

Genome-wide Targeted Mutagenesis in Rice Using the CRISPR/Cas9 System

Dear Editor,

Since the completion of the rice (Oryza sativa) genomesequencing project, a major goal of rice research has been the functional characterization of all annotated genetic loci in various biological processes. One of the most efficient and widely-used strategies for studying gene function is genetic mutagenesis. Several rice mutant libraries have been generated in the past decade, providing a wealth of resources for plant research (Chang et al., 2012). CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats-associated nuclease 9) has recently emerged as a powerful tool for rice research and breeding (Cong et al., 2013; Feng et al., 2013; Miao et al., 2013; Shan et al., 2013; Sun et al., 2016). The technology provides an effective method of introducing targeted insertions and deletions (indels) at specific sites in the genome that result in loss-of-function alleles. Because the targeting specificity of CRISPR/Cas9 is conferred by a 20-bp short guide RNA (sgRNA), it can be easily generated on a large scale by array-based synthesis of oligonucleotide libraries. Several genome-scale CRISPR/ Cas9 mutagenesis systems have been established for mammalian cells using such synthesized sgRNA libraries, allowing effective genome-scale loss-of-function genetic screening (Shalem et al., 2014; Wang et al., 2014). However, to our knowledge, such powerful genome-scale mutagenesis systems have not been successfully applied for plant research. Here, we developed a pooled approach for genome-scale mutagenesis of genes in rice using an sgRNA library. A total of 91 004 targeted loss-offunction mutants were generated, which provides a useful resource for rice research and breeding.

On the basis of highly efficient CRISPR/Cas9-mediated mutagenesis in rice, we set out to explore the feasibility of applying this technology to perform genome-scale mutagenesis in rice to generate a library of targeted loss-of-function mutants. The main point was to use pooled sgRNA-expressing binary plasmids to generate a library of transgenic rice plants. The sgRNA(s) integrated with T-DNA in the genome of each transgenic plant would serve as a distinct DNA barcode to indicate the gene(s) targeted for mutagenesis.

To design an sgRNA pool targeting all the 39 045 non-TE loci of rice (MSU7), all target sites of CRISPR/Cas9 were computationally derived from the whole genome (Supplemental Figure 1). As shown in Figure 1A, sgRNAs against 5' constitutive coding exons were selected. To minimize off-target cleavages, only sgRNAs that satisfied stringent conditions were chosen. In consideration of the potential complications caused by GC content and cleavage position in its targeting locus, two to three sgRNAs were selected for each gene (Figure 1B and Supplemental Methods). In the end, we designed an sgRNA library with 88 541 members, targeting 34 234 genes with an

average coverage of 2.59 sgRNAs per gene (Figure 1C and Supplemental Table 1). All sgRNAs were classified into 96 groups according to the priority or annotated functions of their target genes (Figure 1C and Supplemental Table 2). The 20-bp sgRNA sequences in each group were flanked with additional nucleotides to facilitate amplification group by group from the synthesized oligonucleotide pool (Figure 1D; Supplemental Figure 2 and Supplemental Table 3). The mutation frequency of CRISPR/Cas9 vector BGK03 has been shown to be as high as \sim 80%, similar to other vectors (Shan et al., 2013; Ma et al., 2015; Xie et al., 2015). In order to avoid self-ligation of BGK03 during insertion of sgRNAs into the digested vector, the toxic gene ccdB was inserted between two Bsal sites (Figure 1D and 1E). The PCR result verified that negative selectivity of ccdB could improve the accuracy of plasmid construction to almost 100%, thus ensuring the efficiency and reliability of library preparation (Supplemental Figure 3). Using this modified vector, a genome-scale mutagenesis library of rice (RGKO-ALL) and three separated sublibraries (RGKO#2, #34 and #66, Supplemental Table 2) were constructed. Differing from sgRNA libraries for transient screening in mammalian cells, the RGKO libraries were used for generating stable transgenic plants. To confirm that the majority of these plasmids in the libraries were correct, all libraries were carefully tested. A total of 1109 Escherichia coli colonies were sequenced individually during library construction using Sanger sequencing, and all four pooled plasmid libraries were further verified with nextgeneration sequencing (NGS). The results suggest that more than 90% of plasmids in the libraries were correct and covered more than 99% of the designed sgRNAs, indicating their usability for further experiments (Figure 1F and Supplemental Table 4).

To assess the mutation frequency of RGKO, a total of 62 plasmids isolated from E. coli colonies of RGKO#2 were used individually for rice transformation. Of 1488 stable transgenic seedlings were regenerated from hygromycin-resistant calli, targeting 62 genes. A total of 364 seedlings were genotyped using Sanger sequencing, and the results showed that 315 seedlings contained indels at the targeted sites, indicating an 86.5% mutation frequency using RGKO libraries (Figure 1G). As an initial evaluation of our pooled approach, sublibraries RGKO#2, #34, and #66 were used for transformation. To keep the uniformity of sgRNAs in pooled Agrobacterium, a key modification for rice transformation was made that millions of Agrobacterium colonies from electroporation were directly used for rice transformation. A total of 5132 transgenic plants were generated from these RGKO libraries. And a random sampling survey indicated that 35 of 41 (85.4%) plants tested

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.

