Genotyping by sequencing of rice interspecific backcross inbred lines identifies QTLs for grain weight and grain length

Article *in* Euphytica · January 2018

DOI: 10.1007/s10681-018-2119-1

CITATIONS
14

READS
374

8 authors, including:

Dharminder Bhatia
Punjab Agricultural University
49 PUBLICATIONS **516** CITATIONS
SEE PROFILE

Rod A Wing
The University of Arizona
519 PUBLICATIONS **41,189** CITATIONS
SEE PROFILE

Yeisoo Yu
DNACARE
184 PUBLICATIONS **20,650** CITATIONS
SEE PROFILE

Kapeel Chougule
Cold Spring Harbor Laboratory
27 PUBLICATIONS **439** CITATIONS
SEE PROFILE

Euphytica (2018)214:41
https://doi.org/10.1007/s10681-018-2119-1

Genotyping by sequencing of rice interspecific backcross inbred lines identifies QTLs for grain weight and grain length

Dharminder Bhatia · Rod A. Wing · Yeisoo Yu · Kapeel Chougule · Dave Kudrna · Seunghee Lee · Allah Rang · Kuldeep Singh

Received: 31 July 2017 / Accepted: 22 January 2018

Abstract Grain weight and grain length are the most stable components of rice yield and important indicators of consumer preference. Considering the potentials of wild rice and to enhance the rice yields to meet the increasing demands, 185 Backcross Inbred Lines (BILs) in the background of *O. sativa* ssp. *indica* cv. PR114, including 63 *rufi*-BILs derived from *O. rufipogon* IRGC104433 and 122 *glumae*-BILs from *O. glumaepatula* IRGC104387 were evaluated for mapping QTLs for yield and yield component traits using Genotyping by Sequencing (GBS). Phenotypic evaluation of BILs in three seasons spanning two locations revealed significant differences compared with recurrent parent. BILs which did not show significant differences for any trait under investigation, or similar based on pedigree, were excluded from GBS. Some *glumae*-BILs had to be excluded from mapping QTLs due to less sequence information. A custom designed approach for GBS data analysis identified 3322 informative SNPs in 55 *rufi*-BILs and 3437 informative SNPs in 79 *glumae*-BILs. QTL mapping identified one QTL for thousand grain weight (*qtgw5.1*), two for grain width (*qgw5.1*, *qgw5.2*) and one for grain length (*qgl7.1*) in *rufi*-BILs. In the *glumae*-BILs, three QTL for thousand grain weight (*qtgw2.1*, *qtgw3.1, qtgw6.1*) and two for grain length (*qgl3.1, qgl7.1*) were identified. Most of the grain weight and width QTL showed positive additive effect contributed by wild species allele, whereas the grain length QTL showed positive additive effect contributed by recurrent parent allele. Based on their physical position, none of the QTLs were found similar to previously cloned QTLs. QTLs for grain traits identified from low yielding wild relatives of rice reveals their significance in improving further the rice yields and widen the genetic base of cultivated rice.

Keywords Backcross inbred lines (BILs) · *O. rufipogon* · *O. glumaepatula* · Grain weight · Grain length · Genotyping by sequencing (GBS) · Quantitative trait loci (QTL)

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10681-018-2119-1) contains supplementary material, which is available to authorized users.

D. Bhatia (✉) · A. Rang
Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana 141004, India
e-mail: d.bhatia@pau.edu

D. Bhatia · K. Singh
School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana 141004, India

D. Bhatia · R. A. Wing · Y. Yu · K. Chougule · D. Kudrna · S. Lee
Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ, USA

Present Address:
K. Singh
National Bureau of Plant Genetic Resources, New Delhi, India

Published online: 31 January 2018

Introduction

Increasing the yield potential of rice is one of the prime objectives to overcome impending food crises caused by a large increase in the world population. A quantum jump in the rice yields was observed during the 1960–1970s with improvements in harvest index and plant architecture through the use of semi-dwarfing genes, known as era of "Green Revolution". Thereafter, significant improvement in rice yield came from incorporating resistance to diseases, insect pests and better agronomic practices. However, rice yield growth has fallen by 2.3% per year during 1970s–1990s to 1.5% during the 1990s and to $< 1.0\%$ during the first decade of present century (Khush 2013). To further enhance the yield potential of rice over that of existing high yielding cultivars, there is a need to unfold the genetic determinants of grain yield. However, because grain yield is a complex trait and is the outcome of many processes and factors, direct analysis of grain yield has not given ideal results. Therefore, it is necessary to conduct a proper dissection and focused studies on yield components instead of yield itself as a whole.

The grain yield of rice can be dissected into four major components: productive tiller number per plant, spikelets per panicle, percent spikelet fertility, and grain weight. All these yield components express continuous phenotypes and are governed by number of genes called quantitative trait loci (QTL). Out of these, grain weight is mainly determined by grain width, grain length and degree of grain filling. Grain weight and grain length are are the most stable components of rice yield and important indicators for consumer preference. Both grain weight and grain length are highly heritable, thus making them useful for genetic analysis (Chauhan 1998).

Due to limited variability in the germplasm with which breeders are working, a major source of genetic variability for these traits can come from its wild relatives. In the recent past, a common wild rice accession, *O. rufipogon* IRGC105491, has been utilized to improve the yield component traits of cultivated rice (McCouch et al. 2007; Cheema et al. 2008; Jin et al. 2009; Imai et al. 2013). However, immense genetic variability is still not characterized in thousands of accessions of the twenty-two wild species of rice. In this case, interspecific backcross inbred lines (BILs), derived through limited backcrossing a wild-by-cultivated cross with the cultivated parent, are useful tools to introgress and map QTLs for yield component traits. Since BIL carries only a small fraction of the wild species genome, most fertility problems are eliminated and yield-associated traits can be measured more precisely (Eshed and Zamir 1995). BILs might get an advantage over advanced backcross-QTL (AB-QTL) mapping populations, because (1) utilization of large number of wild species accessions using limited backcross strategy (Bhatia et al. 2017); (2) the presence of fairly uniform background between lines and relatively rapid and straight forward utilization for commercial plant breeding (Jeuken and Lindhout 2004).

In rice, numerous studies have been conducted to genetically map QTL for grain traits, and thousands of QTL have been detected over the past several decades. Among these, more than 300 QTL that control grain length, grain width and grain weight have been identified (www.grameme.org). A large number of QTL have also been identified from *O. rufipogon* IRGC105491, where *O. rufipogon* alleles have shown to contribute to increased grain weight (Xiao et al. 1998; Moncada et al. 2001; Thomson et al. 2003; Septiningsih et al. 2003; Xie et al. 2006, 2008; Wickneswari et al. 2012). A few major grain weight and grain length QTL have also been cloned (Fan et al. 2006; Song et al. 2007; Shomura et al. 2008; Li et al. 2011). For example, *gw3.1*, a QTL for grain length and grain weight was identified in ten independent studies involving different bi-parental populations, and was later cloned as *GS3* (Fan et al. 2006) which suggest that QTL for these traits are stable in different genetic backgrounds and environments, and are suitable targets for grain yield improvement (Li et al. 2004).

Conventional QTL analysis remains time-consuming and labour-intensive, mainly because it requires the development of polymorphic markers for linkage analysis. With recent developments in next-generation sequencing (NGS) technology, genome complexity reduction methods such as genotyping by sequencing (GBS), have emerged as powerful genotyping platforms. For example, GBS can be used to discover thousands of markers across almost any genome of interest in number of individuals in a population in a single and simple experiment (Davey et al. 2011). Many plant breeders and geneticists are still using sparse molecular marker data due to extremely resource limited programs (Rosyara et al. 2009). The

low coverage sequencing and multiplexing of samples in GBS lowers the genotyping cost and the resultant NGS data provide large number of markers for immediate applications in gene/QTL discovery and genomic-assisted breeding (Thomson et al. 2012). Flexible, rapid and low cost GBS protocols have proved its worth in many applications ranging from marker discovery to linkage/association mapping in a number of crops to address many questions of plant breeding and genetics (Poland and Rife 2012; Bhatia et al. 2013).

