# Variations in Hd1 proteins, *Hd3a* promoters, and *Ehd1* expression levels contribute to diversity of flowering time in cultivated rice

**Yasuyuki Takahashi[a], Kosuke M. Teshima[b], Shuji Yokoi[a,1], Hideki Innan[b], and Ko Shimamoto[a,2]**

[a]Laboratory of Plant Molecular Genetics, Nara Institute of Science and Technology, Ikoma, Nara 630-0101, Japan; and [b]The Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan

Rice is a facultative short-day plant, and molecular genetic studies have identified the major genes involved in short-day flowering. However, the molecular mechanisms promoting the diversity of flowering time in cultivated rice are not known. We used a core collection of 64 rice cultivars that represent the genetic diversity of 332 accessions from around the world and studied the expression levels and polymorphisms of 6 genes in the short-day flowering pathway. The RNA levels of *Heading date 3a* (*Hd3a*), encoding a floral activator, are highly correlated with flowering time, and there is a high degree of polymorphism in the Heading date 1 (Hd1) protein, which is a major regulator of *Hd3a* expression. Functional and nonfunctional alleles of *Hd1* are associated with early and late flowering, respectively, suggesting that *Hd1* is a major determinant of variation in flowering time of cultivated rice. We also found that the type of *Hd3a* promoter and the level of *Ehd1* expression contribute to the diversity in flowering time and *Hd3a* expression level. We evaluated the contributions of these 3 factors by a statistical analysis using a simple linear model, and the results supported our experimental observations.

cereal | florigen | natural variation | photoperiod | polymorphism

**R**ice (*Oryza sativa* L.) has evolved during the last 8,000 to 10,000 years of domestication and breeding (1, 2). A major reason for the spread of rice cultivation to a wide range of geographical regions, and for the increases in yield, is the diversification of flowering time (1). In general, rice is known as a short-day plant that induces transition from the vegetative phase to the reproductive phase when it senses a decrease in day length. The molecular genetic pathway for short-day flowering in cultivated rice (Fig. 1*A*) is relatively well characterized (3–5). Signals from light and circadian clocks are received by *OsGI*, the rice orthologue of *Arabidopsis GIGANTEA*, and it regulates expression of *Heading date 1* (*Hd1*) and *OsMADS51* (6–8). *Hd1* and its *Arabidopsis* orthologue *CONSTANS* encode zinc-finger type transcriptional activators with the *CO*, *CO*-like, and *TOC1* (CCT) domains (9). *Hd1* regulates *Heading date 3a* (*Hd3a*) expression (7, 9, 10). *Hd3a* is a rice orthologue of *Arabidopsis FLOWERING LOCUS T* (*FT*), and these genes recently were shown to encode a mobile flowering signal (11–16). *RICE FLOWERING LOCUS T1* (*RFT1*) belongs to the rice *FT-like* gene family and functions as a floral activator, acting redundantly with *Hd3a* (17, 18). *OsMADS51* encodes a type I MADS-box gene and functions upstream of *Early heading date 1* (*Ehd1*) (8). *Ehd1* encodes a B-type response regulator and acts as an activator of *Hd3a* independently from *Hd1* (19). No clear orthologues of *Ehd1* or *OsMADS51* are found in the *Arabidopsis* genome. Although the genetic pathway for short-day flowering in rice is relatively well understood, the molecular mechanisms generating the diversity of flowering time in cultivated rice are not known. In this study, we analyzed the expression and nucleotide sequences of genes involved in short-day flowering in rice. Our study revealed that allelic variation of *Hd1* is a main source of flowering time diversity in cultivated rice. We also found that diversity in *Ehd1* expression levels and *Hd3a* promoter sequences contributes to diversity in flowering time and the *Hd3a* expression level.

