THESIS FOR DEGREE OF MASTER OF SCIENCE

QTL mapping and candidate gene analysis for pant shape and yield-related traits using whole genome resequncing in rice

BY

HYUN-JUNG YANG

AUGUST, 2013

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

QTL mapping and candidate gene analysis for pant shape and yield-related traits using whole genome resequncing in rice

UNDER THE DIRECTION OF DR. NAM-CHON PAEK
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF SEOUL NATIONAL UNIVERSITY

BY
HYUN-JUNG YANG

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY
DEPARTEMENT OF PLANNT SCIENCE

JULY, 2013

APPORVOED AS A QUALIFIED DISSERTATION OF
HYUN-JUNG YANG
FOR THE DEGREE OF MASTER
BY THE COMMITTEE MEMEBERS

AUGUST, 2013

CHAIRMAN

Hee-Jong Koh, Ph.D.

VICE-CHAIRMAN

Nam-Chon Paek, Ph.D.

MEMBER

Suk-Ha Lee, Ph.D.

QTL mapping and candidate gene analysis for pant shape and yield-related traits using whole genome resequncing in rice

HYUN-JUNG YANG

ABSTRACT

Plant breeders have focused on the improvement of plant architecture to break through the stagnated crop yield since the Green Revolution. To identify quantitative trait loci (QTLs) influencing plant shape and yield in rice, eight agronomic traits (days to heading, plant height, tiller number, panicle diameter, panicle length, flag leaf length, flag leaf width and yield per plant) were analyzed. QTL analysis was performed with 178 F7 recombinant inbred lines (RILs) derived from a cross of japonica rice ‘SNU-SG1’ and the Tongil-type high-yield rice ‘Milyang23’. By next generation sequencing (NGS) approach, whole genomes resequencing was applied to the two parent cultivars. Using 125 molecular markers developed by NGS, including 64 simple

sequence repeat, 33 sequence-tagged site and 28 insertion/deletion markers, distributed in 12 chromosomes, we identified 46 QTLs for the eight agronomic traits with the threshold of LOD > 2.8. By analyzing the coding sequences using the NGS data, we found 17 strong candidate genes for 20 main-effect QTLs, which harbor causal SNPs leading to frame shift or premature stop codon in amino acid sequences. It is noteworthy that the application of whole genome sequencing substantially enhanced the efficiency in polymorphic marker development for QTL mapping and candidate gene isolation for the QTL. Our study provides not only advanced approach for QTL analysis, but also the list of qualified QTLs and strong candidate genes for each QTL related to plant shape and yield. This would be useful genetic resources to the breeding of new plant-type high-yielding rice.

Keywords: quantitative trait loci, whole genome resequencing, plant architecture, yield, agronomic trait, rice

Student number: 2011-23483

CONTENTS

LIST OF TABLES AND FITURES

ABBREVIATION

QTL	Quantitative trait loci
RIL	Recombinant inbred lines
SNP	Single nucleotide polymorphism
NGS	Next generation sequencing
WGS	Whole-genome sequencing
CIM	Composite interval mapping
NID	NGS-based insertion/deletion
InDel	Insertions/deletions

INTRODUCTION

Rice is a staple food in many Asian countries accounting for about 60 % of the world population. There has been tremendous concern about impending global food shortage due to the consequence of world population growth. Nevertheless, a per capita increase of global food production was above the increase in population for the decades since 1960 due to enormous increase of cereal crops production (Trewavas 2001). This dramatic increase of crop yield, which has been known as the 'green revolution', was due largely to the development of genetically improved high-yielding crop varieties (Hedden 2003). Among genetic approaches to improve yield potential, ideal-type breeding was applied by modification of plant architecture (morphology); for example, the plant type such as IR24 that enabled green revolution displayed high tiller (shoot) numbers, short plant height and erect leaves. Various genes have contributed to this modification of the rice plant architecture, i.e. a recessive gene, sd1, derived from a Chinese variety, Dee-geowoo-gen, was incorporated to reduce plant height (Suh et al. 1978),

and a photoperiod-insensitive gene, se1, was introduced to increase rice adaptability to a wide range of latitudes (Sano, 1992; Tamura et al. 1998).

Although this genetic modification plays an importantant role in marked increase of crop production, the demand for rice varieties with higher yield potential and greater yield stability is rising due to increasing human population, which is expected to touch 9 billion by 2050, and a changing global climate (Khush, 2005). To further enhance the rice yield potential over that of existing high-yielding cultivars, a new plant type was proposed (Khush 1995). The proposed ideal plant architecture (IPA) included several important characteristics such as a reduction in tiller number, few or no unproductive tillers, an increase in the number of grains per panicle, and thick and sturdy stems (Khush 1995; Li et al., 2012). The genes and QTLs regulating the IPA-related traits have been reported. For example, OsSPL14, which is regulated by OsmiR156, was proposed to regulate IPA-related traits; a point mutation in OsSPL14 perturbs OsmiR156-mediated regulation of OsSPL14 and generates an ‘ideal’ rice plant with a reduced tiller number, increased lodging resistance and enhanced grain yield (Jiao et al.

2010). Marathi et al. (2012) reported 19 novel QTLs and 15 QTL hotspots providing favorable alleles for yield and yield-contributing traits including various traits involved in plant architecture using RILs developed from Pusa1266, a semi-dwarf high-yielding cultivar. Several other QTLs regulating plant shape and yield, including leaf length, leaf width, tillering, panicle length, spikelet fertility and grain size, have been reported (Vergara et al. 1996; Moncada et al. 2001; Septiningsih et al. 2003; Farooq et al. 2010; Liu et al. 2010; Wang et al. 2010; Wang et al. 2011). Although several studies were performed, identification of QTLs and their responsible genes in different genetic background are necessary to modify the plant architecture for the IPA-plant breeding.

QTL mapping and cloning using DNA markers are useful tools to unlock the genetic basis of complex phenotypic variation as well as the application of molecular markers developed from QTL analysis enhances breeding efficiency for the particular agronomic traits. However, traditional approach for QTL analysis is time-consuming and labor-intensive because of low efficiencies in polymorphic marker development for high-resolution map construction and candidate gene identification for the detected

QTL. These limitations in QTL analysis have been trying to overcome by applying whole-genome sequencing (WGS) approach with the advent of the next-generation sequencing technology. A sequencing-based genotyping method uses single nucleotide polymorphisms (SNPs) detected from WGS of a mapping population (Huang et al. 2009). This approach could substantially reduce the amount of time and effort required for genotyping of mapping population. Recently, another new approach named as QTL-seq was proposed for rapid identification of QTL for a given phenotype by WGS of DNAs from two populations each composed of 20-50 individuals showing extreme opposite trait values in a segregating progeny (Takagi et al. 2013). Although these approaches with application of WGS take advantages of time and cost effectiveness for QTL mapping, solutions to increase the efficiency in candidate gene identification has not yet been proposed.

In this study, we mapped QTL and QTL hotspots for eight agronomic traits related to plant shape and yield, and identified 46 QTLs for the eight traits. By WGS of two parents, we identified 17 strong candidate genes for the QTLs related to seven agronomic

traits. Application of WGS approach substantially enhanced the efficiencies in polymorphic marker development for QTL mapping and candidate gene identification compared to conventional approach. The identified QTLs and their candidate genes provided from WGS-based QTL analysis would be useful genetic resources for high-yielding new plant-type rice breeding.

MATERIALS AND METHODS

Mapping population

To identify QTL influencing plant shape and yield, Oryza sativa Tongil-type hybrid cultivar Milyang23, which shows many tillers and high yielding, was crossed with japonica cultivar SNU-SG1, which has less tiller and yield. Among 338 F7 RILs developed from F2 population by single seed decent (SSD), 178 individual lines were used for QTL mapping, which consist of 40 lines showing extreme phenotypic values for each trait. Field management followed normal rice practice in Suwon, South Korea (37o N latitude) in 2011. Seeds were sown on April and seedlings were transplanted on May at spacing of 0.3 m by 0.15 m with one seedling per hill.

Evaluation of agronomic traits

For the phenotypic evaluation of the selected 178 F7 RILs, we measured phenotypic values for all agronomic traits, including days to heading, plant height, tiller number, panicle diameter,

panicle length, flag leaf length, flag leaf width and yield per plant. Days to heading was counted when the first panicle emerged from the flag leaf sheath after seeding in individual plant. Panicle diameter, panicle length, flag lead length, flag leaf width and tiller number were evaluated when panicles were fully emerged. Panicle diameter was measured by a digital caliper at the area of ~0.5 cm below the panicle base. Plant height was measured from the soil surface to the apex of the tallest panicle. On the main panicle, flag leaf length was measured from the junction of leaf blade and sheath to the leaf apex, and flag leaf width was measured at the widest location of the leaf. Tiller number was scored when grains fully matured, while unproductive tillers (tillers without grains) were excluded from counting. Of all the flowered tillers of an individual plant, the longest panicle was measured for panicle length. Yield per plant was scored by measuring grain weight from all panicles in each plant.