The present study was conducted for mapping QTLs for yield and yield component traits using a GBS approach for genotyping BILs derived from two different wild species, *O. rufipogon* IRGC104433 and *O. glumaepatula* IRGC104387, in the background of recurrent parent *O. sativa* ssp. *indica* cv. PR114.

Materials and methods

Experimental material

A set of 185 PR114-BILs were selected based on agronomic performance and variation for yield and yield component traits (Bhatia et al. 2017). Of the 185 PR114-BILs, 63 were derived from *O. rufipogon* IRGC104433, and 122 were from *O. glumaepatula* IRGC104387 using *O. sativa* ssp *indica* cv. PR114 as common recurrent parent for both crosses (Table S1). The PR114-BILs derived using *O. rufipogon* IRGC104433 and *O. glumaepatula* IRGC104387 are referred as *rufi*-BILs and *glumae*-BILs, respectively, and aggregate as BILs further in the manuscript.

Experimental design

Field trials to evaluate the phenotypic performance of the BILs and recurrent parent PR114, were conducted in three growing seasons in two different locations. The two locations were Punjab Agricultural University (PAU), Ludhiana, India situated at 30.91°N, 75.85°E longitude and latitude (North-West of India), and National Rice Research Institute (NRRI), Cuttack, Odisha at 20.27°N, 85.52°E longitude and latitude (South East of India) respectively. The three growing seasons consists of two main rice growing seasons (April–September) of 2011 and 2012 at PAU, and an off-season (November–April) of year 2011–12 at NRRI. All BILs were evaluated in a square lattice (14×14) design with two replications and plot sizes of 1.95 m^2 at PAU, and 1.275 m^2 at NRRI. BILs were planted with a plant-to-plant and row-to-row distance of 15 cm × 20 cm, and cultivated following the recommended package of practices of each region. A high yielding variety of PAU named as PAU201 was also included as check, but comparisons of phenotypic performance of BILs were made with recurrent parent PR114. Details of the pedigree of each BIL are given in Table S1.

Evaluation of BILs for yield component traits

Observations on plot yield and twelve yield component traits were recorded for each BIL as well as the check. Yield component traits included: days to 50% flowering (DF), plant height (PH), tillers/plant (TL), panicle length (PL), flag leaf length (FLL), flag leaf width (FLW), spikelet/panicle (SP), percent fertility (PF), thousand grain weight (TGW), grain length (GL), grain width (GW), along with plot yield (PY). Briefly, DF was evaluated as the average number of days when 50% of the tillers of each plot flowered; PH was measured on ten random plants from ground level to the tip of the tallest panicle and averaged; TL were counted from ten competitive plants; PL was measured from panicle base to the tip of the panicle; FLL was measured from base to tip of the flag leaf after booting stage, while FLW was taken from the middle of flag leaf; SP was recorded by counting the total number of grains (including sterile spikelets) from ten panicles taken from ten random plants; PF was estimated by dividing fertile grains with SP; TGW was recorded by weighing 1000 fully filled grains counted from a bulk of ten panicles; GL and GB was measured with a Dial Thickness Gauge (Baker Mercer, India), and plot yield (PY) was recorded on a whole plot basis. Analysis of Variance (ANOVA) and least square difference ($p \leq 0.05$) for all the traits was calculated with SAS version 9.2 (SAS institute, Cary, NC) using *Proc Lattice*. Differences between adjusted means of the recurrent parent and each BIL was compared with LSD values to identify the BILs that were significantly different for the target traits. Significant BILs from each season were manually compared to identify consistent performing BILs for each trait.

GBS library preparation

BILs which did not show significant differences with the recurrent parent for any trait under investigation, or were very closely related based on pedigree, were excluded from GBS library construction. A total of 55 *rufi*-BILs and 101 *glumae*-BILs along with the parents, each replicated eight times, were used for GBS. We used a modified GBS method with two restriction enzymes. Briefly, total genomic DNA of all the BILs and their corresponding parents were isolated from leaf tissue using a CTAB (cetyl trimethyl ammonium bromide) extraction method (Murray and Thompson 1980), with the chloroform-isoamyl alcohol purification step repeated twice to assure good quality DNA. DNA fragments were generated by restriction enzyme digestion with HindIII and MspI, followed by ligation with dual-indexed Illumina P5 and P7 adapters (5 bp index sequences for P5 and 6 bp index for P7 adapter). A set of 24 P5 (HindIII digested end) and 16 P7 (MspI digested end) Illumina adaptors containing different index sequences were designed, which in combination is sufficient to tag 384 different DNA samples in a GBS library (Table S2). Adaptor specific primers were used to amplify the GBS fragments. Fragments which had P5 adaptor sequence ligated at one end, and P7 adaptor sequences at other were the only ligation products that could be amplified. After quantification, equimolar concentration of each PCR amplified sample was pooled into one tube. Smaller fragments (primer dimers or with less information) were removed from the pooled GBS library using 1:0.7 × AmpureXP Beads (exclude < 200 bp GBS constructs) before sequencing. The Agilent High Sensitivity DNA assay kit was used for analysis of DNA fragments with a Bioanalyzer (Agilent Technologies), and qPCR was used to determine the concentration of each pooled library. After quantification, each DNA pool was diluted to 2nmoles/µl and loaded onto two lanes of an Illumina HIseq 2500.

High-throughput genotyping of BILs

GBS data was analysed individually for each BIL population with a custom designed method (Fig. S1). Pair end reads of 100 nucleotides of each individual were sorted by unique index combination. The adaptor sequence including index sequences were trimmed by custom script. The updated genome sequence of *O. sativa* ssp. *indica* cv 93-11 (Gao et al. 2013) was used as the reference *indica* genome in this study. Pair end reads of each parent were aligned to the reference genome using BWA with a threshold quality score of 20 (Li and Durbin 2009). SNPs were called using both Samtools (Li et al. 2009) and GATK (McKenna et al. 2010) and SNPs common in both algorithms were identified. SNPs were then filtered based on read depth ≥ 2 and associated quality scores ≥ 30. Final SNPs called in the parents after this step were replaced in the 93-11 genome sequence using a Perl script, *PseudoMaker*, implemented in SEG-Map (Zhao et al. 2010) to construct the fake pseudomolecules of each parent. Parental SNPs were compared to identify SNPs polymorphic between two parents of each BIL population. Thereafter, short reads of each individual BIL in the population were aligned to fake pseudomolecule of each parent and SNPs were called with the same procedure as with the parents. Here, BIL SNPs that did not corresponds to polymorphic parental SNP positions, were rejected as these were with 93-11 sequence. BIL SNPs were then filtered with $\leq 8\%$ missing data points and the final SNPs obtained at this point were called informative SNPs. Any heterozygous SNP was considered as missing data point due to low confidence SNP call in low coverage GBS data. Informative SNPs were provided genetic positions on each chromosome using the Kosambi mapping function, and plotted based on their physical and genetic distance on the twelve linkage groups. Graphical genotypes was prepared using informative SNPs to see the extent of introgression of the wild donor species in the background of recurrent parent, and recovery of the recurrent parent genome after two backcrosses. To identify the position of QTL for yield and yield component traits, QTL mapping was conducted using inclusive composite interval mapping implemented in QTL IciMapping software version 3.1 (Meng et al. 2015) with population code 11, informative SNPs and phenotypic data of target traits across all the seasons. Significant thresholds were found with 1000 permutations in stepwise regressions with $p < 0.001$. Location of QTL was described according to LOD (logarithm of odds) value. The contribution rate (PVE) was estimated as the percentage of variance explained by each QTL in proportion to the total phenotypic variance. Additive effect was estimated to find the positive or negative effect on the target trait.

