CellPress
PARTNER JOURNAL

Molecular Plant
Letter to the Editor

Genome-Wide Sequencing of 41 Rice (*Oryza sativa* L.) Mutated Lines Reveals Diverse Mutations Induced by Fast-Neutron Irradiation

Dear Editor,

Fast-neutron (FN) irradiation has been used to create mutagenized collections of many plant species (Bolon et al., 2014). FN-induced mutagenesis has clear advantages: it is an efficient means of saturating the genome, and it does not involve time-consuming plant transformation or tissue culture. In rice, most mutant collections, although highly valuable, were generated using either T-DNA insertion or transposon tagging approaches that often induce mutations unlinked to the insertion and complicating analysis (Wang et al., 2013). Another disadvantage of some of these collections is that they were created in rice varieties that are photoperiod-sensitive and often have long generation times.

FN irradiation affects DNA by damaging base groups, creating strand breaks, and making other changes (Gual et al., 2011). Multiple approaches have been taken to study the consequences of FN irradiation on plant genomes, including the polymerase chain reaction (PCR), comparative genomic hybridization, and next-generation sequencing (Belfield et al., 2012; Bolon et al., 2014). These investigations have revealed the presence of single base substitutions (SBSs), deletions, insertions, and duplications induced by FN irradiation. Despite these advances, the genome-wide consequences of FN irradiation have not been fully characterized, due either to the limited resolution of the approach or to the limited number of samples analyzed. In particular, structural variants (genomic variants larger than 1 kb) have not been systematically examined.

To accelerate functional genomic studies of rice and to investigate the effects of FN irradiation on the rice genome, we generated an FN-mutagenized population in X.Kitaake, a derivative of the *Japonica* rice variety Kitaake (Supplemental Information S1 and Figure 1A). Kitaake is neutral to photoperiod changes and has a very short life cycle (9 weeks), which allows researchers to grow four generations in the greenhouse each year (Jung et al., 2008). We treated 10 000 seeds with 20 Gy of FN irradiation. Approximately 9000 seeds germinated, and 7333 were fertile. M_2 seeds derived from individual M_1 plants were collected and stored (Supplemental Table 1).

To investigate the effects of FN irradiation on the rice genome, we sequenced 41 M_3 lines and one non-irradiated parental line using the Illumina Hiseq 2000 system (Supplemental Information S2). On average, 196 million reads of each line were mapped onto the Nipponbare reference genome, giving a sequencing depth of 52 fold (Supplemental Table 2). 97.5% of the reference genome is covered by 10 or more reads from each sequenced line. We applied multiple complementary variant-calling approaches to identify genomic variants (Supplemental Information S3 and S4). Based on mutations defined in the variant-calling algorithms (Supplemental Information S3), a total of 2418 FN-induced homozygous and heterozygous mutations were detected in these 41 rice lines, including 1273 SBSs, 864 deletions, 145 insertions, 82 inversions, 49 translocations, and five tandem duplications (Figure 1B). For example, translocations and insertions are different in the orientation and location of anomalous read pairs. The origin of the translocated DNA fragment is known but the origin of inserted DNA usually cannot be identified. Owing to the nature of variant calls made by our algorithms and the limited sequence depth that we have generated (Supplemental Information S4), our results contain only tandem duplications, not dispersed duplications, and contain only genotype (homozygosity/heterozygosity) information for SBSs as well as small deletions and insertions (Supplemental Data 1). The sizes of FN-induced mutations range from 1 bp to 18 Mb. We detected a greater diversity of FN-induced mutations than that reported by other groups (Belfield et al., 2012; Bolon et al., 2014).

To gain additional insight into FN-induced mutagenesis in plants, each type of FN-induced mutation was further analyzed. Similar to the results of a study of FN-induced mutations in *Arabidopsis* (Belfield et al., 2012), we found that FN-induced SBSs were the most abundant mutation type identified (Supplemental Data 1). More transitions than transversions were observed in both rice and *Arabidopsis*. Transversions occur more frequently in FN-induced SBSs than in the SNPs present between X.Kitaake and Nipponbare, resulting in a reduced transition/transversion (Ti/Tv) ratio (Figure 1C and Supplemental Data 2). Moreover, these SBSs have a much higher proportion of C > T mutations than the X.Kitaake SNPs. A total of 163 genes are mutated by SBSs (Figure 1B and Supplemental Data 3). The sizes of FN-induced insertions and deletions range from 1 bp to 678 kb, with small ($\leq$ 10 bp) deletions and insertions accounting for 73.2% of all deletion and insertion events (Supplemental Data 1 and Figure 1D). Single base deletions account for 26.3% of all deletions. There are 73 deletions larger than 1 kb. In total, 918 genes are affected by deletions and insertions in the 41 M_3 lines. Among the 82 observed inversions, 80 are paracentric and two are pericentric (Supplemental Data 1). In our study, inversions and translocations affect 64 and 49 genes, respectively (Figure 1B and Supplemental Data 3). Five tandem duplications were detected, which increased the copy number of 90 genes (Supplemental Data 3). FN-induced complex mutations were

