Exome sequencing of ion-beam-induced mutants facilitates the detection of candidate genes responsible for phenotypes of mutants in rice

Short title: Exome sequencing of ion-beam-induced rice mutants

Yutaka Oono[1*], Hiroyuki Ichida[2], Ryouhei Morita[2], Shigeki Nozawa[3], Katsuya Satoh[1], Akemi Shimizu[4], Tomoko Abe[2], Hiroshi Kato[4, #], Yoshihiro Hase[1].

[1] Department of Radiation-Applied Biology, Takasaki Advanced Radiation Research Institute (TARRI), Quantum Beam Science Research Directorate, National Institutes for Quantum and Radiological Science and Technology (QST), Takasaki, Gunma, JAPAN.

[2] Ion Beam Breeding Team, Nishina Center for Accelerator-Based Science, RIKEN, Wako, Saitama, JAPAN.

[3] Research Planning Office, Quantum Beam Research Directorate, QST, Takasaki, Gunma, JAPAN.

[4] Radiation Breeding Division (RBD), The National Agriculture and Food Research Organization (NARO), Hitachi-ohmiya, Ibaraki, JAPAN.

[#] Current address: Genetic Resources Center, NARO, Tsukuba, Ibaraki, JAPAN.

*Corresponding Author:

E-mail: ohno.yutaka@qst.go.jp (YO)

Abstract

Ion beams are physical mutagens and used for plant and microbe breeding. They are thought to cause mutations via a distinct mechanism from chemical mutagens or gamma rays. Characteristics of ion-beam-induced mutations have been analyzed using marker genes; however, little is known about the extent of the mutations induced by ion beams at a genomic level. To understand the properties of ion beam-induced mutations at a genomic level, we conducted whole-exome sequencing of rice DNA. DNA extracted from carbon-ion-beam-induced rice mutants were fragmented and captured with a custom probe library covering 66.3 M bases of all exons and miRNAs. A total of 56 mutations, including 24 single nucleotide variations, 23 deletions, and 5 insertions, were detected in 5 mutant rice lines (2 dwarf and 3 early-heading-date mutants). The mutations were distributed among all 12 chromosomes, and the average mutation frequency in the M1 generation was estimated to be 2.6×10^{-7} per base. Many single base insertions and deletions were associated with homopolymeric repeats, whereas larger deletions up to 7 base pairs were more linked to polynucleotide repeats in the DNA sequences of the mutation sites. Among 56 mutations, 6 (1.2 mutations per line on average) were classified as high-impact mutations that caused a frame shift or loss of exons. A gene that was functionally related to the phenotype of the mutant was disrupted by a high-impact mutation in 4 of the 5 lines, suggesting that whole-exome sequencing of an ion-beam-irradiated mutant could facilitate detection of a candidate gene responsible for the mutant phenotype.

Introduction

Ion beams are charged particles that are derived from particle accelerators using electromagnetic fields. As with other ionizing radiations, ion beams cause damage to DNA molecules in living organisms and have been used as physical mutagens for plant and microbe breeding [1–3]. Ion beams are characterized by the deposition of a high energy transfer per unit length (linear energy transfer, LET) and are believed to induce mutations as a consequence of distinct biological effects from low LET radiation such as gamma-rays and electrons. In fact, during the screening of mutants from irradiated explants of carnations, ion beams induced a wider variety of mutants in terms of flower color and shape than gamma-rays and X-rays [4]. The appearance of more flower color variations in the ion-beam-irradiated population has also been observed in chrysanthemums [1]. In contrast, no remarkable difference in the mutation spectrum between ion-beam and gamma-ray irradiation has also been observed [5].

Characterization of ion beam-induced mutations in plant DNA was conducted via several approaches using Arabidopsis, a model plant for plant molecular genetics. The most common approach for characterizing germline mutations is the isolation of mutants deficient in well-characterized marker genes responsible for visible phenotypes such as seed color (*tt*) and leaf trichome morphology (*gl*), followed by analysis of the DNA sequence of the marker gene [6–9]. The isolation of mutants or tissue sectors of well-characterized marker genes is also useful for characterizing ion-beam-induced somaclonal mutations by detecting the loss of heterozygosity [10–12]. An alternative method that can eliminate the isolation of mutant plants is the use of the *rpsL* (*Escherichia coli* ribosomal protein small subunit S12) mutation detection system [13,14]. In this system, plants containing the *rpsL* transgene are irradiated, followed by genomic DNA

extraction from irradiated transgenic plants and introduction into *E. coli*. Mutated nonfunctional *rpsL* DNA fragments can be recovered from drug-resistant *E. coli* colonies. The results of these experiments using marker genes suggested that 1) ion beams induce various types of mutations in plant DNA, such as base(s) substitutions, DNA insertions and deletions (InDels), inversions (INVs), and chromosomal translocations, 2) regarding germline mutations, ion beams induce a higher ratio of rearrangement [InDels (> 100 bp), inversions, translocations, and total deletions of the marker gene] than electron beams, which induce mostly point-like mutations [base substitutions and InDels (< 100 bp)] [7], 3) the size of deletion induced by ion beams correlates positively with the degree of LET [11,15,16], and 4) most mutants obtained by the pollen-irradiation method carried large deletions reaching up to > 6 Mb, most of which are not transmissible to the next generation [10].

Recent advances in genome analysis technologies have allowed us to analyze mutations more quantitatively at the genomic level without bias arising from genomic positioning and the functional significance of the DNA sequences of a specific marker gene. The mutation-accumulating experiment combined with whole genome analysis of next-generation sequencing revealed the estimated spontaneous mutation rate in Arabidopsis to be 7×10^{-9} base substitutions per site per generation [17]. Whole genome sequencing of ion beam-irradiated Arabidopsis DNA suggested that 200-Gy carbon ion beams increased the mutation rate by nearly 47-fold compared with the spontaneous mutations [18]. In addition, Arabidopsis whole genome sequencing has demonstrated that the frequency and type of mutations induced by ion beams are affected by the LET of ion beams [19] and the physiological status of irradiated tissues [20].

Rice is one of the most important crops for humans and is a major target crop for ion-beam breeding [21–25]. Several genes that cause mutant phenotypes were successfully cloned by map-based cloning from ion-beam-induced rice mutants, and the mutation sites in the genes were characterized [22,25]. However, due to a relatively large genome size compared with the dicotyledonous model plant Arabidopsis, the characterization of ion beam-induced mutations at the genomic level was not achieved in rice until Ichida et al. [26] recently established a multiplexed and cost-efficient whole-exome sequencing procedure in rice. In the present work, we employed the whole-exome sequencing procedure to analyze the properties of induced mutations in selected rice mutants generated with carbon ion beams accelerated using the Azimuthally Varying Field (AVF) cyclotron of the Takasaki Ion Accelerators for Advanced Radiation Application (TIARA) at TARRI, QST. We also demonstrated that the limited number of induced mutations with a potentially high impact on protein function quickly narrowed down the candidate genes responsible for the mutant phenotypes.

Materials and Methods

Plant Materials and ion beam irradiation

Rice seeds (*Oryza sativa* L. cv Nipponbare, NPB) harvested in RBD, NARO (Hitachi-ohmiya, Japan) were used in this work. Water contents in the rice seeds were adjusted to 12~13% by keeping the rice seeds in a plant growth chamber (BIOTRON LPH200, NK system, Japan) set at 24 °C and 60% RH for 4 days just before irradiation. Then, the seeds were irradiated with 10 to 150 Gy of 25.9 MeV/u $^{12}C^{6+}$ ions (LET on surface: 76 keV/μm)

accelerated by an AVF cyclotron in TIARA, TARRI, QST (Takasaki, Japan). The irradiated seeds were sown on soil to grow plants (M1 plants) in a greenhouse at TARRI, QST. The dose response for the survival rate and plant height at one month after sowing were measured. The seeds treated with 40 Gy of $^{12}C^{6+}$ ions were used further genetic screening of mutants. M2 or M3 seeds were harvested from every individual mature M1 or M2 plant, respectively. For selection of mutant candidates, M2 and M3 plants were grown in the greenhouse at TARRI, QST or in an experimental paddy field at RBD, NARO (36°52'45"N 140°39'81"E). Some mutant lines were backcrossed to NPB. The resulting seeds (F1) were germinated and grown in the greenhouse to generate seeds of F2 plants, which were used for linkage analysis between the mutations and the mutant phenotype.

DNA extraction, whole exome capturing, and sequencing

Genomic DNA for whole exome capturing was extracted from rice leaves using MagExtractor –Plant Genome– (Toyobo, Japan) according to a protocol supplied by the manufacturer. One microgram of genomic DNA from individual samples was fragmented using a Covaris S220 instrument (Covaris inc, USA). The resulting DNA fragments were purified with AMPure XP reagent (Beckman Coulter, USA). Barcoded sequencing libraries were prepared using NextFlex Rapid DNA-Seq Kits (Bioo Scientific, USA) according to the manufacturer's protocol. After double-sided size-selection with AMPure XP, pre-capture amplification was performed for 12 cycles with the supplied primers 1 and 2. The Os-Nipponbare-Reference-IRGSP-1.0 sequences and annotations were used as the reference (available from https://rapdb.dna.affrc.go.jp/download/irgsp1.html; as of 2015/3/31). A custom whole-exome-capturing probe library, which covers 66.3 Mb of

107,616 rice genome regions consisting of all predicted exons in the IRGSP-1.0 annotations, including 100 bp each of the 5'- and 3'-franking regions as well as known miRNA regions including 700 bp of the upstream and 300 bp of the downstream regions, was designed and synthesized as a SeqCap EZ Developer Library (Roche Diagnostics, USA). After 16–20 h of hybridization at 47 °C, the hybridized DNA fragments were captured using streptavidin-coupled magnetic beads (Dynabeads M-280 streptavidin, Thermo Fisher Scientific, USA), washed, and amplified by PCR using KAPA HiFi Hotstart Ready Mix (KAPA Biosystems) with the same primers described above. The whole exon-enriched libraries were sequenced on a HiSeq 4000 instrument (Illumina, USA) in PE100 mode.

Data analysis was performed with the bioinformatics pipeline described elsewhere [26]. Briefly, the resulting sequencing reads were mapped to the reference sequences using the BWA-MEM program version 0.7.15 (http://bio-bwa.sourceforge.net) with default parameters, followed by manipulation with SAMtools version 1.3.1 (http://samtools.sourceforge.net/), Picard software package version 2.3.0 (https://broadinstitute.github.io/picard/) for sorting the reads, marking and removing duplicates, and GATK software package version 3.7.0 (https://software.broadinstitute.org/gatk/) for local realignment and base quality recalibration. Variant calling was carried out with a combination of the GATK, Pindel version 0.2.5a8 (http://gmt.genome.wustl.edu/packages/pindel/user-manual.html), and Bedtools version 2.26.0 (https://bedtools.readthedocs.io/en/latest/) programs. To remove intra-cultivar polymorphisms between the reference sequences and our parental NPB line used for the irradiation experiments, the raw variant calling results among the mutants were compared, and the variants that were present in only one mutant line were extracted

as the line-specific (real) mutations. For the GATK and Pindel results, the detected line-specific mutations were further filtered out by the following criterion: 1) a mutation that is supported with less than 10 reads was removed to eliminate false-positives due to an insufficient number of mapped reads, 2) a homozygous mutation was removed if reads harboring the same mutation accounted for more than 5% of the total reads in equal or more than half of homozygous lines for the other allele (unmutated allele). Mutations detected with each program were merged together by the normalized chromosomal positions (start and end coordinates), and duplicated mutation records that were detected by multiple programs were removed from the final result. In the Bedtool analysis, regions covering less than 20% in one individual line but 100% in all the other lines were extracted. To increase reliability, regions with less than an average of 10 reads in all lines with a wild type homozygous allele were excluded. The predicted effect and location of the mutations were annotated using the SnpEff program version 4.1b (http://snpeff.sourceforge.net/). After variant calling against the whole genome sequence, only mutations located within the exome target regions of whole-exome capturing were extracted and outputted in Variant Calling Format (VCF). The identified mutations were checked visually using the Integrative Genomics Viewer (IGV) (http://www.broadinstitute.org/igv/).

To confirm the mutation by Sanger sequencing, genomic DNA was amplified by PCR using Takara Ex Taq polymerase (Takara Bio, Japan) followed by the dye terminator cycle sequencing reaction with the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, USA). The resulting samples were analyzed using a 3500 Genetic Analyzer (Thermo Fisher Scientific, USA). The primers used for PCR and Sanger sequencing are listed in S1 Table.

Estimation of the genome-wide mutation frequency

Mutation frequencies (MFs) were calculated by dividing the total number of mutations detected in individual plant DNA by the total number of nucleotides in the captured target region (66,258,908 bases). Assuming all mutations arise as heterozygotes in the M1 generation and are simply inherited according to Mendelian characteristics, all homozygous mutations in M2 and later generations are transmitted to the next generation, but a quarter of the heterozygous mutations are not transmitted, the number of mutations occurred in M1 plants (N_1) is estimated as

$$N_1 = 2^n / (2^{n-1} + 1) \times Nn$$

where n is the generation and Nn is the number of mutations in the Mn diploid genome. For comparison with the whole genome sequencing results of other researchers in the referenced articles, the number of mutations was divided by the size (number of bases) of the reference genome described in the articles.

