Targeted gene modification in *Fragaria vesca* mediated by the CRISPR/Cas9 system

Carmen Martín-Pizarro[1], David Posé[1]

1. *Instituto de Hortofruticultura Subtropical y Mediterránea, Universidad de Málaga–Consejo Superior de Investigaciones Científicas, Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, 29071 Málaga, Spain.*

Introduction

Genome editing is becoming an important biotechnological tool for gene function analysis and crop improvement, with the CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeat-CRISPR associated protein 9) system being the most widely used mechanism. The natural CRISPR/Cas9 system has been reduced to two components: a single-guide RNA (sgRNA) for target recognition via RNA-DNA base pairing, which is commonly expressed using a promoter for small-RNAs (U6 promoter), and the Cas9 endonuclease for DNA cleavage (1).

To optimize the CRISPR/Cas9 system for strawberry plants, we generated constructs using a U6 promoter from the woodland strawberry (*Fragaria vesca*). We also engineered the coding sequence for the *Streptococcus pyogenes Cas9* endonuclease (*SpCas9*) by modifying it to match the plant codon usage frequencies. To validate the CRISPR/Cas9 system in *F. vesca*, we designed sgRNAs directed against the floral homeotic gene *APETALA3.* This gene was selected because *ap3* mutations induce clear developmental phenotypes in which petals and stamens are missing or partially converted to sepals and carpels, respectively (2) .

Results

1. Selection of the U6 promoter from *Fragaria vesca* and plant codon usage optimization of *Streptococcus pyogenes Cas9* endonuclease (*SpCas9*).

