Transgenic Plant Core Facility, Agricultural Biotechnology Research Center, Academia Sinica, Taipei, Taiwan

## Background

The “Green Revolution” of the 1960s led to a great increase in global production of rice (*Oryza sativa*) (Hargrove and Cabanilla 1979; Khush 1999) and wheat (*Triticum aestivum*) (Chapman et al. 2007). The newly developed high-yielding varieties of these two important staple crops fed the rapidly expanding world population. The key trait contributing to this higher productivity in both rice and wheat was the “semi-dwarf” trait, which is associated with the plant hormone gibberellic acid (GA) and reduces culm length without compromising yield. The new semi-dwarf varieties could withstand heavy winds and rainfall. In addition, they were resistant to lodging following treatment with a large amount of nitrogen fertilizer (Khush 1995).

The key Green Revolution variety of rice was IR8, also known as miracle rice; its *semi-dwarf 1* gene (*sd1*) encodes GA<sub>20</sub> oxidase-2 (GA<sub>20</sub>ox-2; Os01t0883800) (Sasaki et al. 2002; Ashikari et al. 2002; Spielmeier et al. 2002). The parental lines of IR8 were Dee-Geo-Woo-Gen (DGWG), a Taiwanese native semi-dwarf *indica* landrace, and Peta, a tall Indonesian *indica* landrace known for its high eating quality. IR8 was bred by the International Rice Research Institute (IRRI) in late 1966 (Hargrove and Cabanilla 1979). Both IR8 and DGWG contain a 383-bp deletion in GA<sub>20</sub>ox-2 (*sd1*), affecting parts of exons 1 and 2, that leads to the loss of function (LOF) of this gene (Sasaki et al. 2002; Ashikari et al. 2002; Spielmeier et al. 2002).

Single-nucleotide polymorphisms (SNPs) also occur in other *sd1* mutants. The Japanese native variety Jikkoku has a Gly (G) 94 to Val (V) substitution (Kikuchi et al. 1985). Two semi-dwarf varieties developed by  $\gamma$ -ray radiation have SNPs in *sd1*: Reimei has an Asp (D) 349 to His (H) substitution from the *japonica* cultivar Fujimino (Futsuhara et al. 1967), and Calrose 76 has a Leu (L) 266 to Phe (F) substitution as compared to the US *japonica* cultivar Calrose (Foster and Rutger 1978). The SNPs in the GA<sub>20</sub>ox-2 gene in these accessions thus represent three independent alleles leading to the semi-dwarf phenotype (Sasaki et al. 2002; Ashikari et al. 2002). A survey of additional semi-dwarf rice accessions revealed further alleles, including a 2-bp deletion causing a frameshift at amino acid (aa) 130 in Aijiao-Nante (*indica*); an SNP in Zhayeqing 8 (*indica*) resulting in a substitution of Pro (P) 240 to Leu (L); and another SNP in 9311 (*indica*), replacing Tyr (Y) 342 with a stop codon and causing early translational termination (Tyr342\*) (Asano et al. 2007). More recently, a new *sd1* allele in Pusa 1652, a  $\gamma$ -ray-induced mutant of the short-grain aromatic landrace Bindli in India, was reported to introduce a premature stop codon (Tyr300\*) (Bhuvaneswari et al. 2020).

The *SD1* gene has been revealed to be a target of selection during early rice domestication. Quantitative trait

locus (QTL) mapping was conducted to identify the genetic basis of culm length using backcross inbred lines derived from a cross between the *japonica* rice variety Nipponbare and the *indica* variety Kasalath (Asano et al. 2011). High-resolution mapping with approximately 5,000 plants segregating at the *qCL1* locus indicated that the target gene was *SD1*, and that the functional nucleotide polymorphisms corresponded to amino acid substitutions: Glu (E) and Gln (Q) at residues 100 and 340 in Nipponbare, but G and Arg (R), respectively, at the same residues in Kasalath. The EQ variant of the double polymorphism exhibits lower enzymatic function. Studies focusing on landraces indicated that the GR haplotype is present in most *indica* and wild rice accessions, whereas EQ is predominant in *japonica* rice (Asano et al. 2011). A recent report indicated that the GR-EQ SNPs have existed in ancient rice populations since before domestication (Alam et al. 2025).

An analysis of the genealogy of the rice Green Revolution gene, including the DGWG and Jikkoku alleles (Nagano et al. 2005), suggested that the DGWG and Jikkoku alleles derive from different lineages. By examining the 383-bp deletion, the authors determined that the DGWG allele was present in several lines, including the IRRI *indica* varieties IR8 and IR36, the Taiwanese landrace Liu-T’ou-Tu, and two wild rice (*O. rufipogon*) accessions collected in southern China: W1718 and W1944. Because these wild rice accessions have unique polymorphisms in the region surrounding the *SD1* locus, the authors suggested that the DGWG allele might have originally been present in the wild progenitors rather than being introgressed from high-yielding varieties into the nearby wild rice (Nagano et al. 2005).

GA<sub>20</sub>ox-2 catalyzes the late steps of GA biosynthesis (Ashikari et al. 2002; Spielmeier et al. 2002; Pimenta Lange and Lange 2006). This enzyme is responsible for the early 13-hydroxylation pathway of GA biosynthesis in vascular plants (Toyomasu et al. 1997; Spielmeier et al. 2002; Pimenta Lange and Lange 2006), which starts with the transformation of GA<sub>53</sub> to GA<sub>44</sub>, GA<sub>19</sub>, and GA<sub>20</sub>, ultimately yielding the bioactive form GA<sub>1</sub>. That is, SD1 is involved in the reaction of GA<sub>12</sub> and GA<sub>53</sub> to generate the bioactive GA precursors GA<sub>9</sub> and GA<sub>20</sub>. The contents of individual GA compounds in equivalent elongating stem segments vary between tall and semi-dwarf rice plants, including Kyeema (tall, *SD1*) versus Doongara (dwarf, *sd1*) and Calrose (tall, *SD1*) versus Calrose76 (dwarf, *sd1*) (Spielmeier et al. 2002). In particular, the content of GA<sub>53</sub>, the initial substrate of GA<sub>20</sub>-oxidase activity, is significantly higher in the semi-dwarf rice varieties (Doongara and Calrose76) than in their tall counterparts. Additionally, the contents of GA<sub>20</sub>, the major product of the pathway, and bioactive GA<sub>1</sub> are lower in

**Table 1** Haplotypes of *SD1*

| Haplotype | Allele name         | Amino acid changes        | Position(s) on chromosome 1 | Number of accessions in the 3K dataset <sup>a</sup> | References                                 |
|-----------|---------------------|---------------------------|-----------------------------|-----------------------------------------------------|--------------------------------------------|
| 1         | <i>GR</i>           | Gly100 and Arg340         | 38,382,764 and 38,385,057   | 1235                                                | Asano et al. 2011                          |
| 2         | <i>EQ</i>           | Glu100 and Gln340         | 38,382,764 and 38,385,057   | 773                                                 | Asano et al. 2011                          |
| 3         | <i>Jikkoku</i>      | Gly94Val                  | 38,382,746                  | 4                                                   | Sasaki et al. 2002                         |
| 4         | <i>Calrose76</i>    | Leu266Phe                 | 38,383,363                  | 7                                                   | Sasaki et al. 2002; Spielmeier et al. 2002 |
| 5         | <i>Reimei</i>       | Asp349His                 | 38,385,083                  | 9                                                   | Sasaki et al. 2002                         |
| 6         | New allele          | 1279-bp deletion          | 38,381,879–38,383,157       | 10                                                  | Current study                              |
| 7         | <i>DGWG</i>         | 383-bp deletion           | 38,382,762–38,383,144       | 450                                                 | Sasaki et al. 2002                         |
| 8         | <i>9311</i>         | Tyr342* (early stop)      | 38,385,064                  | 21                                                  | Asano et al. 2007                          |
| 9         | <i>Ajjiao-Nante</i> | Arg130fs (2-bp insertion) | 38,382,846–38,382,847       | 15                                                  | Asano et al. 2007                          |
| 10        | <i>Zhayeqing 8</i>  | Pro240Leu                 | 38,383,286                  | 50                                                  | Asano et al. 2007                          |
| 11        | <i>ER</i>           | Glu100 and Arg340         | 38,382,764 and 38,385,057   | 65                                                  | Current study                              |
| 12        | <i>bm</i>           | Tyr300* (early stop)      | 38,384,938                  | N/A                                                 | Bhuvaneswari et al. 2020                   |
| Total     |                     |                           |                             | 2639                                                |                                            |

<sup>a</sup>Includes only accessions with high sequence quality (reads with quality  $\geq$  Q20). Accessions carrying only type 1 or type 2 mutations were identified first; type 3–12 mutations were classified separately because these accessions may share the same amino acid substitutions at positions 100 or 340 as type 1 or 2 mutations. \*, termination codon. fs, frame shift.

the semi-dwarf varieties than in the other varieties (Spielmeier et al. 2002).

In the IR8 breeding history reports, the IRRI provided a blueprint for the variety in the 1961–62 Annual Report: “a combination of short, stiff culms bearing erect, moderately sized, dark-green leaves; responsiveness in yield to fertilizer; mid-season maturity and in most cases, photoperiod sensitivity to permit double cropping practices.” And, in 1963, they performed selection of  $F_2$  plants from the Peta–DGWG cross in which “only short, early maturing plants were saved” (Khush et al. 2001; Hargrove and Coffman 2006). A few key genes are thought to affect rice sensitivity to photoperiod, such as *Heading date 1* (*Hd1*), which encodes a B-box zinc finger protein and is the ortholog of Arabidopsis (*Arabidopsis thaliana*) *CONSTANS* (*CO*) (Yano et al. 1997, 2000); *Hd3a*, which encodes a phosphatidylethanolamine binding protein and is the ortholog of Arabidopsis *FLOWERING LOCUS T* (*FT*) (Tamaki et al. 2007); and *Grain number, plant height, and heading date 7* (*Ghd7*), an important regulator of heading date and yield potential (Xue et al. 2008). More than 10 loss-of-function alleles of *hd1* have been reported (Takahashi et al. 2009; Yano et al. 2000; Doi et al. 2004; Wu et al. 2020).

In the present study, we explored the mutation types of *SD1* and their evolution in rice by analyzing changes in the *SD1* sequence using the 3000 (3K) Rice Genome sequence database. Together with the 10 haplotypes discovered previously, we summarized 12 haplotypes, some of which have been used extensively in modern breeding. We studied the possible origins of some of these mutations. Using information about amino acid changes

**Table 2** Percentages of *GR*, *EQ*, and *ER* at aa positions 100 and 340 in the 3K panel<sup>a</sup>

|           | <i>indica</i> | <i>aus</i> | <i>admix</i> | <i>japonica</i> | <i>aromatic</i> |
|-----------|---------------|------------|--------------|-----------------|-----------------|
| G100/R340 | 80.7          | 73.8       | 61.9         | 0               | 17.9            |
| E100/Q340 | 0.8           | 1.3        | 19.0         | 89.0            | 82.1            |
| E100/R340 | 5.9           | 16.3       | 4.8          | 0.7             | 3.4             |

<sup>a</sup>Only traditional landraces in the 3K panel were used for analysis, and only homologous alleles were counted

associated with the semi-dwarf trait, we also performed structural predictions to evaluate the effects of these changes on GA<sub>20</sub>-oxidase activity. We further suggest that the photoperiod-insensitive trait from DGWG also played an important role in the Green Revolution. Our findings shed light on the origins of the semi-dwarf and photoperiod-insensitive traits of rice that emerged during the Green Revolution.