Results and Discussion

Preparation of the ion-beam irradiated rice population and mutant selection

We chose the rice cultivar Nipponbare (NPB) in this work because its high-quality whole genome sequence is publicly available [28]. Since the water content of a seed affects its radiation sensitivity [29], we adjusted the water contents of the seeds to 12 ~ 13% before irradiation. To estimate the biological effectiveness of the carbon ion beam on the seeds

of NPB, dose response curves for the survival rate and plant height were generated (Fig 1A). Because a shoulder dose or slightly lower dose in the survival curve is empirically considered as the optimum for obtaining a large number of mutants [1,5,21], the absorbed dose chosen for the genetic screen in this work was 40 Gy, approximately 2/3 of the shoulder dose (approximately 60 Gy) for the survival curve and slightly over the shoulder dose for the dose response curve of plant height.

Fig 1. Estimating the effect of carbon ion beams on rice.

(A) Dose response for survival rate and relative plant height of NPB rice seeds irradiated by $^{12}C^{6+}$ ion beams. Mean values against the mean value of unirradiated controls (0 Gy) are plotted. Error bars indicate SE. Data are from 5 replicates with 10 seedlings per treatment.

(B) Phenotypes of chlorophyll-related mutants. The number of these mutants in the M2 seedling was counted to estimate the mutagenic effect of the 40 Gy irradiation of carbon ion beams used in this work.

To estimate the mutagenic effect of 40 Gy irradiation of carbon ion beams, the M2 seeds (4 ~ 10 seeds / M1 line) from 2,039 individual M1 plants were sown on soil, and chlorophyll mutants were searched at the seedling stage. In this pilot study, several types of chlorophyll mutants, showing a pale or white leaf color, were identified (Fig 1B). The appearance rate of the chlorophyll mutants was 6.6% (134 / 2,039 M1 lines), suggesting that the carbon-ion-beam irradiation effectively induced mutations in the rice genome as expected. We also screened other types of mutants with an altered phenotype from this M2 population in the green house and paddy field. In this work, we focused on two dwarf

(lines “885” and “3098”) and three early-heading-date (lines “786-5”, “IRB3517-3”, and “IRB3790-2”) mutants that were confirmed to show the same phenotypes in the next (M3) generation (Fig 2) for further experiments. The mutant lines “885” and “3098” were identified in the greenhouse, whereas lines “IRB3517-3”, and “IRB3790-2” were found in the paddy field. Line “786-5” was initially found in the green house as a mutant with wide leaves, and later, the early heading phenotype was detected in the paddy field. Homozygosity of the mutant phenotypes was confirmed in the M3 generation of all the lines except “IRB3790-2”, for which we failed to obtain a homozygous population even in the M4 generation. Leaf tissues excised from an M3 plant of each line except “3098”, for which the leaf was from an M2 plant, were subjected to genomic DNA extraction for whole-exome sequencing analysis.

Fig 2. Dwarf and early-heading-date mutants analyzed in this work.

(A) Comparison of plant height (length of culms and panicles) among the two dwarf mutants [“885” (M4) and “3098” (M3)] and NPB control. The plant height of mature plants was recorded in the paddy field. Error bars indicate SD (n = 8~9) for plant height. *P* values in two sample *t* tests were $< 10^{-14}$ for both NPB vs “885” and NPB vs “3098”.

(B) Left panel: Phenotype of the dwarf mutant, “3098”. (3 middle lanes) and controls (most left and right) grown in the paddy field. Center panel: Magnified photo of “3098”. Right panel: Comparison of the shape of the panicles of NPB control and “3098”.

(C) Heading date of NPB, “786-5”, “IRB3517-3”, and “IRB3790-2” grown in the paddy field. One-month-old seedlings were transferred to the paddy filed on May 20, 2016. The date of first heading in the population was observed and recorded. The M4, M3, and M3 populations were used for “786-5”, “IRB3517-3”, and “IRB3790-2”, respectively.

(D) Left panel: The wide leaf phenotype of the "786" seedling.
Right panel: The NPB control and "786" rice growing in the paddy field. The photograph was obtained on August 12, 2016, when all "786" rice had completed initiation of heading, but not the NPB rice.

Results of the whole-exome captured sequencing in rice

We subjected all five mutants to whole-exome capture and massively parallel sequencing analysis to comprehensively identify the mutations introduced by the carbon-ion beam irradiation. By whole-exome capture followed by Illumina sequencing, we obtained the sequence data for 61.0 ~ 69.9 million unique reads with 6.08 ~ 6.96 billion unique bases aligned to the reference genome (Table 1). The number of bases mapped on the target region was 2.38 ~ 2.78 billion. The mean target coverage was 36 ~ 42, and the percentage of all target bases achieving 10 times or greater coverage was 95.4 ~ 96.0% (Table 1).

Table 1. Summary of the whole-exome captured sequencing

	Line					Mean	SD
	3098	786	885	IRB3517-3	IRB3790-2		
Reference genome size (bases)	373,245,519	373,245,519	373,245,519	373,245,519	373,245,519		
Size of target (bases)	66,258,908	66,258,908	66,258,908	66,258,908	66,258,908		
No. of reads aligned	65,509,961	61,048,713	69,002,757	69,858,886	67,703,204	66,624,704	3,521,809
No. of bases aligned	6,527,192,732	6,082,274,964	6,874,937,097	6,959,746,867	6,745,199,762	6,637,870,284	350,806,394
No. of bases mapped on targeted region	2,684,758,131	2,384,714,814	2,637,344,060	2,776,001,880	2,734,422,333	2,643,448,244	153,731,509
Mean target coverage	40.52	35.99	39.80	41.90	41.27	39.90	2.32
% target bases 10X [1]	0.956	0.954	0.959	0.960	0.957	0.957	0.002

[1] The fraction of all target bases achieving 10X or greater coverage

Through the pipeline described in Materials and Methods, GATK, Pindel and Bedtools programs initially called 61 line-specific mutations in the target regions. By manual verification of VCF files, two or three mutations that were overlapped or located adjacent to one another were considered as one mutation and categorized as a replacement (RPL) (mutation #12, 13, 37 in Table 2, S2 Table and Fig 3). In the Bedtools program, one large deletion was suggested on chromosome 3 of the "IRB3517-3" genome. This large deletion was also called by the Pindel program (mutation # 38). Consequently, the number of mutations was reduced to 56 (Table 2 and 3). To verify the accuracy of the mutation calling in the pipeline, DNA fragments corresponding to the 25 selected mutations (Table 2), including the junction sites of the large deletion (mutation #38) and the inversion (mutation #52), were amplified by PCR and reanalyzed by Sanger sequencing. The sequences of all the selected mutations were confirmed as they were called by the pipeline, except the 3-bp shift at one of the junction sites of the #52 inversion, suggesting that the results of our exome analysis were reliable, at least in regard to false positives.

Table 2. List of mutations[1]

Mutation No.	Line	CHR No.	Type after evaluation[2]	Annotation	Genotype [3]	Verification [4]
1	3098	chr02	INS (1)		0|1	
2	3098	chr02	DEL (1)		0|1	
3	3098	chr03	SNV	APG6/CLPB-P/CLPB3.	0|1	
4	3098	chr03	SNV	DCL2A.	0|1	
5	3098	chr04	INS (7)		1|1	✓
6	3098	chr05	SNV	Diacylglycerol acylCoA acyltransferase.	0|1	
7	**3098**	**chr05**	**DEL (128)**	**Guanine nucleotide-binding protein alpha-1 subunit.**	**0|1**	✓
8	3098	chr08	DEL (1)		0|1	
9	3098	chr08	INS (1)		0|1	
10	3098	chr12	DEL (1)		0|1	
11	786	chr01	SNV	Glutelin family protein.	0|1	
12 [5]	786	chr01	RPL		1|1	✓
13 [5]	786	chr03	RPL	predicted protein.	1|1	✓
14	786	chr03	SNV		1|1	
15	**786**	**chr03**	**DEL (5)**	**Phytochrome B.**	**1|1**	✓
16	786	chr06	SNV		1|1	
17	786	chr07	SNV		1|1	
18	**786**	**chr10**	**DEL (1)**	**Expansin/Lol pI family protein.**	**1|1**	✓
19	**786**	**chr12**	**DEL (1)**	**Nucleotide-binding, alpha-beta plait domain containing protein.**	**0|1**	
20	885	chr01	DEL (3)		1|1	
21	885	chr01	DEL (1)		0|1	
22	885	chr01	SNV		1|1	✓

23	885	chr01	DEL (46)	Soluble inorganic pyrophosphatase	1|1	✓
24	885	chr02	DEL (1)	casein kinase I isoform delta-like.	0|1	
25	885	chr03	SNV		0|1	
26	885	chr03	DEL (1)		1|1	✓
27	885	chr04	SNV		1|1	✓
28	885	chr05	SNV	Nucleotide-binding, alpha-beta plait domain containing protein.	0|1	
29	885	chr05	INS (1)		0|1	
30	885	chr05	SNV	DNA glycosylase/lyase 701.	1|1	✓
31	885	chr08	SNV		1|1	
32	885	chr08	DEL (6)		1|1	✓
33	885	chr08	DEL (4)		1|1	✓
34	885	chr08	SNV	Hypothetical conserved gene.	1|1	✓
35	885	chr10	DEL (6)	CYP71Z8	1|1	✓
36	IRB3517-3	chr01	SNV	Cell division protein ftsH	0|1	
37 [5]	IRB3517-3	chr01	RPL		1|1	✓
38	**IRB3517-3**	**chr03**	**DEL (33.6k)**	**5 ORFs are deleted [6]**	**1|1**	**✓**
39	IRB3517-3	chr04	SNV		1|1	
40	IRB3517-3	chr06	DEL (1)		0|1	
41	IRB3517-3	chr06	DEL (6)	Heavy metal ATPase.	1|1	✓
42	IRB3517-3	chr07	DEL (8)		1|1	✓
43	IRB3517-3	chr08	SNV	Helix-loop-helix DNA-binding domain containing protein.	0|1	
44	IRB3517-3	chr09	SNV	Phosphate/phosphoenolpyruvate translocator.	1|1	✓
45	IRB3517-3	chr09	SNV		0|1	✓
46	IRB3517-3	chr11	SNV	Hypothetical gene.	1|1	
47	IRB3517-3	chr11	SNV	NB-ARC domain containing protein.	1|1	✓

48	IRB3517-3	chr12	DEL (6)	Heterochromatin protein.	0|1	
49	IRB3790-2	chr01	DEL (1)		0|1	
50	IRB3790-2	chr03	INS (1)		0|1	
51	IRB3790-2	chr04	SNV		1|1	
52	**IRB3790-2**	**chr06**	**INV (535k)**	**HEADING DATE 1** [6]	**0|1**	✓
53	IRB3790-2	chr07	SNV	Esterase precursor	0|1	
54	IRB3790-2	chr08	DEL (4)		1|1	✓
55	IRB3790-2	chr09	DEL (2)		1|1	
56	IRB3790-2	chr12	SNV		1|1	✓

[1] Mutations putatively having a high impact on protein function are indicated in bold.

[2] The size of InDels is indicated in parentheses.

[3] "0|1" is heterozygous and "1|1" is homozygous.

[4] Mutations verified by Sanger sequencing are marked with a checkmark. Primer sequences are shown in S1 Table.

[5] Independent mutation calls are unified and counted as one RPL mutation because they are overlapped or located consecutively (Fig 3).

[6] Evaluated manually.

Table 3. Summary of the number of mutations and MF

Line name	Phenotype	Generation	# of mutation				MF	Estimated MF in M1	# of high impact mutation	Type of high impact mutation
			Total	SNV	InDel	Others				
3098	Dwarf Small grains	M2	10	3	7	0	1.5E-07	2.0E-07	1	128-bp DEL
786	Early heading date Wide leaves Broken flag leaf	M3	9	4	3	2	1.4E-07	2.2E-07	3	1-bp DEL x2, 5-bp DEL
885	Dwarf Colum bending	M3	16	7	9	0	2.4E-07	3.9E-07	0	
IRB3517-3	Early heading date	M3	13	7	5	1	2.0E-07	3.1E-07	1	33.6-kb DEL
IRB3790-2	Early heading date	M3	8	3	4	1	1.2E-07	1.9E-07	1	535-kb INV
Total			56	24	28	4			6	
Average			11.2					2.6E-07	1.2	

Fig 3. Nucleotide sequences in the mutation site of 3 RPLs (#12, 13, and #37) and junction sites of the 535-kb INV (#52).

Sequences in the NPB control and mutants are shown at the top and bottom, respectively. The underlined regions indicate the nucleotides of "Reference" and "Alternatives" in S2 Table. For mutations #12 and #37, altered nucleotides between the NPB control and mutants are shown in red. The dinucleotide repeats in control DNA of #12 and #37 mutation sites are indicated by top-opened brackets. For #52, green and blue letters are overlapped and deleted nucleotides in junction sites. Inverted sequences, excluding the overlapped nucleotides, are shown in red.

