Plant Science 242 (2016) 187–194

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Both *Hd1* and *Ehd1* are important for artificial selection of flowering time in cultivated rice

Fu-Jin Wei[a,b,*], Yuan-Ching Tsai[a], Hshin-Ping Wu[a], Lin-Tzu Huang[a], Yu-Chi Chen[c], Yi-Fang Chen[d], Cheng-Chieh Wu[a,e], Yi-Tzu Tseng[a,e], Yue-ie C Hsing[a,*]

[a] *Institute of Plant and Microbial Biology, Academia Sinica, Nankang, Taipei 115, Taiwan*
[b] *Department of Agronomy, National Taiwan University, Taipei 106, Taiwan*
[c] *Taiwan International Cooperation and Development Fund, Taipei 111, Taiwan*
[d] *Soil and Water Conservation Bureau, Council of Agriculture, Nantou 540, Taiwan*
[e] *Institute of Botany, National Taiwan University, Taipei 106, Taiwan*

ARTICLE INFO

Article history:
Received 27 April 2015
Received in revised form 31 July 2015
Accepted 4 September 2015
Available online 10 September 2015

Keywords:
Domestication
Heading date
Introgression
Next-generation sequencing
Rice

ABSTRACT

Rice is a facultative short-day plant, and it requires a photoperiod shorter than the critical day length to get flowering. Sensitivity to photoperiod has been suggested as a major selection target in cultivated or weedy rice. The modern rice varieties in Taiwan may be cultivated twice a year. These varieties contain loss-of-function of two important flowering-time related genes, *Heading date 1* (*Hd1*) and *Early heading date 1* (*Ehd1*), and are mainly from a mega variety, Taichung 65. However, the parental lines of this variety were sensitive to photoperiod, thus, how Taichung 65 loss its sensitivity is a mystery. In this study, we used accession-specific single nucleotide polymorphism analysis to reveal the gene flow that occurred between different rice accessions decades ago and demonstrate that two landraces introgressed during the breeding process, which led to the loss of photoperiod sensitivity. Both *Hd1* and *Ehd1* may be important during artificial selection for flowering time, especially in a subtropical region such as Taiwan. This is a good example of introgression playing important roles during rice domestication.

1. Introduction

Introgression or hybridization may play an important role in crop domestication. For rice, gene flow and genetic isolation with several domestication-related genes has been used to reveal the dynamics of rice domestication [1,2]. One of the early domestication traits found was shattering (SH), a quantitative trait loci (QTL) that controls cell-wall degradation and therefore, established the abscission layer between the grain and panicle. *SH4* is a putative transcription factor [3] and *SH1* encodes a homeobox transcription factor [4]. Because *SH4* is present in many *indica* and *japonica* rice accessions, the *SH4* allele was considered to have appeared and been selected before the differentiation of *indica* and *japonica* from a common wild rice ancestor or selected very early during the domestication of one rice subpopulation and then introgressed to others. Heerwaarden and colleagues used single nucleotide polymorphism (SNP) datasets for a large number of accessions of both teosinte and maize and demonstrated gene flow between maize and its wild relatives [5]. A recent spontaneous DNA introgression was found from domesticated polyploid wheat into its distantly related, wild tetraploid *Aegilops peregrine,* with the stabilization of this sequence in wild populations despite their not having homologous chromosomes. Thus, DNA may spontaneously introgress between homoeologous genomes of species of the tribe Triticeae [6].

Flowering time, also known as heading date for rice, is an important trait during rice domestication [7]. Flowering time is affected by photoperiod (day length) and also temperature [8]. As a facultative short-day plant, rice flowering is promoted under short-day conditions and delayed under long-day conditions, with the critical day length around 13 h [9,10]. Many recent reviews provided detailed information on the regulation of rice flowering [11–13]. A few key genes have been suggested to affect flowering time,

Abbreviations: BVG, basic vegetative growth; FNP, functional nucleotide polymorphism; IGV, integrative genomics viewer; LTR, long terminal repeat; QTL, quantitative trait loci; SNP, single nucleotide polymorphism; TC65, Taichung 65; TNG67, Tainung 67; VCF, variant call format.

* Corresponding author.
E-mail addresses: masker@gate.sinica.edu.tw (F.-J. Wei), botsai@gate.sinica.edu.tw (Y.-C. Tsai), hpwu@gate.sinica.edu.tw (H.-P. Wu), melissa@gate.sinica.edu.tw (L.-T. Huang), yc.chen@icdf.org.tw (Y.-C. Chen), r94621121@ntu.edu.tw (Y.-F. Chen), ccwu1212@gate.sinica.edu.tw (C.-C. Wu), jillts@gate.sinica.edu.tw (Y.-T. Tseng), bohsing@gate.sinica.edu.tw (Y.-i.C. Hsing).

http://dx.doi.org/10.1016/j.plantsci.2015.09.005
0168-9452/

including *Heading date 1* (*Hd1*), a B-box zinc finger protein and the ortholog of Arabidopsis CONSTAINS [14]; *Early heading date* (*Ehd1*), a B-type response regulator with no ortholog in Arabidopsis [15]; and *Hd3a*, a phosphatidylethanolamine-binding protein and the ortholog of Arabidopsis FT [16].

Using a core collection of 64 rice cultivars that represent the genetic diversity of 332 accessions from around the world, Takahashi et al., demonstrated that the variation in *Hd1*, *Ehd1*, and *Hd3a* contributed to the flowering time diversity in cultivated rice [17]. They suggested that Hd1 proteins, *Hd3a* promoters, and *Ehd1* expression levels together contributed to the diversity of flowering time. Later on, using 60 accessions of *Oryza sativa* and 38 accessions of *Oryza rufipogon*, this group showed that *Hd3* was highly conserved over time, whereas *Hd1* underwent human selection. Thus, they suggested *Hd1* was a possible target for domestication in the heading-date trait [7]. The flowering strategy of US weedy rice was studied recently and again *Hd1* was found important, although the *Hd1* haplotype alone could not fully explain the differences in flowering phenotype [18].

In temperate regions such as Japan, modern rice varieties such as Nipponbare are transplanted into the field in early June and harvested in late fall, for only one cropping season each year, and the yield is reasonable. In subtropical regions like Taiwan, modern local rice varieties such as Tainung 67 (TNG67) may be planted twice a year, with the first cropping season from February to June and the second from July to November. That is, seeds harvested from the first cropping season may germinate and be planted the next month. However, if Nipponbare were planted in Taiwan, it would flower very early or very late, depending on the transplanting time, so grain yield would be affected. However, how Taichung 65 (TC65, abbreviated as T65 in some papers), a variety that may be grown two times a year in Taiwan, lost its photoperiod sensitivity even though both of its parental lines are sensitive to photoperiod, has been a puzzle for a long time.

