UNDER PEER REVIEW

Original Research Article

Marker aided selection of yield-enhancing QTL *yld2.1* into restorer KMR3 and fine mapping it to a genomic region of 62 kb on chromosome 2

Abstract

Yield-enhancing QTLs are becoming major tools in plant breeding, their identification and characterization leads to understanding of genes and the physiological pathways involved in the complex trait, yield. In our previous study (Marri et al 2005), one major effect QTL (yld2.1) for yield and related traits was identified on chromosome 2 and was mapped to the interval between RM262 and RM263 (21 cM, ~5MB) using an advanced backcross test cross family derived from the cross between IR 58025A and O.rufipogon (IC 22015) accession. We aimed at fine mapping the yld 2.1 for identifying the gene/s responsible for increased yield in restorer line KMR3 background. 55 BC3F3 near-isogenic lines (NILs) representing 960 plants were used in fine mapping of yld2.1 with 8 polymorphic SSR markers. QTL analysis in BC3F3 population using single marker analysis, interval mapping and composite interval mapping for 9 yield and related traits narrowed the yld2.1 to a region of 62kb between RM3688 and RM3762 markers. Introgresssion lines with RM3688 and RM3762 have shown significant increase in yield per plant, grain weight per panicle, number of tillers per plant, panicle length and number of panicles. Their grain yield per plant was 17.6 – 20.72 % higher than that of ILs with rest of the regions and KMR3. 62kb region has 10 predicted genes consisting of 4 putative expressed genes including a fertility restorer A homologue like gene, 1 gene encoding 18S pre-ribosomal assembly protein and 5 genes encoding unknown expressed proteins. Significant increase in five important yield and related traits suggests that gene/s underlying this QTL should involve in multiple physiological pathways directly or through epigenetic interactions.

UNDER PEER REVIEW

Keywords: QTL, *Yld2.1*, KMR-3, Fine mapping, marker aided selection, Rice, IR 58025

Introduction

Increasing world rice (*Oryza sativa* L.) production is vital to meet dietary demands of the growing global population. Rice cultivars with higher yield potential must be developed to reduce the gap between yield potential and average farm yields (Peng et al., 1999). Oryza is an agronomically important genus containing spp with highly diverse morphological characteristics (Clayton and Renvoize 1986) included in this genus is cultivated rice (*Oryza sativa* L.) which constitutes the more than half of the worlds population, while other wild species of Oryza genome are potential reservoirs of useful genes. Common wild rice *O. rufipogon* is the wild ancestor of cultivated rice.

With the advent of DNA molecular markers, QTL mapping has become a routine strategy for the discovery of genes involved in complex quantitative traits. Thousands of QTL have been mapped for important agronomical traits in rice. Although primary mapping populations including F_2, recombinant inbred lines (RILs) and doubled haploid lines (DHs) have been widely used for QTL mapping in rice (Fan et al. 2005; Marri et al. 2005), QTL can only be localized to a genomic region (confidential region) rather than a locus in those populations. Following the primary mapping, advanced populations such as near isogenic lines (NIL) and chromosome segment substitution lines(CSSL) can be used to map QTL to a locus as a Mendalian factor by blocking the genetic background noise (Wang et al. 2004). Based on this strategy, several QTLs have been isolated in tomato (Fraryet al. 2000) and rice (Yano et al. 2000; Takahashi et al. 2001) in recent years. The common aspect in all the QTL cloning work is to exploit high quality NILs as advanced mapping populations.

UNDER PEER REVIEW

In the long run, development of high yield varieties is one of the most important goals in rice cultivation. Unfortunately, yield improvement efficiency is deemed to be very low due to its complex property affected by number of spikelets per panicle, 1,000-grain weight and tillers per plant. Of these factors, SPP was shown to be highly correlated with yield and acts as a crucial component in determining rice yield (Hua et al. 2002). Therefore, dissection of its genetic basis would be of great value in breeding high yield variety. So far, QTLs for SPP have been mapped in lots of populations, which were derived from inter-sub specific and intra specific crosses. Although many QTL controlling SPP were mapped in rice, only *Gn1a*, has been cloned. *Gn1a*, controlling grain productivity in rice, was elucidated to be a gene encoding cytokinin oxidase/dehydrogenase (*OsCKX2*), an enzyme that degrades the phytohormone cytokinin. Reduced expression of *OsCKX2* causes cytokinin accumulation in inflorescence meristems and increases the number of reproductive organs, resulting in enhanced grain yield (Ashikari et al. 2005). He et al 2007, constructed nearly isogenic lines (NILs) that differ only at a single quantitative trait locus (QTL) and fine-mapped the yield-improving QTL *qGY2–1* to a 102.9-kb region on rice chromosome 2. NILs identified the haplotypes of a leucine-rich repeat receptor kinase (*LRK*) gene cluster, the allelic variation of gene expression in hybrids at the *LRK* locus that associated with the yield QTL ranged from unequal expression of the two alleles simultaneously to expression of a single allele regardless of the parent-of-origin. They suggested allelic variations in structure and expression points to LRK gene cluster to be good candidate for the source of yield QTL.

In our previous studies, a major QTL responsible for major yield traits was mapped on chromosome2 flanked by SSR markers RM262 and RM263 (Marri et al. 2005). Marri et. al (2005) have used 251 BC2 test cross families and commercial hybrid KRH2 (IR58025A X

UNDER PEER REVIEW

KMR3) and evaluated them in a RBD (Randomized Block Design) trial. They have found 39 QTLs for different yield related traits, and two of them (*yld 2.1* and *yld 8.1*) were major effect QTLs. Both these yield QTLs were from *O.rufipogon*. The *yld2.1(* Interval mapping-LOD 35.33, R^2-50.47, additive effect- -238.51, composite interval mapping- LOD31.92, R238.46, Additive effect- -216.04) the *yld2.1* controlling an increase in number of panicles, grain number per plant, yield per plant, harvest index and plot yield respectively. In the present study, we selected *yld2.1* QTL & aimed to identify Near Isogenic Lines for sub-QTLs in the region *yld2.1* flanked by RM262 – RM263 (fig 1), followed by fine mapping of the candidate sub-QTL.

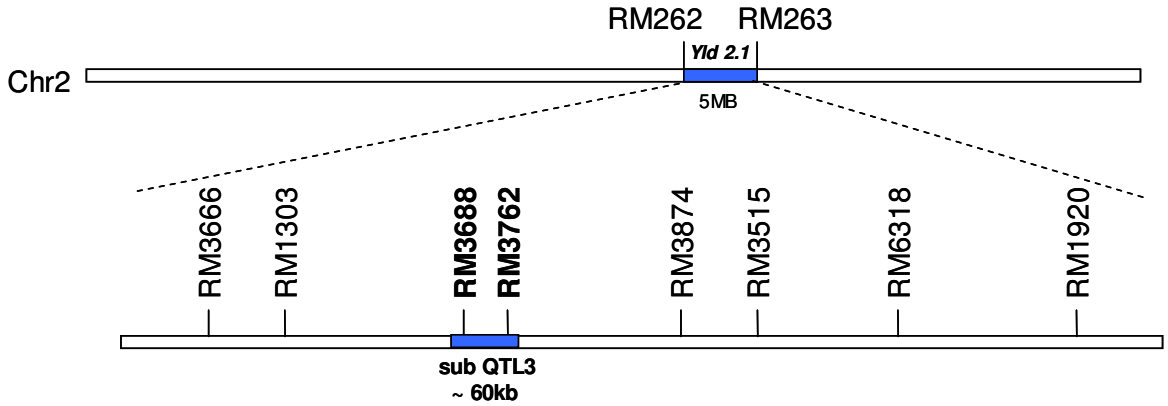


Fig 1: Depiction of *yld2.1* on chromosome 2.