## Results

**Hd3a Expression Is Highly Correlated with Flowering Time in Cultivated Rice.** To analyze the molecular mechanisms controlling variation in flowering time of cultivated rice, we used a core collection of 64 cultivars (20). This collection was developed from an original set of 332 cultivars to study the molecular diversity present in cultivated rice, and it represents 91% of the genetic diversity present in the original cultivars, as determined by restriction fragment length polymorphism analysis. Flowering times of cultivars in the collection were highly diverse, ranging from 45 days to 153 days after sowing (Fig. 1*B*). To identify the genes that are the major determinants of this diversity, we first analyzed the mRNA levels of the 6 flowering time genes shown in Fig. 1*A* in the core collection cultivars and looked for possible relationships between their expression levels and flowering times. Expression levels of *OsGI*, *Hd1*, and *OsMADS51* were uniform among the cultivars, and little correlation with flowering time was found (Fig. 1 *C–E, I*). In contrast, expression levels of *Ehd1*, *Hd3a*, and *RFT1* were highly variable (Fig. 1 *F–H*). Among these genes, RNA levels of *Hd3a* were most closely correlated with flowering time (Fig. 1 *C–H*), indicating that higher *Hd3a* expression levels were associated with earlier flowering time. A similar correlation also was observed in a later stage of plant development [supporting information (SI) Fig. S1]. Expression levels of *Ehd1* correlated moderately with flowering time (Fig. 1 *F* and *I*). These results suggest that mRNA levels of *Hd3a* strongly affect the flowering time of cultivated rice and are consistent with previous results showing that Hd3a is a major floral activator in rice (9, 7, 18). Functional associations between the genes were detected also (Fig. 1*I*). The expression of *OsMADS51* was correlated with that of *OsGI*, consistent with the recent finding that *OsMADS51* expression is regulated by *OsGI* (8). Moreover, *Hd3a* expression was correlated with *Ehd1* expression, as shown previously (19). Thus, genetic variations that are able to create diversity in *Hd3a* mRNA levels are the major determinants of variation in flowering time in rice. Candidates for such factors are the regulatory region of *Hd3a* and variations in the functions of 2 genes that regulate *Hd3a* expression, *Hd1* and *Ehd1* (Fig. 1*A*).
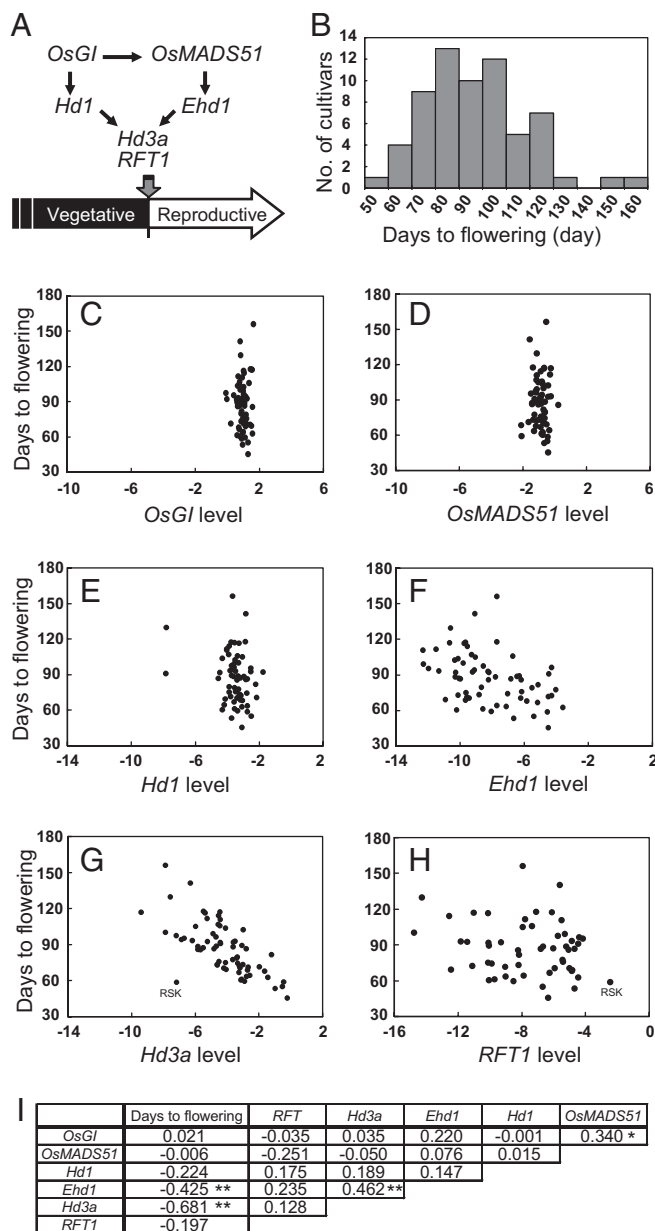
**Fig. 1.** Variation in flowering time and relationships between RNA levels of flowering time genes and flowering times among cultivated rice. (*A*) A schematic model of the short-day flowering pathway in rice. (*B*) Distribution of mean flowering time under short-day conditions (10 h light, 14 h dark) in the core collection. (*C–H*) Correlation of flowering time with RNA levels of (*C*) *OsGI*; (*D*) *OsMADS51*; (*E*) *Hd1*; (*F*) *Ehd1*; (*G*) *Hd3a*; (*H*) *RFT1*. RNA levels in leaves of 35-day-old plants were determined by real-time RT-PCR and are shown as natural logarithms. RSK, cv. Ryou Suisan Koumai. (*I*) Pearson's correlation coefficient between flowering time and gene expression level. ** indicates $P < 0.001$, and * indicates $P < 0.01$.

**Nucleotide Polymorphisms in the Promoter and Coding Regions of *Hd3a*.** First, we investigated the diversity in the regulatory region of *Hd3a* among the cultivars in core collection. The 2-kbp region of the *Hd3a* promoter previously was shown to be sufficient for proper regulation of *Hd3a* expression (7, 12). Therefore, nucleotide polymorphisms were analyzed over a 2-kb region including the promoter and the 5′ UTR region, and 7 types of sequences were identified (Fig. 2*A*). They were classified into 2 groups according to their sequences: Types 1 and 2 in Group A and Types 3–7 in Group B (Fig. 2*B*). Genes with Group B promoters

were expressed at significantly higher levels than those with Group A promoters, suggesting an association between promoter sequence types and expression levels (Fig. 2*C*). Thus, the *Hd3a* promoter sequence is a potential factor generating diversity in flowering time. We analyzed several potential *cis*-elements in the promoter. We found no nucleotide changes in the CCAAT box (21). The *Golden2*, *Arabidopsis RESPONSE REGULATOR* (*ARR*) and *Chlamydomonas* regulatory protein of P-starvation acclimatization response (*Psr1*) (GARP) DNA-binding domain of Ehd1 potentially can bind the *Hd3a* promoter region via the ARR1 binding element (19). Three potential ARR1 binding sites were found in the *Hd3a* promoter, but a nucleotide change was found in an element at −1932 bp. No changes in the other 2 ARR1 binding elements were detected (Fig. 2*A*). Thus, no major alterations in potential regulatory sequences of the *Hd3a* promoter were identified. An association between the promoter type and expression level (Fig. 2*C*) may be caused by other unidentified sequences or by other loci that are tightly linked with these polymorphic sites.

We found 6 types of *Hd3a* coding sequences in the core collection (Fig. 2*D*). No variations were found in the amino acids surrounding the potential ligand binding pocket or the external loop domain (residues 130–143), which is proposed to be important for FT protein function (22) (Fig. 2*D*). Therefore, we concluded that the Hd3a protein function is highly conserved among rice cultivars in the core collection.

**Nucleotide Polymorphisms in *Ehd1* and *OsMADS51*.** We examined possible diversity in the *Ehd1* coding region and found an amino acid substitution (G219R) in the GARP domain of a Chinese cultivar, DANYU (Type 2) (Fig. 2*E*). This Gly-to-Arg substitution was described previously and was shown to decrease the DNA binding activity of Ehd1 (19). In the receiver domain of Ehd1, 3 amino acid residues, $D_1$, $D_2$, and K, are known to be important for protein function (19, 23). In all the cultivars except DANYU, no amino acid substitutions that might affect protein function were found in either the receiver or the GARP domains. Thus, the Ehd1 protein function is highly conserved in cultivated rice.

Because mRNA levels of *Ehd1* were correlated moderately with those of *Hd3a* (Fig. 1 *F* and *I*), genes functioning upstream of *Ehd1* potentially could provide variations in *Ehd1* expression. Therefore, we also analyzed nucleotide diversity in *OsMADS51* and identified 3 allele types. We found 6 amino acid substitutions, 2 in the MADS domain and 4 in other regions of the protein. The Type 2 alleles also carried a frame-shift mutation at the C-terminal end (Fig. 2*F*). However, these allelic variations of *OsMADS51* were not correlated with *Ehd1* mRNA levels (Fig. S2). We also investigated nucleotide polymorphism in the *Ehd1* promoter regions. Analysis of more than 2 kbp in the upstream regions identified 9 types of sequences (Fig. S3). We analyzed the relationship between the allelic variation of *Ehd1* promoter regions and the *Ehd1* expression levels using Scheffé's F test, but we observed no significant difference.