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

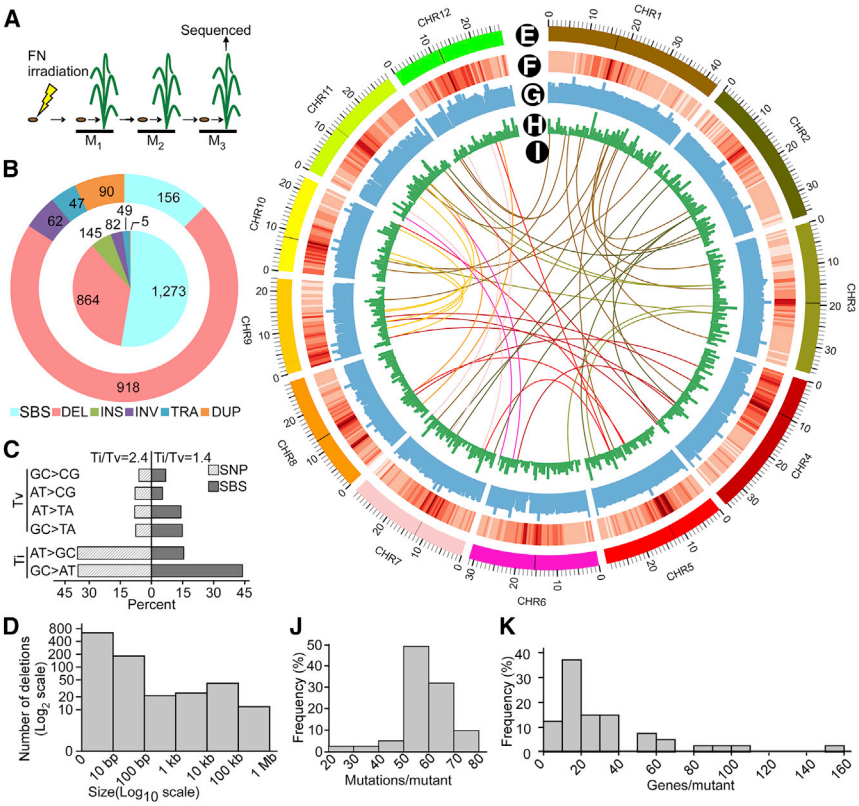

Figure 1. Genome-wide Characterization of Mutations in 41 FN-Induced Rice Mutant Lines.

(A) Strategy for generating and sequencing the rice FN-mutagenized population. Seeds from individual M_1 plants were collected separately and stored. Genomic DNA isolated from a single M_3 plant was used in sequencing.

(B) Representation of the diversity of mutations induced by FN irradiation. The inner pie chart indicates the number of each type of mutation in the 41 sequenced M_3 lines. The outer circle shows the number of genes affected by each type of mutation. SBS, single base substitutions; DEL, deletions; INS, insertions; INV, inversions; TRA, translocations; and DUP, tandem duplications.

(C) Comparisons between FN-induced SBSs and the SNPs present between X.Kitaake and the Nipponbare reference genome. Ti, transitions; Tv, transversions.

(D) Size distribution of FN-induced deletions the 41 rice mutant lines. The x axis shows deletion size in the log10 scale, and the y axis depicts the number of deletion events in the log2 scale.

(E) Representation of the 12 rice chromosomes on an Mb scale. The centromere is indicated by a black line.

(F) Representation of repetitive sequences in the reference genome in non-overlapping windows (window size = 500 kb). The darker the color, the higher content of repetitive sequence. Repeat data were derived from RGAP version 7.

(G) The sequencing depth of the 42 rice lines, including the 41 M_3 lines and the nonirradiated parental line X.Kitaake.

(H) Genome-wide distribution of FN-induced mutations in non-overlapping 500-kb windows. The highest bar shown in chromosome 1 equates to 17 mutations/500 kb.

(I) Translocations. Each line is colored according to the smaller number chromosome that the translocation involves.

(J) The frequency of the number of FN-induced mutations per mutant line in the 41 M_3 lines.