Materials and Methods

Development of NILs:

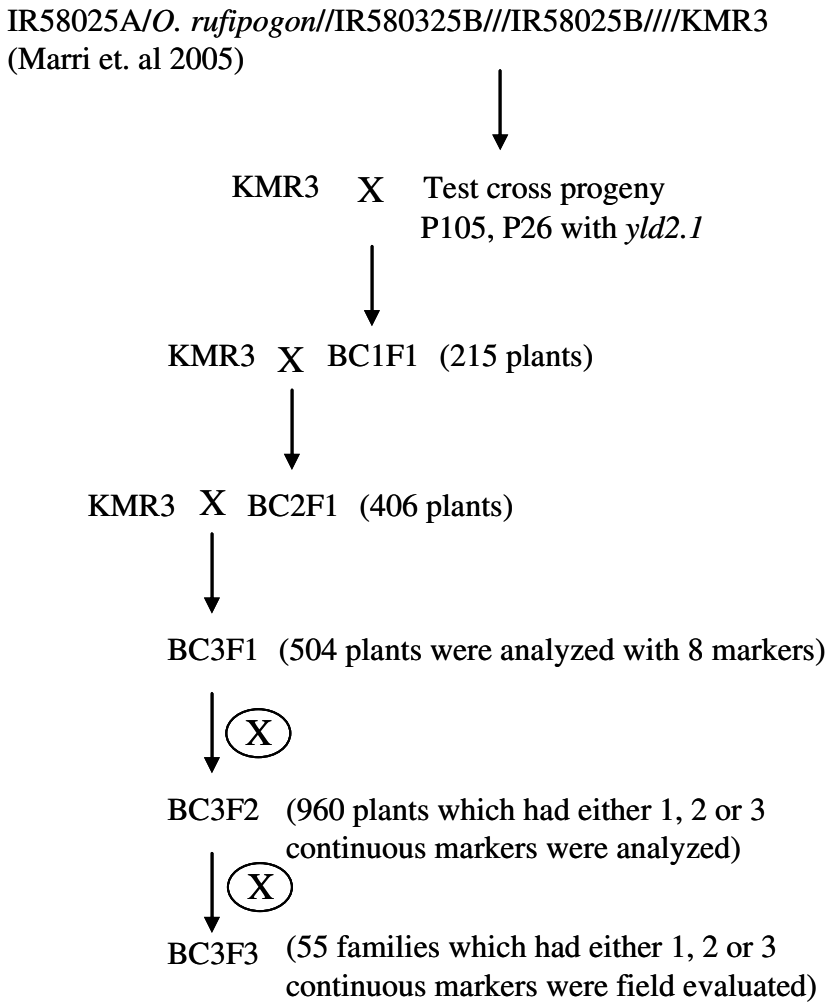
In our previous study, one QTL, *yld2.1*, was mapped to the interval between RM262 and RM263 on chromosome 2 using an advanced backcross test cross family derived from the cross between IR 58025A and *O.rufipogon*(IC 22015) accession, which was selected based on genetic diversity as the wild donor for this study. An accession of *O.rufipogon* which was genetically moderately

UNDER PEER REVIEW

87 distant from cultivar rice and IR58025A (a commercial CMS line) was used as a recurrent

88 parent, KMR3 a restoration line in popular hybrid KRH2 was used as a tester (Marri et al. 2005).

89



90

91

92 Fig 2: Flow diagram of the crosses made and number of plants selected for narrowing down into

93 QTL *yld2.1*.

94 Current study has started with IR58025A/*O. rufipogon*//IR580325B///IR58025B////KMR3

95 test cross progeny, we have selected plants P-105 and P-26 from the 251 BC2 testcross progeny

96 which were used for mapping. These two plants had QTL *yld2.1*, which was checked by

UNDER PEER REVIEW

presence of flanking markers RM262 and RM263. P-105 and P-26 were crossed with KMR3, this gave 215 BC1F1 plants. In this BC1F1 population genotypic selection was done using flanking markers for *yld2.1*. Selected BC1F1 plants were back crossed with KMR3, and 406 BC2F1 plants were obtained. Like in BC1F1, genotypic selection for desirable QTL *yld2.1* was done. Selected BC2F1 plants were back crossed again with KMR3 to obtain BC3F1 population of 504 plants. All 504 BC3F1 plants were screened with 8 polymorphic SSR markers lying within *yld 2.1* region and were selfed to obtain BC3F2 population. This BC3F2 population was used for small scale fine mapping of *yld2.1*. From 504 BC3F2 genotyping were resulted 55 plants of homozygous alleles for *O.rufipogon* for all the 8 markers lying within *yld2.1*. The 55 plants are field evaluated in RBD design for phenotype with own kind of checks and KMR3 as overall check.

DNA extraction and SSR analysis:

Fresh leaves of the individual from introgression lines were collected and ground in liquid nitrogen. DNA was extracted from the ground tissues by the CTAB (cetyltrimethyl ammonium bromide) method. A total volume of 10 µl reaction mixture was composed of 2 µl of 50 ng/µl template DNA, 10 mmol Tris–HCl (pH 9.0), 50 mmol KCl, 1.5 mmol MgCl2, 0.1% Triton X-100, 0.2 µM of each forward and reverse primer, 2.5 mM each of dNTP, and 1 U of Taq DNA polymerase (Bangalore Genei). The PCR amplification was performed under the following conditions: Initial denaturation at 94^0C for 5 min, followed by 35 cycles of denaturation at 94^0C for 1 min, annealing at 55^0C for 30 sec, extension at 72^0C for 2 min, followed by the final extension 72^0C for 5 min. Following amplification, the products were checked for polymorphism or marker segregation on an agarose gel and scored for the segregating bands.

Phenotypic evaluation of BC_3F_2 population:

UNDER PEER REVIEW

The 8000 BC_3F_2 plants derived from 400 BC_3F_1 families were grown in an experimental plot at DRR (Directorate of Rice Research) Hyderabad, India, during dry season in 2008. They were planted with a distance of 15 cm X 20 cm for a line X row. From this population we have selected 37 families (each family consist of 24 plants and in some families 48 plants, approximate total of 960 plants) which contained either only 1, 2 or 3 consecutive markers. This selection was to minimize the *O.rufipogon* chromosomal segment, and in future to fine map the specific gene/s contributing to yield increase. Morphological and phenotypic selection was done in the selected 960 plant population, out of this 55 plants showed homozygous *O.rufipogon* alleles at corresponding marker.

