Development and application of *indica–japonica* SNP assays using the Fluidigm platform for rice genetic analysis and molecular breeding

Jeonghwan Seo · Gileung Lee · Zhuo Jin · Backki Kim · Joong Hyoun Chin · Hee-Jong Koh

Abstract Molecular markers are efficient and essential genotyping tools for molecular breeding and genetic analysis of rice. We developed two 96-plex *indica–japonica* single nucleotide polymorphism (SNP) genotyping sets for genetic analysis and molecular breeding in rice using the Fluidigm platform. Informative SNPs between *indica* and *japonica* were selected from SNP data of the Rice Diversity database, HapRice world SNP data of the Q-TARO database, and our 40 rice cultivar resequencing dataset. SNPs in set 1 were evenly distributed across all 12 rice chromosomes at a spacing of 4–5 Mb between adjacent SNPs. SNPs in set 2 mapped to the long genetic intervals in set 1 and included 14 functional or linked SNPs in genes previously cloned and associated with agronomic traits. Additionally, we used the SNP sets developed in this study to perform genetic diversity analysis of various cultivated and wild rice accessions, construction and validation of a subspecies diagnostic subset, linkage map construction and quantitative trait locus (QTL) analysis of a *japonica* × *indica* F_2 population, and background profiling during marker-assisted backcrossing. Furthermore, we identified subspecies-specific SNPs and discuss their distribution and association with agronomic traits and subspecies differentiation. Our results indicate that these subspecies-specific SNPs were present in wild rice prior to domestication. This genotyping system will serve as an efficient and quick tool for genetic analysis and molecular breeding in rice.

Keywords Rice · *Indica* · *Japonica* · SNP · Fluidigm · Subspecies specificity

J. Seo · G. Lee · Z. Jin · B. Kim · H.-J. Koh (✉)
Department of Plant Science, Research Institute for Agriculture and Life Sciences and Plant Genomics and Breeding Institute, Seoul National University, Seoul 08826, South Korea
e-mail: heejkoh@snu.ac.kr

J. H. Chin
Department of Integrative Bio-Industrial Engineering, Sejong University, Seoul 05006, South Korea

Introduction

Rice (*Oryza sativa* L.) consists of two main subspecies, *indica* and *japonica*, which can be discriminated based on the sterility of F_1 hybrids developed from a cross between them. *Indica* rice is adapted to and mostly grown in tropical regions and constitutes most of the global rice production. By contrast, *japonica* rice is intensively cultivated and consumed in temperate regions, such as Korea, Japan, China, and Taiwan, and forms approximately 20% of the global rice production. The domestication history of rice had been a controversial issue. Some studies using whole-genome sequencing technology suggest that rice was domesticated at least twice from two geographically and ecologically discrete subpopulations of *Oryza rufipogon*. Differences in the climate of these geographically discrete locations and in selection for grain-related traits by farmers have

led to morphological and physiological differentiation between *indica* and *japonica* rice (Civán et al. 2016; Civáň et al. 2015; Huang and Han 2015; Khush 1997; Oka 1988). Recently, Choi et al. (2017) revealed Asian rice has multiple origins but only single domestication in China.

In rice genetic improvement programs, hybridization between *indica* and *japonica* has been attempted by breeders for at least 50 years to genetically enhance the yield, grain quality, stress tolerance, and regional adaptation traits through recombination and hybrid breeding. *Indica* and *japonica* rice exhibit unique eco-logical adaptations and agronomically useful traits. However, reproductive barriers between the two sub-species such as hybrid sterility, hybrid breakdown, seg-regation distortion, and restricted recombination hinder cross-breeding attempts between the two subspecies. The exact classification of rice germplasm into these two subspecies could provide essential information for the selection of parents for use in the intersubspecific breeding programs in rice. Morphological and physio-logical characteristics such as grain shape, apiculus hair length, phenol reaction, and potassium chlorate ($KCIO_3$) resistance have been used to classify rice ac-cessions according to these subspecies (Oka 1988).

DNA-based markers are efficient and essential genotyping tools for molecular breeding and genetic analysis of rice since the 1980s. The first DNA marker developed for genetic analysis was restriction fragment length polymorphism (RFLP), based on DNA digestion with restriction endonucleases, followed by DNA hy-bridization with a labeled probe (Lander and Botstein 1989; McCouch et al. 1988; Tanksley et al. 1989). After the development of polymerase chain reaction (PCR), markers such as random amplified polymorphic DNA (RAPD), utilizing arbitrary primers (Kubelik et al. 1990), and amplified fragment length polymorphism (AFLP), based on restriction enzyme digestion and adapter-specific primer (Friters et al. 1995), were used for genotyping. With the development of genome se-quencing technologies, simple sequence repeats (SSRs) (Wu and Tanksley 1993), sequence-tagged sites (STSs) (Inoue et al. 1994), and single nucleotide polymor-phisms (SNPs) (Feltus et al. 2004) have been exploited as DNA markers based on sequence information.

SNPs are more abundant, stable, easy to automate, efficient, and cost-effective markers for genotyping than other DNA markers. Advancements in sequencing tech-nology have led to the development of new genotyping

platforms and the accumulation of diverse sequence data. Moreover, genome-wide SNP assays have taken the place of conventional gel-based PCR markers, such as SSRs and STSs, in rice breeding programs (McCouch et al. 2010). Until recently, genome-wide SNP assays based on several genotyping platforms including Illumina BeadXpress (Chen et al. 2011; Thomson et al. 2011; Yamamoto et al. 2010), Illumina Infinium (Chen et al. 2014; Thomson et al. 2017; Yu et al. 2014), Affymetrix (McCouch et al. 2016; Singh et al. 2015; Zhao et al. 2011a), and the KASP marker system (Cheon et al. 2018; Pariasca-Tanaka et al. 2015; Steele et al. 2018; Yang et al. 2019) have been developed and applied to rice genetic research and molecular breeding.

Previously, we reported subspecies-specific RAPD and STS markers (Chin et al. 2007; Chin et al. 2003), and we have used these markers and SSR markers for genetic diversity analysis, quantitative trait locus (QTL) analysis, mutant gene mapping, and molecular breeding. However, these conventional gel-based PCR markers are laborious, time-consuming, and expensive, as they require a large amount of polymerase and PCR reaction buffer. The Fluidigm SNP genotyping platform has automated PCR and the integrated fluidic circuit (IFC) technology, which automatically mixes PCR reagents through the microfluidic channel networks. In this study, we developed two genome-wide 96-plex Fluidigm SNP sets that distinguish between *indica* and *japonica* rice, based on publicly available rice SNP data and our resequencing data. We used these SNP assays to assess the genetic diversity of rice germplasms, QTL mapping of a population derived from *indica* × *japonica* cross, and background profiling during marker-assisted backcrossing. Furthermore, we propose a subset of di-agnostic SNPs for the classification of rice germplasms according to the two subspecies and discuss the distri-bution of subspecies-specific nucleotide variations and their relationship with agronomic traits and subspecies differentiation.

Materials and methods

Plant materials

A panel of 83 rice germplasms collected from around the world and maintained at the CMB laboratory, Seoul National University, was used for SNP verification and genetic diversity analysis in this study (Table S1). These

rice germplasms were selected based on diverse origin and subspecies according to literatures or previous studies. The F_2 population, consisting of 190 individuals, derived from a cross between Dobong (temperate *japonica*) and TN1 (*indica*) was used for linkage map construction, segregation distortion analysis, and QTL mapping of the days to heading trait. All plants were grown using conventional practices in a paddy field at the experimental farm of Seoul National University, Suwon, Korea (37° 16′ 10.2″ N, 126° 59′ 20.3″ E).

SNP selection and Fluidigm genotype assay design

SNPs were collected from the 44K SNP set (Zhao et al. 2011a) of the Rice Diversity database (http://www.ricediversity.org/data/), the HapRice world SNP data (Yonemaru et al. 2014) of the Q-TARO database (http://qtaro.abr.affrc.go.jp/), and the resequencing data of 40 rice cultivars from CMB lab. Seoul National University generated using Illumina platform based on methods described by Kim et al. (2014). SNPs showing only substitution polymorphism between *indica* and *japonica* varieties and no other SNP in at least 60 bp on either side were selected. To design the Fluidigm SNP genotyping assays, 60–150 bp sequences flanking the selected SNPs on either side were downloaded in the reference genome, and checked for the specificity of the regions using BLAST. Finally, the selected SNPs and flanking sequences were uploaded on the D3 Assay Design (https://d3.fluidigm.com/) website. After confirming the results, the designed assays were ordered. One Fluidigm SNP assay contains allele-specific primer 1 (ASP1), ASP2, locus-specific primer (LSP), and specific target amplification (STA) primer. Information on SNP sets 1 and 2 is summarized in Table S2 and Table S3, respectively.

DNA extraction

Young leaves were sampled from each plant at the tillering stage. Genomic DNA was extracted using the modified cetyltrimethylammonium bromide (CTAB) method, as described by Murray and Thompson (1980). The concentration and purity of DNA samples were measured with a NanoDrop 1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). DNA samples with absorbance ratios at 260 and 280 nm (A260/A280) > 1.8 were diluted to a concentration of 50 ng/μl and used for genotyping.