Molecular markers for genotyping

Genomic DNAs were extracted from leaf tissues at the

maximum tillering stage by a CTAB method (Murray and Thompson 1980). Of 125 simple sequence repeat (SSR) and sequence-tagged site (STS) markers used for genotyping, 64 markers are RM-series that were designed according to Temnykh et al. (2001), and 33 markers are the S-series designed based on the sequence differences between japonica and indica rice, using information available from the Crop Molecular Breeding Lab, Seoul National University (unpublished). The other 28 markers are NGS-based insertion/deletion (NID) markers designed by sequence comparison of insertions/deletions (InDels) between parent genotypes. The size difference of polymorphic NID markers between parents was around 7~11 bp, and flanking sequences were used for primer design of the markers.

Linkage map construction and data analysis

Molecular linkage map was constructed using Mapmaker 3.0 (Lander et al. 1987; Lincoln et al. 1993). Distance between markers was presented in Centimorgan (cM) using Kosambi map fuction (Kosambi, 1944), and order of markers was established by

using three point linkage analyses. QTL analysis was conducted with the composite interval mapping (CIM) implemented in software Qgene 4.3.10 (Joehanes and Nelson 2008). Significance thresholds for QTL detection were determined with 1,000 permutations. The QTLs explained more than 10% of phenotypic effect were defined as main-effect QTLs.

High-throughput whole genome re-sequencing (WGS)

The WGS of the two parents, SNU-SG1 and Milyang23, was performed. The Illumina Genome Analyzer II were used to generate short reads and called by the Sequence Analysis Software of Pipeline version 1.4 of the Genome Analyzer (Illumina, Inc., San Diego, CA). Genomic DNAs were extracted from leaf tissues of 4-week old plants grown in field using QIAquick PCR Purification Kit (Qiagen). Full sequencing was processed in National Instrumentation Center for Environmental Management (NICEM) in Seoul National University, Korea. We ultimately gained 26x and 31x in fold coverage as well as 9 Gbp and 11 Gbp in read length for SNU-SG1 and Milyang23, respectively (Table S1, S2).

Then, the 75 bp paired-end reads of the two parents were mapped to the japonica cv. Nipponabre reference genome. Finally, the low-quality bases (Q score in scale < 20) and sites with conflicting genotypes among reads were excluded, and only the reads aligned to unique locations in the reference genome were used for sequence construction. Total number of SNPs was 3,202,922 which were added up from 163,933 of SNU-SG1 and 3,038,989 of Milyang23.

Identification of candidate genes for the QTLs using NGS data

Of the 46 QTLs, 20 main-effect QTLs with high LOD and R2 value were used for candidate gene analysis. The candidate genes harboring mutated sequences such as SNPs and InDels were identified by following process. First, a reference sequence corresponding each main-effect QTL was transferred from GRAMENE database (http://www.gramene.org/) onto text in Microsoft Word, a word processing program. Additional 0.5 Mb sequences from the two flanking DNA markers covering the target

QTL were analyzed to increase accuracy. Coding sequence regions derived from SNU-SG1 and Milyang23 were interrogated to detect mutated sequences using AutoHotkey macro program (http://www.autohotkey.com/). Among the genes harboring mutated sequences, only the genes whose mutation caused amino acid changes were selected as the candidate genes.

Results

Phenotypic variation in F7 RILs

178 F7 RILs developed from the cross of Milyang23 and SNU-SG1 (Yoo et al. 2007) were used for QTL analysis for agronomic traits. Phenotypic variations of the eight agronomic traits were evaluated in two parental lines and the RILs (Fig. 1). Most of the traits showed approximately normal distribution with bidirectional transgressive segregation except for plant yield, which displayed a biased distribution. A large amount of variation was observed between the two parents for most of the traits, while plant height and tiller number showed less or no significance in difference. Statistical analysis of eight agronomic traits are evaluated and displayed in Table S3.

The correlation among the eight agronomic traits was analyzed in the F7 RIL population (Table 1). Positive correlation was observed between flag leaf length and all the other traits except for flag leaf width, while days to heading displayed significant negative correlation with flag leaf width and tiller number.

Yield per plant was positively correlated with plant height, panicle length, flag leaf length and tiller number, but not with flag leaf width, panicle diameter and days to heading, suggesting that yield is tightly regulated by increased plant height and panicle number in this genetic background.

Table 1. Correlations of the eight agronomic traits analyzed in the F_7 recombinant inbred lines.

	PH	PL	FLL	FLW	PD	TLN	DH
PL	0.64***						
FLL	0.29***	0.31***					
FLW	-0.02	0.02	0.08				
PD	0.11	0.16*	0.20**	0.53***			
TLN	0.32	0.07	0.37***	-0.06	-0.19**		
DH	-0.09	0.00	0.36***	-0.23***	-0.06	-0.24***	
YP	0.47***	0.35***	0.18**	0.04	0.12	0.51***	-0.07

Significance level *** P < 0.001, **0.01, *0.05, respectively

PH plant height, PL panicle length, FLL flag leaf length, FLW flag leaf width, PD panicle diameter, TLN tiller number, DH days to heading, YP yield per plant

Table S3. Descriptive statistics of the eight agronomic traits for the parents and F_7 RILs

Trait	Parents		F_7 RIL population		
	SNU-SG1	Milyang23	Average	Range	SD[a]
Plant height (cm)	93.0	101.0	101.0	52 - 144	22.4
Panicle length (cm)	16.1	28.0	22.0	10 - 34	4.62
Panicle diameter (cm)	1.9	2.4	2.2	0.7 - 4.1	0.48
Flag leaf length (cm)	17.9	33.8	29.0	15 - 51	9.25
Flag leaf width (cm)	1.6	2.0	1.7	1.0 - 2.8	0.35
Tiller number (cm)	10.0	11.9	9.0	2 - 20	4.98

[a] Standard deviation

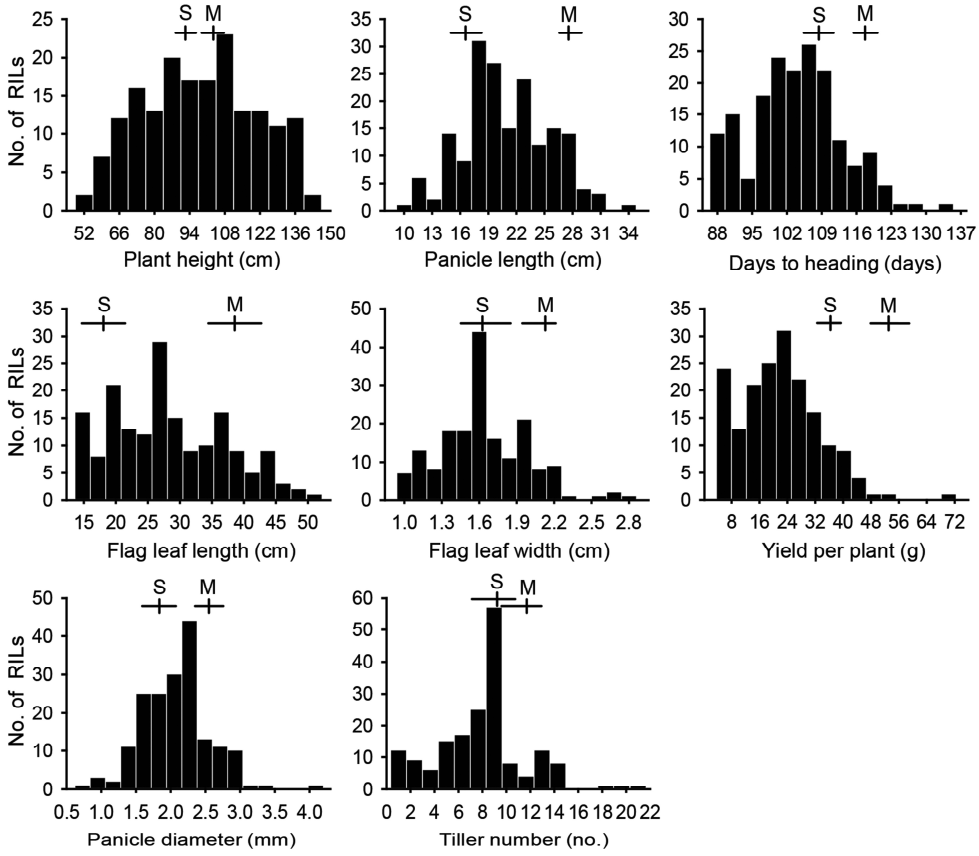

Figure 1. Frequency distribution of the eight agronomic traits in the 178 F_7 RILs derived from the cross of SNU-SG1/Milyang23. The vertical axis of each figure means the number of F_7 RILs. Means and ranges of phenotypic values for each trait were marked at the top of each histogram; S, SNU-SG1 and M, Milyang23

Linkage map construction

Linkage map was constructed using 178 F7 RILs generated from the cross between two parent lines, Milyang 23 and SNU-SG1. 125 polymorphic markers were used for QTL mapping, which include 64 RM-series SSR markers, 33 S-series SNP markers (unpublished) and 28 NID markers. A coarse-scale linkage map was initially constructed over rice whole genome using 97 SSR and STS markers. For further linkage map construction in unlinked regions, 28 NID markers were designed by using sequence data obtained from WGS. The genetic linkage map by using Mapmaker 3.0 covered 2,450.7 cM and consisted of 12 linkage groups (Fig. 2). Average distance between adjacent markers was 18.5 cM, which is less than 20 cM, the minimal compulsory level for QTL mapping (Lander and Botstein 1989).