The 56 mutations consisted of 24 single nucleotide variations (SNV), 23 deletions (DELs), 5 insertions (INSs), 1 inversion (INV), and 3 RPLs (Table 2 and 3). The average number of mutations per line was 11.2 ± 3.3 for all five mutant lines and 11.5 ± 3.7 for the four M3 lines. The mutations were distributed among all 12 chromosomes (Fig 4). The 1 bp DEL (mutation #40) of "IRB 3517-3" and the 535 kb-large INV (mutation #52) of "IRB 3790-2" were located at the same region on the upper arm of chromosome 6. The mutation frequencies (MFs) were calculated to be 1.2×10^{-7} to 2.4×10^{-7} per base, and the MFs in the M1 generation were estimated to be 1.9×10^{-7} to 3.9×10^{-7} per base and, on average, 2.6×10^{-7} per base (Table 3).

Fig 4. Chromosomal location of the 56 mutations detected in this work.

Rice chromosomes are shown as boxes in sky blue with chromosomal numbers at the top and centromeres marked by yellow boxes. The mutation sites are indicated by horizontal

bars in the chromosome boxes with the mutation number (left) and type (right) color-coded by the mutant lines: “3098” (red), “786-5” (blue), “885” (orange), “IRB 3517-3” (magenta) and “IRB 3790-2” (green). Numbers in parentheses indicate the size (bp) of the deletion or insertion. The green box in the upper arm of chromosome 6 shows the 535-kb inverted region of mutation #52. The position of centromeres are based on the Os-Nipponbare-Reference-IRGSP-1.0 database (RAP-DB https://rapdb.dna.affrc.go.jp/).

We compared the estimated MFs to previous genome-wide mutation studies in rice. The whole-exome sequencing of unselected M2 plants obtained from dry rice seeds irradiated with 150 Gy of carbon ions (135 MeV/u, LET: 30 keV/μm) reported that the number of mutations was 9.06±0.37 (average ± standard error) per plant [26]. MF were estimated as 2.2×10^{-7} per base by dividing the number of mutations with 41.75-Mb of the targeted exon region. These values are comparable to the values of the present study. In another case, the exome analysis of rice EMS-mutagenized DNA with 39-Mb target probes detected an average of 70 to 508 mutations per individual M2 plant of 4 independent capture experiments [30]. MFs in M2 plants can be calculated as 1.7×10^{-6} to 1.3×10^{-5} per base. In contrast, whole genome sequencing of 3 independent regenerated rice plants by tissue culture suggested that 2,492, 1,039, and 450 SNVs occurred in the M1 generation, indicating that the average of MF for base substitution in M1 was 3.5×10^{-6} per base [31]. Another whole genome sequencing of regenerated rice plants that were selfed for 8 consecutive generations detected 54,268 DNA polymorphisms, including 37,332 SNPs and 16,936 small INSs and DELs, indicated that the MF was 1.5×10^{-4} per site for total polymorphisms [32]. Conversely, large-scale whole genome sequencing of a mixed population of M2 and M3 of fast neutron (FN) rice mutant correction identified

91,513 mutations in 1,504 lines [33], estimating the MF in this population to be 1.6×10^{-7} per base. The number of detectable mutations by genome sequencing is highly dependent on the experimental conditions of mutagenesis, quality and quantity of massively parallel sequencing, and bioinformatics pipeline for sequence analysis. It should also be noted that the number of mutations per genome estimated by whole-genome sequencing might be larger than that estimated by whole-exome sequencing because most genomic regions are intergenic regions, and a mutation occurring in the intergenic region does not have a direct impact on the function of a protein that may affect the viability of cells and individual rice lines. Although an accurate comparison is difficult for the above reasons, the data published to date along with our results suggest that carbon ion irradiation seems to induce a reduced number of mutations in rice DNA than EMS- and tissue culture-based mutagenesis and a similar number to FN mutagenesis.

The values of MF were also reported in carbon ion-mutagenized Arabidopsis: 3.4×10^{-7} single base mutation per base of an M1 plant with seeds exposed to 200Gy of $^{12}C^{6+}$ ions (43 MeV/u; average LET within samples, 50 keV/μm) [18]. Another report using 17.3 MeV/u carbon ions (surface LET, 107 keV/μm) at TIARA also showed MFs of 2.7×10^{-7} and 3.2×10^{-7} per base in M2 plants generated by dry seed irradiation (125 and 175 Gy, respectively), and 1.8×10^{-7} and 1.6×10^{-7} per base in M2 plants generated by seedling irradiation (20 and 30 Gy, respectively) [20]. Kazama et al. [19] reported 307 and 473 mutations (sum of rearrangements, single base substitutions, and small INSs and DELs) in 8 lines of argon ion- (290 keV/μm, 50 Gy) and carbon (30.0 keV/μm, 400 Gy) ion-mutagenized M3 populations obtained by dry seed irradiation, implying MFs of 3.2×10^{-7} and 5.0×10^{-7} per base, respectively. Therefore, MFs caused by ion-beam irradiation seem to be the same in order of magnitude between Arabidopsis and rice.

Among 56 mutations detected, 24 (43%), 23 (41%) and 5 (9%) mutations were SNVs, DELs, and INSs. Although it is difficult to perform a direct comparison because the absorbed dose for irradiation causing the same biological effect was not adjusted, the ratio of SNVs in this work was similar but slightly lower than the irradiation results of a lower level of LET carbon ion (135 MeV/u; LET 30 keV/μm) reported by Ichida et al. [26], in which more than half (58%) of the detected mutation were SNVs, followed by DELs (37%) and INSs (4%), in 110 independent M2 lines from dry rice seeds irradiated with 150 Gy of carbon ions. The FN mutagenesis of rice seeds also induced SNVs most frequently (48%) and DELs accounting for 35% [33].

The most frequent (38% of the total SNVs) base changes were G to A and C to T (GC → AT) transitions (Fig 5A). Another major base change was the AT → GC transition, which accounts for 33% of the total SNVs. This result is in contrast to the EMS-mutagenized rice DNA, in which most of the mutations were SNVs of GC → AT transitions [30]. Regarding to size of the 28 INSs and DELs (InDels), 14 InDels (50%) were 1 bp in size, and most of the InDels (89%) were less than 10 bp in length, excluding 3 relatively large DELs (mutation #7, #23, and #38), which were 128, 46, and 33.6 kb in length, respectively (Fig 5B).

Fig 5. Summary of small mutations detected in this work.

(A) Type of base change of 24 SNVs. The ratio of transition to transversion (Ti/Tv) was 2.4.

(B) Size distributions of 28 InDels.

Of the 28 InDels, 16 (57%) occurred within nucleotide repeats such as homopolymers that consisted of 3 or more repeats of the same nucleotides, or polynucleotide repeats that consisted of either partial or full repetitive sequences (Table 4). Many of the single-base InDels (8 of the 14 events; 57%) were associated with homopolymers, but none of them were associated with a polynucleotide repeat. In contrast, InDels that were more than 2 bp in length showed a greater linkage to polynucleotide repeats (7 of the 14 events; 50%) than homopolymers (1 of the 14 events; 7.1%). No such repeat was observed in the DEL site that were more than 7 bp in length. A similar trend was observed in the presence of homopolymers or polynucleotide repeats at flanking sequences of small InDels in the genome of carbon ion-beam-irradiated Arabidopsis [18,20]. Hase et al. [34] inferred that distinct mechanisms are involved in between the generation of the single base deletion and larger-sized deletions in carbon ion-irradiated Arabidopsis. They also suggested that much larger size deletions (≧ 50 bp) are generated via another distinct mechanism. The presence or absence of characteristic repeats in the InDel site in our data may also suggest that InDel generation mechanisms are different depending on the InDel size in the irradiated rice genome.

Table 4. List of InDELs and repeat sequences

Mutation Type	Size [1]	Mutation No. [2]	Sequence [3]	Repeat Type
INS	1	1	CAAC**A**ACAT	
		9	TTAA**T**TTTT	homopolymer
		29	AAAG**A**AAAA	homopolymer
		50	ATCA**T**TTTT	homopolymer
	7	5	AGAT**AGA**T**AGA**AGATAGA	full polynucleotide repeat
DEL	1	2	AGCC**a**AAAA	homopolymer
		8	CAAT**g**GTCA	

	10	TTTT**g**GTTA	
	18	CGAC**g**GGAA	homopolymer
	19	GTAG**c**TTTT	
	21	AACT**g**GGGG	homopolymer
	24	GAAT**a**AAGG	homopolymer
	26	CATT**c**CACC	
	40	TCAC**a**AAAA	homopolymer
	49	TCGG**c**TCAG	
2	55	GTTG**ga**ACAA	
3	20	AGATT**aga**AGAAGA	full polynucleotide repeat
4	33	TCCT**gaaa**GAAA	full polynucleotide repeat
	54	GAAC**ttaa**AAAA	homopolymer
5	15	GTGA**atggc**CAGGT	
6	32	TCTC**atcgaa**ATCAA	partial polynucleotide repeat
	35	AGGT**cgcgcc**CGTGC	partial polynucleotide repeat
	41	GCTT**cctgtc**TCAA	partial polynucleotide repeat
	48	GATT**cactac**CACTAC	full polynucleotide repeat
8	42	AGGG**cgtatatt**GATC	
46	23	CCCG**tgtc-----cgca**TATA	
128	7	TAGA**cgtt-----tgat**TCTC	
33.6K	38	ACAT**gagg-----gttg**GTGC	

[1] INSs and DELs are sorted by their size.

[2] The mutation no. is identical to that in Table 2.

[3] Inserted nucleotides and deleted nucleotides are shown in bold uppercase and bold lowercase, respectively. Nucleotide repeats are underlined.

The primary structure of the three mutations (#12, 13, and 37) classified as RPL are shown in Fig 3. In two RPLs, #12 and #37, 3 and 8 nucleotides were replaced with single bases accompanied by 2- and 7-bp deletions, respectively. In contrast, mutation #13 was a dinucleotide substitution, in which GG was substituted with TA. No homopolymer, but a dinucleotide repeat, was observed in the mutation site of #12 and #37 RPLs. In the 535-kb INV (mutation #52), 2-bp overlapped nucleotides in one of the two junction sites along with a 3-bp deletion on both sides were present (Fig 3). Although such a tendency was

observed in RPLs, it was difficult to determine whether these features in primary sequences were related to the process during DNA repair and mutation generation since the number of observed events was too small.

The effect of mutations on gene function was assessed by SnpEff [27]. The results indicated that 5 and 13 mutations caused putatively high (frameshift_variant and exon_loss_variant) and moderate (missense_variant and disruptive_inframe_deletion) impacts on protein function classified according to the SnpEff output (Table 2 and S2 Table). In addition, by manually confirming the break points of an inversion (mutation #52), we noticed that one of the two breakpoints of the inversion was located in a coding region of a gene, *Os06g0275000*. Thus, this inversion was categorized as a high impact mutation. In total, 6 mutations were classified as high impact mutations (an average of 1.2 such high-impact mutations per line) (Table 3). Among them, an exon_loss_variant (mutation #38) consisted of a deletion of a 33.6 kb-DNA region and caused a loss of 5 genes located within this region. Thus, the number of affected genes by 6 potentially high impact mutations was 10 (only 2 genes per line on average). Because high impact mutations cause a probable change in protein function, it is highly possible that they are causal mutations for the phonotypes of mutants. In fact, we found strong candidate mutations possibly having a major role on the mutant phenotype in 4 out of 5 lines, as demonstrated below.

Candidate genes that are responsible for mutant phenotypes

As described above, we isolated two dwarf mutants (lines “3098” and “885”) and three early-heading-date mutants (lines “786-5”, “IRB3517-3”, and “IRB3790-2”) in the

present study (Fig. 2). Through whole-exome sequencing analysis, we successfully identified candidate genes that might be responsible for the phenotypes of the mutants.

The mutant line "3098" was isolated as a dwarf mutant in the M2 population grown in the green house. Homozygosity of the dwarf phenotype was confirmed in M3 progenies grown in the green house as well as in the paddy field (Fig 2A and B). The seeds of line "3098" were small and rounded compared with those of NPB. The whole-exome sequencing analysis of DNA extracted from an M2 plant of line "3098" identified 10 mutations, mutations #1~10 in Table 2, including a 128-bp DEL (mutation #7) spanning the 3rd intron and 4th exon of the guanine nucleotide-binding protein alpha-1 subunit (*GPA1*, also called *RGA1*, and *D1*) gene, which was predicted to cause a truncated protein (Fig 6A). The functional disruption of this gene (*Os05g0333200*) is known to cause the Daikoku dwarf (*d1*), which is characterized by broad, dark green leaves, compact panicles, and short, round grains [35,36]. These phenotypes were apparent and quite similar in line "3098". Although the initial analysis of whole-exome sequences suggested that this mutation (mutation #7) was present as a heterozygote in the M2 plant, PCR analysis using a specific primer set indicated that the mutation was present as a homozygote in the same M2 plant. PCR analysis of 12 F2 plants (6 dwarf and 6 normal-look plants) from a backcross between line "3098" and NPB and subsequent self-crossing showed that all 6 dwarf plants harbored the 128-bp DEL (mutation #7) in homozygotes, whereas the remaining 6 were heterozygotes or wild type. Therefore, the 128-bp DEL in the *D1* gene was thought to be a strong candidate responsible for the dwarf phenotype of line "3098". The discrepancy in zygosity in the results of the initial bioinformatics call (heterozygous) and PCR (homozygous), as well as the phenotype (homozygous), could be due to the presence of highly homologous

DNA sequence encoding the *D1–like* gene (*Os05g0341300*), which is likely a pseudogene, in the same chromosome in the NPB genome [37]. A BLAST search of the 350-bp DNA sequence of the *D1* gene, including the 128 bp-deleted region and ~100-bp upstream/downstream sequences (positions 15612701-15613050 on chromosome 5), indicated 99% identity (3 nucleotide difference in 350 bp) with the region from 16009580 to 16009929 on chromosome 5, in which the *D1–like* gene is located (Fig 7). The nucleotide sequences were completely identical between the *D1* and *D1-like* genes in the 100-base upstream region of the deletion. Therefore, it is impossible to distinguish the right origin of short sequence reads in these regions, and some sequencing reads originating from the *D1–like* gene region were mapped to the deleted region in the *D1* gene. Consequently, the mutation was predicted to be heterozygous.