**Nucleotide Polymorphism of *Hd1*.** In contrast to *Ehd1* and *OsMADS51*, we identified a high degree polymorphism in the *Hd1* sequences, some of which cause frame-shift mutations or create premature stop codons. The *Hd1* alleles in the core collection were grouped into 17 types, and 15 distinct proteins were identified (Fig. 3*A*). In addition to the previously reported 2-bp deletion found in cultivar (cv.) Kasalath (Fig. 3*A*, Type 13) (9), we discovered 3 deletions and 1 SNP that could cause defects in the entire or part of the CCT domains of 8 protein types (Fig. 3*B*). The CCT domain functions as a nuclear localization signal, and the lack of the CCT domain in the *Arabidopsis* CO causes a defect in the protein function (24). Thus, the Hd1 proteins encoded by these 9 allele types are likely to be nonfunctional.
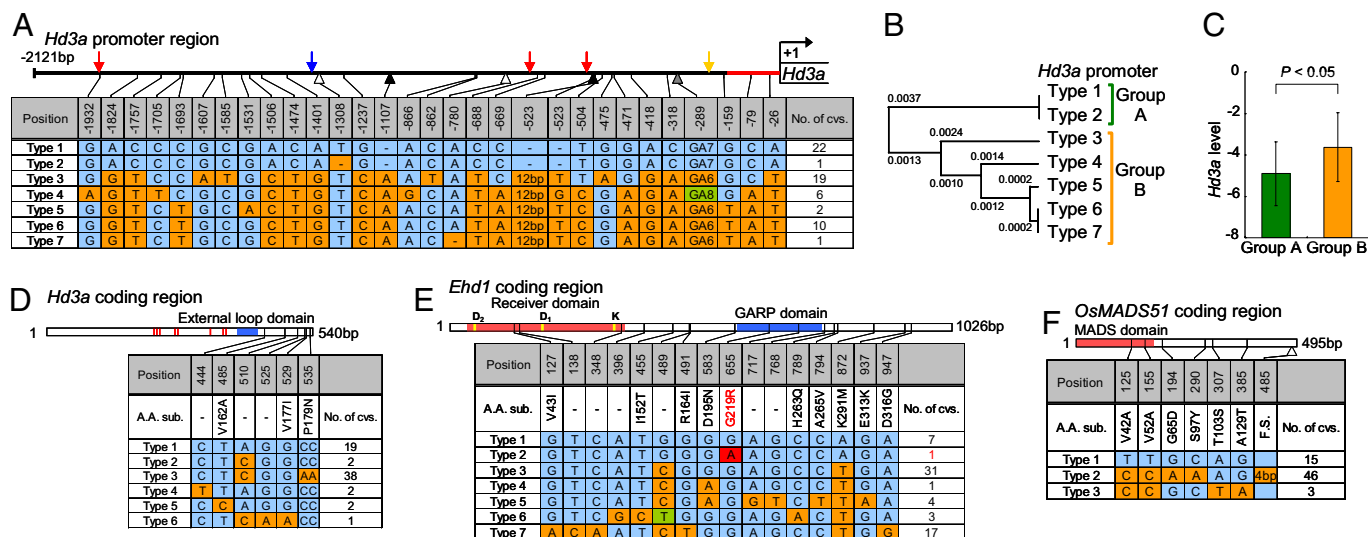
## Fig. 2 (A–F)

**A** *Hd3a* promoter region −2121bp ... +1 *Hd3a*

| Position | -1932 | -1824 | -1757 | -1705 | -1693 | -1607 | -1585 | -1531 | -1506 | -1474 | -1401 | -1308 | -1237 | -1107 | -866 | -862 | -780 | -688 | -669 | -523 | -523 | -504 | -475 | -471 | -418 | -318 | -289 | -159 | -79 | -26 | No. of cvs. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type 1 | G | A | C | C | C | G | C | G | A | C | A | T | G | - | A | C | A | C | C | - | - | T | G | G | A | C | GA7 | G | C | A | 22 |
| Type 2 | G | A | C | C | C | G | C | G | A | C | A | C | A | - | G | - | A | C | A | C | C | - | - | T | G | G | A | C | GA7 | G | C | A | 1 |
| Type 3 | G | G | T | C | C | A | T | G | C | T | G | T | C | A | A | T | A | T | C | 12bp | T | T | A | G | G | A | GA6 | G | C | T | 19 |
| Type 4 | A | G | T | T | C | G | C | G | C | T | G | T | C | A | G | C | A | T | A | 12bp | G | C | G | A | G | A | GA8 | G | A | T | 6 |
| Type 5 | G | G | T | C | T | G | C | G | C | T | G | T | C | A | A | C | A | T | A | 12bp | T | C | G | A | G | A | GA6 | T | A | T | 2 |
| Type 6 | G | G | T | C | T | G | C | G | C | T | G | T | C | A | A | C | A | T | A | 12bp | T | C | G | A | G | A | GA6 | T | A | T | 10 |
| Type 7 | G | G | T | C | T | G | C | G | C | T | G | T | C | A | A | C | - | T | A | 12bp | T | C | G | A | G | A | GA6 | T | A | T | 1 |

**B** *Hd3a* promoter — Phylogenetic tree
- 0.0037
- Type 1 ⟩ Group A
- Type 2 ⟩ Group A
- 0.0024
- 0.0013
- 0.0014
- 0.0010
- 0.0002
- Type 3 ⟩ Group B
- Type 4 ⟩ Group B
- 0.0012
- Type 5 ⟩ Group B
- 0.0002
- Type 6 ⟩ Group B
- Type 7 ⟩ Group B

**C** $P < 0.05$ — *Hd3a* level (Group A vs Group B), y-axis 0 to −8.

**D** *Hd3a* coding region — External loop domain, 1 ... 540bp

| Position | 444 | 485 | 510 | 525 | 529 | 535 | No. of cvs. |
|---|---|---|---|---|---|---|---|
| A.A. sub. | - | V162A | - | - | V177I | P179N | |
| Type 1 | C | T | A | G | G | CC | 19 |
| Type 2 | C | T | C | G | G | CC | 2 |
| Type 3 | C | T | C | G | G | AA | 38 |
| Type 4 | T | T | A | G | G | CC | 2 |
| Type 5 | C | C | A | G | G | CC | 2 |
| Type 6 | C | T | C | A | A | CC | 1 |