Results

Phenotypic evaluation of BILs

ANOVA revealed significant differences in the BILs for yield and yield component traits as compared with recurrent parent in all the seasons (Tables 1, 2; Fig. 1a, b; Fig. S2). However, the observed phenotypic evaluation featured variations for different yield components in *rufi*-BILs and *glumae*-BILs. Desirable variations for DF (short duration) and SPP were observed in *glumae*-BILs (Fig. 2a, b), while no desirable variation for these traits were observed in *rufi*-BILs. Increase in GW accompanied by a decrease in GL was observed in the *rufi*-BILs, while both positive and negative variation for GL and GW were present in *glumae*-BILs (Fig. 2c). Variation for FLL was more prominent in the *glumae*-BILs as compared with *rufi*-BILs (Fig. 2d), while variation for FLW was observed in both *rufi*-BILs and *glumae*-BILs.

Correlation of traits among seasons

Correlation coefficients (r) was calculated for eight different traits evaluated in three seasons to find associations among seasons (Table S3). There were significantly positive correlations among the seasons for almost all yield component traits. The magnitude of correlations were higher in the two different seasons at PAU than between seasons at two locations (PAU and NRRI). The highest magnitude of significant positive correlation was estimated for TGW ranging from 0.84 to 0.91, while lowest was estimated for PF ranging from 0.05 to 0.23. There was no significant correlation for PF between season 1 and 3.

High throughput genotyping of BILs using GBS

Identification of parental SNPs

A total of 78.3 GB of raw sequence data in form of 100 nucleotide pair-end Illumina reads was generated with an average of 0.08X genome coverage and 14.5X allele depth per base for the BIL parents; 0.05X genome coverage and 7.07X allele depth per base for the BILs. More than 80% of the reads of the BIL parents aligned to 93-11 genome, with highest being *O. glumaepatula*. With two different SNP calling algorithms, GATK and Samtools, a total of 41,180 SNPs including 35,708 homozygous SNPs with PR114; 146,848 SNPs including 57,090 homozygous SNPs with *O. rufipogon*; 100,681 SNPs including 93,354 homozygous SNPs with *O. glumaepatula* were identified using 93-11 genome as reference (Table 3). In order to identify SNPs from BILs, we first had to generate fake pseudomolecule of each parent. For making fake pseudomolecules, homozygous SNPs identified between each parent and reference molecule were used.

Identification of informative SNPs

More than 75% of 0.2– > 2 million short reads generated for each BIL aligned to the parental fake pseudomolecules (Fig. S3). For the *rufi*-BILs, a total of 85,483 SNPs were identified with PR114 pseudomolecule, and 97,130 SNPs with the *O. rufipogon* pseudomolecule. Similarly, for the *glumae*-BILs, a total of 149,367 SNPs were identified with PR114 pseudomolecule, and 154,049 SNPs with the *O. glumaepatula* pseudomolecule (Table 4).

SNPs identified with each parent fake pseudomolecule contains a number of SNPs called between BILs and 93-11 genome. In order to identify the informative SNPs (SNPs that belong to BIL population), only those SNPs were kept, which corresponded to the parental polymorphic SNPs positions. Out of 101 *glumae*-BILs, 22 could not be used for identification of informative SNPs and genotyping due to very less sequence information resulting in large amount of missing data points. Thereafter filtering based on missing data points, a total of 3322 informative SNPs in the *rufi*-BILs and 3437 informative SNPs in the *glumae*-BILs were identified. Informative SNPs for both *rufi*-BILs and *glumae*-BILs were uniformly distributed across their genomes, with few major gaps as on chromosomes 1 and 6 based on their physical positions (Fig. S4a, b). Informative SNPs of both the *rufi*-BILs and *glumae*-BILs were given genetic position using "Kosambi" mapping function (Fig. 3a, b). The total genetic map length was 2225.14 cM in *rufi*-BILs, whereas 3076.11 cM in *glumae*-BILs.

Percent wild genome introgression in BILs

Graphical genotyping using informative SNPs (Fig. S5a, b) indicated *O. rufipogon* genome

Table 1 Mean, range and ANOVA of eight yield and yield component traits of BILs evaluated in three growing seasons

Season	Source	DF	PH (cm)	TL	PL (cm)	SPP	PF	TGW (g)	PY (t/h)
SI	PR114[a]	98.8	101.2	9.3	25.3	133.9	83.5	23.2	516.60
	Range[b]	82.5–127.5	85.0–125.5	7.1–13.1	21.0–28.7	94.5–157.6	74.1–90.7	19.4–30.7	223.83–749.83
	Genotype MS	19373.0[c]	15589.0[c]	249.6[c]	760.1[c]	58421.0[c]	6718.2[c]	1416.9[c]	18049.0[c]
	LSD ($p \leq 0.05$)	4.7	6.2	1.6	1.2	22.7	6.3	2.4	139.82
SII	PR114	121.8	83.5	8.4	23.4	154.9	88.2	23.5	470.63
	Range	108.6–135.0	66.9–95.4	6.8-10.3	20.1–24.5	91.5–200.8	76.8–95.0	19.5–31.2	200.12–755.85
	Genotype MS	4217.4[c]	6169.4[c]	224.1[c]	213.6[c]	5399.5[c]	154505.0[c]	1455.4[c]	16939.0[c]
	LSD ($p \leq 0.05$)	4.2	8.3	1.4	1.5	30.2	7.1	3.1	159.28
SIII	PR114	100.7	101.0	12.2	23.7	116.1	88.6	23.8	542.30
	Range	76.0–127.5	83.1–125.5	8.9–17.3	20.0–28.0	76.2–144.3	75.1–94.9	20.0–29.5	252.60–763.24
	Genotype MS	18848.0[c]	15589.0[c]	3655.4[c]	7969.5[c]	315529.0[c]	159592.0[c]	13329.0[c]	18561.0[c]
	LSD ($p \leq 0.05$)	3.2	6.2	3.1	1.6	21.7	8.7	2.9	136.52

SI, SIII are two different growing seasons of year 2011 and 2012 at Punjab Agricultural University, Ludhiana and SII season belongs to year 2011–12 at National Rice Research Institute, Cuttack

DF days to 50% flowering, *PH* plant height, *TL* tillers/plant, *PL* panicle length, *SPP* spikelet/panicle, *PF* percent spikelet fertility, *TGW* thousand grain weight, *PY* plot yield

[a,b]Values based on adjusted means of square lattice; [c]Genotype Mean square values for each trait found significant at $p \leq 0.05$ as well as $p \leq 0.01$

Table 2 Mean, range and ANOVA of leaf and grain traits of BILs evaluated in season III

Source	FLL (cm)	FLW (mm)	GL (mm)	GW (mm)
PR114[a]	33.7	14.2	7.60	2.02
Range[b]	25.7–48.9	13.2–17.9	6.10–8.13	1.58–2.60
Genotypes MS	18914.0[c]	2163.4[c]	61.9[c]	763.3[c]
LSD ($p \leq 0.05$)	7.5	1.9	0.24	0.12

GL grain length, *GW* grain width, *FLL* flag leaf length, *FLW* flag leaf width

[a, b]Values based on adjusted means of square lattice

[c]Genotype Mean square values for each trait found significant at $p \leq 0.05$ as well as $p \leq 0.01$

introgressions in the range of 2.4–29.0% in the *rufi*-BILs and 0.4–27% *O. glumaepatula* genome introgressions in the *glumae*-BILs (Table S4, S5). Less than 10% introgression was identified in almost 50% of the *rufi*-BILs. The smallest percent introgression (2.4%) was identified in BIL numbers 2456 and 2402, while largest percent introgression (29%) was identified in BIL number 2432. Similarly, in *glumae*-BILs, the smallest percent introgression (0.4%) was identified in BIL number 2492, while the largest percent introgression (27.5%) was identified in BIL number 2578. Sixty *glume*-BILs out of a total 101 had less than 10% *O. glumaepatula* genome introgression.