Fig 6. Structure of candidate genes and position of mutations.

Black boxes indicate exons, and a white arrow represents the 5' untranslated region. Introns and 3' untranslated regions are indicated by black lines.

(A) Guanine nucleotide-binding protein alpha-1 subunit (*GPA1*) gene. The position of the mutation (#7) in "3098", the 128-bp DEL, indicated by a blank box.

(B) Mutation #15 is the 5-bp DEL located in the first exon of the *PHYTOCHROM B* gene in the "786-5" DNA.

(C) Relative position of the genes (solid arrows) in and near the 33-kb DEL on chromosome 3 of "IRB3517-3" (mutation #38). The *HD16* gene is indicated in red.

(D) One of the break points of the 535-kb INV (mutation #52) in "IRB3790-2" is located in the second exon of the *HD1* gene. Nucleotide and corresponding amino acid sequences of the break point in the NPB control and "IRB3790-2" DNA are shown at the bottom.

The amino acid (R347) of the HD1 protein is marked with the amino acid position number. The green and blue letters are overlapped and deleted nucleotides in the junction sites. Inverted sequences, excluding overlapped nucleotides and altered amino acids, are shown in red. The detailed sequence of the junction sites is shown in Fig 3.

Fig 7. Alignment of the D1 and D1-like gene.

The 350-bp *D1* gene region, including a 128-bp-deleted region (mutation #7 indicated in red letters) plus ~100-bp upstream/downstream sequences (Oschr05 15612701-15613050) and a highly similar *D1–like* gene region (Oschr05 16009580-16009929), are aligned. Distinct nucleotides are indicated by bold letters surrounded by boxes. The numbers on both sides indicate the position on chromosome 5.

The line “885”, a dwarf mutant, was isolated from the M2 population grown in the greenhouse, and its homozygosity was confirmed in the M3 population in both the green house and paddy field (Fig. 2A). Sixteen mutations were identified by whole-exome sequencing (Table 2). Although 4 of them (mutations #28, 30, 34, and 35) were homozygous nonsynonymous mutations (S2 Table), the analysis of F2 lines suggested that none of these four mutations were linked to the dwarf phenotype of the “885” line. It is possible that mutations located outside of the exon regions, presumably promoter regions, could be responsible for the dwarf phenotype observed in this line.

The mutant line “786-5” was initially identified as a mutant with wide leaves in our greenhouse (Fig 2C). The mutant also showed increased leaf blade declination (Fig 2C). The homozygous phenotype was confirmed in the M3 generation. In the paddy field, all

M4 plants (n = 45) started heading 15 to 17 days earlier than the NPB control and showed broken flag leaves (Fig 2D). The whole-exome sequencing analysis of DNA extracted from an M3 plant identified 9 mutations: #11 to #19 in Table 2. Among them, the 5-bp DEL (mutation #15) in the first exon of the *PHYTOCHROME B* gene (*PHYB*; *Os03g0309200*) caused a frameshift and was present in the homozygote (Fig 6B and Table 2). We considered this mutation to be the most likely candidate causing the characteristic morphology and early heading date phenotype in line "786-5" because similar phenotypes have also been reported in a rice *phyB* mutant [38].

The early-heading-date mutant "IRB3517-3" was identified in the M2 population. All M3 plants (n = 45) grown in the patty field showed earlier heading than the NPB control (Fig 2D), indicating that the genotype in the M2 plant was homozygous. The whole-exome sequencing analysis of DNA extracted from an M3 plants revealed 13 mutations (#36 to #48), including 8 homozygous mutations (Table 2). The only high impact mutation identified was a 33.6-kb large deletion on chromosome 3 (mutation #38). The deleted region contained 5 genes, namely, *Os03g0793000*, *Os03g0793100*, *Os03g0793300*, *Os03g0793450*, and *Os03g0793500* (Fig 6C). Among the 5 genes, the *HEDAING DATE 16* (*HD16*; *Os03g0793500*) gene was the most likely candidate for the early heading phenotype in the "IRB3517-3" mutant. The *HD16* gene encodes a casein kinase I protein and is known to act as an inhibitor of rice flowering by phosphorylating and activating the GHD7 protein, one of the major players in the photoperiodic control of rice flowering [39]. It has been reported that a near-isogenic line with decreased kinase activity of the HD16 protein shows early flowering under a natural day length [39].

Another early-heading-date mutant line is “IRB3790-2”. DNA from one of the M3 plants obtained from an M2 plant that exhibited an early heading date (Fig 2D) was subjected to whole-exome sequencing analysis. However, this M2 plant was not a homozygote regarding the early-heading-date phenotype because plants with early and normal heading date phenotype were segregated in the M3 population derived from this M2 plant. We failed to establish a homozygous population even in the M4 generation and had to wait until M5 to obtain a homozygous population for the early heading date. The sequencing results indicated that there were 8 mutations (#49 to #56), including one high impact mutation (mutation #52), which was a 543-kb INV on chromosome 6 (Table 2). As demonstrated above, one of the break points of this INV was located in the second exon of the *HEADING DATE 1* (*HD1*; *Os06g0275000*) gene (Fig 6D), which encodes a zinc-finger type transcription factor and regulates the expression of HEADING DATE 3, a mobile flowering signal [40,41]. Although the position of the break point (R347) is near the C-terminal region of the encoded protein, the truncated HD1 protein generated by the INV lacks the CCT motif (amino acid position from 326 to 369), a conserved domain that is widely present in flowering-related proteins and essential for the protein function of HD1 [42], implying that the truncated HD1 protein produced in “IRB3790-2” was a loss-of-function protein. PCR analysis of selected 8 early and 12 non early-heading-date individuals (scored on August 10, 2018) in the M4 population using primers detecting #52 INV showed that all 8 early heading date plants possesses this INV in homozygous but not other plants, suggesting that this inversion was likely the cause of the early heading date. We did not determine why homozygous *hd1* plants were not obtained until the M4 generation, even though we chose a plant that exhibited the earliest heading date phenotype in the population to obtain seeds for the next generation.

Heterozygous plants (*HD1* / *hd1*) are thought to show an intermediate phenotype between wild type and *hd1* homozygous plants [43]. One possibility is that another mutated gene that affects the heading date or viability may be present in the genome of the selected plants for propagation. Further genetic analysis may reveal additional factors that contribute to a minor extent to heading date determination in this mutant line.

In summary, we have identified likely candidate genes responsible for the mutant phenotypes in 4 of the 5 mutant lines examined. Although further analyses, such as genetic complementation, RNA inactivation, and/or genome editing, are required to confirm the relationship between the candidate genes and mutants' phenotypes, this work showed that the combination of ion-beam mutagenesis and whole-exome sequencing enabled us to quickly narrow down probable candidate genes. In this work, we isolated mutants with remarkable phenotypes that could be very easily identified in a greenhouse or a paddy field. This finding could explain why the candidate genes were largely previously reported genes. Because the average number of high impact mutations per line was very low, it would be quite easy to narrow down candidate genes for the mutant phenotype, even if the responsible gene was an uncharacterized gene, by examining the genetic linkage between the identified mutation and the phenotype.

Although a recent fulfillment of genetic resources in rice enabled us to take a reverse genetic approach, such as the use of transposon- and T-DNA-tagged mutant libraries [44] as well as genome editing techniques [45], to understand the genetic bases of gene function and improve economically important traits in rice, a forward genetics approach with induced mutagenesis is still important to identify genes that cause uncharacterized mutant phenotypes or uncharacterized genes that cause mutant phenotypes. Induced mutagenesis can be applied not only in rice but also in other non-model plants that do

not fulfil genetic information and resources. It would be particularly useful to introduce new phenotypic variations into plants with a specific genetic background or to improve traits of already established varieties used in agricultural production. Unlike tagged mutations, such as transposon- or T-DNA tagging, induced mutagenesis has not been thought to be advantageous to isolate a casual gene because it requires large efforts such as positional cloning. However, recent "mapping by sequencing" techniques that combine mutagenesis techniques and massively parallel sequencing have been employed and promoted for the rapid detection of casual genes in mutants isolated by forward genetics [46]. For example, in the MutMap method [47], one of the mapping by sequencing techniques developed to identify rice genes, the selected mutants generated by EMS mutagenesis were back-crossed, and the F2 generation was obtained. DNA from F2 individuals who expressed mutant phenotype was pooled and subjected to NGS. Then, SNPs that were commonly present in mutant F2 DNA were detected by scoring the SNP index. This technique can secure an accurate evaluation of mutant phenotype because the work produces mutants and their parental lines in almost identical genetic backgrounds and can eliminate the generation of new polymorphic DNA markers for mapping because massively parallel sequencing can detect many SNPs in the F2 population generated by EMS, dramatically shortening the time required to clone a gene. As discussed earlier, ion beams induce a reduced number of mutations than EMS and more effectively cause InDels that often lead to complete inactivation of gene function. Therefore, it can be expected that a reduced number of candidate gene responsible for the mutation phenotype will be produced by ion beam irradiation compared with EMS. Indeed, in this work, we successfully identified candidate genes from 4 mutants in a very efficient manner by evaluating mutation catalogues generated by whole-exome

sequencing. Furthermore, in Arabidopsis, the combination of ion-beam mutagenesis, whole genome sequence, and rough mapping also successfully identified a casual gene of a variegated leaf mutant [48]. Therefore, characterization of ion beam-induced mutants by whole-exome or whole-genome sequencing will enable promotion of the effective isolation of causal genes for mutants. Further efforts to isolate useful mutants by ion-beam irradiation and to detect useful genes from mutants by genome sequencing will facilitate improvement of agronomically important traits in rice as well as in other plants.

Conclusions

The properties of ion beam-induced mutations in rice were revealed by whole-exome sequencing. The average mutation frequency in the M1 generation of ion beam-irradiated rice was estimated to be 2.6×10^{-7} per base. The results suggest that carbon ion irradiation seems to induce a reduced number of mutations in rice DNA than EMS- and tissue culture-based mutagenesis and a similar number to FN mutagenesis. The identification of a small number of high-impact mutation would facilitate detection of candidate genes responsible for the phenotypes of mutants.

References

1. Tanaka A, Shikazono N, Hase Y. Studies on biological effects of ion beams on lethality , molecular nature of mutation , mutation rate , and spectrum of mutation

phenotype for mutation breeding in higher plants. J Radiat Res. 2010;51: 223–233. doi:10.1269/jrr.09143

2. Abe T, Ryuto H, Fukunishi N. Ion beam radiation mutagenesis. In: Shu, Q. Y., Forster, B. P., Nakagawa H, editor. Plant mutation breeding and biotechnology. Oxfordshire UK: CAB International; 2012. pp. 99–106. doi:10.1079/9781780640853.0099

3. Yamaguchi H. Mutation breeding of ornamental plants using ion beams. Breed Sci. 2018/02/17. Japanese Society of Breeding; 2018;68: 71–78. doi:10.1270/jsbbs.17086

4. Okamura M, Yasuno N, Ohtsuka M, Tanaka A, Shikazono N, Hase Y. Wide variety of flower-color and -shape mutants regenerated from leaf cultures irradiated with ion beams. Nucl Instruments Methods Phys Res Sect B Beam Interact with Mater Atoms. 2003;206: 574–578. doi:https://doi.org/10.1016/S0168-583X(03)00835-8

5. Yamaguchi H. Characteristics of Ion Beams as Mutagens for Mutation Breeding in Rice and Chrysanthemums. Japan Agric Res Q JARQ. 2013;47: 339–346. doi:10.6090/jarq.47.339

6. Shikazono N, Yokota Y, Kitamura S, Suzuki C, Watanabe H, Tano S, et al. Mutation Rate and Novel tt Mutants of Arabidopsis thaliana Induced by Carbon Ions. Genetics. 2003;163: 1449–1455. Available: http://www.genetics.org/content/163/4/1449.abstract

7. Shikazono N, Suzuki C, Kitamura S, Watanabe H, Tano S, Tanaka A. Analysis of mutations induced by carbon ions in Arabidopsis thaliana. J Exp Bot. 2005;56: 587–596. doi:10.1093/jxb/eri047

8. Kazama Y, Saito H, Yamamoto YY, Hayashi Y, Ichida H, Ryuto H, et al. LET-dependent effects of heavy-ion beam irradiation in Arabidopsis thaliana. Plant Biotechnol. 2008;117: 113–117.