**E** *Ehd1* coding region — Receiver domain ($D_2$, $D_1$, K), GARP domain, 1 ... 1026bp

| Position | 127 | 138 | 348 | 396 | 455 | 489 | 491 | 583 | 655 | 717 | 768 | 789 | 794 | 872 | 937 | 947 | No. of cvs. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A.A. sub. | V43I | - | - | - | I152T | - | R164I | D195N | G219R | - | H263Q | A265V | K291M | - | E313K | D316G | |
| Type 1 | G | T | C | A | T | G | G | G | G | A | A | G | C | C | A | G | A | 7 |
| Type 2 | G | T | C | A | T | G | G | G | A | A | A | G | C | C | A | G | A | 1 |
| Type 3 | G | T | C | A | T | C | G | G | G | A | G | C | C | T | G | A | 31 |
| Type 4 | G | T | C | A | T | C | G | A | G | A | G | C | C | T | G | A | 1 |
| Type 5 | G | T | C | A | T | C | G | A | G | G | T | C | T | T | A | 4 |
| Type 6 | G | T | C | G | C | T | G | A | G | A | G | A | C | T | G | A | 3 |
| Type 7 | A | C | A | A | T | C | T | G | G | A | G | C | C | T | G | G | 17 |

**F** *OsMADS51* coding region — MADS domain, 1 ... 495bp

| Position | 125 | 155 | 194 | 290 | 307 | 385 | 485 | No. of cvs. |
|---|---|---|---|---|---|---|---|---|
| A.A. sub. | V42A | V52A | G65D | S97Y | T103S | A129T | F.S. | |
| Type 1 | T | T | G | C | A | | | 15 |
| Type 2 | C | C | A | A | A | G | 4bp | 46 |
| Type 3 | C | C | G | C | T | A | | 3 |

**Fig. 2.** Functions of *Hd3a*, *OsMADS51*, and *Ehd1* are conserved among cultivated rice. The nucleotide sequences of the 3 genes in the cultivars in the core collection were compared with those of Nipponbare. Polymorphic nucleotides are indicated by different colors. (*A*) Nucleotide polymorphisms in the *Hd3a* promoter region. Red arrows indicate ARR1 binding elements, the blue arrow indicates the CCAAT box, and the yellow arrow indicates the TATA box. Deletion and insertion sites are shown by open and closed arrowheads, respectively. A GA repeat sequence is indicated by gray arrowheads. The 5′-untranslated region is colored in red. (*B*) Phylogenetic tree of the *Hd3a* promoter created by the unweighted pair-group method using arithmetic average. Numbers indicate genetic distance. (*C*) *Hd3a* RNA levels in cultivars with Group A and Group B *Hd3a* promoters. RNA levels in leaves of 35-day-old plants were determined by real-time RT-PCR and are shown as natural logarithms. Error bars represent SD. (*D*) Nucleotide polymorphisms in the *Hd3a* coding region. Red lines indicate the locations of codons for amino acids that are essential for potential ligand binding pocket. The blue box indicates the external loop domain. (*E*) Nucleotide polymorphisms in the *Ehd1* coding sequence. Asp ($D_1$, $D_2$) and Lys (K) residues are indicated by yellow lines. A single amino acid substitution in the GARP domain, found in cv. DANYU, is shown in red. (*F*) Nucleotide polymorphisms in the *OsMADS 51* coding sequence. A deletion site is indicated by an open arrowhead.

Although all the other alleles contained intact zinc-finger and CCT domains, we found several SNPs and indels that could cause amino acid substitutions in other regions of the protein (Fig. 3*A*). Therefore, we determined whether these Hd1 proteins were functional using a previously described transient assay in rice protoplasts (7). Although *Hd1* activates *Hd3a* promoter activity under short-day conditions *in planta*, overexpression of functional *Hd1* suppresses *Hd3a* promoter activity in this transient assay (7). Protoplasts were transformed with the *Hd3a* promoter-luciferase construct along with 1 of the *Hd1* alleles, linked to the cauliflower mosaic virus (CaMV) 35S promoter, and we evaluated Hd1 function by measuring the suppression level of luciferase activity. In the reporter assay, the putatively functional *Hd1* alleles showed decreased luciferase activity when compared with the nonfunctional *Hd1* alleles (Fig. 3*C*), suggesting that these 8 types of *Hd1* alleles encode functional Hd1 proteins. In conclusion, we found that *Hd1* has a much higher level of polymorphism than the other genes examined in this study (Tables S1 and S2).

**Correlation of the Allelic Diversity in *Hd1* and Flowering Time.** To determine whether the allelic diversity of *Hd1* was correlated with variations in the flowering times of cultivated rice, we grouped all the cultivars into those with functional and nonfunctional *Hd1* alleles and compared their *Hd3a* mRNA levels with flowering times (Fig. 3*D*). Cultivars that carry functional *Hd1* alleles tended to show higher *Hd3a* expression levels and earlier flowering times, whereas those carrying nonfunctional *Hd1* alleles tended to show lower *Hd3a* expression levels and later flowering times (Fig. 3*D*). The differences in *Hd3a* expression levels and flowering times between the functional and nonfunctional *Hd1* alleles were statistically significant by a student's *t* test. These results suggest that the nucleotide polymorphisms observed in *Hd1* might be one of the main causes for the diversity of flowering time in cultivated rice. Cultivars with fu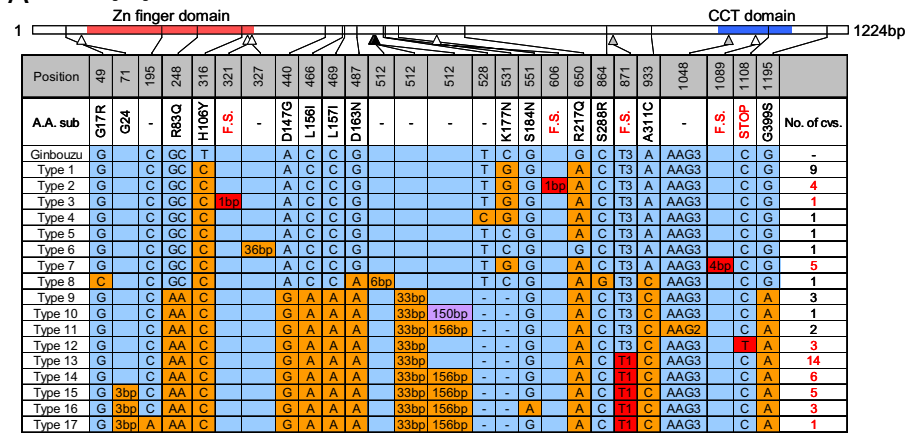nctional and nonfunctional *Hd1* alleles were mapped to various regions of Asia (Fig. 3*E*). Results indicate that both groups of alleles are widely distributed, indicating that they have spread to many regions during the last 8,000 to 10,000 years.