In the present study, we used SNPs obtained from whole genome sequencing to examine how TC65 gained different photoperiod sensitivity. We found that introgression must have occurred decades ago from upland rice accessions grown nearby and that TC65 had four accessions in its pedigree. This is a good example of introgression playing important roles during rice domestication.

2. Materials and methods

2.1. Plant materials and growth conditions

Rice TC65 seeds were obtained from Professor Motoyuki Ashikari (Nagoya University, Japan). The seeds of TNG67, Kameji, Shinriki and all Taiwan landraces as well as modern varieties were obtained from the Plant Germplasm Division, Taiwan Agriculture Research Institute, Taiwan. The breeders' seeds harvested decades ago were used for the modern varieties.

For DNA extraction, the plants of single seed descent were cultivated until tillering stage in an Academia Sinica greenhouse under natural light. Healthy leaves without insect damage from one plant were harvested, frozen under liquid nitrogen and stored at −80 °C. For heading date and agronomic trait analysis, the plants were cultivated in the experimental paddy field under natural light at agricultural research institutes or stations, including Academia Sinica (Taipei, 2012), Taiwan Agriculture Research Institute (Taipei, 1961, 1962), Taiwan Agriculture Research Institute (Taichung, 2012, 2014), and Tainan District Agricultural Research and Extension Station (Tainan, 2014). Twenty plants per accession, with two duplicates in Tainan, were transplanted. Heading date was defined as the time when the first panicle appeared from the node in half of these plants.

2.2. Genomic DNA extraction, PCR analysis, and sequencing

Genomic DNA was extracted from leaves by using a DNeasy Plant Mini Kit (Qiagen). After quality assessment, the genomic DNA was randomly fragmented and size-fractionated. DNA fragments with desired lengths were gel-purified. For whole-genome resequencing, paired-end libraries with 450–500-bp inserts were constructed and sequenced on a GA2 or HiSeq2000 system (Illumina). Sequence data were deposited in the NCBI Sequence Read Archive. The functional SNP (FNP) for *Hd1*, *Ehd1*, *qSH1* and *SH4* were amplified by PCR followed by capillary sequencing analysis. Primer sequences are in Table S1.

2.3. SNP and indel calling

Adaptor sequences, low-quality bases and reads <20 bp long were discarded. The trimmed paired reads were then aligned to the reference rice Nipponbare genome sequence (IRGSP v1.0). SAMtools and VCFtools [19] were used to manipulate and transform the SAM and variant call format (VCF) [19] file format. To detect SNPs and small indels, we used the command lines in the section "EXAMPLES" in the SAMtools manual without any restriction on depth or mapping quality. The information on SNP and small indels was recorded in VCF files, compared by "vcf-isec" to classify sample-specific or intersection variants, then imported into Integrative Genomics Viewer (IGV) [20,21] to show alleles or genotypes.

All split VCF files were filtered and checked by visualization, then the regional files were output in BED format (.bed). We used snapshot image files from IGV [21] for variant events on the NGS mapped images. Locations of SNPs were then manually validated. The curated SNPs along the 12 chromosomes were plotted by using BasicChromosome module of Biopython [22]. SNPeff [23] was used to calculate the location of each SNP in the genome.

2.4. Estimation of heterozygosity

The VCF files were used to estimate heterozygosity, then the common SNPs for the 5 accessions were filtered out. To screen for heterozygous SNPs, 10 kb was used as a window size. If the number of heterozygous SNPs was greater than two times that of the homozygous SNPs, the block was considered heterozygous. In total, 37,440 blocks occur in the rice genome, and the percentage heterozygosity was then calculated.

2.5. Measurement of agronomic traits

The color of apiculus, presence of awn, flag leaf length and the plant types of the accessions were monitored by visualization. Shattering was measured as the percentage of spikelets that fell from each panicle. Near-infrared spectroscopy was used to estimate seed protein, amylose, moisture content and brown rice score (NIRT Grain Tester AN-900, Kett Electric Lab, Tokyo). The resistance to seedling blast was monitored as R (resistant), M (moderate), and S (susceptible). The resistance to sheath blight was screened by degree of severity of sheath blight (H): $H = 3n_1 + 2n_2 + 1n_3 + 0n_4/3N \times 100$, where $N = n_1 + n_2 + n_3 + n_4$ (n_1 = number of leaves with symptoms in the flag leaf and second, third and fourth leaf blade or sheath; n_2 = number of leaves with symptoms in the second, third and fourth leaf blade or sheath; n_3 = number of leaves with symptoms in the third and fourth leaf blade or sheath and any leaf underneath; and n_4 = number of healthy leaves [24]).

Table 1
Heading date for Taichung 65 (TC65) and its related accessions transplanted in different locations, years and dates (Trans. date).[e]

Location	Year	Trans. date	TC65	Kameji	Shinriki	Muteka	Nakabo	Habun1
Taipei[a]	1961	8-August	64	52	50	68	69	68
Taipei[b]	2012	6-April	81	124	134	86	82	83
Taichung[c]	2014	14-Febauary	88	72	72	84	84	81
Tainan[d]	2014	14-Febauary	81	75	71	81	80	79
Tainan[d]	2014	25-July	62	47	50	61	60	65

[a] Plants cultivated at the experimental farm of the Taiwan Agriculture Research Institute, Taipei.
[b] Plants cultivated at the experimental farm of Academia Sinica, Taipei.
[c] Plants cultivated at the experimental farm of the Taiwan Agriculture Research Institute, Taichung.
[d] Plants cultivated at the experimental farm ofthe Tainan District Agricultural Research and Extension Station, Tainan.
[e] Twenty plants per accession were transplanted on the day indicated. Heading date, half of the plants were heading, from the date of transplantinfto the date of heading.

2.6. Analysis of genetic distance and diversity

The genomic region, including promoter, open reading frame, intron, and UTR, for 10 domestication-related genes in the five accessions was used for analysis of genetic distance and diversity. Nucleotide diversity (π) [25], θw [26], and Tajima's D [27] were calculated by using DNASP v5.0 [28]. Insertion, deletion, and unalignable regions were excluded from the analysis.

3. Results

3.1. Response of TC65 and related races to different transplanting time

The parental lines of TC65 were Kameji and Shinriki, two Japanese landraces [29]. Table 1 lists the heading date for TC65, Kameji and Shinriki transplanted in the first cropping season (February or April) or second cropping season (July or August) in northern (Taipei), central (Taichung), and southern (Tainan) Taiwan. Both landraces headed before or after TC65, with a 1–7-week difference. They showed delayed flowering, by about 1 1/2 months, in the first cropping season (i.e., the one transplanted at early April in 2012). Thus, both Kameji and Shinriki are sensitive to photoperiod and cannot be cultivated twice a year, whereas TC65 can.

