A compendium of nonredundant short Polymerase III promoters for CRISPR applications

Michihito Deguchi, Kayla M. Sinclair, Annie Patel, Mckenna Coile, Maria A. Ortega, William P. Bewg, Chung-Jui Tsai

Warnell School of Forestry and Natural Resources, Department of Genetics, and Department of Plant Biology, University of Georgia, Athens, GA 30602

Corresponding author: Chung-Jui Tsai (cjtsai@uga.edu)

Abstract

Multiplex genome editing via CRISPR is a powerful tool for the simultaneous knockout, activation, and/or repression of distinct genes or noncoding sequences. However, current toolkits for multiplex editing lack diversity. Repeated use of the same promoter in multiple expression cassettes complicates construct assembly and has long been a concern for genetic stability in the host organism. Additionally, using unnecessarily long promoters may increase the genetic load and introduce uncertainties that impact CRISPR efficiency. To address these challenges, we present a collection of short Polymerase III (Pol III) promoters of diverse origins to support increasingly sophisticated genome editing applications in dicots.

Pol III promoters, such as U6 and U3, are widely used to express guide RNAs (gRNAs) in CRISPR studies. Published plant U6 and U3 promoters range from 79 bp to over 700 bp (Table S1), with the longer promoters being severalfold larger than the transcripts they control. Given that the upstream sequence element (USE, TCCCACATCG) and TATA box (TTTATATA) necessary for Pol III transcription are located close to the transcription start site (Waibel and Filipowicz, 1990), we reasoned that minimal Pol III promoters would be effective. To test this idea, we generated a deletion series of the experimentally validated *Medicago truncatula MtU6.6* promoter (352 bp, Zhou et al., 2015) for CRISPR-Cas9 editing of *mEGFP* (monomeric enhanced green fluorescent protein) in *mEGFP*-expressing wild tobacco (*Nicotiana benthamiana*) and hybrid aspen (*Populus tremula* × *P. alba* INRA 717-1B4). The 70 bp *MtU6.6* promoter was as functional as the 352 bp promoter, based on the loss of mEGFP signal in both root tips and leaf trichomes of wild tobacco (Figure 1a). Amplicon sequencing confirmed that all four *MtU6.6* promoter lengths (352, 189, 104, and 70 bp) had similar editing efficiencies in wild tobacco and hybrid aspen, whereas the promoterless control was ineffective (Figure 1b, Data S1).

We next investigated whether the 70 bp length is universally effective for diverse U6 and U3 promoters. We oligo-synthesized 14 short U6 or U3 promoters (Figure 1c) from *Arabidopsis thaliana*, chicory (*Cichorium intybus*), rubber tree (*Hevea brasiliensis*), apple (*Malus domestica*), poplar (*Populus trichocarpa*), and grapevine (*Vitis vinifera*) for binary vector assembly to test in wild tobacco. All but two (*AtU6.1* and *HbU6.1*, see below) showed efficient editing (Figure 1d). The *AtU6.1* promoter also failed to work in hybrid aspen, but the other four 70 bp promoters we tested were functional. These results suggest that the 70 bp length is sufficient for gRNA expression across a variety of dicot species, with some exceptions.

One of the exceptions, *AtU6.1*, was also ineffective in previous studies (Fan et al., 2015; Li et al., 2021), suggesting a genetic basis for defective Pol III promoters. Close examination revealed sequence variations within the USE of *AtU6.1* and *HbU6.1*, as well as the TATA box of *HbU6.1* (Figure 1c). To demonstrate the causality between USE and/or TATA sequence and promoter activity, we first utilized erroneous promoter clones from multiplex Gibson assembly of pooled

