

Original Research Article

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

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. Introgression 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.

Comment [DN1]: Of

Comment [DN2]: Delete

Comment [DN3]: The

Comment [DN4]: Delete

Comment [DN5]: Delete

Comment [DN6]: previously

Comment [DN7]: delete

Comment [DN8]: Delete, no reference in abstract

Comment [DN9]: Italic

Comment [DN10]: Delete

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

25 **Introduction**

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

34 With the advent of DNA molecular markers, QTL mapping has become a routine strategy for the
 35 discovery of genes involved in complex quantitative traits. Thousands of QTL have been mapped
 36 for important agronomical traits in rice. Although primary mapping populations including F₂,
 37 recombinant inbred lines (RILs) and doubled haploid lines (DHs) have been widely used for
 38 QTL mapping in rice (Fan et al. 2005; Marri et al. 2005), QTL can only be localized to a
 39 genomic region (confidential region) rather than a locus in those populations. Following the
 40 primary mapping, advanced populations such as near isogenic lines (NIL) and chromosome
 41 segment substitution lines(CSSL) can be used to map QTL to a locus as a Mendalian factor by
 42 blocking the genetic background noise (Wang et al. 2004). Based on this strategy, several QTLs
 43 have been isolated in tomato (Fraryet al. 2000) and rice (Yano et al. 2000; Takahashi et al. 2001)
 44 in recent years. The common aspect in all the QTL cloning work is to exploit high quality NILs
 45 as advanced mapping populations.

Comment [DN11]: Replace with Rice

Comment [DN12]: ONE SPACE

Comment [DN13]: One space

Comment [DN14]: italic

Comment [DN15]: no space

47 In the long run, development of high yield varieties is one of the most important goals in rice
 48 cultivation. Unfortunately, yield improvement efficiency is deemed to be very low due to its
 49 complex property affected by number of spikelets per panicle, 1,000-grain weight and tillers per
 50 plant. Of these factors, SPP was shown to be highly correlated with yield and acts as a crucial
 51 component in determining rice yield (Hua et al. 2002). Therefore, dissection of its genetic basis
 52 would be of great value in breeding high yield variety. So far, QTLs for SPP have been mapped
 53 in lots of populations, which were derived from inter-sub specific and intra specific crosses.
 54 Although many QTL controlling SPP were mapped in rice, only *Gn1a*, has been cloned. *Gn1a*,
 55 controlling grain productivity in rice, was elucidated to be a gene encoding cytokinin
 56 oxidase/dehydrogenase (*OsCKX2*), an enzyme that degrades the phytohormone cytokinin.
 57 Reduced expression of *OsCKX2* causes cytokinin accumulation in inflorescence meristems and
 58 increases the number of reproductive organs, resulting in enhanced grain yield (Ashikari et al.
 59 2005). He et al 2007, constructed nearly isogenic lines (NILs) that differ only at a single
 60 quantitative trait locus (QTL) and fine-mapped the yield-improving QTL *qGY2-1* to a 102.9-kb
 61 region on rice chromosome 2. NILs identified the haplotypes of a leucine-rich repeat receptor
 62 kinase (*LRK*) gene cluster, the allelic variation of gene expression in hybrids at the *LRK* locus
 63 that associated with the yield QTL ranged from unequal expression of the two alleles
 64 simultaneously to expression of a single allele regardless of the parent-of-origin. They suggested
 65 allelic variations in structure and expression points to LRK gene cluster to be good candidate for
 66 the source of yield QTL.

67 In our previous studies, a major QTL responsible for major yield traits was mapped on
 68 chromosome2 flanked by SSR markers RM262 and RM263 (Marri et al. 2005). Marri et. al
 69 (2005) have used 251 BC2 test cross families and commercial hybrid KRH2 (IR58025A X

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 77 RM262 – RM263 (fig 1), followed by fine mapping of the candidate sub-QTL.

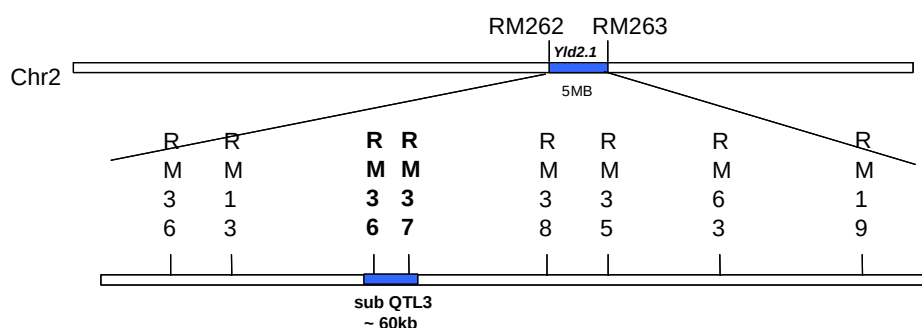


Fig 1: Depiction of *yld2.1* on chromosome 2.