About 60 upland rice accessions (Table S2) were collected from Taiwanese aboriginal villages at an early stage of Japanese colonial time, around 1900. They were since propagated (renewed) about every 10 years by rice breeders, and seeds were stored in the germplasm center. We previously studied the agronomic traits and domestication-related genes of these landraces [30]. Some landraces are sensitive to photoperiod and some are not. Thus, we used the tool Targeting-induced Local Lesions or PCR analysis to check six domestication-related genes controlling HD—*Ehd1*, *Ehd2*, *Hd1*, *Hd3a*, *Hd6*, and *Se5* [30]—and demonstrated large variations in sequence for some of these genes. Surprisingly, some of the accessions, cultivated in the mountain area for thousands of years, showed TC65 haplotypes for *Hd1* and *Ehd1*. Why they contain identical sequence changes as TC65, which appeared in the 1930s, is a mystery.

The most obvious difference in the TC65 haplotype of *Hd1* was an 1.9-kb insertion at exon 2 [15]. Therefore, TC65 *Hd1* is a recessive trait with the loss-of-function HD1 protein. Fig. S1 shows the gel analysis for 65 accessions of this region. The *Ehd1* gene for TC65 had a nucleotide substitution: one Gly was changed to Arg, which abolished the protein function [15]. Table S2 lists the TC65 haplotypes for *Hd1* and *Ehd1* for three modern varieties, including TNG67, Nipponbare, and IR64, and many landraces. Eleven upland rice accessions had the TC65 *Hd1* haplotype and seven had the TC65 *Ehd1* haplotype; three—Muteka, Nakabo, and Habun1—had both. Most accessions are *japonica* rice, with one *indica* race mutated in *Hd1* and one in both *Hd1* and *Ehd1*. We then checked the heading date of these three accessions in three locations and under four transplanting times (Table 1). All showed similar flowering time as TC65. In addition, we checked the heading date of the accessions with only mutated *Hd1* or *Ehd1*, as well as four wild types (*Hd1Ehd1*) (Table S3). The flowering time difference from TC65 was from 1 month for *Hd1* lines to 3 months for *Ehd1* lines. Thus, the accessions were not the source of the introgression.

Table 2
Comparison of single nucleotide polymorphism (SNP) rates for five accessions and TC65.

	Polymorphism (SNP)[a]	Shared SNPs[b]
Kameji	584,536	35,972
Shinriki	116,783	44,688
Muteka	387,196	70,161
Nakabo	168,048	63,732
Habun1	2,717,967	34,027

[a] SNP rate for each accession with TC65. The numbers are raw data from the SAMtools output without manual checking.
[b] Shared SNP rate for each accession with TC65. The numbers are curated data from the SAMtools output with manual checking, as indicated in Methods.

3.2. Resequencing and SNP calling

We resequenced the genomes of the rice varieties/races related to the TC65 pedigree using the Illumina HiSeq platform. We obtained paired-end sequence data averaging 7–18-fold depth, with the coverage for each genome 90–98% (Table S4). The FASTQ files are available in the NCBI Archives [accession nos. SRR1956774 (TC65), SRR1956769 (Kameji), SRR1956772 (Shinriki), SRR2106584 (Muteka), SRR2106585 (Nakabo) and SRR2106586 (Habun1)].

All sequence reads were aligned to the reference Nipponbare genome sequence (IRGSP v1.0) by using BWA [31]. We used SAMtools [32] to call SNPs. Table 2 lists the SNP rates for each accession with TC65. The SNP rates for Shinriki, Kameji, Muteka, and Nakabo were from 100,000 to 600,000, in the same log scale as for other *japonica* varieties. However, Habun1 had >2.7 million SNPs and thus should be an *indica* type. Table 2 also lists the shared SNPs for each accession with TC65. We manually checked these shared ones as indicated in Methods because they are important for studying relationships. A Venn diagram indicated that the specific SNPs shared between TC65 and Muteka, Nakabo, Kameji, and Shinriki were 1754, 641, 421, and 310, respectively (Fig. 1). Thus, Mutake should be the accession most closely related to TC65, followed by Nakabo, Kameji, and Shinriki. Because (1) Habun1 is an *indica* rice and the hybrid with *japonica* should have low fertility and (2) the shared SNP rate with TC65 was relatively low (Table 2), we excluded it from the following discussion.

According to genome annotation (RAPdb, released 2015), 62.7% of all SNPs in TC65 were found in intergenic regions, 11.4% in promoter regions, 5.1% in UTR regions, 10.5% in intron regions, and 10.3% in gene coding regions. The exact number of SNPs in different regions for all five accessions are in Table S5. The SNPs led to

Table 3
Some agronomic traits for TC65 and its related accessions.

	TC65	Kameji	Shinriki	Muteka	Nakabo
Color of apiculus	Light green	Light green	Light green	Dark purple	Light green
Presence of awn	no	no	no	no	no
Shattering	31.20%	26.06%	38.85%	27.27%	43.85%
Brown rice score	60.00 ± 4.32	50.67 ± 2.49	54.33 ± 2.87	55.67 ± 5.73	61.00 ± 0.00
Protein (brown rice)	9.10 ± 0.50	9.70 ± 0.22	9.20 ± 0.36	8.83 ± 0.93	8.80 ± 0.30
Amylose (brown rice)	18.67 ± 0.17	18.50 ± 0.24	18.53 ± 0.39	18.17 ± 0.34	17.37 ± 1.60
Total grain weight	30.86 ± 7.84	20.93 ± 1.64	16.20 ± 6.89	11.22 ± 2.61	27.54 ± 5.27
Flag leaf length	31.0	32.5	22.5	31.0	27.5
Plant type	erect	open	erect	erect	erect
Resistance to seedling blast[a]	S	R	S	RM	RM
Degree of severity to sheath blight[b]	9.6%	38.1%	20.2%	26.2%	5.0%

[a] The resistance to blast was screened at the experimental farm of the Taiwan Agriculture Research Institute at the first cropping season, 1962. S: sensitive, R: resistant, M: moderate.
[b] The resistance to blight was screened at the experimental farm of the Taiwan Agriculture Research Institute at the first cropping season, 1962. The percentage is degree of severity of sheath blight (Yoshimura 1954).