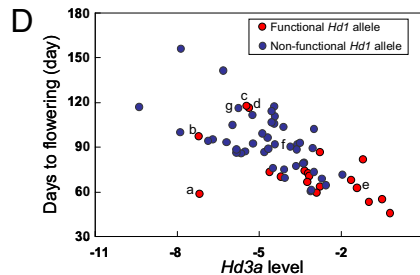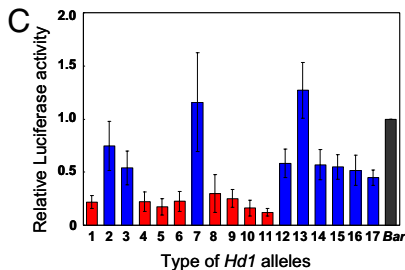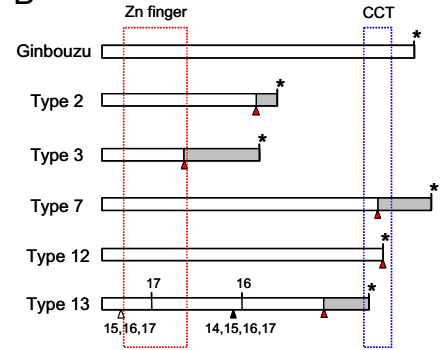**Other Factors Contributing to the Diversity in Flowering Time of Rice.** Although we found significant differences in *Hd3a* expression levels and flowering times between cultivars carrying functional and nonfunctional *Hd1* alleles, we also detected variations in each group, suggesting that other factors also may contribute to diversity in *Hd3a* expression levels and flowering times. We found that *Ehd1* expression levels showed a moderate correlation with *Hd3a* expression level ($r = 0.462$, $P < 0.001$), and this correlation was little affected by the allelic variations of *Hd1* (Table S3). *Ehd1* was shown to induce *Hd3a* expression independently from *Hd1* (19). Therefore, variation in *Ehd1* expression levels also may contribute to the diversity of *Hd3a* expression levels and flowering times. This idea also is supported by the distribution of cv. NIPPONBARE (*Hd1*, *Ehd1*), *hd1* mutant (*hd1*, *Ehd1*), and cv. Taicung 65 (*hd1*, *ehd1*) in Fig. 3*D*.

**Linear Model Analysis of Flowering Time and *Hd3a* Expression Level in Cultivated Rice.** The relative contribution of various genetic factors to the phenotype variation in flowering time was assessed by applying a simple linear model that incorporates 3 genetic factors: *Hd1* functionality, *Hd3a* promoter type, and *Ehd1* expression level. This analysis revealed that the independent contributions of these genetic factors and the interactions between them account for 26.6% of the entire variance. The relative contributions of *Hd1*, *Hd3a*, and *Ehd1* comprise 44.8%, 16.7%, and 38.5% of the genetic variance, respectively, indicating that the largest single contributor to flowering time variation is the difference between functional and nonfunctional *Hd1* alleles. When phenotype was assessed by the expression level of *Hd3a*, 44.8% of the phenotypic variance could be explained by the 3 genetic factors; their relative contributions were 24.7%, 48.3%, and 27.0%, respectively, suggesting that variation in the

**A** *Hd1* coding region

Zn finger domain — CCT domain — 1 ... 1224bp

| Position | 49 | 71 | 195 | 248 | 316 | 321 | 327 | 440 | 466 | 469 | 487 | 512 | 512 | 512 | 528 | 531 | 551 | 606 | 650 | 864 | 871 | 933 | 1048 | 1089 | 1108 | 1195 | No. of cvs. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A.A. sub | G17R | G24 | – | R83Q | H106Y | F.S. | – | D147G | L156I | L157I | D163N | – | – | – | – | K177N | S184N | F.S. | R217Q | S288R | F.S. | A311C | – | F.S. | STOP | G399S | |
| Ginbouzu | G | C | GC | T | | | | A | C | C | G | | | | | T | C | G | G | C | T3 | A | AAG3 | | C | G | – |
| Type 1 | G | C | GC | C | | | | A | C | C | G | | | | | T | G | G | A | C | T3 | A | AAG3 | | C | G | 9 |
| Type 2 | G | C | GC | C | | | | A | C | C | G | | | | | T | G | G | 1bp | A | C | T3 | A | AAG3 | C | G | 4 |
| Type 3 | G | C | GC | C | | 1bp | | A | C | C | G | | | | | T | G | G | A | C | T3 | A | AAG3 | | C | G | 1 |
| Type 4 | G | C | GC | C | | | | A | C | C | G | | | | | C | G | G | A | C | T3 | A | AAG3 | | C | G | 1 |
| Type 5 | G | C | GC | C | | | | A | C | C | G | | | | | T | C | G | A | C | T3 | A | AAG3 | | C | G | 1 |
| Type 6 | G | C | GC | C | | | 36bp | A | C | C | G | | | | | T | C | G | G | C | T3 | A | AAG3 | | C | G | 1 |
| Type 7 | G | C | GC | C | | | | A | C | C | G | | | | | T | G | G | A | C | T3 | A | AAG3 | 4bp | C | G | 5 |
| Type 8 | C | C | GC | C | | | | A | C | C | A | 6bp | | | | T | C | G | A | G | T3 | C | AAG3 | | C | G | 1 |
| Type 9 | G | C | AA | C | | | | G | A | A | A | 33bp | | | | – | – | – | A | C | T3 | A | AAG3 | | C | A | 3 |
| Type 10 | G | C | AA | C | | | | G | A | A | A | 33bp | 150bp | | | – | – | – | A | C | T3 | A | AAG3 | | C | A | 1 |
| Type 11 | G | C | AA | C | | | | G | A | A | A | 33bp | 156bp | | | – | – | – | A | C | T3 | A | AAG2 | | C | A | 2 |
| Type 12 | G | C | AA | C | | | | G | A | A | A | 33bp | | | | – | – | – | A | C | T1 | A | AAG3 | | T | A | 3 |
| Type 13 | G | C | AA | C | | | | G | A | A | A | 33bp | | | | – | – | – | A | C | T1 | A | AAG3 | | C | A | 14 |
| Type 14 | G | C | AA | C | | | | G | A | A | A | 33bp | 156bp | | | – | – | – | A | C | T1 | A | AAG3 | | C | A | 6 |
| Type 15 | G 3bp | C | AA | C | | | | G | A | A | A | 33bp | 156bp | | | – | – | – | A | C | T1 | A | AAG3 | | C | A | 5 |
| Type 16 | G 3bp | C | AA | C | | | | G | A | A | A | 33bp | 156bp | | | – | A | – | A | C | T1 | A | AAG3 | | C | A | 3 |
| Type 17 | G 3bp | A | AA | C | | | | G | A | A | A | 33bp | 156bp | | | – | – | – | A | C | T1 | A | AAG3 | | C | A | 1 |

**B** Zn finger — CCT. Ginbouzu, Type 2, Type 3, Type 7, Type 12, Type 13 (17, 16; 15,16,17; 14,15,16,17). *, premature stop codon.