Materials and Methods

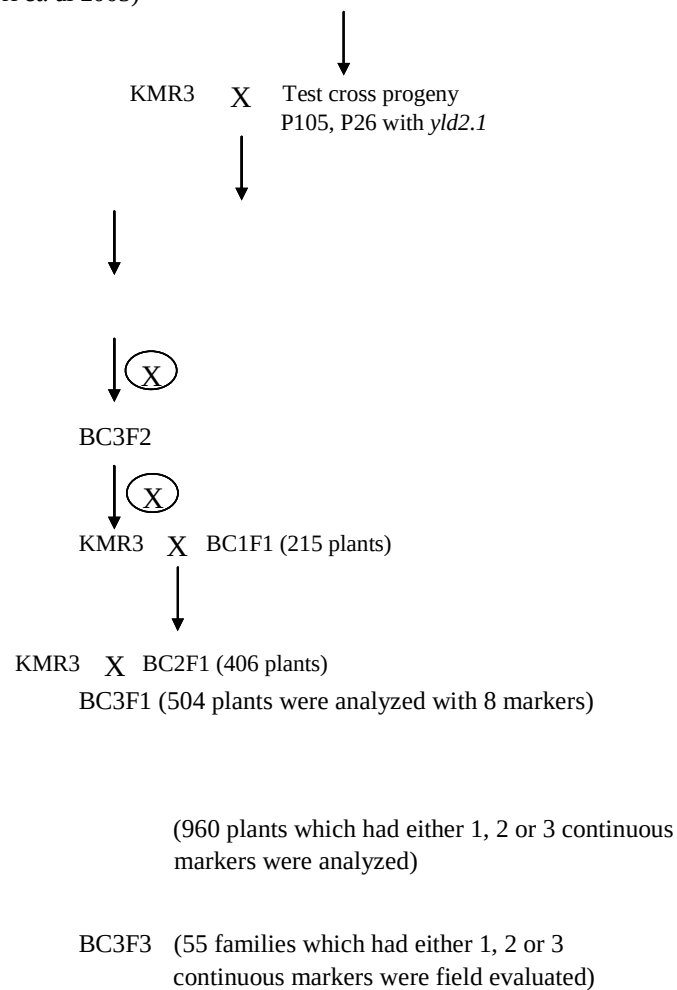
Development of NILs:

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

Comment [DN16]: delete

86 as the wild donor for this study. An accession of *O. rufipogon* which was genetically moderately
 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

IR58025A/*O. rufipogon*//IR580325B///IR58025B///KMR3
 (Marri et. al 2005)



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

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

108 **DNA extraction and SSR analysis:**

109 Fresh leaves of the individual from introgression lines were collected and ground in liquid
 110 nitrogen. DNA was extracted from the ground tissues by the CTAB (cetyltrimethyl ammonium
 111 bromide) method. A total volume of 10 µl reaction mixture was composed of 2 µl of 50 ng/µl
 112 template DNA, 10 mmol Tris-HCl (pH 9.0), 50 mmol KCl, 1.5 mmol MgCl₂, 0.1% Triton
 113 X100, 0.2 µM of each forward and reverse primer, 2.5 mM each of dNTP, and 1 U of Taq DNA
 114 polymerase (Bangalore Genei). The PCR amplification was performed under the following
 115 conditions: Initial denaturation at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C
 116 for 1 min, annealing at 55°C for 30 sec, extension at 72°C for 2 min, followed by the final
 117 extension 72°C for 5 min. Following amplification, the products were checked for
 118 polymorphism or marker segregation on an agarose gel and scored for the segregating bands.

119 **Phenotypic evaluation of BC₃F₂ population:**

Comment [DN17]: references???

120 The 8000 BC₃F₂ plants derived from 400 BC₃F₁ families were grown in an experimental plot at
 121 DRR (Directorate of Rice Research) Hyderabad, India, during dry season in 2008. They were
 122 planted with a distance of 15 cm X 20 cm for a line X row. From this population we have
 123 selected 37 families (each family consist of 24 plants and in some families 48 plants,
 124 approximate total of 960 plants) which contained either only 1, 2 or 3 consecutive markers. This
 125 selection was to minimize the *O.rufipogon* chromosomal segment, and in future to fine map the
 126 specific gene/s contributing to yield increase. Morphological and phenotypic selection was done
 127 in the selected 960 plant population, out of this 55 plants showed homozygous *O.rufipogon*
 128 alleles at corresponding marker.