## Results

### Many *sd1* haplotypes are present in rice accessions

In the current study, we discovered previously unknown null *sd1* alleles from the 3K Rice Genome sequencing data (3K-RGP 2014), a 1,279-bp deletion variant lacking the promoter, exon 1, and intron 1, and an E100/R340 (*ER*) variant in addition to the previously described *GR* and *EQ*. Based on this information and the previously discovered haplotypes, we classified the *SD1* gene into 12 haplotypes (Table 1 and S1). Using the 3K dataset and the wild rice pan-genome data (Zhao et al. 2018), we confirmed that *GR* is present in most *indica* and all wild rice accessions, while *EQ* is present in most *japonica* accessions (Table 2). Thus, we defined *GR* as the type 1 (wild

type) *SD1* haplotype (Fig. 1a). There were no *SD1*-GQ-type alleles (another combination of G/E and R/Q) in the natural population.

The remaining *sd1* haplotypes were categorized as SNPs, short indels, or long deletions. As compared with the wild type, types 2, 3, 4, 5, 8, 10, 11, and 12 had SNPs causing non-synonymous aa residue changes or early stop codons (i.e., E100/Q340, G94V, L266F, D349H, Y342\*, P240L, E100/R340, and Y300\*, respectively). Type 9 contained a 2-bp indel that led to a frameshift at aa position 130. The 383-bp deletion of the type 7 *sd1* haplotype abolished parts of exons 1 and 2 as well as intron 1. The type 6 haplotype lacked the promoter region, exon 1, and intron 1 due to a 1,279-bp deletion (Fig. 1b). Integrative Genomics Viewer (IGV) snapshots of the 383- and 1,279-bp deletions in *SD1* showed that their 3' ends were only 13 bp apart—at positions 38,383,145 and 38,383,158 of chromosome 1, respectively (Fig. S1)—although their 5' ends were separated by ~900 bp.

To evaluate the phenotypic impact of these diverse *sd1* haplotypes, we analyzed plant height data from the IRRI SNP-SEEK phenotype database (<https://snp-seek.irri.org/>) (Mansueto et al. 2017; Alexandrov et al. 2015). This database provides plant height information for rice cultivated at the IRRI, alongside extensive phenotypic data for approximately two-thirds of the 3K accessions. Data on culm length at heading were retrieved from the database and grouped by *SD1* haplotype. Accessions excluding types 3 through 10 were categorized as *GR*, *EQ*, or *ER*. Analysis of variance (ANOVA) revealed significant variation in plant height among the *sd1* haplotypes (Table S2). Multiple comparisons demonstrated that plants carrying types 2–10 differed significantly in height from the wild type (type 1) (Fig. 2 and Table S3). The significant height reduction in plants with type 2, which lacked other structural mutations, confirmed that the *GR*-to-*EQ* substitution alone represents a mutation with reduced function, exerting a clear effect on plant height at heading. In contrast, type 11 (carrying the *ER* allele) exhibited the same

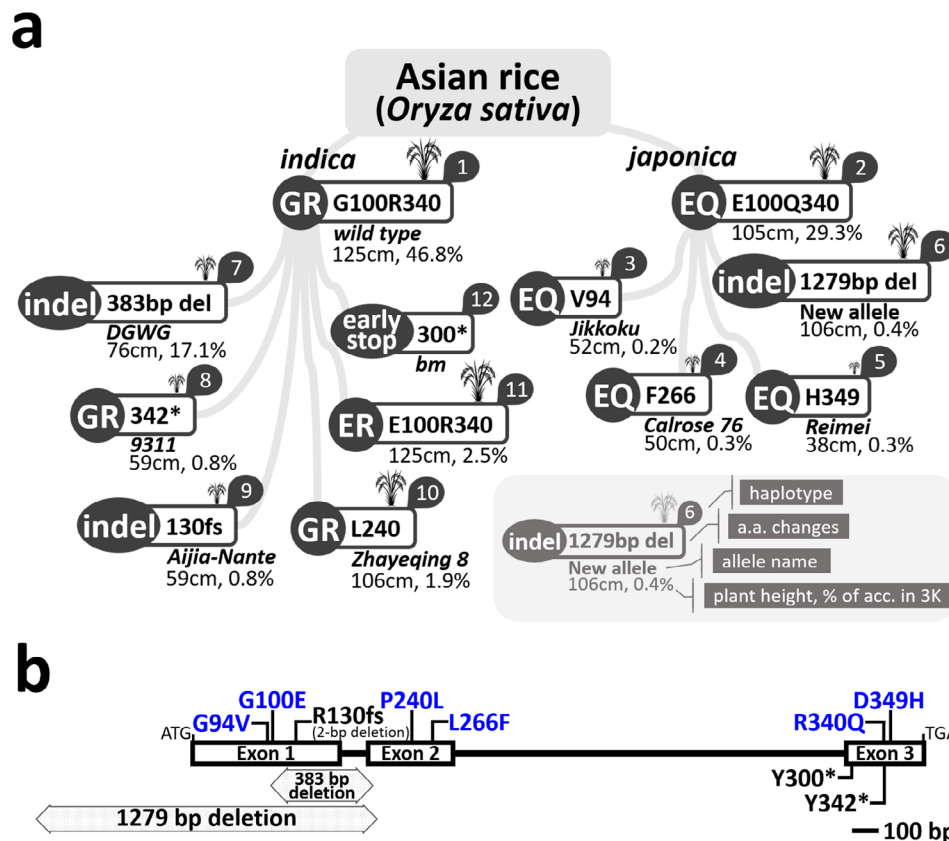


**Fig. 1** Overview of *SD1* haplotypes and mutation sites in rice. **(a)** Schematic representation of the 12 *SD1* haplotypes identified in this study. Haplotypes are grouped according to subspecies (*indica* and *japonica*). Each type is labeled with the corresponding amino acid change or deletion, allele name, average plant height, and frequency in the 3K Rice Genome dataset. GR represents the wild type allele (G100/R340), whereas EQ and ER denote alternative combinations at residues 100 and 340. Indel mutations and premature stop codons are indicated. **(b)** Gene structure of *SD1* (*GA20ox-2*) and positions of mutation sites. The *SD1* gene contains three exons and two introns. The deleted sequences in the type 6 and 7 *SD1* haplotypes are represented by horizontal gray bars. The amino acid changes caused by single-nucleotide polymorphisms or short deletions of type 2–5 and 8–10 *SD1* haplotypes are shown. Amino acid substitutions are in blue. Bar, 100 bp

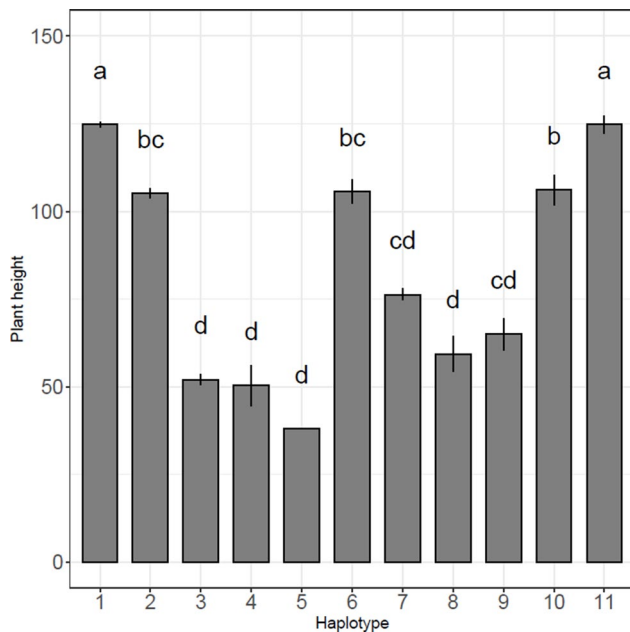


**Fig. 2** Plant height of the 11 *SD1* haplotypes. Plants were grown at the International Rice Research Institute, and plant height (culm length) was measured during the heading period. The phenotype information was downloaded from the SNP-SEEK website (<https://snp-seek.irri.org/>). Sample number and detailed statistical data are shown in Table 1, Table S2, and Table S3. Type 1 is the wild type (*GR* type). Data are mean  $\pm$  SD. Different letters above the bars indicate significant differences at  $P < 0.05$

plant height as the wild type, indicating that this mutation did not alter plant stature.

To explore the presence of *GR* and *EQ* and even the intermediate allele *ER* in natural rice populations, we investigated the variants of the *SD1* gene present in the 3K panel. Only homozygous SNPs were counted. No *japonica* accessions and six *aromatic* accessions featured the *GR* type; approximately 40 *indica* and *aus* accessions featured the *EQ* type. Details about the total number of accessions and traditional landraces with these alleles are shown in Table S4. Notably, some rice accessions harbored the *SD1-ER* allele (type 11): specifically, 6 *japonica*, 2 *aromatic*, 40 *aus*, and 154 *indica* accessions. These alleles were quite rare in landraces: 2 were found in *japonica*, 1 in *aromatic*, 13 in *aus*, and 38 in *indica* rice.

Accessions with type 3–5 mutations were the shortest plants among all groups. These were semi-dwarf, less short than previously reported dwarf mutants (Futsuhara et al. 1967; Foster and Rutger 1978; Kikuchi et al. 1985). However, type 3–5 plants grown in the IRRI paddy field were considerably shorter than other *sd1* mutants (types 6–10). There were likely several factors contributing to these height differences. Plant height was measured at 80% heading. Since photoperiod sensitivity determines the number of days to heading, the temperate *japonica* lines grown in tropical regions headed significantly earlier (~2 months after planting) than most other *sd1*

mutant types (at least 3 months), which directly led to a shorter final plant height (Fig. 2 and Table S3).

Beyond these physiological factors, it is important to note that potential inconsistencies in seed purity within germplasm collections could complicate further analysis. For instance, among the two Aijiao-Nante accessions sequenced in the 3K dataset, only IRIS\_313-12273 had the type 9 (2-bp) deletion. The other accession (CAAS-B060) contained the 383-bp deletion (type 7). Similar discrepancies were observed in Taichung Native 1 (TN1). As the first modern semi-dwarf variety derived from a cross between DGWG and Tsai-Yuan-Chung, TN1 is expected to inherit the type 7 null allele from DGWG. However, of the two TN1 accessions in the 3K dataset, only CX162 carried the 383-bp deletion, while CX270 did not. Furthermore, although the famous Japanese variety Koshihikari is known to lack the 383-bp deletion (Tomita 2009), the 3K accession CAAS\_CX330 was unexpectedly grouped into *sd1* type 7.