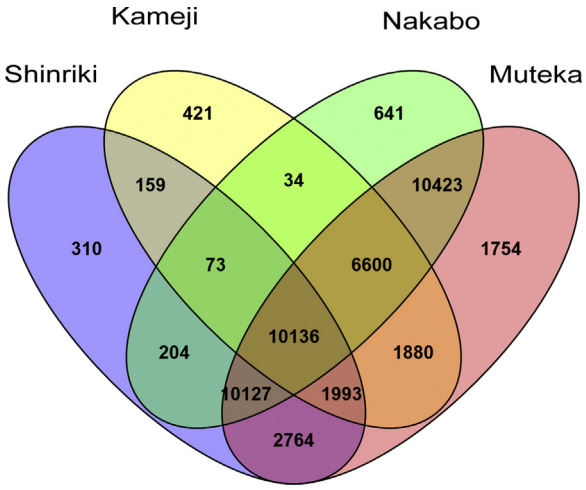

Fig. 1. Venn diagram for shared single nucleotide polymorphisms (SNPs) between Taichung 65 (TC65) and Shinriki, Kameji, Nakobo, and Muteka.

an early stop for 36 proteins in TC65 (shown in Table S6). Half of these proteins are hypothetical proteins. Transcription factors and kinases occupy half of the remainder.

3.3. Genome analysis of TC65 and its related races

Fig. 2 illustrates regions in the rice genome showing a TC65 genome similar to that of Kameji, Shinriki, Muteka, or Nakabo. All sequences were aligned to the Nipponbare sequence (IRGSP v1.0). Each region is 100 kb. Panel A shows the 1600–1700 K region of chromosome 4, and the SNP patterns in these 5 lanes indicate that the TC65 genome is similar to that of Kameji. Panel B is a 16,260–16,360 K region of chromosome 3, and the SNP patterns indicate TC65 genome similar to Shinriki. Panel C is a 1520–1620 K region of chromosome 6, and the SNP patterns indicate TC65 genome similar to Muteka. Panel D is an 8540–8640 K region of chromosome 3, and the SNP patterns indicate TC65 genome similar to Nakabo. Panel E is a 12,920–13,020K region of chromosome 7, and the SNP patterns indicate TC65 genome similar to the region with shared SNP patterns for both Muteka and Nakabo.

Fig. 3 shows the distribution of the accession-specific SNPs and several domestication-related genes for the 12 chromosomes of TC65. The Shinriki-specific region (red, i.e., shared by TC65 and Shinriki only) is larger than the Kameji-specific region (green). However, more genomic regions are Muteka-specific (orange), Nakabo-specific (blue), or shared by Muteka and Nakabo (purple). The white region indicates areas with shared SNPs among TC65, Kameji, Shinriki, Muteka, or Nakabo. Thus, Kameji and Shinriki were not the only parental lines. Muteka and Nakabo were also in the TC65 pedigree; that is, introgression occurred during the breeding process. Because the white area occupies about half of the genome, we also plotted the regions shared by four (any combination of four), three, two, or one accessions using a grey scale (Fig. S2). The white area in this figure indicates regions where all accessions and Nipponbare share the same sequences. We plotted several domestication-related genes (Fig. 3). For the heading-date—related genes, *Hd1, Ehd1, Ehd2* and *Ehd3*, all located at a TC65, Muteka and Nakabo shared region, and *Hd3a* located at a shared region in TC65 and Muteka. Thus, many of the flowering genes in TC65 locate in genome regions from Muteka and/or Nakabo.

In addition, we found 530 TC65-specific SNPs which were not present in Nipponbare and other four accessions. We suggested these sequence changes mainly came from spontaneous mutation during the last 90 years.

Fig. 2 shows that the upland rice Muteka contains heterozygous SNPs in one entire 100-kb region (panel B, lane 4). To estimate the heterozygosity percentage for TC65 and its relatives, we used a block method. The heterozygosity was 2.0%, 16.7%, 16.1%, 48.1%, 11.5%, and 5.3% for TC65, Kameji, Shinriki, Muteka, Nakabo, and Habun1, respectively. Thus, some landraces still contain heterozygous regions in their genome, with the highest frequency for the upland rice Muteka.

To reveal the genetic relationships of TC65 with the related accessions, SNPhylo [33] was used to plot the relationship of TC65 and related accessions as well as another 22 *japonica* and 12 *indica* rice accessions (Fig. 4). The materials included Nipponbare, TC65, Kameji, Shinriki, Muteka, Nakabo, thirteen modern *japonica* and *indica* Taiwanese varieties and 11 upland rice landraces. TC65 is indeed very close to the two upland rice landraces Muteka and Nakabo but more distant from Shinriki or Kameji. Habun1 locates at the *indica* clade. In addition, TC65 is grouped with many modern Taiwanese varieties, all starting with T in the abbreviation, and very distant from five other upland *japonica* rice landraces (Montana, Matara, Chuan4, Nakairitsu, and RuiYan).

3.4. Trait improvement in TC65 from its pedigree

In addition to heading date, we analyzed nine other traits of TC65 and its related accessions (Table 3), including plant type, seed shattering, grain quality (brown rice score and protein content), grain yield, as well as disease resistance (against seedling blast and sheath blight). The apiculus color of Muteka is dark purple, as for wild rice, whereas the other four accessions have a light-green apiculus. The degree of shattering is higher for Shinriki and Nakabo than TC65, Kameji, and Muteka. TC65 has good