Fluidigm genotyping

Fluidigm genotyping was performed using the BioMark™ HD system (Fluidigm, San Francisco, CA, USA) and 96.96 Dynamic Array IFCs (Fluidigm), according to the manufacturer's instructions, at the National Instrumentation Center for Environmental Management (NICEM), Seoul National University (Pyeongchang, Korea). Specific target amplification (STA) was performed prior to SNP genotyping analysis to increase the probability of success of the SNP type assay. First, a 10× STA primer pool was prepared comprising mixture of 2 μl of STA primer for each of the 96 markers, 2 μl of LSP for each of the 96 markers, and 16 μl of DNA suspension buffer. For each of the 96 samples, STA was conducted in a total volume of 5 μl per reaction, which contained 2.5 μl of 2× multiplex PCR master mix (Qiagen, Hilden, Germany), 0.5 μl of the 10X STA primer pool, 0.75 μl of PCR-certified water, and 1.25 μl of genomic DNA with the following thermal cycling conditions: 15 min at 95 °C, followed by 14 cycles of a 2-step amplification including 15 s at 95 °C and 2 min at 60 °C. Then, STA products were diluted 1: 50 in DNA suspension buffer and used for the SNP type assay.

To perform the SNP type assays using the 96.96 IFC, the assay mix and sample mix were prepared. The assay mix contained 1.5 μl of PCR-certified water, 2.5 μl of 2X assay loading reagent, and 1.0 μl of the SNP type assay mix, which was taken from the mixture of 3 μl of each ASP, 8 μl of each LSP, and 29 μl of DNA suspension buffer. The sample pre-mix contained 360 μl of 2× Fast Probe Master Mix (Biotium, Fremont, CA, USA), 36 μl of SNP type 20× sample loading reagent, 12 μl of SNP type 60× reagent, 4.3 μl of 50× ROX dye (Invitrogen, Waltham, MA, USA), and 7.7 μl of PCR-certified water. Subsequently, the sample mix was prepared by mixing 2.5 μl of each STA product and 3.5 μl of the sample pre-mix in each well of 96-well plates. Finally, 5 μl of each sample mix and 4 μl of each assay mix were loaded into 96 sample inlets and 96 assay inlets of the 96.96 IFC, respectively. The SNP type assays were performed in series using two machines, the IFC controller HX (Fluidigm) and the BioMark™ HD system (Fluidigm), according to the manufacturer's instructions, using the following thermal cycling conditions: 95 °C for 15 s, 64 °C for 45 s, and 72 °C for 15 s, with a touchdown of − 1 °C per cycle from 64 to 61 °C, followed by 34 cycles of 95 °C for 15 s, 60 °C for 45 s,

and 72 °C for 15 s. The genotyping results were acquired using the Fluidigm SNP Genotyping Analysis software. All base calls were manually verified, and any errors in homozygous or heterozygous clusters were curated.

SS score calculation

The concept of subspecies specificity (SS) score was described by Chin et al. (2007) as below.

SS score of each SNP

$$= \frac{\text{(Total number of expected alleles in each subspecies)}}{\text{(Total number of varieties tested)}} \times 100 \ (\%)$$

For example, if a SNP has a SS score 100% in 50 varieties tested, it means the SNP showed all expected alleles to all of 50 varieties without any exception. In addition, a SNP with SS score equal to 98 means 49 varieties showed expected alleles and one variety showed unexpected allele.

Genetic diversity analysis

A model-based Bayesian approach implemented in the software package STRUCTURE 2.3.4 (Pritcharda et al. 2009) was used to estimate the number of subpopulations ($K = 2$ to $K = 10$) among 83 accessions with an admixture model and correlated allele frequencies. Simulations were run with 100,000 burn-in period, 100,000 Markov Chain Monte Carlo (MCMC) repeats, and ten independent iterations. The optimal number of subpopulations corresponds to the highest peak in the Delta K graph (Evanno et al. 2005), and germplasms with > 60% membership probabilities were assigned to each subpopulation. Bar plots of STRUCTURE analysis were modified using the STRUCTURE Plot V2.0 (http://omicsspeaks.com/strplot2/) (Ramasamy et al. (2014). Analysis of basic marker statistics of 190 SNPs including MAF, heterozygosity, and PIC was performed using PowerMarker V3.25 (Liu and Muse 2005). PowerMarker V3.25 was also used to calculate the genetic distance based on CS Chord (Cavalli-Sforza and Edwards 1967) and to construct an un-weighted pair group method with an arithmetic mean algorithm (UPGMA) and a neighbor-joining (NJ) dendrogram, which were visualized in Molecular Evolutionary Genetics Analysis version 7.0 (MEGA7; (Kumar et al. 2016).

Linkage map construction, QTL analysis, and segregation distortion analysis

Linkage map construction and segregation distortion and QTL analyses were conducted using the QTL IciMapping 4.1 software (Meng et al. 2015). First, the BIN functionality (binning of redundant markers) was used to remove redundant markers. Markers with more than 10% missing data were also removed. The output file generated in the binning step was used for linkage map construction with the MAP functionality. The Kosambi mapping function was used to calculate genetic distances in centimorgans (cM) (Kosambi 1944). QTL mapping was conducted using the BIP functionality (QTL mapping in biparental populations). The inclusive composite interval mapping of additive (ICIM-ADD) QTL method was used to detect additive QTLs, and a walk speed of 1.0 cM and stepwise regression probability of 0.001 were chosen for QTL detection. Significant logarithm of the odds (LOD) threshold value was calculated for each QTL using 1000 permutations at $P = 0.05$. A LOD score of 3.78 was used to determine significant additive QTLs. The inclusive composite interval mapping of digenic epistatic (ICIM-EPI) QTL method was used to identified digenic epistatic QTLs, and a walk speed of 1.0 cM and stepwise regression probability of 0.001 were chosen for QTL detection. A significant epistatic LOD threshold value of 5 was set manually. Segregation distortion analysis was carried out using the SDL functionality (segregation distortion locus mapping), using default settings.

Results

Development of two 96-plex *indica–japonica* SNP sets

We developed two sets of 96-plex *indica–japonica* SNP genotyping assays covering the whole rice genome using the Fluidigm platform. In the first set (set 1), we selected 96 nucleotides polymorphic between *indica* varieties and temperate *japonica* from the 44K SNP data (95) and our resequencing data (1) (Table S2). SNPs in set 1 were mapped to all 12 rice chromosomes, with 4–5 Mb physical distances between adjacent SNPs. Then, we selected additional 82 SNPs from the HapRice world SNP data (30) and our resequencing data (52), comprising set 2, to fill the long genetic intervals between adjacent SNPs in set 1 and the monomorphic region between Dobong (temperate *japonica*) and TN1

(*indica*) on chromosomes 2, 6, and 11 (Table S3). Set 2 included 14 functional or linked SNPs in cloned genes accounting for natural variation in agronomic traits, including hybrid sterility, plant architecture and heading date, and grain traits such as grain color, shape, shattering, and quality. These 14 genes were selected for rapid screening and profiling of interest genes account for agronomic traits. All 192 SNPs from sets 1 and 2 were evenly distributed in the rice genome, and the number of SNPs on each chromosome varied from ten (chromosome 10) to 23 (chromosome 1). The average physical and genetic intervals between adjacent SNPs were 2.1 Mb and 8.4 cM, respectively (Fig. 1; Table 1; Table S4). These 192 SNPs include44d 95 SNPs from the 44K SNP data, 53 SNPs from our resequencing data (Crop Molecular Breeding [CMB] laboratory), 30 SNPs from HapRice data, and 14 SNPs from cloned natural variation associated genes (Fig. S1a). Among these 192 SNPs, 131 were located in intergenic regions, 27 were non-synonymous, 14 were located in 3′ untranslated regions (3′-UTRs), 12 were synonymous, and eight were located in introns (Fig. S1b).

## SNP verification and genetic diversity analysis

Previously, our genetic diversity analysis of rice varieties adapted to the environmental conditions in Africa using *indica–japonica* SNP set 1 revealed that CG14 (*Oryza glaberrima*) is genetically closely related to *japonica*, whereas NERICA (New Rice for Africa) varieties, developed from a cross between *O. glaberrima* and *O. sativa*, are closely related to *indica* (Megersa et al. 2016). After the development of *indica–japonica* SNP set 2, we genotyped 83 diverse germplasms using both SNP sets. Among the 192 SNPs, 190 were polymorphic among these 83 accessions (Fig. S2); however, of the two remaining SNPs (both in set 2), cmb0232.7 was monomorphic, and cmb0909.6 showed low base call quality. Therefore, these two SNPs were excluded from further analysis. The list of 190 SNPs and their diversity information, such as major allele frequency (MAF), gene diversity, heterozygosity, polymorphic index content (PIC), and SS score, are listed in Table S5.

