Accepted Manuscript

Construction of a genome-wide mutant library in rice using CRISPR/Cas9

Xiangbing Meng, Hong Yu, Yi Zhang, Fengfeng Zhuang, Xiaoguang Song, Songsong Gao, Caixia Gao, Jiayang Li

PII: S1674-2052(17)30172-7
DOI: 10.1016/j.molp.2017.06.006
Reference: MOLP 478

To appear in: MOLECULAR PLANT
Accepted Date: 11 June 2017


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

All studies published in MOLECULAR PLANT are embargoed until 3PM ET of the day they are published as corrected proofs on-line. Studies cannot be publicized as accepted manuscripts or uncorrected proofs.
Construction of a genome-wide mutant library in rice using CRISPR/Cas9

Xiangbing Meng¹, ⁵, Hong Yu¹, ⁵, Yi Zhang², ⁴, Fengfeng Zhuang³, Xiaoguang Song¹, Songsong Gao³, Caixia Gao², Jiayang Li¹, ⁴

¹ State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.
² State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China
³ Beijing ViewSolid Biotechnology, Beijing, China
⁴ University of the Chinese Academy of Sciences, Beijing 100049, China
⁵ These authors contributed equally to this work.

Correspondence should be addressed to J. L (jyli@genetics.ac.cn) and C. G. (cxgao@genetics.ac.cn).
Rice (*Oryza sativa* L.) is one of the world’s most important staple crops, and a powerful model system for studying monocot species because of its relatively small genome, rich genomic resources, and a highly efficient transformation system. With the complete genome of rice sequenced, the challenge of the post-genomic era is to systematically analyze the functions of all rice genes. Gene knockout is a frequently used and effective strategy for achieving this. Thus, generation of large-scale mutants at the whole genome level is of great value both for functional genomics and for genetic improvement of rice. Traditionally, large numbers of mutants are produced by physical, chemical or biological mutagenesis. Mutants created by these methods have made enormous contributions to plant basic research and crop improvement. T-DNA insertion (Jeon et al., 2000), TILLING (targeting induced local lesions in genomes) (Till et al., 2007) and RNAi (RNA interference) (Wang et al., 2013) are the three most common methods of performing genetic studies. T-DNA insertion and TILLING are time-consuming and labor-intensive, and large mutagenized populations must be generated to ensure sufficient genome-wide coverage. For T-DNA insertion, the insertions occur at random and often in intergenic and noncoding regions; for TILLING, it is difficult to identify the targeted mutations for the observed phenotypes. And RNAi reduces the expression of targeted genes rather than knocking them out. Recently, a simple and highly efficient genomic engineering tool, the CRISPR (Clustered Regularly Interspaced Palindromic Repeats)/Cas9 system, has been developed; this can create small insertions and deletions (indels) in specific target genes and has been applied to many organisms. Because of its easy and convenient characters, some CRISPR/Cas9 mutant libraries have been developed for genome-wide mutation screens in cultured eukaryotic cells (Shalem et al., 2015). However, no large-scale CRISPR/Cas9 mutant libraries have yet been generated in higher plants. Here, we report the construction of a high-throughput CRISPR/Cas9 mutant library in rice, and demonstrate its application to identifying gene functions and its potential use for genetic improvement.

Effective single-guide RNAs (sgRNA) targeting specific genomic sites using the CRISPR/Cas9 system are usually 20 bp in length and are followed by a protospacer adjacent motif (PAM) sequence (NGG). For rational design of highly specific sgRNAs, the 12-bp seed sequence of the sgRNA should match only one site in the target genome, so that there is only
a low possibility of off-target editing (Doench et al., 2014). Using this criterion, we searched
the rice genome and identified 1,535,852 target sites located in the exon regions of 52,916
rice genes. The CRISPR/Cas9 system induces double strand breaks and generally produces
small indels, which often cause frameshifts in protein-coding sequences. Hence, to generate
loss-of-function mutations efficiently, sgRNA target sites were designed in exons near the
beginning of ORFs just downstream of start codons. We chose the first two identified sgRNA
target sites in each candidate gene and selected 12,802 genes highly expressed in rice shoot
base tissue (rice expression profiles database RED) and 25,604 corresponding sgRNAs to
generate a large-scale mutant library.

