Genet Resour Crop Evol (2014) 61:121–142
DOI 10.1007/s10722-013-0019-8

RESEARCH ARTICLE

Domestication and association analysis of *Hd1* in Chinese mini-core collections of rice

Xin Wei · Weihua Qiao · Nannan Yuan · Youtao Chen · Rongsheng Wang · Lirong Cao · Wanxia Zhang · Qingwen Yang · Hanlai Zeng

Received: 26 February 2013 / Accepted: 25 June 2013 / Published online: 15 August 2013

Abstract *Hd1* is one of the major photoperiod genes with high degree of polymorphisms and contributes to rice (*Oryza sativa* L.) flowering in different light conditions. Ninety-two rice landraces and 111 accessions of common wild rice (*O. rufipogon* Griff.) from the mini-core collections in China were selected and sequenced to analyze the domestication process and association of *Hd1* with rice flowering. Association analysis revealed that three insertions and two deletions in the coding region of *Hd1* are highly correlated with flowering time in the short-day condition. Phylogenetic analysis suggested that the polymorphisms of *Hd1* in wild rice are related to the distributions. Haplotype analysis indicated that *Hd1* in most *O. sativa* L. ssp. *indica* Kato and *O. sativa* L. ssp. *japonica* Kato landraces evolved from different *O. rufipogon* groups containing functional *Hd1* and that most *aus* varieties were domesticated from *O. rufipogon* containing the long-insertions *Hd1*, suggesting multiple origins of *Hd1*. Moreover, *O. rufipogon* which contained the long-insertions *Hd1* could only be found in Southern China, implying that Southern China might be one of the domestication centers of *O. sativa*. This study provides much significant information to aid further understanding of the domestication process of *Hd1*.

Keywords Association analysis · Domestication · *Hd1* · Mini-core collections · *Oryza rufipogon* · *Oryza sativa*

Introduction

Rice (*Oryza sativa* L.) is the most significant crop, feeding approximately half the population of the world. It is believed that rice was domesticated from common wild rice (*O. rufipogon* Griff.) in East Asia about 10,000 years ago (Doebley et al. 2006). Common wild rice is photoperiod sensitive, flowers at short-day condition, and mainly exists in South and Southeast Asia (Vaughan et al. 2003). However, *O. sativa* has been spread all around the world, and some can even flower in long-day condition. Previous studies have suggested that the allelic variation of *Heading date 1* (*Hd1*) in *O. sativa* might be one possible reason for this phenomenon (Yano et al. 2000; Tsuji et al. 2008; Takahashi et al. 2009; Huang et al. 2012a).

Electronic supplementary material The online version of this article (doi:10.1007/s10722-013-0019-8) contains supplementary material, which is available to authorized users.

X. Wei · W. Qiao · N. Yuan · Y. Chen · R. Wang · L. Cao · W. Zhang · Q. Yang (✉)
Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
e-mail: qwyang@mail.caas.net.cn

X. Wei · H. Zeng (✉)
College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China
e-mail: zenghl@mail.hzau.edu.cn

Hd1 is a major photoperiod-sensitive gene and plays a significant role in regulating rice flowering. Previous research revealed that clock genes receive signals from light and circadian clocks and regulate expression of *Hd1* (Yano et al. 2000). Then *Hd1* regulates the expression of a mobile flowering signal florigen, *Hd3a*, and controls the flowering of rice (Tsuji et al. 2008). *Hd1* is regarded as the major determinant of the variation in flowering-time diversity in cultivated rice, and the high degree of polymorphisms in *Hd1* is believed to contribute to the regulation of rice flowering time (Takahashi et al. 2009). Although more than 60 cultivated rice accessions have been sequenced and plenty of nucleotide variations have been obtained (Takahashi et al. 2009; Fujino et al. 2010; Huang et al. 2012a), which variations are significantly associated with flowering time is not yet completely clear. Thus, the association of *Hd1* with flowering time was analyzed in the present study to detect these significant variations.

Traditionally, *O. sativa* is divided into two subspecies: *O. sativa* L. ssp. *indica* Kato and *O. sativa* L. ssp. *japonica* Kato. These two rice subspecies can be distinguished by both DNA markers and morphologic characteristics. The origin of *indica* and *japonica* has been studied for decades, but still with quite varied conclusions (Londo et al. 2006; Gao and Innan 2008; Molina et al. 2011; Huang et al. 2012b; Wei et al. 2012a). However, according to rice diversity research, *O. sativa* can be divided into several groups. Not only *indica* and *japonica*, but also *aus* should form one group of *O. sativa* (Garris et al. 2005; Kovach et al. 2007; Zhao et al. 2010). *Aus* is known as pre-kharif rice or autumn rice which grows mainly in India and Bangladesh. It is usually planted during May to August and harvested in autumn. It was selected from other varieties based on the character of earliness. Earliness is the main character of *aus*, and most varieties of *aus* are light insensitive. Few studies about the domestication of *aus* have been reported. A recently published paper which analyzed the domestication of *O. sativa* by genome-wide patterns and rich material from a wide range suggests that ancient *japonica* was first domesticated from *O. rufipogon* around the middle area of the Pearl River in Southern China and that *indica* was subsequently developed from crosses between ancient *japonica* and local wild rice as the initial cultivars spread into Southeast and South Asia, then initial *indica* diverged into *indica* and *aus* (Huang et al. 2012c). In the present study, *japonica*, *indica*, and *aus* varieties were collected and sequenced to analyze the phylogeny and domestication process of *Hd1* in these groups.

Hd1 shows intriguing variations of functions at 31°N and 23.5°N (Izawa 2007). It promotes flowering in the short-day condition in geographical regions south of 31°N but represses flowering in the long-day condition in areas north of 31°N, while south of the Tropic of Cancer (TOC, 23.5°N), *Hd1* tends to be nonfunctional and results in late flowering due to its insensitivity to photoperiod (Huang et al. 2012a). In the present study, all the *aus* varieties were from tropical areas south of the TOC, but the *japonica* varieties ranged from 10°N to 45°N and the *indica* varieties ranged from 5°N to 30°N. Therefore, *japonica* could be divided into tropical *japonica*, subtropical *japonica*, and temperate *japonica* according to the geographical divisions at 31°N and the TOC. Also, *indica* could be divided into tropical *indica* and subtropical *indica* according to the geographical division at the TOC. The diversity of *Hd1* in different groups and the relationship between *Hd1* in geographical regions are also investigated.

In the present study, 92 accessions of *O. sativa* and 111 accessions of *O. rufipogon* were collected and sequenced. All cultivated samples are landraces (pure-line varieties developed by farmers without artificial intercrossing). The landraces and *O. rufipogon* were collected from their mini-core collections in China. The mini-core collections were identified from China National Genebank, which included 50,526 landraces and more than 10,000 accessions of *O. rufipogon*, by morphological traits and simple sequence repeat (SSR) markers. Through a hierarchical sampling strategy, the mini-core collections of *O. sativa* and *O. rufipogon* retained more than 70 % of the morphological variation in all germplasm collections (Zhang et al. 2011). Sixty *O. sativa* accessions originating from other countries have been sequenced and published (Fujino et al. 2010). These sequences are also included in our analysis. In total, *Hd1* of 263 accessions of *O. sativa* and *O. rufipogon* with high diversity has been sequenced and analyzed. Such study with rich and representative materials can provide much significant information to aid further understanding of the domestication process of *Hd1*.

Materials and methods

Sampling and phenotypic data collection

The materials used in this study included 92 accessions of cultivated rice (Table 1), 111 accessions of *O. rufipogon*, and 1 accession of *O. barthii* A. Chev. (Table 2). The selected 46 *indica* and 46 *japonica* landraces were from 23 different provinces in China. To distinguish *indica* and *japonica*, we investigated all individuals by Cheng's index method (Lu et al. 2009), which has been popularly used in China. *O. rufipogon* was procured from Guangzhou and Nanning Wild Rice Field Genebanks, except for the wild rice of Yunnan Province, which was collected directly from distribution sites. The *O. rufipogon* accessions were investigated carefully throughout their whole lifespan to remove the individuals which had significant gene flow from *O. sativa* from the samples. Currently, *O. rufipogon* only exists in seven provinces in South China, and samples originating from all seven provinces were included in our research. *O. barthii* was provided by the International Rice Research Institute and used as the outgroup sample. The geographic localities of *O. rufipogon* and landraces from China sampled in our research are shown in Fig. 1. Sixty *O. sativa* accessions originating from other countries were sequenced (Table 3), and the sequences were added in the following analysis.

All landraces were planted in late November at Sanya, Hainan (southernmost China) from 2009 to 2011 and grew mostly under short-day conditions. Ten plants were transplanted in a single row with 20 cm between plants and 30 cm between rows of different accessions. Field management was performed following normal agricultural practices. Heading date was defined as the days from sowing to the appearance of 50 % panicles.

DNA extraction, PCR amplification, cloning, and sequencing

Total DNA extraction and polymerase chain reaction (PCR) amplification were conducted generally following the methods of our previous studies (Wei et al. 2012a; Qiao et al. 2012). Primers of *Hd1* are listed in Table S1. Almost the entire gene region of *Hd1* was sequenced, including the promoter, exon 1, intron, exon 2, and 3′-untranslated region (UTR) (Fig. 2). Sequencing was performed by an ABI 3730 automated sequencer (Applied Biosystems, USA). Initially, all of the samples were directly sequenced, but if the accession was a heterozygote, the PCR product was ligated into the EASY vector (Transgen, China). Independent plasmid DNA was then selected randomly, and at least four clones were sequenced so that the sequences of both alleles would be obtained. Because Taq errors did occur, when polymorphisms were only found in one of the accessions, this accession was resequenced with the cloning step to ensure these polymorphisms were not false. Because heterozygous individuals exist in *O. rufipogon*, two sequences of alleles were obtained for some wild samples.

Population structure analysis

Twenty-four simple sequence repeat (SSR) markers were used to detect the structure of the cultivated samples: RM529, RM522, RM526, RM211, RM411, RM60, RM518, RM348, RM574, RM274, RM508, RM412, RM427, RM172, RM339, RM408, RM553, RM321, RM484, RM239, RM224, RM479, RM247, and RM463. One exists in each short and long arm of the 12 rice chromosomes. PCR was performed as described above, and PCR products were separated on 6 % polyacrylamide denaturing gels to determine the alleles of each marker. The STRUCTURE 2.3.2 program (Falush et al. 2003) was used to infer the population structure with burn-in of 100,000, run length of 100,000, and a model allowing for admixture and correlated allele frequencies. Number of subpopulations K from two to ten was tested, and ten independent runs yielded consistent likelihoods of the population structure for each K.

DNA sequence analysis, neutrality test, and association analysis

The DNA sequences were aligned using the ClustalX program (Thompson et al. 1997) and manually adjusted in BioEdit (Hall 1999). The number of segregating sites (S), the number of haplotypes (h), the haplotype diversity (Hd), and two parameters of nucleotide diversity, namely π (Nei 1987) and Watterson's estimator from S (θ_w) (Watterson 1975), were calculated by DnaSP version 5.0 (Rozas 2009).

