RESEARCH ARTICLE | APRIL 09 2019

Development of CRISPR/Cas9 plasmid for multiple sites genome editing in oil palm (*Elaeis guineensis* Jacq.) FREE

Victor Aprilyanto; Chris Darmawan; Condro Utomo ✉; Tony Liwang

AIP Conf. Proc. 2099, 020002 (2019)
https://doi.org/10.1063/1.5098407

Development of CRISPR/Cas9 Plasmid for Multiple Sites Genome Editing in Oil Palm (*Elaeis guineensis* Jacq.)

Victor Aprilyanto, Chris Darmawan, Condro Utomo[a)], and Tony Liwang

Biotechnology Department, Plant Production and Biotechnology Division, PT SMART Tbk.
Jl. Raya Cijayanti, Bogor, West Java, 16810, Indonesia.

[a)]Corresponding author: condro.utomo@sinarmas-agri.com

Abstract. Genome editing technology via CRISPR/Cas9 system is a versatile technique with numerous potential applications, particularly in agriculture. In this study, we attempted to develop a CRISPR/Cas9 plasmid containing four sgRNA to allow multiple gene editing in the oil palm genome. In the first step, we used an in silico approach to finding the optimum 20-nt guides from four gene regions across oil palm genome. These guides were later joined with a promoter and tracr-RNA fragment to construct a 472 bp module, and together with three tetranucleotide linkers and restriction sites at both terminals gave an insert of length 1 918 bp. This insert was then incorporated into CRISPR/Cas9 vector, and the final plasmid was sequence validated.

Keywords: *Elaeis guineensis* Jacq., CRISPR/Cas9, genome editing, golden gate assembly, sgRNA.

INTRODUCTION

The utilization of CRISPR/Cas9 system in genome editing has gained widespread application in various fields such as medicine and agriculture [1–4]. The versatility of this system has been applied from knockout or edit a gene, regulate gene expression, until creating single nucleotide polymorphisms [5]. In plant genome editing, two delivery methods of CRISPR/Cas9 system currently applied are particle bombardment and *Agrobacterium* transformation. Current development of *Agrobacterium*-mediated CRISPR/Cas9 plasmid transformation contains only one single guide RNA (sgRNA) which targets a single locus in the target genome. It is therefore advantageous if such plasmid could be developed to target more than one locus.

With more discovery of genes responsible for oil palm agro-economical traits, editing one locus in the genome might not give the expected outcome if several genes govern the trait. *Agrobacterium*-mediated transformation rarely permits the entry of multiple plasmids into plant cells [6], further increasing technical difficulties of plasmid transformation into oil palm. One strategy is to develop a modified CRISPR/Cas9 system consisting of several target modules. Each module will consist of an expression promoter, a 20-nt specific guide, and tracr-RNA sequence. These modules will be joined next to each other to create one long insert containing multiple modules. Normally, the joining of such modules would involve restriction sites which often will be left intact in the joined insert if ordinary type-II restriction enzymes were employed. It could halt further ligations particularly if same restriction sites were co-applied on different modules. Golden Gate Assembly/GGA [7] is a method to assemble DNA fragments without introducing restriction sites in the fragment joints. This technique utilizes a mix of *BsaI* restriction enzyme and T4 DNA ligase. *BsaI* is a type IIs restriction enzyme which cleaves DNA at a defined distance from their non-palindromic asymmetric recognition sites [8]. The cleavage of *BsaI* will leave four-base overhangs which could be anchored to other fragments via base overhang complementarity. It means that the technique will also preserve fragment order during the assembly of all joined fragments.

1st International Conference on Bioinformatics, Biotechnology, and Biomedical Engineering (BioMIC 2018)
AIP Conf. Proc. 2099, 020002-1–020002-6; https://doi.org/10.1063/1.5098407
Published by AIP Publishing. 978-0-7354-1829-5/$30.00

This study attempted to develop a CRISPR/Cas9 system containing an insert of joined four sgRNA modules. Each module would consist of a promoter followed by a sgRNA template. Our strategy employed amplification of each promoter-sgRNA module and joined these modules into one large insert using GGA. The insert was cloned, and the resulting CRISPR/Cas9 plasmid was validated for in parallel would be validated through sequencing.

