Studies on the identification of QTL/gene for agronomically important traits in rice (*Oryza sativa* L.)

2012. 3

岩手大学大学院
連合農学研究科
寒冷圏生命システム学専攻
(岩手大学)

阿部　陽

Table of contents

Chapter 1: General introduction..... 1

Chapter 2: *OsGA20ox1*, a major QTL for plant height of seedlings in rice.... 6

2-1: Introduction..... 7

2-2: Materials and Methods..... 8

2-3: Results.....13

2-4: Discussion.....30

Chapter 3: Genome sequencing reveals agronomically-important loci

in rice using a novel method MutMap.....34

3-1: Introduction.....35

3-2: Materials and Methods.....36

3-3: Results.....40

3-4: Discussion.....65

3-5: Supplementary data.....72

Chapter 4: General discussions.....77

References.....87

List of publications.....98

Acknowledgments.....99

Chapter 1

General introduction

Rice (*Oryza sativa* L.) is one of the most important food crops in the world, providing over 21% of the calorific needs of the population of the world and up to 76% of the calorific intake of the population of Southeast Asia (Fitzgerald *et al.* 2009). Food shortage is becoming a serious global problem because of the ever-increasing world population and the effects of global climate change on food production. In addition, the demand for bio-fuels as alternative energy sources is expected to rise, and this will limit the availability of land for cultivating food crops (Brown and Funk 2008). To meet the expanding food demands, 50% increase in crop production will be needed by 2025 (Khush 2001). To this end, breeding of the major crops such as rice is vital.

Plant breeding has been carried out mainly by the selection of naturally occurring variants. Since the beginning of rice cultivation some 10000 years ago, various traits of the crop have been continuously improved through the introduction of naturally occurring useful alleles by spontaneous and/or artificial crossings (Yamamoto *et al.* 2009). Many important agronomic traits, such as heading date, plant height, grain yield and biotic/abiotic stress resistance/tolerance, are not discrete but quantitative characters in nature. These complex traits are controlled by multiple genes known as quantitative trait loci (QTLs) originally derived from natural variations (Yano 2001). QTL analysis aims to identify positions of QTLs, and has been employed as a powerful approach to discover agronomically useful genes (Yano *et al.* 2000; Takahashi *et al.* 2001; Kojima *et al.* 2002; Ashikari *et al.* 2005; Konishi *et al.* 2006; Fujino *et al.* 2008; Fukuoka *et al.* 2009; Miura *et al.* 2010; Sugimoto *et al.* 2010). Identifying existing genetic variations by QTL analysis is important as it provides the basis for rice

improvement through conventional breeding approaches and makes rice breeding by marker-assisted selection (MAS) possible. To date, however, useful QTL information has not been fully exploited in rice breeding programs. Therefore, more effort is needed in QTL isolation to understand and utilize complex traits relevant to rice breeding.

Recent progresses in rice genomics should accelerate its breeding to develop improved varieties in a shorter time period. In 2002, the draft genome sequences of two subspecies, *O. sativa* ssp. *japonica* (cv. Nipponbare) and *O. sativa* ssp. *indica* 93-11, were reported (Goff *et al.* 2002; Yu *et al.* 2002). Subsequently the International Rice Genome Project finalized and published the complete genome sequence of Nipponbare cultivar (IRGSP 2005). This achievement has provided us with a vast amount of information on the rice genome and paved the way for detailed genetic analyses of the crop. By using this genomics information, researchers have now succeeded in isolating some important QTL genes, which have the potential to greatly improve rice production (Ashikari *et al.* 2005; Konishi *et al.* 2006; Fujino *et al.* 2008; Fukuoka *et al.* 2009; Miura *et al.* 2010; Sugimoto *et al.* 2010). Several of the isolated QTLs with major effects have already been introgressed into elite rice cultivars by MAS. However, QTLs with minor effects remain poorly investigated (Yamamoto *et al.* 2009). R. A. Fisher (1918) and many other researchers have demonstrated that the correlation of quantitative traits between relatives is explained by the combined action of alleles at many loci, each having a relatively small effect. To enable plant breeding by MAS, it is important to identify the locus or chromosome region harboring each gene contributing to an improved trait. However, compared to genes with major effects that determine discrete characteristics,

allelic substitutions at genes causing minor effects only result in subtle changes in the phenotype. Consequently, cloning of genes controlling agronomic traits is not straightforward.

To understand the genetic architecture of agronomic traits, the separation of QTLs affecting complex traits into individual loci is important (Yamamoto *et al.* 2009). In crops, the isolation of minor effect genes and QTLs conventionally involves crossing two inbred lines showing contrasting phenotypes to generate Recombinant Inbred Lines (RILs). Each RIL is a proxy-clone with a high degree of homozygosity that allows measurements of a large number of individuals per line with reduced measurement errors, something that cannot be readily achieved with an F_2 segregating population. Once the RILs are scored, the association between phenotypic measurements and DNA markers of known chromosome locations is then used to infer the approximate positions of the genes. Identification of multiple loci from the RILs is then followed by a series of backcrosses to generate near-isogenic lines (NILs) carrying a single QTL. NILs are subsequently used to isolate individual genes by conventional map-based cloning. The generation of RILs and NILs is time-consuming; hence large-scale isolation of genes contributing minor phenotypic effects remains difficult.

On the other hand, artificially generated mutants have been used for gene isolation thorough map-based cloning. This method involves generating an F_1 hybrid via a cross between a mutant and a distantly related accession. The use of distant cross is important to secure the availability of DNA markers for linkage analysis. For recessive mutants, the next step is the examination of each individual exhibiting the

phenotype of interest among the F_2 population, followed by assessment of the linkage between the phenotype and DNA markers. Once the linkage analysis narrows down the chromosomal region containing the mutation to a relatively short interval, candidate genes within the interval are sequenced one by one using conventional Sanger sequencing until the mutation is finally identified. Overall, these processes are time-consuming and labor intensive. Furthermore, since QTLs existing between mutant and another parental line confounds the scoring of the mutant phenotype in an F_2 progeny, identification of causal mutation has not been a straightforward procedure.

Recent innovations in DNA sequencing technology have made whole genome sequencing (WGS) possible, and WGS has been applied to the identification of causal genes for mutant phenotypes in several model organisms (Schneeberger *et al.* 2009; Austin *et al.* 2011; Uchida *et al.* 2011; Zuryn *et al.* 2010; Ashelford *et al.* 2011). This so-called next-generation sequencing (NGS) technology might avail new opportunities for QTL isolation and crop breeding using artificially generated mutants that show subtle changes in a quantitative traits of agronomical importance.

The objectives of this study were twofold, and both are related to the improvement rice by breeding. Chapter 2 deals with the identification of QTLs for seedling vigor, which is one of the determinants of seedling establishment in rice direct seeding method in Northern Japan. In Chapter 3, the development of a novel method of rapid gene isolation using NGS technology is described.

Chapter 2

OsGA20ox1, a major QTL for plant height of seedlings in rice

2-1 Introduction

Direct seeding of rice (*Oryza sativa* L.) is becoming popular in Japan as it involves lower labor costs compared to the conventional transplanting of seedlings from nursery. However, yield loss mainly due to slow seedling growth is a common problem associated with direct seeding. Delayed emergence of rice seedlings from paddy water surface has been shown to greatly increase seedling mortality (Peterson *et al.* 1978), thereby causing significant reductions in yield. In addition, there is a higher risk of yield loss due to competition with weeds associated with direct seeding than transplanting (Rao *et al.* 2007). Seedling vigor is mainly determined by the rate of germination and rapid early growth of both the shoot and roots (Williams and Peterson 1973; Sasahara *et al.* 1986). However, this poses a major problem in temperate regions where low temperature limits early seedling growth (Redona and Mackill 1996a). As a result, success in direct sowing of germinated seeds in puddled soils, which is becoming a common practice in countries like Japan, depends on having cultivars with good seedling vigor for stable stand establishment.

Many important agronomic traits, including seedling vigor, are generally controlled by complex interactions of multiple genes known as quantitative trait loci (QTLs; Yano 2001; Yano and Sasaki 1997; Ashikari and Matsuoka 2006). Natural variation in QTLs accounts for a wide range of phenotypic differences (Yano and Sasaki 1997). To understand the genetic architecture of agronomic traits, separation of QTLs affecting complex traits into individual loci is important. QTL analysis aims to identify

approximate positions of QTLs, and has been successfully used to identify loci or genes for complex traits (Yamamoto *et al.* 2009; Miura *et al.* 2011). Some QTLs for seedling vigor in rice have been identified by QTL analysis (Onishi *et al.* 2007; Redona and Mackill 1996b; Cui *et al.* 2002; Zhang *et al.* 2005a,b; Zhou *et al.* 2007). However these results have not been always consistent in terms of the exact locations of the QTLs, although the QTLs have been mapped to specific regions in the genome.

The current study was initiated with the main aim of identifying QTLs for seedling vigor in the cultivar Dunghan Shali (*O. sativa* L. subsp. *japonica*). Dunghan Shali was identified as one of the few cultivars showing higher seedling vigor among our rice cultivars and breeding lines, and was then used as a potential donor parent to incorporate this important trait into elite cultivars. A set of recombinant inbred lines (RILs) derived from a cross between Kakehashi (*Oryza sativa* L. subsp. *japonica*) and Dunghan Shali was used to map QTLs affecting seedling vigor. We also produced NILs carrying a major QTL in an elite cultivar background and evaluated its contribution to the enhanced seedling vigor. Here we report the chromosomal positions of the QTLs and the major QTL *qPHS3-2,* as well as identification of the main gene most likely responsible for the difference in seedling height between Dunghan Shali and Kakehashi.

2-2 Materials and Methods

Plant materials

To produce RILs, we selected Kakehashi and Dunghan Shali as parental rice cultivars. Kakehashi, a temperate *japonica* variety, is adapted to the relatively cooler northern Japan climate. Dunghan Shali is a temperate *japonica* variety grown in Hungary and shows superior seedling vigor. The F_1 plants generated by crossing Kakehashi with Dunghan Shali using the former as female parent were self-pollinated, and ~ 300 F_2 seeds were obtained. Each F_2 seed was maintained as RIL by single-seed descent method, resulting in a total of 250 RILs of F_7 generation.

Development of NILs

We used the rice cultivar Iwatekko (*japonica*) as the recurrent parent to generate NILs harboring the target QTL region, *qPHS3-2*, from Dunghan Shali. Iwatekko is a promoted cultivar in Iwate prefecture (Japan), and is preferred for its superior food quality as well as for its tolerance to cool temperature at the booting stage. Iwatekko was first crossed to Dunghan Shali, and the resulting F_1 plants were backcrossed four times to Iwatekko using marker-assisted selection (Yano and Sasaki 1997; Yano 2001).

Evaluation of seedling height

The seeds of all lines used in this study were imbibed in water for 5 days at 12°C and then forced into sprouting for 2 days at 25°C. Fifteen seeds per line were sown on a cell plug tray (cell count: 8*16; tray size: 600*300mm; depth of cell: 45mm) , and the plants were

germinated and grown in a greenhouse where the temperature was controlled at 25°C but under natural light conditions. Seedling height (length between the coleoptile node to the tip of longest leaf blade) was measured at 14 days after sowing.

Genotyping of RILs and NILs with DNA markers

For genotyping the RILs, we used a total of 88 Simple Sequence Repeat (SSR) markers that covered the 12 rice chromosomes from the lists provided by McCouch *et al.* (2002) and IRGSP (2005). Procedures for DNA extraction and SSR marker polymorphism detection were described previously (Matsubara *et al.* 2008; Hori *et al.* 2010). An additional set of 11 SSR (McCouch *et al.* 2002; IRGSP 2005), two insertions and deletions (InDels) and two cleavage amplified polymorphic sequence (CAPS) markers (Table 2-1) located on the long arm of chromosome 3 and showing polymorphisms between Dunghan Shali and Iwatekko were used to develop the NILs.

Map construction and QTL mapping

The linkage map was constructed using Mapmaker/EXP 3.0 (Lander *et al.* 1987), and the Kosambi map function was used to calculate genetic distances (Kosambi 1994). QTL analysis was performed using composite interval mapping as implemented by the Zmapqtl program (model 6) provided by the software package QTL Cartographer version 2.5 (Basten *et al.* 2005; http://statgen.ncsu.edu/qtlcart/WQTLCart.htm). A threshold

Table 2-1 Primers used in this study.

Marker name	Upper primer sequence	Lower primer sequence	Purpose
ID-INDEL-01	GATAATGAACCTGGA ACTGC	TTTCAGATGAGCAAA GGATT	mapping
ID_INDEL_02	TGATGAGAGTGAACC AGCGTA	ACCGGGTGTAGTCAA GTTGC	mapping
ID_CAPS_01	TCTGCACGTTTTTGT TGTCC	TTGTGCATGCAACAG CAGTA	mapping CAPS (*Hpy188*I)
ID_CAPS_02	CACAATGAAGAGGA GCAGCA	ACGATGACATCCCAC ATCCT	mapping CAPS (*Sty*I)
OsGA20oxReal	TGTCGCTGGAGCTGA TGGA	TAGTTGAGGCGCATG ATGGA	real-time RT-PCR
OsActin-1Real	CCCCCATGCTATCCTT CGT	GGCCGTTGTGGTGAA TGAGT	real-time RT-PCR

SSR markers designated by RM numbers are not listed here. For details on these markers readers are referred to IRGSP 2005.

value (r = 0.05) was determined by 1000 permutation test.

Gene expression analysis

Total RNA was isolated from embryo or shoots sampled from plants grown under continuous light at 7, 12, and 17 days after start of imbibition (DAI) of the seeds using an RNA isolation kit (RNeasy plant mini, QIAGEN). This was followed by DNase treatment to remove contaminating DNA from the extracts. Two micrograms of total RNA were converted to cDNA with RTase (TOYOBO).

To quantify the expression levels of transcripts, real-time RT PCR analysis was performed using SYBR green (QIAGEN) on a Step-One Plus Real-Time PCR system (Applied Biosystems). Expression levels were normalized with the *OsActin1* gene. The primer sequences used in this study are provided in Table 2-1.

Quantification of the endogenous gibberellins

About 80 mg fresh weight sample, which was composed of leaf sheath and the stem but without the leaf blades and the roots, was subjected to the analysis. The concentrations of gibberellins were measured with MS-probe modification method as previously described (Kojima *et al.* 2009). In brief, the extract was passed through an Oasis HLB column (Waters, Milford, MA, USA), and gibberellin-containing fraction was obtained by solid-phase extraction with an Oasis MCX column (Waters). After modification with

MS-probe, the derivatized gibberellins were measured with a liquid chromatography-tandem mass chromatography system (AQUITY UPLC™ System/XEVO-TQS; Waters) with an ODS column (AQUITY UPLC BEH C_{18}, 1.7 μm, 2.1 X 100 mm, Waters). Data were processed by MassLynx™ software with QuanLynx™ (version 4.0, Waters).

2-3 Results

Phenotypic variation

A clear difference in seedling height was observed between the two parental lines grown at 25°C for 14 days (Fig.2-1a). On average, seedlings of Kakehashi attained a height of 20.3cm (±1.3), whereas those of Dunghan Shali reached about 29.4cm (±2.9). The RILs showed a continuous distribution with respect to seedling height from 14.1 to 31.7cm, suggesting that this trait is controlled by multiple QTLs (Fig.2-1b).

QTL analysis for plant height of seedling

We constructed a genetic linkage map, based on RILs of the F_6 generation, using a total of 88 SSR markers. The segregation ratios of the two genotype classes fit the expected Mendelian ratio of 1:1 in most loci. However, segregation distortion was observed for two loci located each on chromosome 3 (RM232-RM7365) and 11 (RM7221-RM5926),

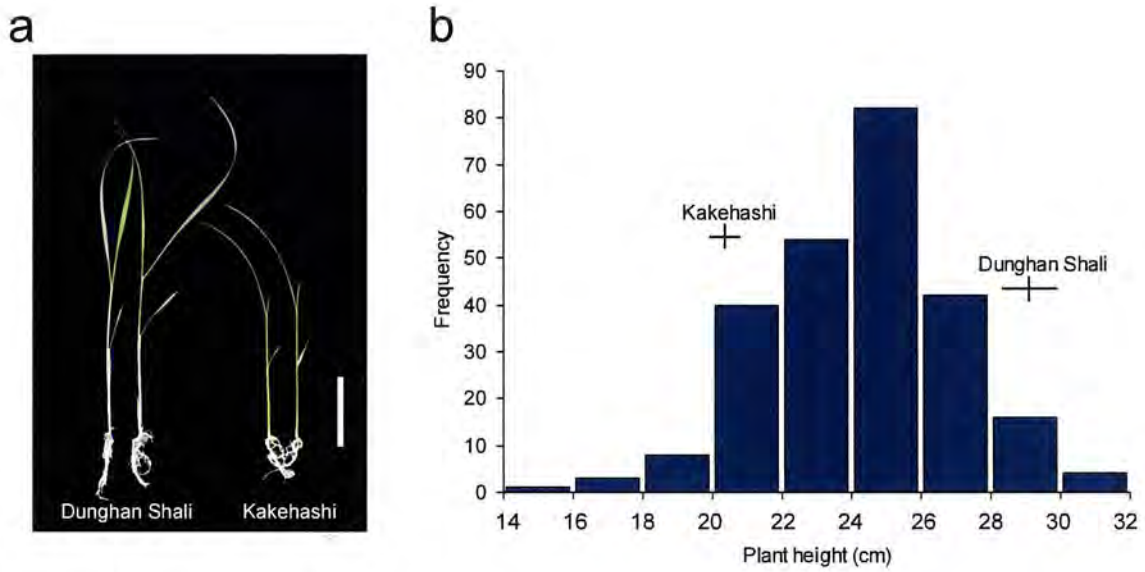

Figure 2-1 Seedling vigor of Dunghan Shali, Kakehashi and RILs.

(a) Gross morphology at 14 days after sowing. Scale bar, 5 cm. **(b)** Frequency distribution of seedling height of 250 RILs germinated and grown at 25°C for 14 days. Vertical and horizontal bars indicate the mean value and standard deviation, respectively.

for which all RILs had the Kakehashi type alleles. Segregation distortion has been frequently observed in populations derived from intraspecific crosses of rice, and certain genomic regions related to segregation distortion have been uniquely reported to particular crosses (Matsubara *et al.* 2011). The region showing segregation distortion on chromosome 3 (RM232-RM7365) is the same location as the one identified in several previous studies (Matsubara *et al.* 2011). Nevertheless, the region on chromosome 11 has not been reported before.

Composite interval mapping was performed to identify the QTLs associated with seedling height, based on 250 RILs of the F_7 generation. A major QTL (LOD=17.5) controlling plant height was identified on chromosome 3 with an average contribution of 26.2% to the phenotypic variation (Fig. 2-2 and Table 2-2). This QTL, designated as *qPHS3-2* (QTL for Plant Height of Seedling 3-2), was located within a 12.2 cM interval flanked by two SSR markers, RM1373 and RM7389, on the distal end of the long arm of chromosome 3. Three additional QTLs, *qPHS1* on chromosome 1, *qPHS3-1* on chromosome 3 and *qPHS4* on chromosome 4, were detected each accounting for 7.2, 6.6 and 5.2% of the total phenotypic variation, respectively (Fig. 2-2 and Table 2-2).

Production of NILs and fine-mapping of *qPHS3-2*

To evaluate the effects of individual QTLs separately, it is necessary to produce NILs for each QTL in a common genetic background (Paterson *et al.* 1988). We produced NILs carrying the major QTL from Dunghan Shali in an Iwatekko genetic background by a

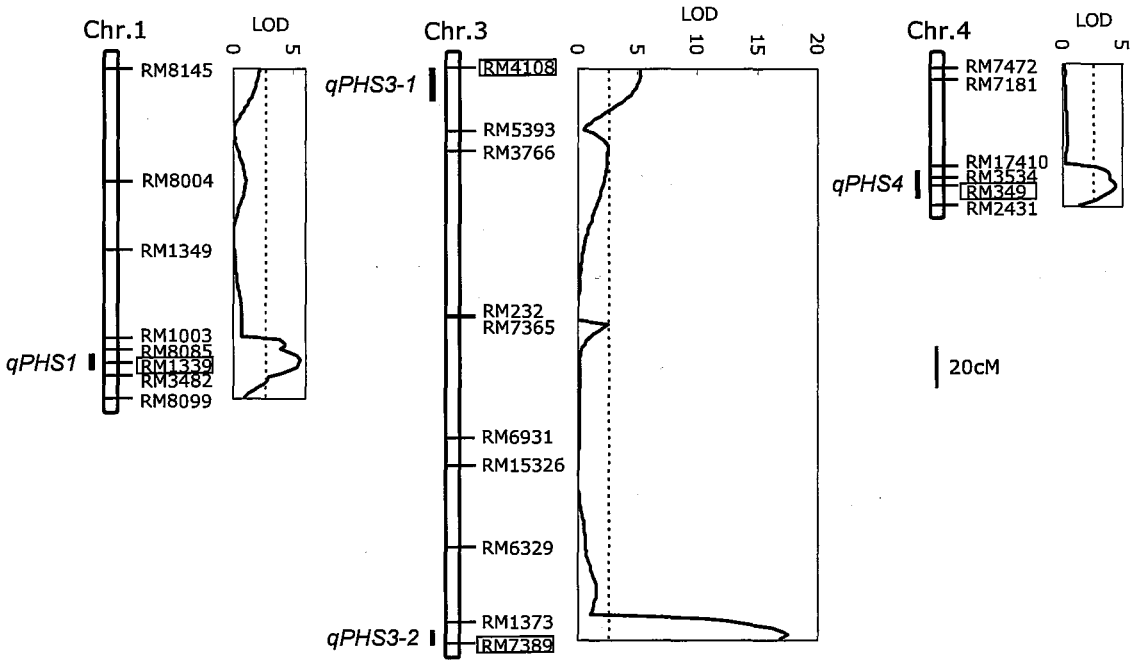

Figure 2-2 Linkage map of the RILs derived from a cross between Kakehashi and Dunghan Shali showing location of the putative QTLs with LOD score.

The markers in boxes represent those nearest to their respective QTLs.

Table 2-2 QTLs for height of seedlings detected in the RILs derived from a cross between Kakehashi and Dunghan Shali.