The homozygous population of 55 plants, representing all the eight markers was field evaluated in Kharif'2008 at DRR. Yield traits measurements were obtained using a complete randomized block design, with each block having three rows (24 plants per row) and in two replications.

Measuring yield related traits

Leaf length (LL) is from the bottom to the tip of the flag leaf in cm, leaf width (LW) is width of the middle portion of the flag leaf in cm, plant height (PH) is from bottom to the tip of the longest panicle in cm, tiller number per plant (NT) is total number of tillers per plant, number of panicles (NP) is total number of panicles in one plant, panicle length (PL) is from neck to last spikelet of main panicle in cm, 50% flowering (50%F) is no. of days taken for 50% flowering in that family, grain weight per panicle (GWP) is grain weight in the main panicle in grams and yield per plant (YLDP) is total weight of filled grains per plant in grams.

Results and Discussions:

Genotypic analysis of 504 BC3F1 plants with 8 SSR primers lying within *yld2.1* region:

UNDER PEER REVIEW

All 504 BC3F1 plants were tested with 8 polymorphic SSR markers for the presence of *O.rufipogon* alleles within the QTL *yld2.1* (fig. 3). In these 8 markers, RM3762 shown highest introgression of 46.63%, followed by RM1920 with 41.07% and lowest was RM3666 with 7.54% in 504 BC3F1 plants. Table 2 shows the no. of plants for each marker with *O. rufipogon* allele at that marker and % distribution of corresponding marker in 504 plants. As shown in table 3, out of 504 BC3F1 plants 402 plants had at least one of these 8 markers. Out of these 402 plants, 6 plants had all the 8 markers, 5 plants had 7 markers, 28 plants had 6 markers, 41 plants had 5 markers, 53 plants had 4 markers, 84 plants had 3 markers, 79 plants had 2 markers and 106 plants only one marker. For fine mapping a QTL, plants with shortest introgression would be beneficial, for this reason we have selected plants with only 1, 2 or 3 consecutive markers having *O. rufipogon* allele. These plants were devoid of *O. rufipogon* at rest of the markers and had *O. sativa*. There were 36 plants which had only 1, 2 or 3 consecutive markers for *O.rufipogon* allele. These 36 plants were continued for selfing along with rest of the population. This selection was to minimize the *O.rufipogon* chromosomal segment, so that we could reach our goal of narrowing down into *QTL yld2.1* and in future to fine map the specific gene/s contributing to yield increase.

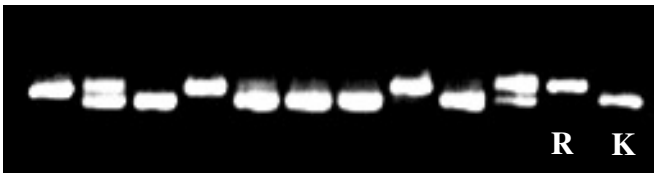


Fig 3: Segregation of RM 3688 in BC3F1 introgression lines. The extreme right lanes show *O.rufipogon* (R) and KMR3 (K) alleles respectively

Genetic and phenotypic analysis of BC3F2 population:

From the 269 BC3F2 lines, 960 plants from 36 lines of BC3F2, which have 1, 2 or 3 consecutive markers, were tested for their homozygosis of their respective markers.

UNDER PEER REVIEW

55 plants were homozygous *O.rufipogon* segments for one of the 8 markers. Out of 55 plants one plant had RM 3666, 8 plants had RM1303, 4 plants had RM3688, 2 plants had RM3762, 9 plants had both the RM3688 and RM 3762, 2 plants had RM3874, 7 plants had both the RM 3874 and RM 3515, 3 plants had RM 3515, 17 plants had RM 6318, one plant had both the RM6318 and RM 1920 and one plant had RM 1920 markers Respectively

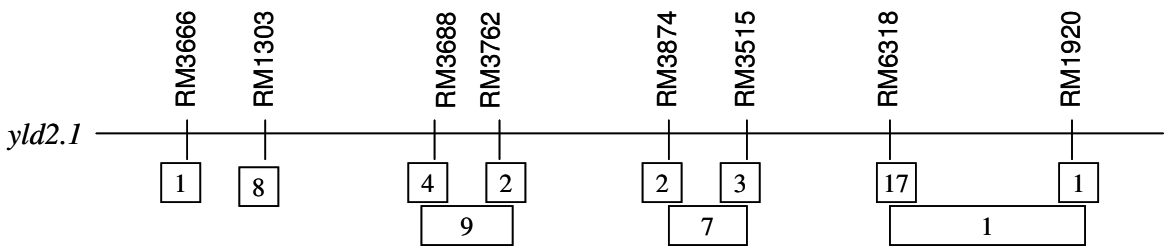


Figure 4: No. of BC3F2 plants with homozygous *O. rufipogon* allele at corresponding marker/s.

Morphological data of these 55 homozygous BC3F2 plants grown in field was recorded and BC_3F_2 plants indicated that 9 plants with *O.rufipogon* allele at 3rd sub QTL (RM 3688 and RM 3762) showed transgressive segregation for grain weight per plant, tiller number per plant. Fertility restoration was tested using IR58025A as CMS line and the KMR3 plants having one of the 8 markers individually. The results on hybrids indicated that plants with 3rd sub-QTL restored fertility to a greater extent compared to the plants with other sub QTLs.

In 55 BC3F2 homozygus plants Sub-QTLs were analyzed for single marker analysis using QTL cartographer (v2.5). For analyzing marker-QTL association for grain yield per plant and determining the precise location of the sub QTLs, A LOD score of 2.5 was used as the threshold for detecting the location of QTLs and 1000 permutations. LOD peaks for significant sub QTLs

UNDER PEER REVIEW

were used to position the QTL on the linkage map. 55 homozygous plants with different sub QTLs indicated a strong association of third sub-QTL with positive transgressive variation for yield per plant (table 4).

Evaluation of BC3F3 population for yield and related traits:

The homozygous population of 55 lines constituting 960 plants, representing all the eight markers was field evaluated in Kharif'2008 at DRR. Yield traits measurements were obtained using a complete randomized block design, with each block having three rows (24 plants per row) and in two replications. A total of nine yield and related traits were measured in this population. Fig.5 shows the data for six traits, number of tillers per plant (nt), number of panicles per plant (np), yield per plant (yldp), grain weight per panicle (gwp), plant height (ph) and panicle length (pl) in plants with corresponding markers. Average of total number of plants for a corresponding marker was used in this analysis. Except for plant height, rest of the five traits showed higher values in plants containing either RM3688 or RM3672 and in plants with both of them.