9. Kazama Y, Hirano T, Saito H, Liu Y, Ohbu S, Hayashi Y, et al. Characterization of highly efficient heavy-ion mutagenesis in Arabidopsis thaliana. BMC Plant Biol. 2011;11: 161. doi:10.1186/1471-2229-11-161

10. Naito K, Kusaba M, Shikazono N, Takano T, Tanaka A, Tanisaka T, et al. Transmissible and Nontransmissible Mutations Induced by Irradiating Arabidopsis thaliana Pollen With γ-Rays and Carbon Ions. Genetics. 2005;169: 881–889. doi:10.1534/genetics.104.033654

11. Hase Y, Yoshihara R, Nozawa S, Narumi I. Mutagenic effects of carbon ions near the range end in plants. Mutat Res - Fundam Mol Mech Mutagen. 2012;731: 41–47. doi:10.1016/j.mrfmmm.2011.10.004

12. Kazama Y, Ma L, Hirano T, Ohbu S, Shirakawa Y, Hatakeyama S, et al. Rapid evaluation of effective linear energy transfer in heavy-ion mutagenesis of *Arabidopsis thaliana*. Plant Biotechnol. 2012;29: 441–445. doi:10.5511/plantbiotechnology.12.0921a

13. Yoshihara R, Hase Y, Sato R, Takimoto K, Narumi I. Mutational effects of different LET radiations in *rpsL* transgenic *Arabidopsis*. Int J Radiat Biol. 2010;86: 125–131. doi:10.3109/09553000903336826

14. Yoshihara R, Nozawa S, Hase Y, Narumi I, Hidema J, Sakamoto AN. Mutational effects of γ -rays and carbon ion beams on Arabidopsis seedlings. J Radiat Res. 2013;54: 1050–1056. doi:10.1093/jrr/rrt074

15. Hirano T, Kazama Y, Ohbu S, Shirakawa Y, Liu Y, Kambara T, et al. Molecular nature of mutations induced by high-LET irradiation with argon and carbon ions in Arabidopsis thaliana. Mutat Res Mol Mech Mutagen. 2012;735: 19–31. doi:https://doi.org/10.1016/j.mrfmmm.2012.04.010

16. Hase Y, Nozawa S, Narumi I, Oono Y. Effects of ion beam irradiation on size of mutant sector and genetic damage in Arabidopsis. Nucl Instruments Methods Phys Res Sect B Beam Interact with Mater Atoms. Elsevier B.V.; 2017;391: 14–19. doi:10.1016/j.nimb.2016.11.023

17. Ossowski S, Schneeberger K, Lucas-lledó JI, Warthmann N, Clark RM, Shaw RG, et al. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science. 2010;327: 92–94. doi:10.1126/science.1180677

18. Du Y, Luo S, Li X, Yang J, Cui T, Li W, et al. Identification of Substitutions and Small Insertion-Deletions Induced by Carbon-Ion Beam Irradiation in Arabidopsis thaliana. Front Plant Sci. 2017;8. doi:10.3389/fpls.2017.01851

19. Kazama Y, Ishii K, Hirano T, Wakana T, Yamada M, Ohbu S, et al. Different mutational function of low- and high-linear energy transfer heavy-ion irradiation demonstrated by whole-genome resequencing of Arabidopsis mutants. Plant J. 2017;92: 1020–1030. doi:10.1111/tpj.13738

20. Hase Y, Satoh K, Kitamura S, Oono Y. Physiological status of plant tissue affects the frequency and types of mutations induced by carbon-ion irradiation in Arabidopsis. Sci Rep. 2018;8: 1394. doi:10.1038/s41598-018-19278-1

21. Yamaguchi H, Hase Y, Tanaka A, Shikazono N, Degi K, Shimizu A, et al. Mutagenic effects of ion beam irradiation on rice. Breed Sci. 2009;59: 169–177.

22. Ishikawa S, Ishimaru Y, Igura M, Kuramata M, Abe T, Senoura T, et al. Ion-beam irradiation, gene identification, and marker-assisted breeding in the development of low-cadmium rice. Proc Natl Acad Sci U S A. 2012;109: 19166–19171. doi:10.1073/pnas.1211132109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1211132109

23. Mahadtanapuk S, Teraarusiri W, Phanchaisri B, Yu LD, Anuntalabhochai S. Breeding for blast-disease-resistant and high-yield Thai jasmine rice (Oryza sativa L. cv. KDML 105) mutants using low-energy ion beams. Nucl Instruments Methods Phys Res Sect B Beam Interact with Mater Atoms. 2013;307: 229–234. doi:https://doi.org/10.1016/j.nimb.2013.01.088

24. Ichitani K, Yamaguchi D, Taura S, Fukutoku Y, Onoue M, Shimizu K, et al. Genetic analysis of ion-beam induced extremely late heading mutants in rice. Breed Sci. 2014;64: 222–230. doi:10.1270/jsbbs.64.222

25. Morita R, Nakagawa M, Takehisa H, Hayashi Y, Ichida H, Usuda S, et al. Heavy-ion beam mutagenesis identified an essential gene for chloroplast development under cold stress conditions during both early growth and tillering stages in rice. Biosci Biotechnol Biochem. Taylor & Francis; 2017;81: 271–282. doi:10.1080/09168451.2016.1249452

26. Ichida H, Morita R, Shirakawa Y, Hayashi Y, Abe T. Targeted exome sequencing of unselected heavy-ion beam-irradiated populations reveals less-biased mutation characteristics in the rice genome. Plant J. 2019;98: 301–314. doi:doi:10.1111/tpj.14213

27. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms,

SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012/04/01. Landes Bioscience; 2012;6: 80–92. doi:10.4161/fly.19695

28. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. Nature. 2005;436: 793–800. Available: http://dx.doi.org/10.1038/nature03895

29. Kodym A, Afza R, Forster BP, Ukai Y, Nakagawa H, Mba C. Methodology for physical and chemical mutagenic treatments. In: Shu QY, Foster BP, Nakagawa H, editors. Plant Mutation Breeding and Biotechnology. Oxfordshire: CAB International and FAO; 2012. pp. 169–180. Available: www.cabi.org

30. Henry IM, Nagalakshmi U, Lieberman MC, Ngo KJ, Krasileva K V, Vasquez-Gross H, et al. Efficient Genome-Wide Detection and Cataloging of EMS-Induced Mutations Using Exome Capture and Next-Generation Sequencing. Plant Cell. 2014;26: 1382–1397. doi:10.1105/tpc.113.121590

31. Miyao A, Nakagome M, Ohnuma T, Yamagata H, Kanamori H, Katayose Y, et al. Molecular Spectrum of Somaclonal Variation in Regenerated Rice Revealed by Whole-Genome Sequencing. Plant Cell Physiol. 2012;53: 256–264. Available: http://dx.doi.org/10.1093/pcp/pcr172

32. Zhang D, Wang Z, Wang N, Gao Y, Liu Y, Wu Y, et al. Tissue Culture-Induced Heritable Genomic Variation in Rice, and Their Phenotypic Implications. PLoS One. Public Library of Science; 2014;9: e96879. Available: https://doi.org/10.1371/journal.pone.0096879

33. Li G, Jain R, Chern M, Pham NT, Martin JA, Wei T, et al. The Sequences of 1504 Mutants in the Model Rice Variety Kitaake Facilitate Rapid Functional Genomic Studies. Plant Cell. 2017;29: 1218–1231. doi:10.1105/tpc.17.00154

34. Hase Y, Satoh K, Kitamura S, Oono Y. Physiological status of plant tissue affects the frequency and types of mutations induced by carbon-ion irradiation in Arabidopsis. Sci Rep. 2018;8. doi:10.1038/s41598-018-19278-1

35. Fujisawa Y, Kato T, Ohki S, Ishikawa A, Kitano H, Sasaki T, et al. Suppression of the heterotrimeric G protein causes abnormal morphology, including dwarfism, in rice. Proc Natl Acad Sci U S A. 1999;96: 7575–80. doi:10.1073/pnas.96.13.7575

36. Ashikari M, Wu J, Yano M, Sasaki T, Yoshimura A. Rice gibberellin-insensitive dwarf mutant gene Dwarf 1 encodes the α-subunit of GTP-binding protein. Proc Natl Acad Sci. 1999;96: 10284–10289. doi:10.1073/pnas.96.18.10284

37. Chen J, Zhao H, Zheng X, Liang K, Guo Y, Sun X. Recent amplification of Osr4 LTR-retrotransposon caused rice D1 gene mutation and dwarf phenotype. Plant Divers. Elsevier Ltd; 2017;39: 73–79. doi:10.1016/j.pld.2017.01.003

38. Takano M, Inagaki N, Xie X, Yuzurihara N, Hihara F, Ishizuka T, et al. Distinct and cooperative functions of phytochromes A, B, and C in the control of deetiolation and flowering in rice. Plant Cell. 2005;17: 3311–3325. doi:DOI 10.1105/tpc.105.035899

39. Hori K, Ogiso-Tanaka E, Matsubara K, Yamanouchi U, Ebana K, Yano M. Hd16, a gene for casein kinase I, is involved in the control of rice flowering time

by modulating the day-length response. Plant J. 2013/07/25. BlackWell Publishing Ltd; 2013;76: 36–46. doi:10.1111/tpj.12268

40. Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, et al. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell. American Society of Plant Biologists; 2000;12: 2473–2484. Available: https://www.ncbi.nlm.nih.gov/pubmed/11148291

41. Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, et al. Hd3a, a Rice Ortholog of the Arabidopsis FT Gene, Promotes Transition to Flowering Downstream of Hd1 under Short-Day Conditions. Plant Cell Physiol. 2002;43: 1096–1105. Available: http://dx.doi.org/10.1093/pcp/pcf156

42. Takahashi Y, Teshima KM, Yokoi S, Innan H, Shimamoto K. Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to diversity of flowering time in cultivated rice. P Natl Acad Sci USA. 2009;106: 4555–4560. doi:10.1073/pnas.0812092106

43. Yamamoto T, Kuboki Y, Lin SY, Sasaki T, Yano M. Fine mapping of quantitative trait loci Hd-1, Hd-2 and Hd-3, controlling heading date of rice, as single Mendelian factors. Theor Appl Genet. 1998;97: 37–44. doi:10.1007/s001220050864

44. Wang N, Long T, Yao W, Xiong L, Zhang Q, Wu C. Mutant Resources for the Functional Analysis of the Rice Genome. Mol Plant. 2013;6: 596–604. doi:https://doi.org/10.1093/mp/sss142

45. Mishra R, Joshi RK, Zhao K. Genome Editing in Rice: Recent Advances, Challenges, and Future Implications. Front Plant Sci. Frontiers Media S.A.; 2018;9: 1361. doi:10.3389/fpls.2018.01361

46. Schneeberger K. Using next-generation sequencing to isolate mutant genes from forward genetic screens. Nat Rev Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014;15: 662. Available: https://doi.org/10.1038/nrg3745

47. Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, et al. Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol. Nature Publishing Group; 2012;30: 174–178. doi:10.1038/nbt.2095

48. Du Y, Luo S, Yu L, Cui T, Chen X, Yang J, et al. Strategies for identification of mutations induced by carbon-ion beam irradiation in Arabidopsis thaliana by whole genome re-sequencing. Mutat Res Mol Mech Mutagen. 2018;807: 21–30. doi:https://doi.org/10.1016/j.mrfmmm.2017.12.001

Acknowledgements

The authors thank Shoya Hirata, Eenzen Mungunchudur, Toshihiko Sanzen, Hiroki Arai, Shogo Ozawa, Satoshi Kitamura, and members of Research Planning Office, Quantum Beam Research Directorate, QST for their help in harvesting rice seeds and characterizing rice mutants. The bioinformatics analysis was performed using the HOKUSAI supercomputing system, operated by the Information Systems Division, RIKEN, under the project number Q18208.

This work was supported by a grant from the Cabinet Office, under the “Technologies for creating next-generation agriculture, forestry and fisheries” project in the Cross-

ministerial Strategic Innovation Promotion Program (SIP). The grant was administrated by the Bio-oriented Technology Research Advancement Institution, NARO.

Author contributions

Y.O and Y.H designed the experiment. S.N. and Y.H. conducted ion-beam irradiation, mutant screening, and characterization of mutants in the green house. Y.O., S.N., K.S., A.S., H.K., and Y.H. performed mutant screening and characterization of mutants in the paddy field. Whole exome analysis was done by H.I., R.M., and T.A.. Y.O. and H.I. analyzed the data and Y.O. wrote the manuscript with input from H.I. and Y.H. All authors approved the final manuscript.

Supporting Information

S1 Table. List of primer sets.

S2 Table. Detailed list of mutations detected.

