A Method for Producing Multiple-gene Mutations in a Gene Family that Utilizes sgRNA-target Mismatch at the Site Directly 5’ of the PAM

Yan Zhang, Miaomiao Qin, Chunbo Miao, Shenkui Liu

State Key Laboratory of Subtropical Silviculture, School of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou, 311300, China

Joint first authors: Y.Z. and M.Q. contributed equally to this work

* Contributed equally to this work

Corresponding author email: shenkuiiu@nefu.edu.cn; chunbomiao@163.com

Molecular Soil Biology, 2020, Vol.11, No.4 doi: 10.5376/msb.2020.11.0004

Received: 12 Jul., 2020
Accepted: 23 Jul., 2020
Published: 26 Aug., 2020

Copyright © 2020 Zhang et al., This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:


Abstract The CRISPR-Cas9 system uses a small single-guide RNA (sgRNA) to direct the endonuclease Cas9 to recognize and cleave the target DNA. Previous studies in bacteria and mammal cells showed that perfect base pairing between sgRNA and target sites within 10–12 bp directly 5’ of the protospacer adjacent motif (PAM) determined the Streptococcus pyogenes Cas9 (SpCas9, the most widely used Cas9) specificity, whereas mismatches in the other region can be tolerated. Here, through gene editing of rice DST and MIR156 genes, we found that CRISPR-SpCas9 with a mismatch at 20th base (counting from the distal of PAM) of the target sites could still lead to high-frequency gene editing in rice. In rice, these results enrich the rules of target site selection for SpCas9 and will be conducive to the emergence of multiple mutations in the gene family.

Keywords CRISPR-Cas9; SgRNA; Gene family; Multiple-gene mutations

Introduction

Prior to development of the clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated (Cas) technology, researchers used sequence-specific nuclease technologies represented by zinc-finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN) for gene editing in a variety of organisms (Urnov et al., 2010; Fu et al., 2013). Compared with ZFN and TALEN, the CRISPR-Cas9 technology provides a simpler, faster and highly scalable method for in vivo gene editing and is thus more extensively used. The CRISPR-Cas system is an adaptive immune defense system developed by bacteria and archaea over a long period of evolution and can be used against invading viruses and exogenous DNA (Bhaya et al., 2011; Terns and Terns, 2011). The CRISPR-Cas9 system originates from type II CRISPR-Cas systems and has been widely used for animal and plant gene editing. In the CRISPR-Cas9 system, a single guide RNA (sgRNA), a chimeric short single-strand RNA derived from the CRISPR RNA (crRNA) and the transactivating crRNA (tracrRNA), specifically binds to the target DNA through base pairing, and Streptococcus pyogenes Cas9 nuclease (SpCas9) is recruited by the sgRNA to the target sites (immediately upstream of a required protospacer adjacent motif (PAM)) with a matching 20-nt guide sequence within the sgRNA (Ran et al., 2013b; Doudna and Charpentier, 2014).

Studies on CRISPR-Cas9 specificity have shown that the perfect matching of 10-12 bp directly 5’ of the PAM between the sgRNA and the target sites helps guarantee the targeting specificity (Ran et al., 2013a; Heckl et al., 2014; Jiang and Marraffini, 2015). In bacteria, the efficient cleavage of target DNA by SpCas9 requires perfect matching between the sgRNA (or crRNA) and the target site in the 13-bp region directly upstream of the PAM, and a single mismatch in the six nucleotides immediately upstream of the PAM can abrogate SpCas9 targeting (Jinek et al., 2013; Jiang et al., 2013). In mammalian cells (human and mouse cells), the SpCas9 specificity is determined mainly by the 8–12-bp region proximal to the PAM, and an sgRNA with a single-base mismatch within this range can hardly direct Cas9 targeting (Cong et al., 2013; Hsu et al., 2013). During the
CRISPR-Cas-mediated editing of the *Arabidopsis* genome, mismatches between the sgRNA and the target DNA at any base within 12 bp directly upstream of the PAM also disrupt SpCas9 targeting to the target gene (Feng et al., 2014).

Although CRISPR-Cas9 technology has been widely used, the specificity of Cas9 targeting in plants, including rice, has seldom been studied. At present, the induction of multiple-gene mutations in a gene family is achieved mainly by ligating several sgRNAs on the same vector backbone or by performing hybridization experiments at the flowering stage. However, the disadvantage is that the mutation probability is low and the production period is long. Here, through editing of the rice *DST* gene, we found that CRISPR-Cas9 with a mismatch at the 20th base (counting from distal to the PAM) of the target sites could still lead to high-frequency gene editing in rice. We suspected that our finding might be beneficial to the emergence of multiple-gene mutations in a gene family, therefore, we utilized CRISPR-Cas9 to mutate the *MIR156* genes in order to obtain the mutations such as mir156a/b/c. The results showed that single- or multiple-gene mutations occurred only if the last base from the 5'- to the 3'-terminal proximal to the PAM showed a mismatch. Multiple-gene mutations in the gene family play an important role in gene function analysis. Our research provides an alternative for the emergence of multiple-gene mutations.