### The origins of the *SD1* haplotypes

Our above results indicated that four *SD1* haplotypes were specific to *japonica* rice: types 3, 4, 5, and 6. Types 3 and 5 corresponded to *Jikkoku* and *Reimei* from Japan, respectively. Type 4 was *Calrose* 76 from the United States. Ten *japonica* accessions carried *sd1* type 6, which harbored a 1,279-bp deletion. Plotting the geographic origins of type 6 accessions using GPS information (Gutaker et al. 2020) showed that they were all collected in mountainous regions of Laos and Thailand (Fig. S2). To determine whether haplotypes with the longer deletion (type 7), we performed phylogenetic analysis of the upstream and downstream 1-Mb regions of *SD1*. Types 6 and 7 alleles were in different clusters (with the former including only *japonica* accessions and the latter primarily *indica* accessions; Fig. S3; accessions are listed in Table S5). In addition, a SNP (C to T at position 38,383,221) was consistently associated with accessions containing the 383-bp deletion but not the 1279-bp deletion. Hence, we concluded that these two alleles were derived from different lineages. Phylogenetic analysis of the  $\pm 1$ -kb region of *SD1* indicated that these type 6 lines shared an extremely close relationship (Fig. S4, accessions listed in Table S6), which suggests that the mutation occurred recently. Accessions with this allele were tropical *japonica* accessions collected from part of the Indochina peninsula.

There were four types of *sd1* mutations in *indica* rice. The 9311 mutation (type 8) occurred *in agro* in China. *Aijio-Nante* (type 9) harbored a mutation that also occurred *in agro* in the cultivar Nante in China; notably, the name Aijio is translated as low foot (semi-dwarf). Table S1 shows that 50 accessions contained the

*Zhayeqing 8* *sd1* mutation (type 10); 21 were traditional landraces. Given that more than 95% of the traditional landraces were collected from Indonesia and Malaysia, and that phylogenetic analysis of the  $\pm 1$ -kb region of *SD1* indicated that the Indonesian lines were closely related to wild rice accessions (Fig. S5, accessions listed in Table S7), we conclude that type 10 arose in the insular region of Southeast Asia.

DGWW, which has the first null *sd1* allele (type 7) identified, was a Taiwanese *indica* landrace. A previous study analyzing polymorphisms of three molecular markers (KS3, RM8278, and RM1387), located approximately 6–7 cM from the *SD1* locus, suggested that this mutation first arose in wild rice in southern China and was later introgressed into local landraces or varieties (Nagano et al. 2005). Nagano et al. (2005) also noted that the Taiwanese landrace Liu-Tou-Tzu carries DGWW allele. Thus, we searched for accessions with the DGWW allele in the 3K panel and found that another Taiwanese landrace, Hsinchu-Ai-Chueh-Chien, had the same deletion. Through detailed data mining of information about landrace collections in the National Plant Genetic Resources Center (NPGRC, <https://www.npgrc.tari.gov.tw>), Taiwan

Agricultural Research Institute (TARI), we discovered at least three additional semi-dwarf lines, including Ti-Chueh-Wu-K'o, Ai-Tzu-Ch'ung, and Liu-Tou-Tzu (Table S8; Wu et al. 2022). According to historical records, these Taiwanese landraces were introduced from Guangdong during the migration of Han people to Taiwan in the Ming and Qing dynasties, primarily between the 17th and 19th centuries (DAFTPG 1989). We confirmed that Liu-Tou-Tzu, described in Nagano et al. (2005), was identical to Liu-Tou-Tzu in the NPGRC collection but had its name misspelled. In that study, the two wild rice accessions carrying the DGWW mutation were determined to be W1718 and W1944, both from Guangdong, southern China. However, based on downloaded sequencing information (Zhao et al. 2018), only W1718, but not W1944, carried the 383-bp deletion. This may be another example of seed mixing in the stock center.

To reveal the origin of the 383-bp deletion and its presence in wild rice accessions, we chose the five Taiwanese landraces with the type 7 *sd1* mutation, along with other landraces, modern varieties, and wild rice accessions, for further analysis. Figure 3 shows the SNPs from position –1073 to +1199 bp around the *SD1* gene in 35

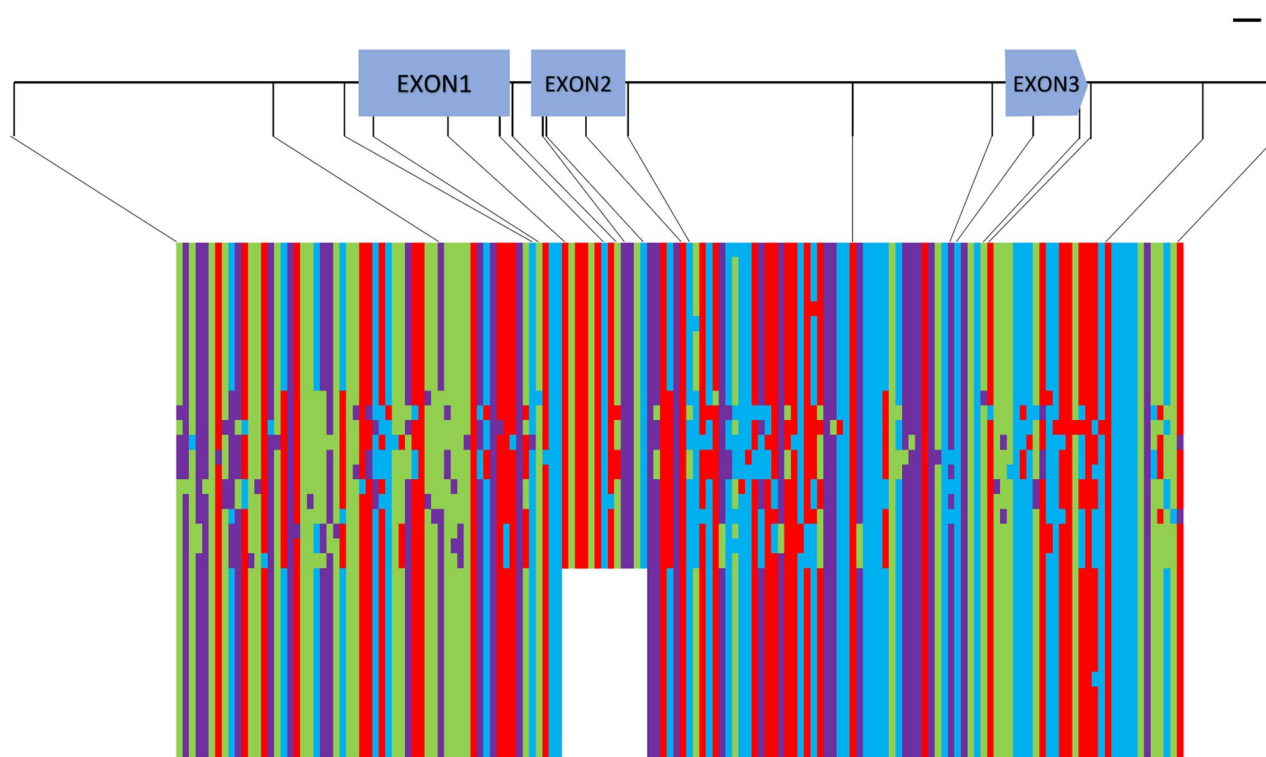


**Fig. 3** Haplotype display of *sd1* type 7. The single-nucleotide polymorphisms (SNPs) from position –1,000 to position +1,000 bp of the gene were plotted. The blue bars at the top indicate three exons. Purple, A; blue, T; green, G; red, C. Accessions are as follows (listed from top to bottom): 10 Taiwanese *indica* landraces: Taichung Woo Gen 2, Chiayi Wu-K'o, Taitung Woo-Li, Ti-Chueh-Hua-Lou, Ti-Chueh-Liu-Chou, Ti-Chueh-Ko-Tzu, Hua-Lou, Wu-Chan, Wu-K'o, Ai-Chueh-Ching-Yu; 13 wild rice accessions: W0123, W1739, W1754, W1777, W2012, W0141, W0170, W1698, W1979, W1943, W3078, W3095, W1718; and 12 accessions with the 383-bp deletion: Taichung Native 1, Hsinchu-Ai-Chueh-Chien, Dee-Geo-Woo-Gen, Ti-Chueh-Wu-K'o, Ai-Tzu-Chung, Liu-Tou-Tzu, Taichung Sen Yu 214, Wagwag Puti, IR64, IR 5494, X23, NON HAI. Information about the collection regions and DNA accession numbers is provided in Table S4

accessions, including landraces or modern varieties with or without the type 7 mutation, as well as 13 wild rice accessions (samples listed in Table S9). The SNP patterns of the 12 landraces and modern varieties with the 383-bp deletion, along with the wild rice accession W1718, are basically identical in the entire 2.7-kb region (including the promoter, three exons, and two introns, as well as the 5' and 3' untranslated regions). In addition, in the region excluding the 383-bp deletion fragment, the 10 Taiwanese *indica* landraces without the type 7 mutation, such as Wu-Chien, Wu-K'o, and Wu-Li, share a pattern similar to that of the landraces and W1718, which contain the 383-bp deletion. However, 12 other wild rice accessions have diverse SNP patterns in this 5-kb region. This suggests that W1718 must share a very close relationship with landraces and varieties with the 383-bp deletion.

To explore the lineage of the type 7 *sd1* mutation, we performed detailed phylogenetic analyses (Fig. S6; Table S10) of the  $\pm 100$ -kb region (Fig. S6A),  $\pm 9$ -Mb region (Fig. S6B), and  $\pm 10$ -Mb region (Fig. S6C) around the *SD1* gene as well as the whole genome (Fig. S6D). The wild rice W1718 (highlighted in purple) was tightly clustered with accessions with the type 7 mutation (highlighted in red) in the  $\pm 100$ -kb region (Fig. S6A); it remained in the clade with Wu-Chien, Wu-K'o, Wu-Li, and accessions with the type 7 mutation in the  $\pm 9$ -Mb region (Fig. S6B).

In the phylogenetic tree for the  $\pm 10$ -Mb region (Fig. S6C), however, this wild rice was separate from the clade containing cultivars, and it was placed in the wild rice clade in the phylogenetic tree based on whole-genome analysis (Fig. S6D).