(A)

```
AtU6-26      ---AACTTTCCATTCGGAGTTTTTGTATCTTGTTTCATAGTTTGTCCCAGGATTAGAATG
FvU6-IV      ----------CGAAACACAATCTTA--CTCATGATCAGTTCCAAGCAGCTAA-GACCAAAC
FvU6-I       GTTTTAGTTTCCTTGAGAATGTAG--GTGTTCGGC-TTTCCCAGAAGCCA---TGGGATA
FvU6-V       CTCCTTATTTCACTCACTT----T--CAAGTCGTA-TTTCCACATCTCAACTGTACAATG
FvU6-II      ----TGATATCAAGTGTGGTAACA--TTAATAATC--TGCGCAGTGACCAT--ATGAAAA
FvU6-III     -----AATTTCACTCTCTCTCT--CTCTCTCT---CTCTCTCTCTCTCT--CTCAT-G
                                                                *

AtU6-26      ATTAGGCATCGAACCTTCAAGAATTTGATTGAATAAAACATCTTCATTC------TTAAG
FvU6-IV      TTTTGTGACAGGGATTTAAAGCTCTTGA----TTCCTAAATCCAAACAGAAAGAACAAAA
FvU6-I       TATAAGAACACAGCTGGATTGGGTTTGA----ATCCTAAAAATTAAAAGGA----TTA--
FvU6-V       TCTAAAAC-ACAGTGTGTAGATCATTCA----TACAAAGTTCGTGTAAAGA----TTATT
FvU6-II      CAAGAGGAAGAACGATTCAGTACTACAGATGAAATTGATAATGTTGACCGTAAATTGAGA
FvU6-III     TATCAAAGCATACGATTCTCTACTTTGTTGTTTCCTG-AATTCAAGATCTAGGATATCTG

AtU6-26      ATATGAAGATAATCTTCAAAAGGCC--CCTGGGAATC------TGAAAGAAGAGAAGCAG
FvU6-IV      AGAAGGGGATGATTTCCATCAGAAGCCATTGGTCATGGTGAAGTTACGGTTACTTGTCTA
FvU6-I       -GAG---TAAAAAGAAAGAAAAGAAAAACAAGCTGGG------CTAAATATATACTTGAA
FvU6-V       CAATGTATATTAACTTGGACAGGTGCCGTAAGCAGTC------AAAATCTGTTCTGTTTTT
FvU6-II      GGATCCAAAACAGTATTACCAAGTCAGATGACAAGGT--------TT-GAACTGATTGTT
FvU6-III     AGATCCATGCCATTTCGAATTTGACCAAGGCCCTGTT--------AT-GACTTTATTG-T
               *      *

AtU6-26      GCCCATTTATATGGGAAAGAACAATAG-TATTTCTTA----------TATAGGCCCATTT
FvU6-IV      TCCCCTT-CAATAGTCCCATACAATTGCGCTTTGAGAGAGACGACACGTGAGGCCCACGT
FvU6-I       GCCCAAT-ACTT---AAACCTTACTA--CTACTTTCATAGAAGACACGTGAGGCGCACGT
FvU6-V       AAACCTT-CATT---CTTCTAGAATT--GCACTTAAAGGGAGGACACGTGACGCGCACAT
FvU6-II      TCTGCCTTATCTTTCTTAGGAGATT---GAACTTTGG--GGTCCAATTGTTAAA-----G
FvU6-III     GCTCAAATCTATAACAGATTGCAAA---CGCCTCTCA--GGCCCAAACAGTCCCCAACTC
                        *        *        *

AtU6-26      AA--GTTGAAAACAATCTTCAAAAGTCCCACATCGCTTAGATAAGAAAACGAAGCTGAGT
FvU6-IV      GC----------TTTTGTTATGATCCCACATCGCAGAAAAGCGAAAAC-AGAAGGCTG
FvU6-I       GCTTTTGTAAAATTGGATCCAAGACTCCCACATCGAAGAAAAGCGAAAAC-AAGAGGCTG
FvU6-V       GCTTTGGTAATATTAGTTCCAAGACTCCCACATCGAAGGA-AGCGAAAAC-AGAAAGCTG
FvU6-II      AATGACATACTTTCAATAGAAACAATCCCACATCGGAAATTAAACTAAGT-AAGCAGTCC
FvU6-III     TTAAAGAGCAAAGGCTACGAAATAATCCCACATCGGAAACCTCTGTCTAC-AAGGACTTC
                          *   *********        *

AtU6-26      TTATATACAGCTAGAGTCGAAGTAGTGATTGTCCCTTCGGGGACATCCGATAAAATTGGA
FvU6-IV      TTTATTTTGGAAGACATGACAGTAGAGCTCGTCCCTTCGGGGACATCCGATAAAATTGGA
FvU6-I       TTCATATAGGAAGACATGGCAGTAGAGCTTGTCCCTTCGGGGACATCCGATAAAATTGGA
FvU6-V       TTCATATATGAGAACATTGCAGTATAGTTCGTCCCTTCGGGGACATCCGATAAAATTGGA
FvU6-II      CTTAAATATTTTGGAATCACATGTATGTTTGTCCCTTCGGGGACATCCGATAAAATTGGA
FvU6-III     TTTATATACAATTGACTCCCATCTAAGCTTGTCCCTTCGGGGACATCCGATAAAATTGGA
               *        *   *      * * **********************************

AtU6-26      ACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCATAAATCGAGAAAT
FvU6-IV      ACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCATAAATCGAGAAAT
FvU6-I       ACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCATAAATCGAGAAAT
FvU6-V       ACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCATAAATCGAGAAAT
FvU6-II      ACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCATAAATCGAGAAAT
FvU6-III     ACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCATAAATCGAGAAAT
             ************************************************************

AtU6-26      GGTCCAAATTTTT
FvU6-IV      GGTCCAAATTTTT
FvU6-I       GGTCCAAATTTTT
FvU6-V       GGTCCAAATTTTT
FvU6-II      GGTCCAAATTTTT
FvU6-III     GGTCCAAATTTTT
             ************
```

USE motif
TATA box
U6 sRNA

AtU6-26 0.20422
FvU6-II 0.2237
FvU6-III 0.19339
FvU6-IV 0.16955
FvU6-I 0.1524
FvU6-V 0.1674

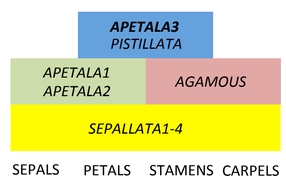
(B)