UNDER PEER REVIEW

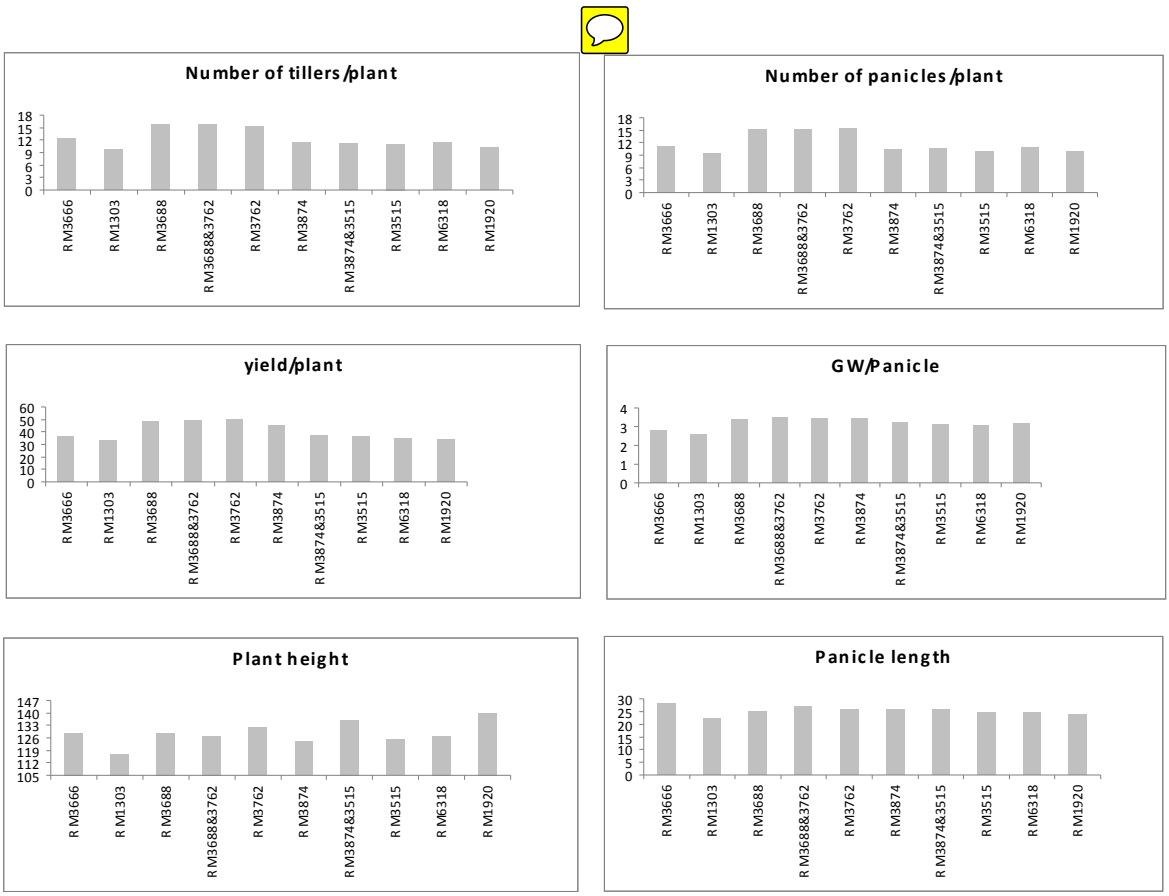


Fig 5: Graphs showing data for 8 yield and related traits in BC3F3 population.

In 55 BC3F3 homozygus population Sub-QTLs were again analyzed using single marker analysis, interval mapping (IM) and composite interval mapping (CIM) using QTL cartographer (v2.5). For analyzing marker-QTL association for 9 trait and determining the precise location of the sub QTLs, interval mapping and composite interval mapping of the QTL cartographer version 2.5 (Botstein et al,) was used. A LOD score of 2.5 was used as the threshold for detecting the location of QTLs and 1000 permutations. LOD peaks for significant sub QTLs were used to position the QTL on the linkage map.

Single marker analysis for nine traits was performed with yield data from BC3F3 population with 8 SSR markers shown in table 5. Seven traits were linked with different significance values to different markers. Markers RM3688 and RM3762 were linked to traits NT, NP, GWP and

UNDER PEER REVIEW

YLDP with highest significance and all these traits have come from *O. rufipogon*. RM6318 was linked to LL, NT, NP and YLDP, but all of these were contribution from *O. sativa*. For each trait at least 3 markers were associated with varying significance.

Table 6 shows the result of QTL analysis using interval mapping (IM) and composite interval mapping (CIM). In this analysis, nine traits were analyzed for QTL mapping. *O. rufipogon* has shown allelic effect on NT, NP, PL, GWP and YLDP in both IM and CIM.

NT and YLDP had shown higher LOD scores and phenotypic variance in both IM and CIM for the region RM1303 – RM3874. RM3666 – RM3874 was the QTL region for NP and GWP and RM3762 – RM3874 was the QTL region for PL. Figure 6 shows the peaks for LOD scores for all the nine traits, where sharpest peak area for five traits (NT, NP, PL, GWP and YLDP) was in the region of RM3688 – RM3762.

