The length of guide RNA and target DNA heteroduplex effects on CRISPR/Cas9 mediated genome editing efficiency in porcine cells

Jiawei Lv, Shuang Wu, Renyue Wei, Yan Li, Junxue Jin, Yanshuang Mu, Yu Zhang, Qingran Kong, Xiaogang Weng, Zhonghua Liu

Key Laboratory of Animal Cellular and Genetic Engineering of Heilongjiang Province, College of Life Science, Northeast Agricultural University, Harbin 150030, China

ABSTRACT

The clustered regularly interspaced short palindrome repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) system is a versatile genome editing tool with high efficiency. A guide sequence of 20 nucleotides (nt) is commonly used in application of CRISPR/Cas9; however, the relationship between the length of the guide sequence and the efficiency of CRISPR/Cas9 in porcine cells is still not clear. To illustrate this issue, guide RNAs of different lengths targeting the EGFP gene were designed. Specifically, guide RNAs of 17 nt or longer were sufficient to direct the Cas9 protein to cleave target DNA sequences, while 15 nt or shorter guide RNAs had loss-of-function. Full-length guide RNAs complemented with mismatches also showed loss-of-function. When the shortened guide RNA and target DNA heteroduplex (gRNA:DNA heteroduplex) was blocked by mismatch, the CRISPR/Cas9 would be interfered with. These results suggested the length of the gRNA:DNA heteroduplex was a key factor for maintaining high efficiency of the CRISPR/Cas9 system rather than weak bonding between shortened guide RNA and Cas9 in porcine cells.

Keywords: CRISPR-Cas systems; guide RNA; heteroduplex; genome editing; swine

INTRODUCTION

The clustered regularly interspaced short palindrome repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) system is a highly efficient tool for genome editing that has been widely used in a variety of species [1-4]. In addition to its applications, researchers always focus on optimization of the CRISPR/Cas9 system, including the activity of Cas9 nuclease [5], design of the guide RNA [6,7], and reaction conditions [8]. Guide RNA is 1 part of the CRISPR/Cas9 system that consists of a scaffold sequence and guide sequence [9]. The guide sequence is designed according to study requirements and is a key variable in its applications [10].

For application of the CRISPR/Cas9 system, a guide sequence must be designed and synthesized [11]. Regardless of the number of guide sequences [12] being designed, a length of 20 nt is usually adopted, and the guide is commonly adjacent to the protoscaler adjacent motif NGG. The length of the guide sequence is a limitation in guide sequence design. Although some studies investigated the impact of the length of guide RNAs on efficiency of the CRISPR/
Conflict of Interest
The authors declare no conflicts of interest.

Author Contributions
Conceptualization: Lv J, Mu Y, Liu Z; Data curation: Lv J, Liu Z; Formal analysis: Wei R; Funding acquisition: Liu Z, Mu Y; Investigation: Lv J, Wu S, Li Y; Methodology: Lv J, Jin J; Project administration: Lv J, Liu Z; Resources: Mu Y, Weng X; Software: Kong Q; Supervision: Liu Z; Validation: Zhang Y; Visualization: Lv J; Writing - original draft: Lv J, Jin J, Zhang Y; Writing - review & editing: Jin J, Liu Z.

Cas9 system [13,14], it is still not clear if the same impact could be applied in porcine cells and no studies have shown how truncated guide RNA sequences influence the efficiency of Cas9. If the relationship between the length of the guide sequence and the efficiency of CRISPR/Cas9 can be clarified, shortened guide RNAs will be applied extensively and the cost of guide sequence synthesis could be decreased, especially in guide RNA library synthesis.

The pig is important to the human food supply and an ideal model for various human diseases [15]. The CRISPR/Cas9 system has been used to generate genome editing pigs [16-18], and is also regarded by virologists as a promising method for the suppression of viral infection in pigs [19,20]. In the present study, we investigated the relationship between the length of the guide sequence and the efficiency of CRISPR/Cas9 in porcine cells. Additionally, we tried to explore the mechanism of how the shortened guide RNA affects the efficiency of the CRISPR/Cas9 system in genome editing.