oligos. We found that 2 nt deletions within the USE (*AtU6.29c13*) or TATA (*PtU6.2c4*) abolished CRISPR editing, whereas a single base deletion in TATA (*PtU6.2c3*) was functional (Figure 1e). Mutations outside the conserved elements (*e.g.*, *AtU6.1c1*, *AtU6.29c14*, *CiU6.6c16*, and *PtU6.1c1*) did not impact promoter activity (Figure 1f). Next, we generated reciprocal mutations between the functional *MtU6.6* and the defective *AtU6.1* targeting the two variable positions within the USE. Mutation of the USE in *MtU6.6* (*MtU6.6m7*$_{C2A/T8G}$) to resemble *AtU6.1* resulted in a complete loss of promoter activity, whereas the reciprocal mutation of *AtU6.1* (*AtU6.1m2*$_{A2C/G8T}$) restored editing (Figure 1e). Furthermore, correction at the eighth position alone (*AtU6.1m3*$_{G8T}$) to resemble the USE of *PtU6.1* was sufficient to restore activity. These results suggest that the A/C wobble at the second base of USE is tolerated and that the T-to-G mutation at the eighth position (T8G) underlies the *AtU6.1* promoter defect. In the case of *HbU6.2*, mutations to restore USE (A10G) and TATA (C3T/G8A) conservation rescued promoter activity, whereas restoration of TATA alone did not (Figure 1e). This suggests that the A-to-G mutation at the tenth position of USE is causal to *HbU6.2* inactivity. Finally, it is worth noting that some variations in the AT-rich TATA box appeared tolerated, such as *PtU6.2* (TTTAAgTA), *MdU3* (cTTATATg), *CiU6.3c8* (TTTATATc), and *CiU6.6* (TTTAcATA), indicative of a less stringent requirement.

The finding that short 70 bp promoters harboring USE and TATA are sufficient for Pol III function, along with the observation that sequence outside these elements is highly variable, led us to propose that the non-conserved space can be exploited to generate novel Pol III promoters. To test this idea, we designed two *MtU6.6* promoter variants with 9-13 dispersed substitutions outside the USE and TATA elements (*MtU6.6m1* and *MtU6.6m4*). A third variant, *MtU6.6m5*, was recovered from multiplex Gibson assembly with an additional substitution compared to *MtU6.6m4* (Figure 1c). All three variants retained promoter activity (Figure 1f), supporting the idea that sequences outside USE and TATA can be altered without affecting function. Our study extends the work of Waibel and Filipowicz (1990), who introduced restriction sites into the non-conserved sequence of *AtU6* promoters without affecting function.

In summary, this study offers several advances and lessons regarding Pol III promoters. First, we provide experimental data supporting the effectiveness of minimal U6 and U3 promoters across a range of dicot species. Second, the USE consensus can be refined as T(C/A)CCACATCG, whereas the TATA consensus is less strict. Third, not all wild-type Pol III promoters are functional across species due to USE sequence variation, as shown for *AtU6.1* (Li et al., 2013) and *HbU6.2* (Dai et al., 2021). The updated USE and TATA consensus and mutagenesis data presented herein can be used to guide the selection of Pol III promoters with broad functionality. Fourth, we exemplified the design of new-to-nature promoters to increase the diversity of the CRISPR toolkit. Given the non-conserved sequence space and the combinatorial nucleotide possibilities, the potential for synthetic Pol III promoters is immense. Finally, while this work focused on dicots, a similar approach is applicable to monocots, whose Pol III promoters harbor additional monocot-specific *cis* elements upstream of USE (Connelly et al., 1994). The compendium of 23 experimentally verified dicot U6/U3 promoters should significantly enhance multiplex genome editing capability for synthetic biology applications.

Acknowledgments

This work was supported by the Center for Bioenergy Innovation (CBI), U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award Number ERKP886 (C.-J.T.) and the Georgia Research Alliance Young Scholar program (K.S.). We thank the assistance of Margot Chen for plant transformation, the Georgia Genomics and Bioinformatics Core (RRID:SCR_010994) for Illumina sequencing, and the Biomedical Microscopy Core for Zeiss LSM-880 imaging.

Conflicts of interests

The authors declare no conflict of interest.

Author contributions

C.-J.T. conceived the research, C.-J.T and M.D. designed the experiments, M.D., K.S., and A.P. executed the experiments with assistance from M.C., M.A.O. and W.P.B., M.D. and C.-J.T. wrote the paper.