QTL	Chr.	Marker interval	Nearest marker	LOD	PVE	AE (cm)
qPHS1	1	RM8085-RM1339	RM1339	5.5	7.2	0.8
qPHS3-1	3	RM4108-RM5393	RM4108	5.3	6.6	0.7
qPHS3-2	3	RM1373-RM7389	RM7389	17.5	26.2	1.5
qPHS4	4	RM3534-RM349	RM349	4.4	5.2	0.6

All genetic parameters were calculated by QTL Cartographer Ver.2.5. The LOD threshold for detection of QTL in the composite interval mapping procedure was 2.6.

LOD Log-likelihood value; *PVE* percentage of total phenotypic variance explained by the QTL; *AE* additive effect of Dunghan Shali allele for plant height of seedling.

repeated backcrossing to Iwatekko. By means of marker-assisted selection using a total of 88 SSR markers, we developed nine BC_4F_3 NILs harboring different portion of Dunghan Shali segments in the long arm of chromosome 3 (Fig. 2-3). The rest of the genome was Iwatekko type for all the nine NILs. Using these NILs, we fine-mapped the major QTL, *qPHS3-2*. The eight NILs (17, 19, 28, 92, 86, 32, 33 and 114) harboring Dunghan Shali segments at the distal region of the long arm of chromosome 3 in an Iwatekko background had significantly larger plant height than Iwatekko (Fig. 2-3). Among the eight NILs, NIL-114 has the shortest DNA segment (797-kbp) derived from Dunghan Shali at the marker interval RM2187-RM16245. On the other hand, the seedling height of NIL-11, which contained only a short Dunghan-Shali segment at its proximal-end and the largest segment of Iwatekko within the candidate region for *qPHS3-2*, was not significantly different from that of Iwatekko seedlings. These results reveal that the *qPHS3-2* is localized within a 797-kbp region in the distil region of the long arm of chromosome 3 (Fig. 2-3).

To narrow down the *qPHS3-2* region, we increased the number of NILs that harbor recombinations between RM16200 and RM16245. Two BC_4F_4 lines, NIL-115-18 and NIL-116-66, had different segments of Dunghan Shali chromosome within the interval RM2187-RM16245 (Fig. 2-4a). Plant heights of the progeny derived from selfing of these two lines (BC_4F_5) did not show significant difference from that of Iwatekko (Fig. 2-4a). Besides, we obtained an additional NIL, NIL-116-43-497, which had Dunghan Shali genome segment in the marker interval between RM16217 and RM7389 in a heterozygous state (Fig. 2-4b). The BC_4F_6 obtained by selfing of this NIL

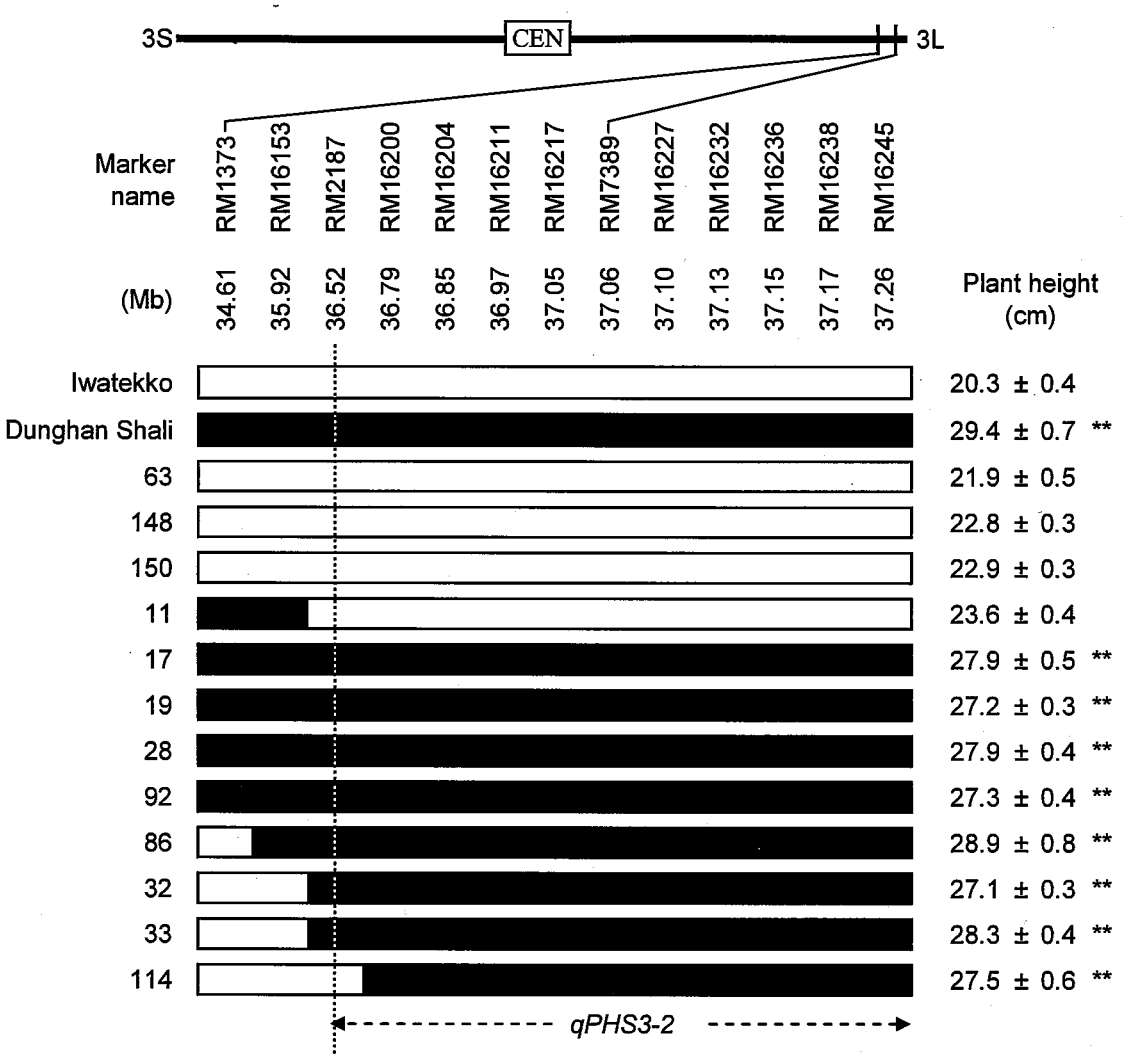

Figure 2-3 Graphical genotypes of nearly isogenic lines (NILs) of BC_4F_3 generation and location of *qPHS3-2*.

White and black boxes indicate homozygous regions for the Iwatekko and Dunghan Shali, respectively. Mean shoot length and standard deviation values were included. ** indicate significant difference (P <0.01) between Iwatekko and Dunghan Shali or NILs by Dunnett's multiple comparison test.

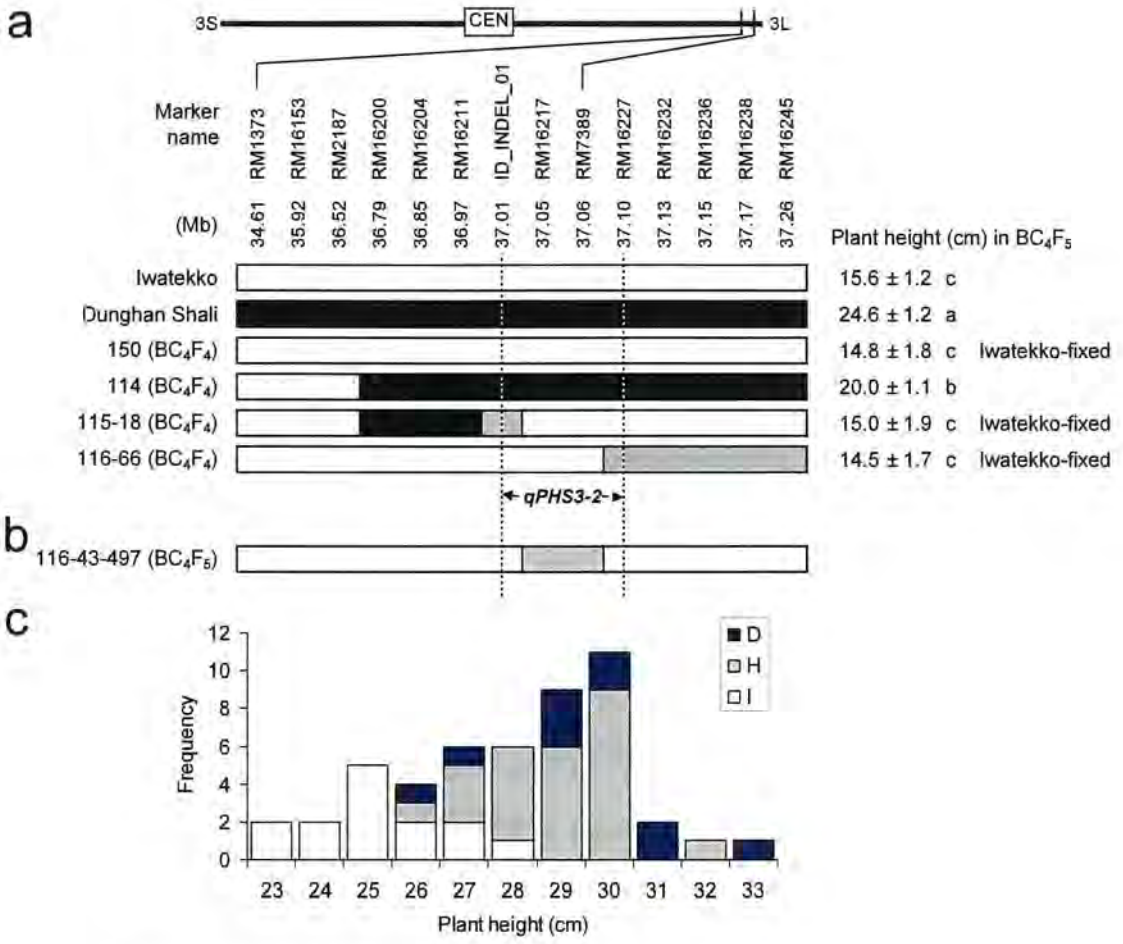

Figure 2-4 Fine-mapping of *qPHS3-2*.

(a) Graphical genotype of Iwatekko, Dunghan Shali and NILs of BC_4F_4 generation. Positions of the Markers in rice genome sequence (Build 5.0; IRGSP 2005) are given. White and black boxes are homozygous for the Iwatekko- and Dunghan Shali-type chromosome segments, respectively. Gray box indicates heterozygous region. Plant height (mean ± SD) was measured in BC_4F_5 progeny that were obtained by self-pollination of each NILs. Letters indicate significant difference among the six lines (P <0.01, Tukey's multiple comparison tests). **(b)** Graphical genotype of NIL-116-43-497 in BC_4F_5. White and gray boxes indicate homozygous and heterozygous for the Iwatekko-type segment, respectively. **(c)** Frequency distribution of seedling height in the BC_4F_6 progeny that was obtained by self-pollination of NIL-116-43-497. D, H and I indicate Dunghan Shali-type homozygous, heterozygous, and Iwatekko-type homozygous, respectively, for RM16217 and RM7389 markers.

exhibited segregation for plant height among the seedlings (Fig. 2-4c). Progeny with Dunghan Shali genome segment in this region in homozygous or heterozygous states showed larger seedling heights than those with Iwatekko-type segment in homozygous state. This result suggested that Dunghan Shali-type allele of the QTL might be dominant over the Iwatekko allele (Fig. 2-4c).

Furthermore, NIL-497-8 with Dunghan Shali genome segment in the marker interval between ID_INDEL_02 and ID_CAPS_02 was selected from NIL-116-43-497 at BC_4F_6 generation. NIL-497-8 harboring only 81 kb segment of Dunghan Shali genome in the Iwatekko background showed a significantly larger seedling height than Iwatekko (Fig. 2-5a). This result strongly suggests that *qPHS3-2* resides in the interval between ID_CAPS_01 and RM16227 with a distance of 81 kb (Fig. 2-5).

Expression of *OsGA20ox1* in seedlings

The 81 kb interval between markers ID_CAPS_01 and RM16227 is predicted to contain 10 genes including *OsGA20ox1*, a gene that encodes gibberellins 20-oxidase-1 (Fig. 2-5b). The gibberellins (GAs) are plant hormones that play important roles in many aspects of plant growth and development, such as seed germination, stem elongation, and flower development (Yamaguchi and Kamiya, 2000, Olszewski *et al.*, 2002). Loss-of-function mutations of the *OsGA20ox2/sd1* gene that is located on chromosome 1 cause semi-dwarfism, resulting in lodging resistance and increased grain yields, and thus are responsible for rice 'Green Revolution' (Sasaki *et al.*, 2002). It is interesting that

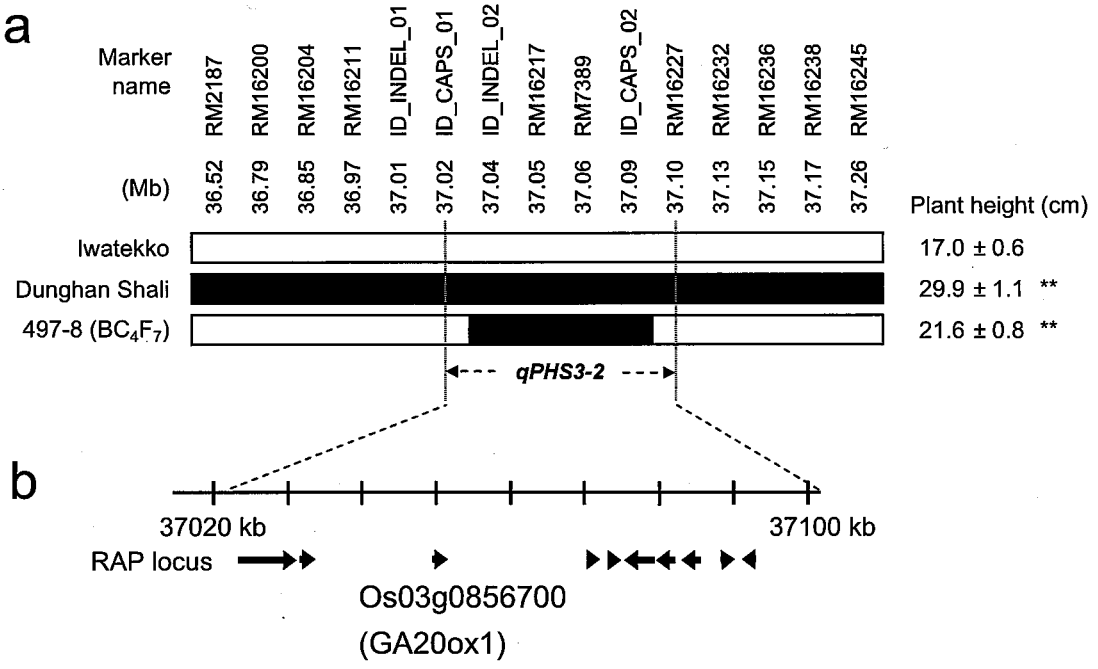

Figure 2-5 High resolution-mapping of *qPHS3-2*.

(a) Graphical genotype of Iwatekko, Dunghan Shali and NILs of BC_4F_7 generation with plant height (mean±SD) at seedling stage. White and black boxes are homozygous for the Iwatekko- and Dunghan Shali-type chromosome segments, respectively. ** indicate significant difference (P <0.01) between Iwatekko and Dunghan Shali or NILs by Dunnett's multiple comparison test. **(b)** Predicted genes, including *GA20ox1*, in the *qPHS3-2* candidate region according to the Rice Annotation Project Database (RAP-DB: Ohyanagi *et al.* 2005).

OsGA20ox1 is located within the 81 kb candidate genomic region of the major QTL for plant height, *qPHS3-2*, detected in the present study (Fig. 2-5b). To evaluate the relationship between seedling height and *OsGA20ox1*, real-time RT-PCR analysis of *OsGA20ox1* gene expression was performed using Iwatekko, Dunghan Shali, NIL-114 and NIL-497-8 (see Fig. 2-4a and Fig. 2-5a) during the initial growth stage. Relative expression levels of *OsGA20ox1* in embryo of seeds maintained at 25°C for 7 DAI (days after start of imbibition) were clearly different between Iwatekko and the other three lines (Fig. 2-6a). All the four lines began to germinate at this stage, but the initiation of germination in Dunghan Shali was clearly earlier as compared to Iwatekko and NIL-114 (Fig. 2-7), as well as compared to NIL-497-8 (data not shown). In addition, the relative expression levels of *OsGA20ox1* in seedlings of Dunghan Shali, NIL-114 and NIL-497-8 at 12 DAI and 17 DAI were significantly higher than that of Iwatekko (Fig. 2-6b). However, *OsGA20ox1* expression levels in seedlings (Fig. 2-6b) were lower than the levels detected in embryo at germination stage (Fig. 2-6a).

We also compared the genomic *OsGA20ox1* sequence between Iwatekko and Dunghan Shali alleles. There was no nucleotide change in the coding region, but there were four DNA changes within the 10 kb region upstream of the translational start site (Fig. 2-8).

Quantification of endogenous GAs in seedling

GA_1 is the primary active GA in the vegetative shoots of rice, and the

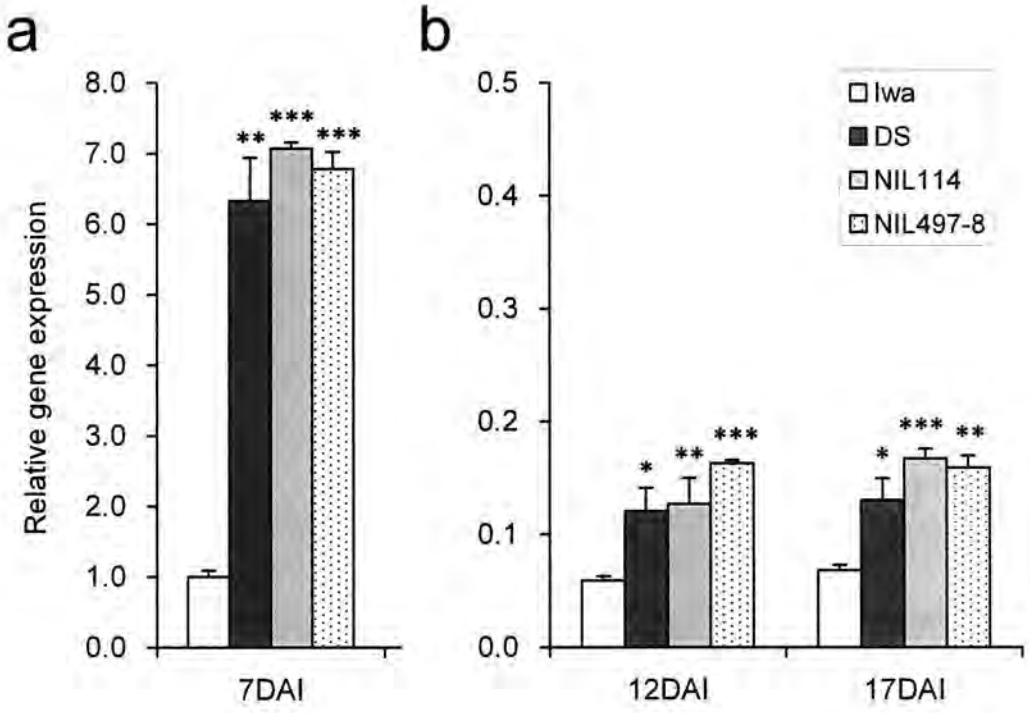

Figure 2-6 Relative expression levels of *OsGA20ox1* during shoot elongation at (a) 7 days, and (b) 12 and 17 days after start of inbibition (DAI).

Transcript abundance was determined by Real-time RT-PCR and normalized to the abundance in Iwatekko at 7 DAI (mean ± SD, n=3). OsActin1 was used as an internal control for normalization. Asterisks indicate significant differences from Iwatekko (Student's t-test, $^{**}P < 0.01$; $^{***}P < 0.001$).

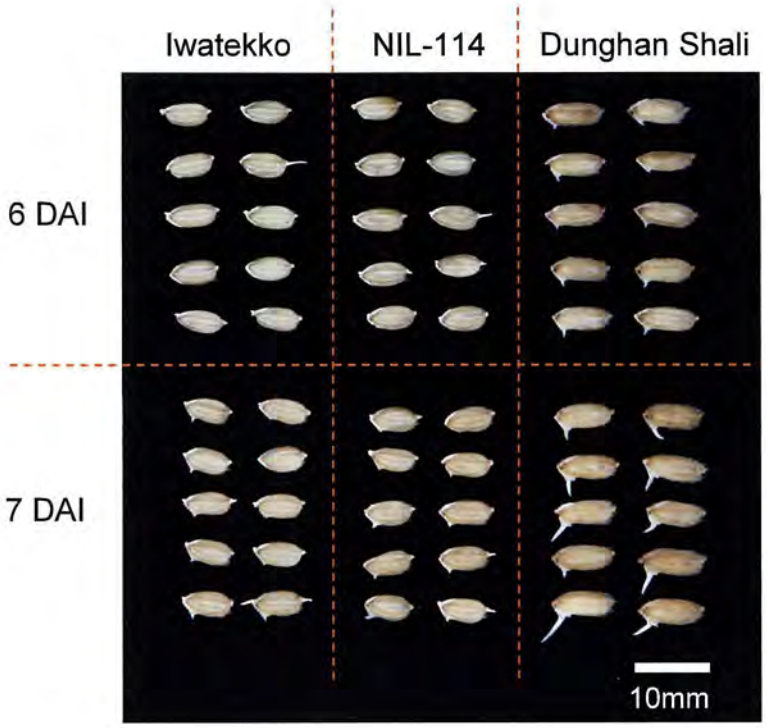

Figure 2-7 Variation in germination rate between Dunghan Shali, Iwatekko and NIL-114.

The germination in Dunghan Shali was visible from about 6 DAI. No difference in germination rate was observed between Iwatekko and NIL-114.

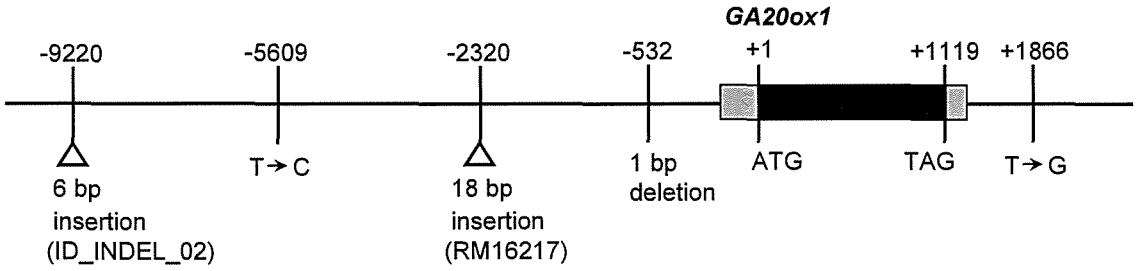

Figure 2-8 Comparison of the DNA sequences of *GA20ox1* between Dunghan Shali and Iwatekko.