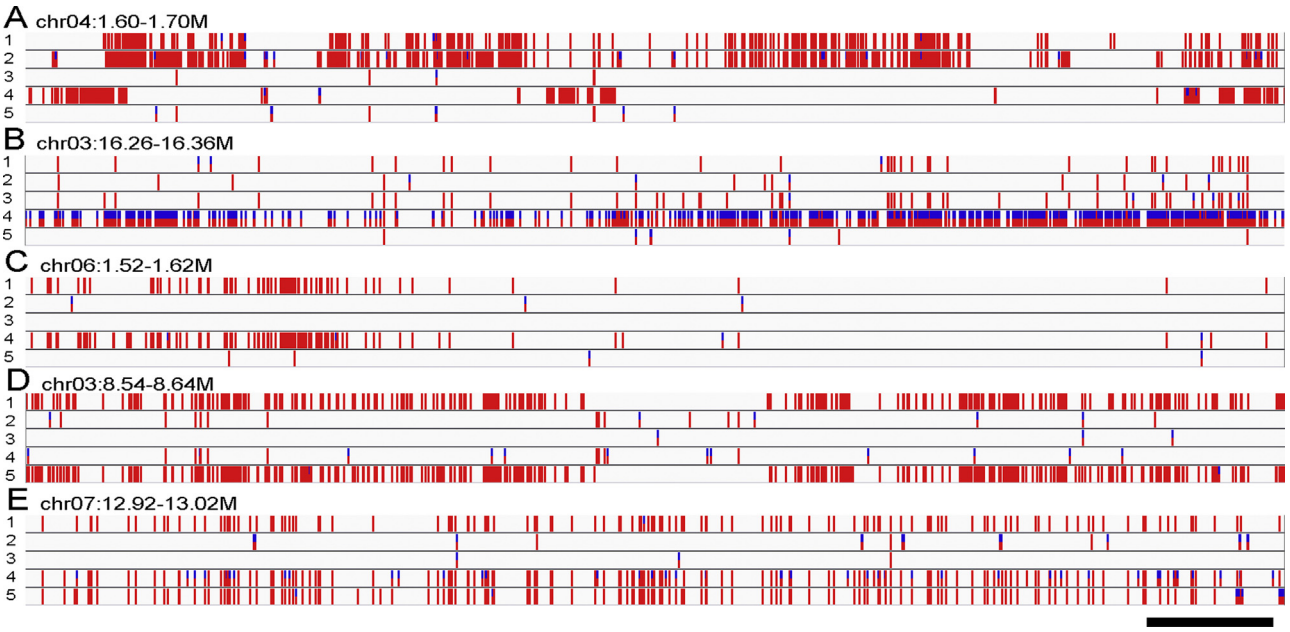

Fig. 2. Five types of introgressions present in TC65 chromosomes. All sequences were aligned to the Nipponbare sequence (IRGSP v1.0). Lane 1, SNPs in the TC65 genome; lane 2, SNPs in the Kameji genome; lane 3, SNPs in the Shinriki genome; lane 4, SNPs in the Muteka genome; lane 5, SNPs in the Nakabo genome. Scale bar: 10 kb.

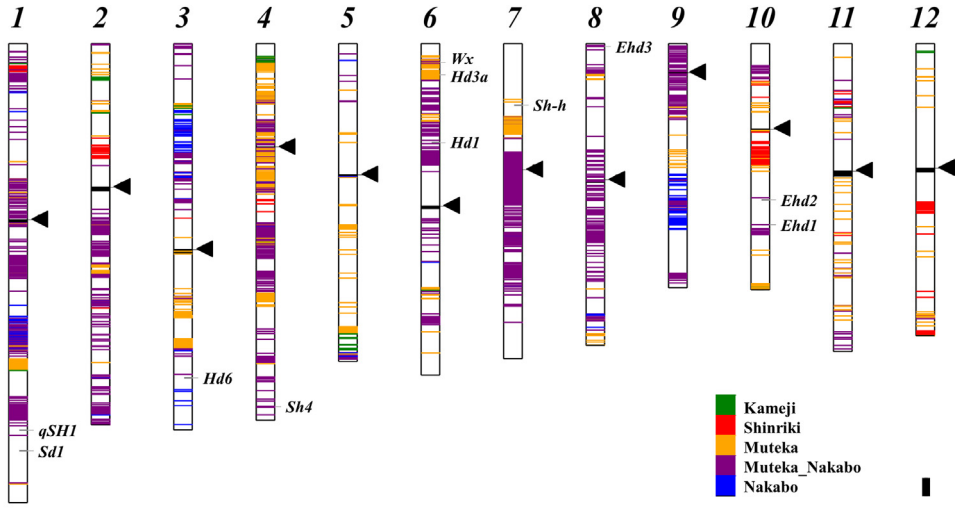

Fig. 3. Distribution of the accession-specific genome block and several domestication-related genes along the 12 chromosomes. The accessions are indicated. The positions of six heading-date–related traits, including Hd1, Hd3a, Hd6, Ehd1, Ehd2, and Ehd3; three shattering-related traits, including qSH1, SH4, and Sh-h; one grain-quality—related trait, Wx; and one plant-height—related trait, Sd1, are marked. Black arrowhead indicates the position of centromeres. Scale bar: 2 Mb.

eating quality (high brown rice score and low protein content), and Nakabo is the potential donor. The seed yield of TC65 is highest among all 5 lines. TC65 is relatively resistant to sheath blight, and the potential donor is Nakabo. Although Kameji, Muteka, and Nakabo are resistant to blast, TC65 is sensitive to this disease. However, the breeding program was conducted from 1925 to 1929, and the disease-resistance screening was performed in 1962. The results might not represent the resistance behavior during the breeding process decades ago because the pathogen races might be different.

Rice seed shattering is controlled by QTL; three are well studied: *qSH1* [4], *SH4* [3], and *Sh-h* [34]. Table 4 lists the haplotypes of shattering genes for each accession used in the study. All five lines have the same genotype (i.e., non-Nipponbare *qSH1* haplotype, and Nipponbare haplotypes for both *SH4* and *Sh-h*) although their shattering behaviors have two types. Therefore, other sequence changes in the *qSH1* gene and/or other unknown shattering genes may control the different phenotypes in the five accessions.

Fig. 3 illustrates that selection footprint of several heading-date—related genes revealed in the genome regions. We further investigated the eating-quality—and disease-resistant–related genes (Fig. S3). The genes related to rice eating and cooking quality were studied recently [35] and were mainly starch-synthesis—related genes. *SSII-1, SSII-2, Sbe3, Sbe4, GBSSII* and *ISA* locate in the genome blocks shared among TC65, Muteka and Nakabo (panel A). However, *Wx, SSI,* and *SSIV-1* locate in the region shared between TC65 and Muteka. Most of the other eating-quality—related genes locate in regions shared among all five accessions. A total of 130 disease-resistant—related genes were collected from the RAP database (released 2015) and plotted in panel B. Five clusters of these genes are mapped to regions shared among TC65, Muteka and Nakabo in chromosome 1 and one region in chro-