MATERIALS AND METHODS

Design of sgRNA and Optimization

A case of oil palm genome editing was used in this study. We used four genes associated with agro-economical traits of oil palm namely, palmitoyl-acyl thioesterase, *EgPATE* (GenBank ID DQ422858 [9]), virescens, *EgVIR* (GenBank ID KJ789862 [10]), lipase *EgLIP* (GenBank ID HE661587 [11]), and shell thickness, *EgSHE* (GenBank ID XM_010909778 [12]) genes. We took the first exon from *EgPATE* and *EgLIP*, third exon from *EgVIR* and second exon from *EgSHE* as the target regions for CRIPSR-Cas9 system. Using a simple python script, we listed all the possible 20-nucleotide guides preceding NGG protospacer adjacent motif (PAM) from both forward and reverse strands. All these 20-nt guides were concatenated with 78-nt tracrRNA sequence and subjected to mFold program [13] to predict any RNA secondary structures and possibility of self-base pairing from each guide. Based on the prediction result and additional criteria of structural self-base pairing and GC content proposed by Xu *et al*. [14] and Liang *et al*. [15], we then selected one best 20-nt guide candidate from each gene to be incorporated into CRISPR/Cas9 plasmid.

Insert Amplification and Assembly

A proposal of multi-target CRISPR/Cas9 plasmid design would contain an insert of joined four modules in the order of *EgPATE-EgVIR-EgLIP-EgSHE*. For insert ligation, to the plasmid backbone, the 5' and 3' ends of this insert should possess *EcoR1* and *HindIII* restriction sites, respectively. Joining four modules, *BsaI* restriction sites ($GGTCTCN_5$) would have to be added to the ends of each module, except the 5'-end *EgPATE* and 3'-end of *EgSHE* since they will harbor *EcoRI* and *HindIII* sites, respectively. *BsaI* restriction sites were employed to join one fragment to another with the respective four free base-linker: AGGT for *EgPATE* to *EgVIR* modules, CAAG for *EgVIR* to *EgLIP* modules, and GCCT for *EgLIP* to *EgSHE* modules. All these restriction sites were introduced into the modules using PCR amplification.

Amplification of each single guide CRISPR/Cas9 plasmid was conducted using a set of primer (Table 1) to obtain 472 bp module containing promoter-sgRNA region (Fig. 1). For PCR amplification, Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, USA) was used. The PCR composition comprised of Q5 2X mastermix 12.5 μL, 10 μM forward primer 1.25 μL, 10 μM reverse primer 1.25 μL, template DNA to final concentration 1 ng · μL^{-1}, and water to total volume of 25 μL. PCR condition was 30 s pre-denaturation at 98 °C followed by 30 cycles of 10 s denaturation at 98 °C, 10 s annealing at 55 °C, and 20 s extension at 72 °C and then 2 min post-extension at 72 °C. Amplicons from each plasmid was purified. Golden Gate assembly (GGA) technique [7] was later employed to seamlessly joining one fragment towards another. According to this scheme, the total length of the insert should be 1 918 bp which consists of four 472 bp fragments, three tetranucleotide linkers, one six-bases *EcoRI* site and another six-bases *HindIII* site each extended by three bases.

Each amplified module was joined one to another in a one-step reaction using Golden Gate Assembly Mix (New England Biolabs, USA) under the following composition: 2 μL Golden Gate buffer, 1 μL Golden Gate Assembly Mix, 2 μL of each module at 40 ng · μL^{-1} concentration and water to bring the total volume reaction to 20 μL. The reaction condition was 30 cycles of 5 min incubation at 37 °C followed by 5 min incubation at 16 °C and then followed by a further 5 min incubation at 55 °C. The result of GGA was run in 1 % agarose gel where a band of size ~2 kb (representing 1 918 bp insert) was gel-extracted.