MATERIALS AND METHODS

Cell culture
PK15 cells and PKpG cells were cultured in Dulbecco’s modified eagle medium (DMEM; Gibco, USA) with 15% fetal bovine serum (FBS; Hyclone, USA), 1% non-essential amino acid (Gibco), 1% penicillin-streptomycin (Gibco), and 2 mmol/L L-glutamine (Sigma, USA). Cells were passaged every 2 day with 0.25% trypsin-ethylenediaminetetraacetic acid (Gibco).

Transfection and selection of enhanced green fluorescent protein (EGFP) positive clone
PK15 cells were cultured to 70–90% confluency, after which pEGFP-C1 plasmid and Lipofectamine 2000 reagent (Invitrogen, USA) were added to the media. Samples were then incubated for 3 day, at which time cells were passaged at a 1:50 ratio. Media were changed to DMEM with 10% FBS and 800 µg/mL G418 (Sigma). After 9 days selected by G418, single cell clones were observed and 1 clone was selected, which all cells expressed EGFP in the chosen clone. This clone was picked and digested into single-cell status by 0.25% trypsin. After 0.25% Trypsin treated transiently, cells were transferred into DMEM with 15% FBS and subcultured. After subcultured and proliferated, the cell line expressing EGFP stably was obtained.

Lentiviral production and infection
Five kinds of plasmids were co-transfected into HEK293T cells with the packaging plasmids psPAX2 and pMD2.G (3:2:1 ratio) (Addgene plasmid 12260 and 12259). The process of transfection was conducted in accordance with the manufacturer protocols specified for the Lipofectamine™ LTX Reagent with PLUS™ Reagent (Invitrogen). After 72 h, viral supernatants were filtered through a 0.45 µm low protein binding membrane (Millipore) and concentrated using an Ultra-15 Centrifugal Filter (100,000 NMWL; Millipore).

When PKpG cells were cultured to be 30–40% confluent, media were changed to DMEM with 10% FBS after which lentiviruses and 0.5% polybrene were added to the media to infect PKpG cells. After 24 h, the media was changed to DMEM with 15% FBS.

Fluorescence-activated cell sorting (FACS) assay
Cells were washed with Dulbecco’s phosphate-buffered saline (Gibco) after dissociation, then subjected to FACS (FACSMelody; BD Biosciences, USA) to separate cells expressing

https://vetsci.org
https://doi.org/10.4142/jvs.2019.20.e23
2/9
Guide RNA length affects the genome editing efficiency

EGFP from those that did not. PK15 cells were sorted as negative controls to build an EGFP negative gate.

**Western blot**
Cells were lysed with 1% Triton X-100 (Sigma) and then subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis, after which proteins were transferred to the polyvinylidene difluoride membrane (Millipore). EGFP primary antibody (ab32146; Abcam) and secondary antibody (A9169; Sigma) were used to detect EGFP, while beta-actin primary antibody (A1978; Sigma) and secondary antibody (A9044; Sigma) were used to detect beta-actin as a loading control. An ECL Chemiluminescent Substrate Reagent Kit (Invitrogen) was used to detect blots of proteins.

**RESULTS**

**Design and strategy**
Guide sequences 19 nt, 17 nt, 15 nt, and 13 nt in length targeting the same site of the *EGFP* gene were designed and synthesized (Fig. 1A). These guide sequences were cloned into the lentiCRISPRv2 vector between the U6 promoter and the guide RNA scaffold (Supplementary Fig. 1). Additionally, we designed a 20 nt guide sequence with no target site in the pig genome as the control. To test the efficiency of these guide RNAs for gene knockout, the porcine PK15

---

**Fig. 1.** Design of the guide RNAs and strategy for CRIPR/Crispr-associated protein 9 system efficiency verification. (A) Schematic of shortened guide RNA. Guide sequences 19 nt, 17 nt, 15 nt, and 13 nt in length targeting the same site of *EGFP* gene (residues Gly35–Tyr39). (B) Fluorescence images of the PKpG cell line. (C) Percentage of EGFP positive cells in each group (n = 3). The FACS results showed that the PKpG cell line achieved EGFP positive rates of 99.7%. PK15 cells were tested as a negative control. (D) The process of EGFP knockout in PKpG cells. The PKpG cell line was the main material used in our study. At 3 day post-infection, puromycin was added and PKpG cells were selected for 5 day. The remaining cells were cultured until 12 day post-infection. A series of tests would be executed 12 day after infection.