To further confirm the lineage, we examined the localizations of the 8- to 10-bp tandem repeats (TRs) around the *SD1* gene applying a previously developed software package that quickly scans TRs and plots them along chromosomes (Wei et al. 2016). The genes of eukaryotes, including rice (IRGSP 2005), contain highly divergent TRs (Gemayel et al. 2010; Zhao et al. 2014). TRs occupy approximately 5% of the Nipponbare rice genome (Korotkov et al. 2021). For instance, there are 6083, 4925, and 4886 copies of 8-, 9- and 10-bp TRs, respectively, in the Nipponbare RefSeq data. In the  $\pm 1$ -Mb *SD1* gene region, we assessed the patterns of TRs that were present only in the current accession but not in the Nipponbare genome, and TRs that were not present in the current accession (but were present in the Nipponbare genome). The patterns for the five Taiwanese landraces (Ai-Tzu-Ch'ung, DGWG, Ti-Chueh-Wu-K'o, Hsinchu-Ai-Chueh-Chien, Liu-T'ou-Tzu) and wild rice W1718 were very similar, whereas each of the four other wild rice accessions from southern China had a unique TR pattern (Fig. 4). Along with identical SNPs in the  $\pm 1$ -kb region

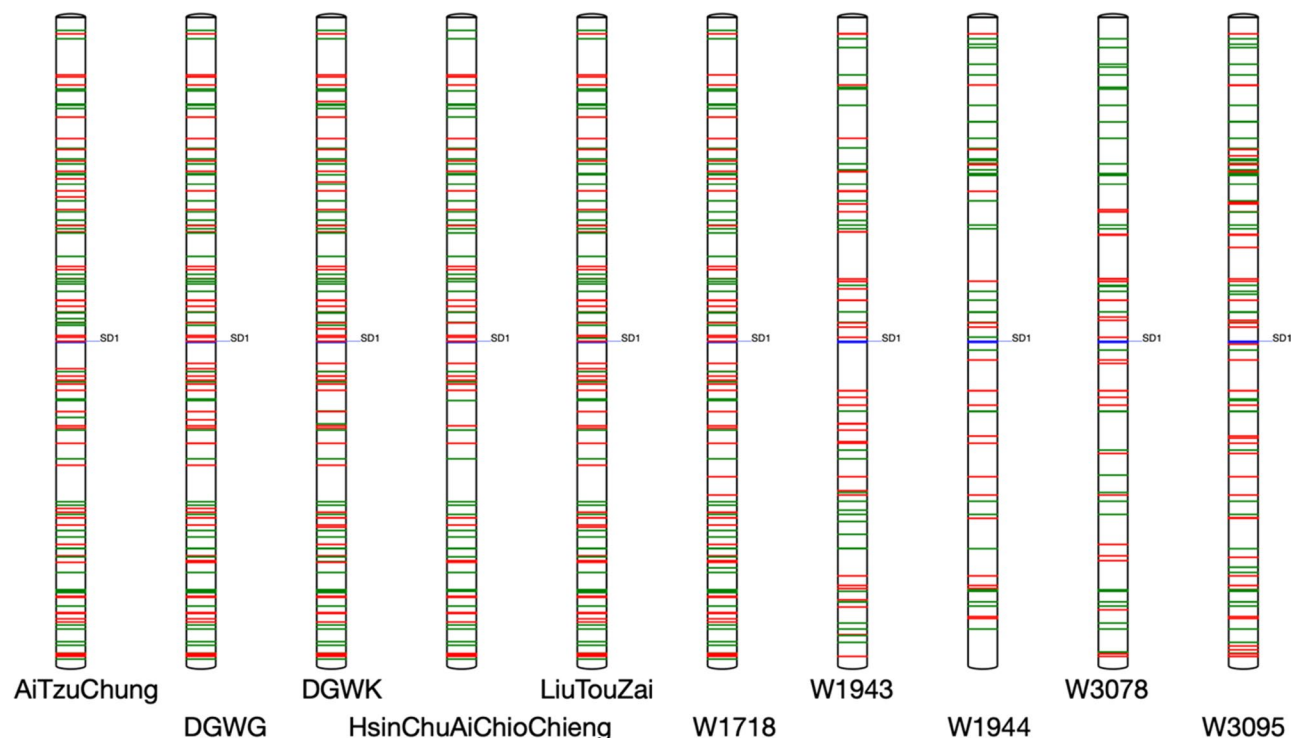


**Fig. 4** Distribution of tandem repeats (TRs) in the *SD1* gene region. The  $\pm 1$ -Mb regions of *SD1* genes in five Taiwanese landraces with type 7 mutations and five wild rice accessions collected in southern China are shown. The accession names are indicated below. The positions of *SD1* genes are marked. Red bands indicate TRs found only in the current accession compared with the Nipponbare genome, and green bands indicate TRs not in the current accession (but detected in the Nipponbare genome). TRs ranging from 8 to 10 bp with 100% identity were included in this plot. Bar, 100 kb

(Fig. 3), high sequence similarity around the end of the long arm of chromosome 1 was present in this specific wild rice accession a long time ago in Guangdong, southern China. These results suggest that the 383-bp deletion originated in the wild rice accession W1718 in southern China and was subsequently introgressed into neighboring landraces.

Our data suggest that the type 7 *sd1* allele occurred *in agro* at least 500 years ago in southern China, before the mass migration of the Han people to Taiwan (DAFTPG 1989). Consequently, when thousands of landraces were introduced to Taiwan during this migration, some already carried the type 7 mutation. In contrast, semi-dwarf landraces carrying the 383-bp deletion were not preserved in southern China, whereas wild rice accessions such as W1718, which were not subject to human selection, continued to propagate and retained this mutation. We explored which type 7 accession(s) were brought to Taiwan by analyzing the nucleotide variation in the  $\pm 200$ -kb region surrounding *SD1* in five type-7 Taiwanese landraces and five wild rice accessions collected from Guangdong, China. Phylogenetic analysis placed Hsinchu-Ai-Chueh-Chien closest to wild rice (Fig. S7, accessions listed in Table S11). A Venn diagram of SNPs across the whole genomes of the five landraces revealed that DGWG and Hsinchu-Ai-Chueh-Chien possessed numerous unique SNPs ( $> 100,000$ ), whereas the other three landraces had fewer unique SNPs ( $< 100,000$ ) (Fig. S8). These results suggest that DGWG and Hsinchu-Ai-Chueh-Chien were the accessions most likely brought to Taiwan by Han migrants.

#### A selective sweep occurred only for the E100/Q340 *sd1* mutation

To test early selection for the *EQ* allele, we analyzed traditional Asian *japonica* landraces that were homozygous for type 2 alleles. In total, 293 accessions were included, representing primarily temperate, subtropical, and tropical *japonica* from Asia (Table S12). Type 4 and type 5 mutations, which were induced by  $\gamma$ -ray treatment, were excluded from the analysis, as were type 8 and type 10 mutations that arose after World War II. Ten accessions carrying the type 6 allele originated from Laos and Thailand, whereas most accessions with the type 10 allele were from Indonesia and Malaysia. A relatively high number of landraces with types 6, 7, and 10 alleles were available and included in the analysis.

We assessed several parameters, including nucleotide diversity ( $\pi$ ; Tajima 1983), Watterson's estimator ( $\theta_w$ ; Watterson 1975), and Tajima's *D* (Tajima 1989), in the  $\pm 1$ -kb and  $\pm 10$ -kb regions. Among the four allele types examined, only type 2 (the *EQ* allele) showed a significantly negative Tajima's *D* ( $P < 0.001$ ) (Table S13), suggesting an excess of low-frequency polymorphisms in

this region. In contrast, Tajima's *D* values for the other allele types were not significantly different from zero in the landrace population.

#### Alignment and analysis of missense mutations reveal the important domain of *GA*<sub>20</sub> oxidase

The previous results showed that different alleles affected plant height. We wanted to pinpoint the location of the mutation sites in the *GA*<sub>20</sub>ox-2 protein structure to understand their impact on protein function. Four *SD1* gene homologs are present in the rice genome: Os01g0883800 (*SD1*), Os05g0421900, Os03g0856700, and Os07g0169700. We generated a protein sequence alignment of the four transcripts, including the positions of GR/EQ as well as each of the missense mutations leading to the semi-dwarf phenotype (Fig. S9). We used the *SD1* protein sequence from Kasalath instead of Nipponbare because we considered *indica* rice to be the wild type in this context. The locations of the type 3, 4, 5, and 10 LOF mutations were highly conserved. In addition, R340 of the low-functionality mutation (type 2) was conserved, but G100 was not: Os01t0883800 and Os05t0421900 had G at this position, whereas two other homologs had Asn (N). The aa changes in these positions, including those denoted 1, 2, 3, 4, and a (Fig. S9), all led to reduced plant height during heading.

We used a structural approach to explore the effects of *GA*<sub>20</sub>ox-2 mutations. Because no crystal structure of *GA*<sub>20</sub>ox-2 has been reported, its enzymatic function was inferred from pea (*Pisum sativum*) *GA*<sub>20</sub>-oxidase, which converts *GA*<sub>19</sub> and 2-oxoglutarate into *GA*<sub>20</sub> and succinate (García-Martínez et al. 1997). Molecular docking was performed to generate a *GA*<sub>20</sub>ox-2/*GA* binding model. The validity of the resulting models was assessed by comparison with relevant reference structures, including *GA*<sub>2</sub>ox3 (PDB ID: 6KU3; Takehara et al. 2020) and thebaine 6-O-demethylase (PDB ID: 5O7Y; Kluza et al. 2018). Based on these evaluations, a *GA*<sub>20</sub>ox-2/succinate/*GA*<sub>20</sub> docking model was selected for further analysis. In addition, a *GA*<sub>20</sub>ox-2/*GA*<sub>20</sub> binding model was generated using the *GA*<sub>2</sub>ox3/*GA*<sub>4</sub> complex (PDB ID: 6KU3) as a template. *GA*<sub>20</sub>ox-2 and *GA*<sub>2</sub>ox3 share 42% sequence similarity, and both oxidases function in *GA* biosynthesis in the cytoplasm. Given this similarity in both sequence and biological function, the *GA*<sub>20</sub>ox-2 structure is likely to closely resemble that of *GA*<sub>2</sub>ox3.

Referring to the *GA*<sub>2</sub>ox3 structure (Takehara et al. 2020), we generated a *GA*<sub>20</sub>ox-2 monomer model using the MODELLER module, in which all *GA*<sub>20</sub>ox-2 mutation sites were labeled (Fig. 5A). This analysis indicated that three of these sites may directly affect oxidase activity. The L266F mutation (type 4) occurred directly in the active site (Fig. 5B). Because phenylalanine has a larger side chain than leucine, this substitution may disturb