UNDER PEER REVIEW

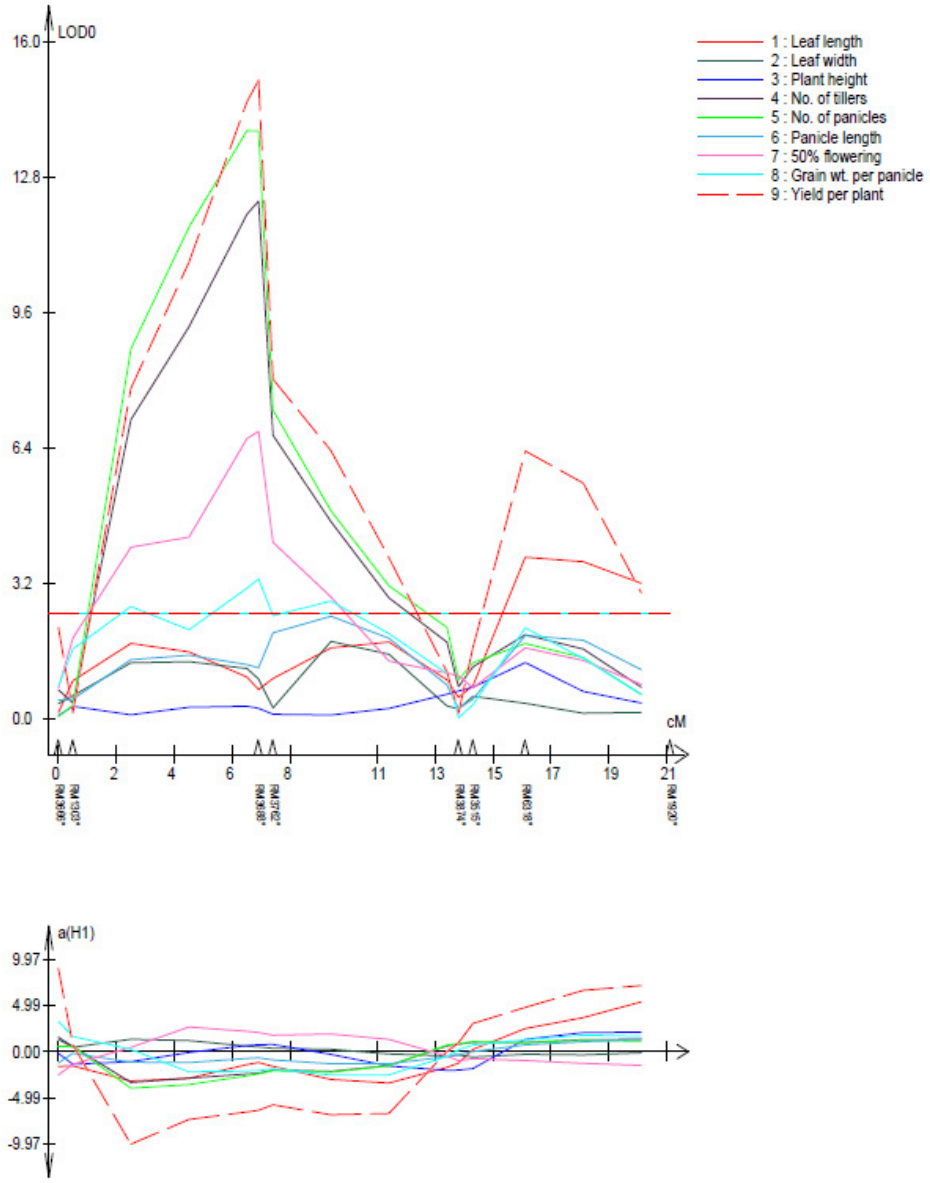


Fig 6. QTL analysis: graph showing additive effect for 9 traits and corresponding markers

QTL mapping provides the genomic region responsible for desirable trait and the size of this region depends on the resolution of mapping. For characterization of candidate gene/s responsible for desirable traits requires high resolution mapping of the region provided by QTL mapping. In this case, a major-effect QTL *yld2.1* for yield and related traits was previously

UNDER PEER REVIEW

mapped by our group (Marri et al 2005), QTL between markers RM262 and RM263 which are about 21cM (5MB) on rice chromosome 2. This region has about 1000 genes, to identify the candidate gene/s for increase in yield and related traits there is need for a mapping population which can give recombinants within this region and further delineate the region to a smaller region to pinpoint a candidate gene/s. In the present study, we selected *yld2.1* QTL for further narrowing down the genomic region & aimed to identify Near Isogenic Lines in the region *yld2.1* flanked by RM262 – RM263, followed by fine mapping of the candidate sub-QTL.

The *yld2.1* had LOD of 35.33 and additive effect of -238.51 in interval mapping, LOD of 31.92 and additive effect of -216.04 in composite interval mapping. Higher LOD score and additive effect for *yld2.1* denotes it's a major effect QTL and fine mapping would provide the key gene/s controlling yield and related traits. According to Marri et al 2005 QTLs *snp2.1*, *np2.1*, *nt2.1*, *gnp2.1*, *gw2.3*, *yldp2.1*, *gnp2.2*, *yldp2.2* and *hi2.1* were mapped in between RM262 and RM263; it is possible that all these traits are controlled by different genes or a single gene which has pleiotropic effect on multiple traits. For any fine mapping a high quality mapping population is essential, in our case NILs (near isogenic lines) were derived from IR58025A/*O. rufipogon*//IR580325B///IR58025B////KMR3 (Marri et. al 2005). Where 2 plants containing *yld2.1* QTL were crossed with KMR3 (a restorer line) and continued back crossing with KMR3 three more times to generate a BC3F3 population. This crossing program was devised to transfer *yld2.1* to restorer line KMR3 and also for fine mapping.

We have used substitution mapping for identifying the region responsible for increase in yield traits. In substitution mapping, the genotypes and trait phenotypes of individual sub-NILs are compared with recipient parent to identify sub-NILs containing genetic factors contributing to specific traits. By comparing sub-NILs with overlapping introgressed regions, chromosome

UNDER PEER REVIEW

253 intervals likely to contain relevant genes can be identified and precisely located. Substitution
254 mapping has been very successful in identifying genes with large phenotypic effects in different
255 species, including tomato (Fridman et al. 2002).
256 BC2F2 and BC2F3 introgression lines derived from *Oryza rufipogon* into an *indica* cultivar
257 Guichao 2 were used in QTL analysis (Tian et al 2006) and QTL *gpa7* was mapped to short arm
258 of chromosome 7 affecting grain number per panicle. Tian et al., (2006) have used two-step
259 substitution mapping for narrowing down *gpa7* to a region of 35kb that contained five predicted
260 genes. Like in our case, *gpa7* QTL region was mapped to multiple panicle traits (length of
261 panicle, primary branches per panicle, secondary branches per panicle, grains on primary
262 branches and grains on secondary branches). Using NILs developed from Jefferson (*O.sativa*) X
263 *O. rufipogon* (IRGC105491), Li et al (2004) fine mapped *gw3.1* QTL to a region of 93.8-kb
264 region from initial region of 31.8 cM. Brouwer et al (2004) used NILs for fine mapping three
265 QTLs (*lb4, lb5b and lb 11b*) for late blight resistance spanning 6.9-cM, 8.8-cM and 15.1-cM on
266 chromosomes 4, 5 and 11 respectively. Using NILs they could limit the three QTLs to smaller
267 regions on the chromosome, which would facilitate the MAS of late blight resistance more easy
268 and precise. Liu et al (2006) fine mapped QTL *SPP1* (spikelets per panicle) on chromosome 1 to
269 a region of 107 kb using NILs. Wan et al (2006) have narrowed down GL-3a QTL controlling
270 grain length to a region of 87.5 kb on chromosome 3 using chromosome segment substitution
271 lines grown under multiple environmental conditions and 2 different genetic backgrounds. Song
272 et al (2007) cloned and characterized the QTL *GW2* responsible for grain width and weight, their
273 fine mapping yielded the RING-type E3 ubiquitin ligase gene as responsible for the trait. This
274 gene negatively controlled the cell division by affecting proteasomes regulated proteolysis. Xie
275 et al 2006 fine mapped a yield-enhancing QTL cluster from *O. sativa* X *O. rufipogon* to a region

UNDER PEER REVIEW

of 37.4 kb on chromosome 9 using a series of BC3F4 NILs. In their case all seven QTLs were additive and came from low yielding *O.rufipogon* (IRGC 105491) parent. They have shown that NILs containing homozygous *O. rufipogon* introgression in the target region have out-yielded NILs containing *O.sativa* allele by 14.2 -17.7 %. They suggested the possibility of a single, pleiotropic gene affecting their 7 traits of interest, which might work as a major regulator of plant development.