References

Connelly, S., Marshallsay, C., Leader, D., Brown, J.W.S. and Filipowicz, W. (1994) Small Nuclear RNA Genes Transcribed by Either RNA Polymerase II or RNA Polymerase III in Monocot Plants Share Three Promoter Elements and Use a Strategy To Regulate Gene Expression Different from That Used by Their Dicot Plant Counterparts. *Molecular and Cellular Biology* **14**, 5910-5919.

Dai, X., Yang, X., Wang, C., Fan, Y., Xin, S., Hua, Y., Wang, K. and Huang, H. (2021) CRISPR/Cas9-mediated genome editing in Hevea brasiliensis. *Industrial Crops and Products* **164**, 113418.

Fan, D., Liu, T., Li, C., Jiao, B., Li, S., Hou, Y. and Luo, K. (2015) Efficient CRISPR/Cas9-mediated targeted mutagenesis in *Populus* in the first generation. *Scientific Reports* **5**, 12217.

Li, G., Sretenovic, S., Eisenstein, E., Coleman, G. and Qi, Y. (2021) Highly efficient C-to-T and A-to-G base editing in a Populus hybrid. *Plant Biotechnol J* **19**, 1086-1088.

Li, J.-F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M. and Sheen, J. (2013) Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol* **31**, 688-691.

Waibel, F. and Filipowicz, W. (1990) U6 snRNA genes of Arabidopsis are transcribed by RNA polymerase III but contain the same two upstream promoter elements as RNA polymerase ll-transcribed U-snRNA genes. *Nucleic Acids Research* **18**, 3451-3458.

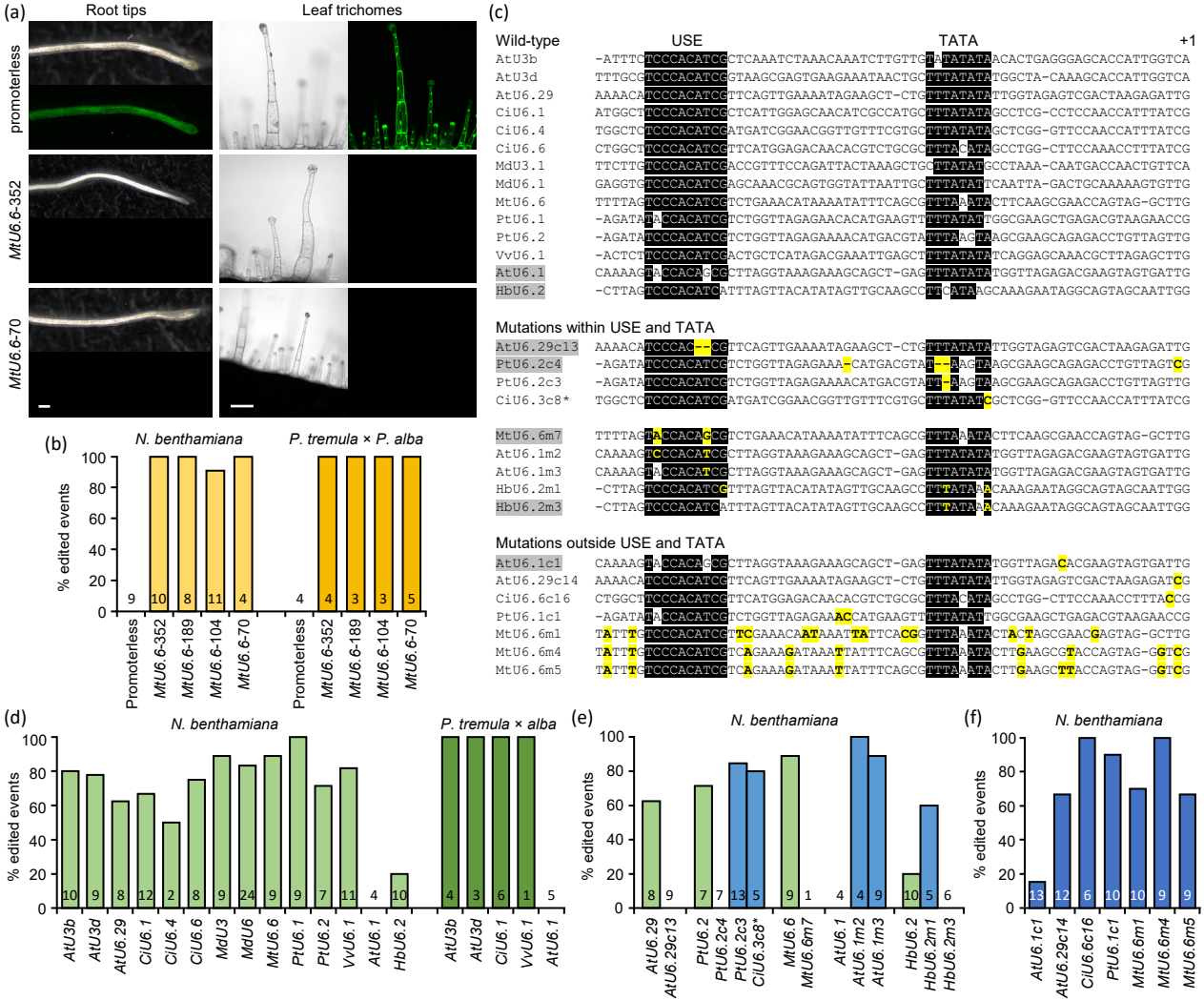
Zhou, X., Jacobs, T.B., Xue, L.-J., Harding, S.A. and Tsai, C.-J. (2015) Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial *Populus* reveals 4-coumarate:CoA ligase specificity and redundancy. *New Phytologist* **208**, 298-301.