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

132 Measuring yield related traits:

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

140 Results and Discussions:

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

Comment [DN18]: Delete

Comment [DN19]: ,

Comment [DN20]: ,cm

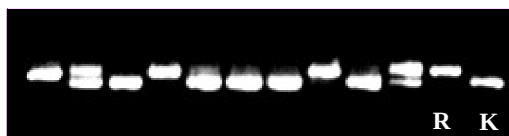
Comment [DN21]: , cm

Comment [DN22]: ,cm

Comment [DN23]: ,cm

Comment [DN24]: , days

142 All 504 BC3F1 plants were tested with 8 polymorphic SSR markers for the presence of
 143 *O.rufipogon* alleles within the QTL *yld2.1* (fig. 3). In these 8 markers, RM3762 shown highest
 144 introgression of 46.63%, followed by RM1920 with 41.07% and lowest was RM3666 with
 145 7.54% in 504 BC3F1 plants. Table 2 shows the no. of plants for each marker with *O. rufipogon*
 146 allele at that marker and % distribution of corresponding marker in 504 plants. As shown in table
 147 3, out of 504 BC3F1 plants 402 plants had at least one of these 8 markers. Out of these 402
 148 plants, 6 plants had all the 8 markers, 5 plants had 7 markers, 28 plants had 6 markers, 41 plants
 149 had 5 markers, 53 plants had 4 markers, 84 plants had 3 markers, 79 plants had 2 markers and
 150 106 plants only one marker. For fine mapping a QTL, plants with shortest introgression would be
 151 beneficial, for this reason we have selected plants with only 1, 2 or 3 consecutive markers having
 152 *O. rufipogon* allele. These plants were devoid of *O. rufipogon* at rest of the markers and had *O.*
 153 *sativa*. There were 36 plants which had only 1, 2 or 3 consecutive markers for *O.rufipogon* allele.
 154 These 36 plants were continued for selfing along with rest of the population. This selection was
 155 to minimize the *O.rufipogon* chromosomal segment, so that we could reach our goal of
 156 narrowing down into QTL *yld2.1* and in future to fine map the specific gene/s contributing to
 157 yield increase.



158
 159 Fig 3: Segregation of RM 3688 in BC3F1 introgression lines. The extreme right lanes show *O.rufipogon* (R) and
 160 KMR3 (K) alleles respectively

161 Genetic and phenotypic analysis of BC3F2 population:

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

Comment [DN25]: delete

Comment [DN26]: results showed that

Comment [DN27]: was

Comment [DN28]: data in

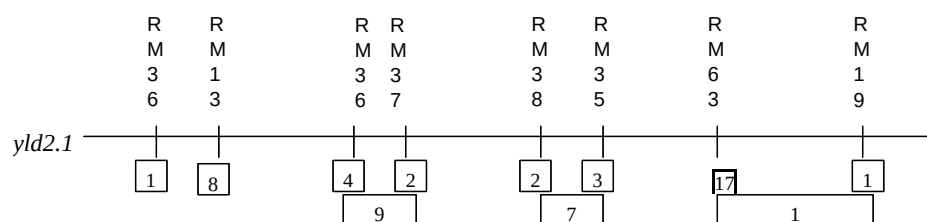
Comment [DN29]: showed

Comment [DN30]: T

Comment [DN31]: Where number, names, marker, molecular weight ???

164 55 plants were homozygous *O.rufipogon* segments for one of the 8 markers. Out of 55 plants one
 165 plant had RM 3666, 8 plants had RM1303, 4 plants had RM3688, 2 plants had RM3762, 9 plants
 166 had both the RM3688 and RM 3762, 2 plants had RM3874, 7 plants had both the RM 3874 and
 167 RM 3515, 3 plants had RM 3515, 17 plants had RM 6318, one plant had both the RM6318 and
 168 RM 1920 and one plant had RM 1920 markers Respectively

169



170

171

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

173

174 Morphological data of these 55 homozygous BC3F2 plants grown in field was recorded and
 175 BC₃F₂ plants indicated that 9 plants with *O.rufipogon* allele at 3rd sub QTL (RM 3688 and RM
 176 3762) showed transgressive segregation for grain weight per plant, tiller number per plant.

177 Fertility restoration was tested using IR58025A as CMS line and the KMR3 plants having one of
 178 the 8 markers individually. The results on hybrids indicated that plants with 3rd sub-QTL restored
 179 fertility to a greater extent compared to the plants with other sub QTLs.