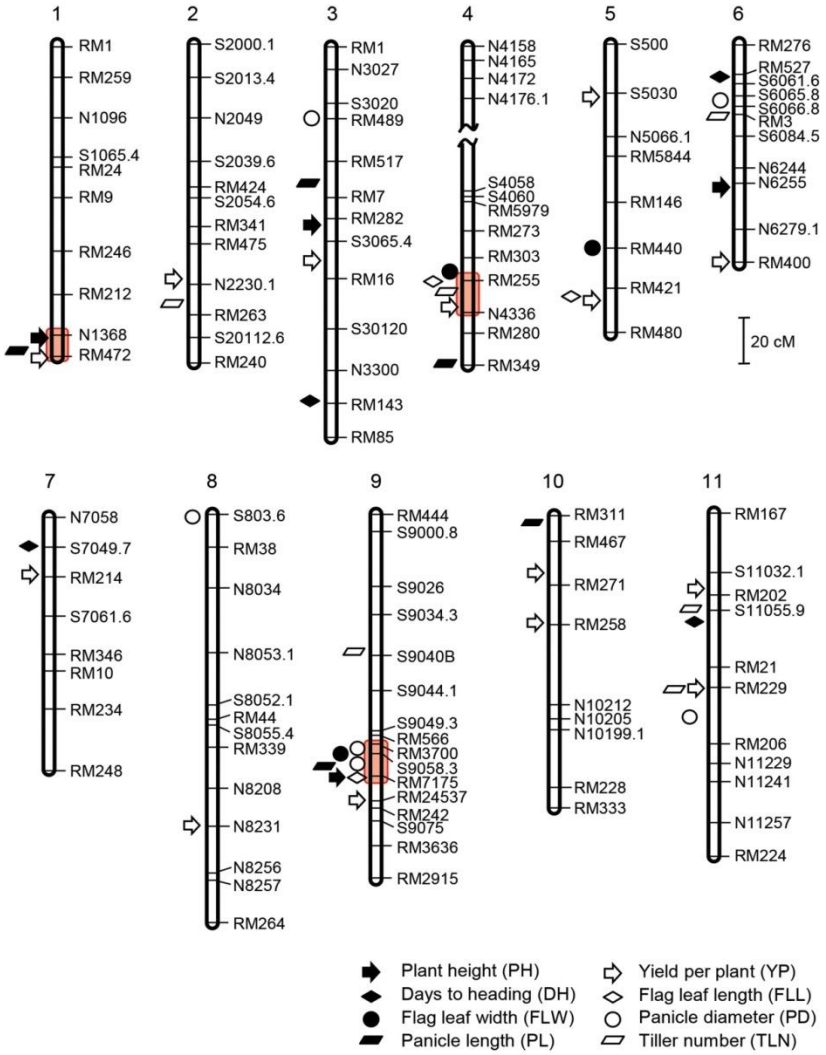

Figure 2. Chromosomal locations of QTLs detected by Q-genes in the F_7 RIL population of SNU-SG1/M23. Chromosomes are numbered at the top and markers are listed on the right of each chromosome. Wave marks in chromosome 4 represent the

unlinked region. Various geometric figures indicate the location of peak LODs of the eight agronomic traits. Chromosome 11 was not shown because QTL were not detected on it.

QTL analysis for the eight agronomic traits

The LOD value was calculated by 1,000 permutation test for each QTL, and LOD threshold for each QTL was ranged from 2.24 to 3.03 with 95% confidence. A total of 46 QTLs was identified for the eight traits, with 7.4 to 33.5% of phenotypic effect (R2) and 2.9 to 15.8 LOD values, across the whole chromosomes except chromosome 12 (Table 2). Note that chromosome 12 was excluded in Fig. 2, because it did not carry any QTL above the threshold. The QTL with highest LOD (15.8) and phenotypic effect (33.5%) was detected for plant height on the long arm of chromosome 1 (Table 2, Fig. 2). Overall, 3 to 14 QTLs were detected for each agronomic trait. Of the 46 QTLs, 20 QTLs explained more than 10% of phenotypic effect and were defined as main-effect QTLs. These main-effect QTLs identified for the eight traits include as follows: Fll4 for flag leaf length, Flw5 and Flw9 for flag leaf width, Dh6 for days to heading, Pd9-1 for panicle diameter, Pl1 and Pl9 for panicle length, Ph1, Ph3 and Ph9 for plant height,

Tln2, Tln11-1 and Tln11-2 for tiller number, and Yp1, Yp2, Yp3, Yp4, Yp9, Yp10-1 and Yp10-2 for yield per plant (Table 2). Notably, the 6 QTLs (Dh6, Pd9-1, Pl9, Ph1, Tln2 and Yp2) displayed more than 15% phenotypic effect. For yield per plant, 14 QTLs above the threshold were detected on all the chromosomes except for chromosome 12; among them, 7 QTLs were main-effect QTLs (Table 2), suggesting that these genetic materials are valuable for the identification of yield-related genes that were spread out over the genome.

QTL hotspots, genomic locations affecting many traits, are biologically important since they may harbor critical regulators for various traits (Neto et al. 2012). Three QTL hotspots were identified on three chromosomes for several traits. Maximum QTL hotspot was detected in a window with around 8 cM in size on the long arm of chromosome 9, which carries four main-effect QTLs (Flw9, Ph9, Pd9-1 and Pl9) for flag leaf width, panicle height, panicle diameter and panicle length and two minor-effect QTLs (Pd9-2 and Fll9) for panicle diameter and flag leaf length (Fig. 2). The QTL hotspots on chromosomes 1 and 4 were identified in the windows ranging in size of 3 to 5 cM mostly for the correlated traits;

i.e. plant height, panicle length and yield per plant for chromosome 1, and flag leaf width, flag leaf length, tiller number and yield per plant for chromosome 4 (Fig. 2).

To compare physical location of the QTL identified in this study and previously reported QTL, we searched the literatures in which mapping populations derived from crosses between indica and japonica cultivars were used for QTL identification. Physical locations of the QTL detected in the current study were compared with those of the QTLs in the literatures, whose physical location was previously determined or calculated by using flaking DNA markers. As a result, 22 out of 46 QTLs identified in this study fell into the chromosomal regions containing the previously identified QTL (Table 3). The matching QTL for the remaining 24 QTL were not found in the literature, indicating that these are potentially novel QTLs. For all the traits used in this study, novel QTLs were identified except for days to heading (Table 2, 3). In addition, the QTL for panicle diameter are all newly discovered. Among all the novel QTLs identified, 12 QTLs were main-effect QTLs with above 10% of phenotypic effect, which include Fll4, Flw5, Pd9-1, Pl9, Ph3, Tln11-1, Tln11-2, Yp2, Yp4, Yp9, Yp10-1 and Yp10-2 (Table 2).

Table 2. QTL identification for the eight agronomic traits by composite interval mapping using 178 F_7 recombinant inbred lines