Two neutrality tests, namely Tajima's D value and Fu and Li's D^*/F^*, were calculated for all loci to test the neutral mutation hypothesis. Tajima's D (1989) is

Table 1 Details of the landraces from China

Name	Origin	No.	Type	Taxa	Protein type	Coding region haplotype	Heading day (Hainan 2009)	Heading day (Hainan 2010)	Heading day (Hainan 2011)	Q1	Q2	Q3
AH_HN-I	Anhui, Huaining, China	11-00389	Landrace	I	T1	H1	87	85	83	0.976	0.003	0.02
AH_TH-I	Anhui, Taihu, China	11-00403	Landrace	I	T2	H2	88	86	87	0.994	0.003	0.003
AH_WH-I	Anhui, Wuhu, China	11-00322	Landrace	I	T2	H2	84	81	80	0.996	0.002	0.002
FJ_XY-I	Fujian, Xianyou, China	13-00816	Landrace	I	T2	H2	76	66	69	0.995	0.003	0.002
GD_GZ-1-I	Guangdong, Guangzhou, China	15-03168	Landrace	I	T2	H2	90	88	84	0.996	0.002	0.002
GD_GZ-2-I	Guangdong, Guangzhou, China	15-03336	Landrace	I	T3[a]	H3	95	92	97	0.992	0.003	0.005
GD_LC-I	Guangdong, Longchuan, China	15-01740	Landrace	I	T3[a]	H3	99	97	97	0.993	0.003	0.004
GD_YD-I	Guangdong, Yingde, China	15-03025	Landrace	I	T2	H2	75	74	77	0.987	0.005	0.008
GX_FeS-I	Guangxi, Fengshan, China	16-09350	Landrace	I	T2	H2	90	89	94	0.905	0.022	0.074
GX_HX-I	Guangxi, Hengxian, China	16-00163	Landrace	I	T4[a]	H4	97	93	92	0.995	0.003	0.002
GZ_CJ-J	Guizhou, Congjiang, China	22-03815	Landrace	J	T5	H5	112	119	115	0.007	0.054	0.939
GZ_FQ-J	Guizhou, Fuquan, China	22-01439	Landrace	J	T5	H5	95	92	82	0.002	0.005	0.993
GZ_HS-I	Guizhou, Huishui, China	22-01615	Landrace	I	T6[a]	H6	77	79	77	0.991	0.004	0.005
GZ_QL-J	Guizhou, Qinglong, China	22-00513	Landrace	J	T7	H7	92	93	91	0.003	0.005	0.992
GZ_SB-J	Guizhou, Shibing, China	22-04053	Landrace	J	T5	H5	96	95	93	0.002	0.004	0.994
GZ_WC-I	Guizhou, Wuchuan, China	22-02148	Landrace	I	T8[a]	H8	95	102	88	0.013	0.063	0.925
GZ_ZF-J	Guizhou, Zhenfeng, China	22-00570	Landrace	J	T7	H7	93	100	86	0.003	0.005	0.992
GZ_ZiY-I	Guizhou, Ziyun, China	22-04637	Landrace	I	T2	H2	91	95	90	0.104	0.007	0.889
GZ_ZN-I	Guizhou, Zhenning, China	22-02754	Landrace	I	T9[a]	H9	109	112	106	0.973	0.022	0.004
GZ_ZY-I	Guizhou, Zunyi, China	22-00040	Landrace	I	T2	H2	82	87	88	0.994	0.003	0.003
HeB_FN-J	Hebei, Funing, China	02-00058	Landrace	J	T10	H10	73	85	80	0.002	0.021	0.976
HeB_LH-J	Hebei, Longhua, China	02-00133	Landrace	J	T11[a]	H11	96	95	97	0.002	0.006	0.992
HeN_HB-I	Henan, Huaibin, China	19-00205	Landrace	I	T2	H2	88	86	84	0.978	0.015	0.007
HeN_ZY-I	Henan, Zhengyang, China	19-00022	Landrace	I	T3[a]	H3	98	104	103	0.995	0.002	0.002
HLJ_HL-J	Heilongjiang, Hailin, China	07-00010	Landrace	J	T12	H12	65	49	53	0.002	0.005	0.993
HLJ_SH-J	Heilongjiang, Suihua, China	07-00109	Landrace	J	T13	H13	66	50	53	0.002	0.025	0.973
HN_BT-I	Hainan, Baoting, China	15-04286	Landrace	I	T2	H2	90	81	83	0.968	0.025	0.007
HN_LD-I	Hainan, Ledong, China	31-00042	Landrace	I	T3[a]	H3	106	109	107	0.727	0.247	0.027
HN_SY-I	Hainan, Sanya, China	31-00032	Landrace	I	T3[a]	H3	113	110	114	0.92	0.068	0.012
HuB_CY-I	Hubei, Chongyang, China	17-00502	Landrace	I	T14[a]	H14	89	78	80	0.996	0.002	0.002
HuB_PQ-I	Hubei, Puqi, China	17-00435	Landrace	I	T2	H2	66	67	57	0.993	0.004	0.003

Table 1 continued

Name	Origin	No.	Type	Taxa	Protein type	Coding region haplotype	Heading day (Hainan 2009)	Heading day (Hainan 2010)	Heading day (Hainan 2011)	Q1	Q2	Q3
HuB_TS-I	Hubei, Tongshan, China	17-00524	Landrace	I	T2	H2	87	77	81	0.995	0.003	0.003
HuB_ZX-I	Hubei, Zhuxi, China	17-01470	Landrace	I	T2	H2	94	91	93	0.996	0.002	0.002
HuN_CS-1-I	Hunan, Changsha, China	18-01067	Landrace	I	T3[a]	H3	93	86	90	0.996	0.002	0.002
HuN_CS-2-I	Hunan, Changsha, China	18-04906	Landrace	I	T2	H2	92	92	91	0.995	0.002	0.002
HuN_LH-J	Hunan, Longhui, China	18-01903	Landrace	J	T15	H15	93	91	93	0.003	0.016	0.981
HuN_QY-J	Hunan, Qianyang, China	18-04082	Landrace	J	T5	H5	97	89	101	0.002	0.006	0.991
JS_JY-J	Jiangsu, Jiangyin, China	09-00530	Landrace	J	T5	H5	72	73	82	0.003	0.008	0.988
JS_NT-J	Jiangsu, Nantong, China	09-01361	Landrace	J	T16[a]	H16	83	88	87	0.013	0.092	0.895
JS_WJ-J	Jiangsu, Wujin, China	09-00724	Landrace	J	T5	H5	68	82	84	0.002	0.05	0.947
JX_DX-I	Jiangxi, Dongxiang, China	12-00589	Landrace	I	T2	H2	69	70	65	0.994	0.003	0.003
JX_GX-I	Jiangxi, Guixi, China	12-00644	Landrace	I	T17	H17	82	73	84	0.981	0.009	0.01
JX_NC-1-I	Jiangxi, Nanchang, China	12-01446	Landrace	I	T18[a]	H18	93	77	94	0.993	0.003	0.005
JX_NC-2-I	Jiangxi, Nanchang, China	12-02254	Landrace	I	T19[a]	H19	91	103	96	0.993	0.004	0.002
JX_NC-3-I	Jiangxi, Nanchang, China	12-02280	Landrace	I	T2	H2	70	76	72	0.996	0.002	0.002
JX_NC-4-I	Jiangxi, Nanchang, China	12-02373	Landrace	I	T2	H2	71	69	73	0.996	0.002	0.002
LN_DD-J	Liaoning, Dandong, China	05-00052	Landrace	J	T20[a]	H20	84	82	89	0.003	0.089	0.907
LN_GX-J	Liaoning, Gaixian, China	05-00024	Landrace	J	T21[a]	H21	80	87	90	0.003	0.006	0.991
SAX_LT-J	Shanxi, Lantian, China	24-00215	Landrace	J	T15	H15	88	78	99	0.002	0.007	0.991
SAX_SY-J	Shanxi, Shanyang, China	24-00195	Landrace	J	T10	H10	81	68	78	0.002	0.019	0.979
SC_GL-I	Sichuan, Gulan, China	20-01452	Landrace	I	T3[a]	H3	93	91	88	0.995	0.002	0.003
SC_GX-J	Sichuan, Gaoxian, China	20-03215	Landrace	J	T22	H22	91	83	97	0.003	0.923	0.075
SC_LX-J	Sichuan, Luxian, China	20-01734	Landrace	J	T23[a]	H23	93	99	85	0.013	0.017	0.971
SC_MN-I	Sichuan, Mianning, China	20-01262	Landrace	I	T2	H2	87	78	79	0.711	0.079	0.209
SC_QX-I	Sichuan, Quxian, China	20-03053	Landrace	I	T2	H2	93	89	81	0.993	0.003	0.004
SC_WQ-I	Sichuan, Wanquan, China	20-02821	Landrace	I	T2	H2	82	80	68	0.995	0.002	0.002
SC_YB-I	Sichuan, Yibin, China	20-02073	Landrace	I	T2	H2	80	76	83	0.995	0.003	0.003
SH_FX-J	Shanghai, Fengxian, China	08-00253	Landrace	J	T5	H5	70	67	80	0.003	0.007	0.99
SH_SH-J	Shanghai, Shanghai, China	08-00036	Landrace	J	T5	H5	70	66	82	0.003	0.009	0.987
SX_TY-J	Shanxi, Taiyuan, China	04-00115	Landrace	J	T24	H24	71	69	75	0.002	0.007	0.99

Table 1 continued

Name	Origin	No.	Type	Taxa	Protein type	Coding region haplotype	Heading day (Hainan 2009)	Heading day (Hainan 2010)	Heading day (Hainan 2011)	Q1	Q2	Q3
TJ_BD-J	Tianjing, Baodi, China	29-00010	Landrace	J	T10	H10	94	86	88	0.003	0.006	0.991
TJ_TJ-1-J	Tianjing, Tianjin, China	02-00294	Landrace	J	T25	H25	69	69	72	0.002	0.007	0.99
TJ_TJ-2-J	Tianjing, Tianjin, China	02-00295	Landrace	J	T10	H10	72	79	71	0.002	0.024	0.973
TW_TB-1-J	Taiwan, Taibei, China	30-00195	Landrace	J	T26	H26	83	88	84	0.154	0.013	0.833
TW_TB-2-J	Taiwan, Taibei, China	30-00206	Landrace	J	T3[a]	H3	98	104	103	0.219	0.009	0.772
TW_TB-3-I	Taiwan, Taibei, China	30-00210	Landrace	I	T3[a]	H3	99	103	95	0.996	0.002	0.002
TW_TB-4-I	Taiwan, Taibei, China	30-00244	Landrace	I	T3[a]	H3	99	105	96	0.996	0.002	0.002
XZ_MT-I	Xizang, Motuo, China	26-00008	Landrace	I	T2	H2	75	91	95	0.985	0.008	0.007
YN_DH-J	Yunnan, Dehong, China	21-00529	Landrace	J	T27[a]	H27	95	101	107	0.003	0.993	0.005
YN_DY-J	Yunnan, Dayao, China	21-04506	Landrace	J	T2	H2	85	83	79	0.958	0.028	0.014
YN_GM-J	Yunnan, Gengma, China	21-00785	Landrace	J	T27[a]	H27	102	101	104	0.003	0.986	0.011
YN_JH-J	Yunnan, Jinghong, China	21-00357	Landrace	J	T27[a]	H27	93	96	97	0.003	0.993	0.004
YN_JP-1-J	Yunnan, Jinping, China	21-01106	Landrace	J	T2	H2	82	79	76	0.985	0.008	0.007
YN_JP-2-I	Yunnan, Jinping, China	21-01120	Landrace	I	T19[a]	H19	93	101	95	0.988	0.008	0.005
YN_LaC-1-J	Yunnan, Lancang, China	21-01970	Landrace	J	T15	H15	92	103	82	0.002	0.024	0.974
YN_LaC-2-J	Yunnan, Lancang, China	21-01989	Landrace	J	T28	H28	96	106	97	0.002	0.965	0.033
YN_LC-J	Yunnan, Lvchun, China	21-01165	Landrace	J	T19[a]	H19	109	102	105	0.989	0.007	0.004
YN_LiC-I	Yunnan, Lincang, China	21-00694	Landrace	I	T2	H2	71	59	56	0.988	0.005	0.006
YN_LL-J	Yunnan, Longling, China	21-04732	Landrace	J	T15	H15	79	71	72	0.002	0.022	0.975
YN_MaL-J	Yunnan, Malong, China	21-04413	Landrace	J	T29	H29	90	95	83	0.025	0.019	0.956
YN_MH-J	Yunnan, Menghai, China	21-02824	Landrace	J	T2	H2	76	86	84	0.927	0.044	0.029
YN_ML-1-J	Yunnan, Menglian, China	21-00083	Landrace	J	T28	H28	93	94	92	0.017	0.977	0.006
YN_ML-2-J	Yunnan, Menglian, China	21-01853	Landrace	J	T30	H30	89	99	83	0.002	0.963	0.035
YN_ML-3-I	Yunnan, Menglian, China	21-01899	Landrace	I	T2	H2	91	97	95	0.996	0.002	0.002
YN_PE-1-I	Yunnan, Puer, China	21-02171	Landrace	I	T19[a]	H19	95	97	98	0.97	0.003	0.026
YN_PE-2-J	Yunnan, Puer, China	21-02224	Landrace	J	T2	H2	87	76	87	0.992	0.004	0.004
YN_XM-J	Yunnan, Ximeng, China	21-02089	Landrace	J	T31	H31	87	79	73	0.005	0.873	0.122
YN_XP-1-I	Yunnan, Xinping, China	21-05072	Landrace	I	T19[a]	H19	104	101	100	0.79	0.202	0.008
YN_XP-2-I	Yunnan, Xinping, China	21-05171	Landrace	I	T6[a]	H6	96	90	97	0.995	0.002	0.003

