nature genetics

Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice

Weiya Xue[1,2], Yongzhong Xing[1,2], Xiaoyu Weng[1], Yu Zhao[1], Weijiang Tang[1], Lei Wang[1], Hongju Zhou[1], Sibin Yu[1], Caiguo Xu[1], Xianghua Li[1] & Qifa Zhang[1]

Yield potential, plant height and heading date are three classes of traits that determine the productivity of many crop plants. Here we show that the quantitative trait locus (QTL) *Ghd7*, isolated from an elite rice hybrid and encoding a CCT domain protein, has major effects on an array of traits in rice, including number of grains per panicle, plant height and heading date. Enhanced expression of *Ghd7* under long-day conditions delays heading and increases plant height and panicle size. Natural mutants with reduced function enable rice to be cultivated in temperate and cooler regions. Thus, *Ghd7* has played crucial roles for increasing productivity and adaptability of rice globally.

Cereal grains are the primary source of food and animal feed globally. Increasing grain yield of cereals has been among the most important goals of crop production and plant science research. To achieve high yield, plants require not only high yielding potential, which is defined by yield component traits, but also an optimal architecture that, to a large extent, is determined by plant height. Moreover, flowering time (or heading date) is a critical determinant enabling crops to adapt to seasonal changes and make maximum use of the temperature and sunlight that are available under specific ecological conditions.

Yield, plant height and heading date are three classes of traits that have been extensively investigated, and hundreds of QTLs have been reported for these traits in the literature (see URLs section in Methods). Yield per plant in rice (*Oryza sativa* L.) is multiplicatively determined by three component traits: number of tillers (panicles) per plant, number of grains per panicle and grain weight. Genes controlling these traits have been recently cloned[1–4]; each of these genes clearly has a distinct biological function. Molecular cloning and functional analyses of several genes associated with plant height in rice and wheat have shown that these genes are mostly related to the synthesis and regulation of the phytohormone gibberellin[5–8]. Extensive studies in *Arabidopsis thaliana* have established that flowering time is controlled by multiple pathways[9], of which the photoperiod pathway acts as a major one. A key component functioning in the photoperiod pathway is the nuclear protein CONSTANS (CO), which promotes flowering under long-day conditions. One of the downstream targets of CO is a protein encoded by the flowering locus T (*FT*), which, together with a bZIP protein FD, controls flowering time by integrating input from various pathways[10–12]. Rice orthologs of both *CO* (*Hd1*) and *FT* (*Hd3a*) have been identified[13–16], and these seem to have parallel functions in regulating flowering time, except that they promote flowering under short-day rather than long-day conditions. In addition, two other genes regulating heading date, *Hd6*, encoding the α-subunit of protein kinase CK2 (ref. 17), and *Ehd1*, encoding a B-type response regulator[18], have also been identified.

Here we show that a gene encoding a CCT (CO, CO-LIKE and TIMING OF CAB1) domain protein functions as a key QTL simultaneously controlling yield, plant height and heading date in rice.

RESULTS

A genomic segment with large effects on multiple traits

Previous studies consistently detected QTLs for number of grains per panicle, plant height and heading date in the interval defined by the two molecular markers R1440 and C1023 on chromosome 7 (R1440-C1023; **Supplementary Fig. 1a** online), using both $F_{2:3}$ and recombinant inbred line (RIL) populations derived from a cross between Zhenshan 97 and Minghui 63, the parents of Shanyou 63, the most widely cultivated rice hybrid in China[19–24]. At all these QTLs, the alleles from Minghui 63 simultaneously increased number of grains per panicle, plant height and number of days to heading. RI50, a RIL containing the chromosome segment R1440-C1023 from Minghui 63 and 74% of the genetic background from Zhenshan 97 from a population of 240 RILs, was backcrossed twice to Zhenshan 97. Self-pollination of the BC_2F_1 plants heterozygous for this fragment resulted in a pair of near-isogenic lines (NILs), designated NIL(zs7) and NIL(mh7), and a heterozygote, NIL(het), with almost all the genetic background of Zhenshan 97 except the introgressed segment. Field examination under natural long-day conditions (day length >14 h) showed that the three genotypes differed significantly in an array of traits (**Fig. 1**, **Table 1** and **Supplementary Tables 1** and **2** online), including heading date, plant height and number of spikelets

[1]National Key Laboratory of Crop Genetic Improvement, National Centre of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China. [2]These authors contributed equally to this work. Correspondence should be addressed to Q.Z. (qifazh@mail.hzau.edu.cn).

Received 25 October 2007; accepted 4 March 2008; published online 4 May 2008; doi:10.1038/ng.143

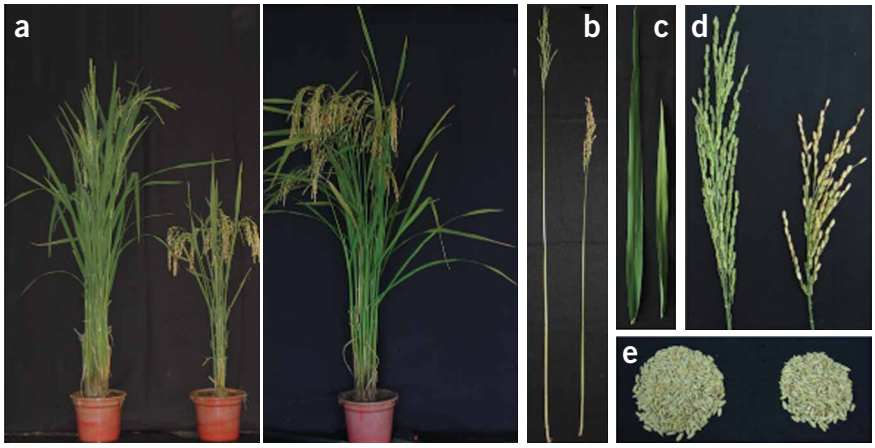


Figure 1 Performance of Zhenshan 97 and NIL(mh7) planted under natural long-day conditions. (**a**) Photos of Zhenshan 97 (middle) and NIL(mh7) (left) taken when Zhenshan 97 reached maturity, and photo of NIL(mh7) taken at maturity (right). (**b**) Main culms of NIL(mh7) (left) and Zhenshan 97 (right). (**c**) Flag leaves of NIL(mh7) (left) and Zhenshan 97 (right). (**d**) Main panicles of NIL(mh7) (left) and Zhenshan 97 (right). (**e**) Grains from the whole plants of NIL(mh7) (left) and Zhenshan 97 (right).

per panicle. Homozygous addition of R1440-C1023 to the Zhenshan 97 genetic background delayed heading by 21.2 days (33.3%), increased plant height by 33.3 cm (40%) and produced 86 (65.8%) more spikelets on the main culm (**Fig. 1** and **Table 1**). Further examination showed that NIL(mh7) out-yielded NIL(zs7) by 8.9 g (50.9%) per plant, primarily because of an increase in number of grains per panicle, although NIL(mh7) had fewer tillers per plant than NIL(zs7) (**Supplementary Table 1**).

An analysis of a BC_3F_2 population of 192 individuals derived by self-pollination of the BC_3F_1 heterozygotes showed that each of the three traits segregated in a single-locus mendelian ratio, and this was confirmed by data from BC_3F_3 families (**Supplementary Fig. 2** and **Supplementary Table 3** online). Large dominant effects were observed in all three traits, such that the heterozygotes were close to the performance of the higher parent. However, the differences diminished when the plants were grown under short-day conditions. We did not detect any significant differences among the three genotypes when we exposed the plants to natural light in the summer for 8 h in the day followed by a 16 h dark period from 33 d after germination till heading (data not shown).

Map-based cloning of *Ghd7*

We field planted the BC_2F_2 population consisting of 8,400 individuals (population 1) under natural long-day conditions in 2003 to examine heading, plant height and grain number. No recombination among the three traits was detected in this population. We self-pollinated plants heterozygous for the introgressed fragment and obtained 30,000 BC_2F_3 plants (population 2), which we planted in the rice-growing season of 2004. We also planted a BC_4F_2 population (population 3), which consisted of 6,240 individuals from a cross between Zhenshan 97 and ZS(tq7), a NIL that has the Zhenshan 97 genetic background except the R1440-C1023 fragment from Teqing that is a high-yielding cultivar, differing from Zhenshan 97 in an array of traits including grain number, heading and plant height (**Supplementary Table 4** online). Again, no recombinant among the three traits was detected in these latter two populations. Given the observation that short-day treatment simultaneously eliminated the phenotypic differences in all three traits among the three genotypes, we inferred that the three traits were controlled by a single locus, hereafter referred to as *Ghd7* (for grain number, plant height and heading date). For ease of accurate scoring, we also used number of spikelets as a measurement for grain number in trait examination.

To precisely map the *Ghd7* locus, we assayed 1,082 plants showing the recessive phenotype for all the three traits (short, early heading and small panicle) from population 1 with two simple sequence repeat markers, RM5451 and RM1135, bracketing the *Ghd7*-containing region (**Supplementary Fig. 1b**). We identified 66 and 94 recombinant individuals, respectively. By further genotyping of the recombinant plants using markers RM3859, RM5436, C39 and RM7110, we localized the *Ghd7* locus to the interval between RM3859 and C39, cosegregating with RM5436 (**Supplementary Fig. 1b**). An assay of 800 plants showing recessive phenotypes from population 3 detected one recombination event between RM5436 and the *Ghd7* locus, and four recombination events between C39 and *Ghd7*, thus delimiting the locus to an interval of 0.31 cM (**Supplementary Fig. 1c**). Mapping the two markers RM5436 and C39 to the finished sequence of Nipponbare[25], we found that this 0.31-cM region corresponded to a DNA fragment of approximately 2,284 kb in the centromeric region of chromosome 7, giving rise to a recombination rate of 7,368 kb per cM. This recombination rate is about 37-fold lower than the genome average of approximately 200 kb per cM.

According to the Rice Annotation Project Database, there are more than 450 predicted genes in this 2,284-kb region. We scanned the predicted proteins and found that one of the genes encodes a protein with homology to the CCT domain (see below), present in the *CO*, *CO*-like (*COL*) and *TOC1* gene products, which have been extensively analyzed and shown to regulate flowering time in *Arabidopsis*[26–28] and heading date in rice[13]. As *Ghd7* had a large effect on heading date, we considered the CCT domain–containing gene as a good candidate for *Ghd7*. We carried out complete sequencing of a BAC clone from a Minghui 63 BAC library[29] and one from a Zhenshan 97 BAC library and found that a 38.3-kb fragment harboring the CCT-containing gene was completely deleted in Zhenshan 97 (data not shown). This finding supported the inference that this CCT-containing sequence was the gene responsible for the *Ghd7* QTL.