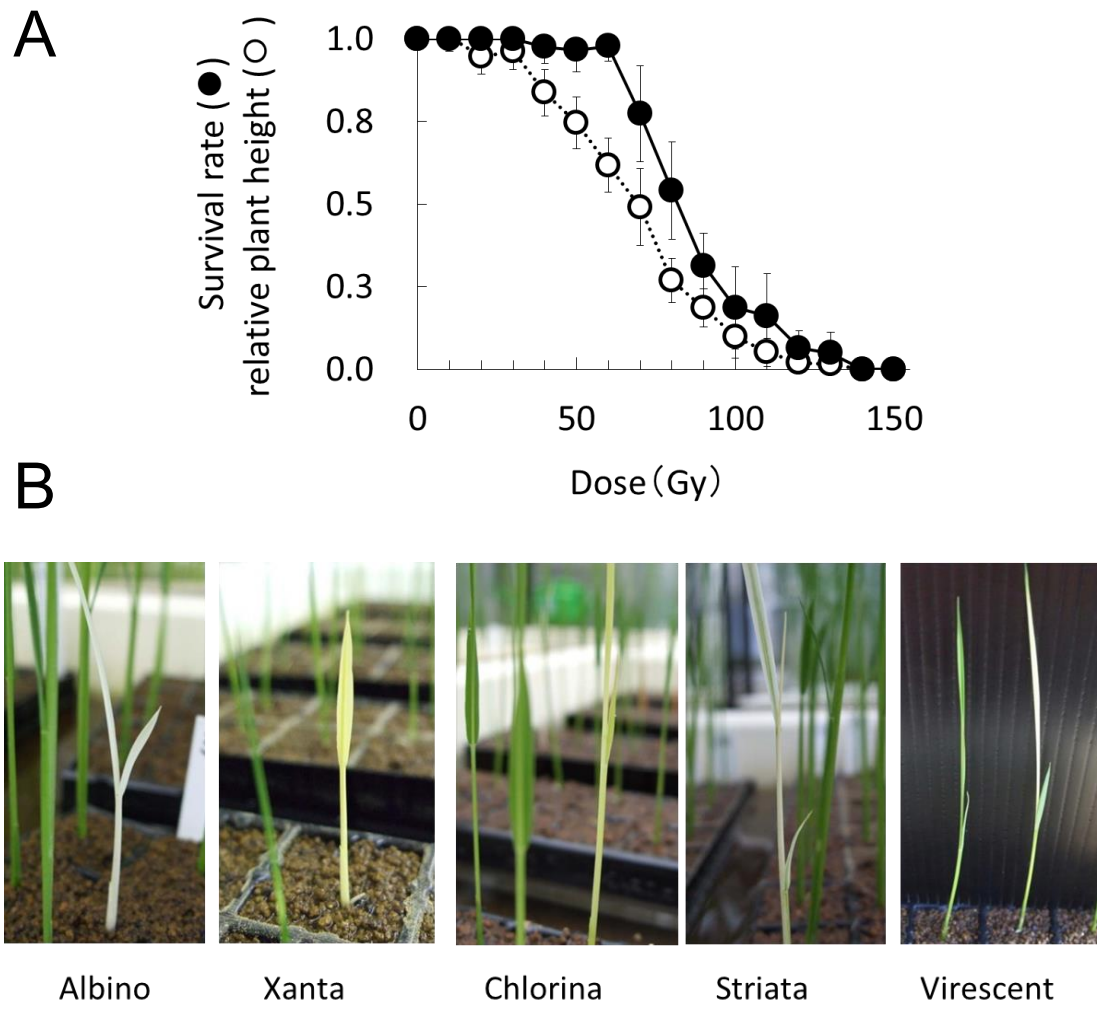
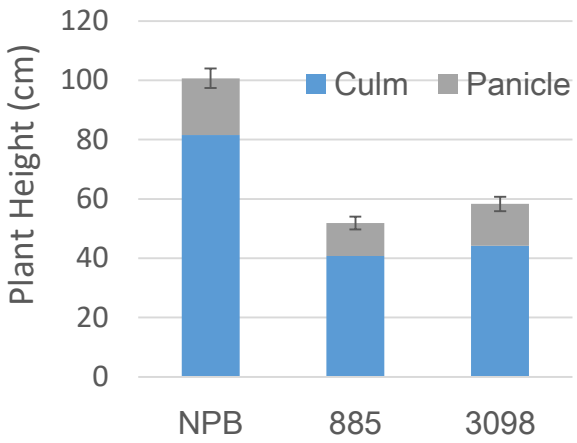

A
Survival rate (●)
relative plant height (○)
1.0
0.8
0.5
0.3
0.0
0
50
100
150
Dose（Gy）
B
Albino
Xanta
Chlorina
Striata
Virescent

Figure 1

A



B

C

D

Line	Heading date*	Days earlier than control	Additional phenotype
NPB	Aug. 14	-	
786-5	Jul. 28	17	wide leaves broken flag leaves
IRB3517-3	Aug. 3	11	
IRB3790-2**	July 28	17	

*Heading date was recorded in the paddy field in 2016.
**Phenotype of early heading date was segregated in the population.

Figure 2

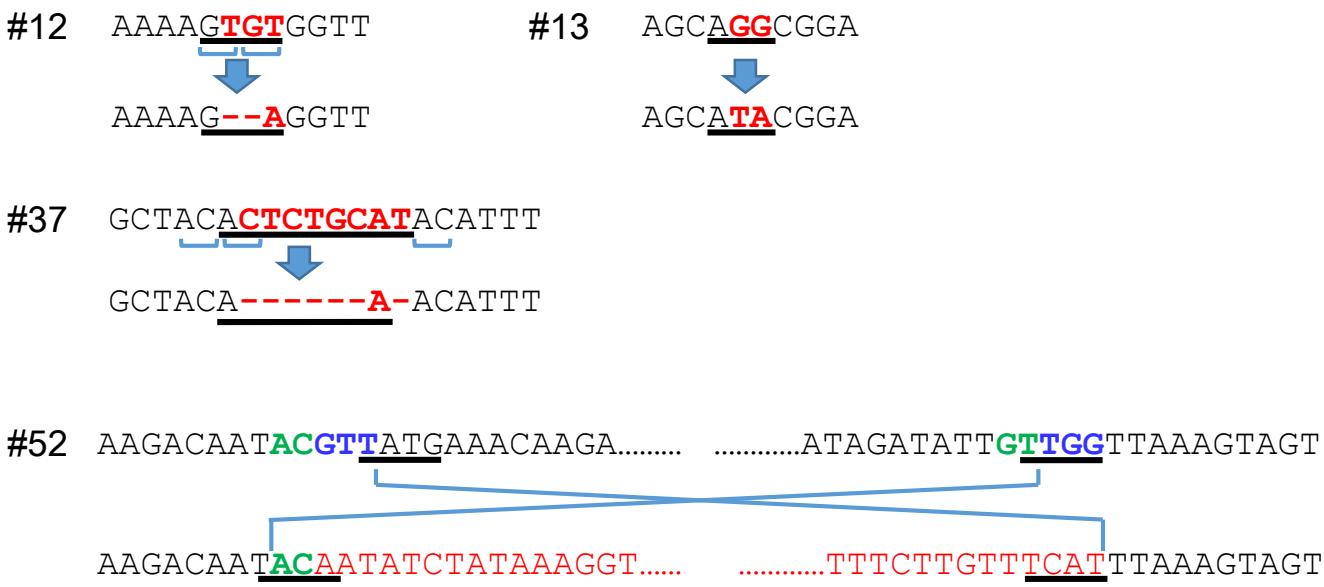

#12
AAAAGTGTGGTT
AAAAG--AGGTT
#13
AGCAGGCGGA
AGCATACGGA
#37
GCTACACTCTGCATACATTT
GCTACA------A-ACATTT
#52
AAGACAATACGTTATGAAACAAGA.....
.....ATAGATATTGTTGGTTAAAGTAGT
AAGACAATACAATATCTATAAAGGT.....
.....TTTCTTGTTTCATTTAAAGTAGT

Figure 3

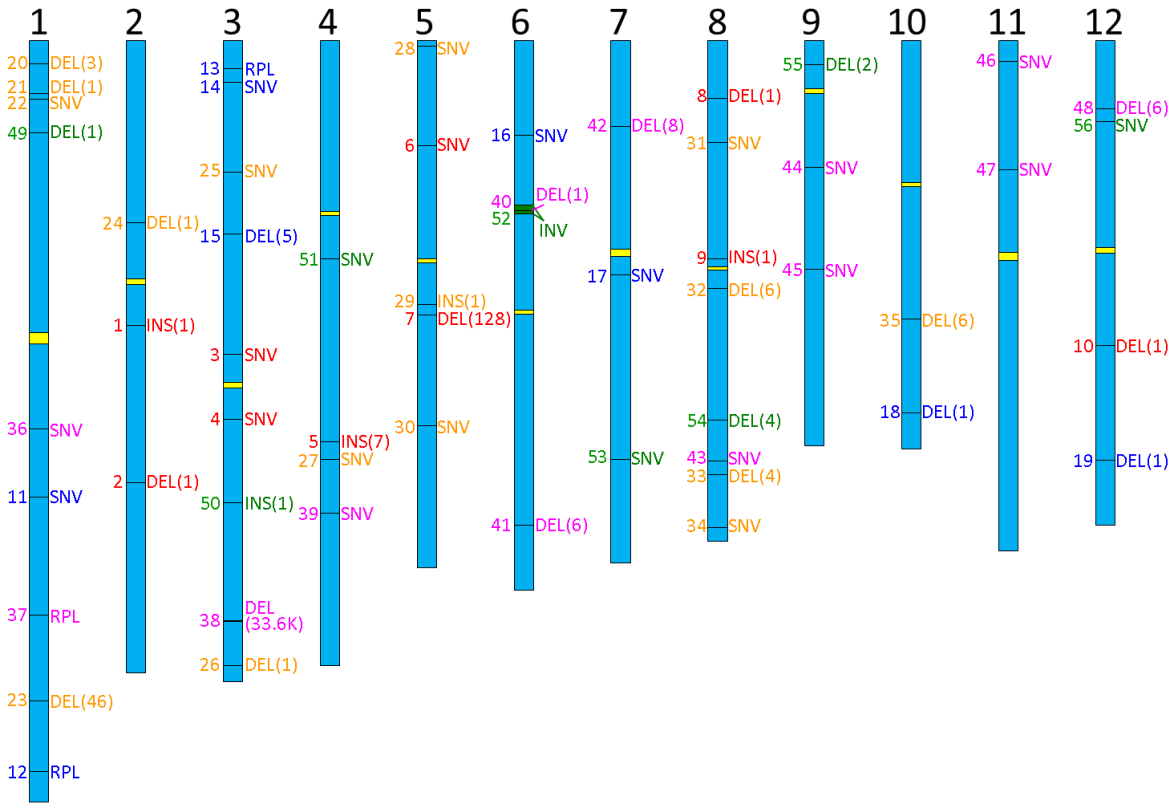

1
20 DEL(3)
21 DEL(1)
22 SNV
49 DEL(1)
36 SNV
11 SNV
37 RPL
23 DEL(46)
12 RPL
2
24 DEL(1)
1 INS(1)
2 DEL(1)
3
13 RPL
14 SNV
25 SNV
15 DEL(5)
3 SNV
4 SNV
50 INS(1)
38 DEL (33.6K)
26 DEL(1)
4
51 SNV
5 INS(7)
27 SNV
39 SNV
5
28 SNV
6 SNV
29 INS(1)
7 DEL(128)
30 SNV
6
16 SNV
40 DEL(1)
52 INV
41 DEL(6)
7
42 DEL(8)
17 SNV
53 SNV
8
8 DEL(1)
31 SNV
9 INS(1)
32 DEL(6)
54 DEL(4)
43 SNV
33 DEL(4)
34 SNV
9
55 DEL(2)
44 SNV
45 SNV
10
35 DEL(6)
18 DEL(1)
11
46 SNV
47 SNV
12
48 DEL(6)
56 SNV
10 DEL(1)
19 DEL(1)