Letter to the Editor Molecular Plant

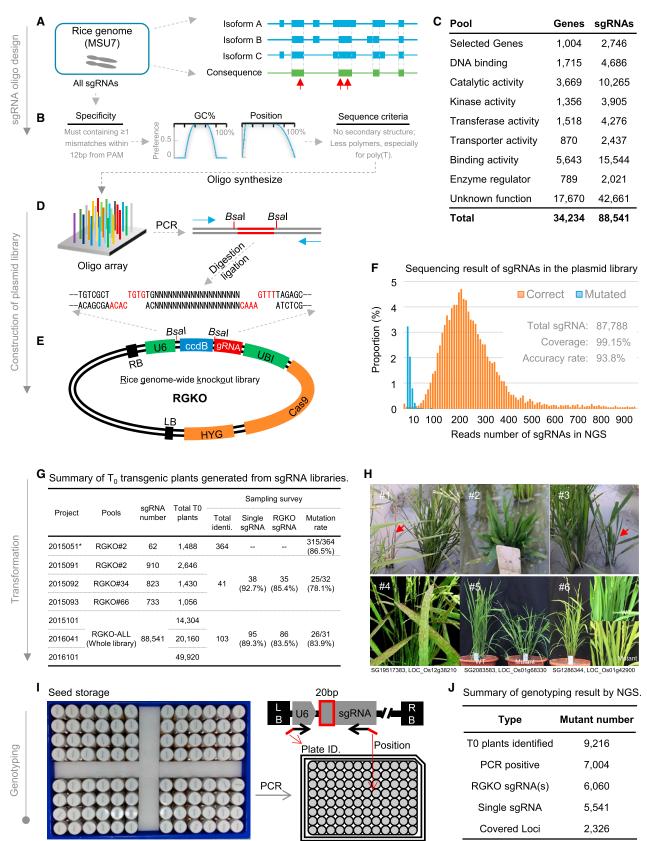


Figure 1. Genome-Scale Mutagenesis of Genes in Rice Using a Pooled sgRNA Library.

(A) Example of sqRNA design. sqRNAs (red arrows) targeting constitutive exonic coding sequences near the start codon were chosen.

⁽B) Pipeline of the sgRNA library design (Supplemental Methods).

Molecular Plant Letter to the Editor

contained correct sgRNAs that belong to the RGKOs, and the mutation frequency at their target loci was about 78.1% (25 of 32 successfully sequenced samples; Figure 1G). These results further demonstrate the reliability of our pooled approach for efficiently generating a targeted mutagenesis library. Thus, the whole library RGKO-ALL was used for conducting gene mutagenesis at the genomic scale. Eventually, a total of 84 384 transgenic plants were generated in three transformation projects, which is equivalent to approximately 1x coverage for all sgRNAs (88 541). A random sampling survey of this expanded library revealed a similar mutation frequency (83.9%). According to the sequencing results from all transformation projects (Figure 1G), most transgenic rice plants contain a "single sgRNA." This single sgRNA does not imply one copy of T-DNA because the single sgRNA may come from multiple copies of T-DNAs with the same sgRNA. Considering that the copy number of T-DNA in each transgenic rice plant is much higher (Chang et al., 2012), we speculate that many multiplyinserted T-DNAs may come from the same Agrobacterium cell during rice transformation. During the growth of T0 transgenic plants in the field, phenotypic alterations possibly due to gene mutations were occasionally observed. Among them, some mutants were lethal or sterile with no progeny (Figure 1H, #1-#3), while some exhibited visible growth defects. As shown in Figure 1H (#4-#6), spotted leaves, increased tiller angle, and altered leaf color were observed. The genes potentially responsible for the phenotypes were easily identified according to the sgRNAs (Figure 1H). Taken together, these data demonstrated the feasibility of this pooled approach for genome-scale mutagenesis of genes. The up to 80% targeted mutagenesis frequency would make the library a useful resource for rice research and breeding.

Although the genotype of each mutant can be conveniently identified using Sanger sequencing, it is challenging and costly when applied to hundreds of thousands of mutants. To solve this problem, a high-throughput genotyping method was developed (Bell et al., 2014). As shown in Figure 1I, all seeds and their genomic DNA samples were stored in 96-well plate format, and PCR primers amplifying the sgRNAs in the T-DNA were tailed with 6-bp additional nucleotides as barcodes. Accordingly, 96 reverse primers were computationally designed, corresponding to the 96 wells in the plate; 96 forward primers tailed with barcode were also synthesized, labeling the plate ID (Supplemental Table 5). As designed, we have conducted 96 × 96 PCR reactions to amplify sgRNAs from 9216 transgenic plants. All PCR solutions were mixed

together for NGS as a single sample. sgRNA(s) of each mutant was distinguished from NGS data by its barcode. As listed in Figure 1J, 7004 samples were successfully identified (PCR positive). The remaining undetected plants may be caused by false positives in rice transformation or failure in PCR amplification. 86.5% (6060) of the identified plants contained sgRNAs belonging to the RGKO library and most of them had a single sgRNA, which is consistent with the Sanger sequencing results (Figure 1G). According to the NGS result, a total of 2326 loci were covered in these identified 5541 plants (Supplemental Table 6). To verify the NGS result, 66 plants were randomly selected for Sanger sequencing. Completely identical results confirmed the high accuracy of this NGS-based high-throughput genotyping method (Supplemental Table 6).