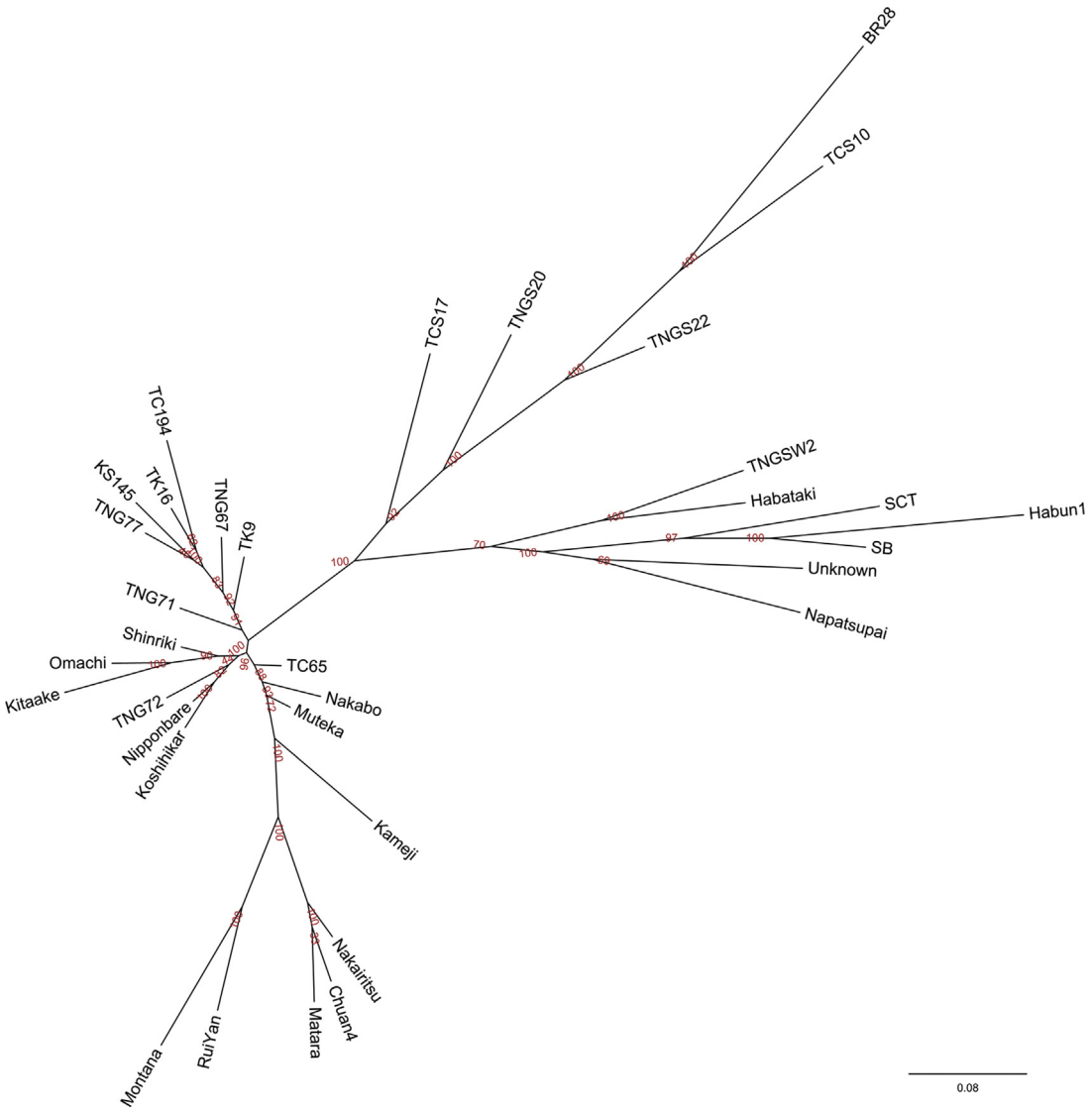

Fig. 4. Phylogenetic tree of several modern rice varieties and landraces. The tree was constructed with SNP data from 34 genomes by using SNPhylo (Lee et al. [33]). The elipse at the left includes 22 japonica types and at the right 12 indica types. The bootstrap values determined with 1000 samples are in red. TK9: Taikang 9; TNG67: Tainung 67; TC194: Taichung 194; TK16: Taikang 16; KS145: Kaoshiung 145; TNG77: Tainung 77; TNG71: Tainung 71; TNG72: Tainung 72; SB: San-Bae; SCT: Shung-Chiang-Tsaw; TNGSW2: Tainung Sen Waxy 2; TNGS22: Tainung Sen 22; TCS10: Taichung Sen 10; TNGS20: Tainung Sen 20; TCS17: Taichung Sen 17.

Table 4
The haplotypes of shattering and heading-date genes for each variety/race used in the study.

Genotype					
	Hd1[a]	*Ehd1* [b]	*qSH1* [c]	*SH4* [d]	*Sh-h* [e]
TC65	TC65	TC65	Non-NP[f]	Nippon[g]	Nippon
Kameji	Nippon	Nippon	Non-NP	Nippon	Nippon
Shinriki	Nippon	Nippon	Non-NP	Nippon	Nippon
Muteka	TC65	TC65	Non-NP	Nippon	Nippon
Nakabo	TC65	TC65	Non-NP	Nippon	Nippon

[a] *Hd1*: LOC_Os06g16370 or Os06g0275000. The functional nucleotide polymorphism (FNP) of TC65 locates at chr06:9338032, with a 1.9-kb insertion at exon 2 (Doi et al. [15]).
[b] *Ehd1*: LOC_Os10g32600, Os10g0463400. The FNP of TC65 locates at chr10:17077589, with the change G>R (amino acid) (Doi et al. [15]).
[c] *qSH1*: LOC_Os01g62920 or Os01g0848400. The FNP of Nipponbare locates at chr01:36461792, with the change G>T at the promoter region (Konishi et al. [4]).
[d] *SH4*: LOC_Os04g49810 or Os04g0670900. The FNP of Nipponbare locates at chr04:34232985, with the change N>K (amino acid) (Li et al. [3]).
[e] *Sh-h*: LOC_Os07g10690 or Os07g0207700. The FNP of Nipponbare locates at chr07:5810885, with A but not T at the splicing site (Ji et al. [34]).
[f] Non-Nipponbare type.
[g] Nipponbare type.

mosomes 6 and 9. Most of the other genes locate in regions shared among all five accessions.

3.5. Selection during breeding and introgression

To reveal the selection during breeding, we calculated nucleotide polymorphism (θ), nucleotide diversity (π) and Tajima's D for 10 domestication-related genes, including four heading-date–related genes *Hd1, Ehd1, Hd3a* and *Hd6*; two shattering-related genes, *qSH1* and *SH4*; one plant-height–related gene, *Sd1*; and three starch-synthesis—related genes, *Wx, Sb3,* and *ISA* (Table 5). Tajima's D values were negative for most of these genes except *Sbe3*. Therefore, many of the genes are under selection.

Because (1) the introgressed region resulted mostly from the region shared between Muteka and Nakabo, (2) the Muteka genome is still highly heterozygous, and (3) some of the Muteka mature seeds are open-hulled (Fig. S4), natural hybridization from Nakabo to Muteka may have occurred before it introgressed into the early generation of the Kameji x Shinriki cross, before TC65 was bred in 1929.

Table 5
Nucleotide diversity of several domesticated-related genes.

Gene	Tajima's D	θw	Nucleotide diversity (Π)
Hd1	−0.94326	0.00546	0.00479
Ehd1	−1.06809	0.00485	0.00414
Hd3a	−1.15043	0.00492	0.00416
Hd6	−1.22187	0.00172	0.00143
qSH1	−1.25334	0.00389	0.00324
SH4	−1.25024	0.00768	0.00640
Sd1	−0.74682	0.00137	0.00122
Wx	−1.23906	0.00260	0.00217
Sbe3	0.05280	0.00456	0.00458
ISA	−1.24874	0.00194	0.00161

π and θw, nucleotide diversity.

4. Discussion

4.1. The TC65 Hd1 haplotype is specific to Taiwan

The loss-of-function *Hd1* gene was revealed in an Aus cultivar Kasalath [14]. The frameshift caused by a 2-bp deletion led to abolishment of the CCT domain and non-functional Hd1. In an intensive study of *Hd1* gene diversity, Takahashi et al., classified the *Hd1* coding region into 17 types for the 68 varieties analyzed [17]. Only one accession was from Taiwan, and its non-function *Hd1* was caused by a 4-bp insertion in the CCT domain. In total, nine allele types were suggested to be nonfunctional, but the TC65 haplotype was not among this long list. It is specific to Taiwan.