Next, we conducted population structure and phylogenetic analyses of all 83 germplasms, based on the genotyping results of 190 SNPs. The population structure was inferred using the STRUCTURE software and the values of K were set from 2 to 10. The value of Delta K was the highest at $K = 2$ (Fig. S3). At $K = 2$, 83 germplasms were classified into *indica* and *japonica*, but some germplasms appeared admixed. At $K = 3$ and $K = 4$, new subgroups emerged within the admixed group and *indica* group, respectively. At $K = 5$, five subgroups could be identified including *indica*, *aus*, *aromatic*, tropical *japonica*, and temperate *japonica*, although six germplasms showed less than 60% of estimated ancestry derived from any single subgroup. At $K = 6$ and $K = 7$, the *indica–temperate japonica* admixture type (Nongan) and *aromatic–tropical japonica* admixture type (Adday Sel) germplasms generated independent subgroups (Fig. 2a). The result of phylogenetic analysis verified the same grouping as the population structure analysis, and the phylogenetic trees showed genetic relationships among the five subgroups identified in population structure analysis (Fig. 2b; Fig. S2). Among these five subgroups, *indica* and temperate *japonica* showed the highest genetic distance. *Aus* and tropical *japonica* showed a close relationship with *indica* and temperate *japonica*. *Aromatic* emerged as the last subgroup at $K = 5$ in structure analysis and clustered in the middle of the other groups in phylogenetic analysis.

Subspecies diagnostic subset

We estimated the subspecies specificity of 190 SNPs using the SS score. To calculate the SS score, we selected 25 high ranking accessions each from the *indica* and temperate *japonica* subgroups, with more than 90% estimated ancestry of each subgroup at $K = 5$ in structure analysis. The average SS score of 190 SNPs was 92.9 (range, 50–100; Table S5). Then, we identified a subspecies diagnostic subset of 21 SNPs with SS score ≥ 98, based on 50 germplasms, and subspecies-specific allele frequency > 95%, based on 2759 *indica* accessions and 1512 *japonica* accessions in the RiceVarMap v2.0 database (http://ricevarmap.ncpgr.cn/v2/; Zhao et al. (2015)). These 21 SNPs were distributed across all 12 rice chromosomes, except chromosome 6, at a frequency of two SNPs per chromosome, except on chromosome 4 (one) (Table 2). We verified the subspecies diagnostic subset on 4729 whole accessions using the Haplotype Network Analysis tool in RiceVarMap v.2.0. The results showed these 21 SNPs accurately classified 4061 (85.6%) out of 4729 accessions into the *indica* and *japonica* groups. Based on this result, we selected a core of six subspecies diagnostic SNPs conserved in each group and re-evaluated them in the same method, as described above. Haplotype network analysis showed that these six core SNPs classified 4553
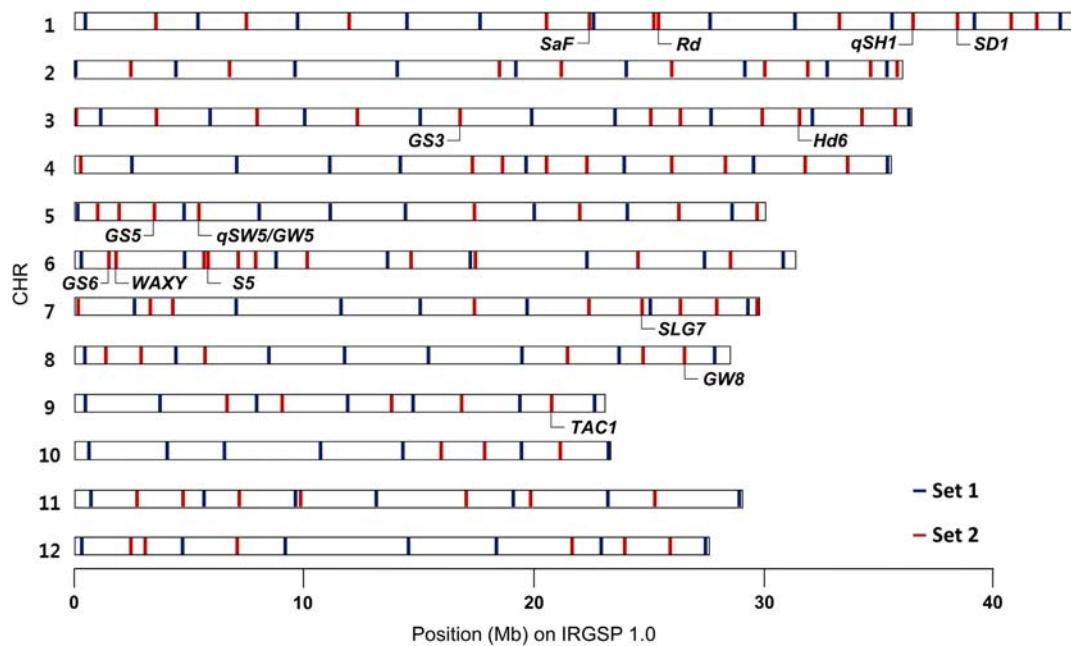
Fig. 1 Physical positions of 192 *indica–japonica* Fluidigm SNPs on the rice genome. Fourteen cloned natural variation genes containing functional or linked SNPs related to agronomic traits are indicated

(96.3%) out of 4729 accessions into the *indica* and *japonica* groups (Fig. S4).

Genetic diversity of wild rice

Genetic diversity analysis of 446 *O. rufipogon* accessions in the Oryza Genome Release 2.0 database (http://viewer.shigen.info/oryzagenome2/) was carried out using subspecies-specific SNPs identified in this study. We downloaded the genotype of 87 SNPs with SS score ≥ 98, based on 50 germplasms, of which 55 SNPs with less than 10% missing data were used for diversity analysis. Of the 55 SNPs, 36 and 19 represented the *indica* and *japonica* major alleles, respectively, in 446 accessions (Table S6). Previously, 446 *O. rufipogon* accessions were classified into three types, Or-I, Or-II, and Or-III, using ~ 5 million SNPs (Huang et al. 2012). However, the neighbor-joining tree based on 55 subspecies-specific SNPs showed four groups, including the *indica*-like group, *japonica*-like group, and two intermediate groups. The *indica*-like groups and two intermediate groups showed more *indica* alleles than *japonica* alleles, whereas in the *japonica*-like group, *japonica* alleles were more abundant. The *japonica*-like group consisted of mainly Or-III type accessions, which originated from southern China and were distributed within a narrow range of relatively higher latitudes. Most of the Or-I type and some of the Or-II type

accessions originated from India and Southeast Asia and were clustered to the *indica*-like group or intermediate I group. The intermediate II group consisted of some Or-II and Or-III accessions from most of the diverse region and a broad range of latitudes (Fig. 3). This suggests that subspecies-specific SNPs existed in *O. rufipogon* before the emergence of *indica* and *japonica* groups during domestication.

Linkage map construction for QTL and segregation distortion analysis

Next, we performed QTL analysis (days to heading trait) and segregation distortion analysis of an F_2 population derived from a cross between Dobong (*temperate japonica*) and TN1 (*indica*). To conduct the QTL analysis, we first genotyped the F_2 population using *indica–japonica* SNP sets 1 and 2 and then constructed a linkage map of the F_2 population using 172 polymorphic SNPs covering all 12 chromosomes. The linkage map spanned a genetic distance of 1791 cM, with an average genetic distance 11.2 cM between adjacent markers, which was sufficient to detect QTLs and analyze segregation distortion (Fig. 4a). Six additive QTLs for days to heading were detected on chromosomes 3, 7, 8, 11, and 12. Dobong alleles on four QTLs (*qDTH3.1*, *qDTH3.2*, *qDTH7*, *qDTH12*) shorten the days to

Table 1 List of functional or linked SNPs in previously cloned genes associated with natural variation

SNP ID	Chr.	Position (bp)	Gene name	Gene ID	SNP location/effect	Gene function
SaF-CT	1	22,376,434	*SaF*	Os01g0578500	Non-synonymous/amino acid change	F-box protein, *indica–japonica* hybrid male sterility
Rd-GA	1	25,383,230	*Rd*	Os01g0633500	Non-synonymous/amino acid change	Similar to dihydroflavonol reductase
qSH1-TG	1	36,461,792	*qSH1*	Os01g0848400	Intergenic/linked	BEL1-type homeobox family, seed shattering (Os01t0848400-01)
SD1-GA	1	38,384,401	*SD1*	Os01g0883800	Intron	Similar to GA C20oxidase2 (Os01t0883800-01); GA 20-oxidase2, GA metabolism (Os01t0883800-02)
GS3-CA	3	16,733,441	*GS3*	Os03g0407400	Non-synonymous/stop gain	Similar to grain length and weight protein
Hd6-AT	3	31,512,460	*Hd6*	Os03g0762000	Non-synonymous/stop lost	Similar to protein kinase CK2, alpha subunit (Os03t0762000-01); casein kinase II alpha subunit, heading date, photoperiod sensitivity (Os03t0762000-02)
GS5-GA	5	3,444,069	*GS5*	Os05g0158500	5′ upstream	Peptidase S10, serine carboxypeptidase family protein
qSW5-AG	5	5,361,396	*qSW5/GW5*	Os05g0187500	5′ upstream	IQ calmodulin-binding motif family protein, positive regulator of brassinosteroid signaling, regulation of grain width and weight
GS6-TG	6	1,465,866	*GS6*	Os06g0127800	3′-UTR	GAI-RGA-SCR (GRAS) family protein, brassinosteroid signaling
WAXY-TG	6	1,765,761	*WAXY*	Os06g0133000	Splicing site	granule-bound starch synthase I, chloroplast precursor (EC 2.4.1.21) (Os06t0133000-01); granule-bound starch synthase 1, starch biosynthesis (Os06t0133000-02)
S5-TC	6	5,761,511	*S5*	Os06g0213100	Non-synonymous/amino acid change	Eukaryotic aspartyl protease domain containing protein
SLG7-GC	7	24,666,135	*GL7/GW7/SLG7*	Os07g0603300	Non-synonymous/amino acid change	TON1 RECRUIT MOTIF (TRM)-containing protein, regulation of grain size and shape (Os07t0603300-01)
GW8-AG	8	26,502,275	*OsSPL16/GW8*	Os08g0531600	Intron	Transcription factor, positive regulator of cell proliferation, control of grain size, shape, and quality (Os08t0531600-01)
TAC1-CT	9	20,731,844	*TAC1*	Os09g0529300	3′-UTR	Similar to leaf angle-associated protein (Os09t0529300-01); regulation of tiller angle (Os09t0529300-02)