Trait	QTL	Chr.[a]	Markers[b]	A^c	LOD	LOD peak (cM)	R^{2d} (%)	Permutation[e] 95%	Permutation[e] 99%
Flag leaf length	*Fll4*	4	RM255	2.08	4.35	202.7	10.7	2.49	2.97
	Fll5	5	RM421-RM480	2.43	3.04	181.8	7.6	2.26	2.81
	Fll9	9	RM7175	-2.64	3.68	143.4	9.1	2.42	2.92
Flag leaf width	*Flw4*	4	RM303-RM255	0.12	3.47	195.3	8.6	2.62	3.50
	Flw5	5	RM146-RM440	-0.13	4.21	145.5	10.3	2.28	3.03
	Flw9	9	S9058.3	0.08	4.12	131.9	10.1	2.61	3.45
Days to heading	*Dh3*	3	N3300-RM143	-2.49	2.88	266.1	7.2	2.62	3.34
	Dh6	6	RM527-S6061.6	5.21	8.83	21.9	20.4	2.46	3.21
	Dh7	7	S7049.7	-0.76	3.01	20.2	7.5	2.35	3.07
	Dh11	11	S11055.9-RM21	2.62	2.99	59.5	7.4	2.62	3.36
Panicle diameter	*Pd3*	3	RM489	0.14	3.75	54.9	9.2	3.03	4.05
	Pd6	6	S6065.8-S6066.8	-0.19	2.96	42.5	7.4	2.60	3.77
	Pd8	8	S803.6-RM38	-0.14	2.87	2.8	7.1	2.38	3.35
	Pd9-1	9	RM3700-S9058.3	-2.28	9.30	129.7	21.4	2.37	3.02
	Pd9-2	9	S9058.3-RM7175	0.12	3.54	143.0	8.8	2.37	3.02
	Pd11	11	RM229-RM206	0.12	2.91	95.0	7.2	2.63	3.73
Panicle length	*Pl1*	1	N1368-RM472	1.43	4.31	224.6	10.5	2.58	3.26
	Pl3-1	3	RM517-RM7	-1.16	3.53	113.0	8.7	2.58	3.26
	Pl3-2	3	S3120-N3300	-1.17	2.88	218.4	7.2	2.58	3.26
	Pl4	4	RM349	0.39	3.34	269.3	8.3	2.56	3.16
	Pl9	9	S9058.3-RM7175	-2.39	9.60	136.1	22.0	2.37	3.02
	Pl10	10	RM311-RM467	0.79	3.15	9.7	7.8	2.24	2.82
Plant height	*Ph1*	1	N1368-RM472	14.79	15.8	222.7	33.5	2.38	3.18
	Ph3	3	RM282- S3065.4	-7.21	4.45	130.6	10.9	2.45	3.06
	Ph6	6	N6255-S6279.1	-6.73	4.03	106	9.9	2.36	2.95
	Ph9	9	RM7175	-8.30	5.68	143.4	13.7	2.39	3.16
Tiller number	*Tln2*	2	N2230_1-RM263	1.45	6.33	184.7	15.3	2.49	3.32
	Tln4	4	RM255-N4336	-1.30	3.36	212.6	8.4	2.85	3.45
	Tln6	6	RM3	0.74	3.79	51.1	9.4	2.64	3.70
	Tln9	9	S9034.4-S9040B	0.37	3.01	76.1	7.6	2.57	3.49
	Tln11-1	11	RM202-S11055.9	-1.18	4.33	48.7	10.7	2.45	3.34
	Tln11-2	11	RM229	-1.22	5.61	94.3	13.7	2.45	3.34
Yield per plant	*Yp1*	1	N1368-RM472	3.99	4.65	224.5	11.3	2.70	3.82
	Yp2	2	RM475-N2230_1	6.50	9.96	169.5	22.7	2.56	3.70

Yp3	3	S3065.4-RM16	-1.55	4.80	162.8	11.7	3.02	4.39
Yp4	4	RM255-N4336	-1.32	6.11	227.7	14.6	2.86	3.94
Yp5-1	5	S5030	0.11	3.67	37.0	9.0	2.29	2.29
Yp5-2	5	RM421-RM480	1.79	3.64	182.0	9.0	2.29	2.29
Yp6	6	S6279.1-RM400	0.89	3.78	162.0	9.3	2.55	3.94
Yp7	7	S7049.7-RM214	-2.21	3.61	35.5	8.9	2.42	4.12
Yp8	8	N8231	0.67	3.30	170.8	8.2	2.53	3.41
Yp9	9	RM24537	0.38	4.44	155.4	10.8	2.77	3.00
Yp10-1	10	RM467-RM271	-0.50	4.38	39.3	10.7	2.40	3.51
Yp10-2	10	RM258	-0.51	4.64	60.4	11.3	2.40	3.51
Yp11-1	11	S11032.1-RM202	-2.62	3.15	41.0	7.8	2.29	3.11
Yp11-2	11	RM229	-2.87	3.41	94.3	8.5	2.29	3.11

[a, b] Chromosome number and marker intervals

[c] Positive and negative values indicate additive effects contributed by SNU-SG1 and Milyang23 alleles, respectively

[d] Phenotype variation rate explained by detected QTL

[e] LOD thresholds raised by permutation tests, p = 0.05 and 0.01

Table 3. Comparison of the QTLs identified in current and previous studies using the mapping populations derived from the crosses of *indica* and *japonica* cultivars

Traits	QTL of this study		QTL of previous studies						
	QTL	Physical position (Mb)	QTL	Marker region	Physical position (Mb)	PT[a]	PS[b]	Mapping parents Japonica(J), Indica(I)	Reference
Days to heading	*Dh3*[c]	30.0-33.1	*Hd6*	R3226	33.2	BC_4F_2	100	Nipponbare(J), Kasalath(I)	Yamamoto et al. (2000)
	Dh6	9.8-11.1	*Hd-1*	R1679	8.9	F_2	186	Nipponbare(J), Kasalath(I)	Yano et al. (1997)
			qHd6-1	S2539-R2123	9.3-11.6	BIL	182	Koshihikari(J), Kasalath(I)	Zhang et al. (2008)
	Dh11	11.9-16.0	*qHd-11*	RM202-RM287	9.0-16.7	RIL	184	Koshihikari(J), Guichao2(I)	Zhang et al. (2008)
Plant height	*Ph1*	36.8-39.6	*QTLph1*	C86-C742	39.5-42.7	BIL	98	Nippnbare(J), Kasalath(I)	Ishimaru et al. (2004)
			Qph1.2	E60551-RM1387	35.0-40.5	F_2	301	Nipponbare(J), IR1545-339(I)	Lin et al (2011)
	Ph6	25.5-27.9	*Qph6.2*	RM3879-RM340	25.9-28.5	F_2	301	Nipponbare(J), IR1545-339(I)	Lin et al (2011)
			Qph6f	RM162-RM30	24.0-27.3	RIL	226	Zhonghui9308, Xieqingzao B(I)	Liang et al (2011)
	Ph9	16.8-18.1	*Qph9*	RM3912-RM278	10.8-27.1	F_2	301	Nipponbare(J), IR1545-339(I)	Lin et al (2011)
Flag leaf length	*Fll5*	23.9-27.3	*qFLL-5*	BIN	27.8-28.6	RIL	150	Nippnbare(J), 9311(I)	Wang et al. (2010)
			QFll5	RM480-RM334	27.5-28.6	RIL	180	IRAT109(J), Zhenshan 97(I)	Yue et al. (2006)
	Fll9	15.9-16.8	*qFll9*	RM24424-RM24434	16.2-16.5	F_2	176	SN265(J), LTH(J)	Shunkun Jiang (2010)
Flag leaf width	*Flw4*	28.5-30.7	*qFlw4*	RM17483-RM17486	31.0-31.2	RIL F_7	190	D50(J), HB277(I)	Chen et al. (2012)
			Qflw4	RM255-RM349	28.5-31.6	RIL $F_{9/10}$	180	IRAT109(J), Zhenshan97(I)	Bing et al. (2006)
	Flw9	15.9	*qFLA9*	RM6839B-	14.5-17.9	BRIL	244	9311(I), Zhenshan97B(I)	Wang et al. (2012)

				RM257					
Tiller number	*Tln2*	23.0-25.8	*Tln2-2*	RM526–RM425	26.6-32.2	SSSL	78	IAPAR9(J), Hua-Jing-Xian74(I)	Liu et al. (2012)
			tp2b	RM561-RM6318	19.8-24.4	BCRIL	244	Zhenshan97B(I), 9311(I)	Zhen et al. (2012)
	Tln4	30.7-33.6	*qTLN4*	RM303-RM349	27.5-32.4	RIL-F_{11}	254	Xiushui 79(J), CBao(J)	Jiang et al. (2012)
			Tln 4-3	RG143-RG620	34.2-34.6	DH[d]	123	Azucena(J), IR64(I)	Yan et al. (1998)
	Tln6	19.4	*pn6*	S6070	20.1	F_2	146	Junambyeo(J), IR71033(I)	Rahman et al. (2008)
	Tln9	10.5-12.6	*Tln9*	RM105–RM278	12.5-19.3	SSSL	91	Hua-Jing-Xian74(I), Amol3(I)	Liu et al. (2012)
Yield per plant	*Yp1*	36.8-39.6	-	C547-C2340	34.4-40	BCRIL	240	Zhenshan97(I), Minghui63(I)	Zhang et al. (2010)
	Yp3	14.4-23.1	-	C1677, C136	15.6, 26.7	BC_1F_7	98	Nipponbare(J), Kasalath(I)	Ishimaru et al. (2001)
	Yp5-1	3.6	*GS5*	RM413-RM574	2.2-3.4	DH	92	Zhenshan97(I), H94(I)	Li et al. (2011)
	Yp5-2	23.9-27.3	-	C624-C246	21.3-27	F_2, F_3	240	Zhenshan97(I), Minghui63(I)	Zhang et al. (2010)
	Yp6	27.9-28.4	*tgw6*	C358	26.1	BC_1F_5	98	Kasalath(I), Nipponbare(J)	Ishimaru et al. (2001)
	Yp7	9.4-12.7	-	C1023-R1440	7.2-16.8	RIL	240	Zhenshan97(I), Minghui63(I)	Zhang et al. (2010)
Panicle	*Pl1*	36.8-39.6	*qPL-1*	RM472–RM104	37.8-40.1	RIL	187	IRAT109(J), Zheshan97B(I)	Liu et al. (2008)
length	*Pl4*	32.4	*QPl4b*	RM349-RM280	32.4-34.9	BC_2F_5	59	Tarome Molaei(J), IR64(I),	Ahamadi et al. (2008)

[a] PT : Population type
[b] PS : Population size
[c] Numbers at the end of QTL represent chromosome number
[d] DH : Double haploid population

NGS data analysis for candidate gene identification

To obtain cultivar-specific nucleotide polymorphism data set, WGS was performed for the two parent cultivars, SNU-SG1 and Milyang23 (see Methods). A japonica cv. Nipponbare, whose whole genomic sequence was determined by International Rice Genome Sequencing Program (2005), was used as a reference sequence. The polymorphic nucleotide sequences including InDels were obtained by comparing the whole genome sequences of SNU-SG1 and Milyang23 with the reference sequence using Pipeline version 1.4 (see Methods). Of the total 363,845 SNPs, 330,317 were identified from Milyang23, and 33,528 from SNU-SG1 by the comparison with the reference sequence. Large numbers of SNPs found in Milyang23 were derived from sequence variation between indica- and japonica-type subspecies. We analyzed all the SNPs derived from SNU-SG1 for candidate gene analysis, and however only InDels capable of impairing protein function were considered for Milyang23 to filter out non-causal sub-specific sequence variation between indica and japonica genomes.