To test this prediction, we generated a construct by placing an 8.1-kb genomic fragment from Minghui 63, containing the predicted gene region, into the vector pCAMBIA1301 (**Supplementary Fig. 3** online). We introduced this construct by *Agrobacterium*-mediated transformation into rice cultivars Hejiang 19, Mudanjiang 8 and Nipponbare, all of which have relatively small panicle, short stature and early heading compared with Minghui 63 or NIL(mh7). Sequence analysis showed that Hejiang 19 and Mudanjiang 8 had an identical mutation resulting in premature termination of the predicted protein, and Nipponbare had a different allele of the gene (see below). We obtained approximately 20 independent transgenic T_0 plants from each recipient line, and we observed the expected phenotype—that is, tall plants with late heading and large panicles—for almost all the transgene-positive plants (**Supplementary Fig. 4** online). We then field-examined T_1 families from T_0 plants with a single-copy transgene, and we observed perfect cosegregation between the transgene and the phenotype (**Table 1**): all transgene-positive plants were tall

Table 1 Performance of NILs and transgene-positive and transgene-negative segregants in the T_1 generation derived from T_0 plants with single-copy transgenes planted under natural long-day conditions

Genotype	Number of plants	Number of days to heading	Plant height in cm	Number of spikelets per panicle[a]
NIL(zs7)	30	63.7 (A,a)	84.2 (A)	130.6 (A,a)
NIL(het)	30	78.6 (B,b)	109.6 (B)	198.8 (B,b)
NIL(mh7)	30	84.9 (B,c)	117.5 (C)	216.6 (B,c)
Hejiang 19	10	52.8 ± 0.7	60.7 ± 0.7	53.3 ± 3.9
Mudanjiang 8	10	53.0 ± 1.4	67.9 ± 1.1	69.3 ± 4.3
Nipponbare	10	75.3 ± 0.8	88.7 ± 0.8	134.5 ± 9.3
Hejiang 19 (+)	25	104.3 ± 1.7	101.3 ± 1.2	161.7 ± 8.3
Hejiang 19 (–)	10	50.8 ± 1.5	60.7 ± 0.7	55.5 ± 2.1
P[b]		1.08×10^{-19}	4.35×10^{-21}	1.58×10^{-9}
Mudanjiang 8 (+)	28	130.9 ± 1.5	106.8 ± 0.9	123.4 ± 3.8
Mudanjiang 8 (–)	11	59.8 ± 1.2	71.6 ± 1.3	65.0 ± 3.1
P[b]		1.68×10^{-26}	3.20×10^{-23}	2.54×10^{-11}
Nipponbare (+)	47	104.7 ± 1.9	88.1 ± 0.7	99.0 ± 3.8
Nipponbare (–)	11	74.0 ± 0.8	84.0 ± 0.7	69.9 ± 1.5
P[b]		1.95×10^{-10}	0.005347	0.000281

Data presented in the first three rows are from the three NILs planted in randomized complete block design with three replications. A, B and C indicate ranking by Duncan test at $P \leq 0.01$; a, b and c indicate ranking by Duncan test at $P \leq 0.05$ (starting from A, B is significantly different from A, and C is significantly different from B). Data presented in other rows are mean ± s.e.m. (+) and (–) indicate transgene-positive and transgene-negative, respectively.
[a]Data for Nipponbare T_1 are based on the average of all panicles of the plant, and data for other materials are based on the panicle of the main culm. [b]Probability resulting from the *t*-test of two genotypic classes in the T_1s.

and late heading with large panicles, and the reverse was the case for all transgene-negative plants. These results confirmed that the CCT-containing gene was the right candidate for *Ghd7*.

Analyses of the T_1 families showed that the phenotypic effects conferred by the transgene are considerable (**Table 1**). For example, the T_1-positive plants of transgenic Hejiang 19 on average headed 104.2 days after sowing, a 105% increase compared to the transgene-negative segregants, and, the height of the positive plants was 101.3 cm, 66.9% taller than that of the negative plants. Additionally, the number of spikelets per panicle (161.7) of the positive plants was almost threefold compared to that of the negative segregants (55.5). We also observed large effects in the T_1 families of transgenic plants from Mudanjiang 8, and significant effects in progenies from Nipponbare, especially with respect to heading date (**Table 1**).

Features of *Ghd7*

We obtained the complete transcribed sequence of *Ghd7* (1,013 bp) by RACE using RNA isolated from seedlings of Minghui 63. The predicted protein consisted of 257 amino acids (**Supplementary Fig. 5** online). We carried out a BLASTp search of the deduced amino acid sequence of the GHD7 protein against the nonredundant database and identified 43–45 contiguous amino acids from 189 to 233 having significant identity with the CCT domain of the CO protein in *Arabidopsis* (74%, $E = 2e^{-9}$), HD1 in rice (62%, $E = 1e^{-8}$) and a number of proteins found in a diversity of plant species that regulate processes such as photoperiodic flowering[13,26,30], vernalization[31], circadian rhythms[27,32] and light signaling[33]. According to previous analyses[31,34], the CCT domain encoded by *Ghd7* (*OsI* by their naming) diverged most compared with that of other CCT domain–containing proteins in rice, wheat, barley and *Arabidopsis*. In addition, GHD7 does not have an obvious B-box zinc finger structure, nor does *Ghd7* have significant homology with *Arabidopsis* genomic sequence outside the CCT domain. Therefore, GHD7 is distinct from all other members of the CCT domain protein family, and *Ghd7* is thus an evolutionarily new gene in the lineage.

It has been shown that the CCT domain of proteins in the CO family functions in nuclear localization[28], and that this domain is similar to yeast HEME ACTIVATOR PROTEIN2 (HAP2), a subunit of the HAP2-HAP3-HAP5 trimeric complex with DNA binding activity[35]. We carried out a PSORT analysis (see URLs section below) of the GHD7 sequence and identified two overlapping motifs of four basic residues, KRKK starting at amino acid 199 and RKKR starting at amino acid 200, which are characteristic of nuclear localization signals (**Supplementary Fig. 5**). To confirm that GHD7 is a nuclear protein, we fused the coding sequence of *Ghd7* with that of green fluorescence protein (GFP) driven by a ubiquitin promoter (*Ubq*::*Ghd7*:*GFP*) and delivered the construct into onion epidermal cells by particle bombardment. The GHD7 protein localized the GFP signal to the nucleus (**Supplementary Fig. 6** online).

Expression pattern of *Ghd7*

We investigated the diurnal expression pattern of *Ghd7* by quantifying the relative abundance of the mRNA in young leaves of NIL(mh7) using quantitative RT-PCR (**Fig. 2a**). The expression of *Ghd7* was much higher under long-day conditions than short-day conditions, and the *Ghd7* transcript was much more abundant during the light period than in dark period, especially under long-day conditions.

We carried out RNA *in situ* hybridization and *Ghd7* promoter–driven GFP transformation to examine sites of *Ghd7* expression. The RNA *in situ* hybridization signals indicated that the gene is expressed in young tissues, such as the apical meristem, developing leaves and the leaf sheaths of the young seedling, the meristem of the root, the epidermal layer of developing stems and the branch-primordia of developing panicles (**Fig. 3**). The GFP signal was strong in the young stem and in vascular tissues in the blades of fully expended leaves, but weak in the root and vascular tissues of the leaf sheath (**Supplementary Fig. 7** online).

Relation of *Ghd7* to genes in the photoperiod pathway

To investigate the relationship of *Ghd7* to genes associated with the circadian clock and those known to be involved in photoperiod flowering, we compared the expression patterns of six additional genes in NIL(mh7) and NIL(zs7): *OsLhy*, *OsPRR* and *OsGI*, associated with the circadian clock[36], and *Hd1*, *Ehd1* and *Hd3a*, known to have key roles in photoperiod-controlled heading[13,14,18].

Clearly, *Ghd7* had no effect on the expression of the clock-associated genes, as they showed identical expression patterns in the two NILs (data not shown), nor did it affect the expression of *Hd1* (**Fig. 2b**). However, *Ghd7* had a considerable effect on the expression of *Ehd1* and *Hd3a*. *Ghd7* suppressed the expression peak of *Ehd1* in

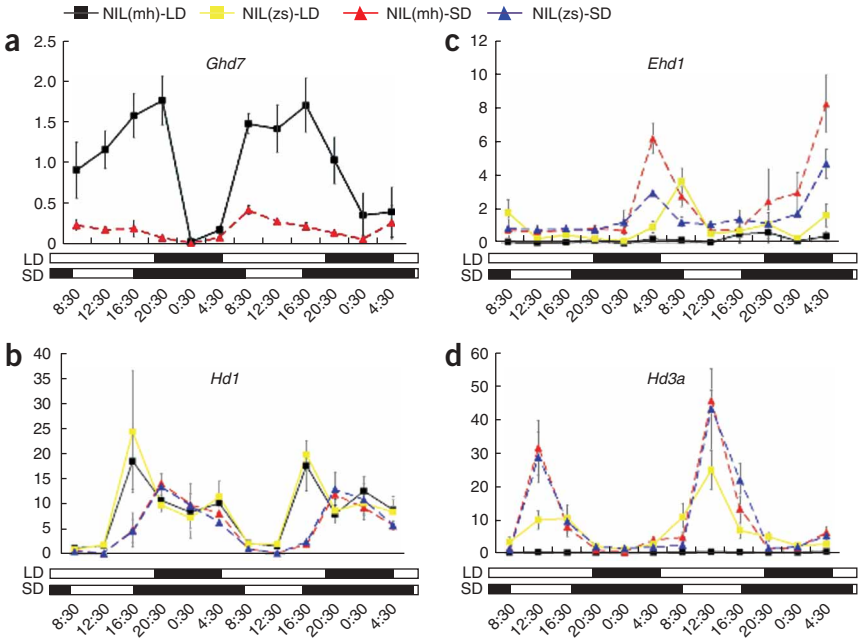


Figure 2 Diurnal expression patterns of NIL(mh7) and NIL(zs7) as indicated by quantitative RT-PCR results. (**a**) *Ghd7*. (**b**) *Hd1*. (**c**) *Ehd1*. (**d**) *Hd3a*. In all panels, the mean of each point is based on the average of three biological repeats calculated using the relative quantification method[44]. The black bars indicate the dark period, and the white bars indicate the light period. The numbers below the bars indicate hours of the day. Error bars, s.e.m.

the early morning under long-day conditions, but slightly promoted *Ehd1* expression in the dark period under short-day conditions (**Fig. 2c**). The expression of *Hd3a* was almost completely suppressed by *Ghd7* under long-day conditions (**Fig. 2d**), whereas *Ghd7* did not affect *Hd3a* expression under short-day conditions. Thus, we concluded that *Ghd7* is upstream of *Ehd1* and *Hd3a* in the photoperiod flowering pathway.