Yield trials using BC3F3 substitution lines confirmed that NILs containing *O. rufipogon* DNA in the target region significantly out-yielded NILs with KMR3 DNA in the same region as well as parental KMR3 controls. The grain yield per plant in the *O. rufipogon* NILs was 17.6 – 20.72 % higher than that of the corresponding KMR3 NILs and KMR3 parent (fig. 7). As shown in figure 6, IL50-7 and IL50-13 (which have region RM3688 – RM3762) have shown better performance for the five listed yield and related traits when compared to both parent KMR3 and plants with other markers. Figure 7 shows the IL50-7 and KMR3 plants, introgression line with RM3688-RM3762 region had more tillers when compared to parent KMR3. This result confirmed that the yield-component QTLs identified in primary QTL analysis (BC3F2) contributed directly to grain yield under field conditions. Our study also confirmed earlier reports about transgressive variation for yield and other traits coming from the *O. rufipogon* parent (Xiao et al. 1998; Moncada et al. 2001; Thomson et al. 2003; Septiningsih et al. 2003, Xie et al. 2008) because alleles from *O.rufipogon* enhanced the yield of the restorer line KMR3. In our mapping RM3688 – RM3762 region also show significance in SMA (single marker analysis), IM and CIM for multiple traits (NT, NP, PL, GWP and YLDP). This indicates RM3688 – RM3762 region has a gene/s that might play an important role in regulation of yield and related traits.

UNDER PEER REVIEW

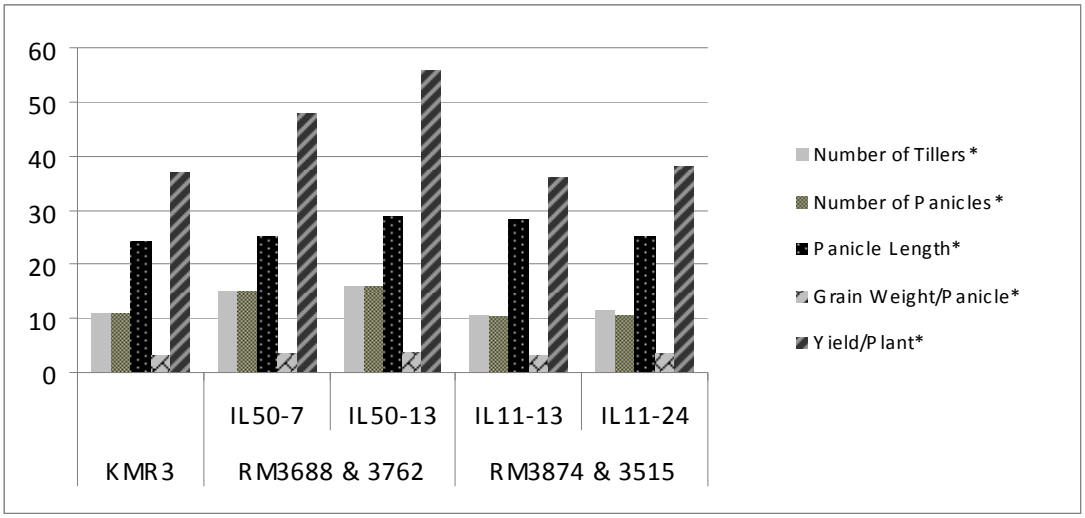


Fig 7: Comparison of five linked traits in plants with RM3688-RM3762, RM3874-RM3515 along with parent KMR3.

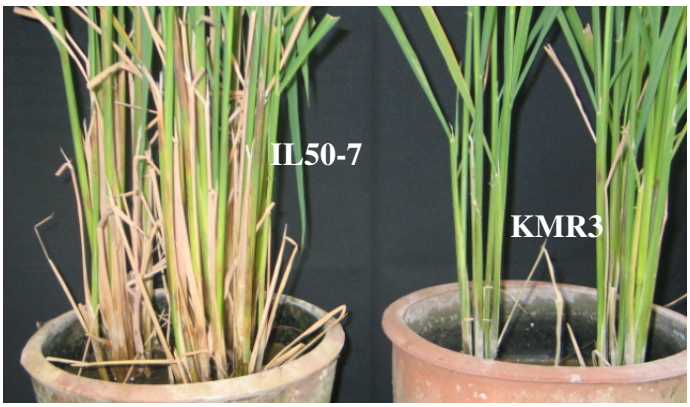


Figure 8: Increased no. of tillers in IL50-7 (RM3688-RM3762) in comparison with KMR3.

QTL analysis by SMA, IM and CIM performed on BC3F3 yield data with 8 SSR markers. SMA has shown minimum of 3 markers for each of the nine traits and as described in results section five traits were derived from *O. rufipogon*. Table 7 summarizes the SMA result, where in NT was linked to RM3688 and RM3762 with high significance and came from *O.rufipogon*, where as RM3874, RM3515 and RM6318 were from *O.sativa*. For most of the traits both RM3688 and RM3762 linked very strongly, which could be an evidence for presence of candidate gene/s within this region. RM6318 was also linked to LL, NT, NP, PL and YLDP, but in this case all of

UNDER PEER REVIEW

them have derived from *O. sativa*. For 50% flowering RM3688 and RM3762 were linked strongly and they came from *O. sativa*, this is probably because cultivated varieties are early flowering when compared wild relatives.

LL – Leaf length, LW – Leaf width, PH – Plant height, NT – No. of tillers, PL – Panicle length, 50%F – 50% flowering, GWP – Grain weight per panicle, YLDP – Yield per plant

Significance pr(F): * = 5%, ** = 1%, *** = 0.1%, **** = 0.01%

S – *Oryza sativa*, R – *Oryza rufipogon*

Interval mapping and composite interval mapping performed has identified QTLs for all nine traits (table 6). Out of the five traits which have come from *O. rufipogon* in IM, RM1303-RM3874 region had shown phenotypic variance of 69% for YLDP and 59% for NT. RM3666-RM3874 region had shown phenotypic variance of 65% for NP and 24% for GWP. RM3762-RM3874 region also showed phenotypic variance of 23% for PL. Above results are similar in the case of CIM as well. As shown in the figure 6, sharpest region in peaks for all the five traits were in the region of RM3688-RM3762. This indicates this region to be the candidate for the increase in yield seen in original *yld2.1* QTL. Along with QTL analysis, yield data in the field and physical observation of plants also points us to the region of RM3688-RM3762 as candidate for the *yld2.1*. This study has narrowed down the yld2.1 QTL from a region of 5MB to a region of RM3688-RM3762 (fig .1).