(K) The frequency of the number of genes affected per mutant line in the 41 M_3 lines.

also detected, including inversion–deletion and inversion–insertion (Supplemental Figures 1 and 2). The complex mutations presented in our study may facilitate further mechanistic analysis of how such complex mutations are generated. On a genome-wide scale, 40% of the 2418 FN-induced mutations overlap genic regions, a number very close to the theoretical 44% if the mutations were randomly distributed. In total, 1284 genes are affected in the 41 sequenced lines. Specifically, SBSs mutate 12.7% of the affected genes (Figure 1B); deletions mutate 71.5%, constituting by far the largest portion of mutated genes. In summary, SBSs constitute the most abundant mutation type, but deletions mutate the largest number of genes.

To assess the genome-wide distribution pattern of mutations, mutations in the 41 sequenced lines were mapped to the Nipponbare reference genome. FN-induced mutations are distributed evenly across the genome (Figure 1H), with the exception of some repetitive regions (Figure 1F). In *Caenorhabditis elegans*, chemically induced single nucleotide variants were found to be almost uniformly distributed across the genome (Thompson et al., 2013). In contrast, the insertions of the most widely used rice and *Arabidopsis* T-DNA insertion mutant collections are biased toward subtelomeric regions (Alonso et al., 2003; Wang et al., 2013). The distribution patterns of mutations in different mutant collections reflect the distinct characteristics of physical/chemical mutagenesis and insertional mutagenesis. The number of mutations per sequenced FN-mutagenized line ranges from 28 to 78, with an average of 59 (Figure 1J and Supplemental Data 1). In comparison, there are 1.4 insertions/line in the T-DNA insertion mutant collection created in the rice variety Dongjin (Jeon et al., 2000) and approximately 390 SBSs/line in a Nipponbare EMS mutant collection (Henry et al., 2014). The number of genes affected in each of the 41 M3 lines ranges from seven to 147, with an average of 31 (Figure 1K and Supplemental Data 3). The large variation of mutated genes per line is due to the presence of large deletions (Supplemental Data 3). For example, 49 genes are disrupted by a 300 kb deletion in line W949 (Supplemental Figure 3). Based on the observed number of mutations in these sequenced lines, we calculated that the probability of identifying a mutation in a gene of interest in the mutant collection containing over 7300 FN-induced lines is 97% (Supplemental Information S6). These results indicate that this collection constitutes a highly valuable genetic resource for forward and reverse genetic studies.

Both homozygous and heterozygous mutations were detected in the M_3 generation. The heterozygous mutations comprise some deleterious mutations. Deleterious mutations decrease organismal fitness, for example by reducing fertility, and will be

therefore more readily removed from the population. This will likely be the case for the two pericentric inversions in lines W1017 and N1202, as these plants produce few seeds. In subsequent generations, as mutant lines carrying deleterious mutations are lost and other heterozygous mutations segregate out, the ratio of structural variants to SBSs will be reduced and the overall mutations will become fewer. In addition, the distribution of mutations will change, becoming less even as deleterious mutations in the gene-rich regions are selectively eliminated. In light of the changes, we assume that FN irradiation causes more deleterious mutations in the M_1 generation than that estimated from our dataset, a conclusion also supported by the fact that 27% of the 10 000 M_1 lines were sterile (Supplemental Table 1).

Mutations detected in this study likely include spontaneous mutations. Using information about spontaneous mutations in rice (Yang et al., 2015), we calculate that an average of 9.6 spontaneous mutations (7.6 SBSs and two small deletions/insertion) occur in each M_3 plant (Supplemental Figure 4). Spontaneous mutations slightly inflate the number of mutations but have minimal effects on non-SBS mutations that affect most of the genes. Subtracting the estimated spontaneous mutations, the data still support the conclusion that SBSs are the most abundant mutation type. More importantly, spontaneous mutations do not affect the utility of the sequenced population, as they were also detected and recorded in each line. To calculate the spontaneous mutation rate more accurately in the rice variety used in this study, it will be necessary to detect accumulated spontaneous mutations from the genome sequences of multiple unmutagnenized rice lines that have self-pollinated over multiple generations.

This study presents the first study of genome-wide profiling of mutations in FN-induced rice mutants. The conclusions garnered from this analysis are likely representative of the types of genomic variants, except dispersed duplications, present in the over 7000 rice mutants that remain to be characterized and of the genomic variants in FN-irradiated collections of other plant species (Bolon et al., 2014). The high-resolution profiles of both small mutations and large structural variants make this collection particularly advantageous. High-throughput phenotyping of the sequenced population will facilitate functional genomic studies in forward genetic screens (Yang et al., 2014). The population can also be used to facilitate reverse genetic studies. By taking advantage of genome-scale functional gene network tools, such as RiceNet (Lee et al., 2015), researchers can identify genes highly predicted to function in a particular biological process. The future availability of the fully sequenced FN-irradiated mutant population will allow researchers to quickly identify rice lines carrying mutations in specific genes and characterize gene function.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

The work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. The work conducted by the US Department of Energy Joint Genome Institute was supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231. Partial funding for this research was provided by NIH (GM55962) and NSF (IOS-1237975) to P.C.R. L.J. was supported by a fellowship from the Education Department in Fujian Province and Xiamen University.