180 In 55 BC3F2 homozygus plants Sub-QTLs were analyzed for single marker analysis using QTL
 181 cartographer (v2.5). For analyzing marker-QTL association for grain yield per plant and

182 determining the precise location of the sub QTLs, A LOD score of 2.5 was used as the threshold
183 for detecting the location of QTLs and 1000 permutations. LOD peaks for significant sub QTLs

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

187 **Evaluation of BC3F3 population for yield and related traits:**

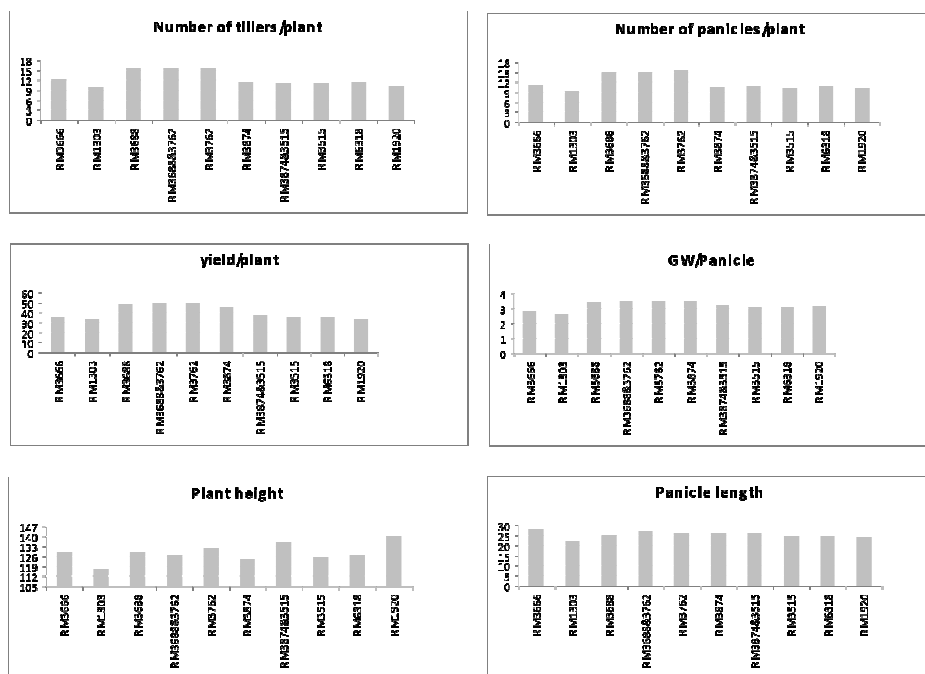
188 The homozygous population of 55 lines constituting 960 plants, representing all the eight
189 markers **was field** evaluated in Kharif'2008 at DRR. Yield traits measurements were obtained
190 using a complete randomized block design, with each block having three rows (24 plants per
191 row) and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

Comment [DN32]: delete

Comment [DN33]: both

198

199



200
201

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

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

210 Single marker analysis for nine traits was performed with yield data from BC3F3 population with
211 8 SSR markers shown in table 5. Seven traits were linked with different significance values to
212 different markers. Markers RM3688 and RM3762 were linked to traits NT, NP, GWP and YLDP

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

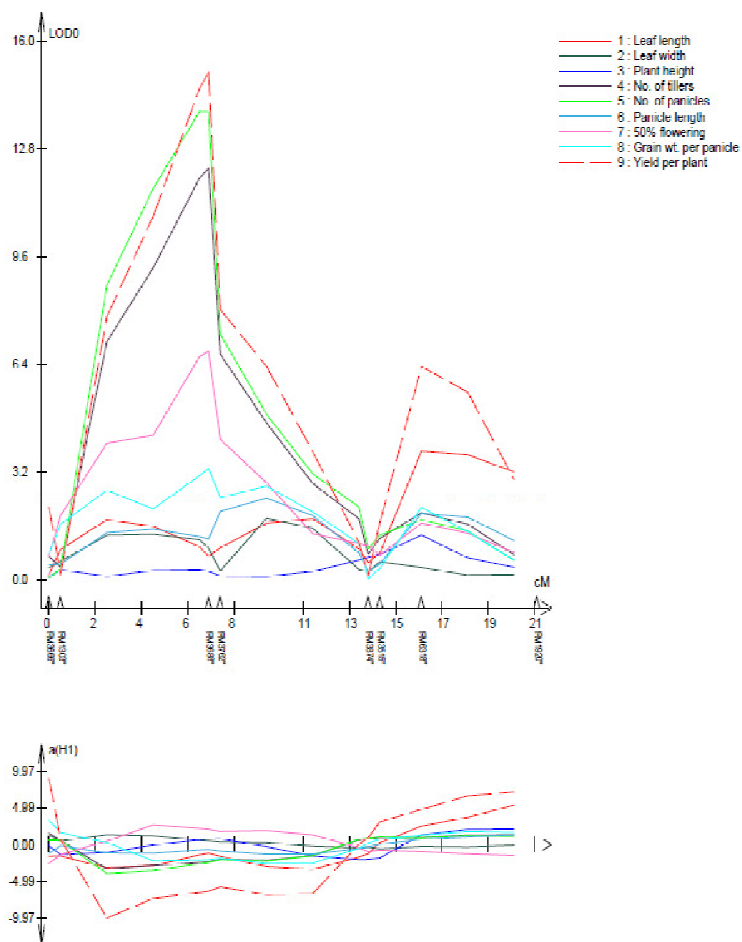
216 Table 6 shows the result of QTL analysis using interval mapping (IM) and composite interval
217 mapping (CIM). In this analysis, nine traits were analyzed for QTL mapping. *O. rufipogon* has
218 shown allelic effect on NT, NP, PL, GWP and YLDP in both IM and CIM.