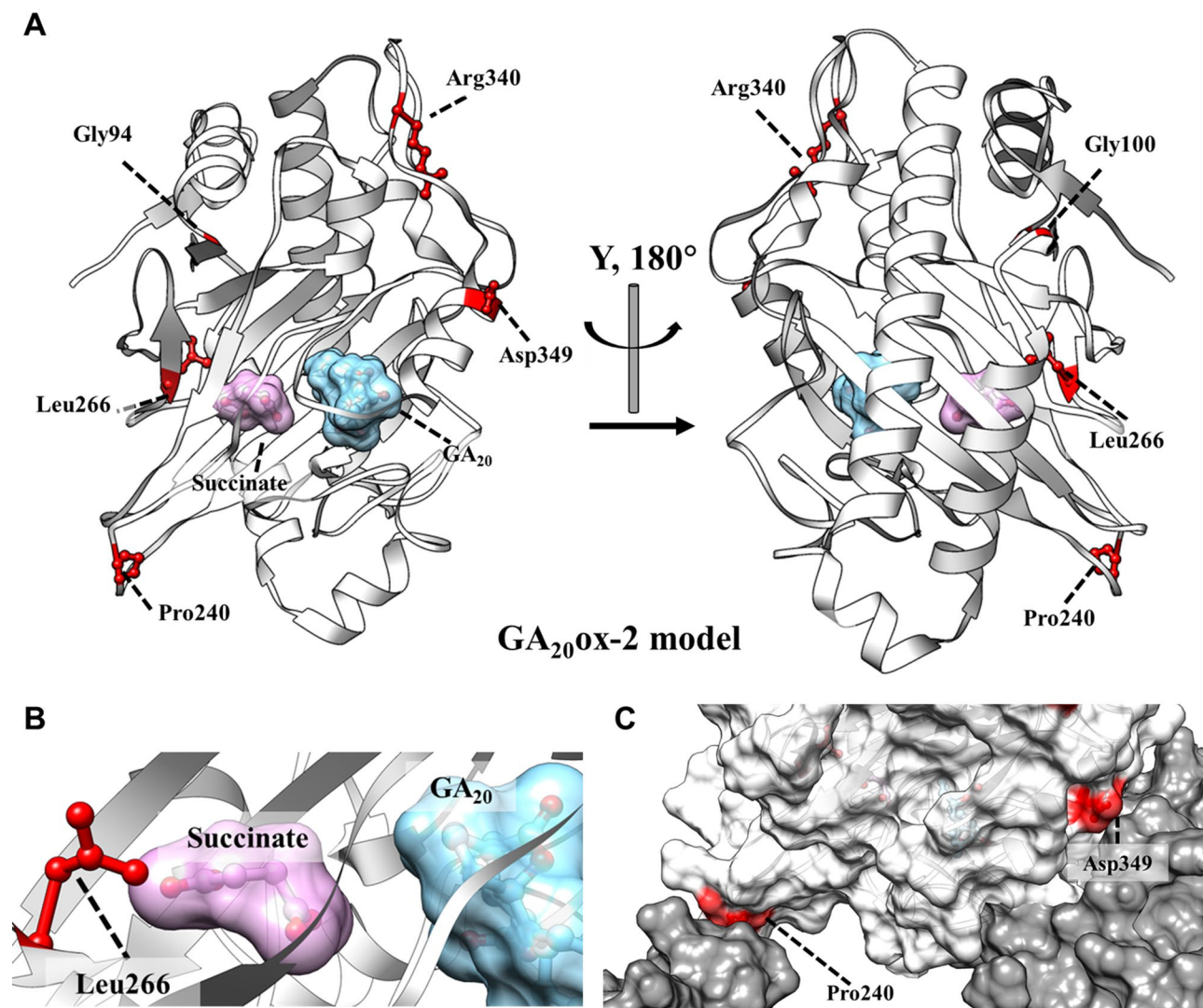


**Fig. 5** Model-based analysis of GA<sub>20</sub>ox-2 encoded by the *SD1* gene. **(A)** Model of two views of GA<sub>20</sub>ox-2 with all mutation sites labeled, generated using the MODELLER module. **(B)** The GA<sub>20</sub>ox-2 active site. The nearby location of Leu266 is indicated. **(C)** The positions of Asp349 and Pro240. The positions of the *SD1* mutations in GA<sub>20</sub>ox-2 are shown in red

the interaction between succinate and GA<sub>20</sub>. In contrast, D349 and P240 were located close to the interface regions of GA<sub>2</sub>ox3 (Fig. 5C). The D349H mutation (types 1 and 5) introduced a charge reversal, while P240L (types 1 and 10) increased hydrophobicity. Both mutations may disrupt intermolecular interactions by altering the local properties at these interfaces.

Apart from the three mutations described above, no further associations between GA<sub>20</sub>ox-2 activity and other mutations were evident from structural analysis. We hypothesize that mutations such as G100E and R340Q in type 2 variants may instead affect *SD1* protein stability.

#### *sd1* and *hd1* mutations are common in modern *indica* varieties

Before the Green Revolution, most rice landraces in Southeast Asia were tall, easily lodged, late maturing, and low yielding (Khush 1987; Khush et al. 2001; Gaur et al. 2020). In contrast, TN1 and IR8 are semi-dwarf and lodging resistant. Consequently, the *DGWW* type has been intensively utilized in rice breeding, accounting for 17% of the accessions in the 3K panel. The next most common was *Zhayeqing 8* (type 10) with 65 accessions (2.5%), while other types (3–6 and 8–9) comprised 66 accessions (Table 1). This suggests that the *DGWW* type was greatly favored by breeders to modulate plant height.

Another advantage of the modern varieties developed since the Green Revolution was their insensitivity to day length. To verify the genotypes of *SD1* and *Hd1* of

**Table 3** Haplotypes of *SD1* and *Hd1* genes in rice landraces related to the Green Revolution

|                        | Origin    | <i>SD1</i> haplotype | <i>Hd1</i> haplotype <sup>a</sup> |
|------------------------|-----------|----------------------|-----------------------------------|
| Hsinchu-Ai-Chueh-Chien | Taiwan    | Type 7               | Type 3                            |
| Ti-Chueh-Wu-Ko         | Taiwan    | Type 7               | Type 7                            |
| Liu-Tou-Tzu            | Taiwan    | Type 7               | Type 7                            |
| Ai-Tzu-Chung           | Taiwan    | Type 7               | Type 7                            |
| DGWG                   | Taiwan    | Type 7               | Type 7                            |
| Peta                   | Indonesia | Type 1               | Type 13                           |
| IR8                    | IRRI      | Type 7               | Type 7                            |
| Tsai-Yuan-Chung        | Taiwan    | Type 1               | Type 7                            |
| Taichung Native 1      | Taiwan    | Type 7               | Type 7                            |

<sup>a</sup>*Hd1* haplotypes were defined by Wu et al. (2020)

major varieties in Green Revolution, we used the haplotypes defined by Wu et al. (2020) to screen the 3K dataset and several Taiwanese landraces. Table 3 lists the allele types of the *SD1* and *Hd1* genes in rice landraces associated with the Green Revolution. DGWG, IR8, Tsai-Yuan-Chung, and TN1 carry type 7 *hd1* alleles, as also reported by Panibe et al. (2021), whereas Peta carries the type 13 *hd1* allele.

We further assessed the distribution of *hd1* haplotypes among 3K dataset accessions harboring LOF *sd1* alleles, specifically quantifying those with *hd1* type 7 or type 13. This dataset includes information on more than 200 accessions each from three rice-growing countries—435 from India, 248 from Indonesia, and 229 from the Philippines—including traditional landraces, which were used for further analysis. To evaluate the importance of the LOF *hd1* allele in modern varieties, we calculated the ratio of non-traditional accessions to the total number of accessions for each country. The *hd1* type 7 allele (represented by the DGWG haplotype) showed an 83.02% increase in frequency from before to after the Green Revolution, in contrast to only a 20.0% increase for the type 13 allele present in Peta (Table S14). According to our previous studies of *Hd1*, both type 7 and type 13 mutations originated in insular Southeast Asia thousands of years ago (Wu et al. 2020). This result indicates that most modern *indica* rice varieties inherited both LOF *sd1* and *hd1* alleles from DGWG.

Discussion

Before the Green Revolution, the plant height of most rice *indica* landraces at the maturing stage was over 160 cm (for example, 180–190 cm for Peta; Griffin 1979) and most were late maturing due to photoperiod sensitivity (Khush 1987). Since the introduction of the LOF *sd1* trait, breeders have intensively used *sd1* alleles to breed new varieties (Hargrove et al. 1988): for example, IR8 and IR64 in the Philippines (Khush et al. 2001) (labeled in yellow in Table S1) and Rasi and Jaya in India (Prasad et al. 2001). However, due to their partially functional *SD1-EQ*

trait, most *japonica* landraces or modern varieties are 100–120 cm tall. *SD1*, one of the genes controlling rice plant architecture, has been an important target for early selection as well as modern breeding. Many alleles of the *sd1* mutants have been studied. In the present study, we classified the mutant alleles we detected into 12 types and explored the origin of each allele. Using the 3K dataset, we also revealed that all modern *indica* varieties but only a few modern *japonica* varieties harbor the mutated alleles. This distribution pattern reflects the intensive selection for both semi-dwarfism and photoperiod insensitivity that was fundamental to the success of the Green Revolution.

Evolutionary and functional significance of the *SD1-EQ* allele in rice domestication and adaptation

At least 62 *dwarf* (*d*) and 15 *semi-dwarf* (*sd*) genes have been cataloged in rice (Kinoshita 1995), indicating that multiple genetic pathways can reduce plant height. However, *sd1* has been the principal semi-dwarf allele incorporated into modern breeding programs. Many other dwarf-related genes influence additional agronomic traits, such as tiller number, leaf greenness, leaf morphology, and spikelet number, causing undesirable pleiotropic effects. Within the *GA<sub>20</sub>* oxidase family, functional differentiation is evident. *GA<sub>20ox-1</sub>* is broadly expressed and influences multiple traits, whereas *GA<sub>20ox-3</sub>* and *GA<sub>20ox-4</sub>* show tissue-specific expression in anthers and panicles, respectively. In contrast, *GA<sub>20ox-2</sub>* (*SD1*) is primarily expressed in the culm and mainly affects stem elongation without substantially altering other major agronomic traits (Sakamoto et al. 2004; Table S15). This tissue-preferential expression may explain why LOF *sd1* reduced plant height with relatively limited pleiotropic effects.

The *SD1-EQ* allele type was chosen and maintained during early *japonica* domestication (Asano et al. 2011; Ashikari et al. 2025; Alam et al. 2025). However, it was not introgressed into *indica* rice with other domestication traits, such as *non-shattering* (*sh4*; Li et al. 2006; Choi and Purugganan 2018) and closed stature (*prog 1*; Tan et al. 2008; Jin et al. 2008). Sanger sequencing of 72 landraces and 42 wild rice accessions revealed that the *EQ* type *SD1* gene was selected thousands of years ago in ancient *japonica* landraces, suggesting that *SD1* was involved in rice domestication (Asano et al. 2011). A comparison of *GA<sub>20</sub>* oxidase activities of recombinant *SD1-GR*, *SD1-EQ*, *SD1-GQ* and *SD1-ER* proteins revealed that approximately three times more *GA<sub>20</sub>* was produced by *SD1-GR* (wild type) than by the other three types, with no significant difference in *GA<sub>20</sub>* production among *SD1-EQ*, *SD1-GQ*, and *SD1-ER* (Asano et al. 2011). The phenotypic records for these *SD1-ER* accessions showed that they had similar plant height to the wild type (Fig. 2,

Table S3): that is, their heights did not correspond to the level of  $GA_{20}ox-2$  activity, as they did among *EQ* (type 2) accessions (Asano et al. 2011). This suggests that environmental conditions, particularly temperature and latitude, may modulate the phenotypic expression of the *GR* and *EQ* allele type.

The *EQ* allele has been proposed to have undergone preferential selection during *japonica* domestication approximately 5,000 years ago (Asano et al. 2011; Alam et al. 2025; Ashikari et al. 2025). In the 3K dataset, 89% of traditional *japonica* landraces are homozygous for the *EQ* allele, whereas approximately 81% of *indica* accessions contain a homozygous *GR* allele, and 12.6% are heterozygous (Table 2). These contrasting frequency patterns are consistent with differential historical processes acting on the *SD1* locus during rice domestication and subsequent breeding. Although demographic factors such as population bottlenecks cannot be excluded, the near fixation of the *EQ* allele in traditional *japonica* landraces, together with previous studies (Zhao et al. 2018; Kou et al. 2020), supports a scenario in which the *EQ* allele was established early during *japonica* domestication.

#### Different spellings for the same lines may cause problems

Liu-Tou-Tzu was misspelled in another paper (Nagano et al. 2005), as noted previously. We noted other similar cases as well. “DG” in DGWG means “low foot (semi-dwarf),” and “WG” (Wu-Chien in NPGRC) means “black epiculus”; its sibling line Wu-Chien has a normal plant height. Another type 7 Taiwanese landrace, Ti-Chueh (semi-dwarf)-Wu-K’o (black glumes), also has a sibling line, Wu-K’o (with normal plant height). Both Wu-Chien and Wu-K’o are landraces whose genomes were sequenced in the 3K project. These two pairs of lines (i.e., DGWG and Wu-Chien; Ti-Chueh-Wu-K’o and Wu-K’o) were popular landraces cultivated in paddy fields in Taiwan during the Ching Dynasty (DAFTPG 1989). They are closely related to another sibling line, Wu-Li (black grain). Another semi-dwarf race, Hsinchu-Ai-Chueh-Chien, was also popular during the same period.