1 Materials and Methods

1.1 Plant materials
ZH11 and XS134 seeds were used in this study to obtain calli. Seeds were provided by C.M. from State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University.

1.2 Target site selection
The sequence of *DST* target site: 5'-'CGTGACGAGACATCCAC-3'; the sequence of *MIR156a* target site: 5'-CAGAAGAGATGACACACAG-3'; the sequence of *MIR156b* target site: 5'-CAGAAGAGGTGACAC-3'; the sequence of *MIR156c* target site: 5'-CAGAAGAGATGACACACA-3'.

1.3 Recombinant vector construction
Recombinant vectors were constructed based on the pCAMBIA1301 backbone and the U6 promoter. The sequence of sgRNA1: 5'-GGUGACGAGACCAUCCAC-3'; the sequence of sgRNA2: 5'-GUGACGAGACCAUCCAC-3'; the sequence of sgRNA3: 5'-GAGAAGAGAGUGAGACACAG-3'; the sequence of sgRNA4: 5'-GAGAAGAGAGUGAGACAC-3'.

1.4 Rice genetic transformation and identification of rice transgenic seedlings
Using an *Agrobacterium*-mediated method, the recombinant vectors were electroporated into EHA105 cells respectively, we referred to previous research on rice transgenic process (Chakraborty et al., 2016). Than hygromycin-resistant rice transgenic seedlings were identified and then grown in a greenhouse.

1.5 Sequencing and counting
Eighty-two transgenic seedlings were used for sequencing edited *DST* genes, the prime sequences: FDST: 5'-CAAGGAGCGTTCCGGCTGT-3'; RDST: 5'-CAGCGCGGCAAACCACGCT-3'. And 123 transgenic seedlings were used for sequencing the edited *MIR156a*, *MIR156b* and *MIR156c* genes, the prime sequences: Fa: 5'-GTCAGAATTACGAGGATG-3', Ra: 5'-ATGCTATGTACAGTGGAG-3'; Fb: 5'-GATTCGGCCCTATTCTTTCG-3', Rb: 5'-GTGTTGTGGTTTGGCCGTGGAG-3'; Fc: 5'-TACCGTACCATCTCCTCCAC-3'; Rc: 5'-AAATGAGCGAGCATCAGC-3'.

2 Results
2.1 Editing of the rice *DST* gene with sgRNA1 matched to the target site, sgRNA2 mismatched to the target site
The drought and salt tolerance (*DST*) gene negatively regulates leaf blade width and grain number per main
tiller to regulate drought and salt tolerance in rice (Huang et al., 2009; Li et al., 2013). We selected one target site in DST and designed two sgRNAs (sgRNA1 and sgRNA2) to target this site: sgRNA1 shows perfect matching to the target site, and sgRNA2 exhibits a mismatch with the target site at the 20th base directly 5’ of the PAM (Figure 1A).

Using an Agrobacterium-mediated method, two vectors expressing sgRNA1 and sgRNA2 were transformed into the calli of japonica rice variety Zhonghua11 (ZH11). Hygromycin-resistant seedlings from the T0 generation were then sequenced, and the results revealed that sgRNA1 and sgRNA2 induced 82.50% (33/40) and 30.95% (13/42) mutations in DST (Figure 1B).

2.2 Editing of the rice MIR156 gene family with sgRNA3 matched to the target site of MIR156a and sgRNA4 matched to the target site of MIR156b/c

We suspected that our finding that CRISPR-Cas9 with a mismatch at the 20th base (counting from distal to the PAM) of the target sites could still lead to high-frequency gene editing in rice might be beneficial to the emergence of multiple-gene mutations in a gene family. Therefore, we utilized CRISPR-Cas9 to mutate the MIR156 genes, which modulate the plant architecture and grain size to determine the grain yield. 11 gene loci express miR156 in rice (Jiao et al., 2010; Miura et al., 2010). And in MIR156 gene family, the target sites with a mismatch at the 20th base (counting from distal to the PAM) only can be found in MIR156a, MIR156b and MIR156c. We designed an sgRNA (sgRNA3) to target MIR156a and another sgRNA (sgRNA4) to target MIR156b and MIR156c (Figure 2A) because the gene sequences of MIR156a, MIR156b and MIR156c are highly similar. The target sites for sgRNA3 and sgRNA4 showed only one base difference at the 20th base (5’ to 3’) in the pairing region (Figure 2A).

Figure 1 Mutation frequency (%) of sgRNA1 and sgRNA2
Note: (A) SpCas9 target sites in DST; “*”: mismatched bases between the sgRNA and the target sites; “.”: perfectly matched bases between the sgRNA and the target sites; The bases that differ between sgRNA1 and sgRNA2 are highlighted in red; (B) Mutation frequency in DST induced by sgRNA1 and sgRNA2

Using an Agrobacterium-mediated method, two vectors expressing sgRNA3 and sgRNA4 were transformed into the calli of japonica rice variety Xiushui134 (XS134). The hygromycin-resistant seedlings from the T0 generation were then sequenced, and the results showed that the base mismatch in the MIR156a/b/c gene not only does not eliminate the activity of Cas9 but also might cause multiple-gene mutations in other genes (Figure 2B; Figure 2C). The highest mutation rate (58.8% and 51.5%) was obtained with the MIR156abc triple mutation, followed by different double-mutation and single-mutation combinations (Figure 2B; Figure 2C).