Figure 4

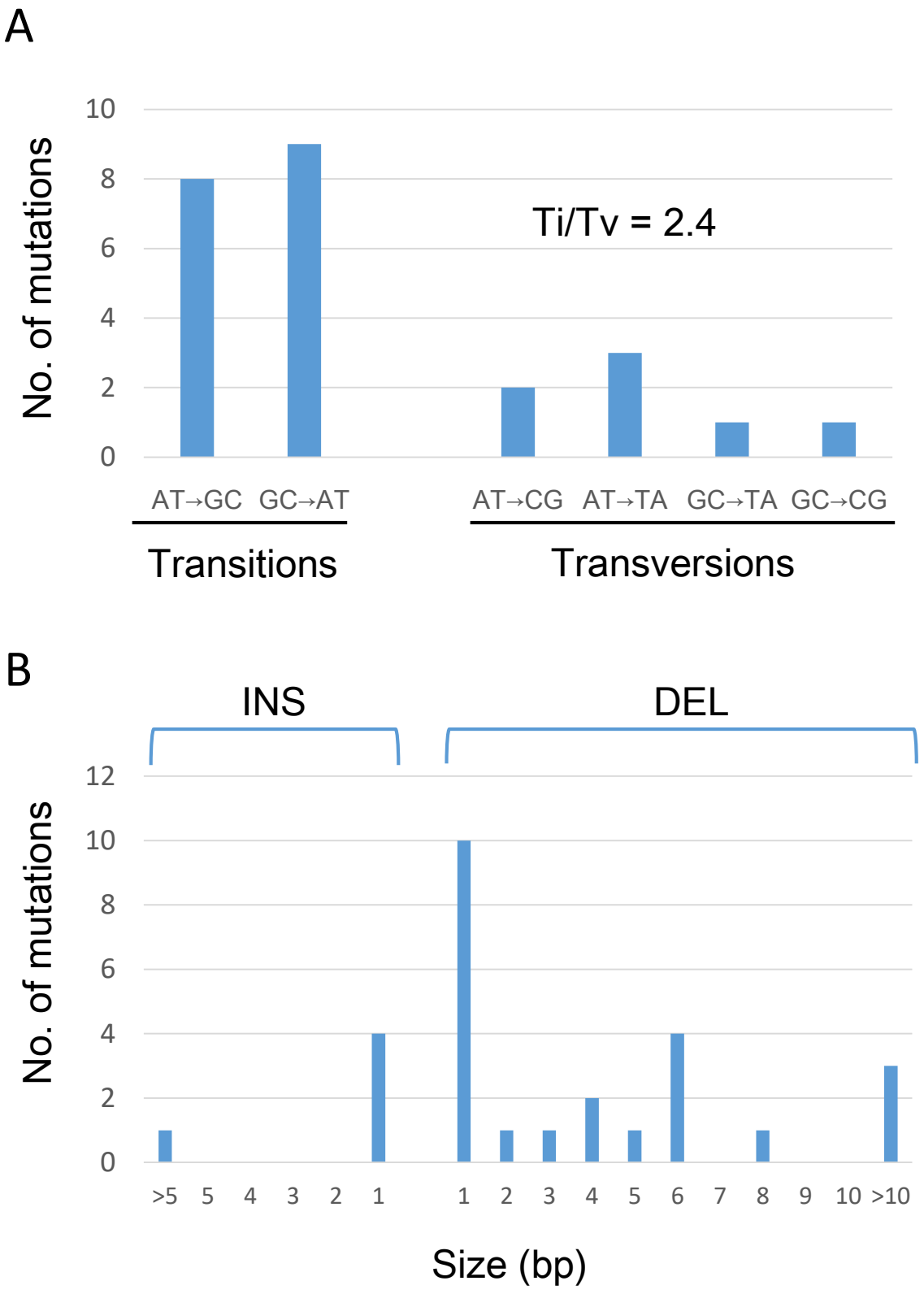

A
No. of mutations
10
8
6
4
2
0
Ti/Tv = 2.4
AT→GC
GC→AT
AT→CG
AT→TA
GC→TA
GC→CG
Transitions
Transversions
B
INS
DEL
No. of mutations
12
10
8
6
4
2
0
>5
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
>10
Size (bp)

Figure 5

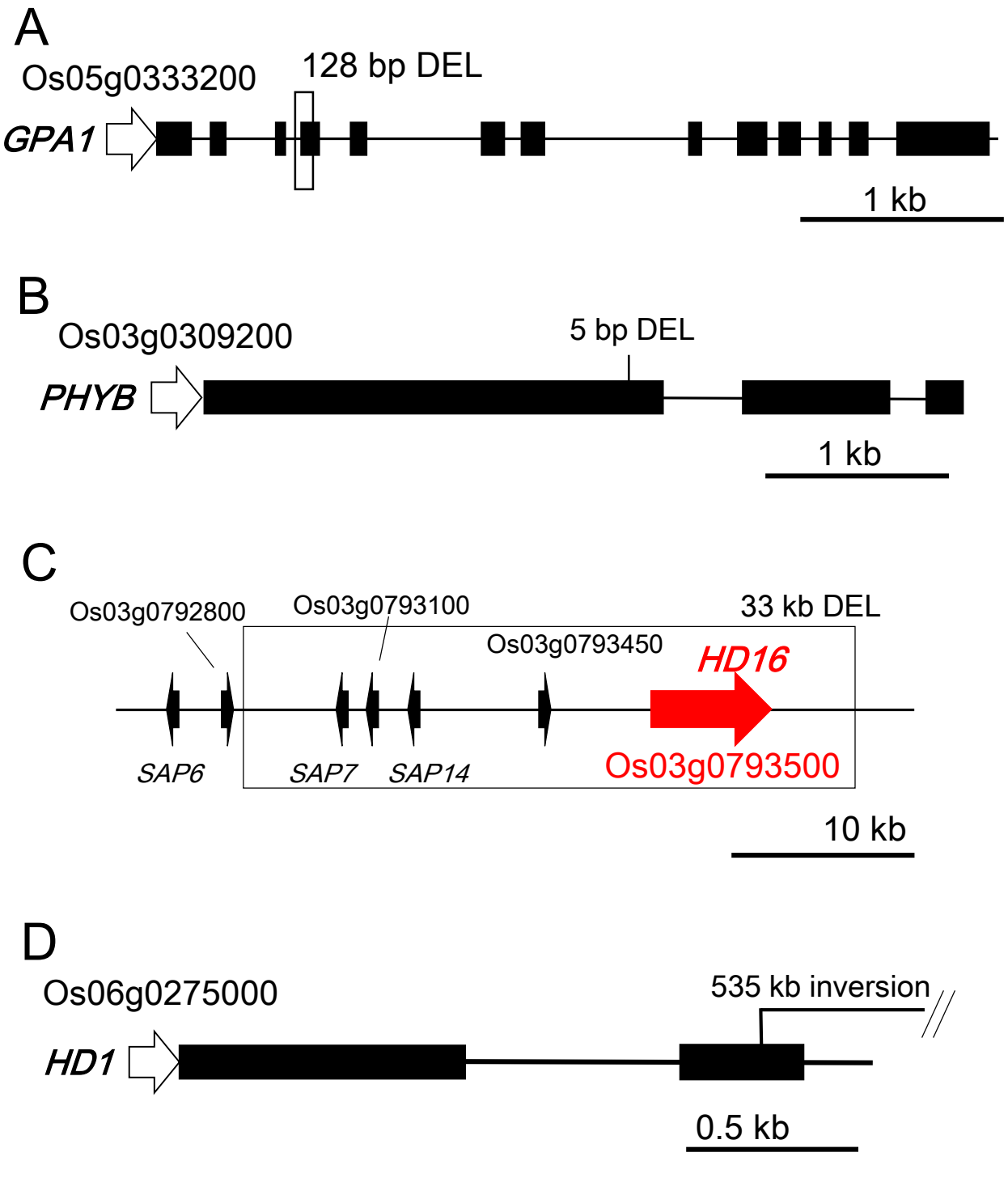

A
Os05g0333200
128 bp DEL
GPA1
1 kb
B
Os03g0309200
5 bp DEL
PHYB
1 kb
C
Os03g0792800
Os03g0793100
33 kb DEL
Os03g0793450
HD16
SAP6
SAP7
SAP14
Os03g0793500
10 kb
D
Os06g0275000
535 kb inversion
HD1
0.5 kb
NPB
AAGTTTGAGAAGACAATACGTTATGAAACA
K F E K T I R Y E T
347
IRB3790-2
AAGTTTGAGAAGACAATACAATATCTATAA
K F E K T I Q Y L *

Figure 6

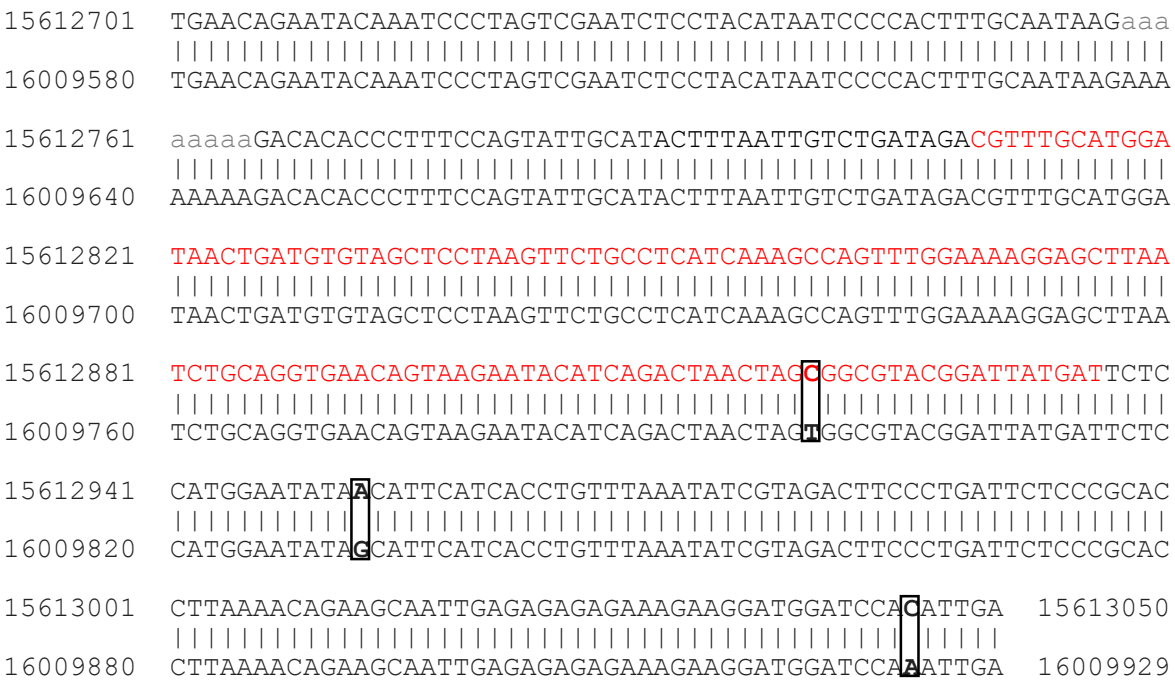

15612701 TGAACAGAATACAAATCCCTAGTCGAATCTCCTACATAATCCCCACTTTGCAATAAGaaa
16009580 TGAACAGAATACAAATCCCTAGTCGAATCTCCTACATAATCCCCACTTTGCAATAAGAAA
15612761 aaaaaGACACACCCTTTCCAGTATTGCATACTTTAATTGTCTGATAGACGTTTGCATGGA
16009640 AAAAAGACACACCCTTTCCAGTATTGCATACTTTAATTGTCTGATAGACGTTTGCATGGA
15612821 TAACTGATGTGTAGCTCCTAAGTTCTGCCTCATCAAAGCCAGTTTGGAAAAGGAGCTTAA
16009700 TAACTGATGTGTAGCTCCTAAGTTCTGCCTCATCAAAGCCAGTTTGGAAAAGGAGCTTAA
15612881 TCTGCAGGTGAACAGTAAGAATACATCAGACTAACTAGCGGCGTACGGATTATGATTCTC
16009760 TCTGCAGGTGAACAGTAAGAATACATCAGACTAACTAGTGGCGTACGGATTATGATTCTC
15612941 CATGGAATATAACATTCATCACCTGTTTAAATATCGTAGACTTCCCTGATTCTCCCGCAC
16009820 CATGGAATATAGCATTCATCACCTGTTTAAATATCGTAGACTTCCCTGATTCTCCCGCAC
15613001 CTTAAAACAGAAGCAATTGAGAGAGAGAAAGAAGGATGGATCCACATTGA 15613050
16009880 CTTAAAACAGAAGCAATTGAGAGAGAGAAAGAAGGATGGATCCAAATTGA 16009929