On the basis of the amplitudes of oscillations of the transcript abundance of these three genes (*Ghd7* < *Ehd1* < *Hd3a*), as shown in **Figure 2**, we tentatively propose that *Ghd7* is upstream of *Ehd1*, which in turn is upstream of *Hd3a* in the pathway. As *Ghd7* is distinct from all other members of the CCT-domain protein family identified thus far[34], and as *Ehd1* does not have an ortholog in the *Arabidopsis* genome[18], the *Ghd7-Ehd1-Hd3a* relation may represent a distinct pathway that does not exist in *Arabidopsis*.

Ghd7 also has a large effect on stem growth

Ghd7 also had marked effects on stem growth and development. *Ghd7* transgene–positive and NIL(mh7) plants (referred to as *Ghd7*-positive plants) had more nodes, a longer upper-most internode and thicker stems than their transgene-negative counterparts (**Figs. 1** and **4a** and **Supplementary Table 2**). An examination of the basal internode showed that the stems of *Ghd7*-positive plants consisted of many more layers of larger cells than the negative plants (**Fig. 4b**), accounting for the increase in thickness of the stems. Notably, the *Ghd7*-positive plants also had much better developed vascular systems and mechanical tissues (**Fig. 4c**), indicating increased transport capacity and mechanical strength, which are crucial for high yield and lodging resistance. Indeed, *Ghd7*-positive plants were much more resistant to lodging than the negative plants (data not shown), although the positive plants were much taller. The observation of better developed vascular systems and mechanical tissues is consistent with the hybridization signals detected surrounding the epidermal layer and vascular bundles (**Fig. 3d**). However, there was not much difference in cell size between *Ghd7*-positive and negative plants in the upper-most internode (**Fig. 4d**), although the length of this internode differed the most between the positive and negative plants. Thus, the increased height of *Ghd7*-positive plants is mostly due to increased cell number rather than cell length.

Ghd7 effects vary with genetic background

We compared the performance of the NILs and the transgenic progenies from the three recipient lines, Hejiang 19, Mudanjiang 8 and Nipponbare, and found that the effects of *Ghd7* varied with genetic background (**Table 1**). The effect of this gene was smaller in transgenic progenies of Nipponbare compared with those in the backgrounds of Hejiang 19, Mudanjiang 8 and Zhenshan 97.

To assess how the genetic backgrounds might have modulated the effect of *Ghd7*, we obtained DNA sequences of *Ghd7*, together with those of *Hd1*, *Ehd1* and *Hd3a*, for Hejiang 19, Mudanjiang 8, Nipponbare and Zhenshan 97. The GHD7 protein sequence of Nipponbare differs from that of Minghui 63 by four amino acids (**Supplementary Table 5** and **Supplementary Fig. 5** online), and shows apparently reduced phenotypic effect (**Table 1** and **Supplementary Table 4**). As described above, the *Ghd7*-containing genomic fragment is completely deleted in Zhenshan 97. The predicted sequences of GHD7 in Hejiang 19 and Mudanjiang 8 have an identical premature termination (**Supplementary Table 5** and **Fig. 5**), and thus are presumably nonfunctional. Comparison of *Hd1* sequences showed that Zhenshan 97 has a large deletion plus a large insertion relative to Nipponbare, and that the *Hd1* sequences of Hejiang 19 and Mudanjiang 8 are identical, having four insertion and/or deletions (**Supplementary Fig. 8** online). Both of these two alleles are unlikely functional. The predicted protein sequence of HD3A of Zhenshan 97 differed from Nipponbare in the last amino acid at the C terminus (**Supplementary Fig. 9** online). The same difference was observed previously in the comparison between Nipponbare and Kasalath[14]. Again, Hejiang 19 and Mudanjiang 8 have identical sequences that differ from Nipponbare by one amino acid. These variant amino acids may not have notable effects on function. The EHD1 sequences of the four cultivars are identical (data not shown).

Thus, *Ghd7* showed large phenotypic effects in genetic backgrounds with nonfunctional alleles of *Ghd7* and *Hd1* (Zhenshan 97, Hejiang 19 and Mudanjiang 8), but its effect was reduced in a background with functional alleles of both genes (Nipponbare).

Geographical distribution of naturally occurring alleles

We obtained the *Ghd7* sequences for 19 genotypes, including cultivars, landraces and parents of elite rice hybrids from rice growing areas representing a wide geographic range of Asia (**Supplementary Table 5** and **Supplementary Fig. 10** online). We also obtained the sequence for an accession of the common wild rice (*Oryza rufipogon* Griff.)

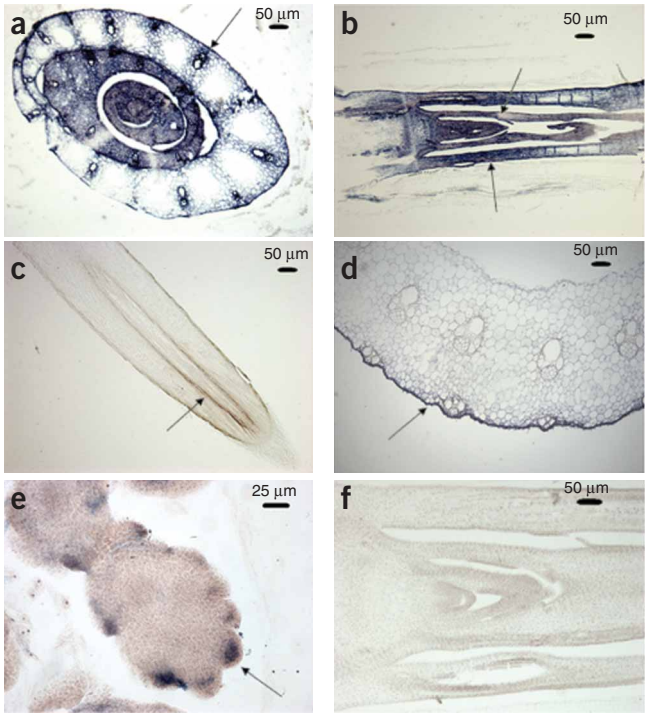


Figure 3 Expression of *Ghd7* in various tissues detected by *in situ* hybridization with an antisense probe using various tissues of Minghui 63. (**a**,**b**) Transverse (**a**) and longitudinal (**b**) sections of shoot apex at 7 days after germination showing strong expression in the apical meristem, developing leaves and vascular tissues of the sheath. (**c**) Longitudinal section of the root at early panicle developing stage. (**d**) Transverse section of a stem at booting stage. (**e**) Young panicle at the secondary branch primordium stage. (**f**) A negative control using the sense probe. Samples shown **c**–**e** were taken at 17:00–18:00 h in long-day conditions. Arrows indicate the hybridization signals.

from Myanmar. Comparison of the predicted protein sequences identified five alleles that showed a clear geographic distribution (**Supplementary Table 5** and **Supplementary Figs. 5** and **10**). *Ghd7*-1 was found in rice varieties from the tropics, subtropics and areas with hot summers and long growing seasons, including the most popular high-yield cultivars and parents of elite hybrids in China and Southeast Asia. The wild rice accession also had this allele, indicating that *Ghd7*-1 is probably the original wild-type allele. *Ghd7*-2, differing from *Ghd7*-1 by four amino acids, was found in varieties of temperate *japonica* from Japan and northern China. This allele seemed to have a smaller phenotypic effect than *Ghd7*-1 (**Table 1** and **Supplementary Table 4**). *Ghd7*-3 occurring in Teqing is distinct from all the other alleles in that it differs by three amino acids from both *Ghd7*-1 and *Ghd7*-2, with one variant amino acid found only in this variety. The phenotypic effect of *Ghd7*-3 seems to be the same magnitude as that of *Ghd7*-1 (**Table 1** and **Supplementary Table 4**). The remaining two alleles (*Ghd7*-0 and *Ghd7*-0a) were nonfunctional. *Ghd7*-0, with the *Ghd7* locus completely deleted, was found in two varieties grown as early rice in a system with two-rice crops in central and southern China, in which the growth duration for early rice is short. *Ghd7*-0a, resulting from a premature termination in the predicted coding region, was found in two varieties from the Heilongjiang Province of northeastern China, which has cool summers and a short growing season.

DISCUSSION

A number of genes individually controlling number of grains per plant, heading date and plant height have been cloned recently using map-based cloning approaches[2,5–8,13,14,17,18,31]. Compared with the genes reported previously, *Ghd7* is notable in that it has large pleiotropic effects on an array of traits, including grain number, heading date and plant height. Such pleiotropic effects may provide an explanation for QTL hot spots that were observed in many studies (Gramene, see URLs section below).

CCT-domain proteins have been reported to have crucial roles in regulating processes such as photoperiodic flowering[13,26,30], vernalization[31], circadian rhythms[27,32] and light signaling[33]. The results of the present study indicate that GHD7 has a key role in photoperiod flowering by regulating the putative *Ehd1-Hd3a* pathway. Strong expression of *Ghd7* in the vascular tissues of leaves is consistent with the role of *CO*-like genes in flowering regulation. However, we also observed expression of *Ghd7* in organs that may not be directly related to photoperiod-controlled flowering, such as roots that do not seem to have a role in flowering, leaves and meristems in the early seedling stage that are much too early for flower induction, and the stems and panicles that already passed the stage of phase transition.