Aminoacid	Codon	Frequency hSpCas9	Frequency pSpCas9
ALA	GCT	2,9	25,0
	GCA	2,1	15,0
	GCC	50,7	8,6
	GCG	0,7	7,9
CYS	TGT	0,0	0,7
	TGC	1,4	0,7
ASP	GAT	13,6	47,8
	GAC	57,1	22,8
GLU	GAA	26,4	40,0
	GAG	50,7	37,1
PHE	TTT	8,6	22,8
	TTC	36,4	22,1
GLY	GGA	8,6	19,3
	GGT	0,7	17,1
	GGG	2,1	7,9
	GGC	40,0	7,1
HIS	CAT	0,7	14,3
	CAC	22,8	9,3
ILE	ATT	7,1	27,1
	ATC	60,0	23,6
	ATA	0,0	16,4
LYS	AAG	87,8	60,0
	AAA	27,1	55,0
LEU	CTT	0,0	35,0
	TTG	0,0	29,3
	CTC	2,1	18,6
	TTA	0,0	0,0
	CTG	103,5	11,4
	CTA	0,0	11,4
MET	ATG	15,7	15,7
ASN	AAT	9,3	25,7
	AAC	40,7	24,3
PRO	CCT	5,7	10,0
	CCA	2,1	9,3
	CCG	0,7	5,0
	CCC	18,6	2,9
GLN	CAA	1,4	21,4
	CAG	36,4	16,4
ARG	AGA	20,7	12,8
	AGG	3,6	16,4
	CGT	0,0	11,4
	CGA	0,0	0,0
	CGG	30,0	0,0
	CGC	2,1	15,7
SER	TCT	5,0	15,0
	TCA	0,0	10,7
	AGT	0,7	8,6
	AGC	33,5	7,1
	TCC	15,0	7,1
	TCG	0,0	5,7
THR	ACT	0,7	16,4
	ACA	6,4	14,3
	ACC	39,3	9,3
	ACG	0,7	7,1
VAL	GTT	0,0	22,1
	GTG	50,7	13,6
	GTC	2,9	10,0
	GTA	0,0	7,9
TRP	TGG	5,0	5,0
TYR	TAT	5,0	20,0
	TAC	34,3	19,3
END	TGA	0,0	0,7
	TAA	0,7	0,0
	TAG	0,0	0,0

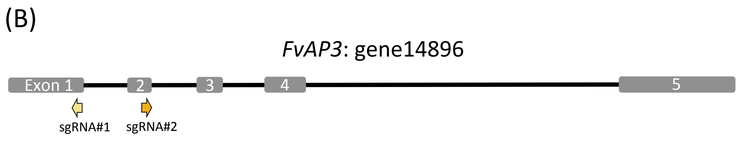
Fig. 1. (A) Alignment of the commonly used *U6-26* promoter from *Arabidopsis thaliana* (AtU6-26) (3) with the orthologous U6 promoters from *Fragaria vesca*. Yellow and red boxes: conserved promoter domains. Grey box: U6 sRNA sequence. The *FvU6-III* promoter was selected for our construct.
(B) Comparison of the codon usage frequency for the optimized human hSpCas9 endonuclease (4) with the pSpCas9 endonuclease for plants that we engineered for this study.

2. Target selection, construct design and transient expression.

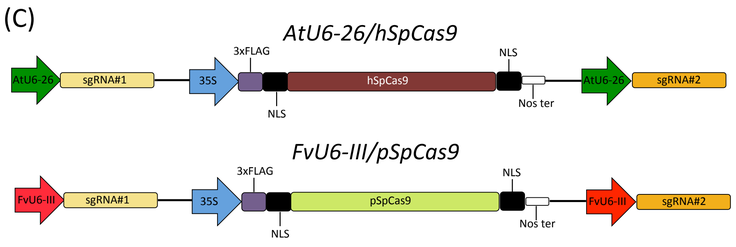
(A)



(B)



(C)