Table 1 continued

Name	Origin	No.	Type	Taxa	Protein type	Coding region haplotype	Heading day (Hainan 2009)	Heading day (Hainan 2010)	Heading day (Hainan 2011)	Q1	Q2	Q3
YN_YY-J	Yunnan, Yuanyang, China	21-01257	Landrace	J	T2	H2	88	72	82	0.971	0.005	0.025
YN_ZK-J	Yunnan, Zhenkang, China	21-03879	Landrace	J	T9[a]	H9	97	93	95	0.615	0.342	0.043
ZJ_WX-J	Zhejiang, Wuxing, China	10-00463	Landrace	J	T5	H5	75	72	71	0.003	0.031	0.966

Q1, Q2, and Q3 represent the genomic components of each individual that originated from subpopulation 1, 2, and 3, respectively. I: *O. sativa* L. ssp. *indica* Kato; J: *O. sativa* L. ssp. *japonica* Kato

[a] Nonfunctional proteins

based on the discrepancy between the mean pairwise differences (π) and Watterson's estimator (θ_w), whereas Fu and Li's D^*/F^* (Fu and Li 1993) relies on the differences between the numbers of polymorphic sites in external and internal groups. In these two tests, negative values indicate an excess of low-frequency polymorphisms, whereas positive values indicate an excess of intermediate variants.

Association between the phenotypes and sequences was analyzed by TASSEL 2.1 (Bradbury et al. 2007). Single-nucleotide polymorphism (SNP)/indel–trait associations were identified by generating a general linear model (GLM). Linkage disequilibrium (LD) was estimated by DnaSP using standardized disequilibrium coefficients and squared allele–frequency correlations (r^2) for pairs of SNP loci.

Phylogenetic analysis

The haplotype networks of the coding region were constructed by mutational steps with NETWORK 4.6 (Bandelt et al. 1999). Only major haplotypes containing two or more individuals were used. The phylogenetic relationships among the major haplotypes of the coding region and all haplotypes of the whole region were constructed by neighbor-joining (NJ) analysis (Saitou and Nei 1987) using MEGA 4.0 (Tamura et al. 2007). Gaps were treated as missing values, and these sites were excluded from the data matrix. In the NJ analysis, we followed Kimura's two-parameter model (Kimura 1980). The nonparametric bootstrap test was performed to quantify the confidence of internal nodes with 1,000 replications.

Results

Nucleotide diversity and neutrality analysis

The whole genomic DNA sequences of *Hd1* from the 92 cultivated and 111 wild accessions were sequenced, and the nucleotide diversity of *Hd1* in them was analyzed. High degree of polymorphisms and long insertions and deletions were detected in *Hd1*. The gene lengths ranged from 2,811 to 3,476 bp because of the long indels. In total, 63 indels and 82 SNPs were found. In different regions of *Hd1*, the polymorphisms varied. As shown in Table 4, the pairwise nucleotide diversity parameter (π) and the level of the Watterson

Table 2 List of wild samples used in the study

Name	Origin	Species	No.	Whole region haplotype	Coding region haplotype	Protein type
FJ_ZP-1-W	Fujian, Zhangpu	R	YD5-0007	h21&h22	H2&H15	T2&T43
FJ_ZP-2-W	Fujian, Zhangpu	R	YD5-0029	h23	H16	T44&T45
FJ_ZP-3-W	Fujian, Zhangpu	R	YD5-0033	h23	H16	T44&T45
FJ_ZP-4-W	Fujian, Zhangpu	R	YD5-0051	h24&h25	H15&H17	T43&T46
FJ_ZP-5-W	Fujian, Zhangpu	R	YD5-0070	h26	H2	T41
GD_CH-W	Guangdong, Conghua	R	YD1-0701	h5	H2	T47&T48
GD_DQ-W	Guangdong, Deqing	R	YD1-0618	h27	H2	T41
GD_EP-1-W	Guangdong, Enping	R	YD1-0323	h28	H2	T41
GD_EP-2-W	Guangdong, Enping	R	YD1-2523	h29	H16	T44
GD_EP-3-W	Guangdong, Enping	R	YD1-2536	h30	H2	T2
GD_FG-1-W	Guangdong, Fogang	R	YD1-1812	h2	H2	T49
GD_FG-2-W	Guangdong, Fogang	R	YD1-2802	h31	H2	T41&T50
GD_HD-1-W	Guangdong, Huidong	R	YD1-1289	h32&h33	H18&H19	T51&T52[c]
GD_HD-2-W	Guangdong, Huidong	R	YD1-1373	h29	H16	T44&T53[c]
GD_HF-W	Guangdong, Haifeng	R	YD1-1683	h34&h35	H2&H16	T48&T54
GD_HeY-1-W	Guangdong, Heyuan	R	YD1-1566	h36	H2	T41
GD_HeY-2-W	Guangdong, Heyuan	R	YD1-1568	h27&37	H2	T41&T55[c]
GD_HX-1-W	Guangdong, Huaxian	R	YD1-0646	h3&h38	H2&H16	T44&T55[c]
GD_HX-2-W	Guangdong, Huaxian	R	YD1-0669	h3	H2	T41&T55[c]
GD_HY-1-W	Guangdong, Huiyang	R	YD1-1109	h5&h39	H2	T55[c]&T56
GD_HY-2-W	Guangdong, Huiyang	R	YD1-1111	h8	H4	T32[c]
GD_HZ-W	Guangdong, Huizhou	R	YD1-0883	h5&h40	H2	T41&T57
GD_KP-1-W	Guangdong, Kaiping	R	YD1-0484	h41	H20	T58
GD_KP-2-W	Guangdong, Kaiping	R	YD1-2549	h42	H2	T59
GD_PN-W	Guangdong, Puning	R	YD1-1779	h43	H2	T41&T60
GD_QJ-1-W	Guangdong, Qujiang	R	YD1-1841	h44	H16	T44
GD_QJ-2-W	Guangdong, Qujiang	R	YD1-1845	h34	H16	T44
GD_QY-W	Guangdong, Qingyuan	R	YD1-1800	h45&h46	H2&H16	T61&T62
GD_RH-W	Guangdong, Renhua	R	YD1-1851	h47	H21	T63
GD_SX-1-W	Guangdong, Suixi	R	YD1-0183	h29	H16	T44&T64
GD_SX-2-W	Guangdong, Suixi	R	YD1-0196	h38&h48	H16&H22	T44&T65
GD_SX-3-W	Guangdong, Suixi	R	YD1-2471	h42	H2	T41

Table 2 continued

Name	Origin	Species	No.	Whole region haplotype	Coding region haplotype	Protein type
GD_SX-4-W	Guangdong, Suixi	R	YD1-2480	h38	H16	T44
GD_TS-1-W	Guangdong, Taishan	R	YD1-0374	h49&h50	H16&H23	T44&T66
GD_TS-2-W	Guangdong, Taishan	R	YD1-0390	h49	H16	T44&T53
GD_YC-W	Guangdong, Yangchun	R	YD1-2521	h49$h51	H2&H16	T44&T60
GD_ZJ-1-W	Guangdong, Zijin	R	YD1-1579	h35&h52	H2&H24	T41&T67
GD_ZJ-2-W	Guangdong, Zijin	R	YD1-1591	h53&h54	H3&H4	T5&T28
GD_ZJ-3-W	Guangdong, Zijin	R	YD1-1610	h6	H3	T5
GX_BY-W	Guangxi, Binyang	R	YD2-0887	h36	H2	T48
GX_FS-1-W	Guangxi, Fusui	R	YD2-2606	h3	H2	T60
GX_FS-2-W	Guangxi, Fusui	R	YD2-2662	h38	H16	T54
GX_GG-1-W	Guangxi, Guigang	R	YD2-2094	h3&h55	H2&H25	T41&T68
GX_GG-2-W	Guangxi, Guigang	R	YD2-2194	h38&h56	H16&H26	T69&T70
GX_GP-1-W	Guangxi, Guiping	R	YD2-0583	h29	H16	T44
GX_GP-2-W	Guangxi, Guiping	R	YD2-0642	h5	H2	T48
GX_GX-1-W	Guangxi, Guixian	R	YD2-0435	h5	H2	T41
GX_GX-2-W	Guangxi, Guixian	R	YD2-0513	h8	H4	T34
GX_HX-W	Guangxi, Hengxian	R	YD2-0704	h5	H2	T41
GX_LA-1-W	Guangxi, Longan	R	YD2-0812	h35&h57	H2&H27	T41&T71
GX_LA-2-W	Guangxi, Longan	R	YD2-0824	h58	H2	T41
GX_LA-3-W	Guangxi, Longan	R	YD2-0841	h59	H2	T72
GX_LA-4-W	Guangxi, Longan	R	YD2-0848	h60	H2	T41
GX_LA-5-W	Guangxi, Longan	R	YD2-0857	h60	H2	T41
GX_LB-1-W	Guangxi, Laibin	R	YD2-1340	h29	H16	T44
GX_LB-2-W	Guangxi, Laibin	R	YD2-1349	h29	H16	T44
GX_PN-1-W	Guangxi, Pingnan	R	YD2-2372	h61	H26	T70
GX_PN-2-W	Guangxi, Pingnan	R	YD2-2380	h29	H16	T44
GX_RX-1-W	Guangxi, Rongxian	R	YD2-2335	h62	H28	T73
GX_RX-2-W	Guangxi, Rongxian	R	YD2-2360	h4&h63	H2&H29	T41&T74
GX_TY-W	Guangxi, Tianyang	R	YD2-0986	h64	H2	T41
GX_WM-W	Guangxi, Wuming	R	YD2-0680	h65	H2	T2
GX_WX-1-W	Guangxi, Wuxuan	R	YD2-1443	h3	H2	T48
GX_WX-2-W	Guangxi, Wuxuan	R	YD2-1476	h5	H2	T55