For the identification of candidate genes for each QTL, 20 main-effect QTLs with above 10% phenotypic effect for the eight agronomic

traits were selected out of total 46 QTLs identified. Accuracy was increased by analyzing additional 0.5 Mb sequences from two flanking DNA markers covering each QTL region. To find out specific locations of polymorphic sequences in the individual genes, these SNP sequence regions were blasted to the Nipponbare reference sequence, which was obtained from Gramene database. By these sequence analyses, 96 genes were found to carry 1,261 nucleotide sequence polymorphisms in open reading frame within the 20 main-effect QTL regions (data not shown). Among them, only the 497 SNPs which caused amino acid variation were considered as candidate genes. Conclusively, by application of WGS approach, we isolated 96 candidate genes carrying at least one amino acid polymorphism for eight agronomic traits.

Table S1. Information of total reads and resequencing data coverage of SNU-SG1 and Milyang23 over the reference genome

Target genome	Raw data		Trimmed[a] data			Coverage (X)
	Reads No.	Reads length[b]	Reads No.	Reads length[b]	%	
SNU-SG1	115,761,304	14,007,117,784	90,178,800	9,758,448,160	69.7	26
Milyang23	1147,258,190	17,818,240,990	111,632,782	11,858,868,551	67	31

[a] Trimming necessary to create a data file for analysis by comparison to known databases of DNA sequences

[b] Unit is base-pair (bp)

Table S2. The profiles of whole genome re-sequencing of the two parents, SNU-SG1 and Milyang23 analyzed with Nipponbare as the reference genome

No. of Chr.	Nipponbare Genome (bp)[a]	SNU-SG1 re-sequencing data			Milyang23 re-sequencing data		
		Number of reads	Reads length (bp)[b]	Cover age (X)	Number of reads	Reads length (bp)[b]	Cover-age (X)
1	43 261 740	10 284 529	1107 933 161	25.61	11 935 581	345 654 426	28.96
2	35 954 743	9 137 367	976 530 820	27.16	10 176 270	302 336 982	29.71
3	36 192 742	8 351 861	899 751 566	24.86	10 037 792	292 200 125	29.11
4	35 498 469	9 107 142	977 272 852	27.53	12 072 277	430 618 121	35.67
5	29 737 217	6 940 309	745 809 402	25.08	8 803 794	272 477 424	30.95
6	30 731 886	7 363 630	791 960 702	25.77	8 601 868	252 378 807	29.34
7	29 644 043	6 850 936	738 136 671	24.9	7 633 219	205 867 916	26.97
8	28 434 780	6 831 347	736 176 454	25.89	8 491 613	266 381 900	31.37
9	22 696 651	5 068 564	545 173 557	24.02	5 927 914	162 187 727	27.36
10	22 685 906	6 292 860	678 308 589	29.9	8 808 716	360 012 223	40.87
11	28 386 948	6 143 634	662 267 497	23.33	6 859 901	173 624 094	25.31
12	27 566 993	6 818 252	734 109 024	26.63	9 291 276	328 539 519	35.36
ave[c]				25.87			30.91

[a] Nipponbare reference genome (GenBank accession-AACV000000000.1)

[b] Length of the consensus sequence

[c] Total number except for the column of coverage and sequencing depth, which are average value

Strong candidate genes for the main-effect QTLs

Among the 96 candidate genes identified by sequence analysis using NGS data, 17 genes were selected as strong candidates, because they harbor the casual SNPs leading to premature stop codon or frame shift in their amino acid sequences (Table 4). These 17 strong candidate genes were identified for all agronomic traits except for panicle diameter, with one to seven genes for each trait; single genes for panicle length, plant height, flag leaf width and days to heading, two genes for flag leaf length, seven genes for tiller number, and eight genes for yield per plant (Table 4). Of the six candidate genes for Tln2, five genes were identified in 2.8 Mb genomic region on chromosome 2, and especially the three genes (Os02g40190, Os02g40200 and Os02g40240) in the region of Tln2 have been known to function as receptor-like protein kinase precursor (Table 4). Interestingly, one candidate gene (Os01g68510), encoding peptide transporter 2 (PTR2), was found in QTL hot spot zone located on the distal part of chromosome 1, on which three QTLs for panicle length, yield per plant and plant height, were identified (Table 4). The largest effect QTL Ph1 for plant height was also detected with 33.5% R2 value in this hot spot of chromosome 1 (Table 2, 3). Short Panicle

1 (SP1) encoding a putative peptide transporter (PTR) has been reported to regulate inflorescence branch elongation; a knock-down mutant sp1 is defective in rice panicle elongation, producing short-panicle phenotypes (Li et al. 2009). In addition, the Arabidopsis nitrate transporter NRT1.7 (Fan et al. 2009), an ortholog of PTR2, plays an important role in source-to-sink remobilization of nitrate. These suggest that PTR2 is a promising candidate gene related to the three QTLs detected for panicle length, plant height and yield per plant in this study. Another candidate gene, Heading date 1 (Hd1), was found in the region of main-effect QTL Dh6 (R2 = 20.4%) for days to heading on chromosome 6. Hd1 controls flowering by regulating Hd3a expression in response to photoperiod (Yano et al. 2000; Hayama et al. 2003). By sequence analysis, we found 4-bp deletion in the Milyang23 alleles which causes premature stop codon (Fig. 3c).

We further investigated the orthologs of these candidate genes in model plant Arabidopsis using Gramene (http://www.gramene.org/) and TAIR (http://www.arabidopsis.org/) databases, and found five orthologs of the genes (Table 4). CONSTANS (CO), a regulator of the florigen, and NRT1.7, a low affinity nitrate transporter, are the Arabidopsis orthologs of Hd1 and PTR2, respectively. The other

genes, At4g25820, At3g53180 and At2g07050, are found to be the orthologs of Os04g51450, Os10g31820 and Os11g18194, respectively. At4g25820, an ortholog of Fll4- and Yp4-related candidate gene (Os04g51450) has been known to encode Xyloglucan Endotransglycosylase 9 (XTR9), which involves carbohydrate metabolic process (Paul and Janet 1999). Yp10-related gene (Os04g51450) is an ortholog of At3g53180 encoding Nodulin/Glutamine Synthase-like protein, and At2g07050, an ortholog of Tln11-2-related gene Os11g18194, was reported to encode Cycloartenol Synthase 11 (Corey et al. 1993), which functions in biosynthesis of brassinosteroids (Asami and Yoshida 1999). These information for promising candidate genes would be useful genetic resources for the further isolation of each QTL-associated genes.