Nucleotide changes are shown for Dunghan Shali using Iwatekko as the reference.

early-13-hydroxylation pathway is considered to be predominant (Kobayashi *et al.* 2000). In addition, the rice *OsGA20ox1* enzyme prefers GA_{53} to GA_{12} as a substrate (Toyomasu *et al.* 1997). Therefore we measured endogenous concentrations of the 13-hydroxylated GAs (GA_{53}, GA_{44}, GA_{19}, GA_{20}, GA_1 and GA_8) in rice seedlings without leaves and roots sampled at 18 DAI. Compared with Iwatekko, we detected significantly higher levels of GA_{20}, the final product of GA20ox activity, in Dunghan Shali and NIL-114, which is consistent with the *GA20ox1* expression levels in these lines (Fig. 2-9). In Dunghan Shali and NIL-114, the levels of active GA (GA_1) and its catabolic GA (GA_8) were also significantly higher than in Iwatekko (Fig. 2-9).

Effects of *qPHS3-2* on the plant characteristics at maturity

Enhanced seedling growth of NILs harboring *qPHS3-2* prompted us to study its effect on plant gross morphology at the mature stage. Thus, we measured culm length (plant height), panicle length, and panicle number in NIL-114 and Iwatekko in two consecutive seasons, 2010 and 2011 (Table 2-3). We also made additional measurements on internode length, number of grains per panicle in 2011. Culm length of NIL-114 was longer than that of Iwatekko in both seasons, but the difference was statistically significant only in the 2011 experiment. The longer culm length in NIL-114 appears to be due to the proportional elongation of all the internodes, although only the contributions of the 2nd and 4th internodes were significant. Although the differences are subtle, this result suggested that Dunghan Shali allele of *qPHS3-2* enhances not only

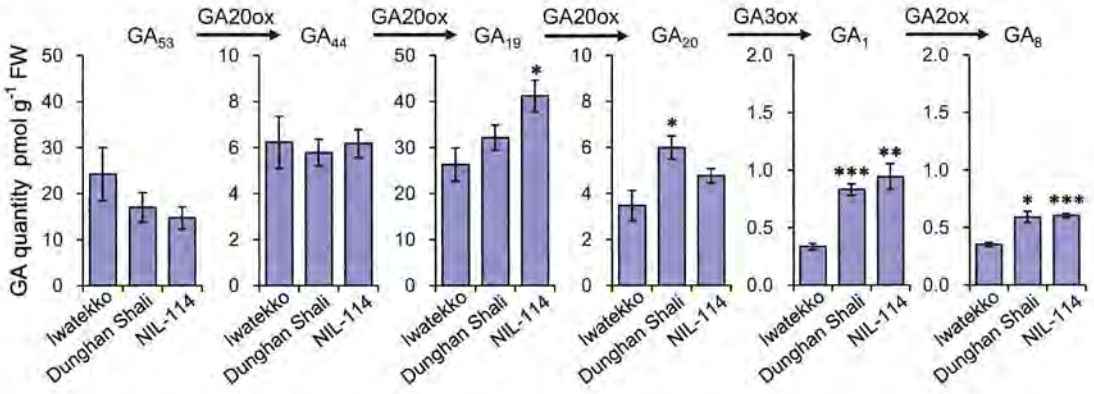

Figure 2-9 Endogenous concentration of GAs in seedlings without leaves and roots at 18 days after start of imbibition.

The GA biosynthesis pathway (Olszewski *et al.* 2002) is shown above graphs. GA values are mean ± SE (n=4). Asterisks indicate significant differences from Iwatekko (Student's t-test, **P* <0.05; ***P* <0.01; ****P* <0.001).

Table 2-3 Phenotypic characteristics of NIL-114 at the mature stage

	Culm length (cm) [a)]	Internode length (cm) [a,b)]					Panicle length (cm) [a)]	No. of panicles per hill	No. of grains per panicle
		I	II	III	IV	V			
2010 (n=8)									
NIL-114 (BC_4F_6)	85.0±1.1	-	-	-	-	-	20.9±0.6	10.4±0.5	-
Iwatekko	82.5±1.9	-	-	-	-	-	18.9±0.3	10.6±0.7	-
	ns	-	-	-	-	-	**	ns	-
2011 (n=10)									
NIL-114 (BC_4F_7)	91.4±0.5	39.2±0.8	24.5±0.4	16.3±0.4	9.5±0.4	2.0±0.5	19.5±0.2	16.9±0.8	84.7±1.4
Iwatekko	85.0±1.0	38.4±0.7	23.1±0.5	14.7±0.7	7.8±0.6	1.0±0.3	20.7±0.4	15.6±0.7	83.4±1.7
	***	ns	*	ns	*	ns	*	ns	ns

a) The longest culm was sampled from hills showing optimum growth for measurements on culm length, internode length and panicle length. Seedlings were transplanted to a paddy field in 2010 and 2011, and the planting pattern was one and three plant per hill, respectively. Asterisks indicate significant differences (Student's t-test, $*P < 0.05$; $***P < 0.001$).

b) Each internode is numbered from top to bottom, such that the uppermost internode just below the panicle is the first (I).

growth at the seedling stage but also promotes internode elongation, thereby resulting in longer culms at the mature stage. On the other hand, we could not find significant differences between NIL-114 and Iwatekko with respect to the other traits measured, i.e. panicle length, panicle number and grain number (Table 2-3).

2-4 Discussion

Seedling vigor is one of the major determinants for stable stand establishment in direct seeding of rice in temperate regions. Previous genetic analyses have revealed that seedling vigor is controlled by several genes (Onishi *et al.* 2007; Redona and Mackill 1996b; Cui *et al.* 2002; Zhang *et al.* 2005a,b; Zhou *et al.* 2007). However, the genes responsible for the enhanced growth of Dunghan Shali seedlings have not been addressed yet. In this study, we identified QTLs for seedling vigor, and demonstrated that the QTL, *qPHS3-2*, that was located on chromosome 3 explains a major part (26.2%) of the total phenotypic variation in the RILs population that was derived from a cross between Kakehashi and Dunghan Shali (Fig. 2-2 and Table 2-2).

Our results show that the height of seedlings in Dunghan Shali is controlled by a QTL with a major effect, *qPHS3-2*, and additional QTLs with minor effects, *qPHS1*, *qPHS3-1* and *qPHS4* (Table 2-2). The major QTL, *qPHS3-2*, was localized to a 12.2 cM interval between the markers RM1373 and RM7389 on chromosome 3. Four QTLs for plant height have been identified by studies involving a cross between Lemont (*O. sativa* L. subsp. *japonica*) and Teqing (*O. sativa* L. subsp. *indica*) (Zhang *et al.* 2005b). Of

these, one QTL was detected near the marker RM148 that was located on chromosome 3, and the Teqing allele in this QTL is associated with the larger plant height (Zhang *et al.* 2005b). In addition, another QTL, *q2LSL3* (leaf sheath length at seedling stage) was detected using RILs derived from a cross between *O. rufipogon* (from India) and a *japonica* strain (Onishi *et al.* 2007). Based on their chromosomal locations, these QTLs appear to correspond to the *qPHS3-2* detected in this study. However further analyses are required to prove the allelic relationship between these QTLs.

In this study, we used a map-based cloning strategy to identify the location of *qPHS3-2* and narrow down its region to 81 kb interval on the long arm of chromosome 3 (Fig. 2-4 and Fig. 2-5). In this region, there are 10 putative genes that were annotated by RAP-DB (Ohyanagi *et al.* 2005) including *OsGA20ox1* (Fig. 2-5). GA20ox (gibberellin 20-oxidase) is an enzyme that normally catalyzes the penultimate steps in GA biosynthesis. In rice, four genes encoding isoforms of GA20ox (*OsGA20ox1*, *OsGA20ox2*, *OsGA20ox3* and *OsGA20ox4*) have been identified. *OsGA20ox2,* located on chromosome 1, is well known as 'Green Revolution gene', and loss-of-function mutation in this gene, *sd1*, causes semi-dwarf phenotype (Sasaki *et al.* 2002; Ashikari *et al.* 2002). In addition, a previous study reported that rice plant statue is regulated not only by *OsGA20ox2* (*SD1*) but also by *OsGA20ox1* (Oikawa *et al.* 2004). The expression levels of *OsGA20ox1* in seedlings of Dunghan Shali, NIL-114 and NIL-497-8, an Iwatekko NIL with a replacement of *qPHS3-2* by Dunghan Shali allele, were significantly higher than that of Iwatekko (Fig. 2-6). Corroborating to this observation, the endogenous GA_1 levels of seedlings in Dunghan Shali and NIL-114 are higher than

that of Iwatekko (Fig. 2-9). These results strongly suggest that the major QTL, *qPHS3-2*, that is involved in the control of height of seedlings corresponds to *OsGA20ox1*. However, we could not detect any nucleotide substitutions in the coding region of *OsGA20ox1* between Dunghan Shali and Iwatekko (Fig. 2-8). We hypothesize that *qPHS3-2* is caused by an unidentified DNA sequence change(s) between Dunghan Shali and Iwatekko that results in differential expression levels of *OsGA20ox1*. Consistent with this hypothesis, a previous study showed that T-DNA insertion in the upstream of *OsGA20ox1* gene (*O. sativa* L., cv. Dontokoi) enhanced its transcription, resulting in the increase in the endogenous GA_1 levels (Oikawa *et al.* 2004). We detected several nucleotide substitutions as well as an insertion/deletion in the region upstream of the *OsGA20ox1*, which might have caused the observed differences in expression levels of the gene (Fig. 2-8).

A GUS activity analysis using the rice cultivar Nipponbare (*japonica*) has revealed that the promoter of *OsGA20ox1* produces low GUS activity in young seedlings (Kaneko *et al.* 2003). In this study, however, we observed the existence of varietal differences in the expression level of *OsGA20ox1* in rice seedlings. Recently, Yano *et al.* (2012) conducted a QTL analysis for plant height at the initial growth stage using backcross inbred lines of Koshihikari (*japonica*) and Habataki (*indica*). Combining QTL-analysis with microarray expression profiling, they detected two major QTLs, *qEPD1* and *qEPD2*, corresponding to *OsGA20ox2* and *OsGA20ox1*, respectively. They concluded that *OsGA20ox1* and *OsGA20ox2* (*SD1*) function during initial growth, but *OsGA20ox1* plays a dominant role in increasing plant height at this stage. Our present

result corroborates with the finding of Yano *et al.* (2012). Together, we propose that *qPHS3-2*, most probably corresponding to *OsGA20ox1*, could be used for enhancing yield rice grown by direct seeding method. The Dunghan Shali allele of *qPHS3-2*, the molecular markers and the NIL developed in this study should be useful to improve seedling vigor in rice breeding programs in temperate regions worldwide.

Chapter 3

Genome sequencing reveals agronomically-important loci in rice using a novel method MutMap

3-1 Introduction

Recent developments in DNA sequencing technology have made whole genome sequencing (WGS) possible in various organisms. WGS has been applied to the identification of causal genes for mutant phenotypes in several model organisms (Schneeberger *et al.* 2009; Austin *et al.* 2011; Uchida *et al.* 2011; Zuryn *et al.* 2010; Ashelford *et al.* 2011). In Arabidopsis, the SHOREmap (Schneeberger *et al.* 2009) and Next Generation Mapping (NGM: Austin *et al.* 2011) methods employ WGS to identify mutations responsible for given phenotypes. In both methods, mutants in the Colombia (Col) ecotype background are crossed to a distantly related ecotype Landsberg (Ler) that functions as a mapping line, and derived F_2 progeny are scored for their phenotypes. DNA from the F_2 progeny exhibiting the mutant phenotype is bulked and sequenced for identification of the genomic regions derived from the Col mutant parent. These methods, modern day adaptations of bulked-segregant analysis (BSA: Michelmore *et al.* 1991; Giovannoni *et al.* 1991), use Single Nucleotide Polymorphisms (SNPs) present between distant parents (Col and Ler) as DNA markers to locate the causal mutations. However, F_2 progeny derived from such a distant cross show a wide phenotypic variation due to segregation of a large number of loci, making these methods unsuitable for the isolation of minor effect genes. Therefore, SHOREmap and NGM have been confined so far to the isolation of genes causing major phenotypic effects in Arabidopsis, and, to date crop genes have not been isolated using these methods.

In this study, we have generated large-scale mutant line populations of an elite

Japanese rice cultivar and applied MutMap, a novel method of rapid gene isolation employing a cross of the mutant to wild-type parental line. We have used MutMap to localize genomic positions of rice genes controlling agronomically-important traits including semi-dwarfism. Given that the mutant plants and associated molecular markers can be immediately made available to plant breeders, this approach has the potential to dramatically accelerate crop breeding and genetics.

3-2 Materials and Methods

Rice mutant lines

Immature embryos of the cultivar "Hitomebore" were mutagenized by immersing the panicles in 0.015% EMS solution overnight (see Rakshit *et al.* 2010 for details). The resulting M1 plants were self-pollinated, and M2 seeds were obtained. M2 plants were further self-pollinated to obtain M3 progeny, resulting in a total of 12,000 M3-M4 lines.

Whole genome sequencing of bulked DNA

DNA was extracted from 100 mg of fresh rice leaves using the DNeasy Plant Mini Kit (QIAGEN Sciences, Maryland, USA). DNA was quantified using the Quant-iT PicoGreen dsDNA Reagent and Kits (Invitrogen, Oregon, USA). To make bulked DNA from F_2 progeny, DNA from F_2 individuals was mixed in an equal ratio. Five

micrograms of the mixed DNA was used for preparation of libraries for Illumina sequencing according to the protocol for the Paired End DNA sample Prep Kit (Illumina, San Diego, CA, USA). The libraries were used for cluster generation on a flow cell and sequenced for 76 cycles on an Illumina Genome Analyzer IIx. Base calling and filtering of low quality bases were performed using sequence control software (SCS) real-time analysis, BCL converter and the GERALD module (Illumina, San Diego, CA, USA).

Alignment of short reads to reference sequences and SNP calling

To identify mutations incorporated by EMS, we generated a reference sequence of the "Hitomebore" wild type genome based on the publicly available "Nipponbare" rice genome sequence (IRGSP 2005). First, we obtained a total of 1,083 million paired-end short reads from "Hitomebore" wild type and 11 mutant lines. These short reads were pooled and aligned with MAQ (Li *et al.* 2008) to the "Nipponbare" reference sequence. Alignment files were converted to SAM/BAM files using SAMtools (Li *et al.* 2009), and applied to a filter pipeline (Kosugi *et al.* manuscript in preparation) to identify reliable SNPs. This filter pipeline was developed to maximize true SNP detection and minimize the false SNP calling by 1) removal of the paired-end reads of an insert size more than 325 bp, 2) calling SNPs only for the genomic regions covered by a minimum of three reads for homozygous SNPs and five reads for heterozygous SNPs and a maximum of three fold of the average read depth over the genome, 3) calling SNPs only on the sites with an averaged Illumina phred-like quality score of $\geqq 20$. This was further optimized

by test data: short reads of "Nipponbare" experimentally obtained by Illumina GAIIx were aligned to the "Nipponbare" reference sequence containing a total of 859,555 artificial nucleotide substitutions at known positions, resulting in the successful calling of 82% of the true SNPs and at the same time keeping the level of false SNP calling to 0.1%. Using this pipeline, we identified 100,819 reliable SNPs between "Hitomebore" reads and the "Nipponbare" reference sequence. Based on this result, we generated a "Hitomebore" reference sequence (DDBJ Project ID67163) by replacing "Nipponbare" nucleotides with those of "Hitomebore" at 100, 819 sites. In order to remove the effect of SNPs irrelevant to the mutant screen, it was important to generate and use a reference sequence of the same wild type Hitomebore line that was used for mutagenesis. We were able to further refine this reference sequence by taking a consensus of cumulative genome sequences of the mutants.

Paired-end sequence reads of bulked DNA of mutant F_2 progeny were aligned to the "Hitomebore" reference sequence, and SNPs were scored. We divided SNPs into two categories: homozygous SNPs and heterozygous SNPs. Homozygous SNPs were defined as SNPs with "SNP index $\geq$ 0.9" and a minimum coverage of the sites of three reads. Heterozygous SNPs were defined as SNPs with SNP index $\geq$ 0.3 and SNP index $<$ 0.9 with a coverage of the position of more than four reads. We further filtered the SNPs with two more steps: (1) removal of common SNPs that are shared by at least two mutant lines, (2) extraction of SNPs that exhibit 'G to A' or 'C to T' transitions, which are known to be the most frequent changes caused by EMS-mutagenesis. After identifying the genomic regions harboring a cluster of SNPs with SNP index =1, we

relaxed the condition of the filter to consider all the SNPs (caused by all the transition as well as transversion) in the region as candidate SNPs for the causal mutation.

SNP index plot regression lines were obtained by averaging the SNP-indices from a moving window of five consecutive SNPs, shifting the window one SNP at a time. The x-axis value of each averaged SNP-index was set at a midpoint between the first and fifth SNP.

Genetic complementation of the Hit1917-pl1 mutant

For the complementation test, a 6,696 bp genomic fragments containing the *OsCAO1* gene with 2kb upstream and 1kb downstream of the transcribed region was amplified from wild-type Hitomebore genomic DNA by PCR and subcloned into the binary vector pGWB1 (Nakagawa *et al.* 2007) to yield pGWB1-OsCAO1. As the control, pCAMBIA-empty was used in this study. For RNAi of the *OsCAO1* gene, a 337 bp *OsCAO1* partial fragment was amplified from wild-type Hitomebore genomic DNA by PCR and subcloned into the binary vector pANDA (Miki *et al.* 2005) to yield pANDA-CAO1. All binary vectors were introduced into Agrobacterium strain EHA105 for rice transformation. Hitomebore plants were transformed as previously described (Toki *et al.* 2006).

3-3 Results

Principle of MutMap

The principle of MutMap is shown in Figure 3-1 and is explained by taking rice as an example. We first mutagenize a rice cultivar "X" that has a reference genome sequence with a mutagen, e.g. ethyl methanesulfonate (EMS). Mutagenized plants of the M1 generation are self-pollinated and brought to the M2 or more advanced generations to make the mutated gene homozygous. Observation of phenotypes in the M2 lines or later generations allows identification of recessive mutants with altered agronomically-important traits including plant height, tiller number, grain number/spike etc. Once the mutant is identified, it is crossed with the wild-type plant of the cultivar "X", the same cultivar used for mutagenesis. The resulting F_1 plant is self-pollinated, and F_2 progeny (>100) are grown in the field for scoring the phenotype. Since these F_2 progeny are derived from a cross between the mutant and its parental wild-type plant, segregation of phenotypes can be unequivocally observed even if the phenotypic difference is small. All the nucleotide changes incorporated into the mutant by mutagenesis are detected as single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) between the mutant and the wild type. Among the F_2 progeny, the majority of SNPs will segregate in a 1:1 ratio for the mutant and wild types. However, the SNP responsible for the change of phenotype is homozygous in the progeny showing the mutant phenotype. If we collect DNA samples from recessive mutant F_2

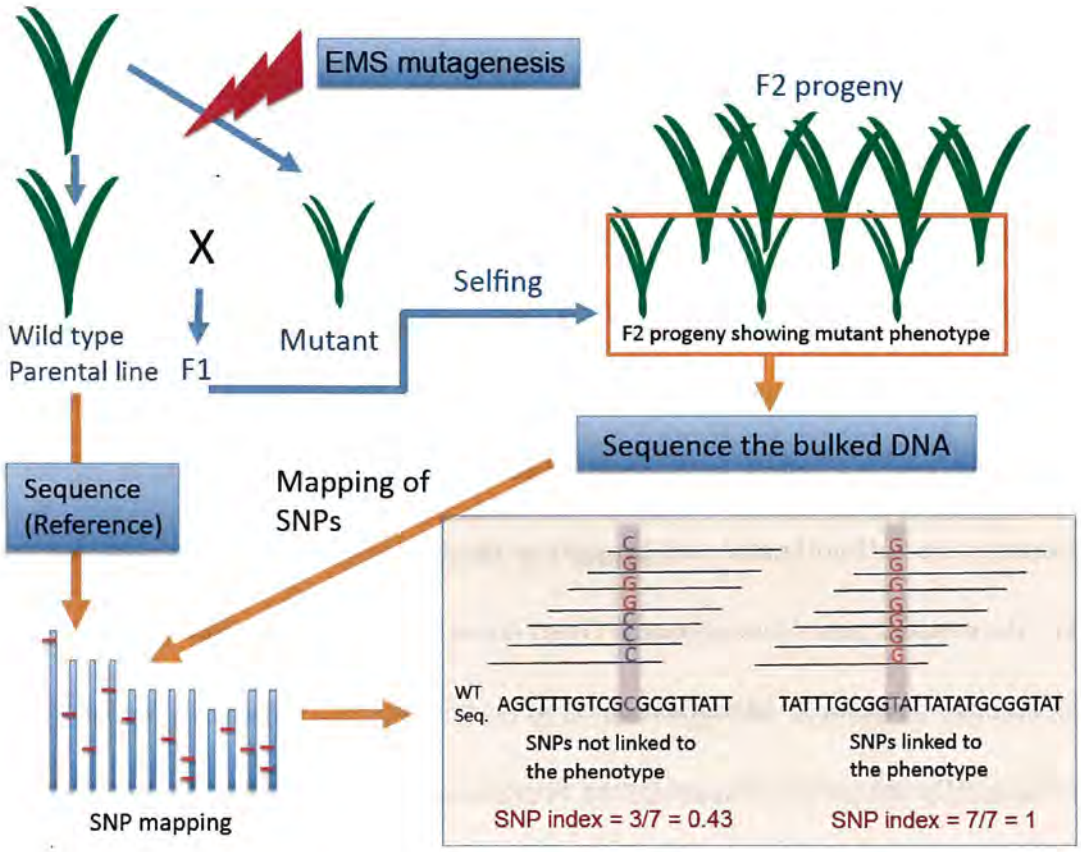

Figure 3-1. A simplified scheme for the application of MutMap to rice.