heading. On the other hand, Dobong alleles on *qDTH8* and *qDTH11* delay the days to heading. Four of these six additive QTLs, such as *qDTH3.1* (25.06%), *qDTH3.2* (9.62%), *qDTH7* (40%), and *qDTH8* (29.28%), representing more than 5% of the phenotypic variation explanation (PVE), contained genes that control heading date, including *DTH3*, *Hd6*, *Ghd7*, and *Ghd8/DTH8*, respectively.

The *qDTH3.2* QTL colocated with the Hd6-AT SNP marker developed on the basis of natural variation available for *Hd6*. Two minor QTLs, *qDTH11* and *qDTH12*, were newly identified in this study (Fig. 4a; Table S7). Additionally, we identified 14 digenic epistatic QTLs. Some regions smaller than 20 cM on chromosomes 6, 7, 8, and 11 showed more than two interactions with other genomic regions (Fig. 4b;
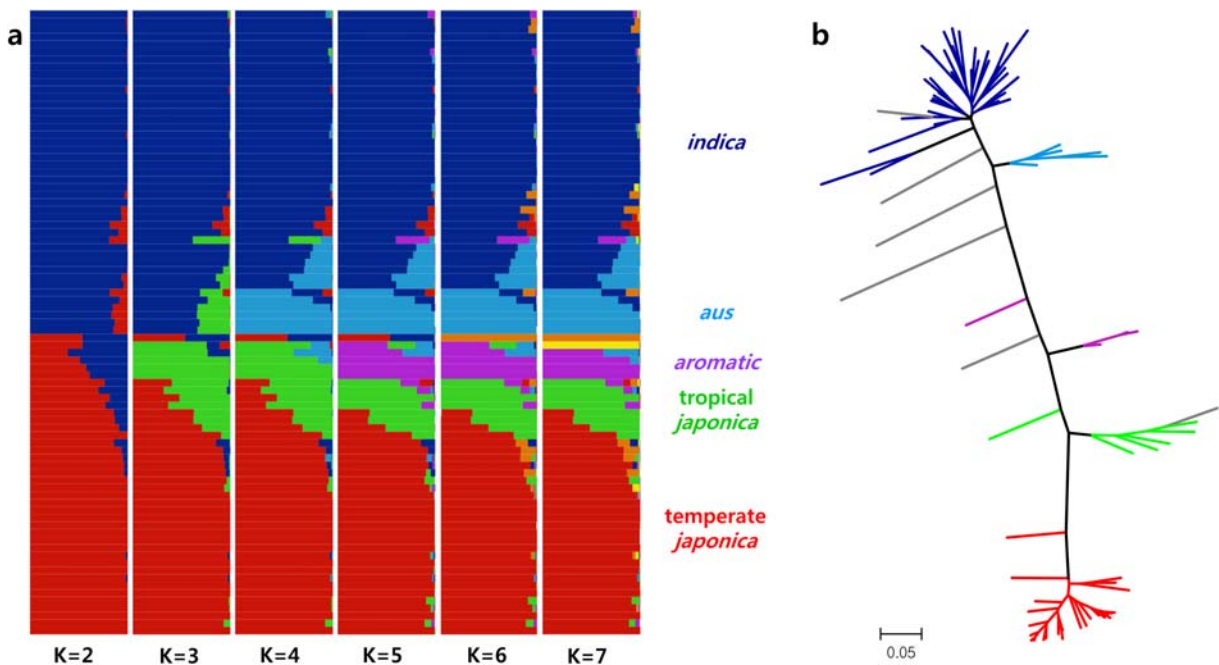
Fig. 2 Genetic diversity analysis of 83 germplasms using 190 *indica–japonica* SNPs. **a** Population structure of rice accessions based on the STRUCTURE output for $K = 2$ to $K = 7$. **b** Neighbor-joining (NJ) tree. Branches of the tree are colored according to the subpopulation assessment in **a** when $K = 5$, Gray branches indicate admixture

Table S8). These QTLs explained transgressive segregation in the F_2 population (Fig. 4c) and could be used to manipulate the heading date in intersubspecific breeding.

Furthermore, 24 markers and eight intervals on chromosomes 3, 4, 5, 6, 8, 10, and 12 showed significant segregation distortion at LOD threshold 2.5 (Table S9; Fig. S5). Four, two, and two intervals were distorted to homozygous TN1, homozygous Dobong, and heterozygous alleles, respectively. All intervals were located near the previously reported loci. For example, two loci, *S9* (Wan et al. 1996) and *hsa2* (Kubo and Yoshimura 2005), distorted to the homozygous Dobong allele have been reported previously on chromosomes 4 and 8, respectively. In the genomic region showing the strongest segregation distortion to the homozygous TN1 allele on chromosome 3, the *Sc* locus, which controls hybrid male sterility and segregation distortion, was cloned recently (Shen et al. 2017). In addition, the *has1* locus, which controls hybrid sterility in F_1 and F_2 populations, was located near the region showing segregation distortion toward the homozygous TN1 allele on chromosome 12; this locus was also cloned (Kubo et al. 2016). The interval distorted to the heterozygous allele on upper side of chromosome 6 was located near the previous reported region (Shanmugavadivel et al. 2013).

Furthermore, intervals showing segregation distortion on chromosomes 4, 10, and 12 were overlapped with loci recently reported by Li et al. (2019). However, no segregation distortion was detected within the low polymorphic regions on chromosomes 2, 6, and 11.

Background profiling during molecular breeding

The introgression of specific genes from the donor germplasm to elite varieties using recurrent backcrossing along with marker-assisted selection is an efficient breeding strategy. Hopum, a high-quality Korean temperate *japonica* variety, showed low spikelet fertility in a cross with *indica* varieties. We have been developing widely compatible Hopum for hybrid rice breeding and as a bridging line in intersubspecific hybridization using the widely compatible *aus* variety Dular as a donor parent. Dular contains neutral hybrid sterility genes, such as *S24* (Zhao et al. 2011b) for pollen sterility and *S5* (Chen et al. 2008; Yang et al. 2012) for embryo sac sterility, in F_1 hybrid of the intersubspecific cross. Marker-assisted selection of the backcross populations (BC_1F_1 to BC_3F_1) has been previously carried out using gene-specific insertion-deletion (InDel) markers of *S24* and *S5*. Two plants were selected from

Table 2 List of 21 subspecies diagnostic SNPs, including six core SNPs, validated using the RiceVarMap database

Chr.	SNP ID[a]	Physical position (bp)[b]	Allele	Reference allele[c]	All *indica* (*n* = 2759)	All *japonica* (*n* = 1512)	Variation ID in RiceVarMap
1	id1004256	5,333,883	AG	G	A (98.9%)	G (98.0%)	vg0105333883
1[b]	*id1022407*	*35,542,178*	*AG*	*G*	*A (99.7%)*	*G (98.3%)*	*vg0135542178*
2	*id2004617*	*9,553,166*	*AC*	*A*	*C (98.8%)*	*A (97.4%)*	*vg0209553166*
2	ah02001499	21,141,852	TG	T	G (98.9%)	T (98.0%)	vg0221141852
3	id3005168	9,975,083	AG	A	G (98.7%)	A (98.4%)	vg0309975083
3	*id3010700*	*23,494,214*	*AG*	*G*	*A (98.7%)*	*G (98.1%)*	*vg0323494214*
4	cmb0420.7	20,508,025	GA	G	A (97.3%)	G (98.9%)	vg0420508025
5	id5000043	101,950	CA	C	A (98.0%)	C (96.2%)	vg0500101950
5	*id5008218*	*19,961,728*	*TG*	*G*	*T (99.2%)*	*G (96.8%)*	*vg0519961728*
7	id7004645	25,022,403	GA	G	A (96.4%)	G (99.7%)	vg0725022403
7	cmb0727.0	26,343,254	TC	T	C (97.3%)	T (95.8%)	vg0726343254
8	id8000140	417,700	AC	A	C (97.6%)	A (97.9%)	vg0800417700
8	id8006751	23,652,821	AG	A	G (98.5%)	A (98.2%)	vg0823652821
9	id9006953	19,338,953	AC	A	C (97.6%)	A (99.3%)	vg0919338953
9	id9007784	22,600,843	AG	A	G (98%)	A (99.3%)	vg0922600843
10	*id10003706*	*14,257,303*	*GA*	*A*	*G (98.7%)*	*A (99.3%)*	*vg1014257303*
10	*ah10001182*	*21,101,395*	*CA*	*C*	*A (98.6%)*	*C (98.2%)*	*vg1021101395*
11	wd11000649	9,570,096	AG	A	G (97.4)	A (95.0%)	vg1109570096
11	id11004341	13,098,539	GA	G	A (95.1%)	G (95.1%)	vg1113098539
12	cmb1226.0	25,884,498	CT	C	T (95.3%)	C (99.2%)	vg1225884498
12	id12010130	27,434,178	AC	C	A (98.5%)	C (99.2%)	vg1227434178