Figure 7

Table S1 List of primer sets

Mutation No.[1]	Line	Target CHR No.	Start	End	Reference[2]	Alternatives[2]	Type of mutation[3]	5' position of primer [4]	Sequence (5' to 3') of primer
5	3098	chr04	2.3E+07	22824954	T	TAGATAGA	INS (7)	chr4 22824682 F	GGCAAACATTACATATAGCATAAAGG
								chr4 22825181 R	GAACTCGCCGACCTCACC
7	3098	chr05	1.6E+07	15612936	ACGT….TGA	A	DEL (128)	chr5 15612143 F	GAACAAGGTAAATGCACAAAGGTATA
								chr5 15613069 R	TTTTAGTTTGAACTTGTTCTCAATG
12	786	chr01	4.2E+07	41549064	GTGT	GA	RPL	chr1 41548923 F	CTGCGCTCTTCAACCGTTC
								chr1 41549206 R	CGGCGTGTTATTGCTCAGT
13	786	chr03	1631231	1631232	GG	TA	RPL	chr3 1630969 F	CACTGCTCGGAGAACACCTT
								chr3 1631462 R	TGGCTGGCTACGTGTGTAGA
15	786	chr03	1.1E+07	11022260	AATGGC	A	DEL (5)	chr3 11022016 F	TCCGGTCACACACAGCTAAG
								chr3 11022491 R	TGTTTCCTCAGATTCCTTGAAGA
18	786	chr10	2.1E+07	21181034	CG	C	DEL (1)	chr10 21180767 F	GCTGTCCATCAGGTTCCTCT
								chr10 21181262 R	ATTCGATGGATGGCTAGCTG
22	885	chr01	3357475	3357475	C	T	SNV	chr1 3357244 F	GCGCAACATGAAAAAGAAAA
								chr1 3357725 R	CCAGCGCAGTAGGAAATGAT
23	885	chr01	3.8E+07	37524065	GTGT….CGC	G	DEL (46)	chr1 37523767 F	GCGCATGTTTGTGAGAGAAG
								chr1 37524250 R	CGAAGCGAGCAATCTCAAGT
26	885	chr03	3.6E+07	35501011	TC	T	DEL (1)	chr3 35500787 F	TGTTATCAGGGGCCTATTGAA
								chr3 35501282 R	CATGACAGATGATTGGAACTTTG
27	885	chr04	2.4E+07	23792560	G	A	SNV	chr4 23792336 F	TTTCTGATCTCGCCTTGGTT
								chr4 23792789 R	ACAGGAGCAAGAACCGGAAG
30	885	chr05	2.2E+07	21895103	C	T	SNV	chr5 21894866 F	CCAGTGGGATTTCCTCTGAA
								chr5 21895323 R	CAAAAAGCAAATGTCCACGA
32	885	chr08	1.4E+07	14099489	CATCGAA	C	DEL (6)	chr8 14099248 F	GTAAGCGCATGGAAATACCG
								chr8 14099767 R	GGACACGGCTGAGAGGAG
33	885	chr08	2.5E+07	24691425	TGAAA	T	DEL (4)	chr8 24691196 F	GCTAAATTTTAAGCACCATGTGA
								chr8 24691653 R	CTGCTCCAACTCTGCATCAA
34	885	chr08	2.8E+07	27686683	T	C	SNV	chr8 27686445 F	TTGCCTTCTAGGCCTCACTC
								chr8 27686938 R	TGCTTTGACAGGAGATGCAG
35	885	chr10	1.6E+07	15810767	TCGCGCC	T	DEL (6)	chr10 15810518 F	CACCGGAACGACCTCACC
								chr10 15810983 R	CTCGCAATACCATGGAGGAC
37	IRB3517-3	chr01	3.3E+07	32645636	ACTCTGCAT	AA	RPL	chr1 32645404 F	TCTGACAAAAATAGCCCCAAA
								chr1 32645884 R	CCATGTTGAAAGGAAACCAGA
38	IRB3517-3	chr03	3.3E+07	33010771	TGAG….GTT	T	DEL (33.6 k)	chr3 32976920 F	TGGCGTCGGGGTTTGTAT
								chr3 33011012 R	AACACCACCACAAAGGAAGC
41	IRB3517-3	chr06	2.8E+07	27521464	TCCTGTC	T	DEL (6)	chr6 27521228 F	CCAGGTCACAACCATGTCAA
								chr6 27521682 R	GCTGGTTTTGAAGCTGCATT
42	IRB3517-3	chr07	4881911	4881919	GCGTATATT	G	DEL (8)	chr7 4881645 F	CACGACGCCGAGAAGTCTAT
								chr7 4882141 R	AAGCCCCTACAAAAGCCAAG
44	IRB3517-3	chr09	7224224	7224224	A	G	SNV	chr9 7223954 F	CCGAGGAAAAGTGCTGAAAG
								chr9 7224446 R	GGGATGGAAGTTTGTTTTCG
45	IRB3517-3	chr09	1.3E+07	13007477	G	A	SNV	chr9 13007214 F	CCTGACCACCAAAAGAAAGG
								chr9 13007708 R	ACGAGGACCTGCTGACAGA
47	IRB3517-3	chr11	7365282	7365282	T	C	SNV	chr11 7365028 F	ACCCATCAGCTGGTCACAAT
								chr11 7365527 R	TTTGCTGTTTGTATGCTTCATTT
52	IRB3790-2	chr06	9338213	9872884	TATG….TTGC	TACA….TCAT	INV (535 k) [5]	chr6 9338078 F	CCTGCTGGAGCAATCAATCT
								chr6 9338377 R	AGCGTCTCATGAGTCCCATC
								chr6 9872652 F	CACTCATTAACAGAAGGGACTGAA
								chr6 9872951 R	GTACTCCCTCCGTCCGAAA
54	IRB3790-2	chr08	2.2E+07	21594853	CTTAA	C	DEL (4)	chr8 21594571 F	CCAGTTTCATCCCAAAAATGA
								chr8 21595070 R	TTTGTGTGTGACATATGAACGA
56	IRB3790-2	chr12	4637905	4637905	G	A	SNV	chr12 4637679 F	ACAAAAGCTGCAGCAGGAGA
								chr12 4638153 R	GTCGCTCCAAGGGTGGAA

[1] Mutation No. is identical to the one in Table 2.

[2] For the large deletions and the inversion, only first and last 4 bases are shown.

[3] Size of InDel is indicated in parentheses.

[4] "F" and "R" indicate orientation of the primers, forward and reverse, respectively.

[5] PCR signal can be detected with the PCR primer sets chr6 9338078 F-chr6 9872652 F and chr6 9338377 R-chr6 9872951R for the INV mutation.

Table S2. Detailed list of mutations detected

Mutation No.[1]	Line	CHR No.	Start	End	Reference[2]	Alternatives[2]	Type after evaluation[3]	Effect[4]	Gene	Annotation	Genotype[5]	Verification[6]
1	3098	chr02	16190324	16190325	C	CA	INS (1)	intergenic_region			0|1	
2	3098	chr02	25123707	25123708	CA	C	DEL (1)	downstream_gene_variant			0|1	
3	3098	chr03	17837150	17837150	C	G	SNV	splice_region_variant&intron_variant	Os03g0426900	[CLPB-C] Similar to APG6/CLPB-P/CLPB3 (ALBINO AND PALE GREEN 6); ATP binding / ATPase.	0|1	
4	3098	chr03	21504174	21504174	T	C	SNV	missense_variant	Os03g0583900	[DCL2A] Similar to Endoribonuclease Dicer homolog 2a.	0|1	
5	3098	chr04	22824947	22824954	T	TAGATAGA	INS (7)	upstream_gene_variant			1|1	✓
6	3098	chr05	5977789	5977789	C	T	SNV	missense_variant	Os05g0196800	Similar to Diacylglycerol acylCoA acyltransferase.	0|1	
7	3098	chr05	15612808	15612936	ACGT....TGAT	A	DEL (128)	frameshift_variant&splice_acceptor_variant&splice_region_variant&splice_region_variant&intron_variant	Os05g0333200	[D1] Guanine nucleotide-binding protein alpha-1 subunit (GP-alpha-1).	0|1	✓
8	3098	chr08	3299965	3299966	TG	T	DEL (1)	downstream_gene_variant			0|1	
9	3098	chr08	12412765	12412766	A	AT	INS (1)	downstream_gene_variant			0|1	
10	3098	chr12	17352209	17352210	TG	T	DEL (1)	intergenic_region			0|1	
11	786	chr01	25949685	25949685	T	A	SNV	missense_variant	Os01g0644600	Glutelin family protein.	0|1	
12[7]	786	chr01	41549062	41549064	GTGT	GA	RPL	upstream_gene_variant			1|1	✓
13[7]	786	chr03	1631231	1631232	GG	TA	RPL	missense_variant	Os03g0128800	Similar to predicted protein.	1|1	✓
14	786	chr03	2403133	2403133	A	T	SNV	upstream_gene_variant			1|1	
15	786	chr03	11022255	11022260	AATGGC	A	DEL (5)	frameshift_variant	Os03g0309200	[PHYB] Similar to Phytochrome B.	1|1	✓
16	786	chr06	5399955	5399955	A	C	SNV	intron_variant			1|1	
17	786	chr07	13311076	13311076	A	T	SNV	upstream_gene_variant			1|1	
18	786	chr10	21181033	21181034	CG	C	DEL (1)	frameshift_variant	Os10g0542400	Expansin/Lol pI family protein.	1|1	✓
19	786	chr12	23844768	23844769	GC	G	DEL (1)	frameshift_variant	Os12g0577100	Nucleotide-binding, alpha-beta plait domain containing protein.	0|1	
20	885	chr01	1319914	1319917	TAGA	T	DEL (3)	intergenic_region			1|1	
21	885	chr01	3020438	3020439	TG	T	DEL (1)	intergenic_region			0|1	
22	885	chr01	3357475	3357475	C	T	SNV	synonymous_variant			1|1	✓
23	885	chr01	37524019	37524065	GTGT….CGCA	G	DEL (46)	5_prime_UTR_variant	Os01g0866500	Similar to Soluble inorganic pyrophosphatase (EC 3.6.1.1) (Pyrophosphate phospho- hydrolase)	1|1	✓
24	885	chr02	10379090	10379091	TA	T	DEL (1)	5_prime_UTR_variant	Os02g0280400	Similar to casein kinase I isoform delta-like. ; Similar to Dual specificity kinase 1.	0|1	
25	885	chr03	7473040	7473040	T	C	SNV	upstream_gene_variant			0|1	
26	885	chr03	35501010	35501011	TC	T	DEL (1)	upstream_gene_variant			1|1	✓
27	885	chr04	23792560	23792560	G	A	SNV	upstream_gene_variant			1|1	✓
28	885	chr05	318166	318166	G	A	SNV	missense_variant	Os05g0105900	Nucleotide-binding, alpha-beta plait domain containing protein.	0|1	
29	885	chr05	15013747	15013748	G	GA	INS (1)	intergenic_region			0|1	
30	885	chr05	21895103	21895103	C	T	SNV	missense_variant	Os05g0446600	Similar to DNA glycosylase/lyase 701.	1|1	✓
31	885	chr08	5808016	5808016	T	C	SNV	downstream_gene_variant			1|1	
32	885	chr08	14099483	14099489	CATCGAA	C	DEL (6)	intergenic_region			1|1	✓
33	885	chr08	24691421	24691425	TGAAA	T	DEL (4)	downstream_gene_variant			1|1	✓
34	885	chr08	27686683	27686683	T	C	SNV	missense_variant	Os08g0553400	Hypothetical conserved gene.	1|1	✓
35	885	chr10	15810761	15810767	TCGCGCC	T	DEL (6)	disruptive_inframe_deletion	Os10g0439924	[CYP71Z8] Cytochrome P450 family protein.	1|1	✓
36	IRB3517-3	chr01	22075801	22075801	G	T	SNV	missense_variant	Os01g0574400	Similar to Cell division protein ftsH (EC 3.4.24.-).	0|1	
37[7]	IRB3517-3	chr01	32645629	32645636	ACTCTGCAT	AA	RPL	intron_variant			1|1	✓
38	IRB3517-3	chr03	32977149	33010771	TGAG….GTTG	T	DEL (33.6k)	exon_loss_variant&splice_region_variant	Os03g0793000 ~ Os03g0793500	5 ORFs are deleted[8]	1|1	✓
39	IRB3517-3	chr04	26873179	26873179	A	C	SNV	intergenic_region			1|1	
40	IRB3517-3	chr06	9657976	9657977	CA	C	DEL (1)	intergenic_region			0|1	
41	IRB3517-3	chr06	27521458	27521464	TCCTGTC	T	DEL (6)	disruptive_inframe_deletion	Os06g0665800	[HMA9] Similar to heavy metal ATPase.	1|1	✓
42	IRB3517-3	chr07	4881911	4881919	GCGTATATT	G	DEL (8)	upstream_gene_variant			1|1	✓
43	IRB3517-3	chr08	23894794	23894794	A	G	SNV	3_prime_UTR_variant	Os08g0483900	Helix-loop-helix DNA-binding domain containing protein.	0|1	
44	IRB3517-3	chr09	7224224	7224224	A	G	SNV	missense_variant	Os09g0297400	[PPT1] Similar to Phosphate/phosphoenolpyruvate translocator.	1|1	✓
45	IRB3517-3	chr09	13007477	13007477	G	A	SNV	synonymous_variant			0|1	✓
46	IRB3517-3	chr11	1188438	1188438	C	T	SNV	5_prime_UTR_variant	Os11g0126250	Hypothetical gene.	1|1	
47	IRB3517-3	chr11	7365282	7365282	T	C	SNV	missense_variant	Os11g0238000	Similar to NB-ARC domain containing protein.	1|1	✓
48	IRB3517-3	chr12	3883429	3883435	TCACTAC	T	DEL (6)	downstream_gene_variant	Os12g0176800	Similar to Heterochromatin protein (Fragment).	0|1	
49	IRB3790-2	chr01	5274282	5274283	GC	G	DEL (1)	intron_variant			0|1	
50	IRB3790-2	chr03	26252641	26252642	A	AT	INS (1)	upstream_gene_variant			0|1	
51	IRB3790-2	chr04	12393088	12393088	G	A	SNV	intergenic_region			1|1	
52	IRB3790-2	chr06	9338213	9872884	TATG….TTGG	TACA….TCAT	INV (535k)	frameshift_variant[8]	Os06g0275000[7]	Photoperiod-sensitivity-1, HEADING DATE 1[7]	0|1	✓
53	IRB3790-2	chr07	23822057	23822057	T	C	SNV	missense_variant	Os07g0586200	[GELP92] Similar to Esterase precursor (EC 3.1.1.-) (Early nodule-specific protein homolog) (Latex allergen Hev b 13).	0|1	
54	IRB3790-2	chr08	21594849	21594853	CTTAA	C	DEL (4)	upstream_gene_variant			1|1	✓
55	IRB3790-2	chr09	1369829	1369831	GGA	G	DEL (2)	intergenic_region			1|1	
56	IRB3790-2	chr12	4637905	4637905	G	A	SNV	upstream_gene_variant			1|1	✓

[1] Mutation No. is identical to the one in Table 2.

[2] For the large deletions and the inversion, only first and last 4 bases are shown.

[3] Size of InDels is indicated in parentheses.

[4] Mutations putatively make high and moderate impact on protein function are indicated by orange and blue background color.

[5] "0|1" is heterzygous and "1|1" is homozygous.

[6] Mutations verified by sanger sequencing were marked with checkmark. Primer sequences are shown in Table S1.

[7] Independent mutation calls are unified and counted as one RPL mutation because they are overlapped or located consecutively (Figure S1).

[8] Evaluated manually.