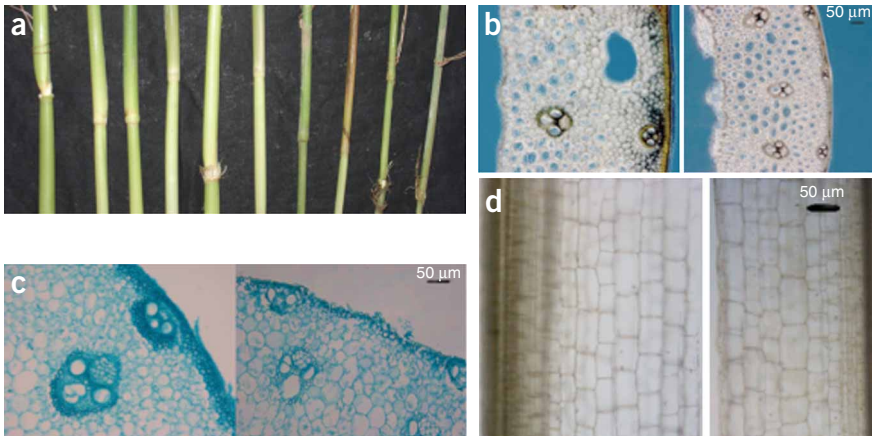


Figure 4 Effects of *Ghd7* on the stem. (**a**) The first nodes and stems above the ground from the main culms for the following varieties (from left to right): HR5 (having the same *Ghd7* allele as Minghui 63), Teqing, Minghui 63, NIL(mh7), ZS(tq7), NIL(hr7) (having the *Ghd7* allele from HR5), Nipponbare, Zhenshan 97, Hejiang 19 and Mudanjiang 8. The first six lines have fully functional alleles of *Ghd7*, Nipponbare has a weaker allele, Zhenshan 97 has the *Ghd7* locus deleted and Hejiang 19 and Mudanjiang 8 have a nonfunctional allele of *Ghd7*. (**b**) Transverse sections of the stems above the basal nodes from the main culms of plants grown in the field under natural long-day conditions: left, NIL(mh7); right, Zhenshan 97. (**c**) Transverse sections of the stems above the basal nodes from the main culms of plants grown in pots in a greenhouse (approximately 12-h day length): left, *Ghd7*-transgenic T_0 plant of Zhenshan 97; right, wild type. (**d**) Longitudinal sections of the stems approximately 2 cm above the upper-most nodes from the main culms of plants grown in the field under natural long-day conditions. Left, NIL(mh7): outer to inner (left to right); right, Zhenshan 97: outer to inner (right to left).

Expression in these tissues corresponded well with increases in the various organs leading to an overall increase in plant size, suggesting the possibility that this gene may have a general role in promoting growth, cell proliferation and differentiation, in addition to photoperiod flowering.

Ghd7 controls heading date through its enhanced expression under long-day conditions to repress the expression of *Hd3a*, likely through *Ehd1*, thus delaying flowering. Of note, the cereal *VRN2* gene, although referred to as a vernalization gene, actually has a day-length control and is only expressed under long-day conditions, when it acts as a repressor of flowering[37,38]. *VRN2* was classified in the same CCT subfamily as *Ghd7* (ref. 31). This suggests that tropical short-day plants that do not use vernalization as a flowering cue, and temperate long-day plants that do use vernalization, actually use related genes to repress flowering in long-day conditions. Such similarity in flowering control provides important clues as to how the temperate cereals might have evolved. Moreover, these genes also act as major sources of adaptive variation, and hence are keys to understanding the spread and success of cereals.

Sequence analysis of allelic variants at the *Ghd7* locus indicated that this locus has contributed greatly to both productivity and adaptability of cultivated rice on a global scale. Asian cultivated rice originated in tropical and subtropical regions of Asia[39]. The functional alleles with strong effects (for example, *Ghd7*-1 and *Ghd7*-3) allow rice plants to fully exploit light and temperature by delaying flowering under long-day conditions in areas with long growing seasons, thus producing large panicles and increasing yield. The mutations giving rise to the *Ghd7*-0 and *Ghd7*-2 alleles, with no or reduced effect of delaying heading under long-day conditions, have had crucial roles by enabling rice to be cultivated under conditions with short growth duration and/or temperate regions. Additionally, the substantial dominant effect (**Table 1** and **Supplementary Tables 1** and **3**) observed in the heterozygote between NIL(zs7) and NIL(mh7) indicated that the *Ghd7* locus has also contributed significantly to the high level of heterosis of Shanyou 63 (refs. 19,40), a widely cultivated hybrid in China in the last more than 20 years.

Yield has been generally regarded as a complex trait that is controlled by multiple genes of small effects. The major effects demonstrated by the NILs and transgenic plants and the cloning of this QTL have fundamental implications for yield improvement, suggesting that yield, like other traits, can also be improved by individually manipulating the component traits using both molecular marker–assisted selection and transformation.

METHODS

Plant materials. We used a total of 19 accessions, including ten *indica* (*O. sativa* L. ssp. *indica*) varieties, eight *japonica* (*O. sativa* L. ssp. *japonica*) varieties and one accession of common wild rice (*O. rufipogon*) (**Supplementary Table 5**).

Field growth of the rice plants. The rice plants examined under natural field conditions were grown in normal rice growing seasons in the Experimental Station of Huazhong Agricultural University, Wuhan, China. Each year, seeds were planted in a seed bed in mid-May, and transplanted to the field in mid-June. The planting density was 16.5 cm between plants in a row, and the rows were 26 cm apart. Field management, including irrigation, fertilizer application and pest control, followed essentially the normal agricultural practice.

Complementation test. A clone numbered 60F11 from the BAC library of Minghui 63 genomic DNA[29] containing the complete *Ghd7* sequence was isolated and digested with restriction endonucleases *Bam*HI and *Eco*RI. A fragment of 8,175 bp in length, containing 2,261 bp upstream the transcription start site and 3,255 bp downstream of the termination site, was recovered by agarose gel electrophoresis. We inserted the fragment into the binary vector pCAMBIA1301 (ref. 41). The construct was introduced into the *Agrobacterium* strain EHA105, and transformation was done as described[42].

We determined the genotypes of the transgenic plants by PCR amplification of the genomic DNA using a pair of primers, SEL-F and SEL-R (**Supplementary Table 6** online), designed according to the sequence downstream the transcription termination site of *Ghd7*. The PCR products of the recipient lines (Mudanjiang 8, Hejiang 19 and Nipponbare) were 133 bp, and that of the transgene was 109 bp, which can be well separated using 2% agarose gel electrophoresis.

RNA *in situ* hybridization. Freshly collected plant tissues were fixed in FAA solution (50% ethanol, 5% acetic acid, 3.7% formaldehyde) at 4 °C overnight, dehydrated with ethanol from 50% to 100%, cleaned by a series of xylene washes from 25% to 100%, and embedded in paraffin at 52–54 °C. We cut microtome sections about 8–12 μm in thickness using microtome YD-1508A, and then we mounted the sections on RNase-free glass slides.

To prepare the probe, we used a pair of primers, insitu-F and insitu-R (**Supplementary Table 6**), to amplify a 513-bp unique sequence of *Ghd7* from a cDNA clone. The fragment was then inserted into the pGEM-T vector (Promega) for RNA transcription. The sense and antisense RNA probes were produced by T7 and SP6 transcriptase labeled with digoxigenin (Roche). RNA *in situ* hybridization and immunological detection were done as described previously[43].

Analysis of diurnal expression patterns. Seeds of NIL(mh7) and NIL(zs7) were sown in pots in a greenhouse. After growing under approximately neutral day-length conditions (12 h of sunlight per day) for 32 d, half of the plants received a short-day treatment (9 h/d) in a phytotron, and the other half received a long-day treatment (15 h/d) in a different phytotron. The light intensity in the phytotrons was set at 10,000 lx, with the temperature set to 33 °C in the light period and 28 °C in the dark period, and the humidity was set at 50%. After entraining for 10 d, young leaves were simultaneously harvested from three different plants as three biological repeats for each treatment, and stored in liquid nitrogen. We collected the samples in 4 h intervals, starting at 08:30 for a total of 48 h.

For quantitative RT-PCR, we isolated total RNA from the leaves using a RNA extraction kit (TRIzol reagent, Invitrogen). About 2 μg total RNA was reverse-transcribed using M-MLV reverse transcriptase (Promega) in a volume of 80 μl to obtain cDNA. We used primers G7rq-F and G7rq-R for amplifying the transcript of *Ghd7*, Hd1-F and Hd1-R for *Hd1*, Ehd-F and Ehd-R for *Ehd1*, Hd3a-F and Hd3a-R for *Hd3a*, LHY-F and LHY-R for *OsLhy*, PRR-F and PRR-R for *OsPRR*, GI-F and GI-R for *OsGI*, and Actin1-F and Actin1-R for *Actin1* as the internal control (see **Supplementary Table 6** for the primers).

We carried out quantitative RT-PCR in a total volume of 25 μl containing 2 μl of the reverse-transcribed product above, 0.25 μM gene-specific primers and 12.5 μl SYBR Green Master Mix (Applied Biosystems) on a Applied Biosystems 7500 Real-Time PCR System according to the manufacturer's instructions. The measurements were obtained using the relative quantification method[44].

RACE. To synthesize the first-strand cDNA, we reverse-transcribed 1 μg total RNA using the protocol provided by the SMART RACE cDNA Amplification Kit (Clontech), with a final volume of 100 μl.

For 5′-RACE, we conducted the first round of amplification according to the protocol provided with the Advantage 2 PCR Kit (Clontech) using primer 593R and the UPM adaptor primer from the kit. For second round of amplification, we diluted the PCR product from the first round 50× and used 1 μl of the dilution as the template. The amplification was conducted using the nest primer 3-4R together with the adaptor primer NUPM provided in the kit. The 3′-RACE was essentially the same, except that primer 2067F was used with UPM in the first round of reaction, 4147F was used with NUPM in the second round of reaction, and the dilution was 30×. The sequences of the primers are listed in **Supplementary Table 6**.

Subcellular localization of GHD7. The *GUS* (β-glucuronidase) gene in the vector pCAMBIA 1391Xa[41] used in this experiment was replaced by *GFP*. The coding sequence of *Ghd7* was amplified with the cDNA clone as the template using the primers Fusin-F and Fusin-R (**Supplementary Table 6**). The PCR

product was inserted into pCAMBIA 1391Xa vector fused with GFP in frame, under the control of the maize Ubq promoter. The fused plasmid co-precipitated with golden particles and was introduced into onion epidermal cells by particle bombardment using the PDS-1000 system (Bio-Rad) at 1,100 psi helium pressure. We obtained the GFP image using a confocal microscope (Leica) after the transformed cells were incubated in the dark for 36 h.