Table 2 continued

Name	Origin	Species	No.	Whole region haplotype	Coding region haplotype	Protein type
GX_XZ-W	Guangxi, Xiangzhou	R	YD2-1400	h3	H2	T2
GX_YL-1-W	Guangxi, Yulin	R	YD2-0223	h29	H16	T54
GX_YL-2-W	Guangxi, Yulin	R	YD2-0252	h66&h67	H2&H30	T41&T75
GX_YL-3-W	Guangxi, Yulin	R	YD2-2029	h3&h38	H2&H16	T60&T69
GX_YN-1-W1	Guangxi, Yongning	R	YD2-2418	h29	H16	T69&T76
GX_YN-2-W	Guangxi, Yongning	R	YD2-2466	h3&h68	H2&H31	T77&T78
GX_YN-3-W	Guangxi, Yongning	R	YD2-2516	h3	H2	T41
HN_DA-1-W	Hainan, Dingan	R	YD1-2383	h69&h70	H2&H32	T79
HN_DA-2-W	Hainan, Dingan	R	YD1-2384	h71	H1	T80&T81
HN_DF-1-W	Hainan, Dongfang	R	YD1-0019	h72	H33	T82
HN_DF-2-W	Hainan, Dongfang	R	YD1-0022	h42	H2	T41
HN_HK-1-W	Hainan, Haikou	R	YD1-0094	h73&h74	H34&H35	T83&T84
HN_HK-2-W	Hainan, Haikou	R	YD1-2442	h42&h75	H2&H36	T85&T86
HN_LD-W	Hainan, Ledong	R	YD1-0009	h76	H4	T28
HN_LG-1-W	Hainan, Lingao	R	YD1-0049	h42	H2	T41
HN_LG-2-W	Hainan, Lingao	R	YD1-0057	h77&h78	H2&H16	T41&T44
HN_LG-3-W	Hainan, Lingao	R	YD1-2391	h42	H2	T41
HN_LG-4-W	Hainan, Lingao	R	YD1-2395	h42	H2	T41
HN_QH-W	Hainan, Qionghai	R	YD1-2368	h79	H37	T87
HN_QS-1-W	Hainan, Qiongshan	R	YD1-2399	h80	H20	T58
HN_QS-2-W	Hainan, Qiongshan	R	YD1-2405	h81	H35	T88
HN_QS-3-W	Hainan, Qiongshan	R	YD1-2426	h82	H38	T89
HN_QS-4-W	Hainan, Qiongshan	R	YD1-2417	h83	H35	T90
HN_QS-5-W	Hainan, Qiongshan	R	YD1-2426	h84	H39	T91
HN_SY-W	Hainan, Sanya	R	YD1-2336	h85&h86	H40&H41	T92&T93
HN_WC-W	Hainan, Wenchang	R	YD1-0087	h87&h88	H42&H43	T94&T95
HN_WN-W	Hainan, Wanning	R	YD1-2354	h89	H44	T96
HuN_CL-1-W	Hunan, Chaling	R	YD6-0136	h62	H28	T73
HuN_CL-2-W	Hunan, Chaling	R	YD6-0139	h62	H28	T73
HuN_CL-3-W	Hunan, Chaling	R	YD6-0162	h90	H28	T73&T97
HuN_CL-4-W	Hunan, Chaling	R	YD6-0183	h62&h91	H28&H45	T73&T98
HuN_JY-1-W	Hunan, Jiangyong	R	YD6-0233	h92	H2	T41

Table 2 continued

Name	Origin	Species	No.	Whole region haplotype	Coding region haplotype	Protein type
HuN_JY-2-W	Hunan, Jiangyong	R	YD6-0084	h38	H16	T44
JX_DX-1-W	Jiangxi, Dongxiang	R	YD4-0020	h93	H28	T73&T99
JX_DX-2-W	Jiangxi, Dongxiang	R	YD4-0022	h93	H28	T99
JX_DX-3-W	Jiangxi, Dongxiang	R	YD4-0049	h93	H28	T73
JX_DX-4-W	Jiangxi, Dongxiang	R	YD4-0076	h94&h95	H28&H46	T73&T100
YN_JH-1-W	Yunnan, Jinghong	R	YD3-0002	h96&h97	H47	T101
YN_JH-2-W	Yunnan, Jinghong	R	YD3-0003	h98&h99	H47	T101
YN_JH-3-W	Yunnan, Jinghong	R	YD3-0007	h97&h100	H47&H48	T101&T102
YN_JH-4-W	Yunnan, Jinghong	R	YD3-0008	h96&h97	H47	T101
YN_JH-5-W	Yunnan, Jinghong	R	YD3-0009	h96&h97	H47	T101
YN_JH-6-W	Yunnan, Jinghong	R	YD3-0011	h97&h100	H47&H48	T101&T102
YN_YJ-1-W	Yunnan, Yuanjiang	R	BC-0046[a]	h42	H2	T103
YN_YJ-2-W	Yunnan, Yuanjiang	R	BC-0049[a]	h42	H2	T103
YN_YJ-3-W	Yunnan, Yuanjiang	R	BC-0081[a]	h42	H2	T103
YN_YJ-4-W	Yunnan, Yuanjiang	R	BC-0085[a]	h42	H2	T103
barthii	Gambia	B	100941[b]			

R: *O. rufipogon* Griff.; B: *O. barthii* A. Chev

[a] Provisional conservation number of Chinese National Germplasm Bank

[b] Conservation number of Genetic Resources Center, International Rice Research Institute

[c] Nonfunctional proteins

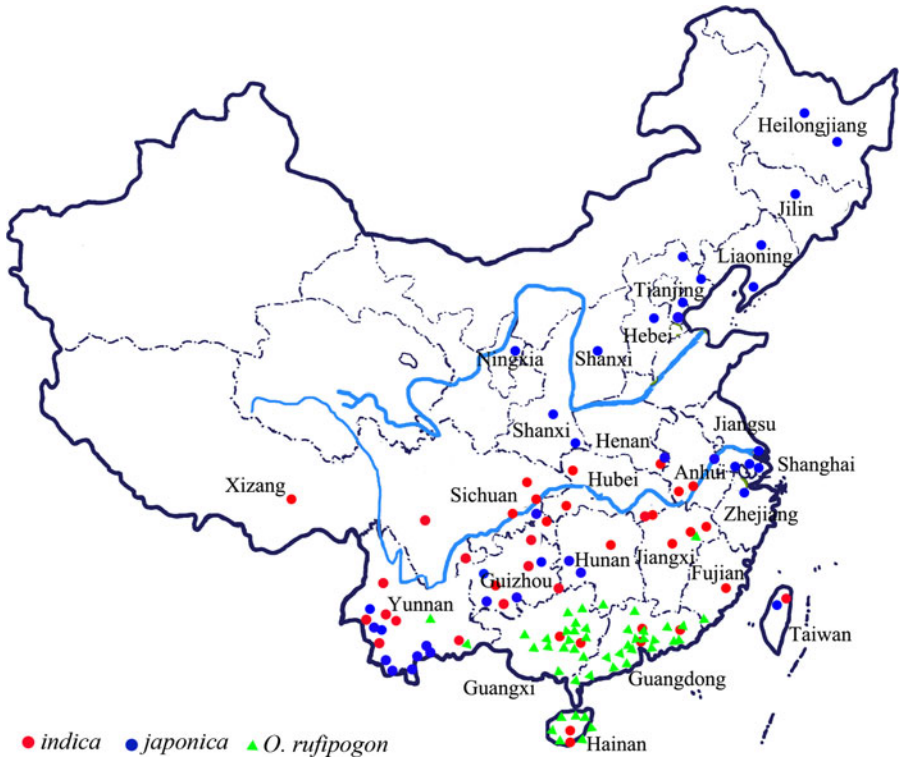


Fig. 1 Geographic origins of the materials from China. *Red circles* indicate *O. sativa* L. ssp. *indica* Kato; *blue circles* indicate *O. sativa* L. ssp. *japonica* Kato; *green triangles* indicate *O. rufipogon* Griff. (Color figure online)

estimator (θ_w) of promoter were higher than for other regions for *O. rufipogon* and *indica*. This result is in line with some other research into another photoperiod gene (*Ghd7*) in rice (Lu et al. 2012). However, for *japonica*, the first exon had the highest diversity. In the neutrality analysis, only the values for *O. rufipogon* were significantly negative, suggesting negative selection or quick population expansion of *O. rufipogon*.

Association analysis

A population structure of the cultivated rice was constructed using the 24 SSR markers (Fig. 3). The structure can be classified into three subpopulations because the highest log-likelihood scores of the population structure were observed when the number of populations was set at 3 ($K = 3$; Fig. S1). The first subpopulation (subpopulation 1) contained 51 accessions, of which 86 % were *indica* varieties; subpopulation 2 contained 8 *japonica* varieties; and subpopulation 3 contained 33 accessions, of which 94 % were *japonica* varieties (Table 1). In fact, when $K = 2$, subpopulations 2 and 3 were in the same cluster (Fig. S2). Further analysis revealed most *japonica* in subpopulation 2 were located south of the TOC, while all *japonica* varieties in subpopulation 3 existed north of the TOC, implying that cluster 2 might be tropical *japonica* and cluster 3 might be temperate *japonica*. LD was detected in the whole genomic region of *Hd1* and fast LD decay was observed (Fig. S3), suggesting weak association between the SNPs in *Hd1*.

Taking the population structure data as covariates (Table 1), we used GLM to identify SNP/indel–trait associations separately for 3 years. SNPs and indels at less than 5 % were excluded. No SNP was found to be related to the flowering date in the association analysis. Five significant associated indels were detected and were the same between different years (Table 5). All insertions and deletions were located in the coding region. S3527 and S4199 were 2- and 4-bp deletions, respectively, both of which would lead to loss of function of *Hd1*. These deletions had also been detected in previous research (Takahashi et al. 2009; Fujino et al. 2010). S1081, S1539, and S1668 were 3-, 33-, and 156-bp insertions, respectively. The long insertions might lead to partial loss of function of *Hd1* (Yano et al. 2000). Thus, S1081, S1539, S1668, S3527, and S4199 were the sites associated with flowering time of rice. All of them were in the coding region and would weaken function of *Hd1*.

Table 3 List of the cultivated samples used in Fujino et al. (2010)

Name	Origin	No.	Type	Taxa	Protein type	Coding region haplotype
NIPPONBARE	Japan	WRC1	Breeding	J	T10	H10
KASALATH	India	WRC2	Landrace	A	T32[a]	H32
BEIKHE	Cambodia	WRC3	Landrace	I	T2	H2
JENA 035	Nepal	WRC4	Landrace	A	T33[a]	H33
NABA	India	WRC5	Landrace	I	T19[a]	H19
PULUIK ARANG	Indonesia	WRC6	Landrace	I	T34[a]	H34
DAVAO 1	Philippines	WRC7	Landrace	I	T3[a]	H3
RYOU SUISAN KOUMAI	China	WRC9	Landrace	I	T2	H2
JINGUOYIN[a]	China	WRC11	Landrace	I	T19	H19
ASU	Bhutan	WRC13	Landrace	I	T19[a]	H19
IR 58	Philippines	WRC14	Breeding	I	T34[a]	H34
CO 13	India	WRC15	Unknown	I	T2	H2
KEIBOBA	China	WRC17	Landrace	I	T9[a]	H9
QINGYU	China	WRC18	Landrace	I	T3[a]	H3
DENG PAO ZHAI	China	WRC19	Breeding	I	T2	H2
SHWE NANG GYI	Myanmar	WRC21	Landrace	I	T3[a]	H3
CALOTOC	Philippines	WRC22	Landrace	A	T35	H35
LEBED	Philippines	WRC23	Landrace	I	T34[a]	H34
PINULUPOT 1	Philippines	WRC24	Landrace	I	T35	H35
MUHA	India	WRC25	Unknown	A	T32[a]	H32
JHONA 2	India	WRC26	Unknown	A	T36[a]	H36
NEPAL 8	Nepal	WRC27	Landrace	A	T9[a]	H9
JARJAN	Bhutan	WRC28	Landrace	A	T9[a]	H9
KALO DHAN	Nepal	WRC29	Landrace	A	T32[a]	H32
ANJANA DHAN	Nepal	WRC30	Landrace	A	T32[a]	H32
SHONI	Bangladesh	WRC31	Landrace	A	T32[a]	H32
TUPA 121-3	Bangladesh	WRC32	Landrace	A	T37[a]	H37
SURJAMUKHI	India	WRC33	Breeding	A	T32[a]	H32
ARC 7291	India	WRC34	Landrace	A	T32[a]	H32
ARC 5955	India	WRC35	Landrace	A	T9[a]	H9
RATUL	India	WRC36	Landrace	A	T32[a]	H32
ARC 7047	India	WRC37	Landrace	A	T32[a]	H32
ARC 11094	India	WRC38	Landrace	A	T32[a]	H32
BADARI DHAN	Nepal	WRC39	Landrace	A	T38	H38
NEPAL 555	India	WRC40	Unknown	A	T32[a]	H32
KALUNEENATI	Sri Lanka	WRC41	Landrace	A	T32[a]	H32
LOCAL BASMATI	India	WRC42	Landrace	A	T32[a]	H32
DIANYU 1	China	WRC43	Breeding	J	T39[a]	H39
BASILANON	Philippines	WRC44	Landrace	A	T35	H35
MA SHO	Myanmar	WRC45	Landrace	J	T27[a]	H27
KHAO NOK	Laos	WRC46	Landrace	J	T27[a]	H27
JAGUARY	Brazil	WRC47	Unknown	J	T34[a]	H34
KHAU MAC KHO	Vietnam	WRC48	Landrace	J	T27[a]	H27
PADIPERAK	Indonesia	WRC49	Landrace	J	T34[a]	H34