QTL mapping for yield component traits

QTL mapping in the *rufi*-BILs with 3322 informative SNPs and twelve target traits identified four QTLs for TGW, GW and GL (Table 5). However QTLs for rest of yield component traits could not be identified. QTL for TGW (*qtgw5.1*) was consistently identified on chromosome 5 in all the seasons (Fig. 4a). The *O. rufipogon* allele at this locus showed positive additive effect in the *rufi*-BILs. Two QTLs for GW (*qgw5.1* and *qgw5.2*) were also mapped on chromosome 5 (Fig. 4b). The *O. rufipogon* allele showed positive additive effect at *qgw5.1*, whereas negative additive effect at *qgw5.2*. The QTL for GL was mapped on chromosome 7 and designated as *qgl7.1* (Fig. 4c). The recurrent parent allele at this locus showed positive additive effect on grain length and negative additive effect on grain width (Table 5).

Similarly, QTL mapping of the *glumae*-BILs with 3437 informative SNP markers identified three QTLs for TGW and two for GL (Table 6). Of the three QTLs for grain weight (*qtgw2.1*, *qtgw6.1*, *qtgw3.1*), the *O. glumaepatula* allele showed positive additive effect on TGW at QTL *qtgw2.1* and *qtgw6.1*, whereas negative effect was observed at *qtgw3.1* (Fig. 5a). Two QTLs for GL were identified on chromosome 3 and chromosome 7 respectively (Fig. 5b). Of these, the *qgl7.1* identified in *rufi*-BILs was also identified in *glumae*-BILs. This QTL share same genomic position and allele effect in both *rufi*-BILs and *glumae*-BILs. The recurrent parent allele at *qgl7.1* showed positive additive effect on the grain length similar to *rufi*-BILs. In addition, one more QTL (*qgl3.1*) was identified in *glumae*-BILs, where recurrent parent allele showed positive additive effect on grain length (Fig. 5b).

Discussion

Rice grain length and grain weight are important to both consumers and farmers because these traits determine the physical appearance of grain, affect cooking quality and grain yield. Grain size and weight has also played an important role in the evolution of cereal crops as humans tended to select for large seed size, and consequently higher grain weight during the early domestication process (Harlan 1992). The wild species of rice though have small grain size and very low grain weight compared with cultivated rice, but contain large amount of cryptic and unutilized variation for these traits. However, the major concern to utilize the wild species is to selectively transfer useful variation while avoiding linkage drag, using a combination of conventional and molecular breeding strategies (Brar and Singh 2011).

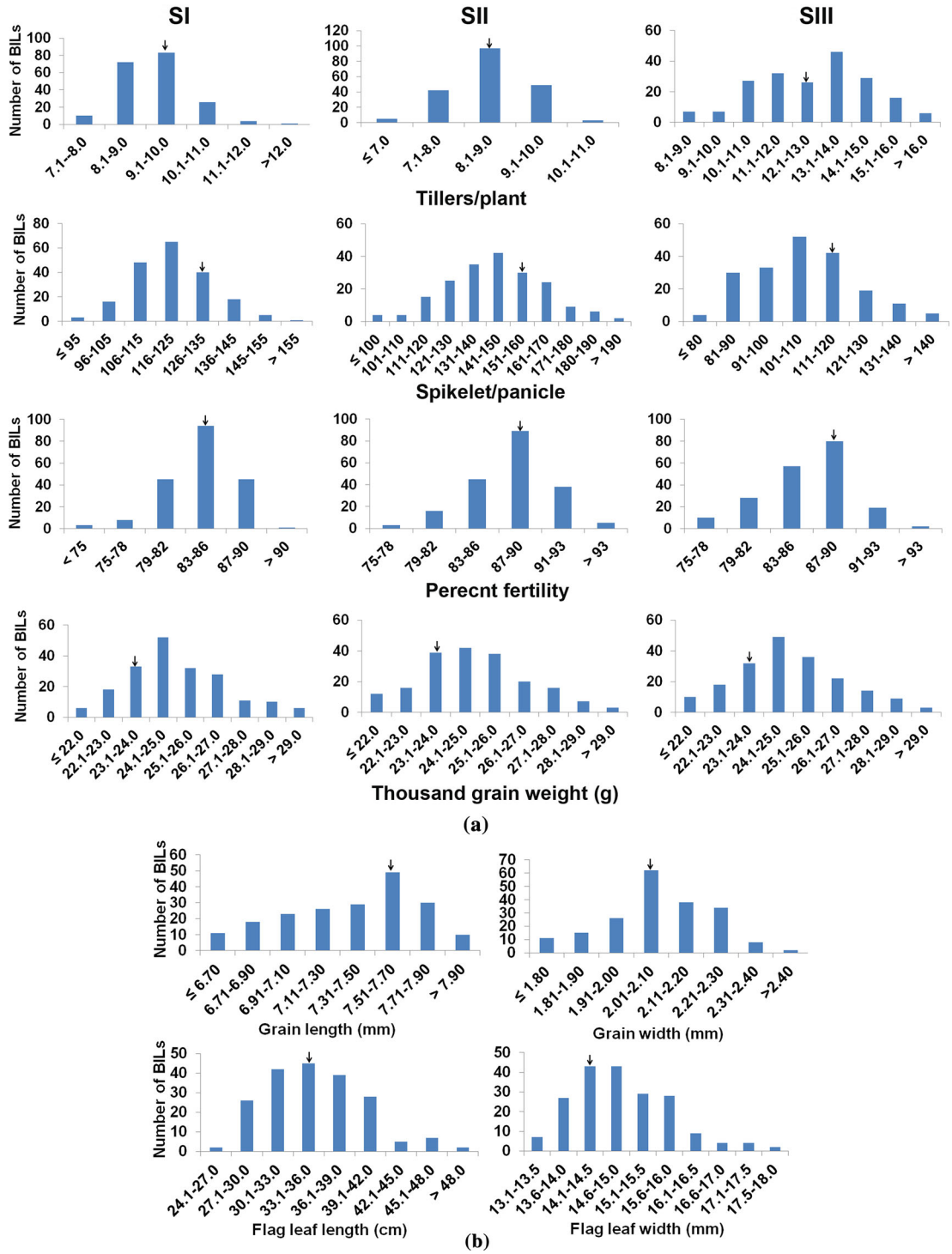

Fig. 1 **a** Graphical plots of phenotypic variation for major yield component traits in BILs. Plotted values are the adjusted means of square lattice design. SI, SII, SIII represent three different seasons. Arrow indicates where the recurrent parent PR114 falls. **b** Graphical plots of phenotypic variation in season III in GW, GL, FLL and FLW in BILs. Arrow indicates where the recurrent parent PR114 falls