URLs. Gramene, http://www.gramene.org; PSORT analysis, http://wolfpsort.org

Accession codes. GenBank: full-length cDNA of *Ghd7*, EU286800; genomic DNA of *Ghd7*, EU286801. IRRI germplasm bank: common wild rice, IRGC80752.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank N. Chua and M. Matsuoka (Nagoya University) for helpful comments on a previous version of the manuscript, and J.B. Wang for field management. This work was supported by the Ministry of Science and Technology and the National Natural Science Foundation of China.

AUTHOR CONTRIBUTIONS

W.X., Y.X., W.T., C.X. and X.L. conducted genetic mapping, gene cloning and data collection; X.W. and L.W. conducted expression analysis; Y.Z. conducted the *in situ* hybridization; H.Z. and S.Y. conducted NIL development; Y.X. and Q.Z. designed and supervised the study; and Q.Z. analyzed the data and wrote the paper. All the authors discussed the results and commented on the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Li, X. *et al.* Control of tillering in rice. *Nature* **422**, 618–621 (2003).
2. Ashikari, M. *et al.* Cytokinin oxidase regulates rice grain production. *Science* **309**, 741–745 (2005).
3. Fan, C. *et al. GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **112**, 1164–1171 (2006).
4. Song, X.J., Huang, W., Shi, M., Zhu, M.Z. & Lin, H.X.A. QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **39**, 623–630 (2007).
5. Peng, J. *et al.* 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* **400**, 256–261 (1999).
6. Ashikari, M., Wu, J., Yano, M., Sasaki, T. & Yoshimura, A. Rice gibberellin-insensitive dwarf mutant gene *Dwarf 1* encodes the alpha-subunit of GTP-binding protein. *Proc. Natl. Acad. Sci. USA* **96**, 10284–10289 (1999).
7. Spielmeyer, W., Ellis, M.H. & Chandler, P.M. Semidwarf (*sd-1*), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *Proc. Natl. Acad. Sci. USA* **99**, 9043–9048 (2002).
8. Itoh, H. *et al.* A rice semi-dwarf gene, *Tan-Ginbozu* (*D35*), encodes the gibberellin biosynthesis enzyme, ent-kaurene oxidase. *Plant Mol. Biol.* **54**, 533–547 (2004).
9. Simpson, G.G. & Dean, C. *Arabidopsis*, the Rosetta stone of flowering time? *Science* **296**, 285–289 (2002).
10. Wigge, P.A. *et al.* Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science* **309**, 1056–1059 (2005).
11. Abe, M. *et al.* FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* **309**, 1052–1056 (2005).
12. Corbesier, L. *et al.* FT protein movement contributes to long-distance signalling in floral induction of *Arabidopsis*. *Science* **316**, 1030–1033 (2007).
13. Yano, M. *et al. 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 (2000).
14. Kojima, S. *et al. 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 (2002).
15. Hayama, R., Yokoi, S., Tamaki, S., Yano, M. & Shimamoto, K. Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* **422**, 719–722 (2003).
16. Tamaki, S., Matsuo, S., Wong, H.L., Yokoi, S. & Shimamoto, K. Hd3a protein is a mobile flowering signal in rice. *Science* **316**, 1033–1036 (2007).
17. Takahashi, Y., Shomura, A., Sasaki, T. & Yano, M. *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proc. Natl. Acad. Sci. USA* **98**, 7922–7927 (2001).
18. Doi, K. *et al. 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 (2004).
19. Yu, S.B. *et al.* Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA* **94**, 9226–9231 (1997).
20. Yu, S.B. *et al.* Identification of quantitative trait loci and epistatic interactions for plant height and heading date in rice. *Theor. Appl. Genet.* **104**, 619–625 (2002).
21. Li, J.X. *et al.* Analyzing quantitative trait loci for yield using a vegetatively replicated F_2 population from a cross between the parents of an elite rice hybrid. *Theor. Appl. Genet.* **101**, 248–254 (2000).
22. Xing, Y., Tan, Y.F., Xu, C.G., Hua, J.P. & Sun, X.L. Mapping and isolation of quantitative trait loci controlling plant height and heading date in rice. *Acta Bot. Sin.* **43**, 721–746 (2001).
23. Xing, Z. *et al.* Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. *Theor. Appl. Genet.* **105**, 248–257 (2002).
24. Hua, J.P. *et al.* Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. *Genetics* **162**, 1885–1895 (2002).
25. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* **436**, 793–800 (2005).
26. Putterill, J., Robson, F., Lee, K., Simon, R. & Coupland, G. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**, 847–857 (1995).
27. Strayer, C. *et al.* Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. *Science* **289**, 768–771 (2000).
28. Robson, F. *et al.* 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 (2001).
29. Peng, K.M., Zhang, H.B. & Zhang, Q.A. BAC library constructed to the rice cultivar "Minghui 63" for cloning gene of agronomic importance. *Acta Bot. Sin.* **40**, 1108–1114 (1998).
30. Turner, A., Beales, J., Faure, S., Dunford, R.P. & Laurie, D.A. The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. *Science* **310**, 1031–1034 (2005).
31. Yan, L. *et al.* The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* **303**, 1640–1644 (2004).
32. Salome, P.A., To, J.P., Kieber, J.J. & McClung, C.R. *Arabidopsis* response regulators ARR3 and ARR4 play cytokinin-independent roles in the control of circadian period. *Plant Cell* **18**, 55–69 (2006).
33. Kaczorowski, K.A. & Quail, P.H. *Arabidopsis* PSEUDO-RESPONSE REGULATOR7 is a signalling intermediate in phytochrome-regulated seedling deetiolation and phasing of the circadian clock. *Plant Cell* **15**, 2654–2665 (2003).
34. Griffiths, S., Dunford, R.P., Coupland, G. & Laurie, D.A. The evolution of *CONSTANS*-like gene families in barley, rice, and *Arabidopsis*. *Plant Physiol.* **131**, 1855–1867 (2003).
35. Wenkel, S. *et al.* CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell* **18**, 2971–2984 (2006).
36. Murakami, M., Tago, Y., Yamashino, T. & Mizuno, T. Comparative overviews of clock-associated genes of *Arabidopsis thaliana* and *Oryza sativa*. *Plant Cell Physiol.* **48**, 110–121 (2007).
37. Dubcovsky, J. *et al.* Effect of photoperiod on the regulation of wheat vernalization genes *VRN1* and *VRN2*. *Plant Mol. Biol.* **60**, 469–480 (2006).
38. Trevaskis, B., Hemming, M.N., Peacock, W.J. & Dennis, E.S. *HvVRN2* responds to daylength, whereas *HvVRN1* is regulated by vernalization and developmental status. *Plant Physiol.* **140**, 1397–1405 (2006).
39. Oka, H.I. *Origin of Cultivated Rice* (Japan Scientific Soc. Press, Tokyo, 1988).
40. Hua, J. *et al.* Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA* **100**, 2574–2579 (2003).
41. Hajdukiewicz, P., Svab, Z. & Maliga, P. The small, versatile pPZP family of Agrobacterium binary vectors for plant transformation. *Plant Mol. Biol.* **25**, 989–994 (1994).
42. Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (*Oryza sativa* L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282 (1994).
43. De Block, M. & Debrouwer, D. RNA-RNA *in situ* hybridization using digoxigenin-labelled probes: the use of high-molecular-weight polyvinyl alcohol in the alkaline phosphatase indoxyl-nitroblue tetrazolium reaction. *Anal. Biochem.* **215**, 86–89 (1993).
44. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}$T method. *Methods* **25**, 402–408 (2001).

Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice

Weiya Xue, Yongzhong Xing, Xiaoyu Weng, Yu Zhao, Weijiang Tang, Lei Wang, Hongju Zhou, Sibin Yu, Caiguo Xu, Xianghua Li, Qifa Zhang

Supplementary tables

Supplementary Table 1 Measurements of yield and yield component traits of the near isogenic lines planted under natural LDs. Thirty plants were measured for each genotype.

Genotype	Number of tillers/plant	Number of grains/panicle	1000-grain weight (g)	Yield/plant (g)
NIL(zs7)	11.8 A	65.3 A	22.6 a	17.5 Aa
NIL(het)	9.5 B	114.9 B	22.6 a	24.8 Bb
NIL(mh7)	8.7 C	132.9 C	22.9 a	26.4 Bc

A, B, C: ranked by Duncan test at $P \leq 0.01$; a, b, c, ranked by Duncan test at $P \leq 0.05$

Supplementary Table 2 Measurements of characters of the main culms. Number of plants measured for each genotype is the same as in Table 1.

	NIL(zs7)	NIL(het)	NIL(mh7)	Hejiang 19(-)	Hejiang 19(+)
Length of internode (cm)*					
1st	31.3±0.3	39.0±1.2	41.2±0.5	25.8±2.2	36.3±4.1
2nd	21.5±0.1	23.5±0.2	23.0±0.6	14.0±0.8	17.1±2.0
3rd	7.2±0.2	14.0±0.9	13.6±0.9	5.5±1.5	11.0±1.6
4th		6.8±0.7	8.0±1.0		9.3±1.8
5th			3.1±0.5		5.7±1.6
6th					2.7±0.7
Panicle length (cm)	22.8±0.5	24.9±0.5	28.8±0.1	14.8±0.8	21.3±1.9
No. primary branches	13.1±0.5	16.2±0.3	16.7±0.2	7.9±0.7	17.2±2.2
No. secondary branches	24.1±1.6	37.1±1.6	40.5±1.3	5.2±1.3	22.2±10.9

*The internode from top to bottom was numbered 1st, 2nd etc.

Data in each cell are presented as mean ± standard error.

Supplementary Table 3 Genetic effects of the QTLs. QTL for heading date, plant height and number of spikelets per panicle assessed using MapMaker/QTL in a BC_3F_2 population segregating for the introgressed *Ghd7* segment in the Zhenshan 97 genetic background

Trait	LOD*	A†	D‡	D/A	Var (%)§
Heading date (days)	102.9	12.3	4.0 C	0.33	91.8
Plant height (cm)	76.7	18.8	12.3 A	0.65	88.9
No. spikelets per panicle	56.1	59.8	28.3 B	0.47	74.2

*Logarithm of odd; †Additive effect (P=0.000000 by t-test for all the three traits); ‡Dominant effect (A, B and C indicate probabilities 0.00001, 0.00095 and 0.019 by t-test, respectively); §Amount of phenotypic variation explained.

Supplementary Table 4 Measurements of the three traits for NILs containing different *Ghd7* alleles.