CRISPR, clustered regularly interspaced short palindrome repeats; EGFP, enhanced green fluorescent protein; FACS, fluorescence-activated cell sorting; PAM, protospacer adjacent motif; FITC-A, fluorescein isothiocyanate A.
cell line was transfected with pEGFP-C1 plasmid (Supplementary Fig. 2A). After being selected by G418, we chose 1 clone that expressed EGFP protein stably (Supplementary Fig. 2B) and cultured 1 cell line denoted PKpG (Fig. 1B). FACS revealed that there were 99.7% EGFP positive cells in the PKpG cell line (Fig. 1C); therefore, this line was used as a rigorous material in this study (Fig. 1D). After lentivirus packaging, we obtained 5 types of lentivirus. PKpG cells were then infected, selected with puromycin, and subjected to a series of tests 12 day after infection.

**High efficiency of CRISPR/Cas9 with 19 nt and 17 nt length guide RNAs**

Twelve days after infection, there was an obvious difference between experimental groups and the control group. Specifically, there was an increase in the number of EGFP negative cells in the 19 nt and 17 nt-EGFPgRNA groups compared with control group, but not in the 15 nt and 13 nt-EGFPgRNA groups (Fig. 2A). These results showed that Cas9 nuclease

![Fig. 2.](https://vetsci.org)

Fig. 2. The efficiency of CRISPR/Cas9 with guide RNAs of different length. (A) Fluorescence images of each group. EGFP protein (green signal) observed at 488 nm and nuclei (blue signal) were marked by Hoechst 33342. Null, the group without any treatment. PK15, wild type PK15 cells. (B) Percentage of EGFP negative cells in each group (n = 3). The FACS results showed EGFP negative rates of each group. PK15 cells were tested as a negative control. (C) Western blotting analysis of the expression of EGFP protein. β-actin were tested as a loading control. (D) and (E) identification of exogenous DNA integrated into genome. Guide RNA region was amplified with a pair of primers: 5′-TAAGGTACCGAGGGCCTATTTC-3′ and 5′-CCAATTCCACTCCTTTCAAGAC-3′. M, DL5000 DNA marker. Guide RNA region polymerase chain reaction amplicon was sequenced. (F) Mutant sequences at the target site generated by Cas9-mediated genome manipulation in 19 nt and 17 nt-EGFPgRNA groups. Red letters are presented protospacer adjacent motif.

CRISPR, clustered regularly interspaced short palindromic repeats; EGFP, enhanced green fluorescent protein; Ins. Mut., insert mutant; Del. Mut., delete mutant.
could contribute to EGFP knockout with a guide of 19 nt and 17 nt-EGFPgRNA. The FACS assay showed that the rates of EGFP negative cells of the 19 nt-EGFPgRNA group and the 17 nt-EGFPgRNA group were 90.23% ± 3.66% and 90.28% ± 3.51%, respectively, which were significantly higher than that of the control group (0.10% ± 0.07%; p < 0.001). However, the rates of EGFP negative cells in the 15nt-EGFPgRNA group and 13nt-EGFPgRNA group were 0.20% ± 0.01% and 0.31% ± 0.17%, respectively (Fig. 2B and Supplementary Fig. 3). Western Blot demonstrated that there was almost no expression of EGFP in the 19 nt-EGFPgRNA group and the 17nt-EGFPgRNA group (Fig. 2C). The guide RNA region was confirmed by Sanger sequencing of the PCR amplicon (Fig. 2D and E), which revealed that the mutation contained insertions and deletions (Fig. 2F) that altered the reading frame. Taken together, these results confirmed that the 19 nt and 17 nt guide RNAs were sufficient to direct Cas9 protein to sequence-specifically cleave target DNA sequences, while the 15 nt and 13 nt long guide RNAs resulted in loss-of-function.