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

Comment [DN34]: data in

Comment [DN35]: showed that

Comment [DN36]: delete



224

225

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

227

228 QTL mapping provides the genomic region responsible for desirable trait and the size of
 229 this region depends on the resolution of mapping. For characterization of candidate gene/s
 230 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

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

238 The *yld2.1* had LOD of 35.33 and additive effect of -238.51 in interval mapping, LOD of 31.92
 239 and additive effect of -216.04 in composite interval mapping. Higher LOD score and additive
 240 effect for *yld2.1* denotes it's a major effect QTL and fine mapping would provide the key gene/s
 241 controlling yield and related traits. According to Marri et al 2004, QTLs *snp2.1*, *np2.1*, *nt2.1*,
 242 *gnp2.1*, *gw2.3*, *yldp2.1*, *gnp2.2*, *yldp2.2* and *hi2.1* were mapped in between RM262 and RM263;
 243 it is possible that all these traits are controlled by different genes or a single gene which has
 244 pleiotropic effect on multiple traits. For any fine mapping a high quality mapping population is
 245 essential, in our case NILs (near isogenic lines) were derived from IR58025A/O.
 246 *rufipogon*//IR580325B//IR58025B////KMR3 (Marri et. al 2005). Where 2 plants containing
 247 *yld2.1* QTL were crossed with KMR3 (a restorer line) and continued back crossing with KMR3
 248 three more times to generate a BC3F3 population. This crossing program was devised to transfer
 249 *yld2.1* to restorer line KMR3 and also for fine mapping.

Comment [DN37]: and

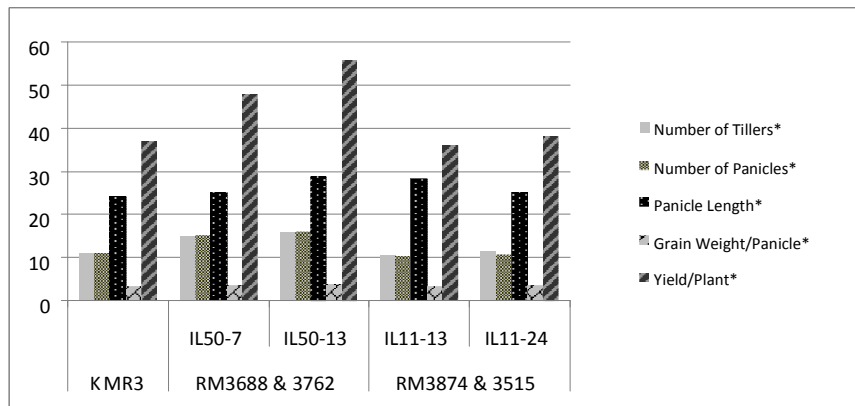
250 We have used substitution mapping for identifying the region responsible for increase in yield
 251 traits. In substitution mapping, the genotypes and trait phenotypes of individual sub-NILs are
 252 compared with recipient parent to identify sub-NILs containing genetic factors contributing to
 253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

intervals likely to contain relevant genes can be identified and precisely located. Substitution mapping has been very successful in identifying genes with large phenotypic effects in different species, including tomato (Fridman et al. 2002).