The Muteka *Hd1* gene was sequenced (accession no. KR230393). The 1901-bp insertion at exon 2 has two identical long terminal repeats (LTRs), of 448 bp. The fragment contains a gag protein and thus is a partial retrotransposon. The Nipponbare genome has two other similar fragments, one in the 12.8-Mb region of chromosome 5 and the other in the 19.8-Mb region of chromosome 8. The mismatch of the two LTRs for chromosome 5 is 12 bp and that for chromosome 8 is 11 bp. All 5 accessions contain these two partial retrotransposons. The integration of this 1.9-kb fragment into the *Hd1* gene must have been a recent event, compared to the ones in chromosomes 5 and 8, because the two LTRs are identical [36]. However, the SNP in *Ehd1* may have been caused by a spontaneous mutation.

Many of the upland rice accessions analyzed are primitive, that is, some have a red pericarp or an extra-long awn (>10 cm) or are shattering. In addition, both *japonica* and *indica* types are included. The *Hd1* and *Ehd1* mutations, especially their combination, provided good traits for rice plants grown in a subtropical region such as Taiwan, with narrow photoperiod range (10.8–13.3 h) and warm weather. These accessions were cultivated, and natural crossing occurred. Because these landraces were cultivated in different tribes or villages featuring different languages in mountain areas and the communication between tribes was restricted, introgression should have occurred a long time ago and the gene flow occurred within different *japonica* rice accessions and from *japonica* to *indica* rice, such as for Habun1, during cultivation in the mountain area. According to Lin [37,38], about 80% of the *japonica* varieties developed after the 1940s in Taiwan were descendants of TC65. Thus, the introgression that occurred during the TC65 breeding process about 90 years ago still affects modern rice breeding.

4.2. Spontaneous outcrossing occurs often in the rice field

Two *japonica* varieties, TNG67 and glutinous Tainung 73 (TNG73), were used to study the transgene escape mediated by pollen dispersal [39]. The outcrossing rate of TNG73 seeds was quite high at 1 m away from the pollen donor TNG67 and decreased to 0% beyond 40 m. In addition, the outcrossing rate differed in different directions and was mainly related to wind direction. Thus, pollen dispersal in the rice field occurred and may have caused introgression. In a recent study from our group (Wei et al., submitted), we found some cases of spontaneous outcrossing in the seeds we used to study somaclonal variation.

4.3. Pedigree resequencing uncovers outcrossing that occurred 90 years ago

The breeding program for TC65 started about 90 years ago. The cross was made in 1924, then the rice grew in the experimental station in the central part of Taiwan for two cropping seasons per year from 1925 to 1929. After 10 generations, the new variety was designated "Taichung 65" at the end of 1929 [29]. We suspected that the 60 upland rice accessions were propagated in the breeders' field during the selection period and introgression occurred in the field.

The TC65 *Hd1* and *Ehd1* genes are located in the shared genome region among TC65, Muteka and Nakabo (Fig. 3). The heading period of TC65, Muteka, Nakabo, and all the modern rice varieties in Taiwan differ by ~20 days between the first and second cropping seasons, that is, about 85 days in the first cropping season and about 65 days in the second cropping season, mainly because the temperature in the early growth period differs in two seasons. The TC65 *Ehd1* locus is a recessive allele [15]. The loss of function of this B-type response regulator led to prolongation of the long basic vegetative growth (BVG) [40]. Taiwan is located in a subtropical region, with annual differential day length (longest to shortest day length) <2.5 h. The loss-of-function *Hd1* led to less sensitivity to photoperiod, and the loss-of-function *Ehd1* led to long BVG, especially for the first cropping season. With the combination of the two alleles, the rice varieties would achieve sufficient vegetative growth before heading and fit well in the subtropical region. This combination led to uniformity in flowering time and the requirement for two cropping seasons per year during the breeding process. However, according to the record, these upland rice accessions were cultivated one cropping season each year at high latitude in the last thousands of years. Even though some had loss-of-function of both *Hd1* and *Ehd1*, these accessions could not have two cropping seasons in mountain areas because of low temperature.

Many previous studies indicated that *Hd1* was a possible target for domestication in heading-date trait. In the present study, we found that both *Hd1* and *Ehd1* may be important during artificial selection for flowering time, especially in the subtropical region.

Acknowledgements

We thank Ms. Lie-hong Wu for technical support and Ms. Laura Smales (BioMedEditing, Toronto, Canada) for English editing. This work was supported by an Academia Sinica (Taiwan) Investigator Award (AS-100-18) to YICH.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.plantsci.2015.09.005

References

[1] S.R. McCouch, M.J. Kovach, M. Sweeney, H. Jiang, M. Semon, The Dynamics of Rice Domestication: A Balance between Gene Flow and Genetic Isolation, in: P. Gepts, T.R. Famula, R.L. Bettinger, S.B. Brush, A.B. Damania, P.E. McGuire, C.O. Qualset (Eds.), Biodiversity in Agriculture, Cambridge University Press, 2012, 2015, p. 19.
[2] T. Sang, S. Ge, Understanding rice domestication and implications for cultivar improvement, Curr. Opin. Plant Biol. 16 (2013) 139–146.