Cloning and Transformation

The gel-extracted 1 918 bp insert was then ligated to pJET 1.2/Blunt vector as a part of CloneJET PCR cloning kit (Thermo Fisher Scientific, USA) and the whole plasmid was transformed to *Escherichia coli* strain DH5α. The selection was conducted by growing the cells in ampicillin-containing (50 mg · L^{-1}) medium. The plasmids were re-

extracted from the growing cells for validation through sequencing. The validated insert was digested from pJET and ligated to CRISPR/Cas9 vector. This CRISPR/Cas9 plasmid was then transformed to *Escherichia coli* strain DH5α competent cells and further selected in medium containing kanamycin (50 mg · L^{-1}). The selected cells then had their plasmids isolated and verified for the presence of 1 918 bp insert through sequencing.

TABLE 1. List of primers used in module construction.

Module	Forward Primer (5' → 3')	Reverse Primer (5' → 3')
EgPATE	AGTGAATTCGACCAAGCCCGTTATT	AGTGGTCTCGACCTGATCTGAAAAAAAGC
EgVIR	AGTGGTCTCGAGGTGACCAAGCCCGTTAT	AGTGGTCTCGCTTGGATCTGAAAAAAAGC
EgLIP	AGTGGTCTCGCAAGGACCAAGCCCGTTAT	AGTGGTCTCGAGGCGATCTGAAAAAAAGC
EgSHE	AGTGGTCTCGGCCTGACCAAGCCCGTTAT	AGTAAGCTTGATCTGAAAAAAAGCA