**C** Relative Luciferase activity (0–2.0) vs Type of *Hd1* alleles (1–17, *Bar*).

**D** Days to flowering (day) (30–180) vs *Hd3a* level (−11 to 1). Legend: red = Functional *Hd1* allele; blue = Non-functional *Hd1* allele. Points labeled a–g.

**E** Geographical distribution map (latitudes 40°, 20°, 0°). Legend: red = Functional *Hd1* allele; blue = Non-functional *Hd1* allele; gray = Not determined.
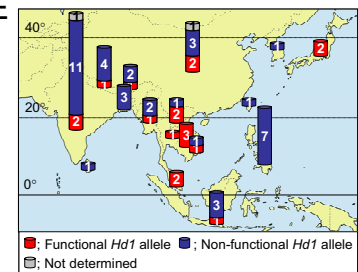
**Fig. 3.** A high degree of polymorphism in the *Hd1* coding sequence and its relationship with flowering time. (*A*) The *Hd1* nucleotide sequences of the cultivars in the core collection were compared with that of cv. Ginbouzu. Polymorphic nucleotides are indicated by different colors. Deletion and insertion sites are indicated by open and closed arrowheads, respectively. The number of cultivars with each type of sequence (Type 1–17) is shown in the column at the right, with the numbers for loss-of-function types in red. Types 1 and 4 are identical, and Types 15 and 17 are identical. F.S., frame shift. (*B*) Structures of loss-of-function Hd1 proteins that lack the entire or part of the CCT domain. The locations of amino acid substitutions that cause loss of protein function are shown by red arrowheads. The open and closed arrowheads indicate deletion and insertion sites, respectively. The numbers below the arrowheads indicate which sequence types have the substitutions. Gray boxes indicate altered amino acid sequences caused by frame shifts. *, premature stop codon. (*C*) Activities of *Hd1* alleles, analyzed by transient luciferase reporter assays using rice protoplasts. Red and blue bars indicate functional and nonfunctional *Hd1* alleles, respectively. The *Bar* gene was used as a vector control. Error bars represent SD; *n* = 12–16. (*D*) Correlation of flowering times with *Hd3a* RNA levels in cultivars that carry functional *Hd1* alleles (*red*) and nonfunctional *Hd1* alleles (*blue*). RNA levels in leaves of 35-day-old plants were determined by real-time RT-PCR and are shown as natural logarithms. a, cv. RYOU SUISAN KOUMAI; b, cv. KHAU TAN CHIEM; c, cv. KEMASIN; d, cv. BINGALA; e, NIPPONBARE; f, *hd1* mutant (NIPPONBARE); g, cv. Taichung 65. (*E*) Geographical distribution of functional (*red*) and nonfunctional (*blue*) *Hd1* alleles. Alleles that were not sequenced are shown by gray columns. The number within each bar indicates the number of cultivars belonging to each *Hd1* allelic group.

*Hd3a* promoter contributes significantly more than the other 2 factors.

These observed differences in the contribution of each component to flowering time and *Hd3a* expression levels may indicate that other factors, such as posttranscriptional regulation of *Hd3a*, also are involved in the generation of flowering time diversity. It should be noted that these statistical analyses were performed using only Indica cultivars (*n* = 45) because of the small sample size available for Japonica cultivars (*n* = 13). We did not pool Indica and Japonica, to avoid any potential bias resulting from the Indica-Japonica population structure.

## Discussion

In this study, we identified 3 factors that generate diversity in flowering time. Moreover, the contributions of these factors were calculated based on linear model analysis. Although we observed a strong correlation between flowering time and *Hd3a* expression level, we found differences in the relative contributions of the 3 genetic factors (the functionality of *Hd1*, the promoter type of *Hd3a*, and the expression level of *Ehd1*) to flowering time and *Hd3a* expression level. A possible explanation of these differences is that the variation in *Hd3a* promoter type directly affects the variation in *Hd3a* expression levels, whereas Hd1 and the expression level of *Ehd1* could

affect flowering time through other unknown factors. *RFT1*, for example, an orthologue of *Hd3a* that has been shown to affect flowering in the absence of functional *Hd3a*, would be an attractive subject for further studies on variations in promoter and protein-coding sequences (18). The current results support the genetic model for flowering time regulation in rice (Fig. 1*A*).

Our results also indicate that other, currently unknown, factors influence flowering time variation in cultivated rice. Several indirect factors such as growth rate, edaphic condition preference, and temperature may be among these unexplained factors. Variation in the activity of *Hd1* alleles might be a factor. We performed a luciferase reporter assay to evaluate the activity of each Hd1 protein (Fig. 3*C*). Although we were able to distinguish functional proteins from nonfunctional proteins, it was difficult to demonstrate accurately differences in activity among Hd1 proteins. Variation of *Hd3a* expression level in each *Hd1* allele could not be explained by the results of the luciferase reporter assay (Fig. 3*C*, Fig. S4), perhaps because of the complex regulation of *Hd3a* expression in rice (7). Allelic variations in the *OsMADS51* and *Ehd1* promoter regions could not explain the variation in *Ehd1* expression levels (Fig. 2*C* and Fig. S2 and S3). Recently, *RID1/Ehd2/OsId1*, a rice orthologue of the maize *ID1*, was reported as a regulator of *Ehd1* (25–27). Maize ID1 is a

transcription factor that can bind to the specific DNA sequence TTTTGTCG/C (28). However, this consensus sequence is not found in the *Ehd1* promoter region, and whether *RID1/Ehd2/OsId1* is a direct target of *Ehd1* promoter remains to be studied. Factors functioning upstream of *Ehd1* that could generate variations in *Ehd1* mRNA levels need to be identified in future work.