[a] Physical position of SNPs was determined from the IRGSP1.0

[b] SNPs in italics represent six core subspecies diagnostic SNPs

[c] Reference allele refers to the Nipponbare allele in the reference genome

the BC_3F_1 population (BC_3F_1-12 and BC_3F_1-17), based on the heterozygous genotype of target genes and the plant phenotype. Whole-genome background of these two BC_3F_1 plants was profiled using 139 polymorphic markers from the *indica–japonica* SNP sets. The results showed that BC_3F_1-12 and BC_3F_1-17 represented 91.7% and 87.4% similarity, respectively, to the genotype of the recurrent parent (Fig. 5). We are tracking the target genes and eliminating undesirable introgressed fragments in advanced progenies based on this result.

Distribution of subspecies-specific markers

Subspecies-specific markers, including 36 SSRs (Neeraja et al. (2006), 67 STSs (Chin et al. (2007), 57 intron length polymorphisms (ILPs) (Zhao et al. (2009), and 34 InDels (Lu et al. (2009), have been reported previously. In this study, 87 SNPs with SS score ≥ 98, based on 50 germplasms, were selected as subspecies-

specific SNPs. In addition, we selected 85 temperate *japonica* and 85 *indica* accessions from the 44K SNP dataset and calculated the SS score of all SNPs. Among the 36,901 SNPs, 3179 SNPs with an SS score of ≥ 98 were detected as subspecies-specific SNPs. All subspecies-specific markers and common gaps were plotted on 12 chromosomes (Fig. 6). Subspecies-specific markers showed an uneven distribution without any positional correlation among markers. However, three common gaps larger than 5 Mb containing no subspecies-specific markers were detected on three chromosomes near or encompassing the centromere in low-recombination regions.

Discussion

In this study, we developed two sets of 96-plex *indica–japonica* SNP sets using the Fluidigm system and used
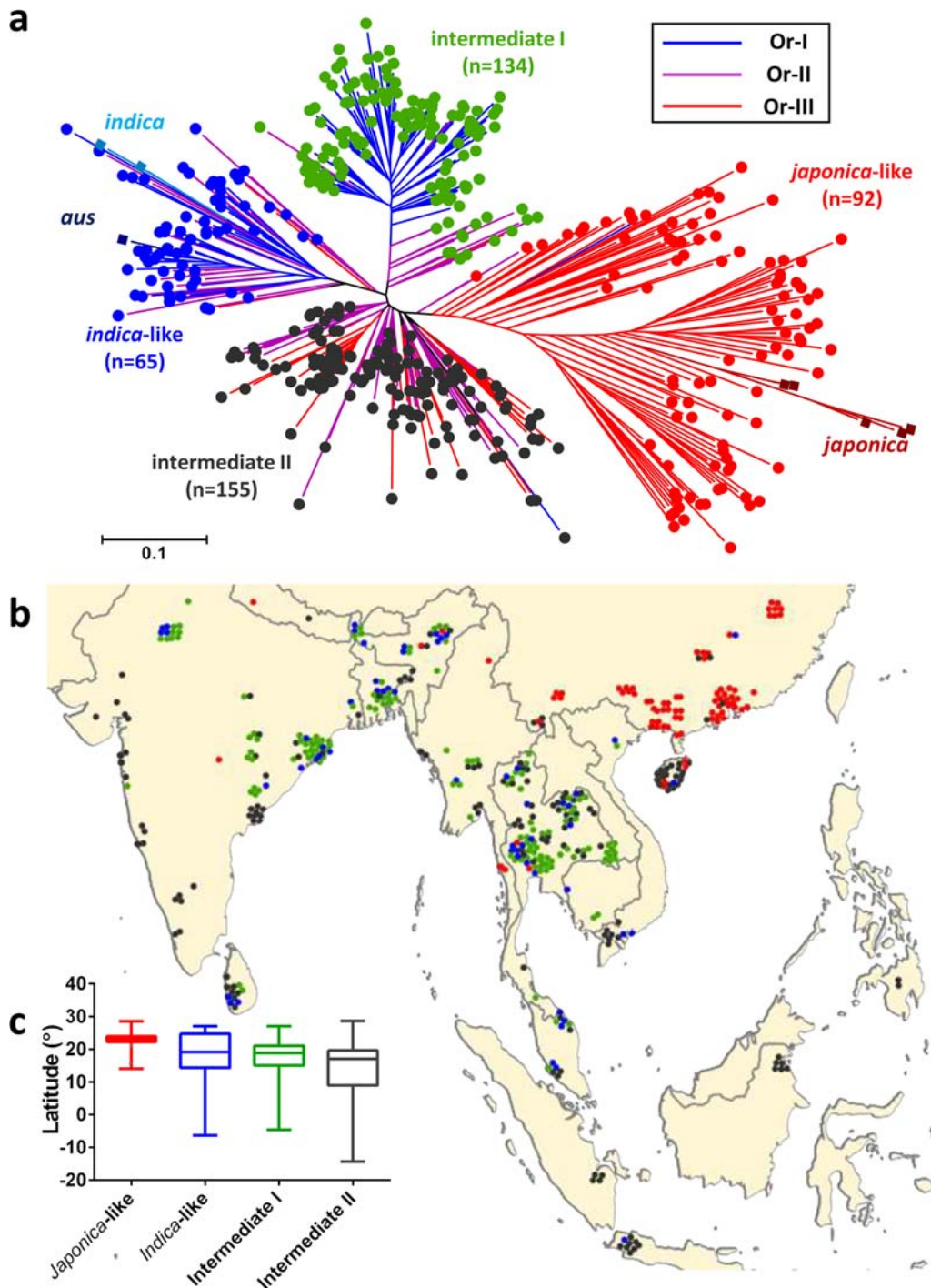
Fig. 3 Genetic diversity analysis of 446 *Oryza rufipogon* accessions based on 55 subspecies-specific SNPs. **a** Neighbor-joining tree of 446 *O. rufipogon* accessions and ten *O. sativa* varieties calculated using 55 SNPs with a subspecies specificity (SS) score ≥ 98. Three groups of *O. rufipogon* accessions defined by Huang et al. (2012) were identified, including Or-I, Or-II, and Or-III, indicated as blue, purple, and red branches, respectively. Dark red, dark blue, and light blue squares indicate *japonica* (including temperate *japonica*, tropical *japonica*, and *aromatic*), *indica*, and *aus* varieties, respectively. Four groups identified in this study, including *japonica*-like, *indica*-like, intermediate I, and intermediate II, are indicated using red, blue, green, and gray circles, respectively. **b** Geographical distribution of the four groups among 446 *O. rufipogon* accessions. **c** Box plot of the latitude distribution of four *O. rufipogon* groups identified in **a**
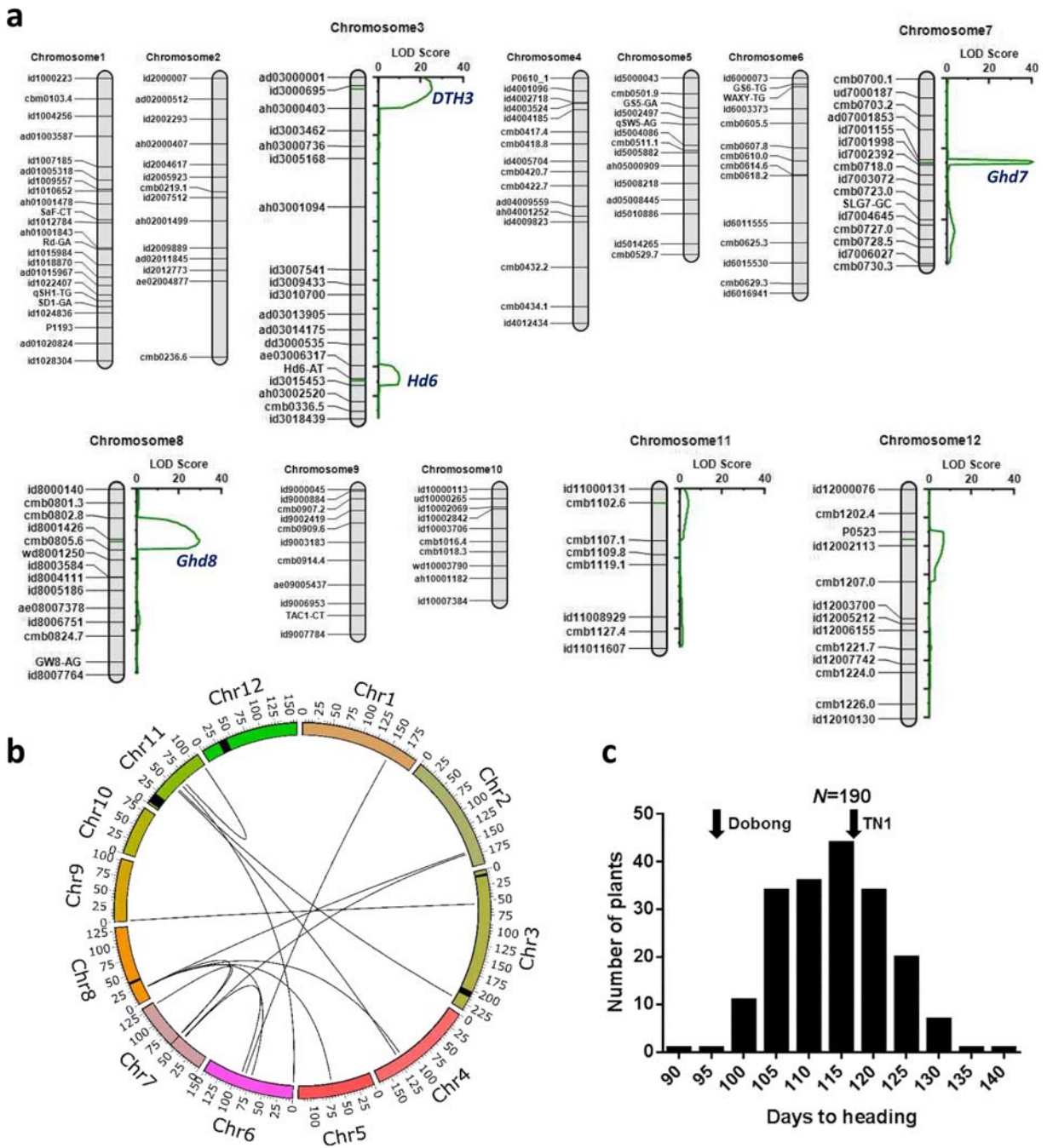
Fig. 4 Linkage map and QTL analysis of days to heading in the Dobong × TN1 F$_2$ population. **a** Linkage map of 12 chromosomes and LOD graph of additive QTLs. Among all chromosomes, only the chromosomes possessing additive QTLs were shown with the LOD graph, and candidate genes at each QTL are indicated. **b** Results of digenic epistatic QTL analysis. Black highlights on chromosomes indicate the confidence interval of each additive QTL. **c** Frequency distribution of heading date in the F$_2$ population