Figure 2 Mutation frequency (%) of sgRNA3 and sgRNA4
3 Discussion

The results of sgRNA1 and sgRNA2 inducing 82.50% (33/40) and 30.95% (13/42) mutations in DST gene editing confirmed that CRISPR-Cas9 with a mismatch at the 20th base (counting from distal to the PAM) of the target sites could still lead to high-frequency gene editing in rice. The results of editing the rice MIR156 gene family with sgRNA3 and sgRNA4 indicated that CRISPR-Cas9 could induce not only a higher mutation efficiency but also single- or multiple-gene mutations in gene family only if the last base from the 5’- to the 3’-terminal proximal to the PAM showed a mismatch. Multiple-gene mutations account for a large proportion.

3.1 Drawing flow chart

I: To summarize the above experimental process, we have drawn the following flow chart. We selected this type of target site with a mismatch only at the 20th base (counting from distal to the PAM) in some genes of the gene family as the target sites. Our method works only when this type of target sites are present.
II: SgRNA-target mismatch only at the site directly 5’ of the PAM.
III: Agrobacterium-mediated method.
IV: Taking the leaves, extracting the DNA, and designing the primers for semi-quantitative PCR experiments.
V: Designing the primers and sequencing positive transgenic seedlings.
VI: Counting probability of different mutations, obtaining multi-spotted seedlings, providing materials for further research (Figure 3).
Figure 3 Outline of the experiment

Note: “*”: Mismatched bases between the sgRNA and the target sites; “.”: Perfectly matched bases between the sgRNA and the target sites

4 Conclusions
In this study, we analyzed the effect of a mismatch at the 20th base (counting from distal to the PAM) of the target sites on mutation rates in rice genes and found that CRISPR-Cas9 with a mismatch at this site can still lead to high-frequency gene editing. These results will be conducive to the emergence of multiple mutations in a gene family if a PAM site with a mismatch only at the 20th base in some genes of the gene family can be found. However, we did not explore the effects of mismatches at other sites proximal to the PAM in rice, which might deserve more attention.

Abbreviations
CRISPR: Clustered regularly interspaced short palindromic repeats
ZFN: Zinc-finger nucleases
TALEN: Transcription activator-like effector nucleases
SgRNA: Single guide RNA
CrRNA: CRISPR RNA
PAM: Protospacer adjacent motif
DST gene: The drought and salt tolerance gene

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.
Accession numbers

The sequence data for the MIR156a, MIR156b, and MIR156c genes can be found in the miRbase database (http://www.mirbase.org/) under the following accession numbers: MIR156a, MI0000653; MIR156b, MI0000654; and MIR156c, MI0000655. The DST sequence can be found in the rice annotation project database (https://rapdb.dna.affrc.go.jp/) under the gene locus identifier Os03g0786400.

Authors’ contributions

Y.Z. and C.M. designed the study and constructed all the vectors, and M.Q. performed the transgenic experiments. M.Q. and Y.Z. analyzed the data and wrote the main manuscript. C.M. and S.L. provided opinions on the paper. All authors have read and approved the manuscript. Y.Z. and M.Q. contributed equally to this work.

Acknowledgments

The study was supported by the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (Grant No. 2016C02056-1).

References

https://doi.org/10.1146/annurev-genet-110410-132430
PMid:22060043

https://doi.org/10.1007/s12298-015-0334-y
PMid:27186018 PMcid:PMC4840146

https://doi.org/10.1126/science.1231143
PMid:23287718 PMcid:PMC3795411

https://doi.org/10.1126/science.1258096
PMid:25430774

https://doi.org/10.1073/pnas.1408022111

https://doi.org/10.1038/nbt.2623
PMid:23792628 PMcid:PMC3773023

https://doi.org/10.1038/nbt.2951
PMid:24952903 PMcid:PMC4160386

https://doi.org/10.1038/nbt.2647
PMid:23873081 PMcid:PMC3969858

https://doi.org/10.1101/gad.1812409
PMid:19651988 PMcid:PMC2720257

https://doi.org/10.1038/nbt.2508
PMid:23360965 PMcid:PMC3748948
https://doi.org/10.1146/annurev-micro-091014-104441
PMid:26209264

https://doi.org/10.1038/ng.591
PMid:20495565

https://doi.org/10.1126/science.1225829
PMid:22745249 PMcid:PMC6286148

https://doi.org/10.1073/pnas.1300359110
PMid:23382237 PMcid:PMC3581943

https://doi.org/10.1038/ng.592
PMid:20495564

https://doi.org/10.1016/j.cell.2013.08.021
PMid:23992846 PMcid:PMC3856256

https://doi.org/10.1038/nprot.2013.143
PMid:24157548 PMcid:PMC3969860

https://doi.org/10.1016/j.mib.2011.03.005
PMid:21531607 PMcid:PMC3119747

https://doi.org/10.1038/nrg2842
PMid:20717154