Genotype*	No. plants	No. days to heading	Plant height (cm)	No. spikelets on the main panicle
Zhenshan 97	14	63.7±0.4	87.8±0.9	124.4±3.0
ZS(het)	27	81.5±0.4	111.2±0.7	204.4±5.3
ZS(tq7)	14	88.7±0.8	113.3±1.2	212.8±8.3
ZS(Nip7)	10	77.8±0.4	110.8±0.8	170.4±5.2

*ZS(tq7) is a NIL developed by introgressing the *Ghd7* fragment from Teqing into the Zhenshan 97 genetic background; ZS(het) is a hybrid between Zhenshan 97 and ZS(tq7); ZS(Nip7) is a NIL developed by introgressing the *Ghd7* fragment from Nipponbare into the Zhenshan 97 genetic background.

Data in each cell are presented as mean ± standard error.

Supplementary Table 5 *Ghd7* alleles found in 19 accessions representing a broad range of rice germplasm in Asia.

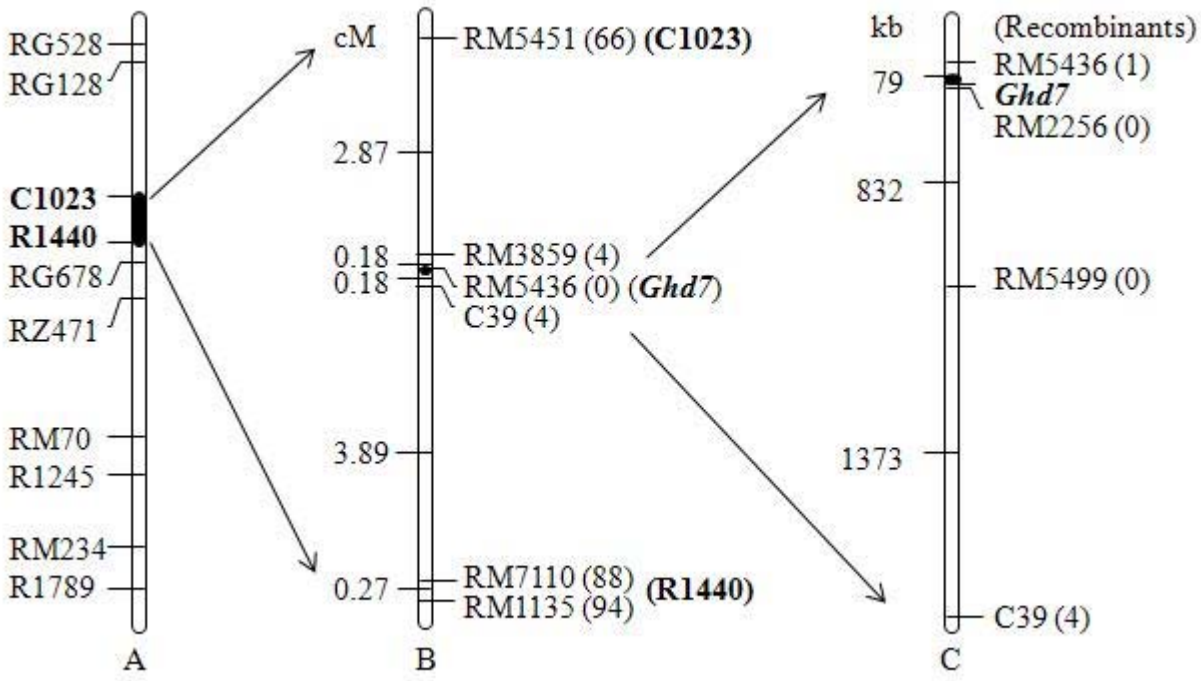
Variety	Class	Allele*	Origin
Minghui 63	Indica	*Ghd7*-1	Fujian, China
9311	Indica	*Ghd7*-1	Jiangsu, China
Wukezhan	Indica	*Ghd7*-1	Fujian, China
Qimei	Indica	*Ghd7*-1	Guangdong, China
Menjiading	Indica	*Ghd7*-1	Hainan, China
TKM5	Indica	*Ghd7*-1	India
IR64	Indica	*Ghd7*-1	Philippines
O. rufipogon (IRGC80752)	Wild	*Ghd7*-1	Myanmar
Slg-1	Japonica	*Ghd7*-2	Japan
Nipponbare	Japonica	*Ghd7*-2	Japan
Zhonghua 11	Japonica	*Ghd7*-2	Beijing, China
Akihikari	Japonica	*Ghd7*-2	Japan
Shennong 606	Japonica	*Ghd7*-2	Liaoning, China
Liming B	Japonica	*Ghd7*-2	Liaoning, China
Teqing	Indica	*Ghd7*-3	Guangdong, China
Zhenshan 97	Indica	*Ghd7*-0	Jiangxi, China
Nanxiongzaoyou	Indica	*Ghd7*-0	Guangzhou, China
Mudanjiang 8	Japonica	*Ghd7*-0a	Heilongjiang, China
Hejiang 19	Japonica	*Ghd7*-0a	Heilongjiang, China

*See Supplementary Figure 5 for sequence information of the alleles.

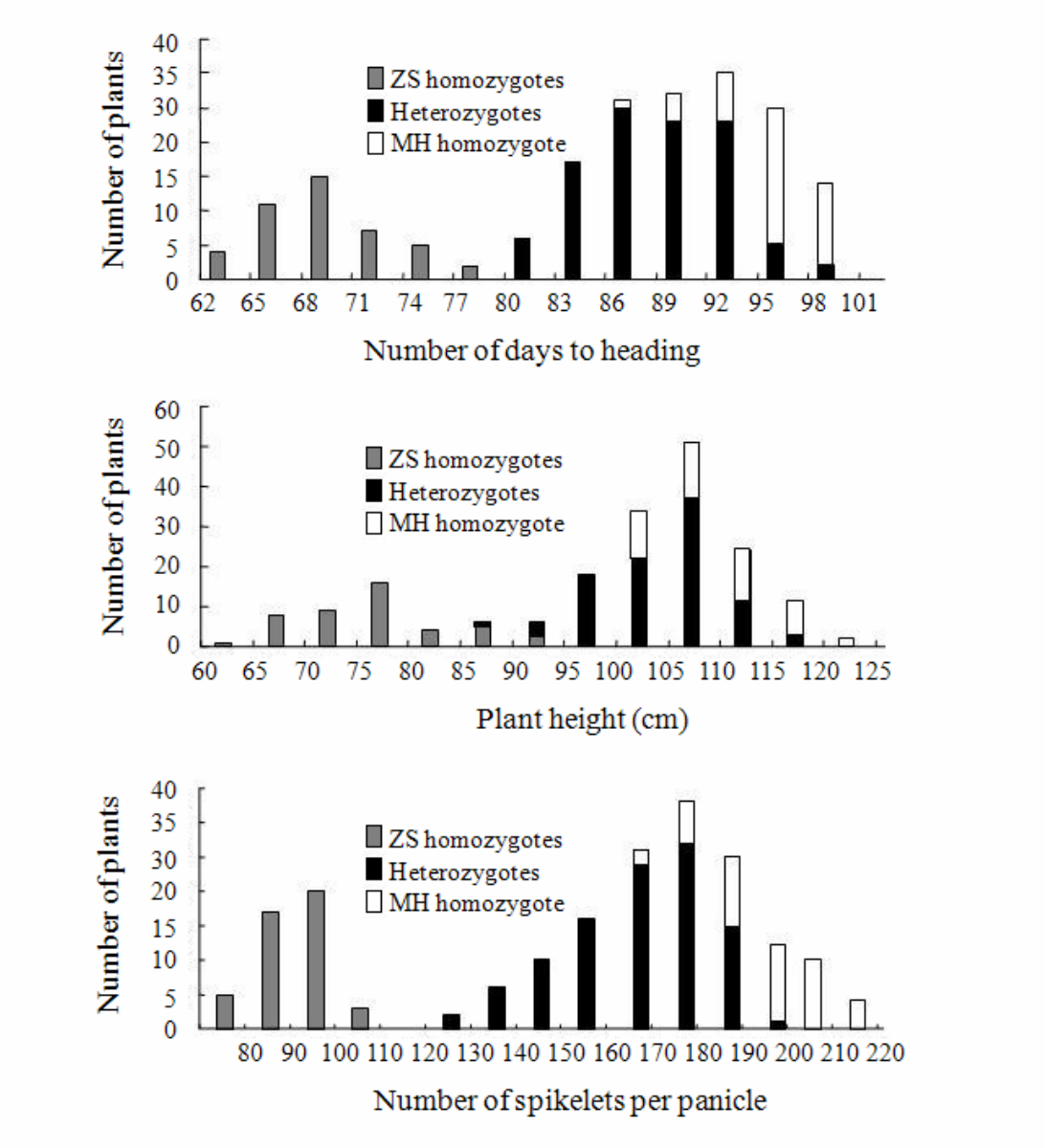
Supplementary Table 6 The sequences of primers used in this study.

Name	Forward (5'-3')	Reverse (5'-3')
Sel	tgcatgcatatacattagct	tgcatgcatatacattagct
Insitu	ttatccgttcatgtcgatgg	accatctccttgggcatcga
G7rq	ttatccgttcatgtcgatgg	gggcctcatctcggcatag
Hd1	tcagcaacagcatatctttctcatca	tctggaatttggcatatctatcacc
Hd3a	gctcactatcatcatccagcatg	ccttgctcagctatttaattgcataa
Actin1	tgctatgtacgtcgccatccag	aatgagtaaccacgctccgtca
593R		accatctccttgggcatcga
3-4R		ttgccgaagaactggaactc
2067F	accagcagccggagaaggat	
4147F	gtcatattgtgggagcacgt	
LHY	cagataaggccgacaccaaac	ggtgtgttggaaccacatg
PRR	ctgctgaacctctggaccca	ggttccgataacgccaactc
GI	tggagaaaggttgtggatgc	gatagacggcacttcagcagat
Promoter	ggatccgacacctgcacgtgaaaatctc	ctgcagcgacatgaacggataaatcaaac
Fusin	ggtaccttatccgttcatgtcgatgg	ggattctctgaaccattgtccaagctc