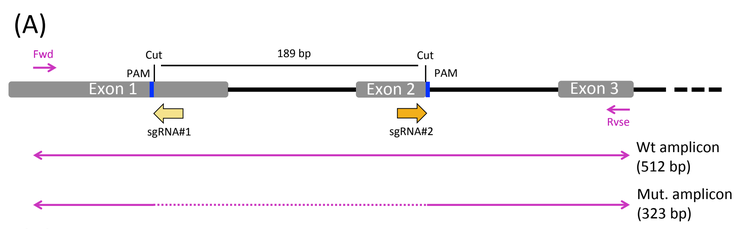


(D)

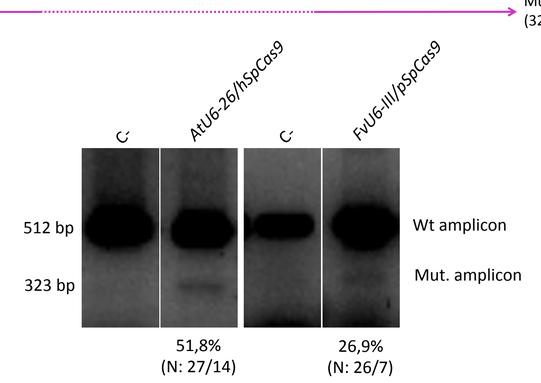
Fig. 2 (A). Arabidopsis ABCE model. The floral homeotic gene *APETALA3* was selected as the target for mutagenesis. (B) Gene structure of *FvAP3* (gene14896) and positions of the two sgRNAs designed for this work. Grey boxes: exons, black lines: introns. (C) Vectors used in this work. The first one (AtU6-26/hspCas9) expresses the two sgRNAs under the *AtU6-26* promoter and uses the human hSpCas9. The second vector (FvU6-III/pSpCas9) expresses the sgRNAs under the FvU6-III promoter and includes the engineered pSpCas9. (D) To induce transient expression, the vectors were transformed into *Agrobacterium tumefaciens* and infiltrated into *F. vesca* fruits using a syringe.

3. Mutagenesis (deletion) efficiency analyses by PCR and sequencing.

(A)



(B)



(C)

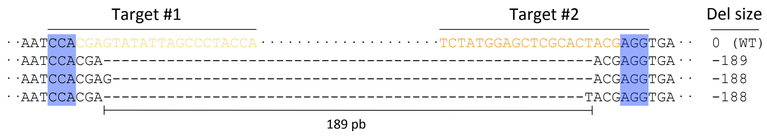


Fig. 3. (A) Diagram of the *FvAP3* gene. Orange arrows indicate the two targets (sgRNAs); blue square: PAM sequence. Cas9 endonuclease cuts 3 nts upstream of the PAM sequence. Cut sites are displayed. Purple arrows: primers used to amplify the region expected to be deleted. Expected amplicon sizes for the WT and the mutant are shown.
(B) Gel electrophoresis of PCR in independent fruits using the flanking primers shown in (A) for the two vectors and a negative control (C-; agro-infiltrated fruits with a vector harbouring *35S:Cas9* but without the sgRNA cassette). 512 bp band: wt allele; 323 bp band: mutant allele. This result shows that the CRISPR/Cas9 deletion system works during transient expression in *F. vesca* fruits, but there is limited transformation and/or mutagenesis efficiency, as evidenced by the strong presence of the wt band. Percentage of fruits with each specific amplification pattern and total number of fruits are shown at the bottom. A lower percentage of deletion was obtained with *FvU6-III/pSpCas9* vector.
(C) Sequences of the mutant amplicons from *AtU6-26/hSpCas9* fruits. Three different types of deletion were found. Cas9 cut either 3 or 4 nts upstream of the PAM sequences (blue box).