A rice cultivar with a reference genome sequence is mutagenized by ethyl methanesulfonate (EMS). The mutant generated, in this case a semi-dwarf mutant, is crossed to the wild-type plant of the same cultivar that was used for the mutagenesis. The resulting F1 is self-pollinated to obtain F2 progeny segregating for the mutant and wild-type phenotypes. Crossing of the mutant to the wild-type parental line ensures the detection of phenotypic differences at the F2 generation between the mutant and wild type. DNA of F2 displaying the mutant phenotype are bulked and subjected to whole genome sequencing followed by alignment to the reference sequence. SNPs with sequence reads composed only of mutant sequences (SNP index = 1; see main text) are closely linked to the causal SNP for the mutant phenotype.

progeny and bulk sequence them with a substantial genome coverage (> 10X coverage), we expect to have 50% mutant-type and 50% wild-type sequence reads for SNPs that are unlinked to the SNP responsible for the mutant phenotype. However, the causal SNP as well as closely linked SNPs should show 100% mutant-type with 0% wild-type reads. SNPs loosely linked to the causal mutation should have > 50% mutant-type and < 50% wild-type reads. If we define the frequency of mutant-type SNPs among the total reads as the SNP index, we expect to have an SNP index equal to 1 (SNP index =1) near the causal gene and an SNP index equal to 0.5 for the unlinked loci. SNP indices can be scanned across the genome to find the region with a SNP index =1, which harbors the gene responsible for the mutant phenotype.

The rate of false positives can be easily assessed because the allelic segregation follows a binomial distribution with a parameter of 0.5 at a SNP with no linkage to the causal SNP. If the sample size (read depth of the site) is 10, the probability of having SNP index = 1 is $p = (0.5)^{\wedge}10=10^{-3}$. Therefore, in a data set with a known number of genotyped SNPs (L), the expected number of clusters of SNPs with SNP index = 1 ($\geq k$) would be approximately $pkL = 10^{-3k}L$. In our case, the maximum estimate of L is 2225, and the probability of observing a cluster of more than four consecutive SNPs with SNP index =1 would be $\leqq 2.3\text{x}10^{-9}$. Statistical considerations of how the number of F_2 progeny to be bulked and the average coverage (depth) of genome sequencing affect the false positive rate and misclassification of phenotypes between mutant and wild-type affects the true positives are given in Supplementary Data that was conducted by Dr. H. Innan.

MutMap applied to rice pale green leaf mutants

We have been maintaining a total of 12,000 EMS-mutagenized rice lines of M3-M4 generations with a background of "Hitomebore", an elite cultivar of Northern Japan (Rakshit *et al.* 2010). Whole genome resequencing of five independent mutants indicated that each line harbors 1499±469 (mean±standard deviation; range 960-2225) SNPs that are different from the wild type (Table 3-1). Using our mutant stock, we set out to isolate genes of agronomic importance. As a proof-of-principle, we applied MutMap to two mutants showing pale-green leaf phenotypes with slightly reduced chlorophyll levels (Hit1917-pl1 and Hit0813-pl2; Fig. 3-2a)

We crossed these mutants to the "Hitomebore" wild-type in 2009, and obtained F_1 progeny. F_1 plants were self-pollinated and > 200 F_2 progeny were obtained for each cross. For both mutants we observed a clear segregation between wild-type and mutant phenotypes in field-grown F_2 progeny, with the ratio conforming to 3:1 (Table 3-2), suggesting that each mutant phenotype was caused by a recessive mutation in a single locus. For each cross, DNA of 20 F_2 progeny showing the mutant phenotype was isolated, and bulked in an equal ratio. This bulked DNA was subjected to whole genome sequencing using an Illumina GAIIx sequencer. We obtained a total of 70 and 133 million sequence reads (75bp) for Hit1917-pl1 and Hit0813-pl2, respectively, corresponding to > 5 Gb in total read length with > 12X coverage of the rice genome (370 Mb) (Table 3-3). These reads were aligned to a reference sequence of "Hitomebore" using the MAQ software (Li *et al.* 2008). Aligned data were passed through a filter to

Table 3-1. Number of reliable SNPs detected between five EMS-mutant lines and wild type Hitomebore.

Hitomebore mutant line	Reference sequence	Number of SNPs detected
Hit0220	Hitomebore	2,225
Hit3069	Hitomebore	1,272
Hit4529	Hitomebore	1,582
Hit5239	Hitomebore	960
Hit5922	Hitomebore	1,455

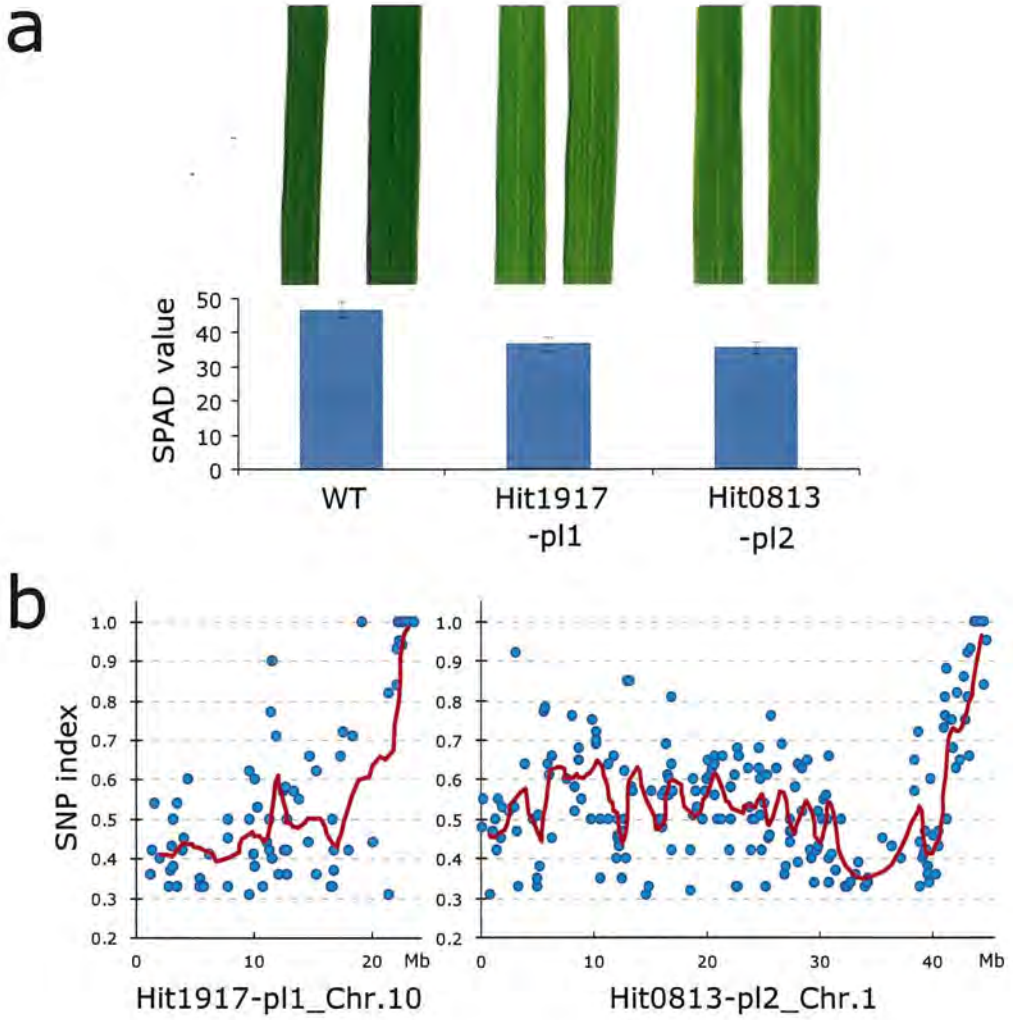

Figure 3-2. Identification of the genomic regions harboring causal mutations for two pale green leaf mutants Hit1917-pl1 and Hit0813-pl2, using MutMap.

a: Leaf color and SPAD values (an estimation of chlorophyll content) of the wild-type Hitomebore (WT) and two mutants (Hit1917-pl1 and Hit0813-pl2). b: SNP index plots for two leaf color mutants (Hit1917-pl1 and Hit0813-pl2) showing chromosomes 10 and 1, respectively. The red regression lines were obtained by averaging the SNP-indices from a moving window of five consecutive SNPs, shifting the window one SNP at a time. The x-axis value of each averaged SNP-index was set at a midpoint between the first and fifth SNP.

Table 3-2. A table showing segregation of wild-type (WT) vs. mutant phenotypes among the F2 progeny of seven mutants as studied by MutMap.

Mutant line	Phenotype	WT type	Mutant type	3:1 segregation $\chi 2$
Hit1917-pl1	Pale green leaves	274	84	0.23 (NS)
Hit0813-pl2	Pale green leaves	143	71	3.47(NS)
Hit1917-sd	Semi dwarf	309	96	0.19(NS)
Hit0746-sd	Semi dwarf	181	40	3.12(NS)
Hit5500-sd	Semi dwarf	185	69	0.31(NS)
Hit5814-sd	Semi dwarf	269	60	4.45 ($P<0.05$)
Hit5243-sm	Male sterility	296	115	0.93(NS)

Table 3-3. Summary of Illumina GAIIx sequencing of bulked DNAs of seven mutant lines*.

Mutant Name	No. of lanes	Conc. (pM)	Reference genome	No. of reads	Total size (Gb)	Coverage (%)	Read depth
Hit1917-pl1	1	8	Hitomebore consensus	69,826,922	5.24	95.2	12.5
Hit0813-pl2	2	8	Hitomebore consensus	133,286,864	10.00	96.0	24.1
Hit1917-sd	1	8	Hitomebore consensus	89,564,924	6.72	84.1	16.6
Hit0746-sd	1	8	Hitomebore consensus	74,738,872	5.61	84.2	14.0
Hit5500-sd	1	8	Hitomebore consensus	69,887,350	5.24	82.4	13.2
Hit5814-sd	1	8	Hitomebore consensus	75,209,760	5.64	84.6	14.2
Hit5243-sm	1	8	Hitomebore consensus	77,124,718	5.78	96.2	14.1

*No. of lanes and Conc. indicate the number of lanes and concentration of the libraries used for Illumina sequencing, respectively. Sequence reads of bulk DNA of mutant F2 progeny were aligned to the Hitomebore consensus sequence as described in the Materials and Methods. Coverage and Read depth were calculated based on the alignment of short reads to the reference genome.

reduce spurious SNP calls caused by sequencing and alignment errors (Materials and Methods). As a result, we found that the mutants harbor 1,001 (Hit1917-pl1) and 1,339 (Hit0813-pl2) transition type (G to A and C to T) SNPs with high quality scores (Table 3-4), presumably caused by EMS mutagenesis (Pienkowska *et al.* 1993).

For each identified SNP, we obtained the SNP index defined as the ratio between the number of reads of a mutant SNP and the total number of reads corresponding to the SNP. The SNP indices were graphically plotted for all 12 chromosomes of rice (Fig. 3-3). As expected, SNP indices were distributed randomly around 0.5 for most part of the genome for the two mutant lines. The noise observed in the dispersion of SNP indices is not unexpected given the large number of SNPs and the stochastic nature of allelic segregation at individual SNP, which follows a binomial distribution with parameter 0.5 (See Supplementary Data). A single unique genomic region harboring a cluster of SNPs exhibiting SNP index = 1 was identified for each mutant (Fig. 3-2b): the Hit1917-pl1 mutant showed a cluster of seven SNPs with SNP index =1 on chromosome 10, while Hit0863-pl2 had a cluster of five SNPs with SNP index =1 on chromosome 1 (Fig. 3-2b, Table 3-4, Fig. 3-3). These results demonstrate that MutMap allows rapid identification of the putative position of a causal mutation responsible for a mutant phenotype.

Identification of the causal SNP of a pale green leaf mutant using MutMap

Taking the example of Hit1917-pl1, a pale green leaf mutant, we further examined the

Table 3-4. Summary of SNPs detected between "Hitomobore" wild-type and bulked F2 DNA of seven mutants.

Mutant line	Number of all SNPs *1		Number of G' A or C' T transition SNPs *3				
	Homo+ Hetero	Homo *2	Homo+ Hetero	Homo *2		No of SNPs with "SNP index" =1	
	All *4	All *4	All *4	All *4	No. of SNPs in a cluster *5	All *4	No. of SNPs in a cluster *5
Hit1917-pl1	10,360	101	1,001	24	10	14	7
Hit0813-pl2	6,187	238	1,339	10	7	5	5
Hit1917-sd	3,432	139	896	10	9	7	6
Hit0746-sd	3,701	121	1,220	13	3	5	4
Hit5500-sd	3,420	131	553	9	4	6	4
Hit5814-sd	2,957	121	448	4	4	4	4
Hit5243-sm	7,791	229	735	21	11	17	11

*1 We define SNP as the site in which ≧30% of sequence reads show the nucleotide to be different from that of the wild type "Hitomebore". They include mutant-type homozygous SNPs (Homo) as well as mutant-type/wild- type heterozygous SNPs (Hetero).

*2 Homozygous SNPs are defined as the SNPs with ≧90% of mutant-type reads among the total reads covering the positions.

*3 From among the all SNPs detected, those caused by G to A or C to T transitions were selected. EMS mutagenesis is known to be predominantly transitions.

*4 SNPs contained in all the chromosomes.

*5 SNPs contained in short genomic stretches with a cluster of SNPs with higher SNP index values (≧0.9). The exact genomic locations of the clusters and details of SNPs therein are given below:

Mutant lines	Regions of SNP cluster with homozygous SNPs (SNP index ≧ 0.9)	Number of SNPs of GtoA or CtoT transition			Number of other SNPs		
		Total SNPs	SNPs in gene ORF	SNPs causing amino acid change	Total SNPs	SNPs in gene ORF	SNPs causing amino acid change
Hit1917-pl1	Chr10: 22179795..23648959	10	5	2	0	0	0
Hit0813-pl2	Chr1: 43488206..44890284	7	3	0	0	0	0
Hit1917-sd	Chr12: 23171444..24815280	9	5	0	1	0	0
Hit0746-sd	Chr8: 27360781..28128481	7	2	1	0	0	0
Hit5500-sd	Chr9: 19720463..20768178	4	2	0	2	1	1
Hit5814-sd	Chr4: 22949403..26036923	4	4	2	4	3	2
Hit5243-sm	Chr8: 7479242..12890545	11	5	2	20	1	1

a

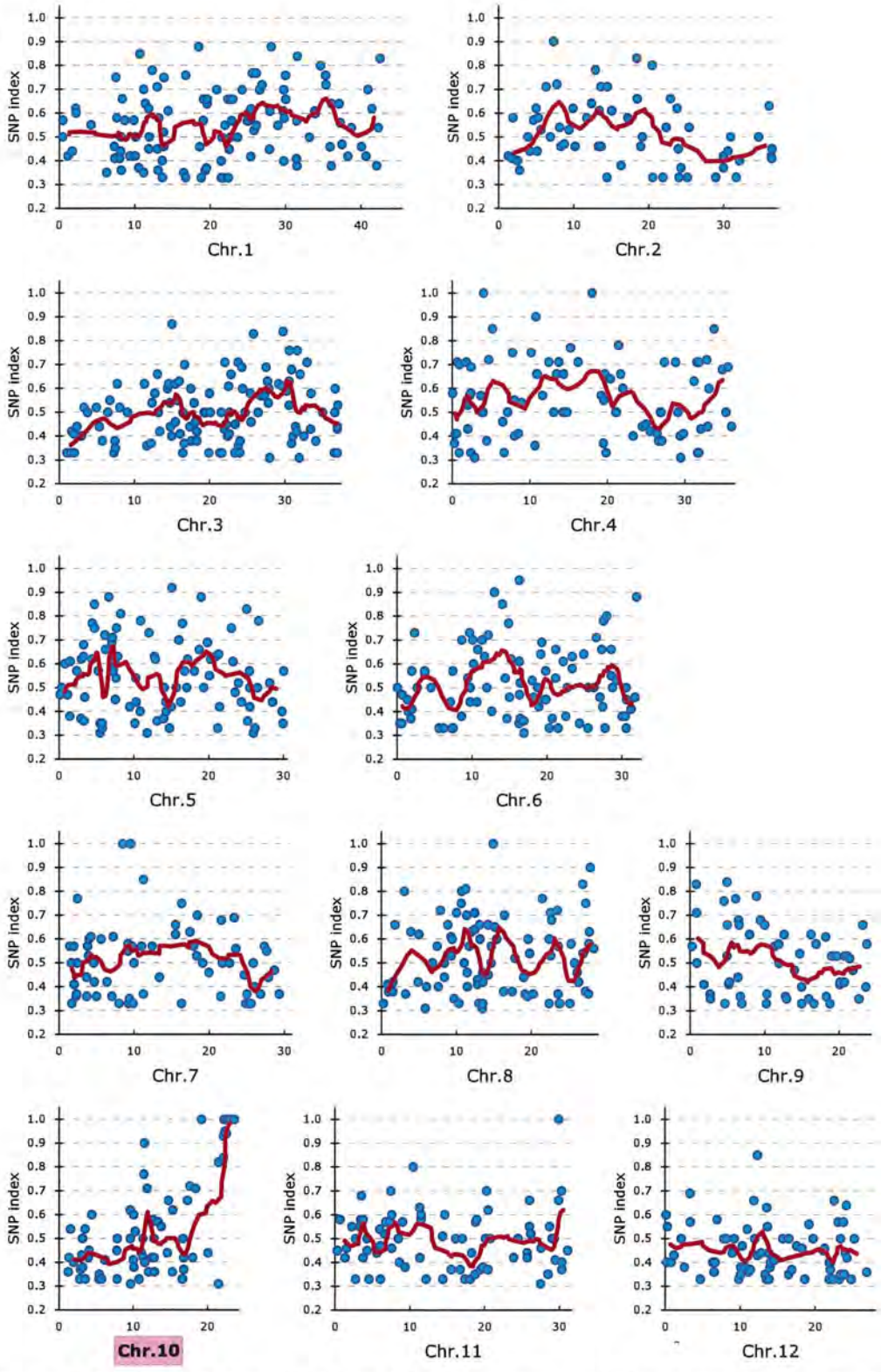

Figure 3-3. SNP index plots for 12 chromosomes of the seven rice mutants (a:Hit1917-pl1, b:Hit0813-pl2, c:Hit1917-sd, d:Hit0746-sd, e:Hit5500-sd, f:Hit5814-sd and g:Hit5243-sm).

Blue points indicate SNP positions and their SNP indices. Red lines are regression lines.

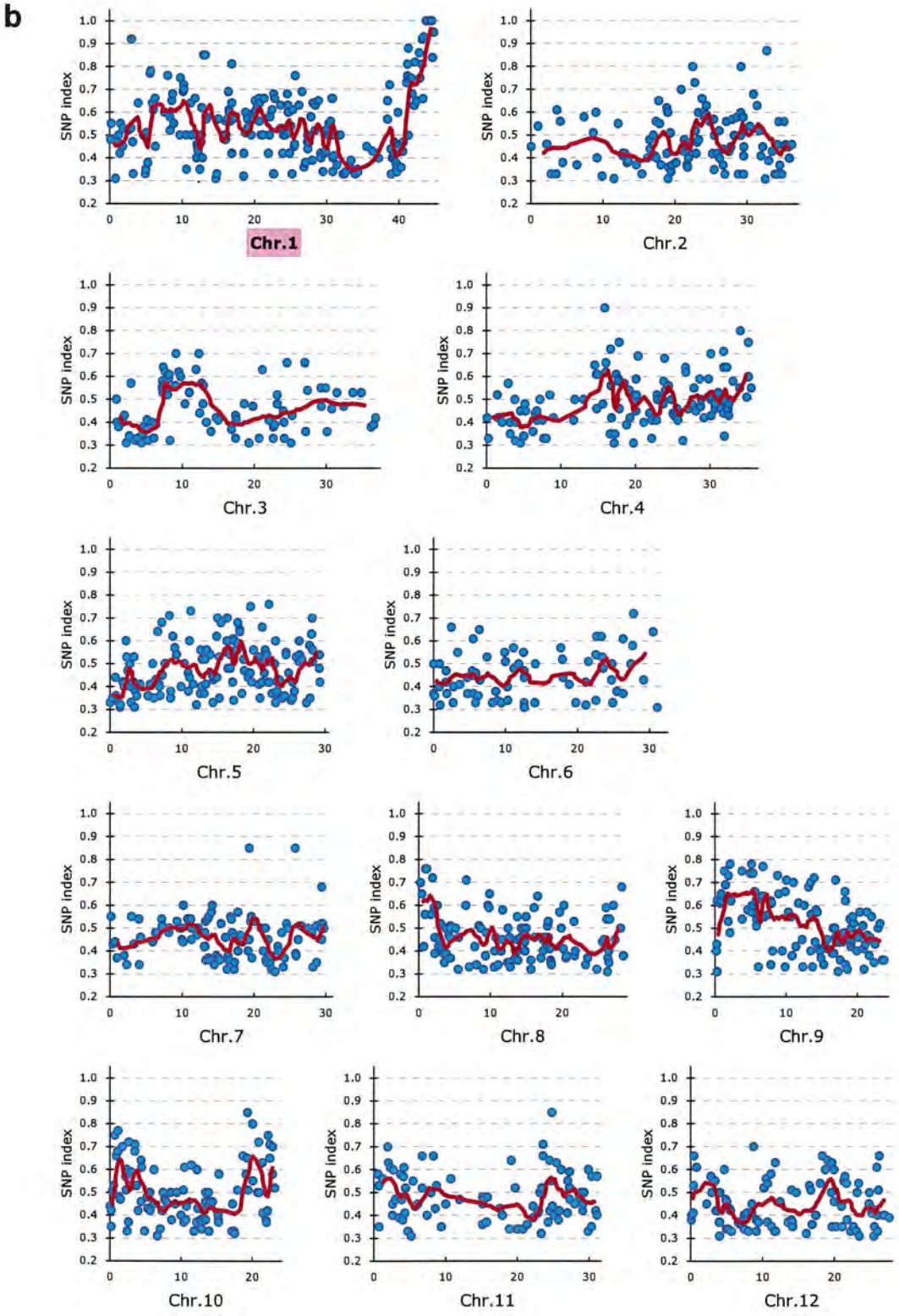
b
SNP index
Chr.1
Chr.2
Chr.3
Chr.4
Chr.5
Chr.6
Chr.7
Chr.8
Chr.9
Chr.10
Chr.11
Chr.12

Figure 3-3b.