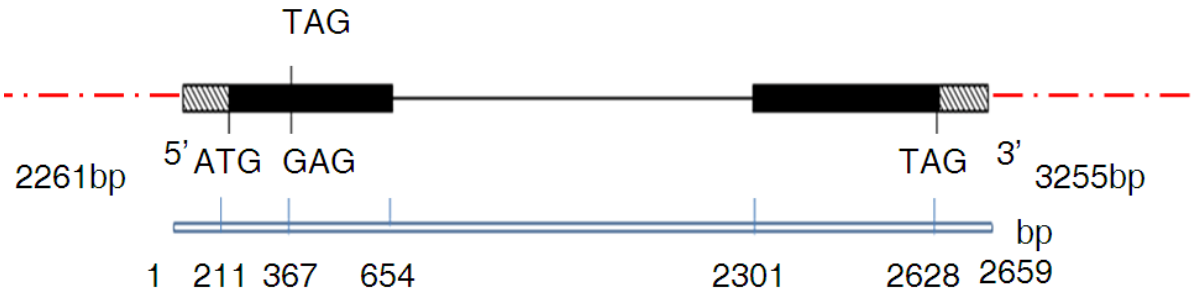
Supplementary figures



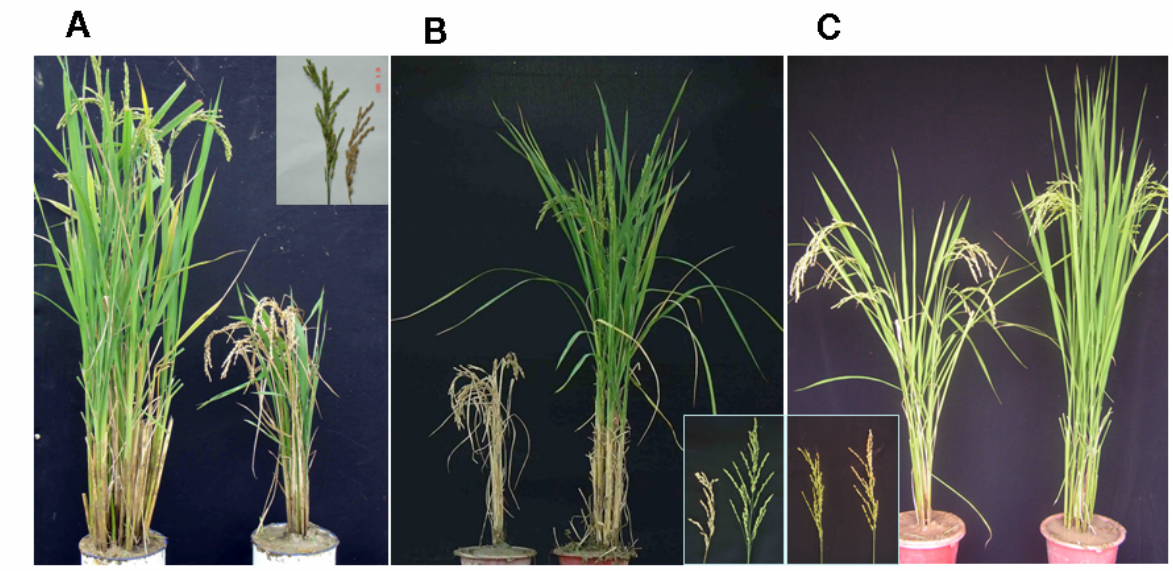
Supplementary Figure 1 Genetic and physical maps of the *Ghd7* locus. (A) Linkage map developed using RILs from a cross between Zhenshan 97 and Minghui 63[24]. (B) Local linkage map of the *Ghd7* surrounding region based on 1082 recessive plants in a BC_2F_2 population from RI50 backcrossed with Zhenshan 97. Numbers in parentheses in the right hand side indicate the numbers of recombinants detected between the markers and the *Ghd7* locus. (C) Physical map of the *Ghd7* genomic region based on 800 recessive plants in a BC_4F_2 population from the backcross series using Teqing as the donor and Zhenshan 97 as the recurrent parent. Numbers of recombinants between the markers and the *Ghd7* locus are indicated in parentheses in the right hand side, and the physical distances are based on the Nipponbare genomic sequence[26].



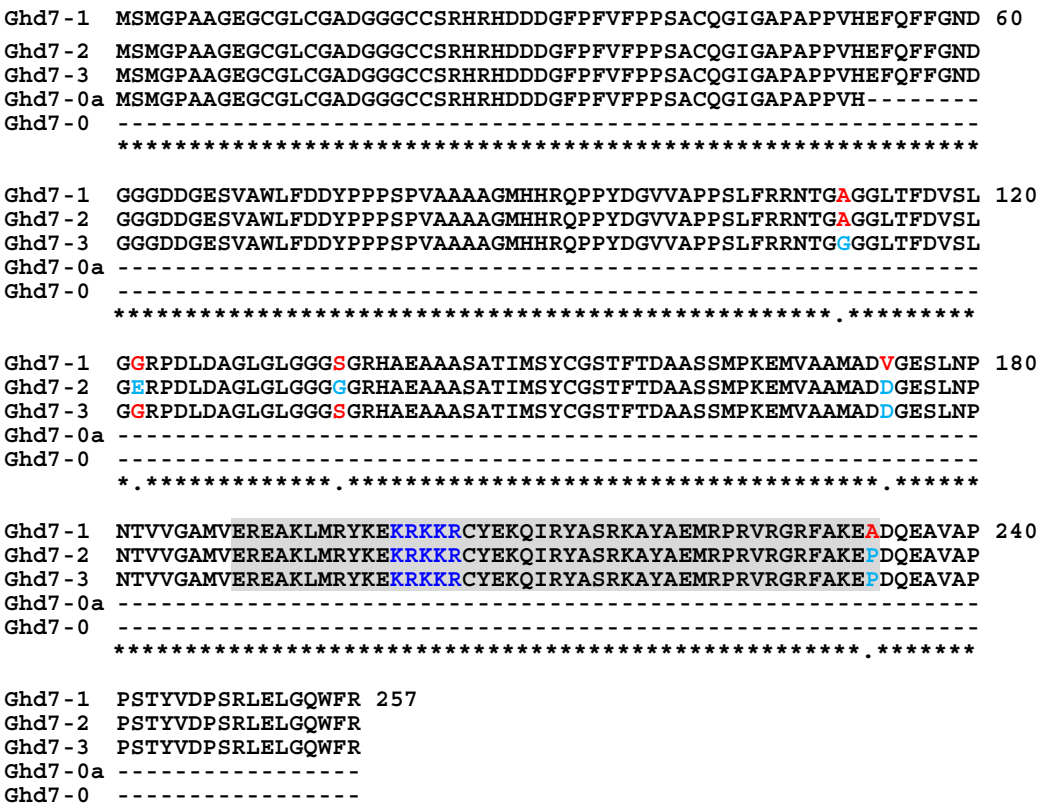
Supplementary Figure 2 Frequency distribution of heading date, plant height and number of spikelets per panicle in the BC_2F_2 population developed by backcrossing RI50 with Zhenshan 97, and confirmed by BC_2F_3 family data.



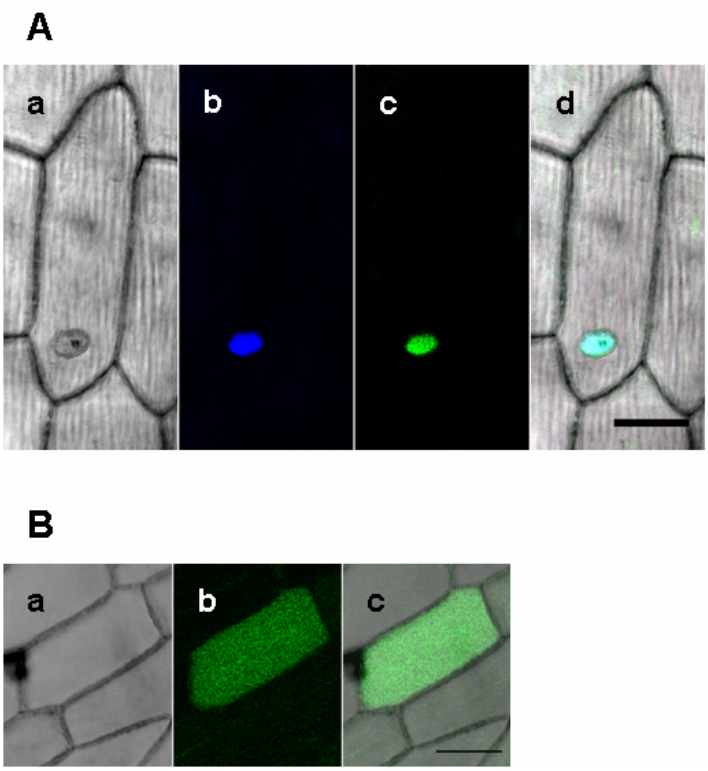
Supplementary Figure 3 A diagram of the 8.1 kb DNA fragment for transformation and the relative location of the *Ghd7* transcript. The upper bar represents the entire 8.1 kb fragment. The thick black bars represent the coding sequence of *Ghd7* with the predicted translation start site (ATG) and stop site (TAG) also indicated underneath, the thin line in between represents the intron. The thick gray bars indicate the 5' and 3' untranslated regions. The red dashed lines in both ends indicate the DNA fragments outside the transcribed sequence. The lower bar indicates the position in bp starting at the first nucleotide in the transcript determined by rapid amplification of the cDNA ends (RACE). TAG above the upper bar indicates a mutation at position 367 changing the codon GAG in Minghui 63 to a stop codon in Mudanjiang 8 and Hejiang 19.



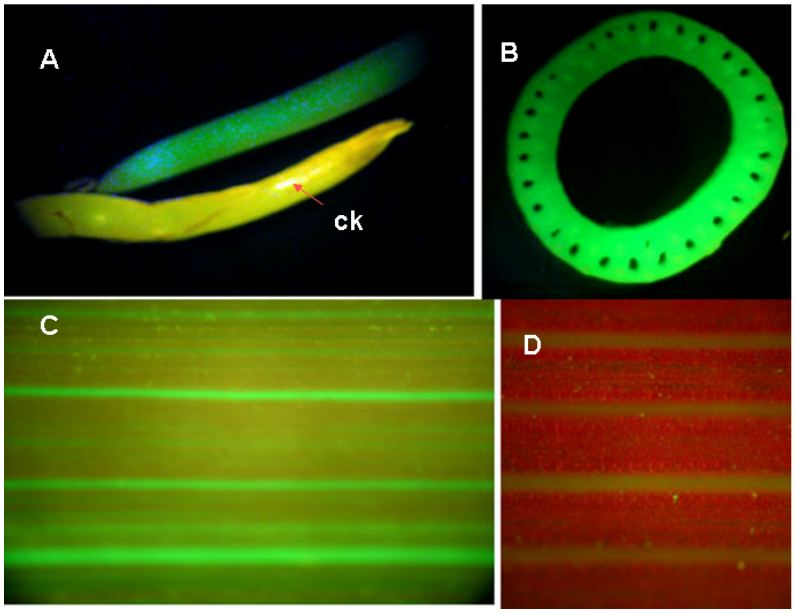
Supplementary Figure 4 Comparison of the *Ghd7* transgenic and wild type plants developed using three recipient lines. (A) Mudanjiang 8: transgenic T_0 (left) and wild type (right). The upper right shows the panicles of transgenic (left) and wild type (right) plants. (B) Hejiang 19: transgenic T_1 (right) and wild type (left). The lower right shows the panicles of transgenic (right) and wild type (left) plants. (C) Nipponbare: right, transgenic T_2; left, wild type. The lower left shows the panicles of transgenic (right) and wild type (left) plants.



