Studies of rice *Hd1* haplotypes worldwide reveal adaptation of flowering time to different environments

Cheng-Chieh Wu

Institute of Plant and Microbial Biology Academia Sinica

Fu-Jin Wei

FourOneFour Studio

Wan-Yi Chiou

Institute of Plant and Microbial Biology Academia Sinica

Yuan-Ching Tsai

Department of Agronomy, National Chiayi University

Hshin-Ping Wu

Institute of Plant and Microbial Biology Academia Sinica

Dhananjay Gotarkar

Institute of Plant and Microbial Biology Academia Sinica

Ha Thi Quynh

Faculty of Agronomy, Vietnam National University

Nguyen Quoc Trung

Faculty of Biotechnology, Vietnam National University

Ming-Hsin Lai

Taiwan Agricultural Research Institute

Yue-le Caroline Hsing (✉ bohsing@gate.sinica.edu.tw)

Academia Sinica <https://orcid.org/0000-0002-2900-7068>

Original article

Keywords: Champa rice, domestication/adaptation, heading date, natural variation, rice, retrotransposon transposition, weedy rice

Posted Date: April 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-23670/v1>

License: (cc) (i) This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Plant domestication/adaptation is a good model for evo-devo studies. Mutations that caused morphological and physiological change, followed by human selection, finally led to improvement of phenotypes suitable for different kinds of environments. Originating from the Yangtze Valley, rice is a facultative short-day plant. Rice spread southward thousands of years ago, but one of the new traits beneficial to crop yield would be loss of sensitivity to photoperiod. That is, if rice could be cultivated 2 or 3 seasons each year, the production would at least double.

Results

We used the sequence information for *Heading date 1 (Hd1)* gene to reveal the relationship of sequence changes and flowering phenotypes of rice in different regions. Seven loss-of-function *hd1* haplotypes were reported previously. By data mining the genome sequencing information in the public domain, we discovered another 3 types. Allele haplotypes are present in sub-tropical and tropical regions, which indicates human selection. Two of these alleles, types 7 and 13, must have occurred early in southern Asia and then were introgressed in many local landraces. According to the rice 3K database, more than one-third of the world's rice accessions contain these 2 loss-of-function haplotypes. We also demonstrate that these haplotypes are present in weedy rice populations, again indicating that these alleles were present in rice cultivation for long time. In comparing the wild rice sequence information, these loss-of-function haplotypes occurred *in agro* but were not from wild rice.

Conclusion

In the current study, we analyzed how sequence changes in a rice flowering-control gene occurred, were selected and were retained during rice cultivation. Many loss-of-function *hd1* alleles have existed in sub-tropical and tropical Asia rice-growing areas for a long time. Some of these haplotypes were present locally, but 2, types 7 and 13, were spread in many regions and are now included in most of the modern varieties in southern Asia.

Background

Rice (*Oryza sativa*) is one of the most important crops in the world. It is a short-day (SD) plant and was domesticated originally in a temperate zone, then brought to sub-tropical and tropical zones with warmer temperature and different photoperiod. One of the main reasons for the spread of rice cultivation to a wide range in Asia as well as the increase in production is the diversification of flowering time (Khush, 1997).

Flowering time, also known as heading date for rice, is an important trait in the cultivation period (Takahashi and Shimamoto, 2011). Flowering time is affected by photoperiod (day length) and also temperature (Luan et al., 2009). As a facultative SD plant, rice flowering is promoted under SD conditions and delayed under

long-day (LD) conditions, with the critical day length around 13 hours (Vergara and Chang, 1985; Itoh et al., 2010). Several recent reviews provided detailed information on the regulation of rice flowering (for instances, Itoh and Izawa, 2013; Tsuji et al., 2013; Shrestha et al., 2014). A few key genes have been suggested to affect flowering time, including *Heading date 1 (Hd1)*, a B-box zinc finger protein and the ortholog of Arabidopsis CONSTANS (CO; Yano et al., 2000); *Early heading date (Ehd1)*, a B-type response regulator with no ortholog in Arabidopsis (Doi et al., 2004); *Hd3a*, a phosphatidylethanolamine-binding protein and ortholog of Arabidopsis *FLOWERING LOCUS T* (FT; Tamaki et al., 2007); and *Grain number, plant height, and heading date 7 (Ghd7)*, an important regulator of heading date as well as yield potential (Xue et al., 2008).

In the temperate area, including Japan, Korea, northern China and Europe, rice crops grow once a year (i.e., they are germinated/transplanted in early summer and harvested in late fall). However, in sub-tropical and tropical areas, where the temperature is suitable for plant growth from February to November or even the whole year round, changes in the sensitivity to day length played important roles during rice domestication/adaptation. If the cropping season each year may be increased from 1 to 2 or 3, the total seed production would double or even triple. Nowadays there are 2 cropping seasons per year in Taiwan and Vietnam and 2 or 3 in the Philippines, Thailand, India, and Pakistan. Thus, these varieties must be insensitive to photoperiod.

Decades ago, a large variation in flowering time, also known as heading date (Hd), was discovered among cultivated rice varieties (Vergara and Chang 1985). That is, the rice flowering is controlled by quantitative trait loci (QTL). Using the progeny derived from a single cross between a *japonica* rice variety, Nipponbare, and an *indica* variety, Kasalath, Yano and colleagues identified a total of 15 QTL for Hd: 5 (*Hd1* to *Hd5*) were mapped by QTL analysis of an F₂ population (Yano et al. 1997), and 3 (*Hd7*, *Hd8*, and *Hd11*) were discovered with BC₁F₅ lines (Lin et al. 1998). Seven loci, *Hd6*, *Hd9*, *Hd10*, *Hd12-Hd15*, were detected by using advanced backcross generations (Yamamoto et al. 2000; Lin et al. 2002). The molecular genetic pathway for 2 model plants, Arabidopsis (an LD plant) and rice (an SD plant), has been addressed in many studies, and the comparison has been discussed recently (e.g., Cho et al., 2017; Hori et al., 2016; Shrestha et al., 2014).

Rice *Hd1* gene was identified as Arabidopsis orthologue *CO* and encodes a zinc-finger type transcriptional activator containing CO, CO-like, and TOC1 (CCT) domains (Yano et al., 2000). It contains 2 exons, with the CCT domain at the second exon. This CCT domain functions as a nuclear localization signal, and the mutant without CCT domain in the Arabidopsis *CO* induced a defect in the protein function (Robson et al., 2001). Rice *Hd1* was also named *Se1*, *SE1* or *Photosensitivity 1*. The functional Hd1 protein is required for suppressing flowering under LD conditions and for promoting flowering under SD conditions. *Hd1* regulates *Hd3a* expression (Takahashi et al., 2009), the rice orthologue of Arabidopsis *FT*, which encodes a mobile flowering signal (Tamaki et al., 2007). However, *Ehd1*, promoting SD flowering in the loss-of-function *hd1*, encodes a B-type response regulator but has no ortholog in the Arabidopsis genome. Variations in Hd1 protein, *Hd3a* promoters, and *Ehd1* expression contribute to the diversity of flowering time (Takahashi et al., 2009).