We found some exceptions in the correlation between *Hd3a* expression and flowering time. In particular, the Chinese cv. Ryou Suisan Koumai (RSK) showed significantly low *Hd3a* expression, but it flowered very early (58.5 days) (Fig. 1*G*). We found that *RFT1* expression in RSK was much higher than in other cultivars (Figs. 1*H* and Fig. S5), suggesting that *RFT1* may complement the low-level *Hd3a* expression. It may be possible to discover novel genes for flowering time regulation in rice, by studying this cultivar. We also found several exceptions: KHAU TAN CHIEM, KEMASIN, and BINGARA have a functional *Hd1* allele but show low *Hd3a* levels and a late-flowering phenotype. However, because we did not find any anomalous values in the *Hd1* and *Ehd1* expression levels, other unknown factors that affect regulation of *Hd3a* expression by Hd1 proteins may be involved.

Natural variation in flowering time has been studied extensively in *Arabidopsis*. These studies have demonstrated that the major determinants of flowering time diversity are present in the vernalization pathway, which is a period of cold temperature that is required to induce flowering. *FLOWERING LOCUS C* (*FLC*), a flowering suppressor gene encoding a MADS-box transcription factor, and *FRIGIDA* (*FRI*), a positive regulator of *FLC* transcription, are the major factors controlling natural variation in flowering time in *Arabidopsis* (29–36).

Unlike *Arabidopsis*, rice is a tropical crop and does not show a vernalization response. Interestingly, no convincing orthologues of *FLC* and *FRI* were found in the rice genome (37). Temperate cereals such as wheat and barley also show vernalization responses; however, the genes involved in these responses are not conserved between *Arabidopsis* and cereals (38). The lack of homologues of *Arabidopsis FLC* in cereals may be consistent with these observations. The existence of variations in the *Ppd-H1* gene, involved in the photoperiod pathway, among European barley cultivars has been described (38), but variations in the *CO/Hd1* orthologues *TaCO* and *HvCO* in wheat and barley have not been reported (39, 40). However, the molecular genetic pathways controlling flowering time in wheat and barley have not yet been elucidated clearly, and the major determinants of flowering time in these species remain to be identified. Nevertheless, it appears that crops and plant species can adopt unique strategies to generate variations in flowering time and that the strategy used is one of the most important characters in the life of plants.

## Materials and Methods

**Plant Materials and Growth Conditions.** The rice core collection and *Tos17*-induced mutant of *Hd1* (NG6082, Nipponbare) were obtained from the National Institute of Agrobiological Science GenBank (20, 41). The plants were grown in climate chambers under short-day conditions with daily cycles of 10 h of light at 30 °C and 14 h of dark at 25 °C. Light was provided by fluorescent white light (400–700 nm, 100 $\mu$mol m$^{-2}$ s$^{-1}$) at 70% humidity. Flowering time was defined as the time when the first panicle appeared from the node. Four to 8 independent plants were used to score flowering time.

**Gene Expression Analysis.** Leaves were harvested from plants grown under short-day conditions, 35 days after germination at the zeitgeber times (ZT) corresponding to the expression peak for each gene, as follows: ZT 0 (*Hd3a* and *RFT*), ZT 8 (*OsGI*, *OsMADS51*), ZT 16 (*Hd1*), and ZT 20 (*Ehd1*). Leaf tissues were ground in liquid nitrogen with the ShakeMaster Auto ver.2.0 (BioMedical Science Inc.). Total RNA was extracted using TRIzol (Invitrogen) and treated with DNase I (Invitrogen). cDNA (20 $\mu$l) was synthesized from 1 $\mu$g of total RNA

using SuperScriptII Reverse Transcriptase (Invitrogen). One $\mu$l of cDNA was used for the real-time PCR performed with the SYBR Green PCR master mix (Applied Biosystems). Data were collected using the ABI PRISM 7000 sequence detection system. All expression levels were normalized by that of *ubiquitin*. Four to 8 independent plants were used for this assay. Displayed data represent means of 2 separate RNA extractions.

**Sequence Analysis.** Intron and promoter regions of each gene were amplified using LA Taq or PrimeSTAR HS DNA polymerase (Takara) from genomic DNA, and PCR products were purified by the MinElute system (Qiagen). These purified PCR fragments were used as templates, and sequencing reactions were performed using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Primers used for this analysis are listed in Table S4. Data were collected using the ABI PRISM 3100 Genetic Analyzer. The data were analyzed using the GENETYX program ver.7 (Genetyx).

**Constructs for Luciferase Reporter Assay.** The *Hd1* coding sequences were cloned from the cultivars in core collection (Table S2). These fragments were amplified by PCR with PrimeSTAR HS DNA polymerase (Takara). *Hd1* fragments were subcloned into pENTR/D-TOPO vector by using the Directional TOPO cloning kit (Invitrogen). Subsequently, the *Hd1* insert of pENTR plasmids was transferred to CaMV 35S::Gateway cassette plasmid, which contained pUC12 backbone, by performing an LR recombination reaction (Invitrogen). The *Hd3a* promoter sequence was cloned from cv. NIPPONBARE. This fragment was subcloned to pUC12 with *pLUC* (*Photinus pyralis*) plasmid using Ligation-Convenience Kit (Nippon Gene). Primers used for this experiment are listed in Table S4.

**Luciferase Reporter Assays.** Rice protoplasts (2 × 10$^5$ cells) were prepared from rice Oc cell cultures and transformed with 1 $\mu$g of *pHd3a::pLUC* (*Photinus pyralis*) as a reporter and 5 $\mu$g of each *Hd1* construct, containing the *Hd1* coding sequence driven by the CaMV 35S promoter using PEG transfection method (42). For a vector control, the bialaphos-resistance gene (*Bar*) was used instead of the *Hd1* coding sequence. After incubation for 12 h, luciferase activities were measured using a Luciferase Assay System (Promega). Luciferase luminescence of each *Hd1* allele was normalized for sample protein level and evaluated as a proportion of that of the vector control.

**Statistical Analysis.** Statistical analyses were performed using Microsoft Excel 2003 or Statcel 2 (OMS). Correlation between flowering time and gene expression level was examined by Pearson's correlation coefficient test. The statistical differences in flowering times and *Hd3a* expression levels between plants with functional and nonfunctional *Hd1* alleles were determined by the 2-tailed student's *t* test, and differences with $P < 0.01$ were considered to be significant. The effect of each haplotype on flowering time and gene expression was examined by ANOVA or Kruskal-Wallis test. If these analyses were significant, post hoc pairwise comparisons (Tukey-Kramer test or Scheffé's F test) were conducted, with the level of statistical significance taken as $P < 0.01$.