Genomic region between RM3688-RM3762 on chromosome 2 in reference sequence *O. sativa* cv. Japonica Nipponbare has about 62kb DNA. There are ten genes/ORFs in the 62kb region (table 8), five genes encode for unknown expressed proteins and 5 genes code for putative expressed proteins. Five putative genes shown similarity to ASC1-like protein, esterase ytx M,

UNDER PEER REVIEW

333 vegetative cell protein gp1 precursor, 18S pre-ribosomal assembly protein and fertility restorer
334 homologue A for orthologous proteins from other species. The sequence annotation using only
335 sequence data of cv. Nipponbare in this study would make it difficult to predict a candidate gene
336 underlying *yld2.1*, which was derived from the *O. rufipogon*. This is mainly due to divergence in
337 sequence level between the cultivated rice and its wild counterpart (Ammiraju et al. 2006).
338 Ammiraju et al. (2006) revealed that the orthologous region of the *Adh1* gene in the *O. sativa*
339 genome was 9.4– 28% larger relative to four wild species including *O. rufipogon* and this size
340 variation in the *Adh1* gene was mainly due to insertion of transposable elements as well as
341 multiple genetic mechanisms. Thus, the candidate gene in our study should be confirmed based
342 on the complete sequence of *O. rufipogon* in the target region and we are sequencing the target
343 region of *O. rufipogon* to determine the gene underlying *yld2.1* in this regard.
344 Most of the studies reporting yield QTLs and their characterization involved regular cultivars as
345 recipients, and their donor as either another high yielding cultivar or wild relatives of rice like *O.*
346 *rufipogon*. Where as current study aims to improve the yield of hybrids using introgression of
347 yield-enhancing QTLs into the parents of the hybrid. We foresee the future with improved
348 restorer (KMR3) lines as parents in developing high yielding rice hybrids. We conclude that the
349 region flanked by RM 3688 and RM 3762 is important in contributing to high yield. Thus the
350 yield enhancing QTL *yld2.1* was narrowed down from 5Mb to 62Kb. This genomic region is
351 being fine mapped further to identify candidate genes for use in MAS or functional analysis.
352
353 **Reference**:
354 Ammiraju J. S., Luo M., Goicoechea J. L., Wang W., Kudrna D., et al. , 2006. The Oryza
355 bacterial artificial chromosome library resource: construction and analysis of 12 deep-

UNDER PEER REVIEW

coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza. Genome Res. 16: 140–147.

Ashikari M, Sakakibara H, Lin S, Ymamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M, Cytokinin Oxidase regulates rice grain production. Science 309: (2005) 741-745.

Botstein D, White RL, Skolnick M, Davis RW, Construction of a genetic linkage map in man using restriction fragment length polymorphism. American Journal of human Genetics 32: (1980) 324-331.

Brouwer D. J., St.Clair D. A., 2004. Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. Theor. Appl. Genet. 108: 628–638

Clayton WD, Phillips SM, Renvoize SA, Gramineae Part II. In: Milne-Redhead E, Polhill RM, eds. Flora of Tropical East Africa. UK: Crown Agents for Oversea Governments and Administrations (1974).

Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Zhang Q, *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(6): (2006) 1164-1171.

Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD, *fw2.2:* a quantitative trait locus key to the evaluation of tomato fruit size. Science 289: (2000) 85-88.

Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D, Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science305: (2004) 1786–1789.

UNDER PEER REVIEW

He G, Luo X, Tian F, Li K, Zhu Z, Su W, Haplotype variation in structure and expression of a gene cluster associated with a quantitative trait locus for improved yield in rice. Genome Research 16: (2007) 618–26.

Hua, J.P., Xing, Y.Z., Xu, C.G., Sun, X.L., Yu, S.B. and Zhang, Q.F. (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. *Genetics*, **162**, 1885–1895.

Li JM, Xiao JH, Grandillo S, Jiang LY, Wan YZ, Deng QY, Yuan LP, Mc Couch SR, QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa*) and African (*O. glaberrima*) rice. Genome 47: (2004) 697-704.

Liu QQ, Li QF, Cai XL, Wang HM, Tang SZ, Yu HX, Molecular marker-assisted selection for improved cooking and eating quality of two elite parents of hybrid rice. Crop Sci46: (2006)2354–2360.

Liu QQ, Li QF, Cai XL, Wang HM, Tang SZ, Yu HX, Molecular marker-assisted selection for improved cooking and eating quality of two elite parents of hybrid rice. Crop Sci46: (2006)2354–2360.

Marri PR, N Sarla, VLN Reddy and EA Siddiq, Identification and mapping of yield and related QTLs from an Indian accession of *Oryza rufipogon*, BMC Genet.6: (2005)33.

Moncada P, Martinez CP, Borrero J, Chatel M, Gauch HJ, Guimaraes E, Tohme J, McCouch SR, Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theoretical and Applied Genetics 102: (2001) 41-52.

UNDER PEER REVIEW

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica, Green revolution genes encode mutant gibberellin response modulators. Nature 400: (1999) 256–261.

Septiningsih EM, Prasetiyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S, McCouch SR, 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*. Theoretical and Applied Genetics 107: (2003b) 1419-1432.

Song XJ, Huang W, Shi M, Zhu MZ, Lin HX A QTL for rice grain width and weight encodes a previously unknown RINGtype E3 ubiquitin ligase. Nat Genet 39: (2007)623–630.

Takahashi Y, Shomura A, Sasaki T, Yano M Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. Proc. Natl. Acad. Sci. USA 98: (2001) 7922–7927.

Thomson MJ, Tai TH, McClung AM, Lai XH, Hinga ME, Lobos KB, Xu Y, Martinez, CP, McCouch SR, 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. Theoretical and Applied Genetics 107: (2003)479-493.

Tian F, Li DJ, Fu Q, Zhu ZF, Fu YC, Wang XK, Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield related traits. Theoretical and Applied Genetics 112: (2006)570–80.

UNDER PEER REVIEW

Wan, X. Y., J. M. Wan, L. Jiang, J. K. Wang, H. Q. Zhai *et al.*, 2006. QTL analysis for rice grain length and fine mapping of an identified QTL with stable and major effects. Theor. Appl. Genet. 112 1258–1270.

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

Xie X, Song MH, Jin F, Ahn SN, Suh JP, Hwang HG, 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*. Theoretical and Applied Genetics 113: (2006) 885–94.

Xu K, Deb R, and Mackill DJ, A microsatellite marker and a codominant PCR-based marker for marker-assisted selection of submergence tolerance in rice. Crop Sci. 44: (2004)248–253.

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

Marker	Forward primer	Reverse primer
RM3666	TGATTTTCAGGGCTGTAGGG	AGTAAAATGCTCCCCATGGC
RM1303	CTGATCTTGGTGAGCGAGTG	TACGGATCAGCACTCAGCAC
RM3688	GTTGAATCAAGCTGTGCAGC	AGCTAGGCAAAGCATGCATG
RM3762	TACCGTAAGGCGCTGGATT	ACGAGGTCCCCCCTCTAAA
RM3874	TGGGTGATCTTAGTTTGGCC	AATGTGCCTGCACATGTCAC
RM3515	GGAAAGAAGATATGCCATGC	AGAGAGAATCAGAAACACCAAC
RM6318	TGCTGCTTCTGTCCAGTGAG	GGATCATAACAAGTGCCTCG

UNDER PEER REVIEW

RM1920	CAAACACAGTGTTGACAGAA	GCTATTGACTTATCCGTTCA

Table1: List of SSR markers used in this study and their sequence.