these to perform various rice genetic analyses. A total of 190 SNPs were used to classify 83 diverse germplasms into the *indica* and *japonica* groups, as these markers were polymorphic between the two subspecies (Fig. 2).

Nevertheless, we could identify five subgroups within these 83 accessions, including *indica*, *aus*, *aromatic*, *tropical japonica*, and *temperate japonica*, similar to a previous study (Garris et al. 2005). In addition, a
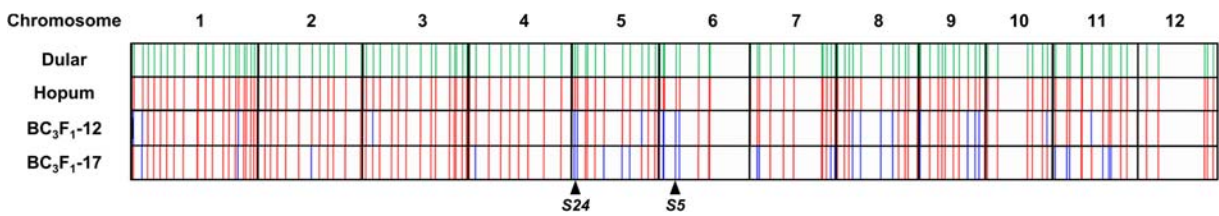
Fig. 5 Genome profiling of two individuals in the BC$_3$F$_1$ population derived from Hopum*4/Dular using *indica–japonica* SNP sets. Black triangles indicate the position of two target genes, *S24* and *S5*, on chromosomes 5 and 6, respectively. Blue bars indicate SNP markers with a heterozygous genotype at genomic segments introgressed from the donor parent Dular. Green and red bars indicate SNP markers homozygous for the Dular and Hopum alleles, respectively

subspecies diagnostic of 21 or six SNPs clearly distinguished between *indica* and *japonica* subspecies (Fig. S4). Based on the objective, either the whole SNP set or a subset can be used for genotyping. For instance, if the objective is to determine whether an unknown accession is an *indica* or *japonica* type, genotyping using the core subset of six SS-SNPs is appropriate; however, if detailed information about the subgroup is needed, 96 or 192 SNPs will be more suitable.

To dissect the genetic diversity of 446 *O. rufipogon* accessions, 55 SS-SNPs were selected, and four groups, such as *indica*-like group, *japonica*-like group, and two intermediate groups, were identified (Fig. 3). Most *japonica*-like accessions were distributed in a narrow range of higher latitudes in southern China, where the seasonal photoperiod is more variable. This implies that photoperiod sensitivity in domesticated *japonica* emerged in the *japonica*-like group before domestication. Additionally, hybrid sterility in the offspring, derived from a cross between *indica* and *japonica*, also existed before domestication. The SNP marker S5-TC, which represents a functional SNP in the second exon of the *S5* gene, causes embryo sac abortion in the *indica* × *japonica* hybrid (Chen et al. 2008). The frequency of the *indica* allele in the *indica*-like group and two intermediate groups was greater than 89%, whereas that of the *japonica* allele in the *japonica*-like group was 73.9%. This suggests that different environmental conditions and reproductive barriers led to the differentiation of *O. rufipogon* into two distinct subspecies before domestication.

A total of 289 out of 446 accessions (64.8%) were clustered in the two intermediate groups, I and II, based on 55 SS-SNPs. Of these, 12 SNPs showed different major alleles between intermediate groups I and II (Table S6). All 12 SNPs showed *indica* and *japonica* alleles as the major allele in intermediate groups I and II, respectively. At 11 out of 12 SNPs (except id10003706), the *japonica* allele was the major allele, and PIC values were relatively high in all accessions. Six SNPs were located within a selective sweep region and were linked to domestication- or yield-related genes, such as *Gn1a*, *Bh4*, *Prog1*, *Ghd7*, *Ehd2/OsID1/RID1*, *SNORKEL 1*, and *SNORKEL 2* (Chen et al. 2016; Chen et al. 2019). Additionally, the SNP marker id5008218 was based on the non-synonymous SNP in *CLA1* (Os05g0408900), which is one of differentially expressed genes associated with the spikelet density QTL *qSD-15* (Chen et al. 2018). These data suggest that subspecific differentiation in the local genome also existed in *O. rufipogon* before the domestication of *indica* and *japonica* subspecies.