In the present study, we analyzed the *Hd1* gene sequence changes by using the 3K rice genome sequence database and sequence information for several local rice accessions as well as weedy rice and wild rice.

Together with the haplotypes discovered previously, in total, there are 9 loss-of-function *hd1* haplotypes. The sequence information was also compared with the flowering date data, along with locations of these rice accessions. In addition, we performed evo-devo analysis of these haplotypes. We show that sub-tropical and tropical Asian countries selected different mutations in this gene during rice cultivation and suggest that some types occurred quite early.

Results

Several gene synonyms for *Hd1*

Flowering time is one of the important traits for rice growth and production and many studies have focused on this topic. Several names for *Hd1* have been used in the rice community. In RAP-DB (Itoh et al., 2007), *Hd1* gene symbol synonyms include *SE1*, *se(t)*, *HD1*, *Fl*, *Lm*, *Se1*, *K*, *Rs*, *Lf*, *Se-1*, *Hd1(t)*, *qHD1(t)*, *OsA*, *OsBBX18*, *Hd1/OsA*, and *OsCCT21*, and gene name synonyms include *PHOTOSENSITIVITY 1*, *Photosensitivity1*, *Photosensitivity 1*, *Photoperiod sensitivity 1*, *Heading date*, *Photoperiod-sensitivity-1*, *HEADING DATE 1*, *Arabidopsis CONSTANS(CO) gene ortholog*, *ortholog of Arabidopsis CONSTANS*, *B-box-containing protein 18*, *CCT domain-containing gene 21*, *CCT (CO, CO-LIKE and TOC1) domain protein 21*, and *CCT domain protein 21*. In the Committee on Gene Symbolization, Nomenclature and Linkage (CGSNL) of Gramene (McCouch, 2008), the gene symbol is *SE1* and the gene name is *PHOTOSENSITIVITY 1*.

Many loss-of-function haplotypes are present in rice accessions

Shimamoto and coworkers identified a high degree of sequence diversity in rice *Hd1* and grouped them into 17 types (Takahashi et al., 2009). The previously recognized Kasalath haplotype (Yano et al., 2000), containing a 2-bp deletion and thus a loss-of-function mutant, was classified as type 13. In addition, the authors discovered 4 new mutations leading to loss-of-function *hd1*, including 3 deletions and 1 single nucleotide polymorphism (SNP). Both types 2 and 3 contained a 1-bp deletion, type 7 a 4-bp deletion, and type 12 an SNP inducing an early stop.

In the first work for *hd1*, Yano et al. reported 2 haplotypes that induced loss-of-function *hd1*: the 2-bp deletion in Kasalath (*hd1*) and the 43-bp deletion (*se1*) in HS66, a mutant of a Japanese landrace Ginbozu located at the short arm of chromosome 6 (Yano et al., 2000). This 43-bp deletion was not grouped previously, and here we classified it as type 18. Yoshimura and coworkers worked on *Ehd1* and *Hd1* genes of a photoperiod-insensitive variety, Taichung65 (TC65) (Doi et al., 2004). The TC65 haplotype of *hd1* contains a 1.9-kb insertion at exon 2. However, this TC65 haplotype was not mentioned in Shimamoto's nomenclature (Takahashi et al., 2009). In the current study, we classified it as type 19. In our previous study, we demonstrated that loss-of-function *hd1* with this 1.9-kb insertion was introgressed from 2 Taiwan landraces, Muteka and Nakabo, to the early breeding generation of TC65 (Wei et al., 2016).

From the 3K (3000 Rice Genome Project, 2014) and *Oryza* Map Alignment Project (OMAP, Stein et al. 2018) rice genome sequencing information, we further discovered 3 new loss-of-function *hd1* types: type 20 was an A-to-T SNP at position 9338273 and thus led to an early stop at Arg368; type 21 consisted of another SNP (C

to T) at position 9337150 and led to an early stop at Gln206; and type 22 is a C-to-A change at position 9337112 and thus an early stop at Ser193.

These types could be grouped into 22 types, with 10 mutations inducing the loss of Hd1 protein function: types 2, 3, 7, 12, 13, 18, 19, 20, 21 and 22 (Table S1). There were 3 categories of mutations: 1) a small insertion/deletion (indel) leading to a frameshift: types 2, 3, 7, 13 and 18; 2) an SNP leading to an early stop: types 12, 20, 21 and 22; and 3) a large (1.9-kb) insertion leading to a frameshift: type 19. Previously when types 1 to 17 were defined, the genome sequences of a landrace Ginbouzu were used (Takahashi et al., 2009). Here we aligned changes to the Nipponbare reference sequence, and Table 1 shows the haplotypes of the loss-of-function *hd1* gene in the 3K rice database as well as its sequence localization and changes.

The heading date information for rice grown in International Rice Research Institute (IRRI) campus, along with much phenotype data, for about two-thirds of the 3K accessions are available in the IRRI SNP-SEEK phenotype database (<https://snp-seek.irri.org/>, Mansueto et al., 2017). Type 18 belongs to a mutation in the mutagen-induced mutant population of Ginbouzu, so no phenotype data are available in SNP-SEEK. Type 22 occurred only in *O. punctata*, an African wild rice. Thus, the heading date information was downloaded and grouped according to 8 haplotypes. The ANOVA results indicate that the heading date for these 8 groups (Table S2) significantly differed from that of the wild type, which did not contain any of these 8 haplotypes. Because only one accession of type 2 contained heading date information, the heading date behavior was analyzed by *t* test for only 7 haplotypes. Figure 1 illustrates the heading date phenotype for these haplotypes and wild type. The data are in Table S3. The day length of Los Baños, on the IRRI main campus, ranges from 11 hr 17 min to 12 hr 57 min year-round; the critical day length of rice is about 12.5 hr. Because all haplotypes tested flowered before the wild type did, we confirmed that the accessions with these 7 types were indeed not sensitive to day length.

Many accessions contained types 7 or 13 as revealed by the 3K genome data

A total of 408 accessions in the 3K genome were type 7 and thus contained loss-of-function *hd1*; 99% are *indica* rice. In total, 51 accessions belong to traditional accessions: 16 were from Indonesia, 5 the Philippines, 4 Bangladesh, 4 Laos, and 3 Vietnam. All are *indica* rice (Tables 2 and 3).

