

# Analysis of carbon ion-induced mutations by exome sequencing of an unselected rice population

H. Ichida,<sup>\*1</sup> R. Morita,<sup>\*1</sup> Y. Shirakawa,<sup>\*1</sup> Y. Hayashi,<sup>\*1</sup> and T. Abe<sup>\*1</sup>

Massively parallel sequencing technology has been utilized in many areas of biology including mutation analysis, by using a large amount of gene and genome sequencing information to achieve a comprehensive and genome-wide analysis. Heavy-ion beams are one of the physical mutagens that are classified as high-LET radiation and are known to induce double strand breaks of DNA in a cell along with its track. The resulting mutations, including deletions, insertions, inversions and base substitutions, that occur on the genome can cause the inactivation and/or temporal change of gene expressions that are necessary for morphogenesis. We have developed a custom-designed oligonucleotide probe library to capture entire exons within the rice genome, which targets a total of 300,746 genomic loci at the same time. We also developed a bioinformatics pipeline to map the sequencing reads to the reference Nipponbare genome sequence and identify reliable mutations in a highly paralleled way by using the “HOKUSAI” parallel computing system operated by the Advanced Center for Computing and Communication, RIKEN. In addition to these previous efforts to make the genome-based mutation detections, we developed a pre-mixed target capture procedure to further reduce the usage of the custom-designed target capture oligonucleotide probes, which takes nearly half of the overall cost in whole exome sequencing, by mixing multiple libraries with different index sequences prior to the target capturing: this reduces the per-sample oligonucleotide probe usage to 1/8th of that of the original protocol.

In the present study, we analyzed a total of 110 independent M<sub>2</sub> lines from carbon-ion beam irradiations (<sup>12</sup>C<sup>6+</sup>, 135 MeV/u, LET: 30 keV/μm, 150 Gy) to Nipponbare rice seeds, the water content of which was adjusted to 13%. The irradiated seeds were grown in a paddy field, and the M<sub>2</sub> seeds were harvested from each line. In each line, 10 to 15 plants were grown in soil, and an equal amount of leaf blades were collected from each plant and subjected to genomic DNA extraction and sequencing library preparation. As a control, 8 pools of non-irradiated Nipponbare plants were also processed in the same manner. A total of 8 libraries were mixed together prior to the target capturing and then sequenced in a half lane on a HiSeq 4000 instrument. The obtained sequencing dataset was processed using our bioinformatics pipeline described above. As a result, the number of mutations within the target region of whole exome capturing was between 3 and 26 in each line (Fig. 1). The average number of

mutations was  $9.06 \pm 0.37$  (average  $\pm$  standard error) per line. There were a total of 997 mutations, which consisted of 573 base substitutions, 372 deletions, 36 insertions, 13 substitutions, and 3 inversions, identified from the irradiated M<sub>2</sub> lines. The percentage of deletions and insertions, against all detected mutations, was 40.9%, which was consistent with the previously described characteristics of mutations induced by highly accelerated heavy-ion beams, which often cause nucleotide substitutions and deletions and insertions of less than 100 bp.<sup>1)</sup> In contrast, no mutation was identified in non-irradiated Nipponbare pools, indicating that the mutations detected from the carbon-ion-irradiated samples were likely to be induced by the mutagenesis.

Based on an interpolation from the proportion of target exon regions against the entire genome (the total length of the target exon regions is 9.12% of the length of the entire genome), roughly 100 mutations are expected to be induced in the entire genome. This might be an underestimation due to the difference in biological significance between protein-encoding exons and other genomic regions, which are mostly intergenic regions and repetitive elements; however, this estimation was of the same order as our previous results (175 to 549 mutations per genome in the 12 mutants analyzed) obtained from the whole genome sequencing of morphological mutants in rice.<sup>2)</sup>

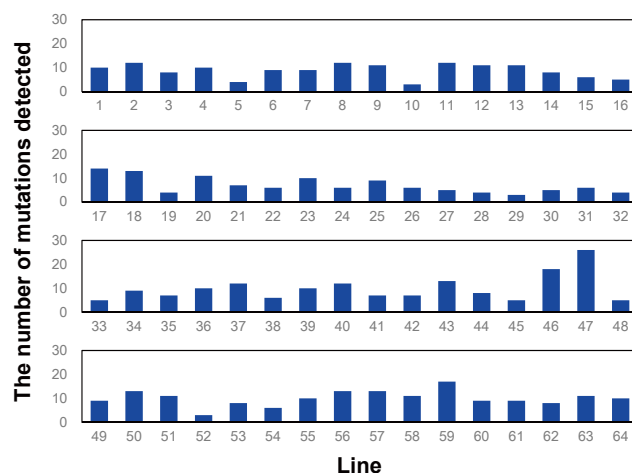


Fig. 1. Number of mutations detected from the carbon-ion irradiated rice M<sub>2</sub> lines (Nos. 1–64).

## References

- 1) Y. Kazama *et al.*, *BMC Plant Biol.* **11**, 161 (2011).
- 2) H. Ichida *et al.*, *RIKEN Accel. Prog. Rep.* **50**, 271 (2017).

<sup>\*1</sup> RIKEN Nishina Center