All these landraces were among the ~1000 rice lines brought to Taiwan from southern China in the late Ming Dynasty by the Han people approximately 400 years ago (DAFTPG 1989). Two Wu-K’o (Chiayi Wu-K’o, Taipei Woo-Co in IRRI) accessions are included in the 3K dataset: a Wu-Chien (Taichung Woo Gen 2) accession and a Wu-Li (Taitung Woo-Li in IRRI) accession. However, different institutes use different spellings for these accessions, such as Wu-Chien in TARI versus Woo Gen in IRRI, and in some cases one institute even uses different spellings for the same accession, such as both Wu-K’o and Woo-Co in IRRI. In summary, the Chinese word for Woo or Wu is exactly the same, but it is spelled differently by various institutes. Likewise, the terms Gen and

Chien rendered in the Roman alphabet for use in English refer to the same Chinese word. Table S16 lists the Chinese and English names of these lines, including their accession numbers in different seed stock centers. This is important information for rice researchers because it may help eliminate the misuse of genetic resources.

#### Many modern *indica* varieties have functional nucleotide polymorphisms in *SD1*

Most *japonica* rice cultivars, including traditional and modern accessions, are shorter than *indica* landraces due to the *GR*-to-*EQ* allele variation. Such effects are reflected in modern breeding programs. Notably, few modern *japonica* varieties have LOF mutation of *SD1* alleles. Approximately 60 *japonica* accessions in the 3K panel contain *sd1* types 3–6, with these 4 alleles originating in *japonica* rice. In Taiwan, both *japonica* and *indica* rice varieties are produced and consumed, and *japonica* × *indica* crosses have often been used in breeding programs. However, no modern *japonica* varieties harbor the DGWG (type 7) mutation, despite the *indica* parental lines used in breeding carrying this allele. Thus, plant height was not a major target during the breeding of most *japonica* varieties. By contrast, all modern *indica* rice varieties contain a type 7–10 *sd1* allele, and more than 400 (35%) modern varieties in the 3K dataset carry the 383-bp deletion. Thus, the *sd1* type 7 allele has been used intensively in breeding programs worldwide, where the semi-dwarf trait was predominantly derived from DGWG or its progeny (Khush et al. 2001; Prasad et al. 2001). Approximately 40 varieties contain *sd1* types 8 or 9, mostly modern varieties in China. Type 10, the allele from Indonesia, was used in ~50 accessions in Southeast Asia. Notably, *indica* rice production comprises ~80% of worldwide rice production. The mutated *SD1* gene is one of the primary reasons for the increase in rice production since the mid-1960s because it leads to semi-dwarf growth habit resulting in lodging resistance and the ability to withstand heavy fertilizer applications.

#### Functional implications of *SD1* haplotypes and applications in genome editing

We observed that plant height varied among *SD1* haplotypes (Fig. 2), consistent with the findings of Ogi et al. (1993). Using isogenic lines of the *japonica* variety ‘Norin 29’, they demonstrated the functional effects of four *sd1* alleles on culm length, internode length, and yield-related traits. When compared with Norin 29 (85.1 cm), SC-TN1 (DGWG type) showed the greatest reduction in culm length (57.5 cm), whereas SC-AJNT (*Aijiao-Nante* type) and SC-SRN (*Jikkoku* type) exhibited a less severe reduction (60.4 cm and 59.4 cm, respectively). (Ogi et al. 1993). Furthermore, Nakamura et al. (1991) analyzed protein profiles in germinating seeds of these lines and identified

differential protein spots: SRP-1 in Norin 29 and SRP-2 in SC-TN1. This suggests that the *sd1* mutation alters downstream protein expression (Nakamura et al. 1991). In this study, we classified the 12 *SD1* haplotypes based on their mutation sites (Table 1), and we observed distinct differences in plant height among the haplotype groups (Fig. 2). The protein modeling results (Fig. 5) further suggested that the positions of these mutations have distinct impacts on *GA<sub>20ox</sub>-2* protein structure and function.

While natural *sd1* alleles have been extensively characterized, recent advancements have enabled the artificial generation of these traits. Two recent studies used CRISPR-Cas9 to generate semi-dwarf rice lines and analyzed the agronomic traits of the edited plants (Hu et al. 2019; Biswas et al. 2020). Hu et al. (2019) targeted the *SD1* gene in two elite landraces, Kasalath (*aus*) and TeTePu. Using 10 or 30 randomly chosen *T*<sub>2</sub> plants grown in paddy fields under natural conditions, the authors demonstrated that gene editing could rapidly create *sd1* mutants. The seed production of the edited plants was better than that of the wild type under modern cultivation, while the desirable agronomic traits of these varieties (i.e., tolerance for low phosphorus in Kasalath and broad-spectrum resistance to diseases and insects in TeTePu) were maintained (Hu et al. 2019).

In contrast, Biswas et al. (2020) edited *SD1* in three elite cultivars: 9815B, JIAODA138, and HUAIDAO1055. The authors investigated plant height and yield across the *T*<sub>3</sub> and *T*<sub>4</sub> generations using 10 independent lines (10 plants each) grown in the paddy field under natural conditions. The mutated plants showed significantly reduced seed production even though they had reduced plant height. Therefore, the authors suggested that the use of the CRISPR-Cas9 system to edit *SD1* can have varying results and still requires more effort and time (Biswas et al. 2020). Considering that Biswas et al. (2020) utilized advanced generations (*T*<sub>3</sub>/*T*<sub>4</sub>), a larger number of independent lines, and more comprehensive statistical analysis, their results may provide a more robust assessment than the earlier findings of Hu et al. (2019).

Given the variable outcomes observed in previous studies, selecting the optimal target site is crucial for successful genome editing. Moreover, different *sd1* alleles result in varying phenotypic effects. The results of our *SD1* protein modeling should therefore be valuable in providing insights for rational design, shedding light on how to precisely manipulate plant height through *SD1* modification.

#### The impact of DGWG on the Green Revolution

Prior to the release of IR8, the earliest semi-dwarf modern rice variety was TN1. Derived from a cross between DGWG and Tsai-Yuan-Chung performed during the second cropping season of 1949, the line designated as “Taichung Native Bred 1” was selected during the second

season in 1952. It was officially released as TN1 in early 1957 and subsequently cultivated widely in Taiwan. Its full genome is available (Panibe et al. 2021). In 1966, 50 tons of TN1 seeds were shipped from Taiwan to India due to a local famine. Notably, TN1 covered a production area of 800,000 ha in India during 1968–1969. Taiwanese semi-dwarf *indica* germplasm served as the foundation for new varieties subsequently developed at the IRRI and elsewhere in Asia (Barker et al. 1985; Hargrove et al. 1988).

Following the era of TN1, IR8 was developed from a cross between Peta and DGWG in 1962. A single plant (the 3rd plant of line 288) in the *F*<sub>5</sub> generation was selected to produce *F*<sub>6</sub> (IR8-288-3) seed and was sent for testing without further selection at the time. In 1966, IR8-288-3 showed dramatic yield increases in the trials and was later named IR8 in the same year (Khush et al. 2001). Khush et al. (2001) reported that IR8 was photoperiod insensitive and matured in 130–135 days; thereafter, the IRRI devoted efforts to producing a uniform seed source of IR8.

However, this duration was not short enough to allow an additional crop season within a year. Therefore, the IRRI continued to improve the rice lines, aiming for a shorter growth duration. Subsequently, IR28 and IR36 matured in 110 days, and the growth duration for IR50 and IR58 was further reduced to 100–105 days (Khush 1987; Khush et al. 2001). Our results demonstrate that the majority of modern varieties possessing the LOF *sd1* allele also carry the LOF *hd1* allele, particularly in the three major indica-producing countries: India, Indonesia, and the Philippines (Table S14). This confirms the continuous selection for early maturity in rice breeding.

Most Green Revolution rice varieties are of the *indica* type and are grown in subtropical and tropical regions, where temperatures are suitable for plant growth year-round. Thus, changes in sensitivity to day length play an important role in their cultivation because this can increase the number of cropping seasons per year and adjust the growing season to escape from or avoid stress caused by seasonal typhoons, monsoons, or drought. As a result of combining the DGWG-derived semi-dwarf trait with photoperiod-insensitivity traits, rice production has increased tremendously since the mid-1960s.

#### Conclusions

This study provides a comprehensive overview of natural variation in *SD1*, revealing 12 haplotypes with distinct evolutionary histories and functional impacts. The GR-to-EQ amino acid changes were confirmed as ancient domestication signatures that became fixed in most traditional *japonica* landraces, whereas other *sd1* mutations arose more recently. In particular, the 383-bp DGWG deletion was preferentially incorporated into

modern *indica* cultivars during the Green Revolution. We also show that the loss-of-function *hd1* allele from DGWG was introgressed into IR8 and many subsequent *indica* varieties, indicating that photoperiod insensitivity was selected alongside reduced plant height to enable early maturity and multiple cropping. Together, our findings clarify the dual genetic foundations of the Green Revolution and provide an evolutionary and functional framework that can guide future rice breeding and genome-editing strategies.

## Materials and methods

### Classification of *SD1* haplotypes and DNA sequence analysis

To identify additional *SD1* haplotypes from the public database, two whole-genome sequencing datasets were used: the Rice 3K project (3K-RGP 2014) and an Asian wild rice dataset (Zhao et al. 2018). Some Taiwanese landraces were also sequenced for comparison. Information including the accession names, seed code numbers, *sd1* types, and DNA accession numbers is provided in Table S8. The paired reads were mapped against the Nipponbare IRGSP-1.0 RefSeq database (Kawahara et al. 2013; IRGSP 2005). Each VCF file of the 3K sequences was first filtered by read depth > 5 and SNP frequency > 0.8. To examine the functional impact of nucleotide variants, rice genome sequencing data were analyzed using SnpEff (Cingolani et al. 2012). Based on genome annotation, these sequence variants were classified according to their location (ORF, intron, and splice sites) and predicted functional impact (missense mutation, frame shift, and early translational termination).

Information in the SAM format was used to determine whether the paired reads were properly aligned based on insert size and orientation. The larger insert size of properly oriented paired reads indicates a possible deletion between the paired reads. To illustrate that the long deletion occurred at exons or the promoter region, snapshot images were produced using Integrative Genomics Viewer (Robinson et al. 2011; Thorvaldsdottir et al. 2013; Robinson et al. 2017) for the variant event in the next-generation sequencing mapped images of the genome for the indicated lines. Accessions carrying exclusively type 1 or type 2 mutations were first identified. Subsequently, types 3 through 12 were classified, as they may share amino acid substitutions at positions 100 or 340 with types 1 and 2.

### Estimating the diversity of different *SD1* haplotypes

DNA sequences were aligned using MUSCLE (Edgar 2004a, b). The -10- to +10-kb region of the *SD1* gene, corresponding to Nipponbare genome chromosome 1 positions 38,372,466 to 38,395,208, was used for analysis. Statistical analysis was performed using DnaSP.v6

(Rozas et al. 2017). Items analyzed included the number of polymorphic (segregating) sites ( $S$ ); total number of mutations ( $\eta$ ); average number of nucleotide differences ( $k$ ); nucleotide diversity ( $\pi$ ; Tajima 1983); theta ( $\theta$ ; per sequence) from  $\eta$  ( $\theta_w$ ); and Watterson's estimator of  $\theta$  (per site) from  $\eta$  ( $\theta_w$ ; Watterson 1975). The Tajima's  $D$  neutrality test (Tajima 1989) was used to test the neutral mutation hypothesis. The  $D$  value is based on the discrepancy between  $\pi$  and  $\theta_w$ . Thus, negative values indicate excess low-frequency polymorphism. These values were calculated after removing missing data and alignment gaps.