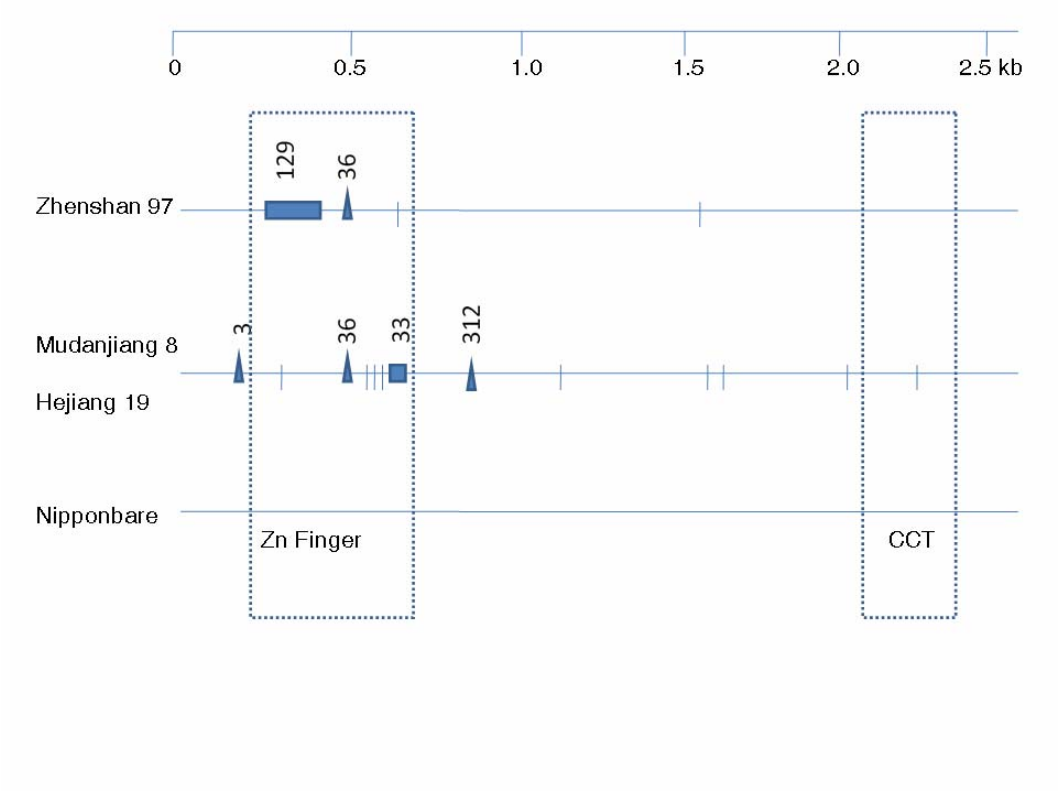
Supplementary Figure 5 The predicted protein sequences of GHD7 for five *Ghd7* alleles. The area shaded in gray (aa189-233) is the putative CCD domain. The stretch of letters in blue indicates the two motifs of putative nuclear localization signals KRKK starting at aa199 and RKKR at aa200 identified by a PSORT analysis (http://wolfpsort.org).



Supplementary Figure 6 Nuclear localization of GHD7. (A) Onion epidermal cells were transformed with *Ubq*::*Ghd7*:*GFP* by particle bombardment, stained with 4'-6-diamidino-2-phenylindole (DAPI) and observed using a confocal microscope under various conditions. (a) A differential interference contrast image of an epidermal cell. (b) The same cell showing the DAPI stained nuclei. (c) The same cell showing the nuclear localization of GHD7-GFP fluorescence. (d) The merged image of a, b and c. (B) Images derived from a transformants with a construct *Ubq*::*GFP*. (a) A differential interference contrast image of an epidermal cell. (b) The localization of the GFP fluorescence. (c) The merged image of a and b. Bar, 50 μm.



Supplementary Figure 7 Pattern of GFP expression driven by the *Ghd7* promoter. The GFP signal is strong in stems at booting stage (B), and blades of fully expanded leaves (C), but weak in root (A), and sheaths of fully expanded leaves (D). For this assay, a 2.45 kb promoter fragment upstream the translation start codon ATG of *Ghd7* was obtained by PCR using primers Promoter-F and Promoter-R (Supplementary Table 8). The PCR product was constructed into pCAMBIA1381a, in which the coding sequence of *GUS* was replaced by *GFP*, and the construct was introduced into Zhonghua 11. A T_2 plant was assayed for GFP signal.



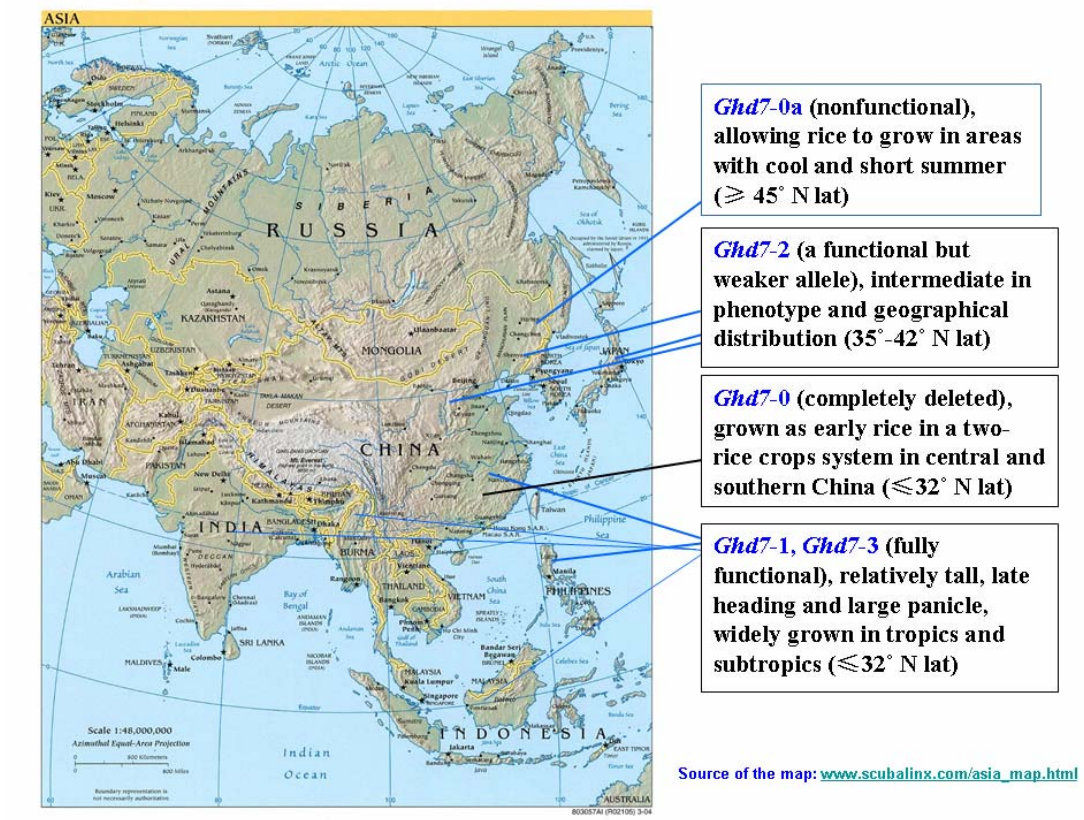
Supplementary Figure 8 Comparison of the coding sequences of *Hd1* in Zhenshan 97, Mudanjiang 8 and Hejiang 19 with reference to Nipponbare. The blocks indicate deletions, and triangles represent insertions. Numbers indicate the numbers of nucleotides inserted/deleted. Vertical thin bars indicate single nucleotide substitutions relative to Nipponbare. The zinc finger and CCT domains are also indicated.

```
Zs  MAGSGRDRDPLVVGRVVGDVLDAFVRSTNLKVTYGSKTVSNGCELKPSMVTHQPRVEVGG  60
Nip MAGSGRDRDPLVVGRVVGDVLDAFVRSTNLKVTYGSKTVSNGCELKPSMVTHQPRVEVGG
Hj  MAGSGRDRDPLVVGRVVGDVLDAFVRSTNLKVTYGSKTVSNGCELKPSMVTHQPRVEVGG
    ************************************************************

Zs  NDMRTFYTLVMVDPDAPSPSDPNLREYLHWLVTDIPGTTAASFGQEVMCYESPRPTMGIH  120
Nip NDMRTFYTLVMVDPDAPSPSDPNLREYLHWLVTDIPGTTAASFGQEVMCYESPRPTMGIH
Hj  NDMRTFYTLVMVDPDAPSPSDPNLREYLHWLVTDIPGTTAASFGQEVMCYESPRPTMGIH
    ************************************************************

Zs  RLVFVLFQQLGRQTVYAPGWRQNFNTKDFAELYNLGSPVAAVYFNCQREAGSGGRRVYN  180
Nip RLVFVLFQQLGRQTVYAPGWRQNFNTKDFAELYNLGSPVAAVYFNCQREAGSGGRRVYP
Hj  RLVFVLFQQLGRQTVYAPGWRQNFKTKDFAELYNLGSPVAAVYFNCQREAGSGGRRVYP
    ************************:*********************************:
```

Supplementary Figure 9 Comparison of the predicted protein sequences of HD3A in Zhenshan 97 (Zs), Mudanjiang 8 and Hejiang 19 (Hj) with reference to Nipponbare. Mudanjiang 8 and Hejiang 19 have identical sequences.



Supplementary Figure 10 Geographical distribution of the *Ghd7* alleles in rice-growing areas of Asia. See Supplementary Table 6 and Fig. 5 for varieties carrying these alleles and their sequence information.