Table 3 continued

Name	Origin	No.	Type	Taxa	Protein type	Coding region haplotype
REXMONT	USA	WRC50	Breeding	J	T34[a]	H34
URASAN 1	Japan	WRC51	Landrace	J	T26	H26
KHAU TAN CHIEM	Vietnam	WRC52	Landrace	J	T26	H26
TIMA	Bhutan	WRC53	Landrace	J	T28	H28
TUPA 729	Bangladesh	WRC55	Landrace	A	T32[a]	H32
MILYANG	Korea	WRC57	Breeding	J	T3[a]	H3
NEANG MENH	Cambodia	WRC58	Landrace	I	T2	H2
NEANG PHTONG	Cambodia	WRC59	Landrace	I	T2	H2
RADIN GOI SESAT	Malaysia	WRC61	Landrace	I	T2	H2
KEMASIN	Malaysia	WRC62	Landrace	I	T28	H28
BLEIYO	Thailand	WRC63	Landrace	I	T40	H40
PADIKUNING	Indonesia	WRC64	Landrace	I	T34[a]	H34
RAMBHONG	Indonesia	WRC65	Landrace	I	T41	H41
BINGALA	Myanmar	WRC66	Landrace	I	T2	H2
PHULBA	India	WRC67	Landrace	J	T42	H42
KHAO NAM JEN	Laos	WRC68	Landrace	J	T28	H28

A: *aus*; I: *O. sativa* L. ssp. *indica* Kato; J: *O. sativa* L. ssp. *japonica* Kato

[a] Nonfunctional proteins

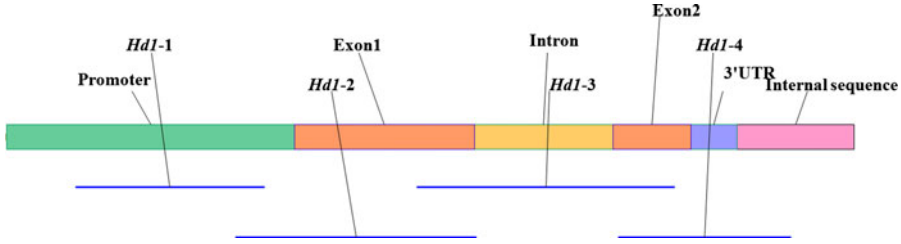


Fig. 2 Structure of *Hd1* and locations of the regions sequenced. The regions amplified and sequenced are shown as *blue lines*. (Color figure online)

Hd1 protein diversity

As revealed in the association study, flowering time of rice is strongly related to the coding region rather than to other regions in *Hd1*. The diversity of *Hd1* in all cultivated samples was analyzed to further investigate the variations. Sixty alleles published in a previous study (Fujino et al. 2010) were also added to the analysis. Forty-two protein types were identified, including 31 protein types in this study and 18 protein types in the previous study (Fig. 4). Besides the 18 protein types reported previously, 24 new protein types were found in the present study. Compared with the sequence of the functional *Hd1* allele from the *japonica* cultivar Ginbouzu (Takahashi et al. 2009), a total of 44 mutation events, including 6 insertions, 2 transposons, 12 deletions, 6 nonsynonymous SNPs, 15 synonymous SNPs, and 3 premature stop codons were detected in the coding region. Twelve mutation events were identified as generating loss-of-function alleles based on their predicted effects on Hd1, including two transposons, three premature stop codons, three 1-bp deletions, one 2-bp deletion, one 4-bp deletion, one 43-bp deletion, and one 144-bp deletion in the end.

All 42 types of coding region in *Hd1* were compared with that of Ginbouzu, as shown in Fig. 4. This clearly showed that the coding region sequences could be divided into two different groups: one contained the 33- or 156-bp insertions and had several different polymorphisms with Ginbouzu, while the other did not contain the long insertions and was more similar to Ginbouzu. As mentioned in the association

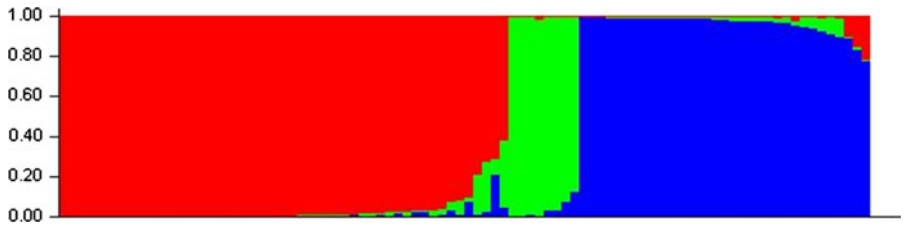


Fig. 3 Structure of the 92 landraces constructed by 24 SSR markers. $K = 3$. Subpopulations are indicated by *different colors*. (Color figure online)

Table 4 Summary of nucleotide polymorphisms and neutrality test

Taxa	Region	*S*	*h*	Hd	$\pi \times 10^3$	$\theta_w \times 10^3$	*D*	D^*	F^*
O. sativa L. ssp. *indica* Kato	Promoter	4	4	0.602	2.37	1.09	2.719***	1.024	1.814*
	Exon 1	7	3	0.300	1.65	1.62	0.056	0.478	0.405
	Intron	5	3	0.27	1.52	1.79	−0.361	0.136	−0.023
	Exon 2	3	4	0.274	0.96	1.21	−0.443	−0.420	−0.500
	Coding	10	6	0.347	1.83	1.93	−0.141	0.177	0.084
	Noncoding	9	6	0.682	1.60	1.11	1.255	0.721	1.070
	Total	19	9	0.705	1.69	1.43	0.590	0.548	0.678
O. sativa L. ssp. *japonica* Kato	Promoter	6	5	0.605	2.04	1.62	0.662	0.324	0.510
	Exon 1	11	9	0.764	4.26	2.77	1.588	0.304	0.882
	Intron	5	4	0.538	3.21	1.79	1.970	0.136	0.837
	Exon 2	1	2	0.502	1.57	0.71	1.659	0.543	0.995
	Coding	12	10	0.766	4.55	2.85	1.780	0.417	1.055
	Noncoding	11	8	0.766	2.15	1.43	1.485	0.304	0.841
	Total	23	17	0.881	3.00	1.94	1.805	0.443	1.134
O. sativa L.	Promoter	5	5	0.655	2.36	1.17	2.172*	1.053	1.678
	Exon 1	11	9	0.677	3.42	2.39	1.130	0.108	0.564
	Intron	6	5	0.437	2.63	1.85	0.958	−0.796	−0.250
	Exon 2	1	2	0.410	1.28	0.61	1.270	0.494	0.844
	Coding	12	11	0.683	3.66	2.47	1.296	1.229	0.734
	Noncoding	11	9	0.785	2.09	1.24	1.813	0.108	0.855
	Total	23	20	0.853	2.65	1.67	1.730	0.217	0.969
O. rufipogon Griff.	Promoter	30	36	0.885	2.87	6.61	−1.637	−0.658	−1.286
	Exon 1	33	33	0.639	2.59	6.41	−1.740	−3.710**	−3.473**
	Intron	17	17	0.599	2.00	4.74	−1.545	−3.365**	−3.198**
	Exon 2	12	22	0.692	2.14	3.77	−1.079	−0.607	−0.937
	Coding	39	40	0.810	2.86	6.24	−1.610	−3.284**	−3.092**
	Noncoding	53	57	0.942	2.20	5.19	−1.754	−2.245	−2.442*
	Total	92	87	0.979	2.45	5.59	−1.764	−3.176**	−3.023**

S: number of segregating sites; *h*: number of haplotypes; Hd: haplotype diversity; π: nucleotide diversity; θ_w: Watterson's parameter for silent sites; *D*: Tajima's *D*; D^* and F^*: Fu and Li's D^* and Fu and Li's F^*, respectively

* $P < 0.05$; ** $P < 0.02$; *** $P < 0.01$

analysis, the association containing the long insertions might have lost part of *Hd1* function, so the first group could be regard as a part loss function group and the second group as a functional group. Both groups contained about half cultivated samples, but the first group included *indica*, *japonica*, and *aus* varieties

Table 5 Results of GLM association analysis

Year	Site	*P*	R^2
2009	1081	0.0076	0.1301
	1539	0.024	0.1011
	1668	0.0444	0.0852
	3527	0.0027	0.1198
	4199	0.0014	0.1353
2010	1081	0.0033	0.1546
	1539	0.0051	0.1438
	1668	0.0103	0.1258
	3527	0.0027	0.1227
	4199	0.0072	0.1
2011	1081	0.012	0.1199
	1539	0.003	0.154
	1668	0.0096	0.1254
	3527	0.0056	0.1042
	4199	0.0028	0.1198

while the second group mainly contained *indica* and *japonica* individuals. This result is in line with the fact that most *aus* varieties are insensitive to photoperiod.

Haplotype analysis

Haplotype networks were constructed by the whole gene region for the samples from the mini-core collections in China (Fig. 5) and the coding region for all accessions (Fig. 6). The whole gene region network contained 101 haplotypes: 20 *O. sativa* haplotypes, 86 *O. rufipogon* haplotypes, and 1 *O. barthii* haplotype. *O. sativa* and *O. rufipogon* shared six haplotypes.

As shown in Fig. 5, *O. rufipogon* samples from the same region tended to be in neighboring haplotypes. The phenomenon of geographical difference in *O. rufipogon* was also detected in previous research (Wang et al. 2008; Wei et al. 2012b). However, obvious association between the distribution of cultivated accessions and the haplotypes had not been observed.