Figure legend

Figure 1. Pol III promoter characterization. (a) Images of *N. benthamiana* root tips or leaf trichomes under brightfield or UV. Scar bars: 1 mm (root tips), 100 µm (trichomes). (b) Editing efficiencies of the *MtU6.6* promoter deletion series. (c) Alignments of 70 bp promoters. Conserved USE and TATA sequences are shaded in black, mutations in yellow, and non-functional promoters in grey. (d) Editing efficiencies of 70 bp promoters from diverse origins. (e) Editing efficiencies of 70 bp promoters with mutations in the USE or TATA (wild-type promoter data from (d) included as reference). (f) Editing efficiencies of 70 bp promoters with mutations outside USE and TATA. Transgenic event numbers are indicated at the bottom of the bars. **CiU6.3* wild-type promoter was not tested.

Table S1. Published U6 and u3 promoters from dicot species.

Data S1. Amplicon sequencing results of transgenic lines


(a)
Root tips
Leaf trichomes
promoterless
MtU6.6-352
MtU6.6-70
(b)
N. benthamiana
P. tremula × P. alba
% edited events
Promoterless
MtU6.6-352
MtU6.6-189
MtU6.6-104
MtU6.6-70
(c)
Wild-type
USE
TATA
+1
AtU3b
AtU3d
AtU6.29
CiU6.1
CiU6.4
CiU6.6
MdU3.1
MdU6.1
MtU6.6
PtU6.1
PtU6.2
VvU6.1
AtU6.1
HbU6.2
Mutations within USE and TATA
AtU6.29c13
PtU6.2c4
PtU6.2c3
CiU6.3c8*
MtU6.6m7
AtU6.1m2
AtU6.1m3
HbU6.2m1
HbU6.2m3
Mutations outside USE and TATA
AtU6.1c1
AtU6.29c14
CiU6.6c16
PtU6.1c1
MtU6.6m1
MtU6.6m4
MtU6.6m5
(d)
N. benthamiana
P. tremula × alba
% edited events
AtU3b
AtU3d
AtU6.29
CiU6.1
CiU6.4
CiU6.6
MdU3
MdU6
MtU6.6
PtU6.1
PtU6.2
VvU6.1
AtU6.1
HbU6.2
(e)
N. benthamiana
% edited events
AtU6.29
AtU6.29c13
PtU6.2
PtU6.2c4
PtU6.2c3
CiU6.3c8*
MtU6.6
MtU6.6m7
AtU6.1
AtU6.1m2
AtU6.1m3
HbU6.2
HbU6.2m1
HbU6.2m3
(f)
N. benthamiana
% edited events
AtU6.1c1
AtU6.29c14
CiU6.6c16
PtU6.1c1
MtU6.6m1
MtU6.6m4
MtU6.6m5