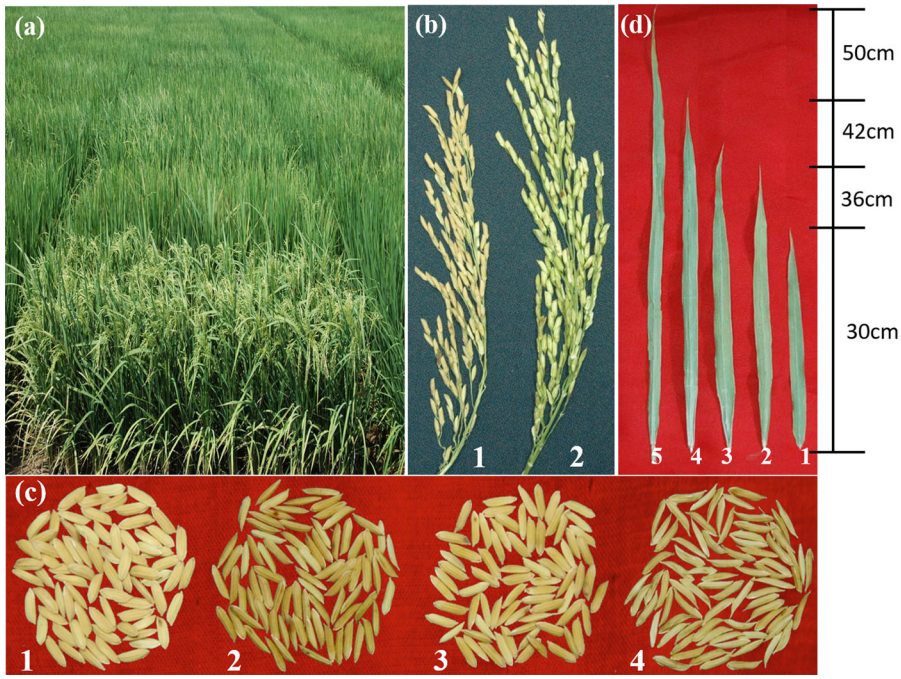

Fig. 2 Variations for yield component traits in BILs based on field evaluation **a** early flowering *glumae*-BIL; **b** increase in SPP of *glumae*-BIL (2) as compared to recurrent parent PR114 (1); **c** variation in grain size in *rufi*-BIL and *glumae*-BILs (3, 4) as compared to PR114 (2); **d** variation for FLL in the *rufi*-BILs (2, 4) and *glumae*-BILs (3, 5) as compared to PR114 (1)

Table 3 Summary of read alignment and SNP calling statistics of BIL parents with *O. sativa* ssp. *indica* cv. 93-11

GBS parameters	PR114	*O. rufipogon* (IRGC104433)	*O. glumaepatula* (IRGC104387)
Total reads	14,005,318	16,107,448	12,810,200
Mapped reads	11,725,905 (83.72%)	13,298,845 (82.56%)	11,339,282 (88.51%)
Common SNPs identified with two algorithms (Samtools and GATK)	41,180	146,848	100,681
Heterozygous SNPs	5472	89,758	7327
Homozygous SNPs	35,708	57,090	93,354

Table 4 SNP calling statistics of *rufi*-BILs and *glumae*-BILs with the fake parent pseudomolecules

GBS parameters	PR114 pseudomolecule	*O. rufipogon* pseudomolecule	PR114 pseudomolecule	*O. glumaepatula* pseudomolecule
SNPs called with Samtools	119,615	131,273	197,744	202,074
SNPs called with GATK	181,987	193,915	334,391	340,261
Common SNPs	85,483	97,130	149,367	154,049
Informative SNPs	2379	8025	6267	17,137
Informative SNPs used for mapping	3322		3437	

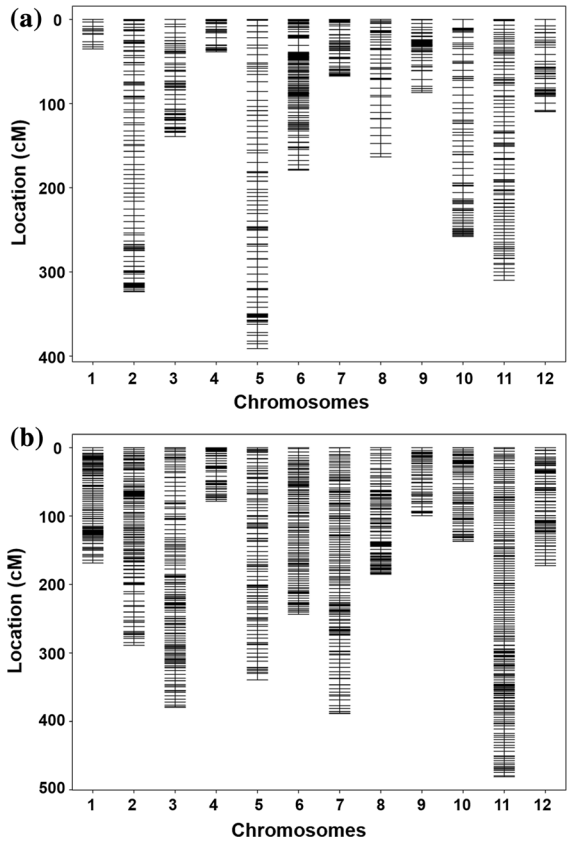

Fig. 3 Genetic maps showing distribution of **a** 3322 informative SNPs of *rufi*-BILs; **b** 3437 informative SNPs of *glumae*-BILs on twelve chromosomes of rice based on Kosambi genetic distance

Importance of BILs and GBS in interspecific crosses

In case of interspecific crosses involving wild rice species, advanced mapping populations like BILs are more preferred as compared to primary mapping populations such as recombinant inbred lines, doubled haploids, backcross and F_2/F_3 populations (Fukuoka et al. 2010; Jacquemin et al. 2013). In the present study, BILs were developed using two wild species for introducing the variability of yield component traits in the background of cultivated rice. During development, BILs were selected for uniform flowering time and plant height so as to maintain uniform background and precise evaluation of yield component traits (Bhatia et al. 2017). Significant variation observed in most of the phenotypic traits clarifies the contribution of the agronomically inferior wild species for yield components traits.

Genotyping by sequencing (GBS) has recently generated high levels of interest within the plant breeding and genetics community. Traditional mapping experiments rely on PCR based markers such as SSR's, which are sparsely located on the genome, thus resulting in low resolution gene/QTL mapping. Genotyping with traditional markers is very time consuming and can take from weeks to months to complete (Spindel et al. 2013). GBS in the current experiment took only a week to prepare the library and a week for sequencing, therefore completing the whole process of genotyping in just 14 days. Our custom designed

Table 5 Summary of QTLs identified in *rufi*-BILs for yield component traits using inclusive composite interval mapping (ICIM)

Trait	QTL	Season	Chromosome	Left marker	Right marker	LOD	PVE (%)	Additive effect/allele effect	Physical region on genome (kb)
TGW	*qtgw5.1*	I	5	Chr05_6059062	Chr05_6145897	9.35	68.00	2.10/ru	86.83
		II	5	Chr05_6059062	Chr05_6145897	9.66	68.67	2.09/ru	
		III	5	Chr05_6059062	Chr05_6145897	8.81	65.95	2.14/ru	
GW	*qgw5.1*	III	5	Chr05_4926531	Chr05_5271075	13.02	74.71	0.13/ru	344.54
	qgw5.2	III	5	Chr05_5664169	Chr05_5696540	4.86	19.43	− 0.09/ru	32.371
	qgw7.1	III	7	Chr07_21416336	Chr07_21779720	6.22	20.46	0.08/ru	363.84
GL	*qgl7.1*	III	7	Chr07_21416336	Chr07_21779720	10.19	53.88	0.26/pr	363.84

PVE % is the percent phenotypic variability explained by particular QTL

TGW thousand grain weight, *GW* grain width, *GL* grain length; *ru O. rufipogon* allele, *pr* PR114 allele