[3] C. Li, A. Zhou, T. Sang, Rice domestication by reducing shattering, Science (New York, N. Y.) 311 (2006) 5.
[4] S. Konishi, T. Izawa, S.Y. Lin, K. Ebana, Y. Fukuta, T. Sasaki, M. Yano, An SNP caused loss of seed shattering during rice domestication, Science (New York, N. Y.) 312 (2006) 1392–1396.
[5] J. Heerwaarden, J. Doebley, W.H. Briggs, J.C. Glaubitz, M.M. Goodman, J. Jesus, S. Gonzalez, J. Ross-Ibarra, Genetic signals of origin, spread, and introgression in a large sample of maize landraces, Proc. National Acad. Sci. U. S. A. 108 (2011) 1088–1092.
[6] A. Cao, L. Xing, X. Wang, X. Yang, W. Wang, Y. Sun, C. Qian, J. Ni, Y. Chen, D. Liu, X. Wang, P. Chen, Serine/threonine kinase gene Stpk-V, a key member of powdery mildew resistance gene Pm21, confers powdery mildew resistance in wheat, Proc. National Acad. Sci. U. S. A. 108 (2011) 7727–7732.
[7] Y. Takahashi, K. Shimamoto, 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 Genet Syst. 86 (2011) 175–182.
[8] W. Luan, H. Chen, Y. Fu, H. Si, W. Peng, S. Song, W. Liu, G. Hu, Z. Sun, D. Xie, C. Sun, The effect of the crosstalk between photoperiod and temperature on the heading-date in rice, PloS One 4 (2009) e5891.
[9] B.S. Vergara, T.T. Chang, The Flowering Response of the Rice Plant to Photoperiod, 4 ed., International Rice Research Institute, 1985, 2015.
[10] H. Itoh, Y. Nonoue, M. Yano, T. Izawa, A pair of floral regulators sets critical day length for Hd3a florigen expression in rice, Nat. Genet. 42 (2010) 635–638.
[11] H. Itoh, T. Izawa, The coincidence of critical day length recognition for florigen gene expression and floral transition under long-day conditions in rice, Mol. Plant 6 (2013) 635–649.
[12] H. Tsuji, K. Taoka, K. Shimamoto, Florigen in rice: complex gene network for florigen transcription, florigen activation complex, and multiple functions, Curr. Opin. Plant Biol. 16 (2013) 228–235.
[13] R. Shrestha, J. Gomez-Ariza, V. Brambilla, F. Fornara, Molecular control of seasonal flowering in rice, arabidopsis and temperate cereals, Ann. Bot. 114 (2014) 1445–1458.
[14] M. Yano, Y. Katayose, M. Ashikari, U. Yamanouchi, L. Monna, T. Fuse, T. Baba, K. Yamamoto, Y. Umehara, Y. Nagamura, T. Sasaki, Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS, Plant cell 12 (2000) 2473–2484.
[15] K. Doi, T. Izawa, T. Fuse, U. Yamanouchi, T. Kubo, Z. Shimatani, M. Yano, A. Yoshimura, Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1, Genes Dev. 18 (2004) 926–936.
[16] S. Tamaki, S. Matsuo, H.L. Wong, S. Yokoi, K. Shimamoto, Hd3a protein is a mobile flowering signal in rice, Science (New York, N. Y.) 316 (2007) 1033–1036.
[17] Y. Takahashi, K.M. Teshima, S. Yokoi, H. Innan, K. Shimamoto, Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to diversity of flowering time in cultivated rice, Proc. National Acad. Sci. U. S. A 106 (2009) 4555–4560.
[18] C.S. Thurber, M. Reagon, K.M. Olsen, Y. Jia, A.L. Caicedo, The evolution of flowering strategies in US weedy rice, Am. J. Bot. 101 (2014) 1737–1747.
[19] P. Danecek, A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, R. Durbin, G. Genomes Project Analysis, The variant call format and VCFtools, Bioinformatics (Oxford, England) 27 (2011) 2156–2158.
[20] J.T. Robinson, H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, J.P. Mesirov, Integrative genomics viewer, Nat. Biotechnol. 29 (2011) 24–26.
[21] H. Thorvaldsdottir, J.T. Robinson, J.P. Mesirov, Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration, Briefings Bioinf. 14 (2013) 178–192.
[22] P.J.A. Cock, T. Antao, J.T. Chang, B.A. Chapman, C.J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M.J.L. de Hoon, Biopython: freely available Python tools for computational molecular biology and bioinformatics, Bioinformatics (Oxford, England) 25 (2009) 1422–1423.
[23] P. Cingolani, A. Platts, L. Wang le, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, D.M. Ruden, 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 (Austin) 6 (2012) 80–92.
[24] S. Yoshimura, On the scala for estimating degree of severity of Sheath blight by Hypochnus Sasakii SHIRAI in rice plant, Ann. Phytopathol. Soc. Japan 19 (1954) 58–60.
[25] M. Nei, W.H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases, Proc. National Acad. Sci. U. S. A. 76 (1979) 5269–5273.
[26] G.A. Watterson, On the number of segregating sites in genetical models without recombination, Theor. Popul. Biol. 7 (1975) 256–276.
[27] F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism, Genetics 123 (1989) 585–595.
[28] P. Librado, J. Rozas, DnaSP v5: a software for comprehensive analysis of DNA polymorphism data, Bioinformatics (Oxford, England) 25 (2009) 1451–1452.
[29] E. Iso, Lectures on Rice Cultivating in Formosa (Taiwan), Taiwan Noyuwai, Taiwan, 1944.
[30] J. -s, Y.- Hsieh, C. Hsing, T. -f, P.J.- Hsu, K.- Li, C.- Li, Studies on ancient rice—where Botanists, Agronomists, Archeologists, Linguists, and Ethnologists Meet, Rice 4 (2011) 178–183.
[31] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics (Oxford, England) 25 (2009) 1754–1760.
[32] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, S. Genome Project Data Processing, The Sequence Alignment/Map format and SAMtools, Bioinformatics (Oxford England) 25 (2009) 2078–2079.
[33] T.H. Lee, H. Guo, X. Wang, C. Kim, A.H. Paterson, SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data, BMC Genomics 15 (2014) 162.
[34] H. Ji, S.R. Kim, Y.H. Kim, H. Kim, M.Y. Eun, I.D. Jin, Y.S. Cha, D.W. Yun, B.O. Ahn, M.C. Lee, G.S. Lee, U.H. Yoon, J.S. Lee, Y.H. Lee, S.C. Suh, W. Jiang, J.I. Yang, P. Jin, S.R. McCouch, G. An, H.J. Koh, Inactivation of the CTD phosphatase-like gene OsCPL1 enhances the development of the abscission layer and seed shattering in rice, Plant J. Cell Mol. Biol. 61 (2010) 96–106.
[35] Z. Tian, Q. Qian, Q. Liu, M. Yan, X. Liu, C. Yan, G. Liu, Z. Gao, S. Tang, D. Zeng, Y. Wang, J. Yu, M. Gu, J. Li, Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities, Proc. National Acad. Sci. U. S A. 106 (2009) 21760–21765.
[36] C. Vitte, O. Panaud, H. Quesneville, LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss, BMC Genomics 8 (2007) 218.
[37] M.S. Lin, Genetic base of Japonica rice varieties released in Taiwan, Euphytica 56 (1991) 43–46.
[38] M.S. Lin, Field uniformity of the Japonica rice region of Taiwan as estimated by relative genetic contribution, TAG Theor. Appl. Genet. 83 (1991) 115–118.
[39] C.S. Tseng, M.T. Wu, H.C. Huang, Y.R. Lin, The Green fence of Chinese hibiscus (*Hibiscus rosa-sinensis* L.) preventing pollen dispersal of transgenic rice (*Oryza sativa*), Plant Prod. Sci. 15 (2012) 100–108.
[40] H. Saito, Q. Yuan, Y. Okumoto, K. Doi, A. Yoshimura, H. Inoue, M. Teraishi, T. Tsukiyama, T. Tanisaka, Multiple alleles at Early flowering 1 locus making variation in the basic vegetative growth period in rice (*Oryza sativa* L.), TAG Theor. Appl. Genet. 119 (2009) 315–323.