We added 25-bp sequences (pVKmp-lib-FP0 and pVKmp-lib-RP0, Supplemental Table
1) derived from the plant expression vector pVKmp-lib (Supplemental Figure 1 and 2,
ViewSolid Biotech) to two ends of the 20-bp guide sequences respectively, for downstream
PCR amplification and Gibson ligation, and synthesized 25,604 oligonucleotides of 70 bp by
array-based oligonucleotide pool synthesis. To clone the synthesized oligo pool into the
CRISPR/Cas9 binary vector pVKmp-lib, which carries the hygromycin B phosphotransferase
(hpt) gene and the Cas9 expression cassette, the oligonucleotides were amplified by PCR
using primer pVKmp-lib-FP1 and pVKmp-lib-RP1 (Supplemental Table 1). The purified
PCR products were ligated into the pVKmp-lib vector by Gibson ligation (Figure 1A), and
the plasmids were transformed into bacterial host competent cells and selected with
kanamycin. More than 1.2×10^6 clones growing on the selection plates were harvested and
combined to prepare the plasmid DNAs constituting the sgRNA library.

To evaluate the accuracy of the large-scale synthesized sgRNA, we randomly selected
275 clones from a total of 1.2×10^6 clones and sequenced them. The sequence data
demonstrated that 36 of the 275 clones had altered sgRNA target sequences (13.1%) and 1
clone harbored no target sequence (0.4%) (Supplemental Table 2); the accuracy of the
synthesized sgRNA was therefore 86.5%. Next, the coverage of the sgRNA library was
assessed, the 303-bp PCR products generated from the plasmid DNA library with primers
pVKmp-F1 and pVKmp-R1 (Supplemental Table 1) were deep-sequenced by high
throughput sequencing. The sequence data showed that 25,265 of the 25,604 sgRNAs were
represented by at least one read (98.7%), and these 25,265 sgRNAs covered 12,786 genes
The majority (82.1%) of the sgRNAs had 11-90 reads, pointing to very high coverage and evenness of the sgRNA library (Figure 1B). These findings indicate that the constructed sgRNAs library is of high quality with good gene coverage and sgRNA accuracy.

Then the plasmid DNA library was transformed to *Agrobacterium tumefaciens* strain EHA105 for rice transformation. We tested the accuracy and coverage of the sgRNA library in *Agrobacterium* by randomly selecting 390 clones and sequenced them. The results showed that 337 of the 390 clones harboring correct sgRNA target sequences (86.41%) (Supplemental Table 2), and from the 337 clones, 331 unique sgRNAs were detected, suggesting that the sgRNA library in *Agrobacterium* is of high accuracy and coverage. The rice variety Zhonghua 11 (*Oryza sativa* L. ssp. *japonica*) was selected as the recipient plant as it can be efficiently transformed and has good fertility, and its genome is fully sequenced. Embryonic calli that were derived from the scutella of mature seed embryos and divided vigorously during subculture were selected for *Agrobacterium*-mediated transformation. Transgenic seedlings generated from single infected calli were regarded as independent transgenic T$_0$ lines, and more than 14,000 independent T$_0$ lines were obtained. To evaluate the quality and coverage of the resulting mutant library, we selected 182 T$_0$ plants at random and sequenced their genotypes. We found that 139 T$_0$ plants harbored single correct sgRNAs, 42 plants harbored incorrect sgRNAs, and one was non-transgenic (Supplemental Table 3). We also found that the 139 “correct” T$_0$ plants contained 136 different sgRNAs; thus only three sgRNAs were represented twice, demonstrating high coverage by the mutant library.

We then focused on the knock-out efficiency of the library using the DSDecode program (Liu et al., 2015); 46 plants were found to be homozygous mutant (33.1%), 15 plants were biallelic mutant (10.8%) and 27 were heterozygous mutant (19.4%) (Supplemental Figure 3 and Supplemental Table 4). Further analysis showed that most were small insertions or deletions at the target site; there were 40.1% (36/88) single-base insertions and deletions (Supplemental Figure 3), which is consistent with previous report (Ma et al., 2015). These results show that the CRISPR/Cas9 system is a powerful tool for constructing mutant libraries in rice.