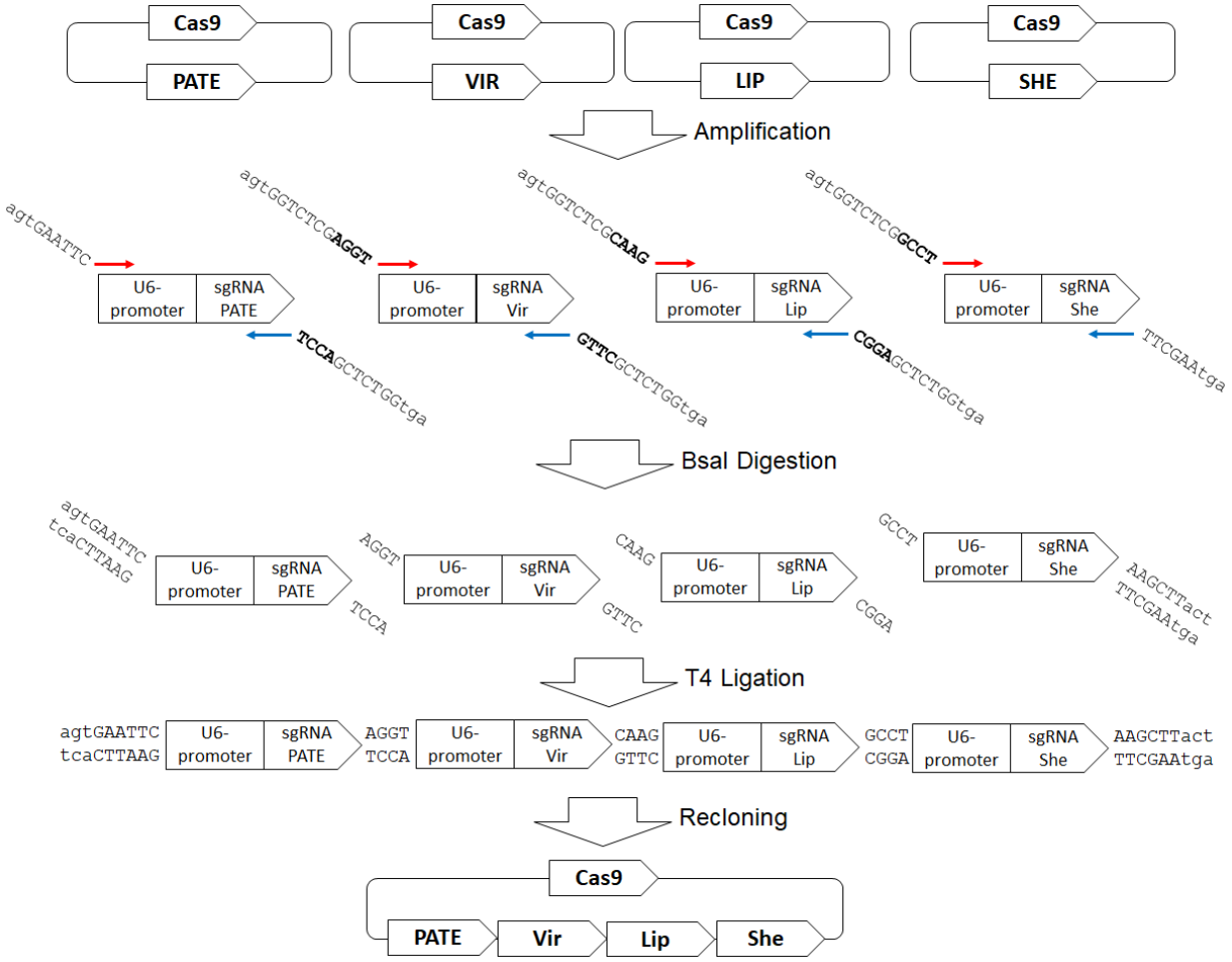


FIGURE 1. Construction steps of multiple sgRNA CRISPR-Cas9 plasmids which involves fragment amplification, assembly, and recloning steps. The assembly step consists of BsaI restriction and T4 ligation which is conducted in a single reaction of GGA kit. Capital base-letters indicate restriction sites while the boldface base-letters indicate four free-base linkers

RESULTS AND DISCUSSION

Design of sgRNA and Optimization

In total 150 sgRNAs had been listed down from four oil palm gene fragments. *EgPATE* exon 1 and *EgVIR* exon 3, the two longest sequences among the four, made up for more than 80 % of total sgRNA (Table 2). However, it is likely that the number sgRNAs which contained no self-base pairing have no direct correlation with total

sgRNA. As shown in Table 2, mFold predicted that only 6.25 % of total sgRNA in *EgVIR* contains no self-base pairing in their first 5' twenty nucleotide bases. EgLIP matched this number although its sequence length was 132 bp shorter than *EgVIR*. *EgSHE* contained the highest fraction of sgRNAs with no self-base pairing, mainly due to its short in length. Since the self-base pairing is a structural phenomenon which involves adjacent bases in a sequence, the order of bases should be considered in calculating the probability of loop formation within the sgRNA sequence. Although choosing which 20nt-guide to be used in genome editing is rather subjective, some studies have given several criteria in choosing a potentially optimum guide. Liang et al. [15] outlined that an optimum 20nt guide for plant genome editing should have a 30 % to 80 % GC content with a maximum of six nucleotides involved in self-base pairing. To further optimize the 20-nt guides, we selected the only sgRNA which does not possess self-base pairing among its 5' 20-nt sequence for module construction (Fig. 2). We did so to anticipate the potential of loop formation which might interfere with the cleavage ability of sgRNA-Cas9 complex.

TABLE 2. sgRNAs from four oil palm genes.

Target	Region	Size (bp)	#guides	#non self-bp guides	Selected Guide Sequence	% GC
EgPATE	Exon 1	513	74	11 (14.86 %)	CATAGCCGCCGAAGAAGAGG	60
EgVIR	Exon 3	442	48	3 (6.25 %)	TCTGACCATAACCATTGCAA	40
EgLIP	Exon 1	310	23	3 (13.04 %)	AGCCTCGTACTCCTCCATTA	50
EgSHE	Exon 2	82	5	2 (40 %)	TGCTCTTTTATACCTCTCGA	40