4. Transcript and protein levels for Cas9 variants.

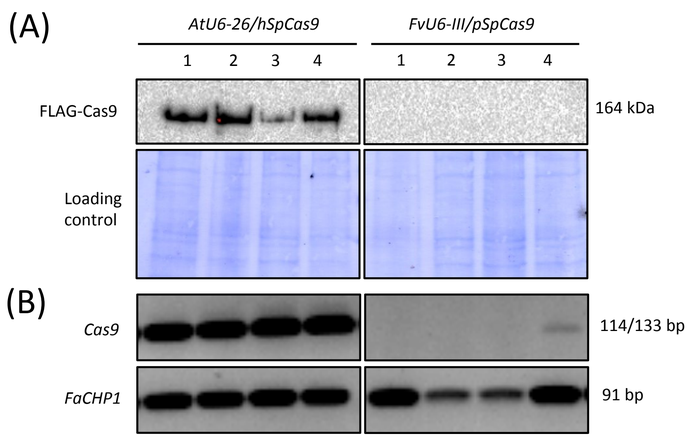


Fig. 4. (A) (*Top*) Detection of the Cas9 variant proteins by Western-blot analysis of the infiltrated fruits that showed a deletion in *FvAP3*. The hSpCas9 protein was detected, but the pSpCas9 protein was not. (*Bottom*) Loading control for nuclear protein extraction. (B) (*Top*) *Cas9* expression analysis by semiquantitative RT-PCR. The *hSpCas9* mRNA was detected, *pSpCas9* showed only weak expression in one sample. This result explains the absence of pSpCas9 protein. (*Bottom*) *FaCHP1* is shown as a loading control.

5. Expression analysis for U6 promoters

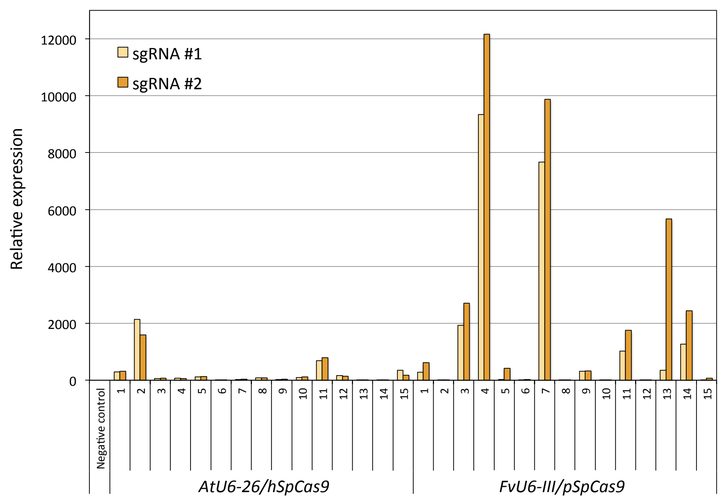


Fig. 5. RT-qPCR analysis of the expression of sgRNAs under the *AtU6-26* and *FvU6-III* promoters. 15 independent fruits infiltrated with each construct and a negative control were analysed. Stronger expression and a greater percentage of fruits expressing the sgRNAs were achieved using the *FvU6-III* promoter. *FaCHP1* was used as housekeeping gene.

Conclusions and perspectives

Our results show that the CRISPR/Cas9 mutagenesis system is functional in *F. vesca*. The mutagenic efficiency was higher for *AtU6-26/hSpCas9* than *FvU6-III/pSpCas9,* which we attribute to impaired *pSpCas9* transcript stability. Nonetheless, the *FvU6-III* promoter showed a higher expression level than *AtU6-26*, which will allow us to further optimize the system. We are currently establishing stable transgenic lines using both the *AtU6-26* and *FvU6-III* promoters to drive expression of the sgRNAs using only the *hSpCas9* endonuclease.

Our work offers a promising tool for genome editing and the functional analysis of genes in strawberry. This tool might represent a more efficient alternative to the sometimes inefficient RNAi silencing methods commonly used in this species.

References

1.Jinek M, et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 337(6096): 816-821.

2.Yang Y, et al (2003) pistillata-5, an Arabidopsis B class mutant with strong defects in petal but not in stamen development. Plant J. 33(1):177-188.

3.Cong L, et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science. 339(6121):819-823.

4.Feng Z, et al (2013) Efficient genome editing in plants using a CRISPR/Cas system. Cell. Res. 23(19):1229-1232.