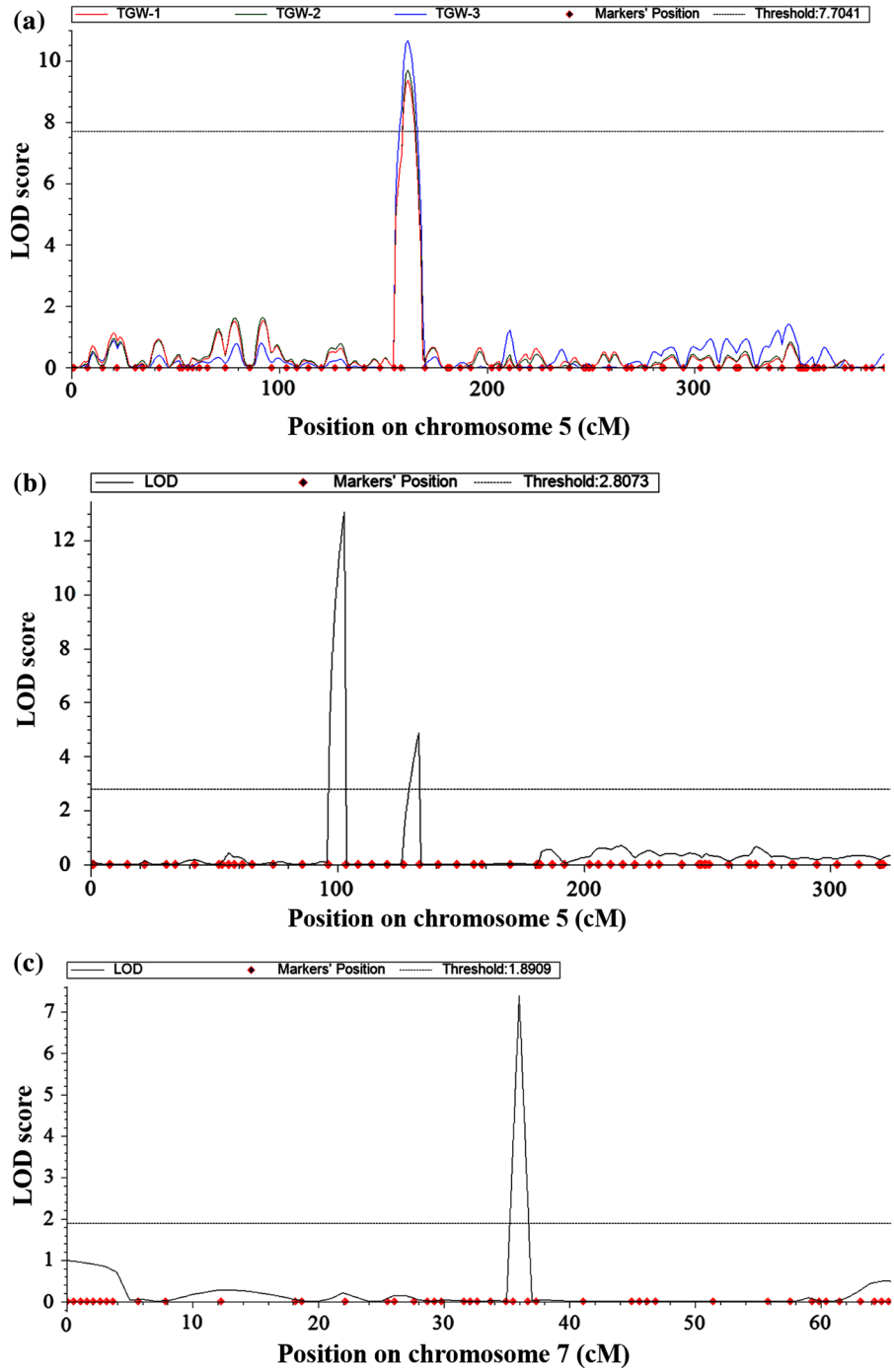

Fig. 4 LOD curves of QTL mapping in *rufi*-BILs **a** TGW for all three seasons; **b** GW, on chromosome 5 and **c** GL on chromosome 7. Horizontal dotted lines are the LOD thresholds. (Color figure online)

approach identified a total of 3322 and 3437 informative SNPs in the *rufi*-BILs and *glumae*-BILs respectively. The number of SNP markers identified with GBS in the current experiment is far more than any traditional marker system used for genotyping. For example, a total of 380 RFLP markers were used by Xio et al. (1998), of which only 28% were found to be polymorphic between *O. rufipogon* and V20A. Similarly, only 157 polymorphic SSR markers were used for mapping of yield and yield component traits in

Table 6 Summary of QTLs identified in *glumae*-BILs for yield component traits using inclusive composite interval mapping (ICIM)

Trait	QTL	Season	Chromosome	Left marker	Right marker	LOD	PVE (%)	Additive effect/allele effect	Physical region on genome (kb)
TGW	*qtgw2.1*	I	2	chr02_8905034	chr02_8996699	5.52	15.44	0.91/gl	91.6
		II	2	chr02_8905034	chr02_8996699	5.11	16.19	0.97/gl	
		III	2	chr02_8905034	chr02_8996699	6.60	23.70	1.38/gl	
	qtgw3.1	I	3	chr03_30178837	chr03_30382043	11.78	31.25	− 1.53/gl	203.2
		II	3	chr03_30178837	chr03_30382043	13.93	46.28	− 1.91/gl	
		III	3	chr03_30178837	chr03_30382043	10.71	33.13	− 1.68/gl	
	qtgw6.1	I	6	chr06_1819184	chr06_1893743	25.99	74.65	3.59/gl	74.5
		II	6	chr06_1819184	chr06_1893743	25.08	72.20	3.86/gl	
		III	6	chr06_1819184	chr06_1893743	23.09	85.50	3.78/gl	
GL	*qgl3.1*	III	3	chr03_18337869	chr03_18418306	10.80	36.05	0.35/pr	80.4
	qgl7.1	III	7	chr07_21552051	chr07_22128152	8.92	40.80	0.26/pr	576.1

PVE % is the percent phenotypic variability explained by particular QTL

TGW thousand grain weight, *GL* grain length, *gl O. glumaepatula* allele, *pr* PR114 allele

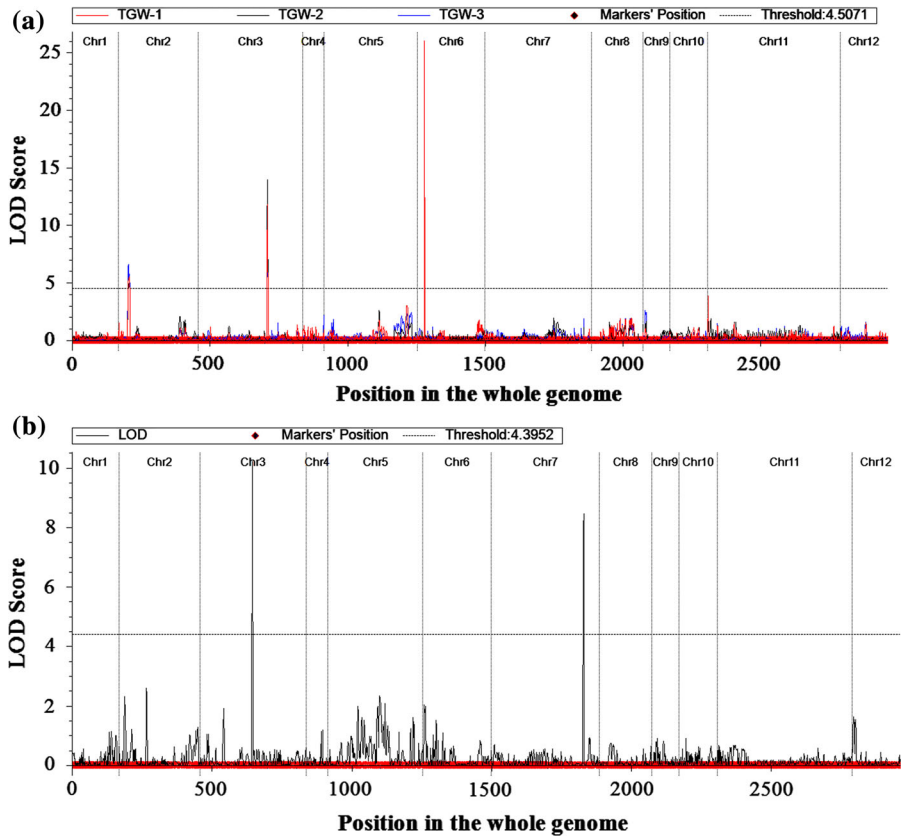

Fig. 5 LOD curves of QTL mapping in *glumae*-BILs for **a** TGW for all three seasons and **b** GL, on whole of genome. Horizontal dotted lines are the LOD thresholds. (Color figure online)

BC_2F_2 families of an *O. sativa* × *O. glumaepatula* interspecific cross (Brondani et al. 2002).