Table4. List of 17 strong candidate genes for the main-effect QTLs obtained by sequence analysis with whole genome re-sequencing data of two parents

QTL	Type[a]	Gene ID	Mutation[b]	Description	Ortholog[c]
Pl1,Yp1,Ph1	Milyang23	Os01g68510	133/601	Peptide transporter PTR2	AT1G69870 (NRT1.7)
Yp2	SNU-SG1	Os02g34800	103/270	Leucine-rich repeat receptor protein kinase EXS precursor	-
Tln2	SNU-SG1	Os02g40100	398/791	DUF869 domain containing protein	-
Tln2	SNU-SG1	Os02g40190	33/240	Receptor kinase	-
Tln2	SNU-SG1	Os02g40200	485/1033	Receptor-like protein kinase precursor,	-
Tln2	SNU-SG1	Os02g40240	675/1002	Receptor-like protein kinase precursor,	
Tln2	SNU-SG1	Os02g40280	273/370	Piwi domain containing protein	-
Tln2	SNU-SG1	Os02g41610	642/745	Expressed protein, Coiled coil	-
Yp3	SNU-SG1	Os03g24690	87/601	Terpene synthase	-
Yp3	SNU-SG1	Os03g29150	49/270	NAD dependent epimerase/dehydratase family protein	-
Fll4, Yp4	SNU-SG1	Os04g51450	35/331	Glycosyl hydrolases family 16	AT4G25820 (XTR9)
Fll4, Yp4	SNU-SG1	Os04g53120	6 A.S.	XA1, NB-ARC domain containing protein	-
Flw5	SNU-SG1	Os05g31230	5 A.S.	N-acetyl transferase ESCO1	-
Dh6	Milyang23	Os06g16370	F.S.	Hd1, CCT/B-box zinc finger protein	AT5G15840 (CO)
Yp9	SNU-SG1	Os09g31000	F.S.	EF hand family protein	-
Yp10	Milyang23	Os10g31820	&823L	FluG protein, similar to Nodulin 6l	AT3G53180 (NODGS)
Tln11-2	Milyang23	Os11g18194	560/761	Cycloartenol synthase	AT2G07050 (CAS1)

[a] Parent type harboring mutant alleles, which were determined by sequence comparison with reference genome *Japonica* cv. Nipponbare

[b] The position of amino acid in which premature stop codon occurred /total protein length, A.S. amino acid substitutions, F.S. frame shift, & stop codon

[c] Accession number of *Arabidopsis* orthologs were obtained from The Arabidopsis Information Resource (TAIR, http://www.arabidopsis.org/)

Validation of QTL region using heterogeneous inbred family-near isogenic lines (H-NILs)

To further validate the QTL regions where these 17 strong candidate genes were identified, we developed H-NILs that segregate for the phenotypes of the corresponding traits (Tuinstra et al. 1997). The genotypic and phenotypic analyses of the progeny H-NILs allow finding the candidate genes underlying QTLs (Routaboul et al. 2012). In 178 F7 RILs, we selected the lines that are heterogeneous for at least one of the flanking marker loci in the QTL regions. By phenotypic analysis using 40 F7:8 individual plants derived from each heterogeneous line, we found several H-NILs for 13 main-effect QTLs that harbor strong candidate genes proposed (Table 5). Most of the H-NILs including F7 RIL #214-derived ones, which showed highest variation in tiller number, displayed transgressive segregation (data not shown). These strongly support that the target genes are localized within the corresponding QTL regions and the strong candidate genes proposed in these regions are highly valuable. Further linkage analysis of the markers and traits in the H-NIL mapping populations would lead to isolation of the genes underlying the QTLs.

Table 5. $F_{7:8}$ heterogeneous inbred family-near isogenic lines (H-NILs) displaying phenotypic segregation for each trait

Trait	QTL	Heterogeneous F_7 RILs[a]
Panicle length	*Pl1*	21, 255
Yield per plant	*Yp1*	255
Plant height	*Ph1*	21, 255
Yield per plant	*Yp2*	17, 21, 184, 236, 255, 266
Tiller number	*Tln2*	8, 17, 21, 72, 148, 236, 255
Yield per plant	*Yp3*	21, 78, 88, 236
Flag leaf length	*Fll4*	21, 78
Yield per plant	*Yp4*	21, 78, 88, 255
Flag leaf width	*Flw5*	72, 78
Day to heading	*Dh6*	17, 298
Yield per plant	*Yp9*	72, 88, 148, 255
Yield per plant	*Yp10*	78, 88, 148, 183, 255
Tiller number	*T1n1-2*	21, 214, 236, 255

[a] Line numbers of heterogeneous F_7 RILs whose progeny H-NILs showed phenotypic segregation.

Discussion

Enhanced efficiency in QTL analysis with the application of next-generation sequencing approach

Agronomically important crop traits are mostly controlled by multiple genes (QTL) with various phenotypic variations. Conventional QTL analysis has been time-consuming and labor-intensive mainly because it requires polymorphic marker development for linkage analysis. To overcome this limitation, NGS technology has recently been introduced as new strategies for genetic mapping by several research groups. A high-throughput method was developed for genotyping recombinant populations by performing whole genome resequencing of individual RILs (Huang et al. 2009). This genome sequence-based genotyping method was faster in genetic map construction and more accurate in determination of recombination breakpoints than marker-based genotyping method (Wang et al. 2011). Another group performed whole genome resequencing of 50 rice cultivars and investigated genome-wide variation patterns in high-quality candidate SNPs obtained from resequencing data of each cultivar (Xu et al. 2012). Although these approaches reduced time

consumption for genotyping, they are still cost-intensive because of high-throughput WGS. Thus, in this study we modified conventional mapping method by applying WGS to only two parental cultivars to improve QTL mapping efficiency.

The most time-consuming procedure in QTL mapping is to develop polymorphic DNA markers for linkage analysis. In this study, we developed NID markers by using WGS data, which were designed based on nucleotide sequence polymorphism between two parents. Only the target regions, ranging from 120 to 250 bp, containing more than 8 bp InDels between two parents were considered to develop NID markers. NID markers showed much higher efficiency (74.5%) in polymorphic marker development compared to those of two conventional SSR (18.1%) and STS (21.1%) markers, which were designed from sequence polymorphism between japonica and indica genomes (Fig. S1). These indicate that application of WGS approach to DNA marker development substantially increased the efficiency in QTL mapping compared to conventional one.

High-resolution mapping to identify candidate genes for detected QTL is another time-consuming and laborious step in conventional QTL analysis. Thus, instead of performing fine mapping to narrow

down candidate QTL region to less than ~100 kb, we analyzed average ~3 Mb sequences within two flanking markers covering the target QTL region to identify candidate genes. To exclude missing candidate genes, we included additional ~0.5 Mb genomic regions extended from the two flanking markers in QTL region for sequence analysis. Average five candidate genes harboring 23 causal SNPs were found in the 3 Mb-QTL region from sequence analysis by comparing genomic sequences of SNU-SG1 and Milyang23 with the reference Nipponbare. In conventional QTL analysis, to obtain these number of candidate genes, fine mapping has to be conducted to narrow down QTL region to 50 to 100 kb through development of large numbers of polymorphic DNA markers in each QTL region as well as development of mapping populations such as heterogeneous inbred family-near isogenic lines (Tisne et al. 2008) and backcross inbred lines (Li et al. 2011), which are time-consuming and laborious work. Furthermore, for candidate analysis in conventional QTL analysis, high resolution-mapping population has to be generated for each QTL, and however, WGS-based QTL analysis enabled to isolate candidate genes for all the QTLs identified from initial mapping population at the same time. Therefore, we propose that this advanced QTL analysis approach by applying WGS to two parent

cultivars is efficient method for candidate gene identification as it skips fine-mapping procedure.

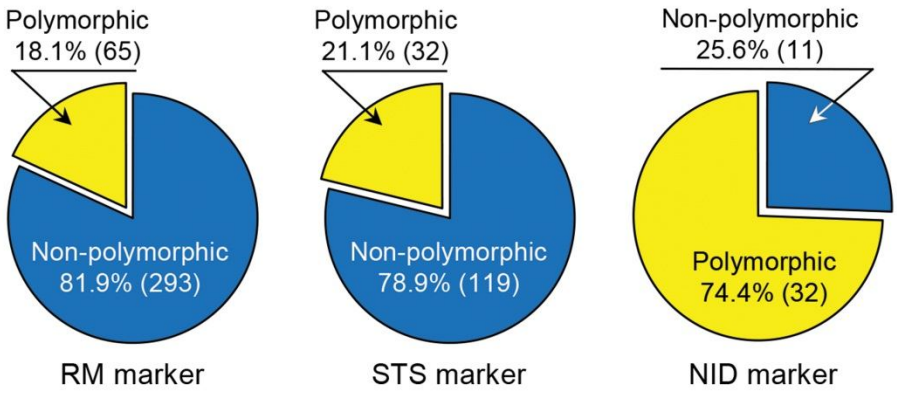

Figure S1. Comparison of efficiencies in polymorphic marker development between NID markers and two conventional SSR and STS markers

Comparison of the identified QTLs with previously reported ones for the agronomic traits

Some QTLs were commonly identified in current and previous studies for plant architecture and yield such as tiller number, panicle diameter and yield per plant. For example, the four QTLs, Tln2, Tln4, Tln6 and Tln9, influencing tiller number shared a similar genomic physical location with some QTLs identified from previous studies (Jian et al. 2012; Liu et al. 2012; Qu et al. 2012; Rahman et al. 2008; Yan et al. 1998) (Table 4). The other two QTLs, Tln11-1 and Tln11-2, have not been previously found in QTL research, indicating that these are noble QTLs. For yield per plant, eight of the total 14 QTLs were newly identified in this study, and the other six QTLs were located in the similar positions to the QTLs found in early studies (Ishimaru 2003; Ishimaru et al. 2001; Li et al. 2011; Xing and Zhang 2010) (Table4). This consistency in the QTL regions associated with the tiller number and yield per plant among studies indicates that the locations of alleles for the traits are conserved across different genetic and environmental backgrounds. On the other hand, all the QTLs for panicle diameter (Pd3, Pd6, Pd8, Pd9-1, Pd9-2 and Pd11) are novel and have no correspondences with previously reported QTLs for this trait (Table 2), indicating that these QTLs may be a potentially new set

of alleles specific for this genetic background.