BC2F2 and BC2F3 introgression lines derived from *Oryza rufipogon* into an *indica* cultivar Guichao 2 were used in QTL analysis (Tian et al 2006) and QTL *gpa7* was mapped to short arm of chromosome 7 affecting grain number per panicle. Tian et al., (2006) have used two-step substitution mapping for narrowing down *gpa7* to a region of 35kb that contained five predicted genes. Like in our case, *gpa7* QTL region was mapped to multiple panicle traits (length of panicle, primary branches per panicle, secondary branches per panicle, grains on primary branches and grains on secondary branches). Using NILs developed from Jefferson (*O.sativa*) X *O. rufipogon* (IRGC105491), Li et al (2004) fine mapped *gw3.1* QTL to a region of 93.8-kb region from initial region of 31.8 cM. Brouwer et al (2004) used NILs for fine mapping three QTLs (*lb4*, *lb5b* and *lb 11b*) for late blight resistance spanning 6.9-cM, 8.8-cM and 15.1-cM on chromosomes 4, 5 and 11 respectively. Using NILs they could limit the three QTLs to smaller regions on the chromosome, which would facilitate the MAS of late blight resistance more easy and precise. Liu et al (2006) fine mapped QTL *SPP1* (spikelets per panicle) on chromosome 1 to a region of 107 kb using NILs. Wan et al (2006) have narrowed down GL-3a QTL controlling grain length to a region of 87.5 kb on chromosome 3 using chromosome segment substitution lines grown under multiple environmental conditions and 2 different genetic backgrounds. Song et al (2007) cloned and characterized the QTL *GW2* responsible for grain width and weight, their fine mapping yielded the RING-type E3 ubiquitin ligase gene as responsible for the trait. This gene negatively controlled the cell division by affecting proteasomes regulated proteolysis. Xie et al 2006 fine mapped a yield-enhancing QTL cluster from *O. sativa* X *O. rufipogon* to a region

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.



299
300 Fig 7: Comparison of five linked traits in plants with RM3688-RM3762, RM3874-RM3515
301 along with parent KMR3.



302
303 Figure 8: Increased no. of tillers in IL50-7 (RM3688-RM3762) in comparison with KMR3.

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

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

314 LL – Leaf length, LW – Leaf width, PH – Plant height, NT – No. of tillers, PL – Panicle length,
315 50%F – 50% flowering, GWP – Grain weight per panicle, YLDP – Yield per plant

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

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

318 Interval mapping and composite interval mapping performed has identified QTLs for all nine
319 traits (table 6). Out of the five traits which have come from *O. rufipogon* in IM,
320 RM1303RM3874 region had shown phenotypic variance of 69% for YLDP and 59% for NT.
321 RM3666-

322 RM3874 region had shown phenotypic variance of 65% for NP and 24% for GWP.
323 RM3762RM3874 region also showed phenotypic variance of 23% for PL. Above results are
324 similar in the case of CIM as well. As shown in the figure 6, sharpest region in peaks for all the
325 five traits were in the region of RM3688-RM3762. This indicates this region to be the candidate
326 for the increase in yield seen in original *yld2.1* QTL. Along with QTL analysis, yield data in the
327 field and physical observation of plants also points us to the region of RM3688-RM3762 as
328 candidate for the *yld2.1*. This study has narrowed down the *yld2.1* QTL from a region of 5MB to
329 a region of RM3688-RM3762 (fig .1).

330

331 Genomic region between RM3688-RM3762 on chromosome 2 in reference sequence *O. sativa*
332 cv. Japonica Nipponbare has about 62kb DNA. There are ten genes/ORFs in the 62kb region
333 (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, vegetative cell protein gp1 precursor, 18S pre-ribosomal assembly protein and fertility restorer homologue A for orthologous proteins from other species. The sequence annotation using only sequence data of cv. Nipponbare in this study would make it difficult to predict a candidate gene underlying *yld2.1*, which was derived from the *O. rufipogon*. This is mainly due to divergence in sequence level between the cultivated rice and its wild counterpart (Ammiraju et al. 2006). Ammiraju et al. (2006) revealed that the orthologous region of the *Adh1* gene in the *O. sativa* genome was 9.4– 28% larger relative to four wild species including *O. rufipogon* and this size variation in the *Adh1* gene was mainly due to insertion of transposable elements as well as multiple genetic mechanisms. Thus, the candidate gene in our study should be confirmed based on the complete sequence of *O. rufipogon* in the target region and we are sequencing the target region of *O. rufipogon* to determine the gene underlying *yld2.1* in this regard.