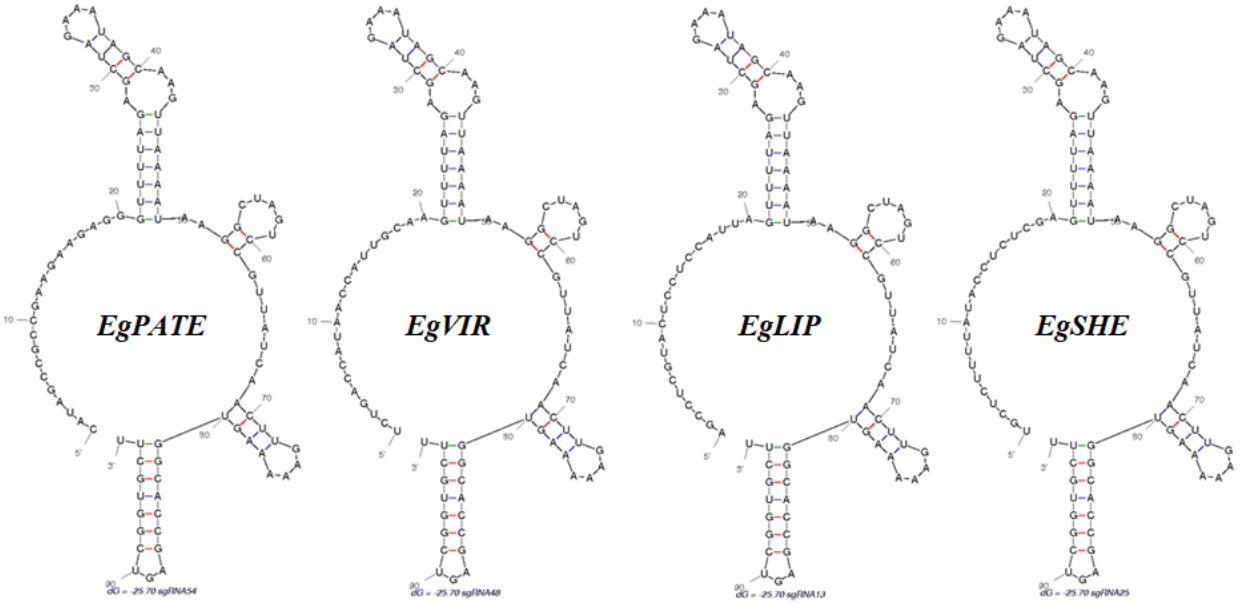


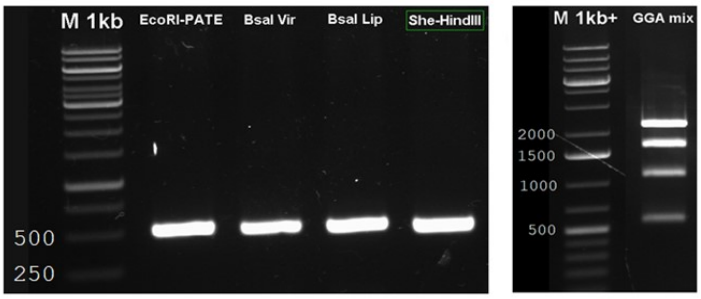
FIGURE 2. The predicted secondary structure of sgRNAs targeting for *EgPATE*, *EgVIR*, *EgLIP*, and *EgSHE* genes

Fragment Amplification and Assembly

For module construction, each of the selected 20-nt guide sequences was combined with 363 bp U6 promoter and 89 bp tracrRNA template sequence. Some joint PCR were used to create a 472 bp U6-20guide-tracrRNA module which was later inserted into a vector. PCR amplification using a set of primer (Table 1) resulted in a 472 bp product from each vector containing *BsaI* restriction sites (Fig. 3a).

Further assembly using GGA kit showed that a longer incubation was needed to increase the concentration of assembled products, which might be related to enzyme efficiency. In our earlier attempt, we followed the suggested assembly protocol for one to four inserts assembly from the manual (NEB E1600 manual) which stated an incubation for one hour in 37 °C and then followed by 5 min incubation in 55 °C. However, this protocol gave a sub-optimal result since the thicker DNA band was still in the ~500 bp size, suggesting that most products were still in the one-module state. The employment of longer incubation with cycles between 37 °C and 16 °C greatly

increases the ~2 kb-size DNA band (Fig. 3b) which is the product of joined four ~500 bp modules. The latter condition favors module assembly might be due to the two above values are the optimum working temperature for *BsaI* and T4 ligase, respectively.