Placements of SNPs on the twelve chromosomes of rice based on their physical and genetic distances revealed uniform coverage across the genome, demonstrating the utility of GBS markers for the precise mapping of target traits. The observed genetic map length particularly in *glumae*-BILs seems to be quite different from original genetic maps developed using SSR markers. Since the physical positions of GBS SNP markers were already known, all the markers were given genetic position on the maps without ordering based on Kosambi mapping function in order to find the QTL regions. In additions, BILs generated using wild species of rice were not appropriate for generating the genetic maps due to segregation distortion. Uniform distribution of GBS SNPs has been observed in number of GBS studies such as wheat and barley (Poland et al. 2012), and rice (Spindel et al. 2013). In our study, SNPs could not be identified on some regions of chromosomes in both the BIL populations, resulting in large gaps in the chromosomes. The paucity of SNPs in these regions could be explained by the fact that BILs contain large percent of recurrent parent genome and differ at certain positions based on introgressed regions.

An important advantage of the BILs used in our study is the presence of a fairly uniform background between lines with few introgressions from the donor parents that enabled more precise estimates of quantitative traits. Based on the graphical genotypes generated with more than 3000 SNP markers in each population, more than 50% of BILs in the *rufi*-BILs and *glumae*-BILs comprise > 90% of the recurrent parent genome. The remaining BILs had 80–90% of the recurrent parent genome except for 3 *rufi*-BILs and 4 *glumae*-BILs. The pattern of introgressions in each BIL was not random, while few clusters of introgressions were identified in each BIL in both populations, which might be the result of the biased selection for the target traits during the generation of BILs.

Uncovering grain size QTLs in BILs

The present study uncovered a number of *O. rufipogon* genomic regions/QTL favorable for grain size traits. For example, GBS markers identified a major QTL (*qtgw5.1*) contributed by *O. rufipogon* on chromosome 5 consistently in all the seasons for increasing TGW spanning a narrow genomic region of 86 kb. Such precision is usually obtained after fine mapping with traditional marker systems. For example, Li et al. (2004) fine mapped *gw3.1* contributed by *O. rufipogon*, to a pericentromeric region of 93.8 kb on chromosome 3. Similarly, *gw8.1* and *gw9.1* were fine mapped to 308 and 37.4 kb genomic regions on rice chromosomes 8 and 9 respectively (Xie et al. 2006, 2008).

A major grain width QTL (*gw5.1*) in the current study responsible for increasing the grain width from 2.0 mm in the recurrent parent to 2.60 mm in *rufi*-BILs was identified near the thousand grain weight QTL locus *qtgw5.1*, indicating that an increase in GW might be responsible for high TGW. Another GW QTL *qgw5.2* responsible for decreasing the grain width was identified in the similar region. This QTL might be a minor QTL or an artifact, because none of the *rufi*-BILs showed significant decrease in grain width as compared to recurrent parent based on phenotypic variation. In earlier studies, two major QTL for grain size, *qsw5* (Shomura et al. 2008) and *GS5* (Li et al. 2011), on chromosome 5 have been cloned. Here, we identified a major QTL (*qtgw5.1*) for grain size, which is located 1.9 Mb from *GS5*, and 0.6 Mb from *qsw5* (data not shown), thus nullifying the possibility that the *O. rufipogon* QTL discovered here is allelic to either of these previously cloned QTL.

One major QTL for grain length (*qgl7.1*) was co-localized on chromosome 7 in both *rufi*-BILs and *glumae*-BILs. The recurrent parent allele at this locus was responsible for increasing the grain length, but decreasing the grain width. Recurrent parent PR114 has long and fine grain, and *qgl7.1* explained both the features.

In *glumae*-BILs, three QTL for TGW were detected on chromosome 2, 3 and 6 as compared to one major QTL in *rufi*-BILs. In earlier studies, major QTL for grain weight and grain size have been cloned such as *GW2* (Song et al. 2007) on chromosome 2, *GS3* (Fan et al. 2006) on chromosome 3 and *TGW6* (Ishimaru et al. 2013) on chromosome 6. The QTL *qtgw2.1* on chromosome 2 in the present study located at a physical distance of approximately 150 kb from *GW2* (data not shown). Another QTL *qtgw3.1* on chromosome 3 was found to be located towards distal end of the long arm of chromosome, whereas the most of the published QTLs for grain traits on chromosome 3 have been mapped near the centromere region (Huang et al.

2013). Similarity, *qtgw6.1* on chromosome 6 in the present study, does not localize in the *TGW6* genomic region (data not shown). Genotyping of *glumae*-BILs uncovered two QTLs for grain length as compared to one in *rufi*-BILs, where the recurrent parent allele was responsible for larger grain length. As mentioned earlier, the *qgl7.1* co-localizes in both populations, therefore confirming its location on chromosome 7 and further narrows down the location of the QTL to a 227 kb region based on comparison of genomic region in both the populations. A similar QTL, *qSS7*, was described in cross between Zhenshan97 and a Cypress chromosomal segment substitution line (Qiu et al. 2012), but it maps approximately 3 Mb from the *qgl7.1* in the present study (data not shown). All QTL identified in this study were shown to explain very high phenotypic variation. This can be explained on the fact that BILs have very high proportions of the recurrent parent genome, and differ only for few introgressed segments from wild species. All the phenotypic variation explained by the QTL is derived from these few wild introgressed segments without any other genetic noise. The *glumae*-BILs also had favourable phenotypic variation for days to 50% flowering. However, consistent QTL for days to 50% flowering could not be identified. This might be due to more biased selection of the BILs toward the recurrent parent, thereby leaving behind insufficient variability for mapping.

O. glumaepatula is a diploid wild species native to the Amazon forest and flooded areas of western Brazil. It is closely related to *O. sativa* (Buso et al. 1998) and is considered to be a potential source of useful genes of agronomic importance. *O. glumaepatula* has yet to be utilized extensively to uncover variation for yield and yield component traits, except for a few earlier reports. Brondani et al. (2002) studied 11 agronomic traits in BC_2F_2 families of the interspecific cross *O. sativa* × *O. glumaepatula* and identified several QTL for target traits; however most of the increasing alleles at identified QTL were contributed by cultivated parent.