**Linear Model Analysis.** To evaluate the relative contribution of the allelic diversity of *Hd1* to the phenotypic variation in flowering time, we applied a simple linear model in which the phenotype of an individual is assumed to be given by $y = G + E + e$, where $e$ signifies biological and technical replicates, and $E$ represents all other genetic and environmental factors. $G$ includes genetic variation caused by the functionality of *Hd1*, the promoter type of *Hd3a*, the expression level of *Ehd1*, and their interactions. We confirmed that there is almost no genetic linkage disequilibrium (LD) among these 3 loci. Although *Hd3* and *Hd1* are located on the same chromosome, the average LD of SNPs between the 2 regions is $r^2 = 0.03$, which is even smaller than the average LD between different chromosomes ($r^2 = 0.06$). However, we assumed a model in which the 3 interact because they belong to the same pathway. The same linear function can be applied using the expression level of *Hd3a* as the phenotype $y$. The statistical analysis was performed using R software.

1. Khush GS (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* 35:25–34.
2. Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127:1309–1321.
3. Hayama R, Coupland G (2004) The molecular basis of diversity in the photoperiodic flowering responses of *Arabidopsis* and rice. *Plant Physiol* 135:677–684.
4. Izawa T (2007) Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice. *J Exp Bot* 58:3091–3097.
5. Tsuji H, Tamaki S, Komiya R, Shimamoto K (2008) Florigen and the photoperiodic control of flowering in rice. *Rice* 1:25–35.
6. Hayama R, Izawa T, Shimamoto K (2002) Isolation of rice genes possibly involved in the photoperiodic control of flowering by a fluorescent differential display method. *Plant Cell Physiol* 43:494–504.
7. Hayama R, et al. (2003) Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* 422:719–722.
8. Kim SL, et al. (2007) OsMADS51 is a short-day flowering promoter that functions upstream of Ehd1, OsMADS14, and Hd3a. *Plant Physiol* 145:1484–1494.
9. Yano 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* 12:2473–2484.
10. Kojima S, et al. (2002) Hd3a, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant Cell Physiol* 43:1096–1105.
11. Corbesier L, et al. (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* 316:1030–1033.
12. Tamaki S, et al. (2007) Hd3a protein is a mobile flowering signal in rice. *Science* 316:1033–1036.
13. Lin MK, et al. (2007) FLOWERING LOCUS T protein may act as the long-distance florigenic signal in the cucurbits. *Plant Cell* 19:1488–1506.
14. Jaeger KE, Wigge PA (2007) FT protein acts as a long-range signal in *Arabidopsis*. *Curr Biol* 17:1050–1054.
15. Mathieu J, Warthmann N, Kuttner F, Schmid M (2007) Export of FT protein from phloem companion cells is sufficient for floral induction in *Arabidopsis*. *Curr Biol* 17:1055–1060.
16. Kobayashi Y, Weigel D (2007) Move on up, it's time for change—mobile signals controlling photoperiod-dependent flowering. *Genes Dev* 21:2371–2384.
17. Izawa T, et al. (2002) Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice. *Genes Dev* 16:2006–2020.
18. Komiya R, et al. (2008) Hd3a and RFT1 are essential for flowering in rice. *Development* 135:767–774.
19. Doi K, 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.
20. Kojima Y, et al. (2005) Development of an RFLP-based rice diversity research set of germplasm. *Breed Sci* 55:431–440.
21. Wenkel S, et al. (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell* 18:2971–2984.
22. Ahn JH, et al. (2006) A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *EMBO J* 25:605–614.
23. Sakai H, Aoyama T, Oka A (2000) *Arabidopsis* ARR1 and ARR2 response regulators operate as transcriptional activators. *Plant J* 24:703–711.
24. Robson F, et al. (2001) Functional importance of conserved domains in the flowering-time gene *CONSTANS* demonstrated by analysis of mutant alleles and transgenic plants. *Plant J* 28:619–631.
25. Wu C, et al. (2008) RID1, encoding a Cys2/His2-type zinc finger transcription factor, acts as a master switch from vegetative to floral development in rice. *Proc Natl Acad Sci USA* 105:12915–12920.
26. Matsubara K, et al. (2008) Ehd2, a rice ortholog of the maize *INDETERMINATE1* gene, promotes flowering by up-regulating Ehd1. *Plant Physiol* 148:1425–1435.
27. Park SJ, et al. (2008) Rice Indeterminate 1 (OsId1) is necessary for the expression of Ehd1 (Early heading date 1) regardless of photoperiod. *Plant J* 56:1018–1029.
28. Kozaki A, Hake S, Colasanti J (2004) The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. *Nucleic Acids Res* 32:1710–1720.
29. Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11:949–956.
30. Johanson U, et al. (2000) Molecular analysis of FRIGIDA, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* 290:344–347.
31. Stinchcombe JR, et al. (2004) A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proc Natl Acad Sci USA* 101:4712–4717.
32. Caicedo AL, et al. (2004) Epistatic interaction between *Arabidopsis* FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *Proc Natl Acad Sci USA* 101:15670–15675.
33. Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu Rev Plant Biol* 55:141–172.
34. Shindo C, et al. (2005) Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. *Plant Physiol* 138:1163–1173.
35. Lempe J, et al. (2005) Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLoS Genetics* 1:109–118.
36. Toomajian C, et al. (2006) A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. *PLoS Biol* 4:e137.
37. Tadege M, et al. (2003) Reciprocal control of flowering time by OsSOC1 in transgenic *Arabidopsis* and by FLC in transgenic rice. *Plant Biotechnol J* 1:336–369.
38. Cockram J, et al. (2007) Control of flowering time in temperate cereals: Genes, domestication, and sustainable productivity. *J Exp Bot* 58:1231–1244.
39. Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of *CONSTANS*-like gene families in barley, rice, and *Arabidopsis*. *Plant Physiol* 131:1855–1867.
40. Nemoto Y, et al. (2003) Characterization and functional analysis of three wheat genes with homology to the *CONSTANS* flowering time gene in transgenic rice. *Plant J* 36:82–93.
41. Hirochika H (2001) Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr Opin Plant Biol* 4:118–122.
42. Chen S, et al. (2006) A highly efficient transient protoplast system for analyzing defence gene expression and protein-protein interactions in rice. *Mol Plant Pathol* 7:417–427.