The subspecies-specific variations, such as SNPs and InDels, showed an uneven distribution pattern in the rice genome (Fig. 6). This phenomenon occurred possibly because of the difference in kind of variations and subspecies variety panel for subspecies-specific marker detection. However, common gaps were identified among all kinds of subspecies-specific variations investigated in this study. Common gaps were located at or near the centromere, i.e., in regions of low-recombination frequency; thus, these regions might represent highly conserved sequences within large linkage blocks. On the other hand, some regions showed rare subspecies-specific variations, despite the high recombination frequency. These regions are likely to be introgressed from other subspecies by the breeder or through natural selection. There are some grain quality related genes, such as *GS6*, *Wx*, and *Alk*, on 1~7 Mb region showing few subspecies-specific variations in chromosome 6. This colocation of agronomic traits and few subspecies-specific variations might imply inter-subspecific introgression. For example, Zhao et al. (2010) showed that the *Wx^b* allele, associated with significantly lower amylose content, was introgressed from temperate *japonica* in some *indica* accessions. In our study, the SNP marker WAXY-TG was designed at the functional T/G mutation in intron 1 of the *Wx* gene, which discriminates
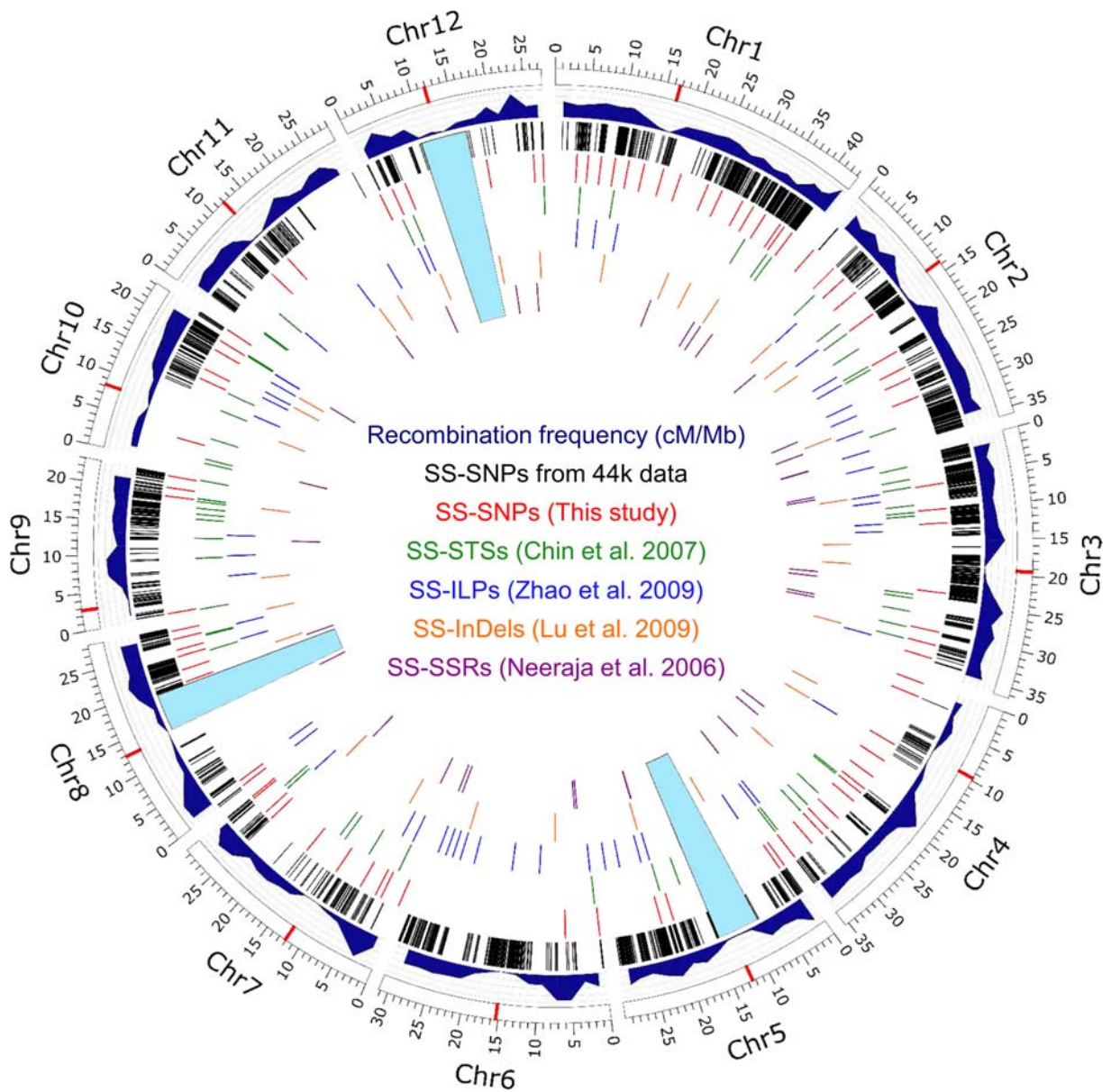
Fig. 6 Genome-wide distribution of subspecies-specific markers and recombination frequency. The outer layer represents 12 chromosomes, and the red line on each chromosome indicates the position of the centromere, based on IRGSP 1.0. Recombination frequency (ranging from 0 to 10 cM/Mb) is shown in the outer circle, based on the rice integrated genetic/physical map data from Harushima et al. (1998) and McCouch et al. (2002). Six types of subspecies-specific markers are indicated in different colors, and regions larger than 5 Mb with no subspecies-specific variation are highlighted in light blue

between the Wx^a and Wx^b alleles (Table 2). Of the 25 *indica* accessions used for the SS score calculation, 14 showed the Wx^b genotype. By contrast, only one of 25 *japonica* accessions showed the Wx^a genotype. In addition, the genotype of the SNP marker GS6-TG developed at the 3′-UTR of *GS6*, which is located 300 kb upstream of *Wx*, showed a similar introgression pattern. Of the 50 accessions, one *japonica* and 11 *indica* accessions harbored the *indica* and *japonica* alleles, respectively. This region was introgressed from *japonica* into *indica* during the development of modern *indica* varieties by breeders; therefore, this region lost subspecies specificity.

High-density SNP arrays such as Illumina Infinium (Yu et al. 2014; Chen et al. 2014; Thomson et al.

2017) and Affymetrix (Zhao et al. 2011a; McCouch et al. 2016; Singh et al. 2015) contain more than 5000 SNPs, which represent a fixed set of pre-determined SNP markers and are suitable for fixed advanced generation populations or natural populations. However, breeders aim to assay only a small number of polymorphic markers linked to target traits during molecular breeding and change the combination of markers used in subsequent generations. The Fluidigm SNP genotyping platform has automated PCR and the IFC technology, which automatically mixes PCR reagents through the microfluidic channel networks. This automated platform can genotype 2304 (48 samples × 48 assays) to 9216 (96 samples × 96 assays) samples in a single run depending on the IFC plates. The IFC-based Fluidigm genotyping system for medium-throughput multiplexing, known as the Dynamic Array, is efficient and quick, and requires a small amount of polymerase and PCR reaction buffer (Wang et al. 2009). When using less than 192 polymorphic SNPs, high-density arrays are more expensive (~$100 per sample) than the Fluidigm SNP set (~$30 per sample). Fluidigm provides three types of sample-assay combination IFCs (48.48, 96.96, and 192.24), which can be changed. Additionally, the Fluidigm SNP genotyping workflow is easy and quick.

Therefore, our *indica–japonica* SNP sets are suitable for genotyping a small number of SNPs and for changing the SNP combination, based on the objective such as target gene screening, genotyping of unfixed early generation populations (e.g., F_2 and BC_nF_1), and gene/QTL pyramiding by marker-assisted backcrossing (MABC). The Fluidigm platform has been used for target SNP genotyping in various rice genetic studies (Duitama et al. 2015; Hyun et al. 2015; Lee et al. 2017). However, a genome-wide SNP set using Fluidigm has not been reported to date. Furthermore, we developed additional SNP assays for specific target regions, based on the differences between *indica* and *japonica* accessions, using the method described in this study. SNP assays developed in this study could be converted to KASP (Kompetitive allele–specific PCR) markers for a small (≤ 24) number of SNPs and/or large (≥ 192) number of samples.

Conclusions

Indica–japonica SNP genotyping sets based on the Fluidigm platform were suitable for genotyping *indica*

and *japonica* germplasms and populations. Our results suggest that subspecies-specific SNPs existed in wild rice before the domestication of *indica* and *japonica* rice. This genotyping system will serve as an efficient tool for genetic analysis and molecular breeding in rice.

Compliance with ethical standards

Conflict of interests The authors declare that they have no competing interests.

References

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis. Models and estimation procedures. Am J Hum Genet 19: 233–257

Chen J et al (2008) A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica-japonica hybrids in rice. Proc Natl Acad Sci 105:11436–11441. https://doi.org/10.1073/pnas.0804761105

Chen H et al (2011) Development and application of a set of breeder-friendly SNP markers for genetic analyses and molecular breeding of rice (Oryza sativa L.) TAG theoretical and applied genetics. Theor Angew Genet 123:869–879. https://doi.org/10.1007/s00122-011-1633-5

Chen H et al (2014) A high-density SNP genotyping array for rice biology and molecular breeding. Mol Plant 7:541–553. https://doi.org/10.1093/mp/sst135

Chen C, Zhiguo E, Lin H-X (2016) Evolution and molecular control of hybrid incompatibility in plants. Front Plant Sci 7. https://doi.org/10.3389/fpls.2016.01208

Chen L et al (2018) Genetic analysis for the grain number heterosis of a super-hybrid rice WFYT025 combination using RNA-Seq. Rice (N Y) 11:37. https://doi.org/10.1186/s12284-018-0229-y

Chen E, Huang X, Tian Z, Wing RA, Han B (2019) The genomics of Oryza species provides insights into rice domestication and heterosis. Ann Rev Plant Biol 70:639–665. https://doi.org/10.1146/annurev-arplant-050718-100320

Cheon K-S et al (2018) Single nucleotide polymorphism (SNP) discovery and kompetitive allele-specific PCR (KASP) marker development with Korean japonica rice varieties plant breed. Biotech 6:391–403

Chin JH, Kim JH, Kwon SW, Cho YI, Piao Z, Han L, Koh HJ (2003) Identification of subspecies-specific RAPD markers in rice. Kor J Breed Sci 35:102–108

Chin J-H et al (2007) Identification of subspecies-specific STS markers and their association with segregation distortion in rice. J Crop Sci Biotechnol 10:175–184

Choi JY, Platts AE, Fuller DQ, Hsing YI, Wing RA, Purugganan MD (2017) The rice paradox: multiple origins but single domestication in Asian rice. Mol Biol Evol 34:969–979. https://doi.org/10.1093/molbev/msx049