### Phylogenetic analysis of *SD1*

Phylogenetic relationships were inferred from nucleotide sequences derived from the selected VCF regions. The VCF files were converted into PHYLIP format using vcf-2phylip v2.6 (Ortiz 2019). The evolutionary history was reconstructed using the Neighbor-Joining (NJ) method (Saitou and Nei 1987). Evolutionary distances were calculated via the Maximum Composite Likelihood method (Tamura et al. 2004), expressed as the number of base substitutions per site. To ensure data robustness, the pairwise deletion option was applied to all ambiguous positions for each sequence pair in the final analysis. Statistical support for internal nodes was assessed through a bootstrap test with 1,000 replicates (Felsenstein 1985). In the final phylogenetic trees, bootstrap values lower than 60% were collapsed to prioritize well-supported clades. All computational procedures were performed in MEGA12 (Stecher et al. 2020, 2026) using multi-threaded parallel processing to enhance computational efficiency.

### Protein modeling of *GA<sub>20</sub>ox-2*

A crystal structure of the gene product of *SD1* (i.e., *GA<sub>20</sub>ox-2*) is not available. Therefore, a homology model was generated using the MODELLER module (Webb and Sali 2021) implemented in the Chimera software package (Pettersen et al. 2004). The crystal structure of *GA<sub>20</sub>ox3* (PDB ID: 6KU3; Takehara et al. 2020) was used as the template because this enzyme utilizes a substrate (*GA<sub>4</sub>*) that is structurally similar to *GA<sub>20</sub>*. The quality of the resulting *GA<sub>20</sub>ox-2* model was assessed using the GA341 score. According to Chimera, the GA341 score reflects model reliability; a score > 0.7 indicates a probability of correct folding greater than 95% (Melo et al. 2002). The *GA<sub>20</sub>ox-2* model yielded a GA341 score of 1.0, suggesting high reliability. For docking, the modeled *GA<sub>20</sub>ox-2* structure was prepared using the Protein Preparation Wizard module of Maestro (Schrödinger), which added hydrogen atoms and assigned atomic charges. The compounds *GA<sub>20</sub>* and succinate were prepared using the Lig-Prep Wizard module of Maestro (Schrödinger). Succinate was first docked into the binding site using Glide XP

(Friesner et al. 2006). The docked position of succinate was considered part of the active site, after which GA<sub>20</sub> was docked into the same binding site.

#### Abbreviations

|       |                                         |
|-------|-----------------------------------------|
| DGWG  | Dee-Geo-Woo-Gen                         |
| Indel | Insertion and deletion                  |
| IRRI  | International Rice Research Institute   |
| LOF   | Loss of function                        |
| NPGRC | National Plant Genetics Resource Center |
| NGS   | Next-generation sequencing              |
| ORF   | Open reading frame                      |
| QTL   | Quantitative trait loci                 |
| TARI  | Taiwan Agricultural Research Institute  |
| TN1   | Taichung Native 1                       |
| TR    | Tandem repeat                           |

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40529-026-00493-3>.

Supplementary Material 1: Table S1. Accession numbers, accession names, collection regions, subspecies, and types of SD1 alleles in the 3K dataset

Supplementary Material 2: Table S2. ANOVA for plant height of the 11 SD1 haplotypes. Table S3. Statistical analysis of plant height for the SD1 haplotypes. Table S4. Distribution of GR, EQ, and ER at aa positions 100 and 340 in the 3K panel. Table S5. Accessions used for phylogenetic analysis of type 7 sd1 alleles. Table S6. Accessions used for phylogenetic analysis of type 6 sd1 alleles. Table S7. Accessions used for phylogenetic analysis of types 6 and 10 sd1 alleles. Table S8. Taiwanese rice landraces used in this study. Table S9. Accessions used for phylogenetic analysis of type 7 SD1 genes. Table S10. Accessions used for phylogenetic analysis of 67 accessions containing type 7 sd1 alleles. Table S11. Accessions used for phylogenetic analysis of 10 Taiwanese and wild rice accessions harboring type 7 sd1 alleles. Table S12. Accessions used to calculate the selection parameters for type 2 SD1 genes. Table S13. Selection parameters for 4 SD1 haplotypes. Table S14. Distribution of hd1 type 7 and type 13 alleles among 3K accessions with loss-of-function sd1 mutations in Asian rice-growing countries. Table S15. Gene IDs, Plant Ontology (PO) and Trait Ontology (TO) annotations of GA20ox homologs in rice. Table S16. Dee-Geo-Woo-Gen (DGWG)-related accessions and their spellings in seed stock centers

Supplementary Material 3: Fig. S1. Types 6 and 7 sd1 alleles shown in Integrative Genomics Viewer images. Fig. S2. Geographic distribution of accessions harboring type 6 sd1 alleles. Fig. S3. Phylogenetic analysis of types 6 and 7 sd1 alleles. Fig. S4. Phylogenetic analysis of accessions harboring type 6 sd1 alleles. Fig. S5. Phylogenetic analysis of accessions harboring types 6 and 10 sd1 alleles. Fig. S6. Phylogenetic analysis of 67 accessions harboring type 7 sd1 alleles. Fig. S7. Phylogenetic analysis of 10 accessions harboring type 7 sd1 alleles. Fig. S8. Venn diagram of single-nucleotide polymorphisms (SNPs) shared among five early landraces with type 7 sd1 mutations. Fig. S9. Alignment of GA20 oxidase homologs in the rice genome

#### Acknowledgements

We thank Ms. Lie-Hong Wu for maintaining the greenhouse plants, Dr. Fu-Jin Wei for assisting with the analysis of plant height data, and Dr. Dong-Hong Wu provided plant materials and related information. We also thank Ms. Laura Smales (BioMedEditing, Toronto, Canada) and Plant Editors for English editing.

#### Author contributions

YICH and WYC designed the study and wrote the manuscript. CCW, YTT, CKL, and WYC performed bioinformatics analysis of the 3K and Taiwan datasets. LTH prepared high-quality genomic DNA. HCW and KCH performed protein modeling.

#### Funding

This project was supported by funding from the Ministry of Science and Technology, Taiwan (MOST 107-2313-B-001-007-MY2), and an Academia Sinica Investigator Award to YICH, and funding from the National Science and Technology Council, Taiwan (NSTC 114-2313-B-005-003-) to WYC.

#### Data availability

All data generated or analyzed during this study are included in this article and the supplementary information files. All seeds mentioned are available in the TARI or IRRI seed stock centers.

#### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 15 December 2025 / Accepted: 10 March 2026

Published online: 26 March 2026

#### References

- Alam O, Gutaker R, Surya N et al (2025) Evolutionary histories of functional mutations during the domestication and spread of *japonica* rice in Asia. *Proc Natl Acad Sci USA* 122:e2514614122. <https://doi.org/10.1073/pnas.2514614122>
- Alexandrov N, Tai S, Wang W et al (2015) SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res* 43:D1023–D1027. <https://doi.org/10.1093/nar/gku1039>
- Asano K, Takashi T, Miura K et al (2007) Genetic and molecular analysis of utility of sd1 alleles in rice breeding. *Breed Sci* 57:53–58. <https://doi.org/10.1270/jsbbs.57.53>
- Asano K, Yamasaki M, Takuno S et al (2011) Artificial selection for a green revolution gene during *japonica* rice domestication. *Proc Natl Acad Sci USA* 108:11034–11039. <https://doi.org/10.1073/pnas.1019490108>
- Ashikari M, Sasaki A, Ueguchi-Tanaka M et al (2002) Loss-of-function of a Rice Gibberellin Biosynthetic Gene, GA20 oxidase (GA20ox-2), Led to the Rice Green Revolution. *Breed Sci* 52:143–150. <https://doi.org/10.1270/jsbbs.52.143>
- Ashikari M, Nagai K, Bailey-Serres J (2025) Surviving floods: Escape and quiescence strategies of rice coping with submergence. *Plant Physiol* 197:kiaf029. <https://doi.org/10.1093/plphys/kiaf029>
- Barker R, Herdt RW, Rose B (1985) The rice economy of Asia, 1st edn. Routledge
- Bhuvaneswari S, Gopala Krishnan S, Ellur RK et al (2020) Discovery of a Novel Induced Polymorphism in SD1 Gene Governing Semi-Dwarfism in Rice and Development of a Functional Marker for Marker-Assisted Selection. *Plants* 9:1198. <https://doi.org/10.3390/plants9091198>
- Biswas S, Tian J, Li R et al (2020) Investigation of CRISPR/Cas9-induced SD1 rice mutants highlights the importance of molecular characterization in plant molecular breeding. *J Genet Genomics* 47:273–280. <https://doi.org/10.1016/j.jgg.2020.04.004>
- Chapman SC, Mathews KL, Trethowan RM, Singh RP (2007) Relationships between height and yield in near-isogenic spring wheats that contrast for major reduced height genes. *Euphytica* 157:391–397. <https://doi.org/10.1007/s10681-006-9304-3>
- Choi JY, Purugganan MD (2018) Multiple Origin but Single Domestication Led to *Oryza sativa*. *G3 Genes/Genomes/Genetics* 8:797–803. <https://doi.org/10.1534/g3.117.300334>
- Cingolani P, Platts A, Wang LL et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w<sup>1118</sup>; iso-2; iso-3. *Fly* 6:80–92. <https://doi.org/10.4161/fly.19695>
- DAFTPG (1989) The history of Taiwan rice development and production. Department of Agriculture and Forestry of the Taiwan Provincial Government
- Doi K, Izawa T, Fuse T et al (2004) *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT-like* gene expression