O. rufipogon acc. IRGC104433 is from Thailand, while *O. glumaepatula* IRGC104387 is from Brazil. *O. rufipogon* and *O. glumaepatula* both carry the 'AA' genome similar to cultivated rice and may carry number of syntenic regions for different traits during evolutionary path. But QTL for thousand grain weight was mapped on chromosome 5 in *O. rufipogon* derived BILs and on chromosome 2, 3 and 6 in *O. glumaepatula* derived BILs. This is due to the fact that 'AA' genome *Oryza* germplasm exhibits remarkable eco-geographic differentiation worldwide, both regionally and locally. Thus, this wild germplasm can be expected to have significant adaptive gene differences among accessions (Vaughan et al. 2003). Different yield component QTLs identified in different wild species raised the attention to judiciously explore more number of wild species accessions present in the germplasm banks. Grain size QTL identified in the present study from low yielding wild relatives of rice reveals their significance in their ability to improve the yield of cultivated rice. Transferring these QTL in the background of cultivated rice will not only increase the yield, but also widen the genetic base of cultivated rice. Combining these yield component QTL in one genetic background will further enhance the magnitude of yield increase of cultivated rice, thus contributing towards food security.

Acknowledgements This work was supported by the Monsanto Beachell Borlaug International Scholarship Programme (MBBISP) to DB, which is gratefully acknowledged.

Author contributions DB, KS, RAW, AR designed and conducted the study, DB did field evaluation; DB, YY, DK prepared GBS library; SL, YY sequenced the GBS library; DB, KC analysed GBS data; DB did QTL mapping; DB, KS, RAW wrote and edited the manuscript.

References

Bhatia D, Wing RA, Singh K (2013) Genotyping by sequencing, its implications and benefits. Crop improv 40:101–111

Bhatia D, Joshi S, Das A, Vikal Y, Sahi GK, Neelam K, Kaur K, Singh K (2017) Introgression of yield component traits in rice (*Oryza sativa* ssp *indica*) through interspecific hybridization. Crop Sci 57:1–17

Brar D, Singh K (2011) *Oryza*. In: Kole C (ed) Wild crop relatives: genomics and breeding resources: cereals. Springer, Berlin, pp 321–365

Brondani C, Rangel P, Brondani R, Ferreira M (2002) QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa)* using microsatellite markers. Theor Appl Genet 104:1192–1203

Buso GSC, Rangel PH, Ferreira ME (1998) Analysis of genetic variability of South American wild rice populations (*Oryza glumaepatula*) with isozymes and RAPD markers. Mol Ecol 7:107–117

Chauhan JS (1998) Inheritance of grain weight, size and shape in rainfed rice (*Oryza sativa).* Indian J Agric Sci 68(1), http://

epubs.icar.org.in/ejournal/index.php/IJAgS/article/view/27174

Cheema K, Bains N, Mangat G, Das A, Brar D, Khush G, Singh K (2008) Introgression of quantitative trait loci for improved productivity from *Oryza rufipogon* into *O. sativa*. Euphytica 160:401–409

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147–1162

Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112:1164–1171

Fukuoka S, Nonoue Y, Yano M (2010) Germplasm enhancement by developing advanced plant materials from diverse rice accessions. Breed Sci 60:509–517

Gao Z-Y, Zhao S-C, He W-M, Guo L-B, Peng Y-L, Wang J-J, Guo X-S, Zhang X-M, Rao Y-C, Zhang C (2013) Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. Proc Nat Acad Sci 110:14492–14497

Harlan JR (1992) Crops and man, 2nd edition, Madison. Am Soc Agron, WI

Huang R, Jiang L, Zheng J, Wang T, Wang H, Huang Y, Hong Z (2013) Genetic bases of rice grain shape: so many genes, so little known. Trends Plant Sci 18:218–226

Imai I, Kimball JA, Conway B, Yeater KM, McCouch SR, McClung A (2013) Validation of yield-enhancing quantitative trait loci from a low-yielding wild ancestor of rice. Mol Breed 32:101–120

Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, Kashiwagi T, Ujiie K, B-i Shimizu, Onishi A (2013) Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. Nat Genet 45:707–711

Jacquemin J, Bhatia D, Singh K, Wing RA (2013) The International *Oryza* Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. Curr Opin Plant Biol 16:147–156

Jeuken MJW, Lindhout P (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. Theor Appl Genet 109:394–401

Jin F-X, Kim D-M, Ju H-G, Ahn S-N (2009) Mapping quantitative trait loci for awnness and yield component traits in isogenic lines derived from an *Oryza sativa*/*O. rufipogon* cross. J Crop Sci Biotechnol 12:9–15

Khush GS (2013) Strategies for increasing the yield potential of cereals: case of rice as an example. Plant Breed 132:433–436

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows Wheeler transform. Bioinformatics 25:1754–1760

Li J, Thomson M, McCouch SR (2004) Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. Genetics 168:2187–2195

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J (2011) Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. Nat Genet 43:1266–1269

McCouch SR, Sweeney M, Li J, Jiang H, Thomson M, Septiningsih E, Edwards J, Moncada P, Xiao J, Garris A (2007) Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. Euphytica 154:317–339

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M (2010) The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

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

Moncada P, Martinez CP, Borrero J, ChÃ¢tel M, Jr Gauch H, Guimaraes E, Tohme J, McCouch SR (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* x *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor Appl Genet 102:41–52

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucl Acids Res 8:4321–4326

Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5:92–102

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253

Qiu X, Gong R, Tan Y, Yu S (2012) Mapping and characterization of the major quantitative trait locus *qSS7* associated with increased length and decreased width of rice seeds. Theor Appl Genet 125:1717–1726

Rosyara UR, Gonzalez-Hernandez JL, Glover KD, Gedye KR, Stein JM (2009) Family-based mapping of quantitative trait loci in plant breeding populations with resistance to Fusarium head blight in wheat as an illustration. Theor Appl Genet 118:1617–1631

Septiningsih EM, Prasetiyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S, McCouch SR (2003) Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. Theor Appl Genet 107:1419–1432

Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nat Genet 40:1023–1028

Song X-J, Huang W, Shi M, Zhu M-Z, Lin H-X (2007) A QTL for rice grain width and weight encodes a previously

unknown RING-type E3 ubiquitin ligase. Nat Genet 39:623–630

Spindel J, Wright M, Chen C, Cobb J, Gage J, Harrington S, Lorieux M, Ahmadi N, McCouch S (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet 126:2699–2716

Tan L, Liu F, Xue W, Wang G, Ye S, Zhu Z, Fu Y, Wang X, Sun C (2007) Development of *Oryza rufipogon* and *O. sativa* introgression lines and assessment for yield related quantitative trait loci. J Integr Plant Biol 49:871–884

Thomson MJ, Tai TH, McClung AM, Lai XH, Hinga ME, Lobos KB, Xu Y, Martinez CP, McCouch SR (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor Appl Genet 107:479–493

Thomson MJ, Zhao K, Wright M, McNally KL, Rey J, Tung C-W, Reynolds A, Scheffler B, Eizenga G, McClung A (2012) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. Mol Breed 29:875–886

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

Wickneswari R, Bhuiyan MAR, Lim LS, Thomson MJ, Narimah MK, Abdullah MZ (2012) Identification and validation of quantitative trait loci for agronomic traits in advanced backcross breeding lines derived from *Oryza rufipogon* x *Oryza sativa* cultivar MR219. Plant Mol Biol Report 30:929–939

Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD, JiMing L, LongPing Y (1996) Genes from wild rice improve yield. Nat (Lond) 384:223–224

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909

Xie X, Song M-H, Jin F, Ahn S-N, Suh J-P, Hwang H-G, McCouch SR (2006) Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between *Oryza sativa* and *Oryza rufipogon*. Theor Appl Genet 113:885–894

Xie X, Jin F, Song M-H, Suh J-P, Hwang H-G, Kim Y-G, McCouch SR, Ahn S-N (2008) Fine mapping of a yield-enhancing QTL cluster associated with transgressive variation in an *Oryza sativa* x *O. rufipogon* cross. Theor Appl Genet 116:613–622

Zhao Q, Huang X, Lin Z, Han B (2010) SEG-Map: A novel software for genotype calling and genetic map construction from next-generation sequencing. Rice 3:98–102