(a) (b)

FIGURE 3. Fragments of size 472 bp amplified from each vectors containing *EgPATE* (EcoRI-PATE), *EgVIR* (BsaI Vir), *EgLIP* (BsaI Lip), and *EgSHE* (She-HindIII), respectively. (b) Golden Gate assembly result of the four 472 bp fragments

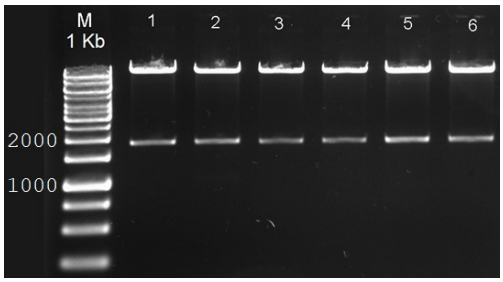


FIGURE 4. (lane 1–6) *EcoRI-HindIII* digestion of CRISPRPL-AMH plasmid gives a product of ~2 kb size which corresponds to a 1 918 bp insert.

AGT**GAATTC**GACCAAGCCCGTTATTCTGACAGTTCTGGTGCTCAACACATTTATATTTATCAAGGAGCACATTGTTACTCACTGCTAGGAGGGAATC
GAACTAGGAATATTGATCAGAGGAACTACGAGAGAGCTGAAGATAACTGCCCTCTAGCTCTCACTGATCTGGGTCGCATAGTGAGATGCAGCCCACG
TGAGTTCAGCAACGGTCTAGCGCTGGGCTTTTAGGCCCGCATGATCGGGCTTTTGTCGGGTGGTCGACGTGTTCACGATTGGGGAGAGCAACGCAGC
AGTTCCTCTTAGTTTAGTCCCACCTCGCCTGTCCAGCAGAGTTCTGACCGGTTTATAAACTCGCTTGCTGCATCAGACTTG**CATAGCCGCCGAAGAA**
GAGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCAGATC**AGGT**
GACCAAGCCCGTTATTCTGACAGTTCTGGTGCTCAACACATTTATATTTATCAAGGAGCACATTGTTACTCACTGCTAGGAGGGAATCGAACTAGGA
ATATTGATCAGAGGAACTACGAGAGAGCTGAAGATAACTGCCCTCTAGCTCTCACTGATCTGGGTCGCATAGTGAGATGCAGCCCACGTGAGTTCAG
CAACGGTCTAGCGCTGGGCTTTTAGGCCCGCATGATCGGGCTTTTGTCGGGTGGTCGACGTGTTCACGATTGGGGAGAGCAACGCAGCAGTTCCTCT
TAGTTTAGTCCCACCTCGCCTGTCCAGCAGAGTTCTGACCGGTTTATAAACTCGCTTGCTGCATCAGACTTG**TCTGACCATAACCATTGCAA**GTTTT
AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCAGATC**CAAG**GACCAAGCC
CGTTATTCTGACAGTTCTGGTGCTCAACACATTTATATTTATCAAGGAGCACATTGTTACTCACTGCTAGGAGGGAATCGAACTAGGAATATTGATC
AGAGGAACTACGAGAGAGCTGAAGATAACTGCCCTCTAGCTCTCACTGATCTGGGTCGCATAGTGAGATGCAGCCCACGTGAGTTCAGCAACGGTCT
AGCGCTGGGCTTTTAGGCCCGCATGATCGGGCTTTTGTCGGGTGGTCGACGTGTTCACGATTGGGGAGAGCAACGCAGCAGTTCCTCTTAGTTTAGT
CCCACCTCGCCTGTCCAGCAGAGTTCTGACCGGTTTATAAACTCGCTTGCTGCATCAGACTTG**AGCCTCGTACTCCTCCATTA**GTTTTAGAGCTAGA
AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCAGATC**GCCT**GACCAAGCCCGTTATTCT
GACAGTTCTGGTGCTCAACACATTTATATTTATCAAGGAGCACATTGTTACTCACTGCTAGGAGGGAATCGAACTAGGAATATTGATCAGAGGAACT
ACGAGAGAGCTGAAGATAACTGCCCTCTAGCTCTCACTGATCTGGGTCGCATAGTGAGATGCAGCCCACGTGAGTTCAGCAACGGTCTAGCGCTGGG
CTTTTAGGCCCGCATGATCGGGCTTTTGTCGGGTGGTCGACGTGTTCACGATTGGGGAGAGCAACGCAGCAGTTCCTCTTAGTTTAGTCCCACCTCG
CCTGTCCAGCAGAGTTCTGACCGGTTTATAAACTCGCTTGCTGCATCAGACTTG**TGCTCTTTTATACCTCTCGA**GTTTTAGAGCTAGAAATAGCAAG
TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCAGATC**AAGCTT**ACT