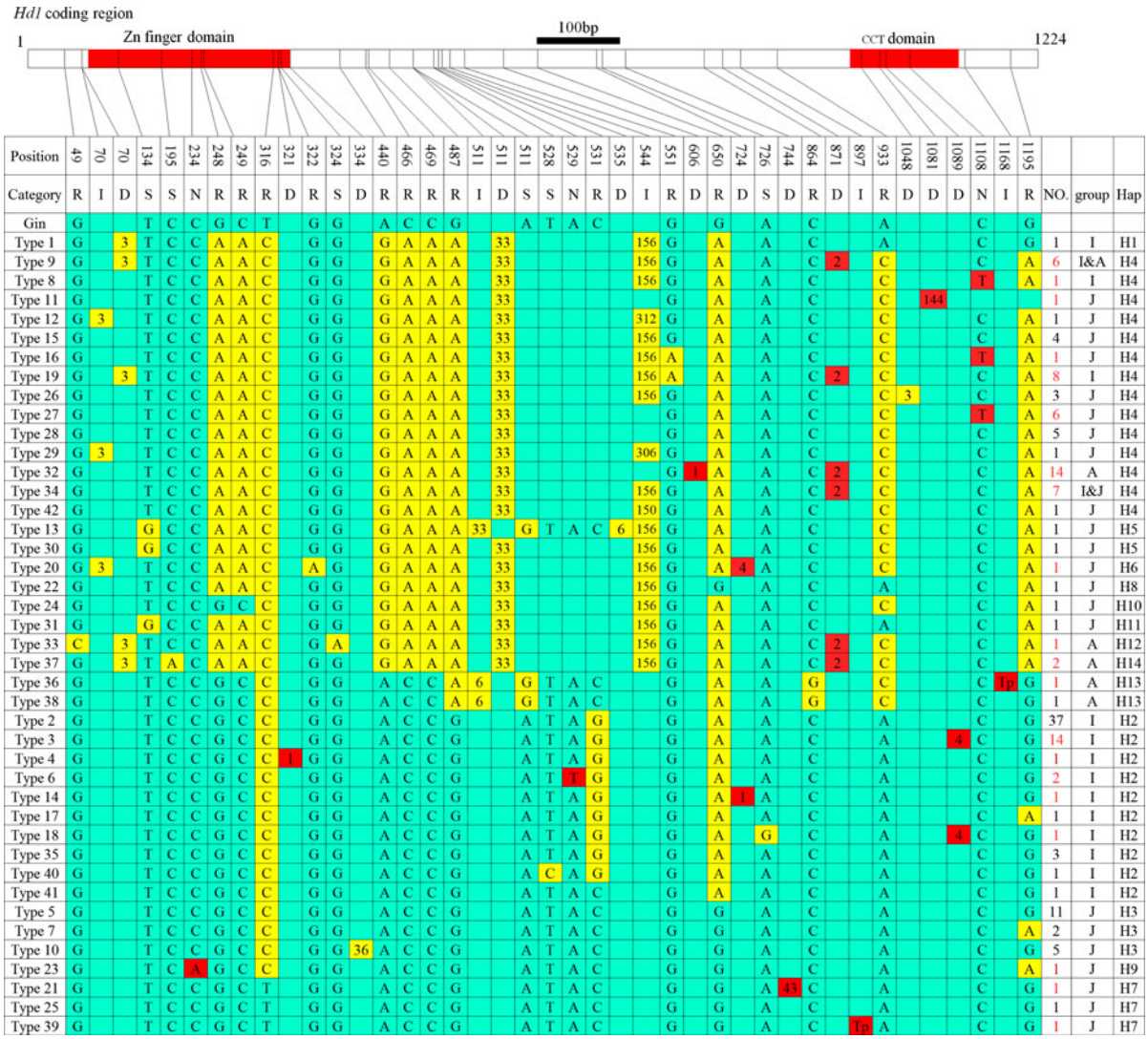
Figure 6 shows the coding region network constructed by 12 haplotypes, including 7 *O. sativa* haplotypes, 8 *O. rufipogon* haplotypes, and 1 *O. barthii* haplotype. All haplotypes were major, containing two or more individuals, except the *O. barthii* haplotype. Among these haplotypes, H1, H2, H3, and H4 were shared by both *O. sativa* and *O. rufipogon*. H1 and H4 belonged to the protein group which had the 33- and 156-bp-long insertions, while H2 and H3 belonged to the other protein group that was similar to Ginbouzu. H1 and H3 contained 1 *indica* accession and 16 *japonica* accessions. H2 contained 52 *indica* accessions, 7 *japonica* accessions, and 2 *aus* accessions. In fact, five *japonica* accessions in H2 were divided into subpopulation 1 in the structure analysis which included most *indica*. Thus, H3 and H2 could be regarded as *japonica* haplotype and *indica* haplotype, respectively. *Indica*, *japonica*, and *aus* could all be found in H4 almost averagely, including 15 *indica* accessions, 27 *japonica* accessions, and 17 *aus* accessions. It is hard to define H4 as an *indica*, *aus* or *japonica* group.

We also used the NJ method to construct the phylogeny of major haplotypes for the coding region (Fig. 7). The phylogenetic result was quite similar to the haplotype network. All branches were divided into two groups: one contained H2, H3, and a major *O. rufipogon* haplotype, while the other one included H1, H4, and other small *O. rufipogon* haplotypes.

Geographic distribution

According to geographical division at 31°N and the TOC, we further divided *O. sativa* into six subpopulations: *aus*, tropical *indica*, tropical *japonica*, subtropical *indica*, subtropical *japonica*, and temperate *japonica*. The relationship between the protein type and the distribution for each subpopulation was analyzed. Figure 8 shows the distribution of Hd1 protein type of the six subpopulations of *O. sativa* in the present study. *Aus* and tropical *japonica* mainly evolved from *O. rufipogon* in haplotype H4; tropical and subtropical *indica* mainly evolved from *O. rufipogon* in H2 with a small part evolved from H4, while subtropical and temperate *japonica* evolved from *O. rufipogon* in H3 and H4. These results reveal that the Hd1 protein type evolved from H4, with part loss of function, was transferred into all subpopulations of *O. sativa*.

Among these subpopulations, *aus* and tropical *japonica* contained the highest percent of Hd1 protein which had lost its function. Even protein type from H4, which might have partly lost its function, would lose its function totally in later artificial selection. In contrast, subtropical *indica* and subtropical *japonica* have the highest percent of functional Hd1 protein. Most of the protein types evolving from H2 and H3 were fully functional, while few of them lost function in later artificial selection.



Position	49	70	70	134	195	234	248	249	316	321	322	324	334	440	466	469	487	511	511	511	528	529	531	535	544	551	606	650	724	726	744	864	871	897	933	1048	1081	1089	1108	1168	1195			
Category	R	I	D	S	S	N	R	R	R	D	R	S	D	R	R	R	R	I	D	S	S	N	R	D	I	R	D	R	D	S	D	R	D	I	R	D	D	D	N	I	R	NO.	group	Hap
Gin	G			T	C	C	G	C	T		G	G		A	C	C	G			A	T	A	C			G		G		A		C			A						C		G			
Type 1	G		3	T	C	C	A	A	C		G	G		G	A	A	A		33						156	G		A		A		C			A						C		G	1	I	H1
Type 9	G		3	T	C	C	A	A	C		G	G		G	A	A	A		33						156	G		A		A		C	2		C						C		A	6	I&A	H4
Type 8	G			T	C	C	A	A	C		G	G		G	A	A	A		33						156	G		A		A		C			C					T		A	1	I	H4
Type 11	G			T	C	C	A	A	C		G	G		G	A	A	A		33							G		A		A		C			C			144					1	J	H4
Type 12	G	3		T	C	C	A	A	C		G	G		G	A	A	A		33						312	G		A		A		C			C						C		A	1	J	H4
Type 15	G			T	C	C	A	A	C		G	G		G	A	A	A		33						156	G		A		A		C			C						C		A	4	J	H4
Type 16	G			T	C	C	A	A	C		G	G		G	A	A	A		33						156	A		A		A		C			C					T		A	1	J	H4
Type 19	G		3	T	C	C	A	A	C		G	G		G	A	A	A		33						156	A		A		A		C	2		C						C		A	8	I	H4
Type 26	G			T	C	C	A	A	C		G	G		G	A	A	A		33						156	G		A		A		C			C	3					C		A	3	J	H4
Type 27	G			T	C	C	A	A	C		G	G		G	A	A	A		33							G		A		A		C			C					T		A	6	J	H4
Type 28	G			T	C	C	A	A	C		G	G		G	A	A	A		33							G		A		A		C			C						C		A	5	J	H4
Type 29	G	3		T	C	C	A	A	C		G	G		G	A	A	A		33						306	G		A		A		C			C						C		A	1	J	H4
Type 32	G			T	C	C	A	A	C		G	G		G	A	A	A		33							G	1	A		A		C	2		C						C		A	14	A	H4
Type 34	G			T	C	C	A	A	C		G	G		G	A	A	A		33						156	G		A		A		C	2		C						C		A	7	I&J	H4
Type 42	G			T	C	C	A	A	C		G	G		G	A	A	A		33						150	G		A		A		C			C						C		A	1	J	H4
Type 13	G			G	C	C	A	A	C		G	G		G	A	A	A	33		G	T	A	C	6	156	G		A		A		C			C						C		A	1	J	H5
Type 30	G			G	C	C	A	A	C		G	G		G	A	A	A		33						156	G		A		A		C			C						C		A	1	J	H5
Type 20	G	3		T	C	C	A	A	C		A	G		G	A	A	A		33						156	G		A	4	A		C			C						C		A	1	J	H6
Type 22	G			T	C	C	A	A	C		G	G		G	A	A	A		33						156	G		G		A		C			A						C		A	1	J	H8
Type 24	G			T	C	C	G	C	C		G	G		G	A	A	A		33						156	G		A		A		C			C						C		A	1	J	H10
Type 31	G			G	C	C	A	A	C		G	G		G	A	A	A		33						156	G		A		A		C			A						C		A	1	J	H11
Type 33	C		3	T	C	C	A	A	C		G	A		G	A	A	A		33						156	G		A		A		C	2		C						C		A	1	A	H12
Type 37	G		3	T	A	C	A	A	C		G	G		G	A	A	A		33						156	G		A		A		C	2		C						C		A	2	A	H14
Type 36	G			T	C	C	G	C	C		G	G		A	C	C	A	6		G	T	A	C			G		A		A		G			C						C	Tp	G	1	A	H13
Type 38	G			T	C	C	G	C	C		G	G		A	C	C	A	6		G	T	A	C			G		A		A		G			C						C		G	1	A	H13
Type 2	G			T	C	C	G	C	C		G	G		A	C	C	G			A	T	A	G			G		A		A		C			A						C		G	37	I	H2
Type 3	G			T	C	C	G	C	C		G	G		A	C	C	G			A	T	A	G			G		A		A		C			A			4			C		G	14	I	H2
Type 4	G			T	C	C	G	C	C	1	G	G		A	C	C	G			A	T	A	G			G		A		A		C			A						C		G	1	I	H2
Type 6	G			T	C	C	G	C	C		G	G		A	C	C	G			A	T	T	G			G		A		A		C			A						C		G	2	I	H2
Type 14	G			T	C	C	G	C	C		G	G		A	C	C	G			A	T	A	G			G		A	1	A		C			A						C		G	1	I	H2
Type 17	G			T	C	C	G	C	C		G	G		A	C	C	G			A	T	A	G			G		A		A		C			A						C		A	1	I	H2
Type 18	G			T	C	C	G	C	C		G	G		A	C	C	G			A	T	A	G			G		A		G		C			A			4			C		G	1	I	H2
Type 35	G			T	C	C	G	C	C		G	G		A	C	C	G			A	T	A	G			G		A		A		C			A						C		G	3	I	H2
Type 40	G			T	C	C	G	C	C		G	G		A	C	C	G			A	C	A	G			G		A		A		C			A						C		G	1	I	H2
Type 41	G			T	C	C	G	C	C		G	G		A	C	C	G			A	T	A	C			G		A		A		C			A						C		G	1	I	H2
Type 5	G			T	C	C	G	C	C		G	G		A	C	C	G			A	T	A	C			G		G		A		C			A						C		G	11	J	H3
Type 7	G			T	C	C	G	C	C		G	G		A	C	C	G			A	T	A	C			G		G		A		C			A						C		A	2	J	H3
Type 10	G			T	C	C	G	C	C		G	G	36	A	C	C	G			A	T	A	C			G		G		A		C			A						C		G	5	J	H3
Type 23	G			T	C	A	G	C	C		G	G		A	C	C	G			A	T	A	C			G		G		A		C			A						C		A	1	J	H9
Type 21	G			T	C	C	G	C	T		G	G		A	C	C	G			A	T	A	C			G		G		A	43	C			A						C		G	1	J	H7
Type 25	G			T	C	C	G	C	T		G	G		A	C	C	G			A	T	A	C			G		G		A		C			A						C		G	1	J	H7
Type 39	G			T	C	C	G	C	T		G	G		A	C	C	G			A	T	A	C			G		G		A		C		Tp	A						C		G	1	J	H7