In traditional mutant libraries the relationship between genotype and phenotype is usually very weak and it is difficult to identify and clone the genes corresponding to
particular phenotypes. It is estimated that the tagging efficiency of a typical rice T-DNA insertion library may be as low as 5–10% (Wei et al., 2016). Even though RNAi mutant libraries can have higher mutation rates (about 47.9% lines with observable phenotypes), mutations usually result from silencing of any one of several members of a gene family rather than of individual target genes (Wang et al., 2013). To evaluate the relationship between genotype and phenotype in the CRISPR/Cas9 mutant library, we harvested all the seeds from these 14,000 T₀ lines and randomly selected 200 lines for phenotyping by planting 24 plants per line in the field. Apparent morphological differences from the wild-type in terms of plant height, tilling, heading time, and leaf color were observed among the progeny of 54 T₀ plants. Sequencing revealed that 32 of the 54 plants harbored the correct target sgRNA vectors and had been edited by the CRISPR/Cas9 system (Supplemental Table 5). For example, 11 individual T₁ plants from the T₀ plant LBM0048 had increased tiller numbers, reduced heights and twisted leaves (Figure 1C, Supplemental Table 6), and Sanger sequencing showed that these T₁ mutants were tad₁ homozygotes, and their phenotype was consistent with that of previously described mutants (Xu et al., 2012). Similarly, four T₁ plants from T₀ plant LBM1027 had damaged leaves with white variegated areas and cell death (Figure 1C, Supplemental Table 6). Further analysis revealed that these plants were noe₁ homozygotes, although their phenotype had previously only been seen when noe₁ mutants were cultivated under strong light (Lin et al., 2012). Also six T₁ plants from T₀ plant LBM0221 had alternating green and yellow cross bands on the leaf blades (Figure 1C, Supplemental Table 6), and were identified as OsCRTISO homozygotes, in perfect agreement with a previous report (Fang et al., 2008).

These examples indicate that genotype and phenotype are much more closely related in the present large-scale CRISPR/Cas9 knock out mutant rice library than in previous mutant resources. Because there is an almost one-to-one relationship between guide sequence and targeted gene among these CRISPR/Cas9 mutants, the CRISPR/Cas9 mutant library requires less laborious mapping to identify interesting genes. Moreover, independent mutant alleles of the same gene can be recognized by their similar phenotypes, so avoiding tedious complementation analysis.

Finally, we tested the off-target effects of the rice mutants, which is a main concern for
the CRISPR/Cas9 system. We computationally predicted the genome-wide potential off-target sites for LBM0048, LBM0221 and LBM1027 using the Cas-OFFinder (Bae et al., 2014) tool in rice. Eight likely off-target sites with three nucleotide mismatches to the recognition site of LBM1027-sgRNA were identified; similarly, five and nine off-target sites were screened for LBM0221 and LBM0048-sgRNA, respectively (Supplemental Table 7). Using sequencing, the off-target effects were measured for T₀ mutants, and the results revealed that none of these sites was mutated among the LBM0048, LBM0221 and LBM1027 mutants (Supplemental Table 7). These data suggested the low off-target effects of the CRISPR/Cas9 mutant library, and the reason may be the careful sgRNA strict filtering and rational design in the beginning.

In summary, we have constructed a large-scale CRISPR/Cas9 mutant library in rice that is of high quality, with good coverage and uniform distribution. Our study demonstrates that CRISPR/Cas9-based screening is a robust method for systematically identifying both genes and mutant phenotypes in rice. The mutant library developed here will be of great value for the study of gene functions in rice and for crop improvement.

FUNDING
This work was supported by grants from the National Key Research and Development Program of China (2016YFD0101800) and the National Natural Science Foundation of China (91635301, 31301087).

AUTHOR CONTRIBUTIONS
X. M., X. S., F. Z., and S. G. performed the experiments; Y. H. conducted the bioinformatics analyses; X. M., Y. H., Y. Z., C. G. and J. L. designed the experiments and wrote the paper.

ACKNOWLEDGEMENTS
No conflict of interest declared.
FIGURE LEGENDS

Figure 1. Construction of a genome-wide mutant library in rice using CRISPR/Cas9.

(A) Schematic illustration of the CRISPR/Cas9 rice library construction.

(B) Results of large scale sequencing of the sgRNA library.

(C) Morphological phenotypes of the CRISPR/Cas9 rice mutants. The T1 mutant plants showing abnormal phenotypes growing at a normal field condition in Beijing. Bar = 24 cm.

REFERENCES


A

Elution → PCR

PCR

Gibson ligation

Purified DNA

PCR product

CRISPR/Cas9 vector library

Mutant plants

B

0 read
1-10 reads
11-90 reads
91-200 reads
>200 reads

82.07%

12.28%

1.23%

1.06%

3.36%

C

Zhonghua 11 (WT)

LB M0048 (tal allele)

LB M1027 (nie allele)

LB M0221 (Oscrtiso allele)