**Length of the gRNA:DNA heteroduplex influences the efficiency of CRISPR/Cas9**

It has been reported that guide sequences containing transversion mismatches could have reduced efficiency [21]. Interestingly, our results showed that there is no function of Cas9 in the 15 nt-EGFPgRNA group. It might be different reason between shortened guide RNA and normal guide RNA with mismatches. No function of Shortened guide RNA might result from low efficiency of guide RNA and Cas9 combine or guide RNA and target DNA combine. To determine if either the length of guide sequence base-pairing with the DNA target or the length of the guide RNA is the key factor influencing the combination of Cas9 and the target site, we designed three 18 nt guide sequences targeting the same site of the **EGFP** gene, including single, double and triple mismatches at the 5’ end (denoted 1-M, 2-M, and 3-M, respectively; Fig. 3A). The results of the FACS assay showed that there was an increase in the number of EGFP negative cells in the 1-M group (Fig. 3B and C); however, there was no significant difference among the 2-M, 3-M, and control group. Based on these results, we inferred the length of the guide sequence base-pairing with the DNA target influenced the efficiency of the CRISPR/Cas9 system during genome editing.

**Shortened guide RNAs with mismatches interfere with the efficiency of the CRISPR/Cas9 system**

To further verify our results, we designed another guide sequence based on the 1-M group, which was denoted 1-M* (Fig. 4A). When compared with the guide sequence of the 1-M group, there was another mismatch in the guide sequence of the 1-M* group. However, the results of the FACS assay showed that there was no significant difference between the 1-M* group and the control group (Fig. 4B and C), while there was a significant difference between the 1-M* group and the 1-M group, suggesting the difference resulted from the additional mismatch. These findings verified our inference that the excessively shortened guide RNA interfered with the efficiency of CRISPR/Cas9 system via the reduction in the length of the gRNA:DNA heteroduplex.

**DISCUSSION**

In this study, we demonstrated that the length of guide RNA has a significant effect on the efficiency of gene knockout in the CRISPR/Cas9 system in PK15 cells. The efficiency of CRISPR/Cas9 was not be affected by gene knockout when the length of the guide RNA was 17 nt or longer; however, it was influenced when the guide RNA was shortened to 15 nt or less.
This influence was complete and resulted in Cas9 nuclease having no effect on the targeting site. Moreover, these effects were caused by changes in the length of the gRNA:DNA heteroduplex. As shown in our study, the shortest guide RNA with normal function could be used for reducing off-target efficiency, because the function could be influenced once there were 1 or more mismatches in the shortest guide RNA. It explained why the shortened guide RNA could reduce the off-target effects [13].

In addition to revealing the relationship between the length of guide RNA and the efficiency of gene knockout, our study provided data that will facilitate the design of guide RNA. With a valid length range, more guide RNA candidates can be chosen for Cas9-mediated genome editing in porcine cells. Furthermore, the cost can be reduced by shortening the guide RNA, especially with respect to guide RNA libraries construction, without reducing the gene knockout efficiency. Indeed, the cost can be decreased by 15% (3/20) when the 17 nt guide RNA library is used instead of the 20 nt library. Therefore, the shortened guide RNA can be
extensively used in application of the CRISPR/Cas9 system. To date, all guide RNA libraries have used only 1 species or cell line [12, 22]. Although a sharable guide RNA library could theoretically be constructed by synthesizing all of the potential guide sequences, the large cost makes it impossible. However, the present study shows that the cost can be reduced greatly using the shortened guide RNA.

Guide sequences containing mismatches have been shown to reduce the efficiency of CRISPR/Cas9-mediated gene knockout [21]. Interestingly, our results showed that Cas9 nuclease had no function in the 15nt-EGFPgRNA group instead of low efficiency gene knockout. Although our results indicated that the length of the gRNA:DNA heteroduplex influences the efficiency of Cas9, the large difference between the 17 nt and 15 nt-EGFPgRNA group implies that the 2 nucleotides might be the key sites in the guide RNA-Cas9 complex. The structure of Cas9 had been reported [9] and the backbone phosphate groups of the guide sequence (nucleotides 15–17) are known to interact with the REC1 domain (Lys510 and Tyr515). Based on the available data, it can be speculated that the REC1 domain of Cas9 may be essential to its ability to recognize and cleave target DNA.

**ACKNOWLEDGMENTS**

We thank Chaqian Jiang, Jiaqiang Wang and Leyun Wang for their technical assistance. We also acknowledge the assistance of all members of Lab217.
SUPPLEMENTARY MATERIALS

**Supplementary Fig. 1**
Structure of the lentiCRISPRv2 vector. Guide RNA was cloned into the lentiCRISPRv2 vector between the U6 promoter and guide RNA scaffold. spCas9 and puromycin selection marker are linked by P2A.