Fig. 4 Nucleotide changes in the coding region of *Hd1* among cultivated rice. Categories *R*, *S*, *N*, *I*, *D*, and *Tp* indicate replacement, synonymous, noncoding site, insertion, deletion, and transposon, respectively. *Numbers* indicate size of insertions or deletions. Numbers in the right column are numbers of cultivars represented in every protein type. *A*, *I*, and *J* indicated *aus*, *O. sativa* L. ssp. *indica* Kato, and *O. sativa* L. ssp. *japonica* Kato. Variations that would not lead *Hd1* to lose function are shown in *yellow*; variations that would lead *Hd1* to be nonfunctional are shown in *red*. Sixty cultivars from a previous study (Fujino et al. 2010) were also included. (Color figure online)

Discussion

Diversity of *Hd1* in *O. sativa* and *O. rufipogon*

The diversity of *O. sativa* and *O. rufipogon* in different genes had been analyzed in our previous studies (Wei et al. 2012a, b; Qiao et al. 2012). The results indicated that the diversity of *O. rufipogon* is much higher than that of *O. sativa* and that the diversity of *indica* is higher than that of *japonica*. In the present study, we also found that the segregating sites and haplotype numbers of *O. sativa* were less than those of *O. rufipogon*. With the haplotype numbers used as a proxy for diversity, common wild rice contained 86 % of the total haplotype diversity, whereas cultivated rice only contained 20 % of the total haplotype diversity, indicating a strong genetic bottleneck during domestication. The diversity of *japonica* was higher than that

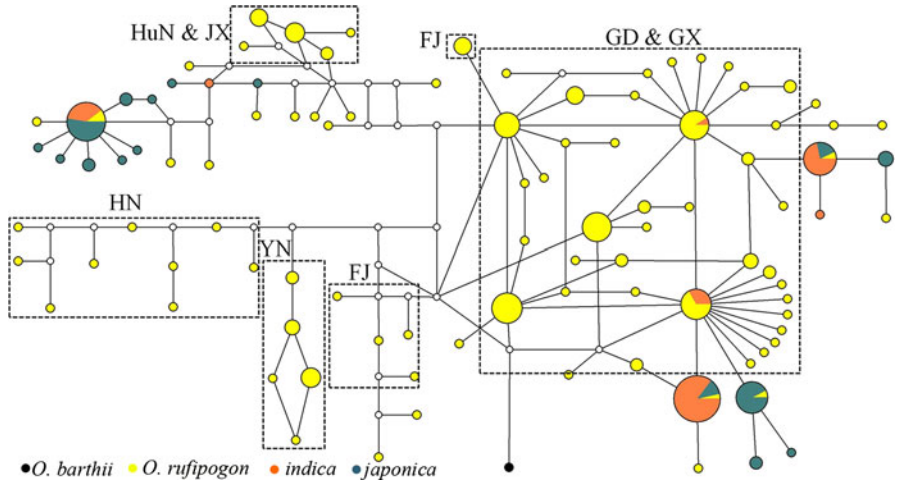


Fig. 5 Haplotype networks of the *Hd1* whole region. Circle size is proportional to the quantity of samples within a given haplotype. *Lines* between haplotypes represent mutational steps between alleles. Colors for species: *yellow*, *O. rufipogon* Griff.; *orange*, *O. sativa* L. ssp. *indica* Kato; *blue*, *O. sativa* L. ssp. *japonica* Kato; *green*, *aus*; *black*, *O. barthii* A. Chev. *FJ* Fujian, *GD* Guangdong, *GX* Guangxi, *HN* Hainan, *HuN* Hunan, *JX* Jiangxi, *YN* Yunnan. (Color figure online)

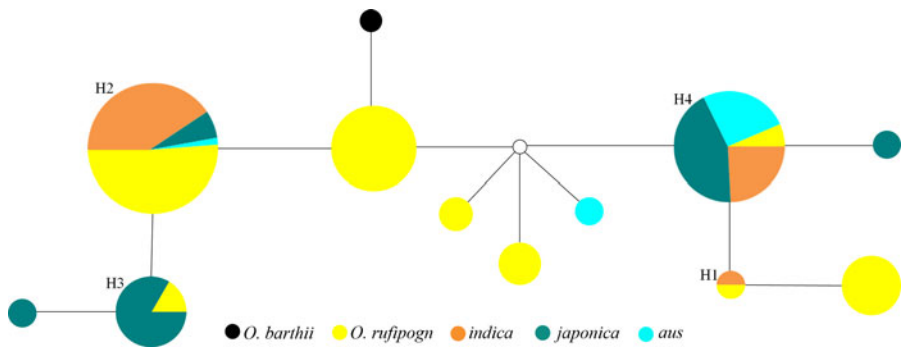


Fig. 6 Haplotype networks of the *Hd1* coding region. Circle size is proportional to the quantity of samples within a given haplotype, and the numbers next to the circles represent the haplotype number. *Lines* between haplotypes represent mutational steps between alleles. Colors for species: *yellow*, *O. rufipogon* Griff.; *orange*, *O. sativa* L. ssp. *indica* Kato; *blue*, *O. sativa* L. ssp. *japonica* Kato; *green*, *aus*; *black*, *O. barthii* A. Chev. (Color figure online)

of *indica* in the whole gene region except in exon 2. This result is quite different from previous research. It might be related to the wider distribution of *japonica*. One possible explanation is that domestication made *japonica* varieties adapt to the varied light conditions in different areas, and more polymorphisms resulted from both natural and artificial selection.

Association analysis of *Hd1*

Association analysis revealed that loss-of-function deletion and long insertion in the coding region contributed to the diversity of rice flowering date and transformed rice from a typical short-day plant to a facultative short-day plant. This result suggests that the coding region was the main target region for artificial selection and that indels that affect Hd1 protein function were the major selection methods.

Besides the previously reported 2 bp and 4 bp in the coding region that lead to loss of function, three other kinds of mutations associated with flowering in short-day condition were also identified. These four kinds of mutations would all delay flowering for rice to adapt to different light conditions; For example, in the *aus* subpopulation, the 2-bp and 4-bp deletions caused its loss of function, reduced the floral expression level, and suppressed its flowering. This variation allows the plant to flower later in the short day and improves field

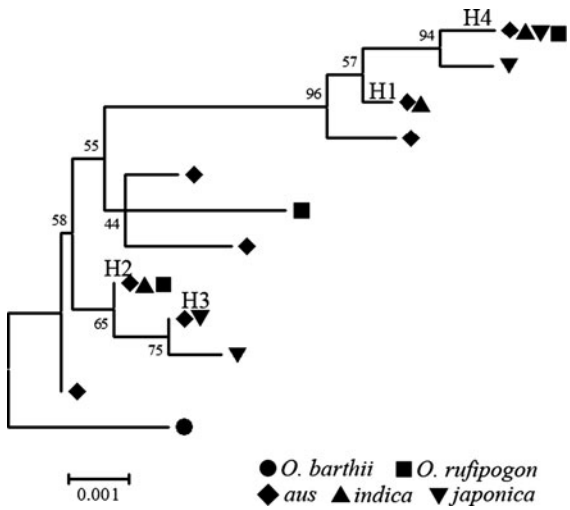


Fig. 7 Phylogenetic tree of *Hd1* coding region. Each haplotype of the loci is indicated by one branch. H1, H2, H3, and H4 are given. Bootstrap values are shown on the trees. The accessions contained in the haplotypes/branches are indicated by different symbols: *O. rufipogon* Griff. alleles by *diamonds*, *aus* alleles by *squares*, *O. sativa* L. ssp. *indica* Kato alleles by *triangles*, and *O. sativa* L. ssp. *japonica* Kato alleles by *inverted triangles*. Trees were rooted with *O. barthii* A. Chev. alleles, indicated by *solid circles*

production. On the other hand, in temperate region where temperate *japonica* varieties are widely cultivated, the light period is usually long. Nonfunctional *Hd1* would not inhibit the expression of *Hd3a* and avoid too late flowering. Additionally, in the subtropical region, both *indica* and *japonica* have high percentage of functional *Hd1*. Thus, to enable them to complete the life cycle in a short summer period in temperate region and to help them avoid too early flowering in tropical area, mutations that would lead to total loss of function of *Hd1* should be transferred into varieties cultivated in these regions. Specific markers could be developed for selection of favorable protein types to meet the demand for varieties in different ecotypes.

Domestication of *Hd1*

Since *O. sativa* and *O. rufipogon* in the same haplotypes had the same nucleotide polymorphisms, *O. sativa* might have evolved from the *O. rufipogon* groups in the same haplotypes rather than the *O. rufipogon* groups in other haplotypes. So, the *O. rufipogon*

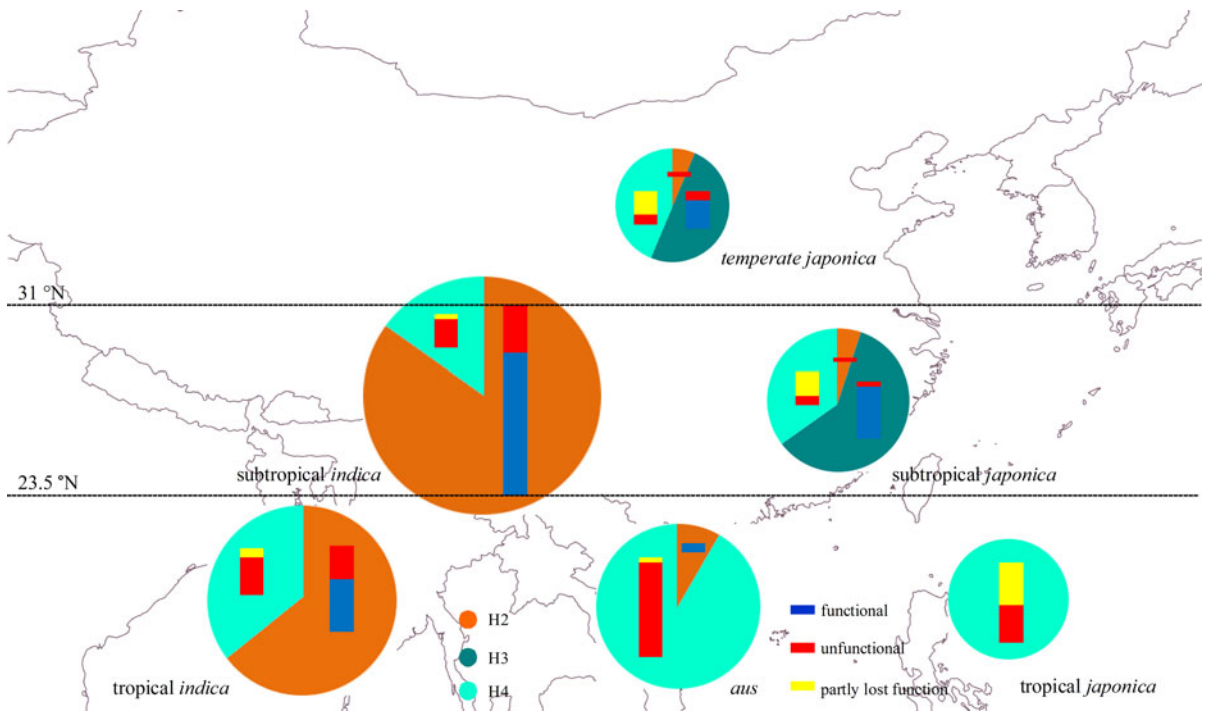


Fig. 8 Hd1 protein type distribution of the six *O. sativa* subpopulations in Asia. Hd1 protein types evolved from H2, H3, and H4 are indicated in *red*, *blue*, and *green*, respectively, in *solid circles*. *Squares* in *red*, *yellow*, and *blue* represent total loss of function, part loss of function, and functional Hd1 protein, respectively. The *size* of the circles and squares is proportional to the quantity of samples. Detailed quantitative information is presented in Table S2. (Color figure onine)

haplotypes shared with *O. sativa* were regarded as direct ancestors of *O. sativa*. Figure 6 reveals that the ancestor *O. rufipogon* could be divided into two groups. One group (H1 and H4) contained individuals including long insertions and tended to be partly nonfunctional, while the other group (H2 and H3) did not include the long insertions and tended to be functional. The functional group could be further divided into an *indica* subgroup (H2) and a *japonica* subgroup (H3) which shared haplotype with *indica* and *japonica*, respectively, indicating that functional *O. rufipogon* diverged into *indica*-like and *japonica*-like groups in natural environment. We concluded that *Hd1* in *indica* and *japonica* were domesticated from that in *indica*-like and *japonica*-like *O. rufipogon* groups, respectively.