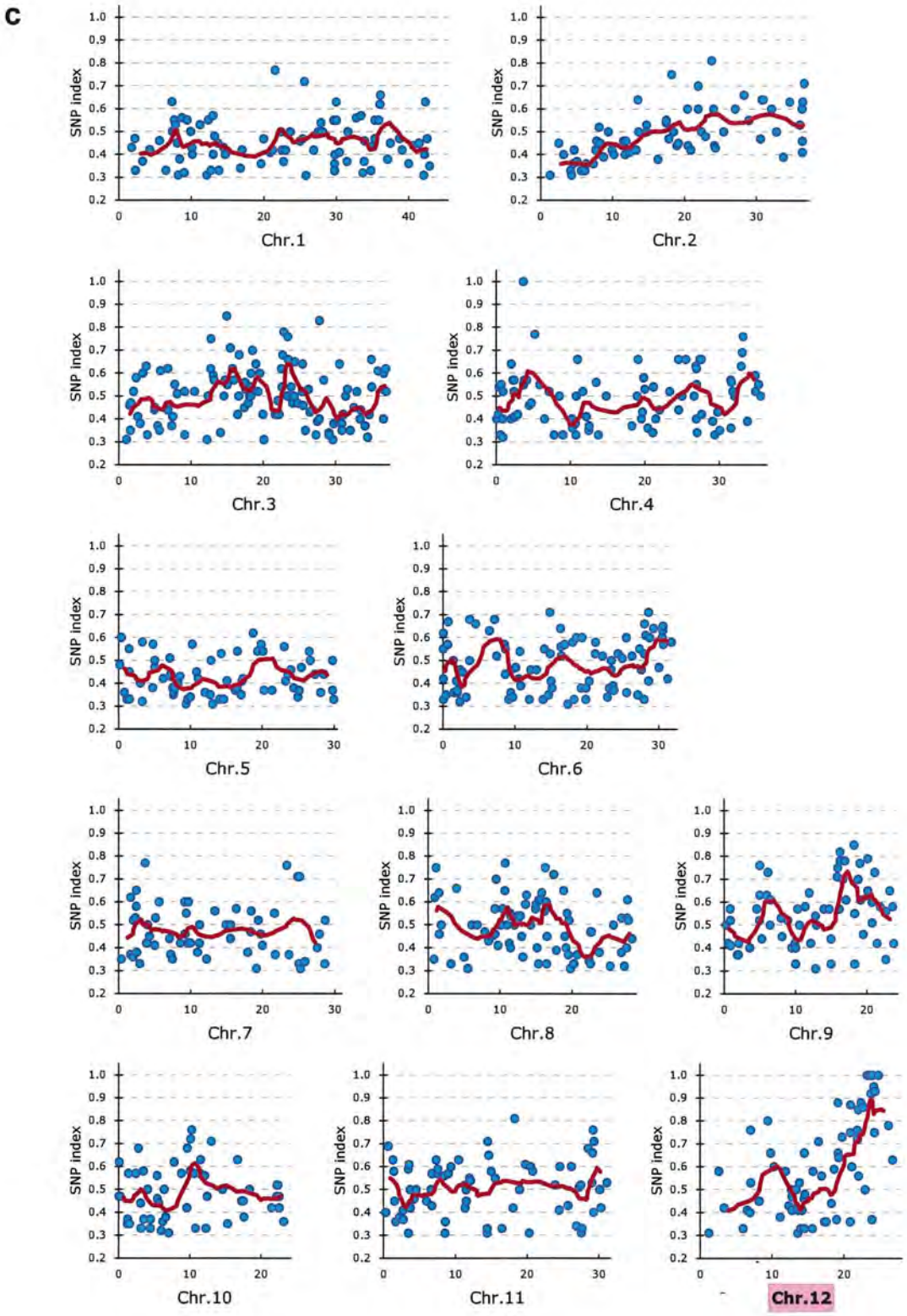
c
SNP index
Chr.1
Chr.2
Chr.3
Chr.4
Chr.5
Chr.6
Chr.7
Chr.8
Chr.9
Chr.10
Chr.11
Chr.12

Figure 3-3c.

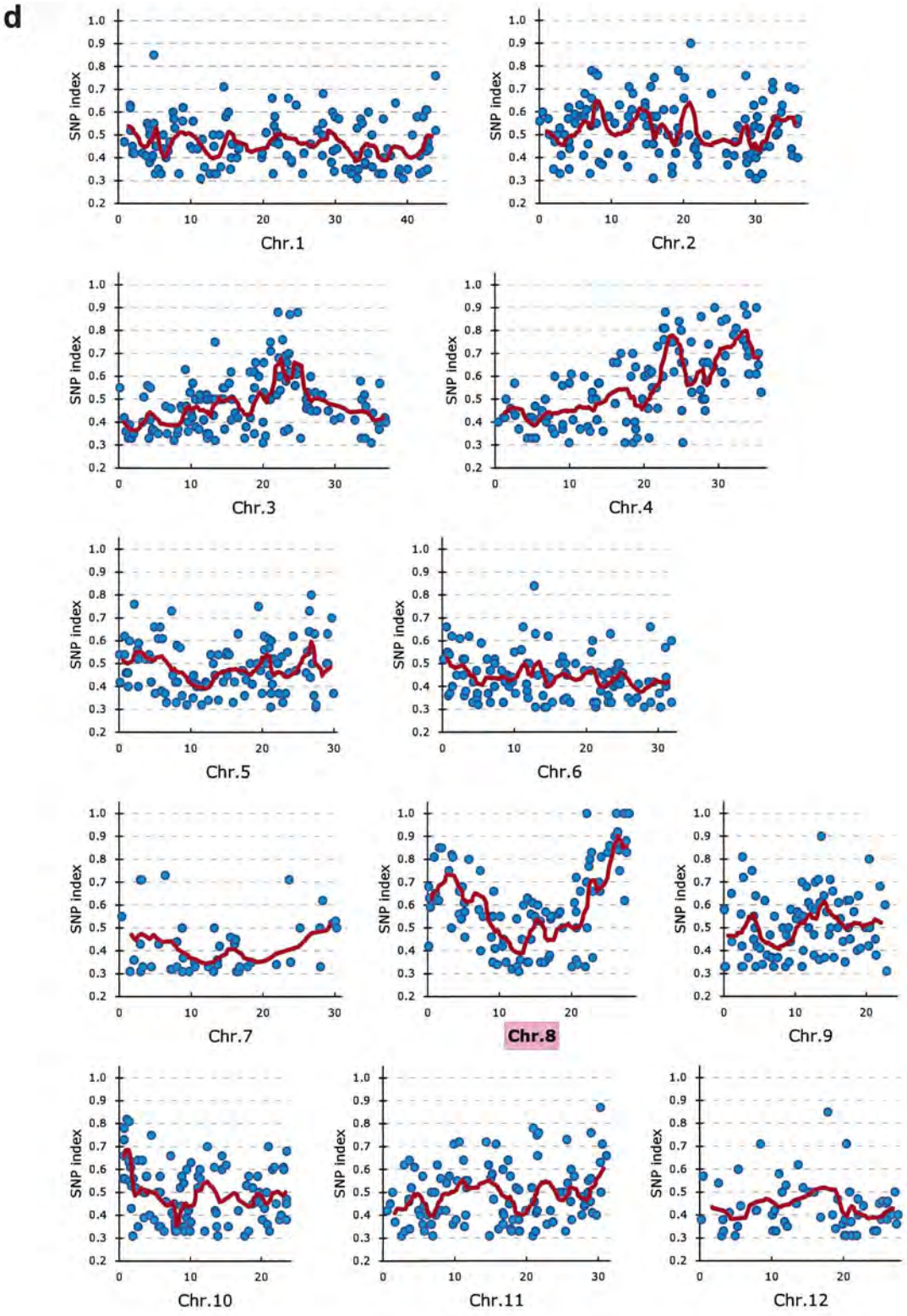
d
SNP index
Chr.1
Chr.2
Chr.3
Chr.4
Chr.5
Chr.6
Chr.7
Chr.8
Chr.9
Chr.10
Chr.11
Chr.12

Figure 3-3d.

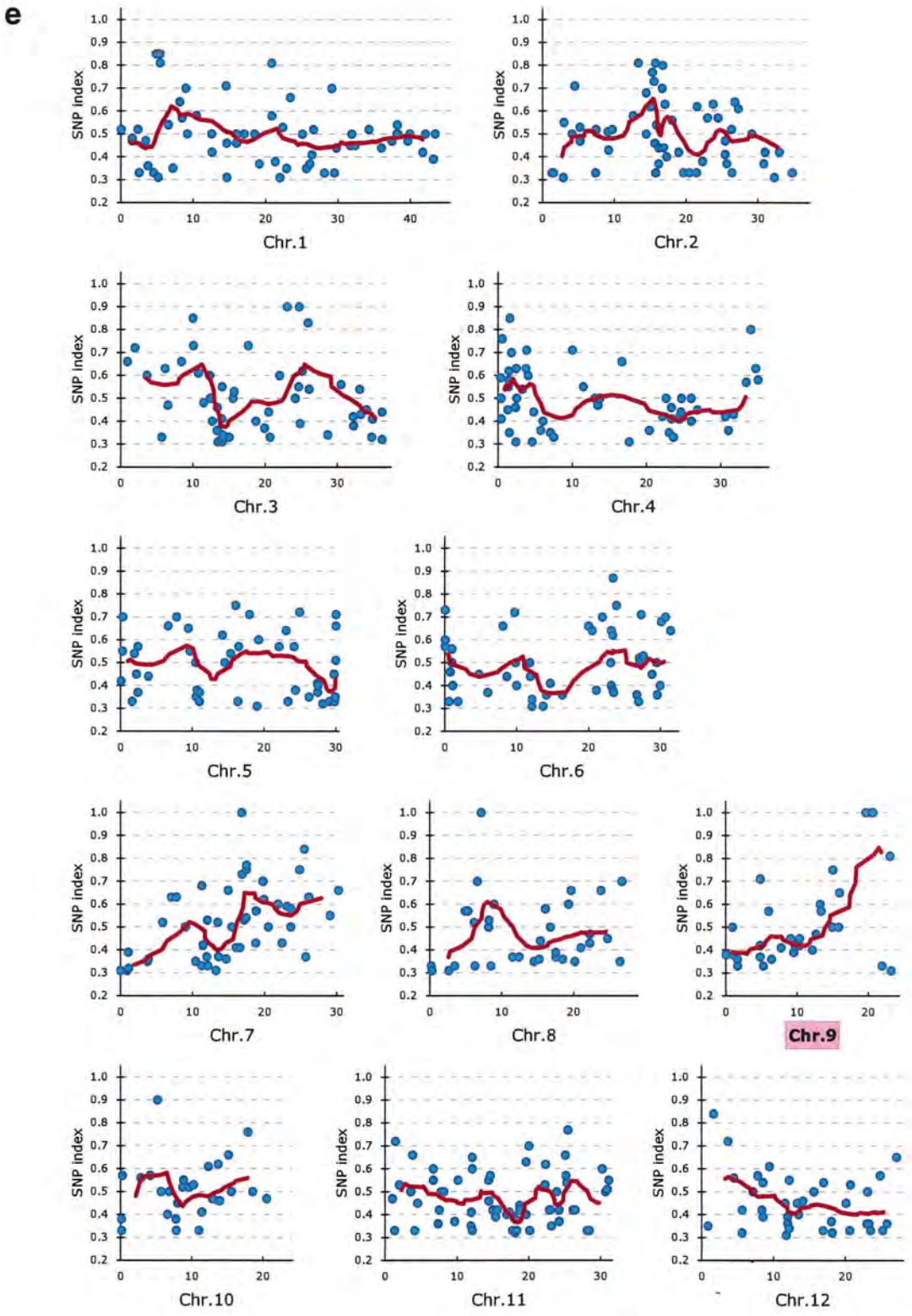
e
SNP index
Chr.1
Chr.2
Chr.3
Chr.4
Chr.5
Chr.6
Chr.7
Chr.8
Chr.9
Chr.10
Chr.11
Chr.12

Figure 3-3e.

f

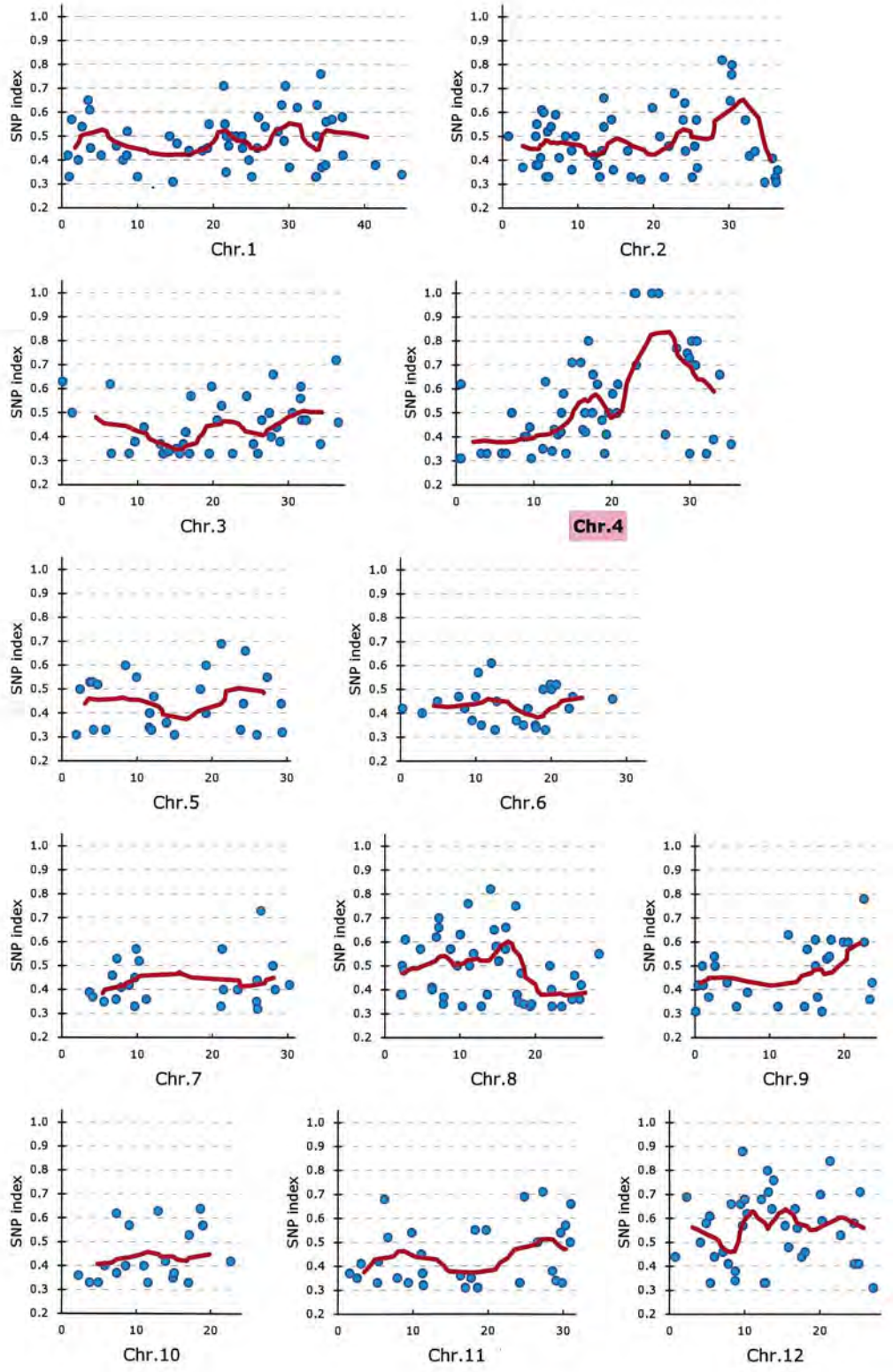
SNP index
Chr.1
Chr.2
Chr.3
Chr.4
Chr.5
Chr.6
Chr.7
Chr.8
Chr.9
Chr.10
Chr.11
Chr.12

Figure 3-3f.

g

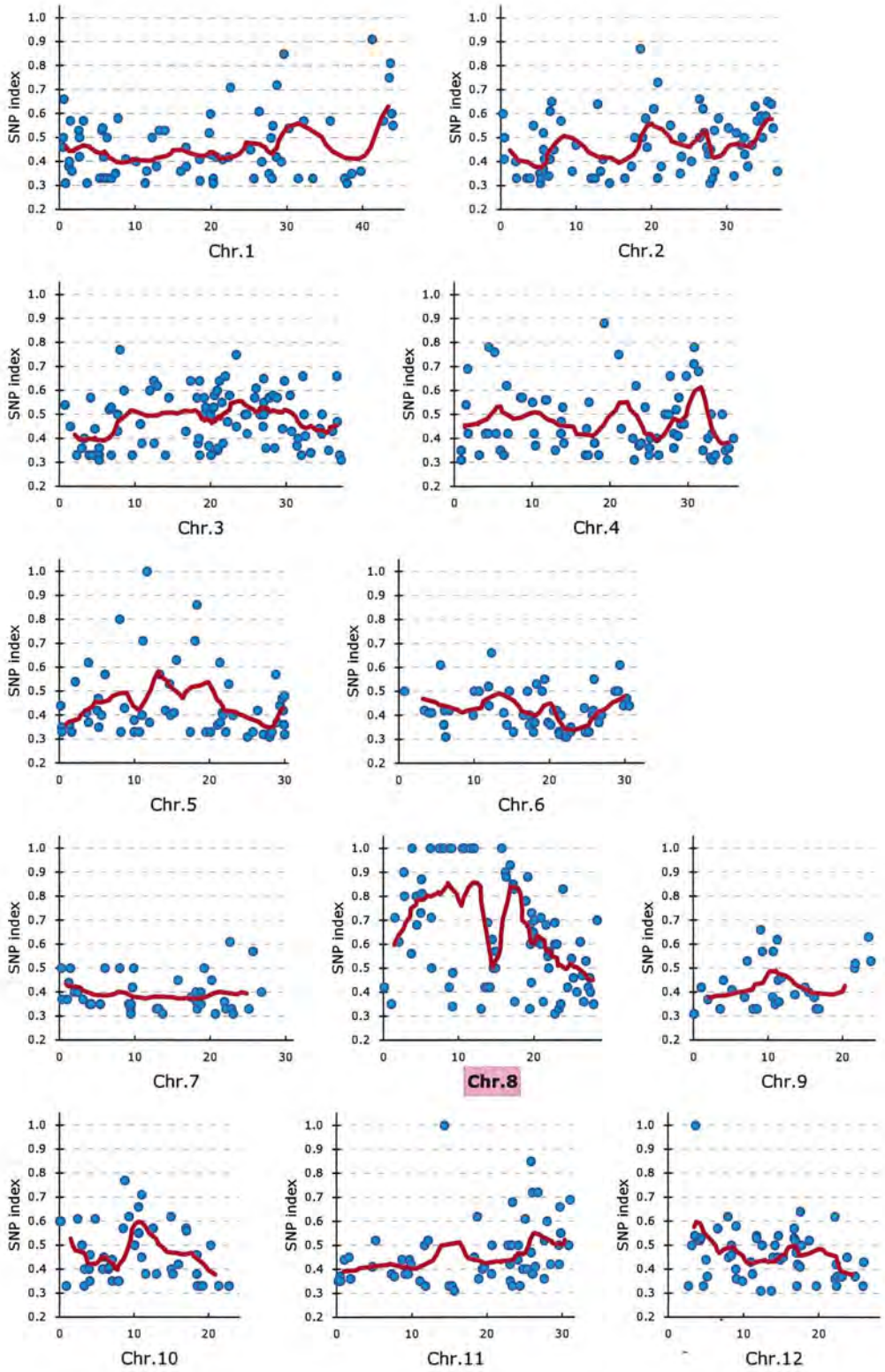
SNP index
Chr.1
Chr.2
Chr.3
Chr.4
Chr.5
Chr.6
Chr.7
Chr.8
Chr.9
Chr.10
Chr.11
Chr.12

Figure 3-3g.

SNPs showing SNP index =1 in detail. Of the seven SNPs showing SNP index = 1, two corresponded to exons of protein coding genes. SNP-22981826 is localized to the gene for chlorophyllide a oxygenase (*OsCAO1*) leading to an amino acid change of Leu253 (codon CTT) to Phe (TTT), while SNP-23202531 corresponded to a Zinc finger-RING/FYVE/PHD-type domain protein leading to a change of Ala106 (GCA) to Val (GTA) (Table 3-5). There is a report of T-DNA insertion knockout of the *OsCAO1* gene (Lee *et al.* 2005) showing that the *OsCAO1* mutant exhibits reduced chlorophyll, similar to our Hit1917-pl1 mutant. Therefore, we hypothesized that Hit1917-pl1 is caused by a non-synonymous substitution in the *OsCAO1* gene.

To verify this hypothesis, we carried out a complementation study by transforming the Hit1917-pl1 mutant with the wild-type *OsCAO1* gene driven by the native promoter. We also made a knockdown mutant of *OsCAO1* by transforming wild-type plants with an RNAi construct targeting the *OsCAO1* gene (Fig. 3-4). The Hit1917-pl1 mutant transformed with wild-type *OsCAO1* expressed both mutant as well as wild-type alleles of *OsCAO1* (Fig. 3-4b), and its phenotype was restored to wild-type (Fig. 3-4a, Fig. 3-5). As expected, the wild-type plant transformed with the RNAi construct for *OsCAO1* showed a reduced level of *OsCAO1* transcripts (Fig. 3-5) and a more enhanced pale green phenotype than Hit1917-pl1 (Fig. 3-4a). These data demonstrate that the Hit1917-pl1 phenotype is caused by the mutation SNP-22981826 identified by MutMap.

Table 3-5. Candidate causal mutations identified by MutMap.