Some candidate genes were previously reported to regulate similar agronomic traits used in this study

By sequence analysis using WGS data, we found 96 candidate genes in all the detected QTL regions related to eight agronomic traits. Of the 96 candidate genes, 17 genes were isolated as the strong candidates, because they carry causal polymorphic SNPs, leading to frame shifts or premature stop codon in amino acid sequence (Table 4). Among the 17 strong candidates, five genes have been previously reported (Table 4). For example, in the Dh6 QTL region, nine genes containing polymorphic SNPs on their coding sequences were initially found by sequencing analysis (data not shown). Of them, 4-bp deletion on the coding region of Hd1 occurred in the Milyang23 allele, leading to premature stop codon (Fig. 3c), and however simple SNPs were found for the other eight genes. Moreover, Hd1 was considered to be the only flowering-related gene based on gene function database in Gremene (http://www.gramene.org/) and RiceXPro (http://ricexpro.dna.affrc.go.jp/). These suggest that Hd1 is the

strongest candidate gene for Dh6 QTL. In the same way, the other 16 strong candidate genes were isolated for six agronomic traits. Interestingly, only one gene PTR2 was isolated to be the strong candidate gene for three QTLs related to panicle length, plant height and yield per plant. Considering the importance of the gene(s) related to this QTL hot spot, we further searched candidate genes by interrogating whole expressed genes in the region with comparison of previously reported genes in the literature, and found SD1 as another promising candidate gene. The coding region of SD1 did not show any polymorphism in nucleotide sequence, but 36 polymorphic sequences were identified in the 1.8-kb promoter region upstream of the translation start site (Fig. 3). SD1 encodes a GA 20-oxidase (GA20ox), an enzyme involved in GA biosynthesis (Monna et al. 2002; Sasaki et al. 2002; Spielmeyer et al. 2002). Loss of function GA20ox mutant showed semi-dwarf and GA-deficient phenotypes (Monna et al. 2002; Spielmeyer et al. 2002). Thus, it is speculated that this sequence polymorphism in the promoter region causes differential expression of SD1 gene in two parent cultivars, resulting in variation in panicle length, plant height and yield per plant. Further, we found another gene GS5 in the region of Yp5-2 (Table 3), which was previously reported to encode a serine carboxypeptidase and function

as positive regulator of grain width, filling and weight (Li et al. 2011). However, no mutation was detected in both ORF and promoter regions of GS5 from the sequence analysis using NGS data, indicating existence of another candidate gene responsible for the yield-related QTL, Yp5-2.

Novel candidate genes whose functions were known in Arabidopsis

The functions of the unknown genes in rice have been predicted from the ones of orthologous genes in Arabidopsis, as molecular and biochemical functions of proteins are generally well conserved among species. For example, the functions of CO and FT, flowering promoters in photoperiod pathway, are well conserved between Arabidopsis and rice and most likely in other species as well (Hayama and Coupland, 2004). Thus, based on Arabidopsis functional information we provide three strong candidate genes for the QTLs related to tiller number, flag leaf length and yield per plant. Os11g18194, located in the region of Tln11-2, is a rice ortholog of Arabidopsis CAS1, encoding a cycloartenol synthase, which is involved in the biosynthesis of brassinosteroids (BR) (Corey et al.

1993). Mutations in the genes involved in the BR biosynthesis pathway modify tiller number (Fujii et al. 1991), leaf size (Morinaka et al. 2006) and yield per plant (Choe et al. 2001). Moreover, increased level of BR causes significant effect on the increasing tiller number and seed weight (Wu et al. 2008). Sequence analysis revealed that Milyang23 allele of the gene carrying one nucleotide insertion produces frame shift, leading to premature stop codon in amino acid sequence. These suggest that OsCAS1 is a promising candidate gene for Tln11-2 related to tiller number. Another candidate gene Os04g51450 located in the regions of Fll4 and Yp4, has an orthologous gene, At4g25820 (XTR9), encoding Xyloglucan endotransglucosylase/hydrolases (XTH), which regulate root growth and cell wall extension in Arabidopsis (Maris et al. 2009). One amino acid deletion occurred in SNU-SG1 allele caused premature stop codon, resulting in generating only 35 amino acids out of total 331 amino acids. This supports possible involvement of Os04g51450 for Fll4 and Yp4 regulating flag leaf length and yield per plant. Lastly, the other candidate gene Os10g31820 found in the region of Yp10 has an Arabidopsis ortholog AT3G53180, which encodes nodulin/glutamine synthetase-like protein (NodGS), a fusion protein composed of an N-terminal amiohydloase domain (homologous to nodulin) and C-

terminal prokaryotic glutamine synthetase type I domain. Downregulation of NodGS by RNAi plant gave rise to a short main root and disrupted root cap development (Doskocilava et al. 2011). As root development is critical for overall plant growth, the interruption of root development might affect rice yield. One amino acid insertion causing frame shift mutation occurred in Milyang23 allele. These suggest that Os10g31820 is a promising candidate gene for Yp10 influencing yield per plant.

In this study, we developed new approach with WGS application to two parental cultivars in DNA marker development for linkage analysis and candidate gene analysis for the detected QTLs. The efficiency was substantially increased in polymorphic DNA marker development by using WGS data of two parents. In addition, the application of WGS to candidate gene analysis greatly reduced the cost and time for candidate gene identification by skipping high-resolution mapping step. By QTL analysis, a total of 46 QTLs were detected for eight agronomic traits. The 26 novel QTLs, three QTL hotspots and 17 strong candidate genes identified in this study would be valuable resources for breeding of high-yielding new plant in rice.

Reference

Asami T, Yoshida S (1999) Brassinosteroid biosynthesis inhibitors. Trends Plant Sci 4:348-353

Choe S, Fujioka S, Noguchi T, Takatsuto S, Yoshida S, Feldmann KA (2001) Overexpression of DWARF4 in the brassinosteroid biosynthetic pathway results in increased vegetative growth and seed yield in Arabidopsis. Plant J 26:573-582

Corey EJ, Seiichi PT, Bartel B (1993) Isolation of an Arabidopsis thaliana gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen. Proc Natl Acad Sci USA 90:11628-11632

Doskocilova A, Plihal O, Volc J, Chumova J, Kourova H, Halada P, Petrovska B, Binarova P (2011) A nodulin/glutamine synthetase-like fusion protein is implicated in the regulation of root morphogenesis and in signalling triggered by flagellin. Planta 234:459-476

Fan SC, Lin CS, Hsu PK, Lin SH, Tsay YF (2009) The Arabidopsis nitrate transporter NRT1.7, expressed in phloem, is responsible for source-to-sink remobilization of nitrate. Plant Cell 21:2750-2761

Farooq M, Tagle AG, Santos RE, Ebron LA, Fujita D, Kobayashi N (2010) Quantitative trait loci mapping for leaf length and leaf width in rice cv. IR64 derived lines. J Integr Plant Biol 52:578-584

Fujii S, Hirai K, Saka H (1991) Growth-regulating action of brassinolide in rice plants. In: Cutler HG, Yokota T, Adam G (eds)

Brassinosteroids. Chemistry, Bioactivity, and Application. American Chemical Society, Washington DC, ACS Symposium Series, Vol. 74, pp 306–311

Hayama R, Coupland G (2004) The molecular basis of diversity in the photoperiodic flowering responses of Arabidopsis and rice. Plant Physiol 135:677-684

Hayama R, Yokoi S, Tamaki S, Yano M, Shimamoto K (2003) Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature 422:719-722

Hedden P (2003) The genes of the Green Revolution. Trends Genet 19:5-9

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068-1076

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793-800

Ishimaru K (2003) Identification of a locus increasing rice yield and physiological analysis of its function. Plant Physiol 133:1083-1090

Ishimaru K, Yano M, Aoki N, Ono K, Hirose T, Lin S, Y., Monna L, Sasaki T, Oshugi R (2001) Toward the mapping of physiological and agronomic characters on a rice function map : QTL analysis and comparison between QTLs and expressed sequence tags. Theor Appl Genet 102:793-800

Jian H, Jiang., Qiang M, Liu., Hong Z, Jian L, Chao L, Wan L, Ni., De L, Hong. (2012) Tiller morphogenesis and mortality were controlled by different loci in rice (Oryza sativa L.). African J Agri Res 7:3964-3977

Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, Qian Q, Li J (2010) Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. Nature Genet 42:541-544

Joehanes R, Nelson JC (2008) QGene 4.0, an extensible Java QTL-analysis platform. Bioinformatics 24:2788-2789

Khush G (1995) Breaking the yield frontier of rice. Geojournal 35:329-332

Khush GS (2001) Green revolution: the way forward. Nat Rev Genet 2:815-822

Khush GS (2005) What it will take to feed 5.0 billion rice consumers in 2030. Plant Mol Biol 59:1-6