In total, 772 accessions contained the type 13 *hd1* allele (i.e., the Kasalath type); 291 belong to traditional landraces: 35 are from Bangladesh, most *aus* rice; 99 from Indonesia, 61 *japonica*, 34 *indica* and 4 intermediate types; 62 from the Philippines, 47 *japonica* and 15 *indica*; 17 from Malaysia, 16 *japonica* and 1 *indica*; 16 from India, all *indica* or *aus*; 13 from Pakistan, all *aus*; 7 from Sri Lanka, 6 *indica* and 1 *japonica*; and 5 from Thailand, 3 *indica* and 2 *japonica*. All are tropical areas. In addition, many accessions belong to tropical *japonica* and are from insular countries such as Indonesia, the Philippines and Malaysia (Tables 2, 3, and S1).

The 1.9-kb insertion in type 19 is a small retrotransposon

The insertion in type 19 is 1901 bp long at exon 2. Dot matrix analysis of this sequence revealed long terminal repeats (LTRs) at both sides, 448 bp in length and with exact identity between the 2 LTRs. In

addition, the central part of this fragment contains a partial pol protein sequence, for the smallest retrotransposon in the rice genome. The cultivated rice genome contains 2 similar retrotransposons, one at chromosome 5 and another at 8, with length of 1895 bp and 1893 bp, respectively. In the Nipponbare genome, the retrotransposon in chromosome 5 is 446 bp for the right LTR and 452 bp for the left LTR, with 97% identity of these 2 LTRs. The retrotransposon in chromosome 8 is 445 bp for the left LTR and 450 bp for the right LTR, with identity 97%.

Other haplotypes occurred in local regions

Only 2 accessions in the 3K collection contained type 2 *hd1*. Both are *indica* rice and were from Vietnam. A total of 21 accessions contained type 3 *hd1*, all *indica*; 11 are from China and 5 from Taiwan. These 5 accessions were brought to Taiwan from southern China in the late Ming Dynasty about 400 years ago and thus must be old landraces. A total of 28 accessions contained type 12 *hd1*, 10 are traditional landraces. These landraces were from Laos, and all are tropical *japonica*. In total, 31 accessions type 19 *hd1*. Most are modern varieties, including TC65, Taichung 179 and Taichung 188. According to our previous study of the *hd1* gene in Taiwan aboriginal rice accessions, 11 contained this haplotype (Wei et al., 2016). A total of 16 accessions contained type 20 *hd1*, all from India and all *indica* or *aus*. Only 3 accessions contained type 21 *hd1*, all from India and all *aus* (Table S1).

Loss-of-function haplotypes also found in Chinese weedy rice accessions and an African wild rice accession

In addition to many landraces and modern varieties, much effort has been extended to resequencing the many weedy rice accessions (Qiu et al., 2017 and Sun et al., 2019 for Chinese ones; Li et al., 2017 for American ones) and wild rice accessions (Huang et al., 2012; Zhao et al., 2018; Stein et al., 2018). The sequence information for 155 Chinese accessions was downloaded from the National Center for Biotechnology Information (NCBI) for *hd1* haplotype analysis. The accessions collected from Liaoning (northeastern part) and Ningxia (northwestern part) are *japonica* and those from Jiangsu (central part) and Guangdong (southern part) are *indica* (Qiu et al., 2017). In all, 25 of the 27 Guangdong accessions contained type 7 *hd1*, one accession did not have any known loss-of-function haplotype and one accession contained heterozygous type 7 *hd1*. However, all 39 Jiangsu accessions and 30 Ningxia accessions contained only functional *Hd1*. For the 59 Liaoning accessions, 2 contained type 7 *hd1* and another 2 type 13; the remaining 55 accessions contained functional *Hd1*. The detailed analysis is in Table S4. Rice production in the Guangdong region involves 2 cropping seasons but only one cropping season in the other 3 regions.

We also performed detailed analysis of *Hd1* haplotypes with the resequencing data for wild rice accessions downloaded from NCBI. Only accessions with at least 10X genome redundancy were used, including those from 1) the *O. rufipogon* pangenome project: W0123, W0141, W0170, W1687, W1698, W1739, W1754, W1777, W1943, W1979, W2012, W3078, and W3095 (Zhao et al., 2018); and 2) the *Oryza* map alignment project: *O. barthii*, *O. glumaepatula*, *O. meidionalis*, *O. nivara*, *O. rufipogon* and *O. punctata* (Stein et al., 2019). A new loss-of-function *Hd1* haplotype was found in *O. punctata* (Table S4). This type 22 is an SNP of C to A at position 9337112, leading to an early stop of Hd1 protein.

Discussion

Insensitivity to photoperiod in tropical and sub-tropical regions has been beneficial to rice production because of 1) double the crop yield per year, 2) escape from drought stress, and 3) avoidance of flooding damage. Thus, it was one of the targets to select for heading gene mutations along with rice cultivation period. In this study, we analyzed the loss-of-function *hd1* mutations that occurred in rice in different regions of Asia by using the whole-genome sequencing data for thousands of accessions.

***Hd1* was suggested to be the main allele associated with adaptation of rice plants to tropical regions**

Rice was domesticated about 8,000 to 10,000 years ago in the Yangtze River region of China, where today there is still one cropping season. Domesticated rice had spread southward thousands of years ago, and one of the new traits beneficial to crop yield would be loss of sensitivity to photoperiod. That is, if rice could be cultivated 2 or 3 seasons each year, production would at least double.

Takahashi and Shimamoto (2011) performed a cDNA sequencing analysis of *Hd1* and *Hd3a* genes in leaf tissues at tillering stage under an SD condition from several rice cultivars and 38 wild rice accessions. Cultivated and wild rice accessions showed no nucleotide changes in *Hd3a*. Also, the 38 wild rice accessions showed no change affecting *Hd1* function, but several changes appeared in cultivated rice. Thus, the authors suggested that *Hd1* was a possible target of selection to generate different flowering-time response in different regions. Kim et al. (2018) selected about 60 diverse rice accessions to analyze the heading date trait of various types, including 4 *aus* accessions, 20 *indica* accessions, 4 tropical *japonica* accessions, 17 temperate *japonica* accessions, and 12 tropical adapted temperate *japonica* accessions. The authors analyzed 7 major flowering genes, including *Hd1* (Yano et al., 2000), *OsPPR37* (Liu et al., 2015), *DTH8* (Wei et al., 2010), *Ghd7* (Xue et al., 2008;), *Ehd1* (Doi et al., 2004), *RFT1* (Komiya et al., 2008) and *Hd3a* (Tamaki et al., 2007) as well as heading behavior under 3 different field conditions in temperate and tropical regions. They concluded that the accessions from tropical/sub-tropical regions preferred the non-functional alleles of *Hd1* but not other flowering genes tested. Therefore, in the present study, we focused on the *Hd1* haplotype.