Mutant lines	Chromosome coordinate	Reference base	Altered base	Type of mutation	Gene annotation and note
Hit1917-pl1	chr10: 22981826	C	T	Missense (L to F)	Chlorophyllide a oxygenase (Os10t0567400)
	chr10: 23202531	C	T	Missense (A to V)	RING-type domain containing protein (Os10t0572500)
Hit0813-pl2	Not identified	-	-	-	No SNPs causing amino acid changes
Hit1917-sd	chr12: 23426603	G	A	Mutation at a splicing junction	Succinyl-CoA synthetase-like (Os12t0572800)
Hit0746-sd	chr8: 27726540	G	A	Missense (V to M)	Conserved hypothetical protein (Os08t0551200)
Hit5500-sd	chr9: 20768178	T	A	Nonsense (L to *)	Cdc48-like protein (Os09t0515100)
Hit5814-sd	chr4: 24054530	A	T	Nonsense (R to *)	Similar to H0418A01.7 protein (Os04t0471400)
	chr4: 24534915	A	T	Missense (E to V)	Leucine-rich repeat-containing protein (Os04t0480500)
	chr4: 25169073	C	T	Missense (L to F)	Conserved hypothetical protein (Os04t0493300)
	chr4: 26036923	C	T	Missense (R to C)	RNA helicase-like protein (Os04t0510400)
Hit5243-sm	chr8: 7491119	C	T	Missense (A to V)	EMBRYO DEFECTIVE 1135 (Os08t0223700)
	chr8: 11640598	C	T	Missense (R to C)	Pentatricopeptide repeat containing protein (Os08t0290000)
	chr8: 12545030	T	A	Missense (I to N)	Peroxidase 40 precursor-like protein (Os08t0302000)

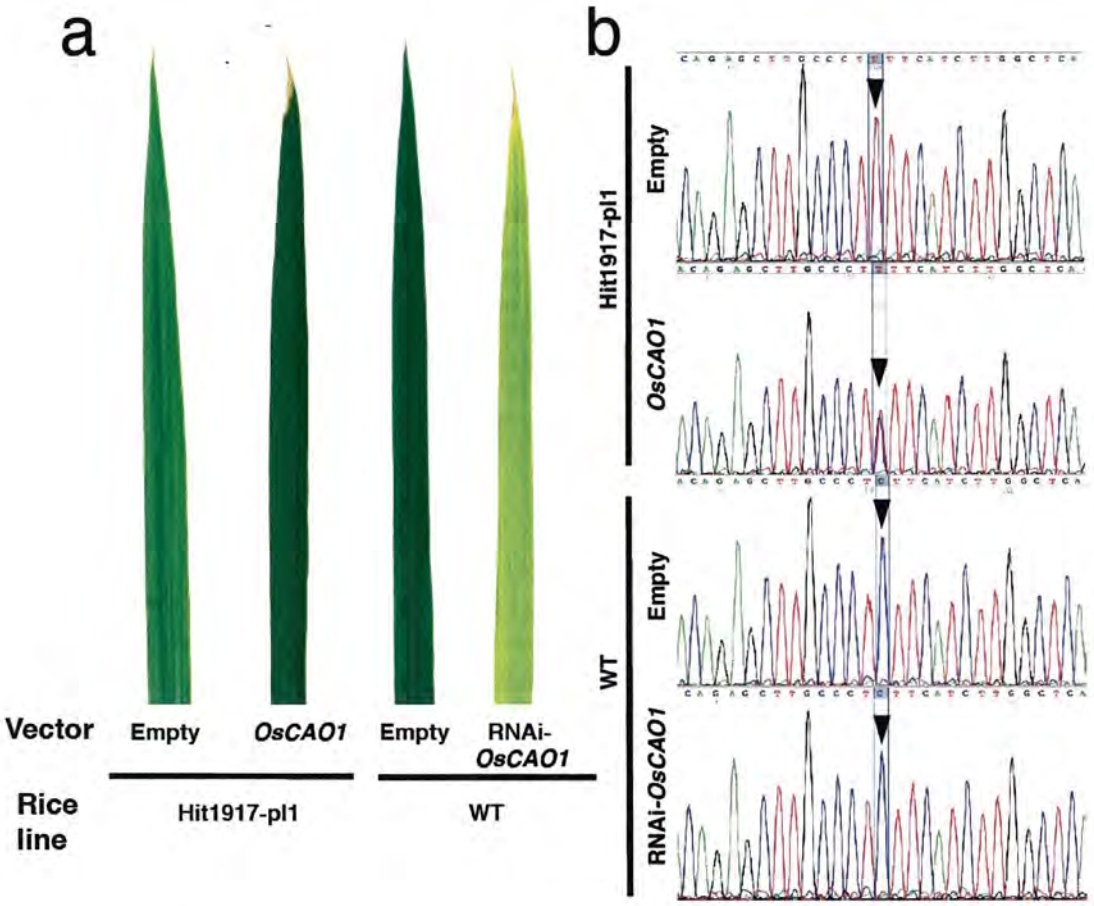

Figure 3-4. Genetic complementation of the Hit1917-pl1 pale green leaf mutant with *OsCAO*.

a. Rice lines and plasmid vectors used for transformation studies (bottom), phenotypes of four transgenic rice lines (top). b. DNA sequencing peak chromatograms of *OsCAO1* cDNA close to the site SNP-22981826 (arrow) obtained from the four individuals shown in a.

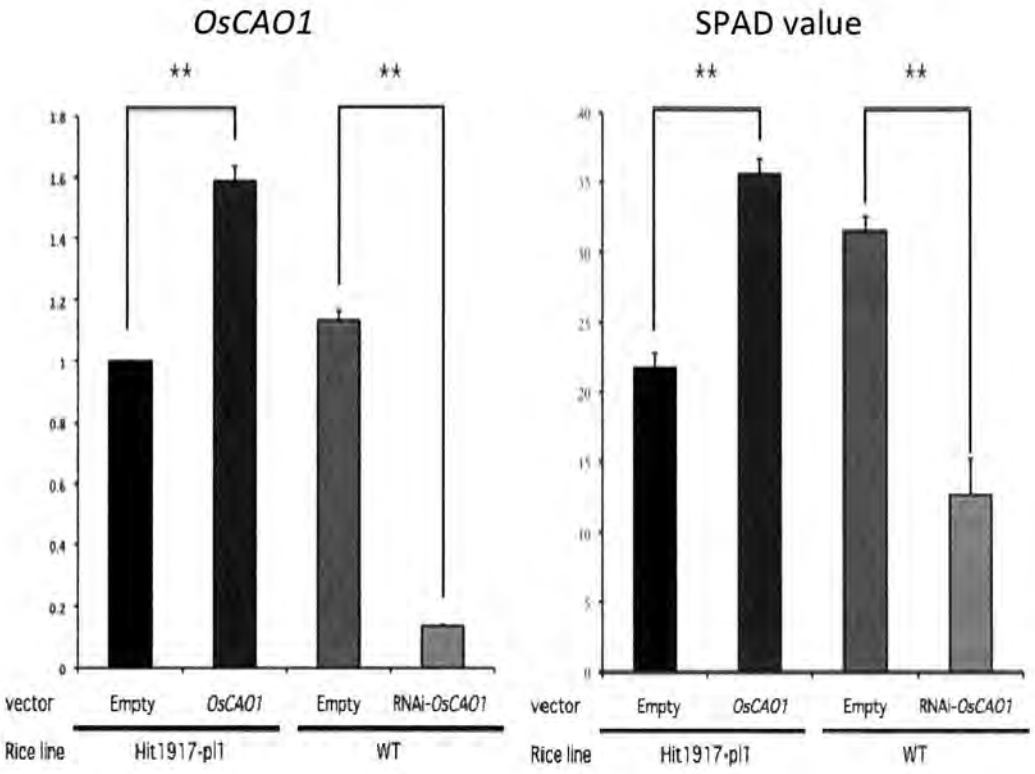

Figure 3-5. Real-time quantitative RT-PCR results of OsCAO1 gene (left) and SPAD values in leaves (right) of four transgenic rice lines.

Hit1917-pl1+Empty, Hit1917-pl1+OsCAO1, WT+Empty and WT+OsCAO1-RNAi.

Application of MutMap to agronomically-important traits

Our rice mutant lines exhibited a wide variation in quantitative traits (Table 3-6). Taking a subset of 1,634 M4 lines, we measured seven traits of agronomic importance (culm length, leaf length, number of panicles per plant, panicle length, husk length, husk width, spikelet numbers per panicle). Out of the seven traits, four (culm length, panicle length, husk length and husk width) showed statistically significant greater variance (F-test, $P < 0.05$) between the mutant lines compared to the wild-type parental "Hitomebore" line, demonstrating that our mutant lines contain a larger genetic variation and are a good resource to isolate genes controlling agronomic traits (Fig. 3-6a).

In our breeding program, we are particularly focusing on plant height (culm length), since a semi-dwarf phenotype increases yield (Sasaki *et al.* 2002). To this end, we applied MutMap to four semi-dwarf mutants (Hit1917-sd, Hit0746-sd, Hit5500-sd, Hit5814-sd) in order to identify the genomic regions responsible for this phenotype. Additionally, a male sterility mutant (Hit5243-sm) was included in the analysis (Fig. 3-6b). Note that phenotypes of pale green leaf in Hit1917 (indicated as Hit1917-pl1) and semi-dwarfism in the same line (Hit1917-sd) are caused by two independent mutations in unlinked loci. All these phenotypes, except that of Hit5243-sm, are subtle and quantitative and are difficult to score in the F_2 generation when the mutants are crossed to an unrelated cultivar. With our MutMap approach, the F_2 segregation was clearly observed (Fig. 3-7), and the segregation ratio suggested that a single recessive gene is involved in governing all the phenotypes (Table 3-2). Twenty F_2 progeny

Table 3-6. Measurements of seven quantitative traits of EMS-treated rice mutant lines derived from the Hitomebore cultivar.

Character	Mutant lines			Wild type			F-test					
	n	Mean	SD	n	Mean	SD	F value	P value	F (0.025)	F (0.975)	F (0.005)	F (0.995)
Culm length (cm)	1634	74.6	5.5	29	77.8	3.1	0.314	0.0004			0.443	1.835
Leaf length (cm)	1634	35.0	2.5	29	34.1	2.3	0.796	0.4672	0.545	1.597		
Number of panicles per plant	1634	11.4	1.9	29	12.0	2.0	1.112	0.6269	0.545	1.597		
Panicle length (cm)	1646	19.3	0.9	27	19.7	0.5	0.366	0.0028			0.427	1.872
Grain length without glumes (mm)	1548	5.1	0.1	27	5.1	0.0	0.073	0.0000			0.427	1.872
Grain width without glumes (mm)	1548	2.8	0.1	27	2.9	0.0	0.243	0.0001			0.428	1.871
Number of spikelets per panicle	1646	113.6	11.9	27	117.8	9.8	0.670	0.2106	0.531	1.621		

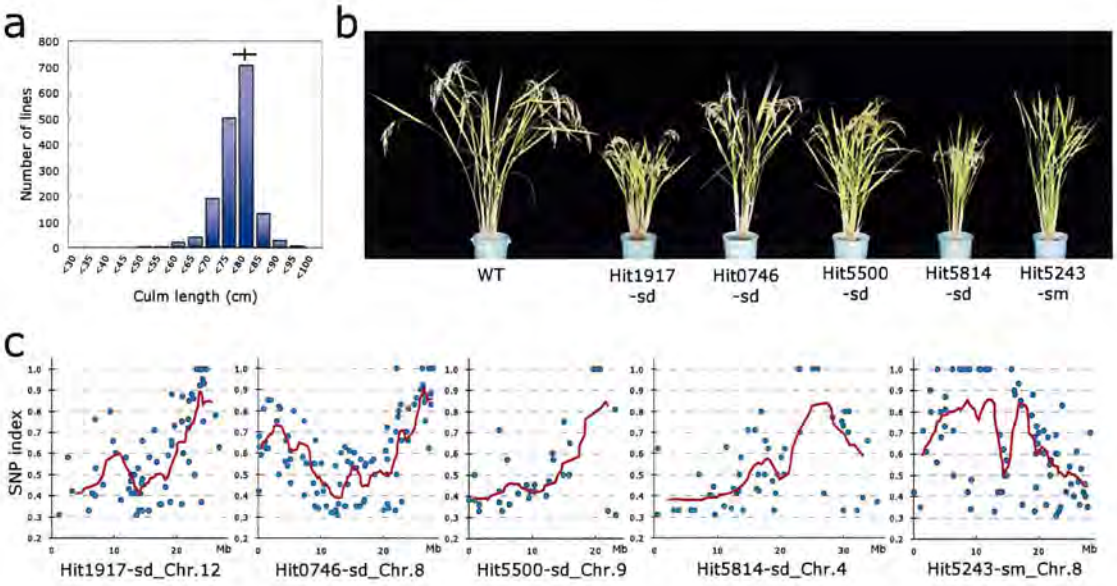

Figure 3-6. Identification of the genomic regions possibly harboring causal mutations for five agronomically useful rice mutants using MutMap.

a. Distribution of culm length of 1,634 Hitomebore mutant lines. Vertical and horizontal bars indicate mean and standard deviation, respectively, of the wild type Hitomebore plants (N = 29). b. Gross morphology of wild type Hitomebore rice (WT) and five mutants (Hit1917-sd, Hit0746-sd, Hit5500-sd, Hit5814-sd and Hit5243-sm). c. SNP index plots for the five mutants. Sometimes multiple peaks (i.e. Hit0746-sd) or a broader peak (i.e. Hit5243-sm) of plots were observed. The former may be caused by an additional mutation affecting viability of F2 and the latter may be attributed to mutation position in recombination-deficient chromosome regions i.e. close to the centromere, but these cases in most cases did not pose a crucial difficulty in identifying the putative regions harboring the causal mutations.

309 : 96 (= 3 : 1)

Figure 3-7. Segregation of wild type (left) and mutant phenotypes (right) in F2 progeny derived from a cross between Hitomebore wild-type and a mutant Hit1917-sd.

showing mutant phenotypes were bulk-sequenced, and the SNP index scored (Table 3-3, 3-4). The resulting SNP-index plots are given in Figure 3-6c and Figure 3-3. In all cases, only a single genomic region contained a cluster of SNPs with SNP index = 1. The average interval of the SNPs within these regions with a SNP index $\geqq 0.9$ was 2.1 Mb and we found at most four SNPs that could have caused non-synonymous changes of protein-coding genes (Table 3-4). We therefore conclude that these regions correspond to locations of the causal mutations responsible for the observed phenotypes (Table 3-5).

3-4 Discussion

Here we have shown that MutMap applied to rice can rapidly identify the genomic region harboring the causal mutation for a given phenotype. The most important feature of MutMap is that it only requires crossing the mutant to the wild-type line used for the mutagenesis followed by one subsequent selfing. As the mutant has been crossed back to its progenitor wild-type, the F_2 progeny will show unequivocal segregation between the mutant and wild-type phenotypes (Fig. 3-7). This contrasts with conventional crossing schemes for gene isolation that involve crosses between genetically distantly related lines. In distant crosses, the parent lines differ in many genes so that segregation of particular phenotypes in the F_2 generation follows a Gaussian distribution, and not a discrete distribution (Falconer and Mackay 1996; Fig. 3-8). In addition, half of all the loci in the F_2 are heterozygous and, therefore, heterosis strongly affects the phenotype. Because of these problems isolation of genes with minor effects as well as QTL has been

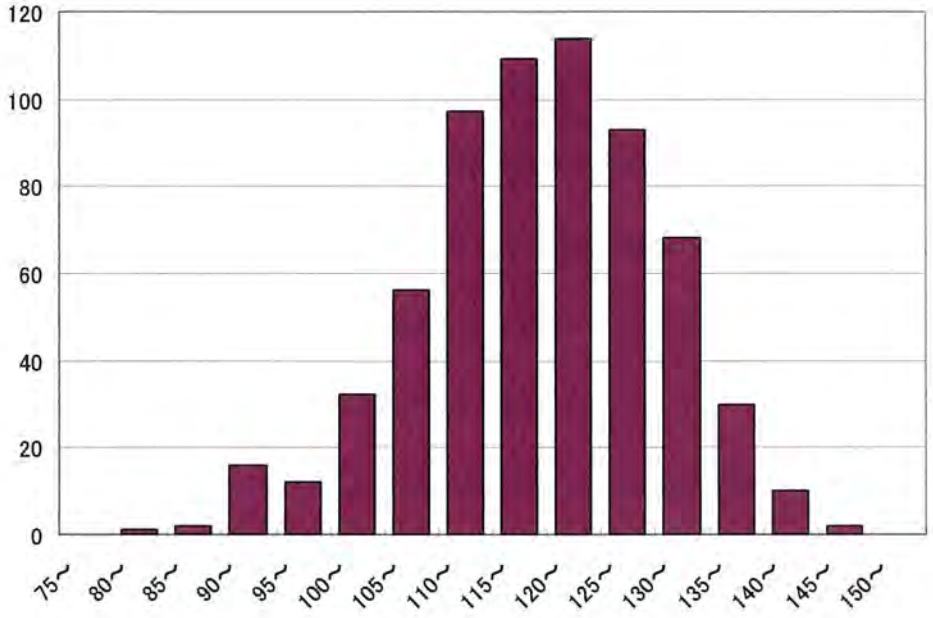

Figure 3-8. Frequency distribution of plant height (x-axis, culm length; cm) of F2 progeny (642 individuals) derived from a cross between Hitomebore wild-type and Kasalath (ssp. *indica*) wild-type.

This cross is commonly used for mapping mutant loci in rice.

carried out by using Recombinant Inbred Lines (RILs) of advanced generations, in which the contribution of an individual gene is separately addressed and the effect of heterosis can be minimized. However, generation of RILs requires time and effort, and even advanced generations of RILs can be difficult to phenotype.

MutMap is technically similar to SHOREmap (Schneeberger *et al.* 2009) and other related methods (Austin *et al.* 2011; Uchida *et al.* 2011) including NGM (Austin *et al.* 2011), based on bulked-segregant analysis of F_2 progeny (Michelmore *et al.* 1991; Giovannoni *et al.* 1991). Typically, a cross is made between a mutant of one ecotype (e.g. Arabidopsis Col) and a wild-type individual of a distantly related ecotoype (e.g. Arabidopsis Ler), and F_2 plants derived from such crosses are bulked and sequenced. The sequencing reads are aligned to the genome sequence of the wild-type parental line (Col) to search for the region with a high frequency of the Col-type SNPs. That is, the SNPs that vary between the distantly related ecotypes (~1 SNP/kb; Schneeberger *et al.* 2009) are used as DNA markers to locate the region harboring the causal mutation. By contrast, MutMap employs SNPs incorporated by mutagenesis as markers to look for the region harboring the mutation responsible for a given phenotype. The small number of SNPs that need to be considered in MutMap (< 2,300) as compared to the SHOREmap/NGM scheme (i.e. 305,002 SNPs between Col and Ler, ftp://ftp.arabidopsis.org/Polymorphisms/Ecker_ler.homozygous/snp.txt) enables reliable alignment between genome sequences and reduces the level of noise in SNP calling. An advantage of having a large number of SNPs, as in SHOREmap/NGM, is that a clear cluster of SNPs with a high SNP index would be expected around the causal SNP.

However, it would be difficult to pinpoint the actual causal SNP because of the large number of SNPs in the cluster. In MutMap, the causal SNP can readily be identified, provided that the region has sufficient sequence coverage (Supplementary Data). In SHOREmap and NGM, only phenotypes with discrete characteristics can be unequivocally scored among the F_2 plants derived from distant crosses, and subtle alterations of quantitative traits cannot be adequately measured. The small number of F_2 progeny used [20] for bulking in MutMap also contrasts to the larger number of progeny required for the application of SHOREmap and NGM to Arabidopsis, namely 500 and >50, respectively (Schneeberger *et al.* 2009; Austin *et al.* 2011). In crop plants, growing such a large number of F_2 progeny in the field can be impractical. The major differences between MutMap and SHOREmap/NGM are summarized in Table 3-7.

MutMap also differs from the method of bulk sequencing of progeny derived from sequential backcrosses (> 4 times) of mutants to the wild-type, that has been applied to Caenorhabditis elegans (Zuryn *et al.* 2010) and Arabidopsis (Ashelford *et al.* 2011). In the latter method, backcrosses remove SNPs from the progeny, and only the regions harboring the causal mutation retain SNPs. The regions harboring the causal mutation for the phenotype are identified by scanning the genome for a region with high SNP density. MutMap is more practical for crops with long generation times since consecutive backcrosses are not required.

It is possible to further simplify MutMap for traits that are easily quantified in the field. Mutagenesis of seeds of an inbred line results in M1 lines. Self-pollination of each M1 plant results in an F_2 generation that segregates for mutant and wild-type

Table 3-7. Comparison of MutMap and SHOREmap/NGM.

Method	Applicable to isolation of genes with small phenotypic effects?	Crossing scheme	Example	DNA markers used for mapping	Number of total SNPs used as marker	Marker interval	Number of F2 progeny used for bulking	Note
MutMap	Yes Suitable for crop species	Mutant x WT cross + selfing → F2	Rice (380 Mb) Hitomebore mutant x Hitomebore WT	SNPs incorporated by mutagenesis	$< 2 \times 10^3$	> 200 Kb	20	This study Use "SNP index"
SHORE map	No	Distant cross + selfing → F2	Arabidopsis (125 Mb) Col mutant x *Ler* WT	SNPs present between ecotypes	3×10^5	~ 1 Kb	500	Schneeberger *et al.* (2009)
NGM	No	Distant cross + selfing → F2	Arabidopsis (125 Mb) Col mutant x *Ler* WT	SNPs present between ecotypes	3×10^5	~ 1 Kb	> 50 (recommended)	Austin *et al.* (2011) Use "Chastity Index"

phenotypes. Provided that the M1 line contains a recessive mutation (designated as a) in the heterozygous state (Aa), there will be a high chance ($1-[0.75]^{20} = 0.997$) of observing the mutant phenotype (aa) among 20 F_2 individuals. Thus, by growing and screening > 20 F_2 individuals per line, it would be possible to obtain the individuals showing the desired mutant phenotypes (aa) as well as their heterozygous siblings (Aa) with the wild-type phenotype. Bulked DNA of mutant-type progeny derived from selfing of a large number of the heterozygous siblings could serve the same function as that of mutant-type F_2 plants derived from the cross between a mutant and its wild-type in the MutMap scheme. The advantage of this method would be that it does not require any crossing, which makes it applicable to crops that are difficult to artificially cross.

In crop plants, identification of QTL has been conventionally addressed by mining natural genetic variation. However, Robertson (Robertson *et al.* 1985) proposed the use of mutants for isolating QTL. We envisage that the application of MutMap to a large mutant collection may yield a treasure trove of information relevant to QTL isolation. One could, for example, first identify a QTL likelihood interval by conventional QTL mapping, then search the MutMap database for mutants that have a related phenotype and that map within the QTL likelihood interval to identify candidate DNA changes worthy of further testing.