Generally, the debate about rice domestication has focused on the origin of *indica* and *japonica*. Some researchers insist on a "single origin" of the two subspecies and suggest that both subspecies were domesticated from one group of *O. rufipogon* in a narrow region (Gao and Innan 2008; Molina et al. 2011; Huang et al. 2012b), while others propose "multiple origins," which means that domestication of the two subspecies occurred independently in different ecological and geographical environments (Cheng et al. 2003; Zhu and Ge 2005; Londo et al. 2006). Thus, our results support multiple origins of *indica* and *japonica*. This conclusion has obtained much support from various rice domestication research (Kovach et al. 2007; Sang and Ge 2007; He et al. 2011; Yang et al. 2011).

As another important group of *O. sativa*, *aus* was somewhat different from *indica* and *japonica* in terms of the domestication of *Hd1*. Most *aus* accessions shared haplotypes with *O. rufipogon* in H4, which was obviously different from that in H2 and H3, containing long insertions that would make the Hd1 protein partly lose its function. Therefore, we concluded that *Hd1* in *aus* were domesticated from *O. rufipogon* group, partly losing its function (H4), while *indica* and *japonica* evolved from functional *O. rufipogon*, which diverged into *indica*-like (H2) and *japonica*-like (H3) groups. Moreover, some varieties of *indica* and *japonica* also existed in H4, indicating that this haplotype of *Hd1* might be widely transferred into *aus*, *indica*, and *japonica* to help them adapt to different photoperiods.

Figure 8 shows the *Hd1* gene in the six subgroups of *O. sativa* in tropical, subtropical, and temperate regions divided by the TOC and 31°N. As shown in Fig. 8, *Hd1* in the six subgroups of *O. sativa* were domesticated from those in different *O. rufipogon* groups, and *Hd1* in *O. sativa* varieties that evolved from that in *O. rufipogon* in H4, which contain the long insertions, could be found in each subgroup of *O. sativa*. This result indicates that *Hd1* which contained the long insertions had been transferred into all subgroups of *O. sativa*. However, the content of the varieties which evolved from the long-insertions *Hd1* in the six subgroups were quite different. Most *Hd1* in *aus* and tropical *japonica* evolved from that in *O. rufipogon* in H4, but only a small percentage of *Hd1* in subtropical *indica* evolved from that in *O. rufipogon* in H4. This phenomenon might result from different light conditions in the tropical and subtropical regions, as explained in the association analysis.

The proportions of functional *Hd1* versus nonfunctional *Hd1* for each subgroup of *O. sativa* are shown in Fig. 8. This shows that *Hd1* in *aus* and tropical *japonica* which existed in the south and most *Hd1* in temperate *japonica* which existed in temperate region lost function totally or partly, and that most *Hd1* in subtropical *indica* and subtropical *japonica* were functional in the subtropical region. This result suggests that *Hd1* tends to be selected to be nonfunctional when spread to tropical and temperate regions. Moreover, even *Hd1* in *O. sativa*, which was supposed to be domesticated from part functional loss (H4) *Hd1* in *O. rufipogon*, was also selected to be nonfunctional. We conclude that artificial selection of nonfunctional mutations occurred after domestication in the breeding of varieties to adapt to the changed light conditions when the varieties were brought into a new environment.

Geographic origin of *Hd1*

It has been proposed that Asian cultivated rice originated from South Asia (Londo et al. 2006), Southern China (Ting 1957), the lower area of the Yangtze River in China (Vaughan et al. 2008; Zong et al. 2007), and Yun-Gui Highland (Liu 1975) based on different evidence, but the molecular evidence was insufficient. In previous study, to determine the relationship between the haplotypes and geographic origin, the whole gene region of *Hd1* was used in the geographic analysis (Wei et al. 2012a). Twenty-one

accessions of common wild rice from Southern China were regarded as the ancestors of *O. sativa* in China, and *O. sativa* might be domesticated from the Pearl River region in Southern China.

Oryza sativa mainly evolved from *O. rufipogon* in H2, H3, and H4. More *O. rufipogon* was included in H2 and H3 than in H4. All *O. rufipogon* accessions in H4 were from Southern China. To confirm whether *O. rufipogon* in H4 only originated from China, we detected the first exon of *Hd1* in 60 accessions of *O. rufipogon* from South Asia and Southeast Asia. Generally, *O. rufipogon* accessions in H4 contained 33- and 156-bp insertions, and their sequences were longer than other samples. However, the insertions in the first exon were not detected in the samples from South and Southeast Asia (Fig. S4). Thus, we conclude that H4 might only originate from Southern China and that Southern China was one of the domestication centers of *O. sativa*. A recently published paper also suggested that rice was first domesticated in the Pearl River region of Southern China (Huang et al. 2012c).

Acknowledgments We thank the Chinese National Germplasm Bank for providing the seeds of *O. sativa*, Guangzhou and Nanning Wild Rice Germplasm Banks for providing the leaves of wild rice, and the International Rice Research Institute for providing seeds of *O. barthii*. This study was supported by the program of Conservation and Sustainable Utilization of Wild Relatives of Crops by China Ministry of Agriculture (201003021).

References

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E (2003) Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Mol Biol Evol 20:67–75

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127:1309–1321

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. Genetics 164:1567–1587

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Fujino K, Wu J, Sekiguchi H, Ito T, Izawa T, Matsumoto T (2010) Multiple introgression events surrounding the *Hd1* flowering-time gene in cultivated rice *Oryza sativa* L. Mol Genet Genomics 284:137–146

Gao LZ, Innan H (2008) Nonindependent domestication of the two rice subspecies *Oryza sativa* ssp. *indica* and ssp. *japonica* demonstrated by multilocus microsatellites. Genetics 179:965–976

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

He Z, Zhai W, Wen H, Tang T, Wang Y, Lu X, Greenberg AJ, Hudson RR, Wu CI, Shi S (2011) Two evolutionary histories in the genome of rice: the roles of domestication genes. PLoS Genet 7:e1002100

Huang CL, Hung CY, Chiang YC, Hwang CC, Hsu TW, Huang CC, Hung KH, Tsai KC, Wang KH, Osada N, Schaal BA, Chiang TY (2012a) Footprints of natural and artificial selection for photoperiod pathway genes in *Oryza*. Plant J 70:769–782

Huang P, Molina J, Flowers JM, Rubinstein S, Jackson SA, Purugganan MD, Schaal BA (2012b) Phylogeography of Asian wild rice *Oryza rufipogon*: a genome-wide view. Mol Ecol 21:4593–4604

Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Yan Z, Liu K, Lu H, Li W, Guo Y, Lu Y, Zhou C, Fan D, Weng Q, Zhu C, Huang T, Zhang L, Wang Y, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan X, Xu Q, Dong G, Zhan Q, Li C, Fujiyama A, Toyoda A, Lu T, Feng Q, Qian Q, Li Y, Han B (2012c) A map of rice genome variation reveals the origin of cultivated rice. Nature 490:497–501

Izawa T (2007) Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice. J Exp Bot 58:3091–3097

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of DNA sequences. J Mol Evol 16:111–120

Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. Trends Genet 23:578–587

Liu Z (1975) The origin and development of cultivated rice in China. Acta Genet Sinica 2:23–29

Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA (2006) Phylogeography of Asian wild rice *Oryza rufipogon* reveals multiple independent domestications of cultivated rice *O. sativa*. Proc Natl Acad Sci USA 103:9578–9583

Lu BR, Cai XX, Jin X (2009) Efficient *indica* and *japonica* rice identification based on the InDel molecular method: its implication in rice breeding and evolutionary research. Prog Nat Sci 19:1241–1252

Lu L, Yan W, Xue W, Shao D, Xing Y (2012) Evolution and association analysis of *Ghd7* in rice. PLoS ONE 7:e34021

Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang P, Jackson S, Schaal BA, Bustamante CD, Boyko AR, Purugganan MD (2011) Molecular evidence

for a single evolutionary origin of domesticated rice. Proc Natl Acad Sci USA 108:8351–8356

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Qiao WH, Chen YT, Wang RS, Wei X, Cao LR, Zhang WX, Yang QW (2012) Nucleotide diversity in *Waxy* gene and validation of single nucleotide polymorphism in relation to amylose content in Chinese microcore rice germplasm. Crop Sci 52:1689–1697

Rozas J (2009) DNA sequence polymorphism analysis using DnaSP. Methods Mol Biol 537:337–350

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sang T, Ge S (2007) Genetics and phylogenetics of rice domestication. Curr Opin Genet Dev 17:533–538

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

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. Proc Natl Acad Sci USA 106:4555–4560

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA 4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Thompson JD, Gibson-Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Ting Y (1957) The origin and evolution of cultivated rice in China. Acta Agr Sinica 8:243–260

Tsuji H, Tamaki S, Komiya R, Shimamoto K (2008) Florigen and the photoperiodic control of flowering in rice. Rice 1:25–35

Vaughan DA, Morishima H, Kadowaki K (2003) Diversity in the *Oryza* genus. Curr Opin Plant Biol 6:139–146

Vaughan DA, Lu BR, Tomooka N (2008) The evolving story of rice evolution. Plant Sci 174:394–408

Wang MX, Zhang HL, Zhang DL, Qi YW, Fan ZL, Li DY, Pan DJ, Cao YS, Qiu ZE, Yu P, Yang QW, Wang XK, Li ZC (2008) Genetic structure of *Oryza rufipogon* Griff. in China. Heredity 101:527–535

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Wei X, Qiao WH, Chen YT, Wang RS, Cao LR, Zhang WX, Yuan NN, Li ZC, Zeng HL, Yang QW (2012a) Domestication and geographic origin of *Oryza sativa* in China: insights from multilocus analysis of nucleotide variation of *O. sativa* and *O. rufipogon*. Mol Ecol 21:5073–5087

Wei X, Wang RS, Cao LR, Yuan NN, Huang J, Qiao WH, Zhang WX, Zeng HL, Yang QW (2012b) Origin of *Oryza sativa* in China inferred by nucleotide polymorphisms of organelle DNA. PLoS ONE 7:e49546

Yang C, Kawahara Y, Mizuno H, Wu J, Matsumoto T, Itoh T (2011) Independent domestication of Asian rice followed by gene flow from *japonica* to *indica*. Mol Biol Evol 29:1471–1479

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) *Hd1* a major photoperiod sensitivity quantitative trait locus in rice is closely related to the Arabidopsis flowering time gene *CONSTANS*. Plant Cell 12:2473–2484

Zhang HL, Zhang DL, Wang MX, Sun JL, Qi YW, Li JJ, Wei XH, Han LZ, Qiu ZE, Tang SX, Li ZC (2011) A core collection and mini core collection of *Oryza sativa* L. in China. Theor Appl Genet 122:49–61

Zhao K, Wright M, Kimball J, Eizenga G, McClung A, Kovach M, Tyagi W, Ali MC, Tung CW, Reynolds A, Bustamante CD, McCouch SR (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PLoS ONE 5:e10780

Zhu QH, Ge S (2005) Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. New Phytol 167:249–265

Zong Y, Chen Z, Innes JB, Chen C, Wang Z, Wang H (2007) Fire and flood management of coastal swamp enabled first rice paddy cultivation in east China. Nature 449:459–462