CRISPR/Cas9 has greatly accelerated research and breeding on plants. Based on this technology, here we provide a detailed pipeline for genome-scale mutagenesis of genes in rice. The high mutation frequency makes the RGKO library a useful resource for rice research and breeding. Although much rice transformation and mutant genotyping work remains to be conducted, the simplicity and effectiveness of this pooled approach make it easy to be expanded. In the future, we could adapt this approach and make use of conserved 20-bp sgRNAs among redundant genes to simultaneously mutate multiple members of a gene family on a genome scale to mitigate the formidable problem of redundancy. Combined with the NGS-based high-throughput genotyping method describe above, this genome-scale mutagenesis system can be applied to other plant species to promote research and breeding.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at Molecular Plant Online.

FUNDING

This project was supported by grants from the National Natural Science Foundation of China (91335203, 31430063).

ACKNOWLEDGMENTS

We thank other colleagues of Biogle for producing transgenic rice. No conflict of interest declared.

Received: April 14, 2017 Revised: June 1, 2017 Accepted: June 11, 2017 Published: June 20, 2017

- (C) Composition of the rice genome-scale mutagenesis (RGKO) library.
- (D) Outline of the sgRNA library construction. Primers indicated by arrows were used for amplification of sgRNAs from the synthesized oligo pool (Supplemental Figure 3).
- (E) Schematic diagram of the binary vector for library construction.
- (F) Sequencing result of sgRNAs in the plasmid library. Plasmid pools containing the whole sgRNA library were identified using next-generation sequencing (NGS). Columns indicate the distribution of sgRNAs.
- **(G)** Summary of T₀ transgenic plants generated from RGKO libraries. Project marked with (*) indicates that rice transformation was conducted using plasmids individually. The others were transformed using pooled *Agrobacterium*. Rice cultivar ZH11 was used for all transformations. Total identi., total mutants identified; single sgRNA, mutants containing only one sgRNA; RGKO sgRNA, sgRNAs belonging to the RGKO library.
- (H) Visible phenotypes of T₀ transgenic seedlings in the field. WT, wild-type. Red arrows indicate the lethal or sterile T0 mutants.
- (I) Genotyping pipeline for the mutant library using NGS. Red lines in arrows indicate the 6-bp additional nucleotides as barcodes.
- (J) Summary of the genotyping result using NGS. PCR positive, successfully amplified and sequenced samples.

Letter to the Editor Molecular Plant

Yuming Lu^{1,3}, Xiao Ye¹, Renming Guo¹, Jing Huang¹, Wei Wang², Jiuyou Tang², Longtao Tan⁴, Jian-kang Zhu³, Chengcai Chu² and Yangwen Qian^{1,*}

¹Biogle Genome Editing Center, Changzhou, Jiangsu Province 213125, China ²State Key Laboratory of Plant Genomics, National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

³Shanghai Center for Plant Stress Biology and Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 201602, China

⁴Key Laboratory of Agro-ecological Processes in Subtropical Region, Institute of Subtropical Agriculture, Chinese Academy of Sciences, Changsha 410125, China

> *Correspondence: Yangwen Qian (qyw@biogle.cn) http://dx.doi.org/10.1016/j.molp.2017.06.007

REFERENCES

- Bell, C.C., Magor, G.W., Gillinder, K.R., and Perkins, A.C. (2014). A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing. BMC Genomics 15:1002.
- Chang, Y., Long, T., and Wu, C. (2012). Effort and contribution of T-DNA Insertion mutant library for rice functional genomics research in China: review and perspective. J. Integr. Plant Biol. 54:953–966.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823.

- Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.L., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y., et al. (2013). Efficient genome editing in plants using a CRISPR/Cas system. Cell Res. 23:1229–1232.
- Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., Wang, B., Yang, Z., Li, H., Lin, Y., et al. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol. Plant 8:1274–1284.
- Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J., Gu, H., and Qu, L.J. (2013). Targeted mutagenesis in rice using CRISPR-Cas system. Cell Res. 23:1233–1236.
- Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343:84–87.
- Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.L., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. Nat. Biotechnol. 31:686–688.
- Sun, Y., Zhang, X., Wu, C., He, Y., Ma, Y., Hou, H., Guo, X., Du, W., Zhao, Y., and Xia, L. (2016). Engineering herbicide-resistant rice plants through CRISPR/Cas9-mediated homologous recombination of acetolactate synthase. Mol. Plant 9:628–631.
- Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science 343:80–84.
- Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci. USA 112:3570–3575.