

Whole genome sequencing of 12 morphological rice mutants from carbon-ion irradiations

H. Ichida,^{*1} Y. Shirakawa,^{*1} R. Morita,^{*1} Y. Hayashi,^{*1} and T. Abe^{*1}

Heavy-ion beams are one of the physical mutagens that are classified in high-LET radiations and 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 previously reported the design and implementation of high-performance bioinformatics pipeline for massively parallel (aka “next generation”) DNA sequences obtained from mutants. We irradiated a carbon-ion beam from RRC ($^{12}\text{C}^{6+}$, 135 MeV/u) on dry and wet seeds of rice (*Oryza sativa* ‘Nipponbare’) plants and isolated morphological mutants in the M₂ generation. Of those mutants, we selected 12 morphological mutants for comprehensive mutation analysis by whole genome sequencing and subsequent bioinformatics analysis.

We utilized the whole-genome shotgun strategy using an Illumina HiSeq X system. Leaf blades were collected from each M₃ plant grown in a paddy field in September 2016, shipped in dry ice, and stored at -80°C until use. Genomic DNA was purified from 100 mg of bulked frozen tissue of each line using MagExtractor plant genome DNA extraction kits (Toyobo, Japan). The purified DNA was quantified and quality-checked, following which it was subjected to random fragmentation by sonication and subsequent processes to attach the adapter sequences necessary for Illumina sequencing. We aimed at an average insert size of 350 bp and a total of 30 billion bases of clean sequences (the sequences that passed filtering by mean quality and read length) from each mutant line. This cor-

responds to approximately 80 times the equivalent of rice genome and is expected to be sufficient for accurate variant calling according to the current recommendations in variant calling programs. The data were analyzed using our bioinformatics pipeline¹⁾ built and operated on the Hokusai GreatWave system operated by the Advanced Center for Computing and Communication, RIKEN.

The statistics of sequencing, mapping, and variant calling are summarized in Table 1. We could obtain a similar amount of sequencing reads as well as other mapping metrics, indicating that the procedure is quite stable and robust. The number of mutations identified by the Genome Analysis Tool Kit (GATK)²⁾ was also similar among the mutants sequenced. Roughly 52,000 to 53,000 mutations were reported by the GATK program. However, as expected, most of those were artifacts from the intra-cultivar variations between the sequenced and our mutagenized Nipponbare cultivars of different origins. Our filtering strategy removed such false positives very effectively: only 175 to 549 mutations were likely to be the actual mutations caused by heavy-ion beam irradiations. These results demonstrate that our approach using whole genome sequencing and the high-performance bioinformatics pipeline is well suited for the mutation analysis in rice.

References

- 1) H. Ichida et al., RIKEN Accel. Prog. Rep. **49**, 254 (2016).
- 2) A. McKenna et al., Genome Res. **20**, 1297–1303 (2010).

Table 1. Sequencing, mapping, and variant calling statistics of whole genome sequencing from 12 rice mutants

Seed condition	Mutant	Irradiation condition		Total reads	Aligned (%)	Mismatch ($\times 10^{-3}$)	Indel ($\times 10^{-5}$)	Num. of mutations	
		LET ^{*1}	Dose (Gy)					Total	Line specific
Wet	C-249	50	15	243,240,586	99.53	3.21	5.00	53,600	191
Dry (9% ^{*2})	7-80	30	50	233,349,445	99.43	3.71	5.10	52,404	175
	7-84	30	50	231,042,177	99.44	3.94	5.00	52,365	202
	7-92	30	50	227,590,774	99.52	3.70	5.00	52,476	210
Dry (13% ^{*2})	14-34	30	150	236,257,550	98.40	3.59	5.10	53,135	276
	14-55	30	150	234,303,120	99.05	4.05	5.00	52,593	193
	14-64	30	150	234,084,341	99.38	3.84	5.00	53,273	273
	14-70	30	150	233,143,578	99.25	3.31	4.80	53,196	294
	15-16	30	175	235,733,926	99.52	3.33	5.00	52,373	404
	15-17	30	175	233,910,879	99.14	3.43	4.90	52,601	549
	15-65	30	175	232,386,380	99.43	3.38	4.90	52,174	280
	15-74	30	175	233,696,581	99.49	3.38	4.90	52,115	258

^{*1} keV/ μm ; ^{*2} water content of the irradiated seed

^{*1} RIKEN Nishina Center