Legend:
ATGC = EcoRI restriction site
ATGC = HindIII restriction site
ATGC = 20-nt guide sequence
ATGC = U6 promoter
ATGC = tracr-DNA sequence
ATGC = tetranucleotide linker

FIGURE 5. Sequencing result of the *EcoRI-HindIII* digested CRISPR/Cas9 plasmid containing joined four-modules insert. Note the four 20-nt guides (underlined letters) which consist the order of *EgPATE-EgVIR-EgLIP-EgSHE*.

It is likely that the utilization of PCR-generated amplicons instead of precloned fragments was the main cause of this assembly inefficiency. Moreover, extending only three bases long on either end of the module was too short for *BsaI* to cleave the fragment ends, thus making the overall amplicon-based assembly rather inefficient. Therefore, we suggest that extending to five or six bases on both ends would be reasonable to achieve efficient assembly reaction. More studies should be addressed to confirm these hypotheses. Ligation of the ~2 kb-size insert into CRISPR/Cas9 vector was confirmed using *EcoRI-HindIII* restriction, which showed a single band of ~2 kb in size (Fig. 4). The sequencing result confirmed that the insert sequence and order was in an exact match with the design (Fig. 5) and thus validated that the multi-target CRISPR/Cas9 plasmid has been successfully developed.

CONCLUSION

In this study, an *Agrobacterium*-mediated CRISPR/Cas9 plasmid containing insert joined from four sgRNA modules have been successfully assembled. It would enable further studies to employ a similar technique in order to create multi-targeted CRISPR/Cas9 plasmid to perform gene silencing or editing in plants.

REFERENCES

1. M-R. Mahmoudian-sani, G. Farnoosh, A. Mahdavinezhad, and M. Saidijam, Biotechnol. Biotechnol. Equip. **32**(2), 286–292 (2018).
2. L. Cai, A. L. Fisher, H. Huang, and Z. Xie, Genes & Diseases **3**(4), 244–251 (2016).
3. X. Liu, S. Wu, and J. Xu, Acta Pharm. Sin. B **7**(3), 292–302 (2017).
4. L. Bortesi and R. Fischer, Biotechnol. Adv. **33**(1), 41–52 (2015).
5. L. G. Lowder, D. Zhang, N. J. Baltes J. W. Paul, X. Tang, X. Zheng, D. F. Voytas, *et al*., Plant Physiol. **169**, 971–985 (2015).
6. M. Dafny-Yelin and T. Tzfira, Plant Physiol. **145**, 1118–1128 (2007).
7. C. Engler, R. Kandzia, and S. Marillonnet, PLoS ONE **3**(11), e3647 (2008).
8. W. Szybalski, S. C. Kim, N. Hasan N, and A. J. Podhajska, Gene **100**, 13–26 (1991).
9. A. Othman, C. Lazarus, T. Fraser, and K. Stobart, Biochem. Soc. Trans. **28**(6), 619–622 (2000).
10. R. Singh, E. L. Low, L. C. Ooi, O. -A. Meilina, R. Nookiah, N. -C. Ting, M. Marjuni, *et al*., Nat. Commun. **5**, 4106–4113 (2014).
11. F. Morcillo, D. Cros, N. Billote, G. -F. Ngando-Ebongue, H. Domenhedo, M. Pizot, T. Cuellar, *et al*., Nat. Commun. **4**, 2160 (2013).
12. R. Singh, E. L. Low, L. C. Ooi, O. -A. Meilina, N. -C. Ting, J. Nagappan, R. Nookiah, *et al*., Nature **500**, 340–344 (2013).
13. M. Zuker, Nucleic Acids Res. **31**(13), 3406–3415 (2003).
14. H. Xu, T. Xiao, C. Chen W. Li, C. A. Meyer, Q. Wu, D. Wu, *et al*., Genome Res. **25**, 1147–1157 (2015).
15. G. Liang, H. Zhang, D. Lou, and D. Yu, Sci. Rep. **6**, 21451 (2016).