Several *Hd1* haplotypes were limited to small regions and some were widely spread

Table 2 summarizes the traditional accessions for each haplotype in each rice-growing country in Asia. As indicated in Results, type 2 was mainly restricted to *indica* rice in Vietnam and type 3 mainly *indica* rice in China and Taiwan. All type 12 mutations were in tropical *japonica* rice from Laos. Type 19 mutation occurred in upland accessions from Taiwan aboriginal villages and was then introgressed in a modern variety TC65. Types 20 and 21 were both from India, the former being *indica* rice and the latter *aus* rice. These accession numbers were relatively small and distributions were limited. Thus, mutations of these 6 loss-of-function haplotypes must have occurred recently during adaptation, and all were in tropical or sub-tropical regions. Most were *indica* or *aus* rice, with one tropical *japonica* (type 12). Figure 2 illustrates the size and frequency of traditional landraces for each haplotype in rice-growing countries in Asia and clearly shows that no or few numbers or types of loss-of-function *Hd1* haplotypes occurred in high-altitude and -latitude regions such as Bhutan, Nepal, Korea and Japan. However, most of the accessions from tropical/sub-tropical regions preferred the non-functional alleles of *Hd1* gene, with multiple haplotypes in the same region.

Types 7 and 13 were old mutations and had introgressed to many landraces

The traditional landraces containing loss-of-function haplotypes 2, 3, 12, 19, 20 and 21 are relatively limited to small areas, which was not the case for types 7 and 13 (Table 2). The numbers for Chinese accessions are inside brackets because of no information on traditional or modern accessions.

Table 3 shows types 7 and 13 *Hd1* alleles in the traditional landraces of each country, including their subspecies. Most of the type 7 accessions are *indica* rice, with more *japonica* rice for type 13 accessions, especially in Indonesia and the Philippines. Both haplotypes are still used in many modern *indica* varieties worldwide, such as most IR series inbred lines. However, the high percentage of *japonica* type 13 in insular southern Asia suggested that the mutation might have occurred first in this area and then introgressed to many *indica* landraces locally, followed by movement to other regions.

TC65 haplotype: retrotransposon integration occurred recently

For the 1.9-kb small retrotransposon integrated into *Hd1* of type 19, because the length and sequences of both LTRs are identical, this transposition into chromosome 6 must have occurred around 2,000 years ago. Two similar retrotransposons at chromosome 5 and 8 had different lengths for the LTR pair, and the sequences were not identical (97% for each), so the transposition into the current location must have occurred a long time ago. In addition, these 2 small retrotransposons were located in chromosomes 5 and 8 in *indica* rice accessions, which further confirmed that they had been present in the rice genome before the split of *japonica* and *indica* rice. For the Asian AA-genome wild rice, *O. nivara* has the chromosome 5 copy only, and *O. rufipogon* has both chromosome 5 and 8 copies.

Champa rice

According to the *History of Song Dynasty*, there was a severe drought in Yangtze and Hui river valley areas and in the eastern and western Zhejiang paddy rice area in 1012 (*History of Song Dynasty* 1, 1343). The emperor Zhenzong (a.d. 992-1022) heard that Champa rice was drought-resistant, early maturing, and non-photoperiod-sensitive (Ho, 1956). He sent governors to Fujian Province and brought 30,000 bushels of Champa seeds and distributed them to farmers in the drought area. This Champa Kingdom was located in what is now central Vietnam. According to the history, as compared with Chinese rice, this Champa rice had longer panicles, no awns, less difference in grain size, could grow well in marginal land; and was drought-resistant and not sensitive to photoperiod (Ho, 1956; *History of Song Dynasty* 2, 1343). Therefore, Champa rice, with many good traits, came from Vietnam to Fujian first, and later to the Yangtze region during the early Song Dynasty in the 11th century. Barker (2011) explored the origin and spread of Champa rice. Susan McCouch, at Cornell University, said that she found a Champa rice in her collection at Cornell and the DNA test indicated it was *aus* rice (Barker, 2011).

Hundreds of landraces were brought from southern China to Taiwan by the Han people during the late Ming to early Ching dynasty about 400 years ago. They were propagated and maintained first by local farmers and then by rice breeders during Japanese colonial time. All these accessions were *indica* and are now stored in the Germplasm Center, Taiwan Agricultural Research Center (TARI). Dee-Geo-Woo-Gen, one of the famous parental lines of IR8, was one of the landraces. These accessions were old landraces in China

hundreds of years ago, but most, including Dee-Geo-Woo-Gen, are not present in the current Chinese germplasm.

We found 7 accessions that contained the term “Champa,” pronounced Chan, in the TARI collection, including Hei Chan, Pai Jih Tung Chan, Pai Ko Pu Chan, Ching Kuo Chan, Wu Chan, Hsin Chu Liu Chan and Lung Ya Chan. All are *indica*. Three contained loss-of-function *Hd1* haplotypes: Lung Ya Chan is type 13, and Wu Chan and Hsin Chu Liu Chan are type 7. Thus, again, types 7 and 13 have been present in Eastern Asia for hundreds or thousands of years.

We argued that the Champa rice is not *aus* because of the following: 1) a group of 826 accessions representing the diversity of Chinese rice germplasm stored in the China National Rice Research Institute was used for genetic diversity and classification studies (Wang et al., 2014). Using 84 nuclear simple sequence repeat markers, the authors identified 3 groups: temperate *japonica*, tropical *japonica* and *indica*, with no *aus* type in the Chinese rice collection. 2) Many landraces with the term “Champa” in the name were brought to Taiwan about 400 years ago. We performed whole-genome sequencing of these “Champa” rice and also 3 landraces from central Vietnam, where the ancient Champa kingdom was. Figure 3 shows the phylogenetic tree of these Champa rice accessions along with some traditional accessions of *japonica*, *indica* and *aus* in Asia. A total of 48 accessions were used, and their types and collection locations are in Table S4. The accessions from central Vietnam (highlighted in blue) and those with the word Chan in the name from Taiwan (highlighted in green) are clustered with *indica* rice but not *aus* (highlighted in red).

To conclude, many *indica* rice accessions were introduced to the Yangtze River region in China about 1000 years ago, according to the *History of Song Dynasty*. Some points can be summarized: 1) the introduction of Champa rice was the beginning of *indica* rice cultivation in China and 2) there may have been no photoperiod-insensitive accession in China during that time.

A final question about Champa rice is Which *Hd1* haplotypes do they have? Table S4 lists the information. Only 3 accessions contained the loss-of-function haplotype: Hsin Chu Liu Chan and Wu Chan contained type 7 and Lung Ya Chan type 13. Some of the current Chinese weedy rice accessions also contained type 7 or 13, again indicating that these 2 loss-of-function alleles have been present in China for long time.