Most of the studies reporting yield QTLs and their characterization involved regular cultivars as recipients, and their donor as either another high yielding cultivar or wild relatives of rice like *O. rufipogon*. Where as current study aims to improve the yield of hybrids using introgression of yield-enhancing QTLs into the parents of the hybrid. We foresee the future with improved restorer (KMR3) lines as parents in developing high yielding rice hybrids. We conclude that the region flanked by RM 3688 and RM 3762 is important in contributing to high yield. Thus the yield enhancing QTL *yld2.1* was narrowed down from 5Mb to 62Kb. This genomic region is being fine mapped further to identify candidate genes for use in MAS or functional analysis.

354

355 **Reference:**

UNDER PEER REVIEW

- 356 Ammiraju J. S., Luo M., Goicoechea J. L., Wang W., Kudrna D., et al. , 2006. The Oryza
357 bacterial artificial chromosome library resource: construction and analysis of 12
358 deepcoverage large-insert BAC libraries that represent the 10 genome types of the genus
359 Oryza. Genome Res. 16: 140–147.
- 360 Ashikari M, Sakakibara H, Lin S, Ymamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q,
361 Kitano H, Matsuoka M, Cytokinin Oxidase regulates rice grain production. Science 309:
362 (2005) 741-745.
- 363 Botstein D, White RL, Skolnick M, Davis RW, Construction of a genetic linkage map in man
364 using restriction fragment length polymorphism. American Journal of human Genetics
365 32: (1980) 324-331.
- 366 Brouwer D. J., St.Clair D. A., 2004. Fine mapping of three quantitative trait loci for late blight
367 resistance in tomato using near isogenic lines (NILs) and sub-NILs. Theor. Appl.
368 Genet. 108: 628–638
- 369 Clayton WD, Phillips SM, Renvoize SA, Gramineae Part II. In: Milne-Redhead E, Polhill RM,
370 eds. Flora of Tropical East Africa. UK: Crown Agents for Oversea Governments and
371 Administrations (1974).
- 372 Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Zhang Q, GS3, a major QTL for grain length and
373 weight and minor QTL for grain width and thickness in rice, encodes a putative
374 transmembrane protein. Theor. Appl. Genet. 112(6): (2006) 1164-1171.
- 375 Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R,
376 Alpert KB, Tanksley SD, fw2.2: a quantitative trait locus key to the evaluation of tomato
377 fruit size. Science 289: (2000) 85-88.

Comment [DN38]: (2006)

Comment [DN39]: (2005)

Comment [DN40]: delete

Comment [DN41]: (1980)

Comment [DN42]: Delete

Comment [DN43]: The same for all please,
authors, year, title, journal

UNDER PEER REVIEW

- 378 Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D, Zooming in on a quantitative trait for tomato
379 yield using interspecific introgressions. *Science* 305: (2004) 1786–1789.
- 380 He G, Luo X, Tian F, Li K, Zhu Z, Su W, Haplotype variation in structure and expression of a
381 gene cluster associated with a quantitative trait locus for improved yield in rice. *Genome*
382 *Research* 16: (2007) 618–26.
- 383 Hua, J.P., Xing, Y.Z., Xu, C.G., Sun, X.L., Yu, S.B. and Zhang, Q.F. (2002) Genetic dissection
384 of an elite rice hybrid revealed that heterozygotes are not always advantageous for
385 performance. *Genetics*, **162**, 1885–1895.
- 386 Li JM, Xiao JH, Grandillo S, Jiang LY, Wan YZ, Deng QY, Yuan LP, Mc Couch SR, QTL
387 detection for rice grain quality traits using an interspecific backcross population derived
388 from cultivated Asian (*O. sativa*) and African (*O. glaberrima*) rice. *Genome* 47: (2004)
389 697-704.
- 390 Liu QQ, Li QF, Cai XL, Wang HM, Tang SZ, Yu HX, Molecular marker-assisted selection for
391 improved cooking and eating quality of two elite parents of hybrid rice. *Crop Sci* 46:
392 (2006) 2354–2360.
- 393 Liu QQ, Li QF, Cai XL, Wang HM, Tang SZ, Yu HX, Molecular marker-assisted selection for
394 improved cooking and eating quality of two elite parents of hybrid rice. *Crop Sci* 46:
395 (2006) 2354–2360.
- 396 Marri PR, N Sarla, VLN Reddy and EA Siddiq, Identification and mapping of yield and related
397 QTLs from an Indian accession of *Oryza rufipogon*, *BMC Genet.* 6: (2005) 33.
- 398 Moncada P, Martinez CP, Borrero J, Chatel M, Gauch HJ, Guimaraes E, Tohme J, McCouch SR,
399 Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza*
400 *rufipogon* BC₂F₂ population evaluated in an upland environment. *Theoretical and*