Marker	RM3666	RM1303	RM3688	RM3762	RM3874	RM3515	RM6318	RM1920
Plants having *O.rufipogon*	38	170	93	235	130	147	164	207
% distribution of marker	7.54	33.73	18.45	46.63	25.79	29.17	32.54	41.07

Table 2: Percentage introgression of *O. rufipogon* allele in 504 BC3F1 plants.

No. of Markers present in the plant	All 8	Any 7	Any 6	Any 5	Any 4	Any 3	Any 2	Any 1	None
No. of Plants	6	5	28	41	53	84	79	106	102
Plants with consecutive markers						8	17	11	

Table 3.Frequency distribution of *O. rufipogon* allele in 504 BC3F1 plants.

Marker	Additive effect	Likelihood ratio	F-calculated	F-crit	Significance
RM3666	-9.410	1.648	1.610	0.210	
RM1303	1.343	0.235	0.227	0.636	
RM3688	-2.953	1.621	1.584	0.214	
RM3762	-6.436	7.213	7.435	0.009	**
RM3874	-1.930	0.424	0.410	0.525	
RM3515	-4.376	2.771	2.737	0.104	
RM6318	5.341	6.859	7.047	0.011	*
RM1920	-1.982	0.141	0.136	0.714	

UNDER PEER REVIEW

Table 4: Result of Single marker analysis in BC3F2 population for yield per plant (YLDP).

Trait	Marker	Additive effect	Likelihood ratio	F-calculated	F-crit	Significance
Leaf length	RM3762	29.147	-1.581	4.372	4.385	0.041*
	RM6318	27.507	2.376	15.239	16.92	0.000 ***
	RM1920	24.374	4.124	6.653	6.815	0.012*
Number of tillers	RM3688	13.62	-2.225	53.397	86.93	0.000 ****
	RM3762	13.707	-1.966	30.867	39.9	0.000 ****
	RM3515	11.92	0.927	4.741	4.772	0.033 *
	RM6318	12.252	0.946	8.251	8.579	0.005 **
Number of panicles	RM3688	13.02	-2.293	56.739	95.698	0.000 ****
	RM3762	13.127	-2.055	33.629	44.682	0.000 ****
	RM3874	11.307	0.874	4.035	4.034	0.050 *
	RM3515	11.237	1.005	5.471	5.543	0.022 *
	RM6318	11.625	0.926	7.674	7.935	0.007 **
Panicle length	RM3688	25.787	-0.644	5.371	5.437	0.024 *
	RM3762	26.003	-0.887	9.345	9.815	0.003 **
	RM3515	25.262	0.717	8.478	8.833	0.004 **
50% Flowering	RM1303	90.323	-1.271	7.621	7.877	0.007 **
	RM3688	88.506	2.006	30.087	38.591	0.000 ****
	RM3762	88.42	1.784	19.229	22.182	0.000 ****
Grain weight/panicle	RM1303	31.151	1.548	6.658	6.82	0.012 *
	RM3688	33.108	-1.923	14.357	15.808	0.000 ***
	RM3762	33.261	-1.83	11.2	11.97	0.001 **
Yield/plant	RM3688	43.184	-6.263	59.479	103.292	0.000 ****

UNDER PEER REVIEW

	RM3762	43.545	-5.727	36.967	50.796	0.000 ****
	RM3515	38.538	2.401	4.25	4.258	0.044 *
	RM6318	38.952	3.977	22.094	26.202	0.000 ****
	RM1920	35.245	5.245	5.181	5.235	0.026 *

Table 5: Result of Single marker analysis in BC3F3 population.

Trait	Chr.	Marker interval	Allelic effect	Interval mapping			Composite interval mapping		
				LOD	R2	Additive	LOD	R2	Additive
Leaf length	2	RM3874-RM1920	*O.sativa*	3.8	27	2.4	3.8	25	2.5
Leaf width	2	RM3762-RM3874	*O.sativa*	1.8	1	0.3	1.8	2	0.3
Plant height	2	RM3762-RM1920	*O.sativa*	1.3	4	1.2	1.3	4	1.4
Number of tillers	2	RM1303-RM3874	*O.rufipogon*	12.3	59	-2.4	12.2	62	-2.4
Number of panicles	2	RM3666-RM3874	*O.rufipogon*	13.7	65	-2.1	13.8	72	-2.2
Panicle length	2	RM3762-RM3874	*O.rufipogon*	2.4	23	-1.4	2.4	23	-1.3
50% flowering	2	RM3666-RM3515	*O.sativa*	6.8	40	1.8	6.8	44	2.0
Grain weight per panicle	2	RM3666-RM3874	*O.rufipogon*	3.2	24	-2.1	3.3	24	-2.0
Yield per plant	2	RM1303-RM3874	*O.rufipogon*	14.4	69	-5.7	15	67	-5.8

Table 6: Result of QTL analysis using interval mapping and composite interval mapping along with allelic effect for 9 yield and related traits in BC3F3 families.

Trait↓/Marker→	RM3666	RM1303	RM3688	RM3762	RM3874	RM3515	RM6318	RM1920
LL	-	-	-	* (R)	-	-	***(S)	* (S)
LW	-	-	-	-	-	-	-	-
PH	-	-	-	-	-	-	-	-

UNDER PEER REVIEW

NT	-	-	**** (R)	**** (R)	-	* (S)	** (S)	-
NP	-	-	**** (R)	**** (R)	* (S)	* (S)	** (S)	-
PL	-	-	* (R)	** (R)	-	-	** (S)	-
50%F	-	** (R)	**** (S)	**** (S)	-	-	-	-
GWP	-	* (S)	*** (R)	** (R)	-	-	-	-
YLDP	-	-	**** (R)	**** (R)	-	* (S)	**** (S)	* (S)

Table:7. Summary of QTL analysis in BC3F3 population.

Gene References TIGR	**IRGSP**	**Genomic Location**	**Description**
LOC_Os02g37070	Os02g0581200	22,395,465-22,396,224	expressed protein
LOC_Os02g37080	Os02g0581300	22,401,835-22,406,455	ASC1-like protein 1, LONGEVITY ASSURANCE HOMOLOG 3, putative, expressed,
LOC_Os02g37090	Os02g0581500	22,420,365-22,423,033.	esterase ytxM, putative, expressed
LOC_Os02g37094	Os02g0581600	22,423,367-22,424,193	expressed protein
LOC_Os02g37100	Os02g0581800	22,426,184-22,426,994.	expressed protein
LOC_Os02g37109	Os02g0581900	22,427,693-22,428,862.	vegetative cell wall protein gp1 precursor, putative, expressed
LOC_Os02g37120	Os02g0582100	22,430,203-22,435,009	expressed protein 18S pre-ribosomal assembly protein gar2-related; similar to unknown protein [Arabidopsis thaliana]
LOC_Os02g37130	Os02g0582200	22,439,111-22,443,133.	expressed protein
LOC_Os02g37140	Os02g0582300	22,446,828-22,451,878.	fertility restorer homologue A, putative, expressed
LOC_Os02g37150	Os02g0582400	22,453,431-22,456,886.	expressed protein

Table 8: List of genes present in the region between RM3688-RM3762