The origin of the loss-of-function mutations in *Hd1* gene

Several independent origins of *Hd1* loss-of-function mutations led us to raise interesting questions about the evolution/adaptation of the *Hd1*-negative phenotype. One novel mutation, type 22, existed only in an African wild rice accession, and none of the other haplotypes preexisted in the wild rice accessions tested. Thus, most mutations, except type 22, did not originate in wild species but occurred *in agro*. For the remaining types in the 3K database (i.e., types 2, 3, 7, 12, 13, 19, 20, and 21), the question is whether the mutations underwent strong selection. From the information on traditional landraces in the 3K database, Table 2 lists their existence in each country. Only total accessions are listed for Chinese accessions because of no information on modern or landraces. Type 2 is only in Vietnam; type 3 is in China, India and Taiwan; type 10 is in Laos; type 19 is mainly in Taiwan; and types 20 and 21 are only in India. However, the mutations of

types 7 and 13 must have occurred a long time ago because the total numbers are high and the distribution is wide, as shown in Table 3.

We then asked about any positive selection on some of these mutations. The distribution of types 12 and 19 are local, with relatively more accessions (≥ 10). We checked selection parameters, including π (Tajima, 1983), θw (Watterson, 1975), as well as Tajima's D (Tajima, 1989), for these types as well as types 7 and 13 to test the neutral mutation hypothesis. *Hd1* gene and the upstream 10-kb and downstream 10-kb regions were used for the calculation. A significant negative D value indicates strong selection. We analyzed 5 haplotypes in 22 regions. Data are summarized in Table 4 and the complete data are in Table S5. Type 7 was positively selected in Indonesia but not in Bangladesh, India, Laos, Malaysia, the Philippines or Taiwan. Thus, this mutation occurred in the *indica* rice accession in Indonesia. Type 12 was present only in Laos and it was positively selected. The mutation of type 19 occurred thousands of years ago in the aboriginal villages in Taiwan (Wei et al., 2016). Eleven accessions contained the 1.9-kb insertion. The parameters shown in Table 4 also indicate positive selection in these lines.

A total of 772 accessions in the 3K database were type 13, and 27% were landraces: 99 were from Indonesia and 61 from the Philippines (Table 3). In addition, more than half were *japonica* even though the current cultivated rice accessions in that area are *indica*. Again Table 4 illustrates highly significant selection ($P < 0.001$) in the landraces of Indonesia and the Philippines, especially in *japonica* rice. The selection was also significant ($P < 0.05$) in Malaysia *japonica* rice. However, the selection was not obvious in Bangladesh or India, etc. (Table S5). Thus, this 2-bp mutation must have occurred in *japonica* accessions in insular countries, either Indonesia or the Philippines, a long time ago, and then in Malaysia. The 3 regions contain some but low numbers of *indica* landraces, so the introgressions occurred later on, and thus some *indica* accessions gained this loss-of-function trait. These *indica* accessions were then brought to the Indochina Peninsula area because most, if not all, were *indica* accessions. This haplotype is present in many modern or traditional races in sub-tropical and tropical countries of eastern and southeastern Asia where there are 2 or 3 cropping seasons.

Some weedy rice or wild rice accessions also contained loss-of-function *Hd1*

Weedy rice (*O. sativa f. spontanea*), also called “red rice”, has been considered a conspecific weed of cultivated rice. Most of them are shattering, with a long awn, brown pericarp, and dark seed hull. Several origins for weedy rice have been suggested, including gene flow resulting from natural hybridization (Ishikawa et al., 2005), evolving directly from within domesticated lineages (Reagon et al., 2010), introgression with local cultivars coupled with selection that maintained weedy identity (Sun et al., 2013), influenced by proximity to reproductively compatible wild and domesticated populations (Song et al., 2014), originated from *indica-japonica* hybridization (Qiu et al., 2014), hybridization of modern *indica/indica* or *japonica/japonica* (He et al., 2017), or evolved from maternal hybrid rice derivatives (Zhang et al., 2015). As mentioned in the Results section, data mining of the Chinese weedy rice accessions illustrates that loss-of-function *Hd1* type 7 occurred in 96% of the accessions collected in Guangdong, a sub-tropical region. However, most of the accessions collected in another 3 locations (all temperate regions) contained

functional *Hd1*. Because of the ecological meaning of the weedy rice, type 7 *Hd1* must have existed in southern China, including Guangdong, for a long time.

With the detailed studies of several wild rice accessions, we also found 1 loss-of-function haplotype present in *O. punctata*, a BB genome African wild rice. Therefore, sequence changes leading to loss-of-function *Hd1* occurred not only in cultivated rice but also in its wild relatives. In addition, the current 8 loss-of-function haplotypes occurred in the field but were not transferred from wild relatives.

Conclusions

Here, we have demonstrated that many loss-of-function *hd1* alleles existed in sub-tropical and tropical Asia rice-growing areas for long time. Some of these haplotypes were present locally, whereas 2, types 7 and 13, were spread in many regions and are now used in most of the modern varieties in southern Asia.

Materials And Methods

Whole-genome sequencing and data interpretation

Genomic DNA from rice plants was extracted from healthy leaves of a single-seed–descent plant by using the DNeasy Plant Mini Kit (Qiagen). After quality assessment, genomic DNA was randomly fragmented and size-fractionated. DNA fragments with the desired lengths were gel-purified. For whole-genome resequencing, paired-end libraries with 450- to 500-bp inserts were constructed and sequenced by using the HiSeq2000 system (Illumina). Adaptor sequences, low-quality bases and reads < 20-bp long were discarded. The trimmed paired reads were mapped against the Os-Nipponbare-Reference-IRGSP-1.0 (IRGSP, 2005; Kawahara et al, 2013). SAMtools and VCFtools were used to manipulate and transform the sequence alignment/map format (SAM) and variant call format (VCF) of the file. To detect SNPs and small indels, we used the command lines in the section “Variant Calling” in “Workflows” of the SAMtools manual without any restriction on depth or mapping quality. The information on SNPs and small indels was recorded in VCF files. The sequence data for all Champa-related landraces were deposited into the NCBI Sequence Read Archive.

Allele Determination of *Hd1* haplotypes

The functional impact of nucleotide variants was analyzed by using the rice genome sequencing data with SnpEff (Cingolani et al., 2012). From the genome annotation, sequence variants were classified according to their location (open reading frame, intron, splice sites, etc.) and predicted functional impact (missense, frame shift, early stop, etc.).

Estimation of diversity of different *Hd1* haplotypes

DNA sequences were aligned by using MUSCLE (Edgar, 2004a, b). The –10- to +10-kb region of *Hd1* genes, corresponding to Nipponbare genome chromosome 6 from 9,326,376 to 9,348,569, was used for the analysis. Statistical analysis involved using DNAsp.v6 (Rozas et al., 2017). Items analyzed included number of polymorphic (segregating) sites, *S*; total number of mutations, *Eta*; average number of nucleotide differences, *k*; nucleotide diversity, π (Tajima, 1983); *theta* (per sequence) from *Eta* θ ; and Watterson’s

estimator of θ (per site) from θ_w (Watterson, 1975). The neutrality test, Tajima's D value (Tajima, 1989), was used to test the neutral mutation hypothesis. The D value was based on the discrepancy between π and θ_w . Thus, negative values indicate excess low-frequency polymorphism. These values were calculated after removing missing data and alignment gaps.