- independently of *Hd1*. *Genes Dev* 18:926–936. <https://doi.org/10.1101/gad.1189604>
- Edgar RC (2004a) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Edgar RC (2004b) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. <https://doi.org/10.1186/1471-2105-5-113>
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783. <https://doi.org/10.2307/2408678>
- Foster KW, Rutger JN (1978) Inheritance of semidwarfism in rice, *Oryza sativa* L. *Genetics* 88:559–574. <https://doi.org/10.1093/genetics/88.3.559>
- Friesner RA, Murphy RB, Repasky MP et al (2006) Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J Med Chem* 49:6177–6196. <https://doi.org/10.1021/jm051256o>
- Futsuhara Y, Toriyama K, Tsunoda K (1967) Breeding of a new rice variety Reimei by gamma-ray irradiation. *Ikushugaku zasshi* 17:85–90. <https://doi.org/10.1270/jsbbs1951.17.85>
- García-Martínez JL, López-Díaz I, Sánchez-Beltrán MJ et al (1997) Isolation and transcript analysis of gibberellin 20-oxidase genes in pea and bean in relation to fruit development. *Plant Mol Biol* 33:1073–1084. <https://doi.org/10.1023/A:1005715722193>
- Gaur VS, Channappa G, Chakraborti M et al (2020) Green revolution' dwarf gene *sd1* of rice has gigantic impact. *Brief Funct Genomics* 19:390–409. <https://doi.org/10.1093/bfpgp/ela019>
- Gemayel R, Vences MD, Legendre M, Verstrepen KJ (2010) Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. *Annu Rev Genet* 44:445–477. <https://doi.org/10.1146/annurev-genet-072610-155046>
- Griffin K (1979) The political economy of agrarian change: an essay on the Green Revolution. Springer
- Gutaker RM, Groen SC, Bellis ES et al (2020) Genomic history and ecology of the geographic spread of rice. *Nat Plants* 6:492–502. <https://doi.org/10.1038/s41477-020-0659-6>
- Hargrove TR, Cabanilla VL (1979) The Impact of Semidwarf Varieties on Asian Rice-Breeding Programs. *Bioscience* 29:731–735. <https://doi.org/10.2307/1307667>
- Hargrove T, Coffman WR (2006) Breeding history. *Rice Today* 5:34–38
- Hargrove TR, Cabanilla VL, Coffman WR (1988) Twenty Years of Rice Breeding. *Bioscience* 38:675–681. <https://doi.org/10.2307/1310869>
- Hu X, Cui Y, Dong G et al (2019) Using CRISPR-Cas9 to generate semi-dwarf rice lines in elite landraces. *Sci Rep* 9:19096. <https://doi.org/10.1038/s41598-019-55757-9>
- International Rice Genome Sequencing Project, Sasaki T (2005) The map-based sequence of the rice genome. *Nature* 436:793–800. <https://doi.org/10.1038/nature03895>
- Jin J, Huang W, Gao J-P et al (2008) Genetic control of rice plant architecture under domestication. *Nat Genet* 40:1365–1369. <https://doi.org/10.1038/ng.247>
- Kawahara Y, De La Bastide M, Hamilton JP et al (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4. <https://doi.org/10.1186/1939-8433-6-4>
- Khush GS (1987) Rice breeding: Past, present and future. *J Genet* 66:195–216. <https://doi.org/10.1007/BF02927713>
- Khush GS (1995) Modern varieties — Their real contribution to food supply and equity. *GeoJournal* 35:275–284. <https://doi.org/10.1007/BF00989135>
- Khush GS (1999) Green revolution: preparing for the 21st century. *Genome* 42:646–655. <https://doi.org/10.1139/g99-044>
- Khush GS, Coffman WR, Beachell HM (2001) The history of rice breeding: IRRI's contribution. In: Rockwood WG Rice Research and Production in 21st Century: symposium honoring, Chandler RF, Jr. International Rice Research Institute, Los Baños, Philippines, pp 117–135
- Kikuchi F, Itakura N, Ikehashi H et al (1985) Genetic analysis of semi-dwarfism in high yielding rice varieties in Japan. *Bull Nat Inst Agri Sci Jap D* 36:125–145
- Kinoshita T (1995) Report of committee on gene symbolization, nomenclature and linkage groups. *Rice Genet Newsl* 12:9–154
- Kluza A, Niedziałkowska E, Kurpiewska K et al (2018) Crystal structure of thebaïne 6-O-demethylase from the morphine biosynthesis pathway. *J Struct Biol* 202:229–235. <https://doi.org/10.1016/j.jsb.2018.01.007>
- Korotkov EV, Kamionskaya AM, Korotkova MA (2021) Detection of Highly Divergent Tandem Repeats in the Rice Genome. *Genes* 12:473. <https://doi.org/10.3390/genes12040473>
- Kou Y, Liao Y, Toivainen T et al (2020) Evolutionary Genomics of Structural Variation in Asian Rice (*Oryza sativa*) Domestication. *Mol Biol Evol* 37:3507–3524. <https://doi.org/10.1093/molbev/msaa185>
- Li C, Zhou A, Sang T (2006) Rice Domestication by Reducing Shattering. *Science* 311:1936–1939. <https://doi.org/10.1126/science.1123604>
- Mansueto L, Fuentes RR, Borja FN et al (2017) Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Res* 45:D1075–D1081. <https://doi.org/10.1093/nar/gkw1135>
- Melo F, Sánchez R, Sali A (2002) Statistical potentials for fold assessment. *Protein Sci* 11:430–448. <https://doi.org/10.1002/pro.110430>
- Nagano H, Onishi K, Ogasawara M et al (2005) Genealogy of the Green Revolution gene in rice. *Genes Genet Syst* 80:351–356. <https://doi.org/10.1266/ggs.80.351>
- Nakamura A, Hirano H, Kikuchi F (1991) Identification and genetic analysis of semidwarfism-related proteins in rice (*Oryza sativa* L). *Theor Appl Genet* 81:376–380. <https://doi.org/10.1007/BF00228679>
- Ogi Y, Kato H, Maruyama K, Kikuchi F (1993) The Effects on the Culm Length and Other Agronomic Characters Caused by Semidwarfing Genes at the *sd-1* Locus in Rice. *Ikushugaku zasshi* 43:267–275. <https://doi.org/10.1270/jsbbs1951.43.267>
- Ortiz EM (2019) vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. <https://doi.org/10.5281/zenodo.2540861>
- Panibe JP, Wang L, Li J et al (2021) Chromosomal-level genome assembly of the semi-dwarf rice Taichung Native 1, an initiator of Green Revolution. *Genomics* 113:2656–2674. <https://doi.org/10.1016/j.ygeno.2021.06.006>
- Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. <https://doi.org/10.1002/jcc.20084>
- Pimenta Lange MJ, Lange T (2006) Gibberellin Biosynthesis and the Regulation of Plant Development. *Plant Biol* 8:281–290. <https://doi.org/10.1055/s-2006-923882>
- Prasad GSV, Prasadara U, Rani NS et al (2001) Indian rice varieties released in countries around the world. *Curr Sci* 80:1508–1511
- RGP K (2014) The 3,000 rice genomes project. *Gigascience* 3:2047–217X. <https://doi.org/10.1186/2047-217X-3-7>
- Robinson JT, Thorvaldsdóttir H, Winckler W et al (2011) Integrative genomics viewer. *Nat Biotechnol* 29:24–26. <https://doi.org/10.1038/nbt.1754>
- Robinson JT, Thorvaldsdóttir H, Wenger AM et al (2017) Variant Review with the Integrative Genomics Viewer. *Cancer Res* 77:e31–e34. <https://doi.org/10.1158/0008-5472.CAN-17-0337>
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Sakamoto T, Miura K, Itoh H et al (2004) An Overview of Gibberellin Metabolism Enzyme Genes and Their Related Mutants in Rice. *Plant Physiol* 134:1642–1653. <https://doi.org/10.1104/pp.103.033696>
- Sasaki A, Ashikari M, Ueguchi-Tanaka M et al (2002) A mutant gibberellin-synthesis gene in rice. *Nature* 416:701–702. <https://doi.org/10.1038/416701a>
- Spielmeyer W, Ellis MH, Chandler PM (2002) Semidwarf (*sd-1*), green revolution rice, contains a defective gibberellin 20-oxidase gene. *Proc Natl Acad Sci USA* 99:9043–9048. <https://doi.org/10.1073/pnas.132266399>
- Stecher G, Tamura K, Kumar S (2020) Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol Biol Evol* 37:1237–1239. <https://doi.org/10.1093/molbev/msz312>
- Stecher G, Suleski M, Tao Q et al (2026) MEGA 12.1: Cross-Platform Release for macOS and Linux Operating Systems. *J Mol Evol* 94:14–18. <https://doi.org/10.1007/s00239-025-10287-z>
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460. <https://doi.org/10.1093/genetics/105.2.437>
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595. <https://doi.org/10.1093/genetics/123.3.585>
- Takahashi Y, Teshima KM, Yokoi S et al (2009) Variations in *Hd1* proteins, *Hd3a* promoters, and *Ehd1* expression levels contribute to diversity of flowering time in cultivated rice. *Proc Natl Acad Sci USA* 106:4555–4560. <https://doi.org/10.1073/pnas.0812092106>
- Takehara S, Sakuraba S, Mikami B et al (2020) A common allosteric mechanism regulates homeostatic inactivation of auxin and gibberellin. *Nat Commun* 11:2143. <https://doi.org/10.1038/s41467-020-16068-0>
- Tamaki S, Matsuo S, Wong HL et al (2007) *Hd3a* Protein Is a Mobile Flowering Signal in Rice. *Science* 316:1033–1036. <https://doi.org/10.1126/science.1141753>

- Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci USA* 101:11030–11035. <https://doi.org/10.1073/pnas.0404206101>
- Tan L, Li X, Liu F et al (2008) Control of a key transition from prostrate to erect growth in rice domestication. *Nat Genet* 40:1360–1364. <https://doi.org/10.1038/ng.197>
- Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14:178–192. <https://doi.org/10.1093/bib/bbs017>
- Tomita M (2009) Introgression of Green Revolution *sd1* gene into isogenic genome of rice super cultivar Koshihikari to create novel semidwarf cultivar ‘Hikarishinseiki’ (Koshihikari-*sd1*). *Field Crops Res* 114:173–181. <https://doi.org/10.1016/j.fcr.2009.05.004>
- Toyomasu T, Kawaide H, Sekimoto H et al (1997) Cloning and characterization of a cDNA encoding gibberellin 20-oxidase from rice (*Oryza sativa*) seedlings. *Physiol Plant* 99:111–118. <https://doi.org/10.1111/j.1399-3054.1997.tb03438.x>
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276. [https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9)
- Webb B, Sali A (2021) Protein Structure Modeling with MODELLER. In: Chen YW, Yiu C-PB (eds) *Structural Genomics*. Springer US, New York, NY, pp 239–255
- Wei F-J, Tsai Y-C, Hsu Y-M et al (2016) Lack of Genotype and Phenotype Correlation in a Rice T-DNA Tagged Line Is Likely Caused by Introgression in the Seed Source. *PLoS ONE* 11:e0155768. <https://doi.org/10.1371/journal.pone.0155768>
- Wu C-C, Wei F-J, Chiou W-Y et al (2020) Studies of rice *Hd1* haplotypes worldwide reveal adaptation of flowering time to different environments. *PLoS ONE* 15:e0239028. <https://doi.org/10.1371/journal.pone.0239028>
- Wu C-C, Liu C-K, Wei F-J, Crop et al (2022) *Environ Bioinform* 18:12–36. [https://doi.org/10.30061/CEB.202212\\_18.0002](https://doi.org/10.30061/CEB.202212_18.0002)
- Xue W, Xing Y, Weng X et al (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat Genet* 40:761–767. <https://doi.org/10.1038/ng.143>
- Yano M, Harushima Y, Nagamura Y et al (1997) Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. *Theor Appl Genet* 95:1025–1032. <https://doi.org/10.1007/s001220050658>
- Yano M, Katayose Y, Ashikari M et al (2000) *Hd1*, a Major Photoperiod Sensitivity Quantitative Trait Locus in Rice, Is Closely Related to the Arabidopsis Flowering Time Gene *CONSTANS*. *Plant Cell* 12:2473–2483. <https://doi.org/10.1105/tpc.12.12.2473>
- Zhao Z, Guo C, Sutharzan S et al (2014) Genome-wide analysis of tandem repeats in plants and green algae. *G3 Genes|Genomes|Genetics*. 4:67–78. <https://doi.org/10.1534/g3.113.008524>
- Zhao Q, Feng Q, Lu H et al (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat Genet* 50:278–284. <https://doi.org/10.1038/s41588-018-0041-z>

## Publisher's note

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