We anticipate that MutMap will greatly facilitate gene isolation and breeding of crops by reducing the time and labor required for identifying agronomically-important genes. Given that DNA sequencing is becoming easier and cheaper, the cost of identifying such genes can be dramatically reduced. However, it is not even necessary

to identify the causal mutation to exploit MutMap in crop breeding. If the causal SNP cannot be identified, the SNPs flanking the regions (those with a SNP index = 1) can be used as DNA markers to immediately start marker assisted selection (MAS) by crossing the mutant to the wild type. Moreover if the mutagenesis is performed in an elite crop cultivar, as in the case of cv. Hitomebore, then the mutants and the associated SNP markers can be immediately made available to breeders to generate new improved varieties. Since MutMap requires relatively deep sequencing coverage of the genome (>10X), its application to crops with very large genome sizes, such as maize, sorghum, soybean, barley, wheat etc., requires further investigation.

3-5 Supplementary Data (This analysis was conducted by Dr. Hideki Innan.)

Evaluation of the effects of number of F2 progeny to be bulked (*n*) and average sequencing coverage (depth) of the genome (*G*) on the levels of false positive SNPs (SNP index = 1).

In our demonstration with rice in the main text, we observed a clear cluster of SNPs with SNP index=1. However, it could be possible to encounter cases where the signature is not clear. Using statistics we would be able to evaluate the probability of occurrence of non-causal SNPs with SNP index =1 by chance. Suppose that n is the number of individuals used for bulk sequencing and G is the genome-wide average of the sequencing coverage. At an unlinked SNP, the number of reads from a single individual follows a Poisson distribution with mean G/n assuming equal density for all individuals in the bulked DNA. Therefore, although the expected SNP index is 0.5 regardless of n, the variance decreases with increasing n. This null distribution of the SNP index can be easily obtained by a simple simulation, while it can also be given by a more complicated analytical expression with a convolution of multiple Poisson distributions. The P-value for a single SNP can be determined from this null distribution. Suppose that there are L SNPs in the entire genome. Then, the statistical cutoff for these multiple SNPs can be determined by correcting for multiple tests, e.g., by the Bonferroni or Sidák corrections, which is denoted by q.

The statistical power largely depends on q; as q decreases the power would increase. Therefore, it is desirable for the null distribution to have a narrow distribution around 0.5. To decrease the variance of the null distribution, we need to increase the sample size and coverage. To demonstrate their effects separately, the effect of n when G is fixed is shown in Figure S1a and Figure S1b represents the effect of G when n is fixed. We found that increasing n significantly decreases the variance of the SNP

index (Figure S1a). The effect of increasing G seems to be relatively minor although a slight decrease of the variance is observed (Figure S1b). These results indicate the importance of increasing n to have a narrow null distribution of SNP index.

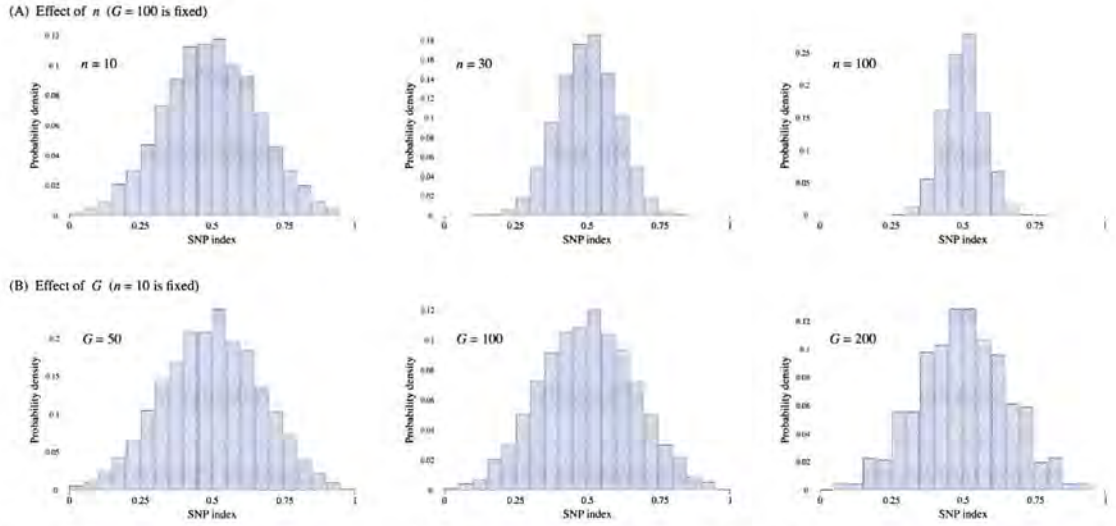

Figure S1 Density distributions of SNP index at an unlinked SNP. (A) The effect of n when $G = 100$. (B) The effect of G when $n = 10$.

Evaluation of the effects of misclassification of the phenotypes of F2 progeny (between mutant- and wild-type) on the SNP-index

Next, consider a causal SNP, which should have SNP index = 1 if the phenotype is correctly determined. Accordingly, the causal SNP can be statistically identified if the statistical cutoff after correcting for multiple tests (q) is smaller than 1. However, this does not hold if there is misclassification of the phenotype. Ideally, all n individuals should have genotype aa at the causal SNP, but suppose j individuals happened to be AA because their phenotype had been wrongly classified. In this case, the expected SNP index at the causal SNP is decreased to j/n. The density distribution of the SNP index is again given by a complicated function of n, j and G, involving multiple Poisson distributions. Figure S2 shows how misclassification affects the distribution. Obviously, as j increases, the distribution of the SNP index shifts left (Figure S2a), indicating the importance of correct classification of the phenotype. Next, if a certain level of misclassification has to be taken into account, we ask how the statistical power can be improved. There are at least two options: (i) to increase the coverage, G, and (ii) to increase the sample size, n. To address this question, in Figures S2b and c, the misclassification rate (j/n) is fixed, and the effects of G and n are explored. It was found that as the coverage (G) increases the variance dramatically decreases (Figure S2b). In contrast, the effect of n appears to be far less significant although we can observe a slight improvement because the probability of SNP index <0.9 decreases with increasing n (Figure S2c). It is suggested that increasing the coverage may be more efficient for increasing the statistical power when there is a possibility of misclassification of the phenotype.

Overall, the success rate for identifying the causal SNP is determined by q, the cutoff value after

correcting for multiple tests, and the expected distribution of the SNP index at the causal SNP. q has a strong positive correlation with the number of SNPs (L), therefore, having too many SNPs might decrease the statistical power. An advantage of having a large number of SNPs is that we can expect a clear array of SNPs with a very high SNP index around the causal SNP, although it would be difficult to pinpoint the real causal SNP from a large number of SNPs in the array. (An extreme case would be NGM). In our demonstration with rice in the main text, we show successful results with approximately $L = 2{,}000$ SNPs. We believe that our method works with an even smaller value of L because the rate of false positives is reduced. Perhaps, the ideal scenario for our method would be just a single SNP that was also the causal SNP. The only concern in a case with a small value of L is that if the causal SNP might not be well sequenced for some reason. In such a situation, additional linked SNPs would help to identify the causal 'region'.

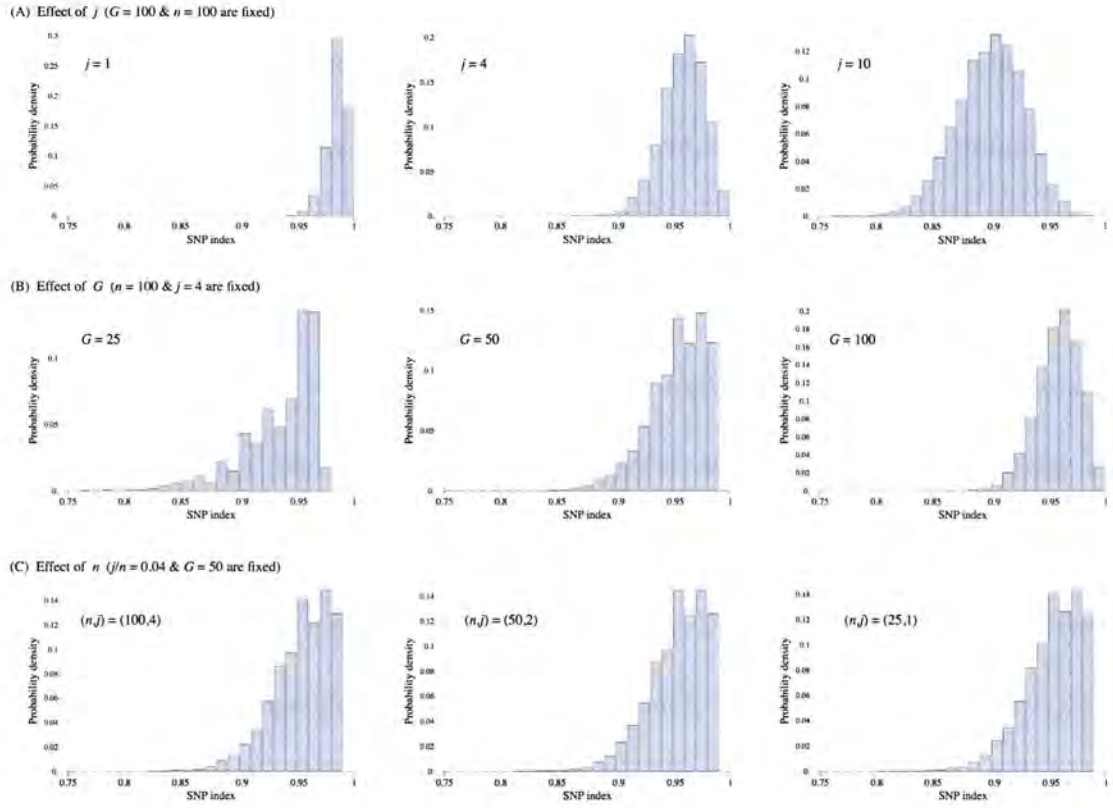

Figure S2 Density distributions of SNP index at the causal SNP. (A) The effect of j when $G = 100$ and $n = 100$. (B) The effect of G when $n = 100$ and $j = 4$. (C) The effect of n when $j/n = 0.04$ and $G = 50$.

Chapter 4

General discussions

Food shortage has become a challenging problem worldwide, not only because of world population growth but also as a result of the adverse effects of global climate change on food production. To address this problem, crop improvement programs that deal with the traits related to tolerance to environment stress, high yield, and lower labor costs are required. To this end, the genes that affect these traits need to be identified and characterized for a better understanding of the traits. Moreover, genetic resources carrying the alleles that enhance agronomic traits must be identified (Takeda and Matsuoka 2008).

The direct seeding of rice (*Oryza sativa* L.) is preferred to transplanting because of its low labor demand, and has been widely practiced throughout the tropics and subtropics. This method has gained popularity only recently in temperate regions like Japan, where low temperatures limit early growth and good stand establishment. Seedling vigor is one of the major determinants of stable stand establishment during direct seeding of rice in temperate regions. In this study, we identified QTLs for enhanced height of seedlings using a cultivar "Dunghan Shali" that shows superior seedling vigor. We showed that the causal gene of a major QTL '*qPHS3-2*' is *OsGA20ox1*, which is an enzyme that normally catalyzes the penultimate steps in GA biosynthesis. Then, to demonstrate the effect of the identified QTL '*qPHS3-2*' in the paddy field condition, we performed field evaluation test using a sowing machine that is commonly used by farmers. Compared with Iwatekko, we observed significantly higher plant height in NIL-114, although leaf age of the two lines was same (Fig. 4-1a,b). In NIL-114, the number of seedlings per square meter was also significantly higher than in

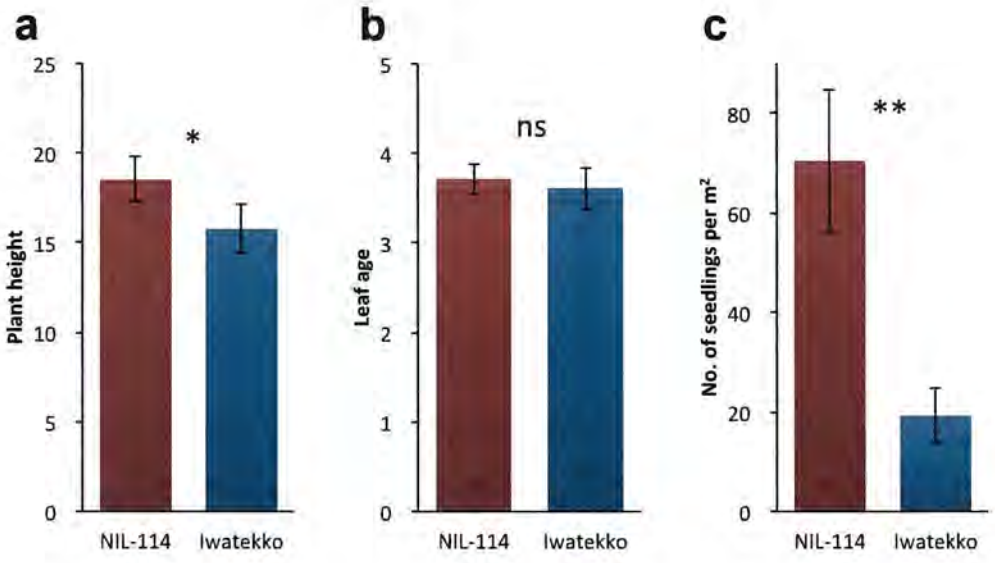

Figure 4-1 Characterizations of seedling of NIL-114 and Iwatekko.

(a) Plant height (cm, n=17 plants). **(b)** Leaf age (n=17 plants). (c) Number of seedlings per square meter (n=3 replications). The seeds used in this experiment were imbibed in water for 3 days at 12-15°C and then forced into sprouting for 12 hours at 28°C. Then, approximately 4kg seeds per 10a were sown by a sowing machine, which is commonly used by farmers, on 11 May 2012. Plant height (length between the coleoptile node to the tip of longest leaf blade), leaf age and number of seedlings were measured at 32 days after sowing. Bars indicate the standard error. Asterisks indicate significant difference (Student's t-test, $*P < 0.05$; $**P < 0.01$).

Iwatekko (Fig. 4-1c, Fig. 4-2). In addition, the seedling establishment rate in NIL-114, which was evaluated by a field test using seeder tape, was 1.7 times higher than that in Iwatekko (Fig. 4-3). These results suggest that *qPHS3-2* is useful for improvement of seedling establishment in direct seeding method. Therefore, we believe that the identified Dunghan Shali allele of *OsGA20ox1* as well as the molecular markers and the NIL developed in this study are invaluable resources for rice breeding programs aimed at improving seedling vigor in temperate regions worldwide.

Mutant lines are also good resources for crop improvement. They are also important materials to understand the molecular mechanism governing important traits in crops. We have generated large-scale mutant line populations of an elite Japanese rice cultivar 'Hitomebore' for the purpose of utilizing them for gene isolation and breeding. However, when a mutant shows only a subtle phenotypic change compared with the wild-type, as is the case with most traits of agronomic importance, identification of the causal mutation using conventional 'map-based cloning' involving distant crosses is difficult. As explained in Chapter 3, the newly developed method, MutMap, circumvents all the problems inherent in distant crosses for rapid identification of the genomic region harboring the causal mutation of a given phenotype even when the change is subtle.

Here I show a result of MutMap application for the isolation of rice gene related to amylose content of grains. In rice breeding programs in Japan, improving eating quality of cooked rice has been one of the most important objectives. Since amylose content of rice grains is negatively correlated with stickiness and score of eating quality

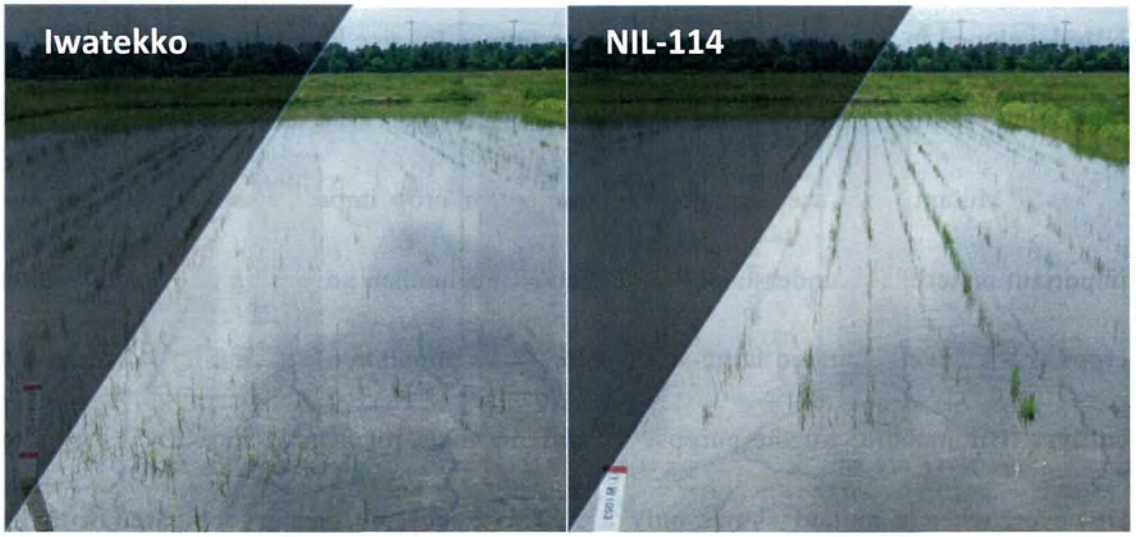

Figure 4-2 Comparison of seedling establishment between Iwatekko (left) and NIL-114 (right).

In NIL-114, seedling rows were clearly visible at 28 days after sowing.

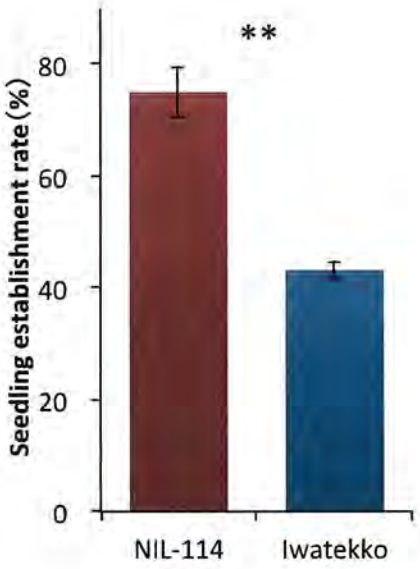

Figure 4-3 Seedling establishment rates in NIL-114 and Iwatekko.

The seeds used in this experiment were wrapped by seeder tape at 47 seeds per 1m. Then, seeds were imbibed in water for 3 days at 12-15°C and then forced into sprouting for 12 hours at 28°C. Then, seeds were sown in paddy field by hand on 14 May 2012. Seedling establishment rate was calculated using the number of seeds and the number of seedlings at 23 days after sowing. Bars indicate the standard error. Asterisks indicate significant difference (Student's t-test, $^{**}P < 0.01$).

in rice varieties (Juliano *et al.* 1965, Juliano 1985, Inatsu 1988), rice breeders in Japan have focused on breeding varieties with a low amylose content of under 15% in the last decade. This resulted in the development of rice varieties including Aya, Milky Queen and Snow Pearl (Kunihiro *et al.* 1993, Ise *et al.* 2001, Higashi *et al.* 1998). However, these varieties exhibit amylose content lower than 12% when cultivated in Tohoku region in Japan, thus are not suitable for daily staple, because of too much stickiness caused by low amylose content. Amylose content of the commonly cultivated varieties in Tohoku region, such as Hitomebore, is usually 17-20%. On the other hand, amylose content of rice grains that is produced in Hokuriku region such as Niigata prefecture is 15-17%. This is because grain amylose content is inversely correlated with the temperature during ripening stage of rice (Asaoka *et al.* 1984, Sano *et al.* 1985, Inatsu 1988). It is thought that difference in growth temperature affects grain amylose content and eating quality so that the rice produced in Hokuriku region is preferred to that in Tohoku region by the consumers.

To improve rice eating quality of a Tohoku variety, we performed screening for amylose content of 2,000 EMS lines to identify mutant line that shows amylose content of 15-17% that is expected to cause a superior eating quality, even in the growth condition of Tohoku region including Iwate prefecture where summer temperature is cool. We identified a mutant line Hit1073 with the 15.7% amylose content that is only 2.5% lower than Hitomebore (18.3%) (Fig. 4-4a). Then, to apply MutMap, we crossed Hit1073 back to Hitomebore WT, and obtained F2 progeny by selfing of the F1 plants. We observed a clear segregation in amylose content for field-grown 76 F2 progeny even though the

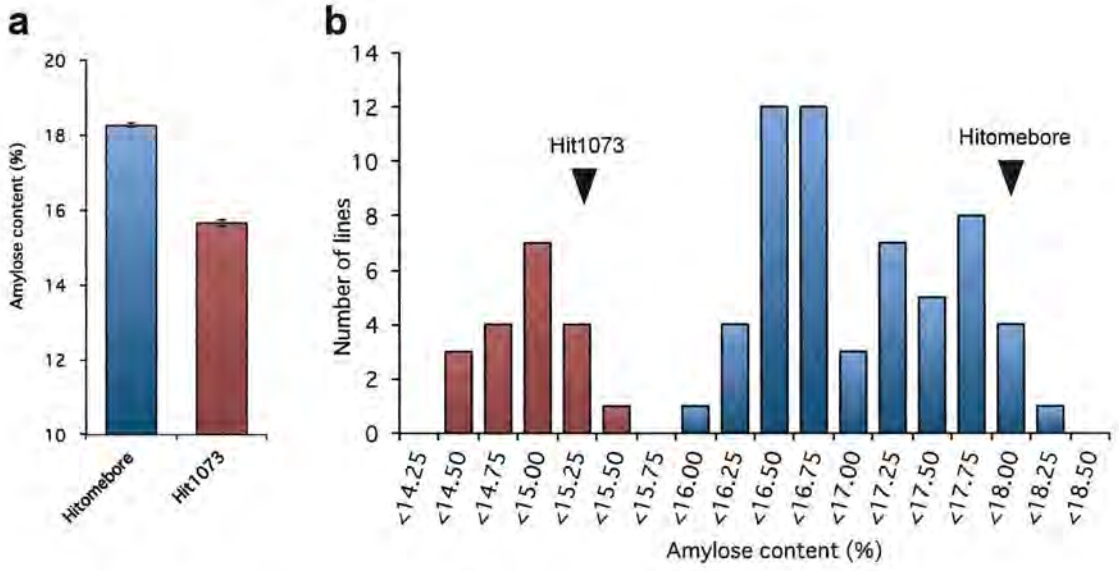

Figure 4-4 Amylose content in Hit1073, Hitomebore and their F_2 progeny.