Phylogenetic analysis

To reveal the position of the Champa rice accessions relative to other Asian rice, we performed a phylogenetic analysis with next-generation sequencing (NGS) data. Table S6 lists the names, types, origins and sequence information for these lines. The clean reads were mapped to the Nipponbare reference genome (IRGSP v1.0) by using BWA v0.7.13-r1126 mem with default parameters (Li and Durbin, 2010, Kawahara et al., 2013). The mapped results were merged and data with low mapping quality ($q < 20$) were removed as BAM files by using Samtools v1.3 (Li et al., 2009, Li, 2011). Picard v2.1.1 MarkDuplicates was used to identify and remove duplicate reads in the same DNA fragments (<http://picard.sourceforge.net>). The Genome Analysis Toolkit v3.5-0-g36282e4 RealignerTargetCreator was used to identify regions around indels, then the Genome Analysis Toolkit IndelRealigner was used for local realignment (McKenna et al., 2010). Samtools and Bcftools were used for variant calling including SNPs and indels with filter by depth and mapping quality. Genetic distance with the p-distances model was calculated, and a neighbor-joining tree was constructed with 1,000 bootstraps by using PHYLIP v3.695 (<http://evolution.genetics.washington.edu/>). MEGA v7 (Kumar et al., 2016) was used to display the phylogenetic tree.

Supplementary Information

Additional file 1: Table S1. Accessions, types and collected nations and source for 10 loss-of-function *hd1* gene in the 3K database.

Additional file 2: Table S2. ANOVA for flowering dates of the *hd1* haplotypes and wild type. **Table S3.** Summary of statistical analysis of flowering dates of the *hd1* haplotypes and wild type. **Table S5.** Selection sweep analysis of 5 *hd1* Types in several countries. The *hd1* gene region and nearby ± 10 kb of traditional landraces were used for calculation.

Additional file 3: Table S4. Accessions, types and collected regions and source for 10 loss-of-function *hd1* gene in Champa rice, weedy rice and wild rice.

Additional file 4: Table S6. Accessions, types, collected regions and sequence accessions used for Champa rice phylogeny analysis.

List Of Abbreviations

CGSNL, Committee on Gene Symbolization, Nomenclature and Linkage; CO, CONSTANS; *Ehd1*, *Early heading date*; FT, *FLOWERING LOCUS T*; *Ghd7*, *Grain number, plant height, and heading date 7*; Hd, heading date; *Hd1*, *Heading date 1*; indels, insertions/deletions; IRRI, International Rice Research Institute; LD, long-

day; LTRs, long terminal repeats; NCBI, National Center for Biotechnology Information; NGS, next-generation sequencing; OMAP, *Oryza* Map Alignment Project; QTL, quantitative trait loci; SAM, sequence alignment/map format; SD, short-day; SNP, single nucleotide polymorphism; TARI, Taiwan Agricultural Research Center; TC65, Taichung65; *VCF*, *variant call format*

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

All data generated or analyzed during this study are included in this article and the supplementary information files.

Competing interests

The authors declare that they have no competing interests

Funding

This project was supported by funding from Ministry of Science and Technology Grant MOST 107-2313-B-001-007-MY2 to YICH, and Academia Sinica Investigator Award to YICH.

Authors' contributions

YICH designed the study and wrote the manuscript. CCW, FYW, WYC, HPW, DG

performed bioinformatics analysis of the 3K data sets. CCW, HTQ, TNQ

analyzed Champa rice sequences; YCT, DG, MHL performed phenotype analysis.

Acknowledgements

We thank Ms. Lie-Hong Wu for maintenance of greenhouse plants and Ms. Laura Smales (BioMedEditing, Toronto, Canada) for English editing.

References

3,000 Rice Genomes Project. (2014) The 3,000 rice genomes project. *GigaScience* 3 (1):2047-2217X-2043-2047

Barker R (2011) The origin and spread of early-ripening champa rice: It's impact on Song Dynasty China. *Rice* 4 (3):184

Cho LH, Yoon J, An G (2017) The control of flowering time by environmental factors. *The Plant Journal* 90 (4):708-719

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6 (2):80-92

Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, Yano M, Yoshimura A (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 & development* 18 (8):926-936

Edgar RC (2004a) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* 5 (1):113

Edgar RC (2004b) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32 (5):1792-1797

He Q, Kim KW, Park YJ (2017) Population genomics identifies the origin and signatures of selection of Korean weedy rice. *Plant biotechnology journal* 15 (3):357-366

History of Son Dynasty 1 (1343) *History of Son Dynasty* 8:151

History of Son Dynasty 2 (1343) *History of Son Dynasty* 173: 4162

Ho P-T (1956) Early-ripening rice in Chinese history. *The Economic History Review* 9 (2):200-218

Hori K, Matsubara K, Yano M (2016) Genetic control of flowering time in rice: integration of Mendelian genetics and genomics. *Theoretical and applied genetics* 129 (12):2241-2252

Huang X, Kurata N, Wang Z-X, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490 (7421):497-501

Huang X, Lu T, Han B (2013) Resequencing rice genomes: an emerging new era of rice genomics. *Trends in Genetics* 29 (4):225-232

International RGSP (2005) The map-based sequence of the rice genome. *Nature* 436 (7052):793

Ishikawa R, Toki N, Imai K, Sato Y, Yamagishi H, Shimamoto Y, Ueno K, Morishima H, Sato T (2005) Origin of weedy rice grown in Bhutan and the force of genetic diversity. *Genetic Resources and Crop Evolution* 52 (4):395-403

Itoh H, Izawa T (2013) The coincidence of critical day length recognition for florigen gene expression and floral transition under long-day conditions in rice. *Molecular plant* 6 (3):635-649