401 Applied Genetics 102: (2001) 41-52.
 402 Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ,
 403 Worland AJ, Pelica, Green revolution genes encode mutant gibberellin response
 404 modulators. Nature 400: (1999) 256–261.
 405 Septiningsih EM, Prasetyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S, McCouch SR,
 406 Identification of quantitative trait loci for yield and yield components in an advanced
 407 backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O.*
 408 *rufipogon*. Theoretical and Applied Genetics 107: (2003b) 1419-1432.
 409 Song XJ, Huang W, Shi M, Zhu MZ, Lin HX A QTL for rice grain width and weight encodes a
 410 previously unknown RINGtype E3 ubiquitin ligase. Nat Genet 39: (2007)623–630.
 411 Takahashi Y, Shomura A, Sasaki T, Yano M Hd6, a rice quantitative trait locus involved in
 412 photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. Proc. Natl.
 413 Acad. Sci. USA 98: (2001) 7922–7927.
 414 Thomson MJ, Tai TH, McClung AM, Lai XH, Hinga ME, Lobos KB, Xu Y, Martinez, CP,
 415 McCouch SR, Mapping quantitative trait loci for yield, yield components and
 416 morphological traits in an advanced backcross population between *Oryza rufipogon* and
 417 the *Oryza sativa* cultivar Jefferson. Theoretical and Applied Genetics 107: (2003)479493.
 418 Tian F, Li DJ, Fu Q, Zhu ZF, Fu YC, Wang XK, Construction of introgression lines carrying
 419 wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.)
 420 background and characterization of introgressed segments associated with yield related
 421 traits. Theoretical and Applied Genetics 112: (2006)570–80.
 422 Wan, X. Y., J. M. Wan, L. Jiang, J. K. Wang, H. Q. Zhai *et al.*, 2006. QTL analysis for rice grain
 423 length and fine mapping of an identified QTL with stable and major effects. Theor. Appl.
 424 Genet. 112 1258–1270.

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

428 Xie X, Song MH, Jin F, Ahn SN, Suh JP, Hwang HG, Fine mapping of a grain weight
 429 quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a
 430 cross between *Oryza sativa* and *Oryza rufipogon*. Theoretical and Applied Genetics 113:
 431 (2006) 885–94.

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

434 Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K,
 435 Umehara Y, Nagamura Y, Hd1, a major photoperiod sensitivity quantitative trait locus in
 436 rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell
 437 12: (2000) 2473–2484.

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

RM1920	CAAACACAGTGTGACAGAA	GCTATTGACTTATCCGTTCA
--------	---------------------	----------------------

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

Marker	RM3666	RM1303	RM3688	RM3762	RM3874	RM3515	RM6318	RM1920
Plants having <i>O.rufipogon</i>	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

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

439

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	

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

441

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	
--------	--------	-------	-------	-------	--

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

Trait	Marker	Additive effect	Likelihood ratio	Fcalculated	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 ****
	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 *

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

445

Trait	Chr.	Marker interval	Allelic effect	Interval mapping			Composite interval mapping		
				LOD	R ²	Additive	LOD	R ²	Additive
Leaf length	2	RM3874-RM1920	<i>O.sativa</i>	3.8	27	2.4	3.8	25	2.5
Leaf width	2	RM3762-RM3874	<i>O.sativa</i>	1.8	1	0.3	1.8	2	0.3
Plant height	2	RM3762-RM1920	<i>O.sativa</i>	1.3	4	1.2	1.3	4	1.4
Number of tillers	2	RM1303-RM3874	<i>O.rufipogon</i>	12.3	59	-2.4	12.2	62	-2.4
Number of panicles	2	RM3666-RM3874	<i>O.rufipogon</i>	13.7	65	-2.1	13.8	72	-2.2
Panicle length	2	RM3762-RM3874	<i>O.rufipogon</i>	2.4	23	-1.4	2.4	23	-1.3
50% flowering	2	RM3666-RM3515	<i>O.sativa</i>	6.8	40	1.8	6.8	44	2.0

Grain weight per panicle	2	RM3666-RM3874	<i>O.rufipogon</i>	3.2	24	-2.1	3.3	24	-2.0
Yield per plant	2	RM1303-RM3874	<i>O.rufipogon</i>	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.

448

Trait./Marker→	RM3666	RM1303	RM3688	RM3762	RM3874	RM3515	RM6318	RM1920
LL	-	-	-	* (R)	-	-	*** (S)	* (S)
LW	-	-	-	-	-	-	-	-
PH	-	-	-	-	-	-	-	-
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	IRGSP	Genomic Location	Description
TIGR			
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

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