Kosambi D (1994) The estimation of map distance from recombination values. Annal Eugenics 12:172-175

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

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181

Li S, Qian Q, Fu Z, Zeng D, Meng X, Kyozuka J, Maekawa M, Zhu X,

Zhang J, Li J, Wang Y (2009) Short panicle1 encodes a putative PTR family transporter and determines rice panicle size. Plant J 58:592-605

Li S, Xie K, Li W, Zou T, Ren Y, Wang S, Deng Q, Zheng A, Zhu J, Liu H, Wang L, Ai P, Gao F, Huang B, Cao X, Li P (2012) Re-sequencing and genetic variation identification of a rice line with ideal plant architecture. Rice 5:18

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

Liu G, Zhu H, Zhang G, Li L, Ye G (2012) Dynamic analysis of QTLs on tiller number in rice (Oryza sativa L.) with single segment substitution lines. Theor Appl Genet 125:143-153

Liu GF, Li M, Wen J, Du Y, Zhang YM (2010) Functional mapping of quantitative trait loci associated with rice tillering. Mol Genet Genomics 284:263-271

Lincoln S, Daly M, Lander E (1993) Constructing genetic linkage maps with MAPMAKER/EXP version 3.0. A tutorial and reference manual. Whitehead institute for biomedical research 3rd (ed) Technical report

Marathi B, Guleria S, Mohapatra T, Parsad R, Mariappan N, Kurungara VK, Atwal SS, Prabhu KV, Singh NK, Singh AK (2012) QTL analysis of novel genomic regions associated with yield and yield related traits in new plant type based recombinant inbred lines of rice (Oryza sativa L.). BMC Plant Biol 12:137

Maris A, Suslov D, Fry SC, Verbelen JP, Vissenberg K (2009) Enzymic characterization of two recombinant xyloglucan endotransglucosylase/hydrolase (XTH) proteins of Arabidopsis and their effect on root growth and cell wall extension. J Exp Bot 60:3959-3972

Moncada P, Martinez C, Borrero J, Chatel M, Gauch J, Guimaraes E, Tohme J, McCouch S (2001) Quantitative trait loci for yield and yield components in an Oryza sativa x Oryza rufipogon BC2F2 population evaluated in an upland environment. Theor Appl Genet 102:41-52

Monna L, Kitazawa N, Yoshino R, Suzuki J, Masuda H, Maehara Y, Tanji M, Sato M, Nasu S, Minobe Y (2002) Positional cloning of rice semidwarfing gene, sd-1: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. DNA Res 9:11-17

Morinaka Y, Sakamoto T, Inukai Y, Agetsuma M, Kitano H, Ashikari M, Matsuoka M (2006) Morphological alteration caused by brassinosteroid insensitivity increases the biomass and grain production of rice. Plant Physiol 141:924-931

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic acids Res 8:4321-4325

Neto EC, Keller MP, Broman AF, Attie AD, Jansen RC, Broman KW, Yandell BS (2012) Quantile-based permutation thresholds for quantitative trait loci hotspots. Genetics 191:1355-1365

Paul C, Janet B (1999) In vitro activities of four xyloglucan endotransglycosylases from Arabidopsis. Plant J 18(4):371-382

Qu Z, Li L, Luo J, Wang P, Yu S, Mou T, Zheng X, Hu Z (2012) QTL mapping of combining ability and heterosis of agronomic traits in rice backcross recombinant inbred lines and hybrid crosses. PloS one 7:e28463

Rahman L, Most SK, Hee J, Kho. (2008) QTL Analysis for Yield Related Traits Using Populations Derived from an indica-japonica Hybrid in Rice (Oryza sativa L.). Czech J Genet Plant Breed 3:93-104

Routaboul JM, Dubos C, Beck G, Marquis C, Bidzinski P, Loudet O, Lepiniec L (2012) Metabolite profiling and quantitative genetics of natural variation for flavonoids in Arabidopsis. J Exp Bot 63:3749-3764

Xing Y, Zhang Q (2010) Genetic and molecular bases of rice yield. Annual review of Plant Biol 61:421-442

Sano, Y (1992) Genetic comparisons of chromosome 6 between wild and cultivated rice. Japanese J Breed 42:561–572

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

Spielmeyer W, Ellis MH, Chandler PM (2002) Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proc Natl Acad Sci USA 99:9043-9048

Suh H, Hue M (1978) The segregation mode of plant height in the

cross of rice varieties. XI. Linkage analysis of the semi-dwarfness of the rice variety 'Tongil'. Korean J Breed 10:1-6

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S, Innan H, Cano L, Kamoun S, Terauchi R (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174-183

Tamura K, Nomura K, Oshima I, Namai H, Yano M, Sasaki T, Kikuchi F (1998) Identification of restriction fragment length polymorphism markers tightly linked to a major photoperiod sensitivity gene, Se-1, and to a blast resistance gene, Pi-zt, in rice. SABRAO J Breed Genet 30:61-67

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441-1452

Tisne S, Reymond M, Vile D, Fabre J, Dauzat M, Koornneef M, Granier C (2008) Combined genetic and modeling approaches reveal that epidermal cell area and number in leaves are controlled by leaf and plant developmental processes in Arabidopsis. Plant Physiol 148:1117-1127

Trewavas AJ (2001) The population/biodiversity paradox. Agricultural efficiency to save wilderness. Plant Physiol 125:174-179

Tuinstra M, Ejeta G, Goldsbrough P (1997) Heterogeneous inbred

family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet 95:1005-1011

Vergara B, Tanaka A, Lilis R, Puranabhavung S (1996) Relationship between grwowth duration and grain yield of rice plants. Soil Sci Plant Nutr 12:31-39

Wang L, Wang A, Huang X, Zhao Q, Dong G, Qian Q, Sang T, Han B (2010) Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. Theor Appl Genet 122:327-340

Wang P, Zhou G, Yu H, Yu S (2011) Fine mapping a major QTL for flag leaf size and yield-related traits in rice. Theor Appl Genet 123:1319-1330

Wu CY, Trieu A, Radhakrishnan P, Kwok SF, Harris S, Zhang K, Wang J, Wan J, Zhai H, Takatsuto S, Matsumoto S, Fujioka S, Feldmann KA, Pennell RI (2008) Brassinosteroids regulate grain filling in rice. Plant Cell 20:2130-2145

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30:105-111

Yan J, Q., J Z, C X, He., M B, P W (1998) Quantitative trait loci analysis for the developmental behavior of tiller number in rice (Oryza sativa L .). Theor Appl Genet 97:267-274

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

Yoo SC, Cho SH, Zhang H, Paik HC, Lee CH, Li J, Yoo JH, Lee BW, Koh HJ, Seo HS, Paek NC (2007) Quantitative trait loci associated with functional stay-green SNU-SG1 in rice. Mol Cells 24:83-94

초록

이 연구는 벼의 초형을 결정하고 수량을 증가시키는 형질관련 유전자를 동정하기 위하여 수행되었다. 초형과 수량을 결정하는 양적형질 유전자들을 분석하기 위하여 자포니카 품종의 SNU-SG1과 고수확량의 통일벼 품종 Milyang23을 교배하여 QTL 분석을 진행하였다. 세대진전을 통해 얻은 178개의 F_7 recombinanat inbred lines (RILs)을 얻고 여기서 벼의 농업형질 중 키, 출수일, 분얼수, panicle 길이, panicle 두께, 지엽 길이, 지엽 두께, 수확량의 8개 농업형질에 대하여 QTL(Quantitative trait loci) 분석을 진행하였다. 유전자 동정 과정 중 QTL mapping 과정 및 유전자 분석 과정의 효율성을 높이고자 각 부모형에 대하여 whole genome sequencing (Illumina사의 Genome Analyser IIx를 이용한 Next Generation Sequencing)을 진행하였다. 벼의 12개 유전자를 QTL mapping하는데 분자 마커 125개가 이용되었고, 이 중 RM marker 64개, STS marker 33개 그리고 차세대 시퀀싱을 이용하여 제작한 28개의 NID(Next generation sequencing InDel) marker가 사용되었다. 그 결과 LOD 2.8 이상인 QTL 46개를 얻었다. 46개의 QTL 중 표현형기여도 기준으로 20개의 main effect QTL을 선정하였으며, 여기서 17개의 유력한 후보유전자를 찾았다. 이 17개 후보 유전자는 SNP에 의해서 단백질 합성이 중단되거나 틀 이동이 일어나 비정상적으로 단백질이 합성되는 경우에 해당된다. QTL 분석에 whole genome sequencing 방법을 도입 함으로써 좀 더 효율적인 분자마커 디자인이 가능 했으며, 또한 벼의 초형 및 수량 형질에 대한 후보유전자 동정이 가능하였다. 이번 연구를 통하여 유용한 형질에 대한 QTL분석 자료를 얻었고 SNP 분석법에 있어 시퀀싱 기술을 이용한 진일보

한 시도였다고 사료된다.