Click here to view

**Supplementary Fig. 2**
Production of the PKpG cell line. (A) PK15 cells were transfected using pEGFP-C1 plasmid. After selection by G418, 1 clone stably expressing EGFP protein was selected and cultured as the PKpG cell line. (B) Fluorescence images of EGFP expressing clone.

Click here to view

**Supplementary Fig. 3**
Representative histograms of EGFP fluorescence from each group. PK15 cells were tested as a negative control to construct a gate.

Click here to view

REFERENCES

1. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. Science 2013;339:819-823.

2. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. Science 2013;339:823-826.

3. Tao L, Yang M, Wang X, Zhang Z, Wu Z, Tian J, An L, Wang S. Efficient biallelic mutation in porcine parthenotes using a CRISPR-Cas9 system. Biochem Biophys Res Commun 2016;476:225-229.

4. Voytas DF. Editorial prerogative and the plant genome. J Genet Genomics 2016;43:229-232.

5. Wright AV, Sternberg SH, Taylor DW, Staahl BT, Bardeles JA, Kornfeld JE, Doudna JA. Rational design of a split-Cas9 enzyme complex. Proc Natl Acad Sci U S A 2015;112:2984-2989.

6. Farboud B, Meyer BJ. Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. Genetics 2015;199:959-971.

7. Zhang H, Zhang X, Fan C, Xie Q, Xu C, Zhao Q, Liu Y, Wu X, Zhang H. A novel sgRNA selection system for CRISPR-Cas9 in mammalian cells. Biochem Biophys Res Commun 2016;471:528-532.

8. Xiang G, Zhang X, An C, Cheng C, Wang H. Temperature effect on CRISPR-Cas9 mediated genome editing. J Genet Genomics 2017;44:199-205.

9. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 2014;156:935-949.

10. Xie S, Shen B, Zhang C, Huang X, Zhang Y. sgRNAcas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. PLoS One 2014;9:e100448.
11. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera MC, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol 2014;32:267-273.

12. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 2014;343:84-87.

13. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol 2014;32:279-284.

14. Zhang JP, Li XL, Neises A, Chen W, Hu LP, Ji GZ, Yu JY, Xu J, Yuan WP, Cheng T, Zhang XB. Different effects of sgRNA length on CRISPR-mediated gene knockout efficiency. Sci Rep 2016;6:28566.

15. Whyte JJ, Prather RS. Genetic modifications of pigs for medicine and agriculture. Mol Reprod Dev 2011;78:879-891.

16. Hai T, Teng F, Guo R, Li W, Zhou Q. One-step generation of knockout pigs by zygote injection of CRISPR/Cas system. Cell Res 2014;24:372-375.

17. Peng J, Wang Y, Jiang J, Zhou X, Song L, Wang L, Ding C, Qiu J, Liu L, Wang W, Liu J, Huang X, Wei H, Zhang P. Production of human albumin in pigs through CRISPR/Cas9-mediated knockin of human cDNA into swine albumin locus in the zygotes. Sci Rep 2015;5:16705.

18. Sato M, Kagoshima A, Saitoh I, Inada E, Miyoshi K, Ohtsuka M, Nakamura S, Sakurai T, Watanabe S. Generation of α-1,3-galactosyltransferase-deficient porcine embryonic fibroblasts by CRISPR/Cas9-mediated knock-in of a small mutated sequence and a targeted toxin-based selection system. Reprod Domest Anim 2015;50:872-880.

19. Lillico SG, Proudfoot C, King TJ, Tan W, Zhang L, Mardjuki R, Paschon DE, Rebar EJ, Urnov FD, Mileham AJ, McLaren DG, Whitelaw CB. Mammalian interspecies substitution of immune modulatory alleles by genome editing. Sci Rep 2016;6:21645.

20. Whitworth KM, Rowland RR, Ewen CL, Trible BR, Kerrigan MA, Cino-Ozuna AG, Samuel MS, Lightner JE, McLaren DG, Mileham AJ, Wells KD, Prather RS. Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. Nat Biotechnol 2016;34:20-22.

21. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science 2016;351:84-88.

22. Bassett AR, Kong L, Liu J. A genome-wide CRISPR library for high-throughput genetic screening in Drosophila cells. J Genet Genomics 2015;42:301-309.