(a) Comparison of amylose content in Hit1073 and Hitomebore. Bars indicate the standard error. **(b)** Frequency distribution of amylose content in 76 F_2 progeny derived from a cross between Hit1073 and Hitomebore. Red bars indicate the total 19 F_2 progeny that showed the mutant phenotype, and were used for bulk sequencing. Arrowheads indicate the mean values for Hit1073 and Hitomebore. To analyze amylose content, 9g polished grains were crushed with a mill. Resulting 50mg rice flour was diluted with 1ml 0.1N NaOH, and was shaken overnight at room temperature. After dilution with 3ml H_2O, amylose content was determined with an Auto Analyzer III (BL-TEK K.K., Japan) with four times measurement per plant.

mutant phenotype was subtle (Fig. 4-4b). We performed bulk sequencing and MutMap application using the 19 mutant-type progeny that showed lower amylose content (Fig. 4-4b). A single unique genomic region harboring a cluster of SNPs exhibiting SNP index = 1 was identified on chromosome 9 (Fig. 4-5). We therefore conclude that this region corresponds to location of the causal mutation responsible for the observed phenotype; 2.5% lower than amylose content as compared to Hitomebore WT. This result demonstrates that MutMap allows rapid identification of the position of a causal mutation responsible for a mutant phenotype even though the phenotype is subtle. Furthermore, we already started breeding of Tohoku variety using Hit1073 mutant by DNA marker assisted selection to improve the eating quality.

I believe that MutMap method as described here is a very powerful tool for accelerating understanding of plant molecular mechanisms and also for enormously enhancing improvement of crop traits including quantitative traits of agronomic importance.

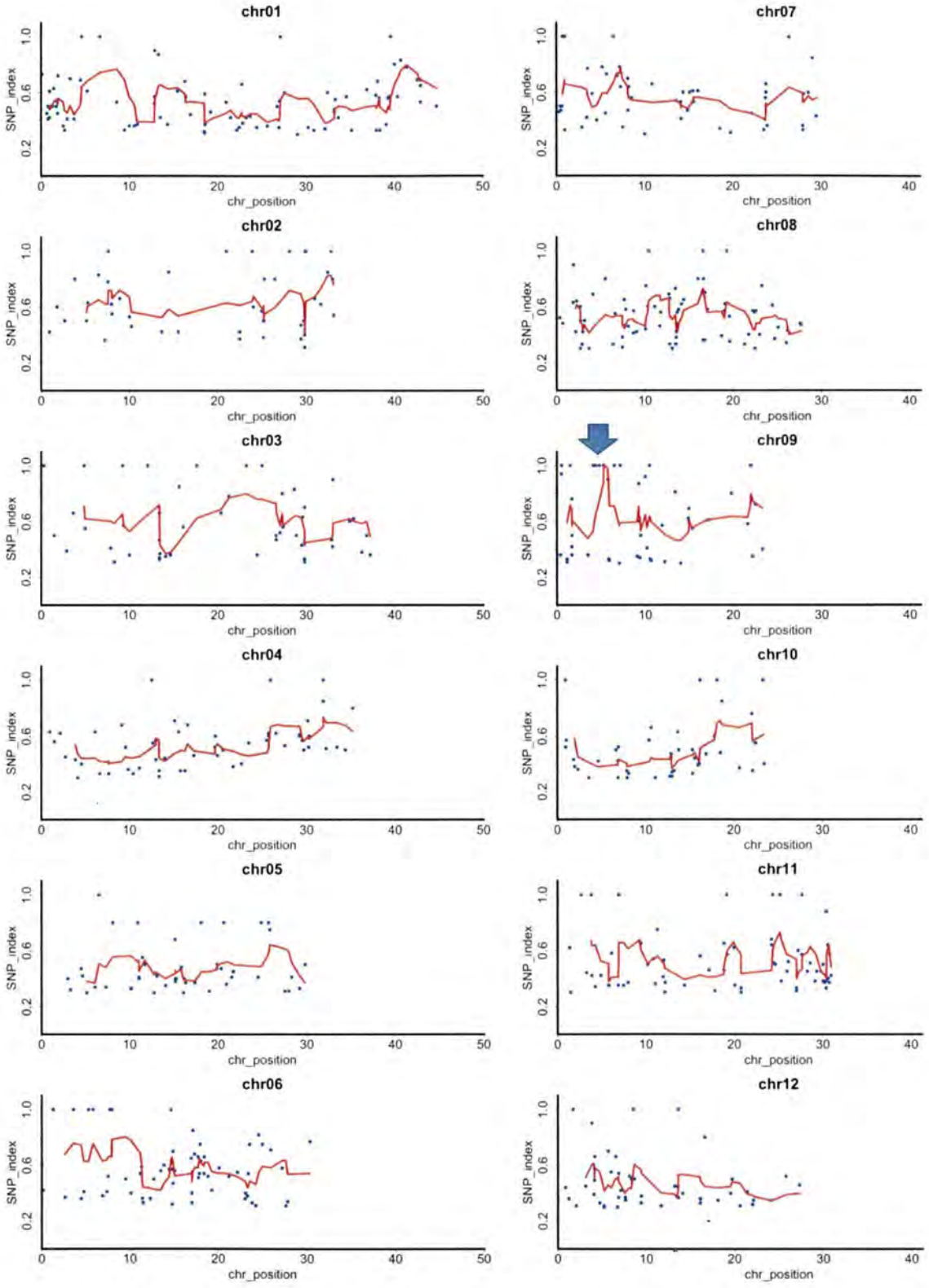

Figure 4-5 SNP index plots for 12 chromosomes of Hit1073.

An arrow indicates the candidate region that likely harbors the causal mutation of Hit1073.

References

Asaoka M, Okuno K, Sugimoto Y, Kawakami J, Fuwa H. (1984) Effect of environmental temperature during development of rice plants on some properties of endosperm starch. Starch 36: 189-193.

Ashelford K, Eriksson ME, Allen CM, D'Amore L, Johansson M, Gould P, Kay S, Millar AJ, Hall N, Hall A (2011) Full genome re-sequencing reveals a novel circadian clock mutation in Arabidopsis. Genome Biol. 12:R28.

Ashikari M, Matsuoka M (2006) Identification, isolation and pyramiding of quantitative trait loci for rice breeding. Trends in Plant Science 11: 344-350.

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

Ashikari M, Sasaki A, Ueguchi-Tanaka M, Itoh H, Nishimura A, Datta S, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M (2002) Loss-of-function of a rice gibberellin biosynthetic gene, *GA20 oxidase* (*GA20ox-2*), led to the rice 'Green Revolution'. Breeding Science 52: 143-150

Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW, McCourt P, Guttman DS (2011) Next-generation mapping of Arabidopsis genes. Plant J. 67: 715-725.

Basten CJ, Weir BS, Zeng ZB (2005) QTL cartographer, version 1.17. Department of Statistics, North Carolina State University, Raleigh, p 189.

Baulcombe D. (2010) Reaping benefits of crop research. Science 327: 761.

Cui KH, Peng SB, Xing YZ, Xu CG, Yu SB, Zhang Q (2002) Molecular dissection of seedling-vigor and associated physiological traits in rice. Theor Appl Genet 105: 745-753.

Falconer DS and Mackay TFC (1996) Introduction to Quantitative Genetics. Pearson/Prentice Hall.

Fujino K, Sekiguchi H, Matsuda Y, Sugimoto K, Ono K, Yano M. (2008) Molecular identification of a major quantitative trait locus, qLTG3-1, controlling low-temperature germinability in rice. PNAS 105(34): 12623-12628.

Fukuoka S, Saka N, Koga H, Ono K, Shimizu T, Ebana K, Hayashi N, Takahashi A, Hirochika H, Okuno K, Yano M. (2009) Loss of function of a proline-containing protein confers durable disease resistance in rice. Science 325: 998-1001.

Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD (1991) Isolation of molecular markers from specific chromosomal interavals using DNA pools from existing mapping populations. Nucl. Acids. Res. 19: 6553-6558.

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C (2010) Food security: The challenge of feeding 9 billion people. Science 327: 812-818.

Higashi T, Saitoh S, Takita T, Yamaguchi M, Sunohara Y, Yokogami N, Ikeda R, Tamura Y, Oyamada Z, Kowata H,Inoue M, Matsumoto S, Kataoka T. (1998) Development of a new rice variety with low amylose contents "Snow Pearl". Tohoku Agric. Res. 51: 1-2 (in Japanese).

Hori K, Sugimoto K, Nonoue Y, Ono N, Matsubara K, Yamanouchi U, Abe A, Takeuchi Y, Yano M. (2011) Detection of quantitative trait loci controlling pre-harvest sprouting resistance by using backcrossed populations of japonica rice cultivars. Theor Appl Genet 120: 1547-1557.

Inatsu O. (1988) Studies on improving the eating quality of Hokkaido rice. Rep. of Hokkaido Pref. Agric. Exp. Stn. 66 (ISSN 0367-6048) (in Japanese with English summary).

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

Ise K, Akama Y, Horisue N, Nakane A, Yokoo M, Ando I, Hata T, Suto M, Numaguchi K, Nemoto H, Furutachi H, Imbe T. (2001) "Milky Queen", A New High-Quality Rice Cultivar with Low Amylose Content in Endosperm. Bull. Natl. Inst. Crop Sci. 2: 39-61 (in Japanese with English summary).

Juliano, BO. (1985) Criteria and test for rice grain quality. In: Juliano, B.O. (ed.) Rice chemistry and technology. Pp 443-513. American Association of Cereal Chemists, Saint Paul.

Juliano, BO, Onate LU, Mundo AM. (1965) Relation of starch composition, protein content, and gelatinization temperature to cooking and eating qualities of milled rice. Food Technology 19: 1006-1011.

Kaneko M, Itoh H, Inukai Y, Sakamoto T, Ueguchi-Tanaka M, Ashikari M, Matsuoka M (2003) Where do gibberellin biosynthesis and gibberellin signaling occur in rice plant?. The Plant Journal 35: 104-115.

Kobayashi M, MacMillan J, Phinney B, Gaskin P, Spray CR, Hedden P (2000) Gibberellin biosynthesis: metabolic evidence for three steps in the early 13-hydroxylation pathway of rice. Phytochemistry 55: 317-321.

Kojima M. Kamada-Nobusada T. Komatsu H. Takei K. Kuroha T. Mizutani M. Ashikari M. Ueguchi-Tanaka M. Matsuoka M. Suzuki K. Sakakibara H. (2009) Highly sensitive and high-throughput analysis of plant hormones using MS-probe modification and liquid chromatography-tandem mass spectrometry: an application for hormone profiling in *Oryza sativa*. Plant and Cell Physiology 50: 1201-1214.

Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M. (2002) Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. Plant Cell Physiol. 43: 1096-1105.

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M. (2006) An SNP caused loss of seed shattering during rice domestication. Science 312: 1392-1396.

Kosambi DD (1943) The estimation of map distance from recombination values. Annals of Human Genetics 12: 172–175.

Kunihiro Y, Ebe Y, Shinbashi N, Kikuchi H, Tanno H, Sugawara K. (1993) A new paddy rice variety "Aya" with good eating quality due to low amylose content development by anther culture breeding. Japan J. Breed. 43: 155-163 (in Japanese with English summary).

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.

Lee S, Kim JH, Yoo ES, Lee CH, Hirochika H, An G (2005) Differential regulation of chlorophyll a oxygenase genes in rice. Plant Mol. Biol. 57: 805-818.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAM tools. Bioinformatics 25: 2078-2079.

Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 18 :1851-1858.

Matsubara K, Kono I, Hori K, Nonoue Y, Ono N, Shomura A, Mizubayashi T, Yamamoto S, Yamanouchi U, Shirasawa K, Nishio T, Yano M (2008) Novel QTLs for photoperiodic flowering revealed by using reciprocal backcross inbred lines from crosses between japonica rice cultivars. Theor Appl Genet 117: 935-945.

Matsubara K, Ebana K, Mizubayashi T, Itoh S, Ando T, Nonoue Y, Ono N, Shibaya T, Ogiso E, Hori K, Fukuoka S, Yano M (2011) Relationship between transmission ratio distortion and genetic divergence in intraspecific rice crosses. Mol Genet Genomics 286: 307-319.

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D,

Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Research 9: 199-207.

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. PNAS 21: 9828-9832.

Miki D, Itoh R, Shimamoto K (2005) RNA silencing of single and multiple members in a gene family of rice. Plant Physiol. 138: 1903-1913.

Miura K, Ashikari M, Matsuoka M (2011) The role of QTLs in the breeding of high-yielding rice. Trends in Plant Science 16:319-326

Miura K, Ikeda M, Matsubara A, Song XJ, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M. (2010) OsSPL14 promotes panicle branching and higher grain productivity in rice. Nat Genet. 42: 545-549.

Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J. Bios. Bioengineer. 104 :34-41.

Ohyanagi H, Tanaka T, Sasaki H, Shigemoto Y, Yamaguchi K, Habara T, Fujii Y, Antonio BA, Nagamura Y, Imanishi T, Ikeo K, Itoh T, Gojobori T, Sasaki T (2006) The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. Nucleic Acids Res. 34: D741-D744.

Oikawa T, Koshioka M, Kojima K, Yoshida H, Kawata M (2004) A role of *OsGA20ox1*,

encoding an isoform of gibberellin 20-oxidase, for regulation of plant stature in rice. Plant Molecular Biology 55: 687-700.

Olszewski N, Sun TP, Gubler F (2002) Gibberellin signaling: Biosynthesis, catabolism, and response pathways. The Plant Cell 14: S61-S80.

Onishi K, Horiuchi Y, Ishigho-Oka N, Takagi K, Ichikawa N, Maruoka M, Sano Y (2007) A QTL cluster for plant architecture and its ecological significance in Asian wild rice. Breeding Science 57: 7-16.

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335: 721-726.

Peterson ML, Jones DB, Rutger J N (1978) Cool temperature screening of rice lines for seedling vigor. Il Riso, Vol. 27:269-274

Pienkowska M, Glickman B.W, Ferreira A, Anderson M, Zielenska M (1993) Large-scale mutational analysis of EMS-induced mutation in the lacI gene of Escherichia coli. Mutat Res 288: 123-131.

Rakshit S, Kanzaki H, Matsumura H, Rakshit A, Fujibe T, Okuyama Y, Yoshida K, Muluneh T, Shenton M, Utsushi H, Mitsuoka C, Abe A, Kiuchi Y, Terauchi R (2010) Use of TILLING for reverse and forward genetics of rice. In: Meksem, K. and Kahl, G. (eds) The Handbook of Plant Mutation Screening: Mining of Natural and Induced Alleles. Pp 187-198. Wiley-VCH

Rao AN, Johnson DE, Sivaprasad B, Ladha JK, Mortimer AM (2007) Weed management in direct-seeded rice. Advances in Agronomy 93: 153-255.

Redona ED, Mackill DJ (1996a) Genetic variation for seedling-vigour traits in rice. Crop Science 36: 285-290.

Redona ED, Mackill DJ (1996b) Mapping quantitative trait loci for seedlimg-vigour in rice using RFLPs. Theor Appl Genet 92:395-402

Robertson DS (1985) A possible technique for isolating genic DNA for quantitative traits in plants. J. Theor. Biol. 117: 1-10.

Sano Y, Maekawa M, Kikuchi H. (1985) Temperature effects on the Wx protein level and amylose content in the endosperm of rice. J. Heredity 76: 221-222.

Sasahara T, Ikarashi K, Kambayashi M (1986) Genetic variations in embryo and endosperm weights, seedling growth parameters and alpha-amylase activity of the germinated grains in rice (*Oryza sativa* L.). Japanese Journal of Breeding 36: 248-261.

Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M (2002) Green revolution: A mutant gibberellin-synthesis gene in rice. Nature 416: 701-702.

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen JL, Weigel D, Andersen SU (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nature Methods 6: 550-551.

Sugimoto K, Takeuchi Y, Ebana K, Miyao A, Hirochika H, Hara N, Ishiyama K, Kobayashi M, Ban Y, Hattori T, Yano M. (2010) Molecular cloning of Sdr4, a regulator involved in seed dormancy and domestication of rice. PNAS 107(13): 5792-5797.

Takahashi Y, Shomura A, Sasaki T, Yano M. (2001) Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. PNAS 98(14): 7922-7927.

Tester M and Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327: 818-822.

Toki S, Hara N, Ono K, Onodera H, Tagiri A, Oka S, Tanaka H (2006) Early infection of scutellum tissue with Agrobacterium allows high-speed transformation of rice. The Plant Journal 46: 969-976.

Toyomasu T, Kawaide H, Sekimoto H, Numers CV, Phillips AL, Hedden P, Kamiya Y (1997) Cloning and characterization of a cDNA encoding gibberellin 20-oxidase from rice (*Oryza sativa*) seedlings. Physiologia Plantarum 99: 111-118.

Uchida N, Sakamoto T, Kurata T, Tasaka M (2011) Identification of EMS-induced causal mutations in a non-reference Arabidopsis thaliana accession by whole genome sequencing. Plant Cell Physiol 52: 716-722.

Williams JF, Peterson ML (1973) Relations between alpha-amylase activity at and growth of rice seedlings. Crop Science 13: 612-615.

Yamaguchi S, Kamiya Y (2000) Gibberellin biosynthesis: Its regulation by endogenous and environmental signals. Plant and Cell Physiology 41: 251-257.

Yamaguchi T, Ebitani T, Omoteno M, Yamamoto Y (2007) Japanese native red rice cultivar, 'Awa-akamai', a genetic resource for the improvement of germination in soil at low temperatures. Breeding Research 9: 103-110 (in Japanese with English abstract)

Yamamoto T, Yonemaru J, Yano M (2009) Towards the understanding of complex traits in rice: Substantially or superficially?. DNA Research 16: 141-154.

Yano K, Takashi T, Nagamatsu S, Kojima M, Sakakibara H, Kitano H, Matsuoka M, Aya K (2012) Efficacy of microarray profiling data combined with QTL mapping for the identification of a QTL gene controlling the initial growth rate in rice. Plant and Cell Physiology 53: 729-739.

Yano M (2001) Genetic and molecular dissection of naturally occurring variation. Current Opinion in Plant Biology 4: 130-135.

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(12): 2473-2484.

Yano M, Sasaki T (1997) Genetic and molecular dissection of quantitative traits in rice. Plant Molecular Biology 35: 145-153.

Zhang ZH, Qu XS, Wan S, Chen LH, Zhu YG (2005b) Comparison of QTL controlling seedling vigour under different temperature conditions using recombinant inbred lines in Rice (*Oryza sativa*). Annals of Botany 95: 423-429.

Zhang ZH, Yu SB, Yu T, Huang Z, Zhu YG (2005a) Mapping quantitative trait loci (QTLs) for seedling-vigor using recombinant inbred lines of rice (*Oryza sativa* L.). Field Crops Research 91: 161-170.

Zhou L, Wang JK, Yi Q, Wang YZ, Zhu YG, Zhang ZH (2007) Quantitative trait loci for seedling vigor in rice under field conditions. Field Crops Research 100:

249-301.

Zuryn S, Le Gras, S, Jamet K, Jarriault S (2010) A strategy for direct mapping and identification of mutations by whole-genome sequencing. Genetics 186: 427-430.

List of publications

Abe A, Takagi H, Fujibe T, Aya K, Kojima M, Sakakibara H, Uemura A, Matsuoka M, Terauchi R (2012) *OsGA20ox1*, a candidate gene for a major QTL controlling seedling vigor in rice. Theoretical and Applied Genetics 125: 647-657..

Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, Innan H, Cano L, Kamoun S, Terauchi R (2012) Genome sequencing reveals agronomically-important loci in rice using MutMap. Nature Biotechnology 30: 174-178.

Okuyama Y, Kanzaki H, Abe A, Yoshida K, Tamiru M, Saitoh H, Fujibe T, Matsumura H, Shenton M, Galam DC, Undan J, Ito A, Sone T, Terauchi R (2011) A multifaceted genomics approach allows the isolation of the rice *Pia*-blast resistance gene consisting of two adjacent NBS-LRR protein genes. The Plant Journal 66: 467-479.

Rakshit S, Kanzaki H, Matsumura H, Rakshit A, Fujibe T, Okuyama Y, Yoshida K, Muluneh T, Shenton M, Utsushi H, Mitsuoka C, Abe A, Kiuchi Y, Terauchi R (2010) Use of TILLING for reverse and forward genetics of rice. In: Meksem K and Kahl G (eds) The Handbook of Plant Mutation Screening: Mining of Natural and Induced Alleles. Pp 187-198. Wiley-VCH

Acknowledgements

本研究をまとめるにあたり、岩手大学大学院連合農学研究科教授（財団法人 岩手生物工学研究センター研究部長）寺内良平博士には、懇切丁寧、かつ熱い御指導、御鞭撻を終始賜りました。心よりの深謝を申し上げます。副査を引き受けてくださった岩手大学農学部附属寒冷バイオフロンティア研究センター教授 上村松生博士、弘前大学農学生命科学部教授 石川隆二博士には貴重な御指導と御助言を賜りました。また、帯広畜産大学地域環境学研究部門教授 加藤清明博士には貴重な御指導と御助言を賜りました。総合研究大学院大学先導科学研究科准教授 印南秀樹博士には MutMap 法の理論にかかる統計的な解析を賜りました。厚くお礼申し上げます。

岩手県農業研究センターの皆様には、試験および調査の実施にあたり多大なる御協力をいただきました。技術部長 畠山均氏、作物研究室長 佐々木力氏には、有意義な御指摘、御助言を賜りました。環境部長 木内豊氏には、当時水稲育種研究室長として社会人入学を快諾していただくとともに、水稲育種に係る技術において多大なる御指導を賜りました。菅原浩視 上席専門研究員をはじめとする先輩研究員、同僚研究員の皆様、菊地徳章 技能員、臨時職員および日々雇用職員の皆様には圃場管理や調査に御協力を賜りました。皆様に感謝の意を表します。

また、試験の遂行にあたり多大なる御指導、御助言、御協力を賜りました岩手生物工学研究センター 小杉俊一博士、吉田健太郎博士、高木宏樹氏、神崎洋之博士、齋藤宏昌博士、Dr. Mulunhe Tamiru Oli、夏目俊氏、藤部貴宏博士、松村英生博士、植村亜衣子氏、宇津志博恵氏、三岡周子氏、神崎英子氏に感謝の意を表します。

最後に、長期にわたり支えていただきました多くの皆様と家族に深謝します。

平成 24 年 3 月　阿部　陽