AUTHOR CONTRIBUTIONS

G.L., M.C., and P.C.R. conceived and designed the experiments. G.L., M.C., R.J., J.A.M., L.J., W.S.S., A.M.L., M.E.V., K.W.B., and J.S. performed the experiments. G.L., R.J., J.A.M., W.S.S., and A.M.L. analyzed the data. G.L., M.C., R.J., and P.C.R. wrote the article.

ACKNOWLEDGMENTS

We thank Patrick E. Canlas, Florian Kraemer, Nhan T. Pham, Kyle C. Jones, and Dr. Syed Mehar Ali Shah for assistance with genomic DNA isolation and seed organization, and Dr. Catherine Nelson for critical reading of the manuscript. No conflict of interest declared.

Received: December 9, 2015
Revised: February 5, 2016
Accepted: March 7, 2016
Published: March 24, 2016

Guotian Li[1,2,6], Mawsheng Chern[1,2,6], Rashmi Jain[1,2,6], Joel A. Martin[3], Wendy S. Schackwitz[3], Liangrong Jiang[1,4], Miguel E. Vega-Sánchez[2,7], Anna M. Lipzen[3], Kerrie W. Barry[3], Jeremy Schmutz[3,5] and Pamela C. Ronald[1,2,]*

[1]Department of Plant Pathology and the Genome Center, University of California, Davis, CA 95616, USA
[2]Feedstocks Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[3]U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA
[4]School of Life Sciences, Xiamen University, Xiamen 361102, China
[5]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA
[6]These authors contributed equally to this article.
[7]Present address: Monsanto Company, Chesterfield Village Campus, Chesterfield, MO 63017, USA
***Correspondence: Pamela C. Ronald (pcronald@ucdavis.edu)**
http://dx.doi.org/10.1016/j.molp.2016.03.009

REFERENCES

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H.M., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301**:653–657.

Belfield, E.J., Gan, X., Mithani, A., Brown, C., Jiang, C., Franklin, K., Alvey, E., Wibowo, A., Jung, M., Bailey, K., et al. (2012). Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. Genome Res. **22**:1306–1315.

Bolon, Y.T., Stec, A.O., Michno, J.M., Roessler, J., Bhaskar, P.B., Ries, L., Dobbels, A.A., Campbell, B.W., Young, N.P., Anderson, J.E., et al. (2014). Genome resilience and prevalence of segmental duplications following fast neutron irradiation of soybean. Genetics **198**:967–981.

Gual, M.R., Milian, F.M., Deppman, A., and Coelho, P.R. (2011). Study of DNA damage with a new system for irradiation of samples in a nuclear reactor. Appl. Radiat. Isot. **69**:373–376.

Henry, I.M., Nagalakshmi, U., Lieberman, M.C., Ngo, K.J., Krasileva, K.V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., Tai, T.H., et al. (2014). Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. Plant Cell **26**:1382–1397.

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.H., Jeong, D.H., Lee, J., Kim, C., Jang, S., Yang, K., Nam, J., et al. (2000). T-DNA insertional mutagenesis for functional genomics in rice. Plant J. **22**:561–570.

Jung, K.H., An, G.H., and Ronald, P.C. (2008). Towards a better bowl of rice: assigning function to tens of thousands of rice genes. Nat. Rev. Genet. **9**:91–101.

Lee, T., Oh, T., Yang, S., Shin, J., Hwang, S., Kim, C.Y., Kim, H., Shim, H., Shim, J.E., Ronald, P.C., et al. (2015). RiceNet v2: an improved network prioritization server for rice genes. Nucleic Acids Res. **43**:W122–W127.

Thompson, O., Edgley, M., Strasbourger, P., Flibotte, S., Ewing, B., Adair, R., Au, V., Chaudhry, I., Fernando, L., Hutter, H., et al. (2013). The million mutation project: a new approach to genetics in *Caenorhabditis elegans*. Genome Res. **23**:1749–1762.

Wang, N.L., Long, T.A., Yao, W., Xiong, L.Z., Zhang, Q.F., and Wu, C.Y. (2013). Mutant resources for the functional analysis of the rice genome. Mol. Plant **6**:596–604.

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., Fang, W., Feng, H., Xie, W., Lian, X., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat. Commun. **5**:5087.

Yang, S., Wang, L., Huang, J., Zhang, X., Yuan, Y., Chen, J.Q., Hurst, L.D., and Tian, D. (2015). Parent-progeny sequencing indicates higher mutation rates in heterozygotes. Nature **523**:463–467.