- Itoh H, Nonoue Y, Yano M, Izawa T (2010) A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. *Nature genetics* 42 (7):635
- Itoh T, Tanaka T, Barrero RA, Yamasaki C, Fujii Y, Hilton PB, Antonio BA, Aono H, Apweiler R, Bruskiewich R (2007) Curated genome annotation of *Oryza sativa* ssp. japonica and comparative genome analysis with *Arabidopsis thaliana*. *Genome research* 17 (2):175-183
- Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6 (1):4
- Khush GS (1997) Origin, dispersal, cultivation and variation of rice. *Plant molecular biology* 35 (1-2):25-34
- Kim S-R, Torollo G, Yoon M-R, Kwak J, Lee C-K, Prahalada G, Choi I-R, Yeo U-S, Jeong O, Jena KK (2018) Loss-of-Function Alleles of Heading Date 1 (Hd1) Are Associated With Adaptation of Temperate Japonica Rice Plants to the Tropical Region. *Frontiers in plant science* 9:1827
- Komiya R, Ikegami A, Tamaki S, Yokoi S, Shimamoto K (2008) Hd3a and RFT1 are essential for flowering in rice. *Development* 135 (4):767-774
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* 33 (7):1870-1874
- Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27 (21):2987-2993
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26 (5):589-595
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16):2078-2079
- Li L-F, Li Y-L, Jia Y, Caicedo AL, Olsen KM (2017) Signatures of adaptation in the weedy rice genome. *Nature genetics* 49 (5):811-814
- Lin H, Ashikari M, Yamanouchi U, Sasaki T, Yano M (2002) Identification and characterization of a quantitative trait locus, Hd9, controlling heading date in rice. *Breeding Science* 52 (1):35-41
- Lin S, Sasaki T, Yano M (1998) Mapping quantitative trait loci controlling seed dormancy and heading date in rice, *Oryza sativa* L., using backcross inbred lines. *Theoretical and Applied Genetics* 96 (8):997-1003
- Liu C, Song G, Zhou Y, Qu X, Guo Z, Liu Z, Jiang D, Yang D (2015) OsPRR37 and Ghd7 are the major genes for general combining ability of DTH, PH and SPP in rice. *Scientific reports* 5 (1):1-11
- Luan W, Chen H, Fu Y, Si H, Peng W, Song S, Liu W, Hu G, Sun Z, Xie D (2009) The effect of the crosstalk between photoperiod and temperature on the heading-date in rice. *PLoS One* 4 (6)

Mansueto L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, Chebotarov D, Sanciangco M, Palis K, Copetti D, Poliakov A (2017) Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic acids research* 45 (D1):D1075-D1081

McCouch SR (2008) Gene nomenclature system for rice. *Rice* 1 (1):72-84

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20 (9):1297-1303

Qiu J, Zhou Y, Mao L, Ye C, Wang W, Zhang J, Yu Y, Fu F, Wang Y, Qian F (2017) Genomic variation associated with local adaptation of weedy rice during de-domestication. *Nature communications* 8 (1):1-12

Qiu J, Zhu J, Fu F, Ye C-Y, Wang W, Mao L, Lin Z, Chen L, Zhang H, Guo L (2014) Genome re-sequencing suggested a weedy rice origin from domesticated indica-japonica hybridization: a case study from southern China. *Planta* 240 (6):1353-1363

Reagon M, Thurber CS, Gross BL, Olsen KM, Jia Y, Caicedo AL (2010) Genomic patterns of nucleotide diversity in divergent populations of US weedy rice. *BMC Evolutionary Biology* 10 (1):180

Robson F, Costa MMR, Hepworth SR, Vizir I, Pinheiro M, Reeves PH, Putterill J, Coupland G (2001) Functional importance of conserved domains in the flowering-time gene *CONSTANS* demonstrated by analysis of mutant alleles and transgenic plants. *The Plant Journal* 28 (6):619-631

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution* 34 (12):3299-3302

Shrestha R, Gómez-Ariza J, Brambilla V, Fornara F (2014) Molecular control of seasonal flowering in rice, arabidopsis and temperate cereals. *Annals of botany* 114 (7):1445-1458

Song BK, Chuah TS, Tam SM, Olsen KM (2014) Malaysian weedy rice shows its true stripes: wild *Oryza* and elite rice cultivars shape agricultural weed evolution in Southeast Asia. *Molecular ecology* 23 (20):5003-5017

Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoechea JL (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nature genetics* 50 (2):285-296

Sun J, Ma D, Tang L, Zhao M, Zhang G, Wang W, Song J, Li X, Liu Z, Zhang W (2019) Population genomic analysis and de novo assembly reveal the origin of weedy rice as an evolutionary game. *Molecular plant* 12 (5):632-647

Sun J, Qian Q, Ma DR, Xu ZJ, Liu D, Du HB, Chen WF (2013) Introgression and selection shaping the genome and adaptive loci of weedy rice in northern China. *New Phytologist* 197 (1):290-299

- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105 (2):437-460
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123 (3):585-595
- Takahashi Y, Shimamoto K (2011) Heading date 1 (Hd1), an ortholog of Arabidopsis CONSTANS, is a possible target of human selection during domestication to diversify flowering times of cultivated rice. *Genes & genetic systems* 86 (3):175-182
- Takahashi Y, Teshima KM, Yokoi S, Innan H, Shimamoto K (2009) Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to diversity of flowering time in cultivated rice. *Proceedings of the National Academy of Sciences* 106 (11):4555-4560
- Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K (2007) Hd3a protein is a mobile flowering signal in rice. *Science* 316 (5827):1033-1036
- Tsuji H, Taoka K-i, Shimamoto K (2013) Florigen in rice: complex gene network for florigen transcription, florigen activation complex, and multiple functions. *Current Opinion in Plant Biology* 16 (2):228-235
- Vergara BS, Chang T-T (1985) The flowering response of the rice plant to photoperiod: a review of the literature. *Int. Rice Res. Inst.*,
- Wang C, Zheng X, Xu Q, Yuan X, Huang L, Zhou H, Wei X, Ge S (2014) Genetic diversity and classification of *Oryza sativa* with emphasis on Chinese rice germplasm. *Heredity* 112 (5):489-496
- Watterson G (1975) On the number of segregating sites in genetical models without recombination. *Theoretical population biology* 7 (2):256-276
- Wei F-J, Tsai Y-C, Wu H-P, Huang L-T, Chen Y-C, Chen Y-F, Wu C-C, Tseng Y-T, Yue-ie CH (2016) Both Hd1 and Ehd1 are important for artificial selection of flowering time in cultivated rice. *Plant Science* 242:187-194
- Wei X, Xu J, Guo H, Jiang L, Chen S, Yu C, Zhou Z, Hu P, Zhai H, Wan J (2010) DTH8 suppresses flowering in rice, influencing plant height and yield potential simultaneously. *Plant physiology* 153 (4):1747-1758
- Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X (2008) Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nature genetics* 40 (6):761
- Yamamoto T, Lin H, Sasaki T, Yano M (2000) Identification of heading date quantitative trait locus Hd6 and characterization of its epistatic interactions with Hd2 in rice using advanced backcross progeny. *Genetics* 154 (2):885-891
- Yano M, Harushima Y, Nagamura Y, Kurata N, Minobe Y, Sasaki T (1997) Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. *Theoretical and Applied Genetics* 95 (7):1025-1032

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y (2000) Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. The Plant Cell 12 (12):2473-2483

Zhang J, Lu Z, Dai W, Song X, Peng Y, Valverde BE, Qiang S (2015) Cytoplasmic-genetic male sterility gene provides direct evidence for some hybrid rice recently evolving into weedy rice. Scientific reports 5:10591

Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nature genetics 50 (2):278-284

Tables

Table 1. Haplotypes of loss-of-function *hd1* genes in the 3K rice database.