Civáň P, Craig H, Cox CJ, Brown TA (2015) Three geographically separate domestications of Asian rice. Nat Plants 1:15164. https://doi.org/10.1038/nplants.2015.164

Civáň P, Craig H, Cox CJ, Brown TA (2016) Multiple domestications of Asian rice. Nat Plants 2:16037. https://doi.org/10.1038/nplants.2016.37

Duitama J, Silva A, Sanabria Y, Cruz DF, Quintero C, Ballen C, Lorieux M, Scheffler B, Farmer A, Torres E, Oard J, Tohme J (2015) Whole genome sequencing of elite rice cultivars as a comprehensive information resource for marker assisted selection. PloS One 10:e0124617. https://doi.org/10.1371/journal.pone.0124617

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome Res 14:1812–1819. https://doi.org/10.1101/gr.2479404

Friters A et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414. https://doi.org/10.1093/nar/23.21.4407

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in Oryza sativa L. Genetics 169:1631–1638. https://doi.org/10.1534/genetics.104.035642

Harushima Y et al (1998) A High-Density Rice Genetic Linkage Map with 2275 Markers Using a Single F_2 Population. Genetics 148:479–4

Huang X, Han B (2015) Rice domestication occurred through single origin and multiple introgressions. Nat Plants 2:15207. https://doi.org/10.1038/nplants.2015.207

Huang X et al (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490:497–501. https://doi.org/10.1038/nature11532

Hyun DY et al (2015) Development of low-temperature germinability markers for evaluation of rice (Oryza sativa L.) germplasm. Mol Breed 35:104. https://doi.org/10.1007/s11032-015-0298-1

Inoue T et al (1994) Sequence-tagged sites (STSs) as standard landmarks in the rice genome. Theor Appl Genet 89:728–734. https://doi.org/10.1007/bf00223712

Khush G (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35:25–34. https://doi.org/10.1023/A:1005810616885

Kim B et al (2014) Defining the genome structure of 'Tongil' rice, an important cultivar in the Korean "Green Revolution". Rice 7:22. https://doi.org/10.1186/s12284-014-0022-5

Kosambi DD (1944) The estimate of map distances from recombination values. Ann Eugenics 12:172–175

Kubelik AR, Williams JGK, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acids Res 18:6531–6535. https://doi.org/10.1093/nar/18.22.6531

Kubo T, Yoshimura A (2005) Epistasis underlying female sterility detected in hybrid breakdown in a Japonica–Indica cross of rice (Oryza sativa L). Theor Appl Genet 110:346–355. https://doi.org/10.1007/s00122-004-1846-y

Kubo T, Takashi T, Ashikari M, Yoshimura A, Kurata N (2016) Two tightly linked genes at the hsa1 locus cause both F1 and F2 hybrid sterility in rice. Mol Plant 9:221–232. https://doi.org/10.1016/j.molp.2015.09.014

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lee G-A et al. (2017) New genetic loci associated with preharvest sprouting and its evaluation based on the model equation in rice Front Plant Sci 8. https://doi.org/10.3389/fpls.2017.01393

Li G et al (2019) Genome-wide dissection of segregation distortion using multiple inter-subspecific crosses in rice. Sci China Life Sci 62:507–516. https://doi.org/10.1007/s11427-018-9452-8

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. https://doi.org/10.1093/bioinformatics/bti282

Lu B-R, Cai X, Xin J (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. https://doi.org/10.1016/j.pnsc.2009.01.011

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815–829. https://doi.org/10.1007/BF00273666

McCouch SR et al (2002) Development and mapping of 2240 new SSR markers for rice. DNA Res 9:199–207

McCouch SR et al (2010) Development of genome-wide SNP assays for rice. Breed Sci 60:524–535. https://doi.org/10.1270/jsbbs.60.524

McCouch SR, Wright MH, Tung C-W, Maron LG, McNally KL, Fitzgerald M (2016) Open access resources for genome-wide association mapping in rice Nat Commun 7. https://doi.org/10.1038/ncomms10532

Megersa A, Seo J, Chin JH, Kim B, Koh H-J (2016) Characterization of selected rice varieties adapted in Africa. Plant Breed Biotechnol 4:297–305. https://doi.org/10.9787/pbb.2016.4.3.297

Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J 3:269–283. https://doi.org/10.1016/j.cj.2015.01.001

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant. DNA Nucl Acids Res 8:4321–4326. https://doi.org/10.1093/nar/8.19.4321

Neeraja CN, Malathi S, Siddiq EA (2006) Subspecies — specific microsatellite markers for rice (Oryza sativa L). J Plant Biochem Biotechnol 15:39–41. https://doi.org/10.1007/bf03321899

Oka HI (1988) Origin of cultivated rice. Japan Scientific Societies Press, Tokyo

Pariasca-Tanaka J, Lorieux M, He C, McCouch S, Thomson M, Wissuwa M (2015) Development of a SNP genotyping panel for detecting polymorphisms in Oryza glaberrima/O. sativa interspecific crosses. Euphytica 201:67–78. https://doi.org/10.1007/s10681-014-1183-4

Pritcharda JK, Wena X, Falushb D (2009) Documentation for structure software: Version 2.3

Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG (2014) STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. SpringerPlus 3:431. https://doi.org/10.1186/2193-1801-3-431

Shanmugavadivel PS et al (2013) Mapping quantitative trait loci (QTL) for grain size in rice using a RIL population from basmati × indica cross showing high segregation distortion. Euphytica:1–16. https://doi.org/10.1007/s10681-013-0964-5

Shen R, Wang L, Liu X, Wu J, Jin W, Zhao X, Xie X, Zhu Q, Tang H, Li Q, Chen L, Liu YG (2017) Genomic structural variation-mediated allelic suppression causes hybrid male sterility in rice. Nat Commun 8:1310. https://doi.org/10.1038/s41467-017-01400-y

Singh N et al (2015) Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. Sci Rep 5: 11600. https://doi.org/10.1038/srep11600

Steele KA, Quinton-Tulloch MJ, Amgai RB, Dhakal R, Khatiwada SP, Vyas D, Heine M, Witcombe JR (2018) Accelerating public sector rice breeding with high-density KASP markers derived from whole genome sequencing of indica rice. Mol Breed 38:38. https://doi.org/10.1007/s11032-018-0777-2

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/Technology 7:257–264. https://doi.org/10.1038/nbt0389-257

Thomson MJ et al (2011) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. Mol Breed 29:875–886. https://doi.org/10.1007/s11032-011-9663-x

Thomson MJ et al (2017) Large-scale deployment of a rice 6 K SNP array for genetics and breeding applications. Rice 10:40. https://doi.org/10.1186/s12284-017-0181-2

Wan J, Yamaguchi Y, Kato H, Ikehashi H (1996) Two new loci for hybrid sterility in cultivated rice (Oryza sativa L). Theor Appl Genet 92:183–190

Wang J et al (2009) High-throughput single nucleotide polymorphism genotyping using nanofluidic Dynamic Arrays. BMC Genomics 10:561. https://doi.org/10.1186/1471-2164-10-561

Wu K-S, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225–235. https://doi.org/10.1007/bf00280220

Yamamoto T, Nagasaki H, Yonemaru J-i, Ebana K, Nakajima M, Shibaya T, Yano M (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11:267

Yang J et al (2012) A killer-protector system regulates both hybrid sterility and segregation distortion in rice. Science 337:1336–1340. https://doi.org/10.1126/science.1223702

Yang G et al (2019) Development of a core SNP arrays based on the KASP method for molecular breeding of rice. Rice (N Y) 12:21. https://doi.org/10.1186/s12284-019-0272-3

Yonemaru J, Ebana K, Yano M (2014) HapRice, an SNP haplotype database and a web tool for rice. Plant Cell Physiol 55: e9. https://doi.org/10.1093/pcp/pct188

Yu H, Xie W, Li J, Zhou F, Zhang Q (2014) A whole-genome SNP array (RICE6K) for genomic breeding in rice. Plant Biotechnol J 12:28–37. https://doi.org/10.1111/pbi.12113

Zhao X, Yang L, Zheng Y, Xu Z, Wu W (2009) Subspecies-specific intron length polymorphism markers reveal clear genetic differentiation in common wild rice (Oryza rufipogon L.) in relation to the domestication of cultivated rice (O. sativa L.). J Genet Genomics 36:435–442. https://doi.org/10.1016/S1673-8527(08)60133-2

Zhao K et al (2010) Genomic diversity and introgression in O. sativa reveal the impact of domestication and breeding on the rice genome. PloS One 5:e10780. https://doi.org/10.1371/journal.pone.0010780

Zhao K et al (2011a) Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2:467. https://doi.org/10.1038/ncomms1467

Zhao ZG et al (2011b) Molecular analysis of an additional case of hybrid sterility in rice (Oryza sativa L.). Planta 233:485–494. https://doi.org/10.1007/s00425-010-1313-8

Zhao H et al (2015) RiceVarMap: a comprehensive database of rice genomic variations. Nucl Acids Res 43:D1018–D1022. https://doi.org/10.1093/nar/gku894