NB position	9336853	9337102	9337112	9337150	9337242	9338004	9338032	9338220	9338243	9338273
Types of a.a. changes	F.S.*	F.S.	STOP	STOP	F.S.	F.S.	F.S.	F.S.	STOP	STOP
Nipponbare seq †	C	G	C	C	-	TTT	-	AAGA	C	A
Type 2	1bp-del									
Type 3	1bp-del									
Type 7	4bp-del									
Type 12	T									
Type 13	2bp-del									
Type 18	43bp-del									
Type 19	1.9kb									
Type 20	T									
Type 21	T									
Type 22	A									

Nipponbare position, nucleotide sequence changes and types of protein changes are indicated.

*F.S.: frame shift

†: The position on Os-Nipponbare-Reference-IRGSP-1.0

Table 2. Traditional accessions of 8 haplotypes in rice-growing Asian countries.

	Total	Traditional	Allele group								
			2	3	7	12	13	19	20	21	other
Bangladesh	186	104	–	–	4	–	35	–	–	–	65
Bhutan	19	19	–	–	2	–	2	–	–	–	15
Cambodia	59	45	–	–	–	–	1	1	–	–	43
China*	481		–	(11)	(120)	–	(50)	(16)	–	–	(284)
India	435	73	–	2	3	–	16	–	9	3	40
Indonesia	248	203	–	–	16	–	99	–	–	–	88
Japan	55	11	–	–	–	–	–	–	–	–	11
Korea	34	19	–	–	1	–	–	–	–	–	18
Laos	126	84	–	–	4	10	–	–	–	–	70
Madagascar	66	16	–	–	–	–	8	–	–	–	8
Malaysia	75	55	–	–	2	–	17	–	–	–	36
Myanmar	75	21	–	–	1	–	–	–	–	–	20
Nepal	44	27	–	–	–	–	3	1	–	–	23
Pakistan	34	24	–	–	1	–	13	–	–	–	10
Philippines	229	97	–	–	5	–	62	–	–	–	30
Sri Lanka	54	40	–	–	1	–	7	–	–	–	32
Taiwan*	30	23	–	5	4	–	3	11	–	–	0
Thailand	147	103	–	–	1	–	5	–	–	–	97
Vietnam	55	28	2	–	3	–	1	–	–	–	22
Total	2457	990	2	7	48	10	272	13	9	3	

Most are from the 3K database and we add some accessions from Taiwan aboriginal villages (Wei et al., 2016).

*Because of no information about traditional/modern types for Chinese accessions, the total numbers are shown in brackets.

Table 3. The distribution of japonica and indica rice in the traditional accessions from different countries.

	Type 7				Type 13			
	Total	Traditional	<i>japonica</i>	<i>indica</i>	Total	Traditional	<i>japonica</i>	<i>indica</i>
Bangladesh	5	4	–	4	78	35	–	35
Bhutan	2	2	2	–	2	2	1	1
Cambodia	–	–	–	–	1	1	1	–
China*	(120)	–	(2)	(118)	(50)	–	(9)	(41)
India	40	3	1	2	108	16	–	16
Indonesia	25	16	1	15	122	99	65	34
Laos	4	4	–	4	2	–	–	–
Madagascar	1	–	–	–	25	8	3	5
Malaysia	8	2	2	–	26	17	16	1
Myanmar	3	1	–	1	3	0	–	–
Pakistan	1	1	–	1	19	13	–	13
Philippines	72	5	–	5	105	62	47	15
Sri Lanka	5	1	–	1	14	7	1	6
Thailand	7	1	–	1	7	5	2	3
Taiwan	7	4	–	4	3	3	–	3
Vietnam	10	3	–	3	5	1	–	1

Only types 7 and 13 are illustrated. Most are 3K data and we added some accessions from Taiwan aboriginal villages (Wei et al., 2016).

* Because of no information about traditional/modern types for Chinese accessions, the total numbers are shown in brackets.

Table 4. Selection sweep analysis for 4 *hd1* types in several countries.

Haplotype	Country	Subspecies	Number of accessions	Π	θ_w	Tajima's <i>D</i>	Significance level of <i>D</i>
Type 7	Bangladesh	<i>indica</i>	4	0.00065	0.00066	-0.15923	P > 0.10
Type 7	Indonesia	<i>Indica</i> and <i>japonica</i>	16	0.00138	0.0027	-2.12711**	P < 0.01
Type 7	Laos	<i>indica</i>	4	0.00044	0.00044	-0.15777	P > 0.10
Type 12	Laos	<i>japonica</i>	10	0.00115	0.00179	-1.75625*	P < 0.05
Type 13	Bangladesh	<i>indica</i>	35	0.00294	0.00603	-1.95450*	P < 0.05
Type 13	India	<i>indica</i>	16	0.00320	0.00267	0.86860	P > 0.10
Type 13	Indonesia	<i>Indica</i> and <i>japonica</i>	99	0.00120	0.00547	-2.65204**	P < 0.001
Type 13	Indonesia	<i>japonica</i>	61	0.00137	0.00534	-2.65105**	P < 0.001
Type 13	Philippines	<i>Indica</i> and <i>japonica</i>	62	0.00125	0.00432	-2.52823**	P < 0.001
Type 13	Philippines	<i>japonica</i>	46	0.00137	0.00432	-2.50592**	P < 0.001
Type 19	Taiwan	<i>japonica</i>	11	0.00205	0.00328	-1.81600*	P < 0.05

The *hd1* gene region and nearby ± 10 kb of traditional landraces were used for calculation.

Figures

Figure 2

Flowering date of the 7 loss-of-function haplotypes and wild type. These plants were grown on the IRRI campus and the flowering date information was downloaded from the SNP-SEEK website. Sample number and detailed statistical data are in Table S1.

Figure 4

Geographical distribution of functional and nonfunctional Hd1 alleles. The number within each pie indicates the number of cultivars belonging to each Hd1 allele group. Color indicates different haplotypes and the pie size indicates the sample size. Only the traditional landraces in the 3K database were used for the plot. Type 2, lime green; type 3, sky blue; type 7, yellow; type 12, green stripe; type 13, navy; type 19, brown; type 20: yellowish green; type 21, light blue; other (without loss-of-function mutation), gray. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

- [TableS6.xlsx](#)
- [TablesS2350417.docx](#)
- [TableS6.xlsx](#)
- [TableS4weedyrice.xlsx](#)