A multi-purpose toolkit to enable advanced genome engineering in plants
Tomáš Čermák, Shaun J. Curtin, Javier Gil-Humanes, Radim Čegan, Thomas J. Y. Kono, Eva Konečná, Joseph J. Belanto, Colby G. Starker, Jade W. Mathre, Rebecca L. Greenstein, Daniel F. Voytas


Corresponding author: Daniel F. Voytas, voytas@umn.edu

Review timeline:

<table>
<thead>
<tr>
<th>Submission</th>
<th>Dec. 8, 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Decision</td>
<td>Jan. 28, 2017 revision requested</td>
</tr>
<tr>
<td>1st Revision</td>
<td>April 24, 2017</td>
</tr>
<tr>
<td>2nd Decision</td>
<td>April 25, 2017 acceptance pending, sent to science editor</td>
</tr>
<tr>
<td>Final acceptance</td>
<td>May 16, 2017</td>
</tr>
<tr>
<td>Advance publication</td>
<td>May 18, 2017</td>
</tr>
</tbody>
</table>

REPORT: (The report shows the major requests for revision and author responses. Minor comments for revision and miscellaneous correspondence are not included. The original format may not be reflected in this compilation, but the reviewer comments and author responses are not edited, except to correct minor typographical or spelling errors that could be a source of ambiguity.)

TPC2016-00922-LSB 1st Editorial decision – revision requested Jan. 28, 2017

We have received reviews of your manuscript entitled "A multi-purpose toolkit to enable advanced genome engineering in plants." Thank you for submitting your best work to The Plant Cell. The editorial board agrees that the work you describe is substantive, falls within the scope of the journal, and may become acceptable for publication pending revision, and potential re-review.

We ask you to pay attention to the following points in preparing your revision. The reviewers provide a number of points that should be followed. One of the most critical issues is that the evidence for gene targeting (Figure 9) is not well-developed and lacking definitive evidence for the nature of the potential targeting events. This point will need to be bolstered for publication in The Plant Cell. Further, it is not clear how this part of the manuscript overlaps (or not) with the experiments in Gil-Humanes et al. listed in the references as "in press". If you decide to provide a revision, please supply this manuscript. One approach would be to delete this experiment although therein lies part of the argument for novelty of this manuscript. Reviewer #2 also noted that the experiments involving mutagenesis are not robust enough to determine efficiencies in the best manner and this point needs to be addressed. Lastly, the references are riddled with typos and omissions that need to be corrected.

See attached file for comments on the figures. The font sizes on some components are likely to be too small once figures are sized for publication. Also, note that 0 on graphs should not be written as 0.0

Note that the sampling and nature of "biological replicates" should be described precisely (i.e. different plants, parts of plants, pooled tissue, independent pools of tissue, sampled at different times, etc). The reader should know exactly what was sampled; what forms the basis of the calculation of any means and statistical parameters reported. This is also necessary to ensure that proper statistical analysis was conducted.

Please contact us if there are ambiguous comments or if you wish to discuss the revision.

-------------------------------Reviewer comments:

[Reviewer comments shown below along with author responses]
Reviewer comments and author responses:

Reviewer #1:

Importance of findings. This is an informative paper from a lab that has been pioneering genome editing of plants for multiple years. The described toolkit provides multiple functionalities and appears to work well. The paper describes experiments that contribute empirical data to the editing field and multiple interesting findings that will facilitate genome editing in diverse monocot and dicot plant species. The data comparing the efficiencies of different strategies for expressing multiple guide RNAs (Csy4 vs. tRNA vs. ribozyme and other systems) are particularly informative.

Quality of experiments. The development of the components of the toolkit is described in great detail. In general, experiments are well carried out. However, the analysis of the transgenic plants could be interpreted better (see below). Data for germ line transmission is minimal.

Points in favor of publication. The discussion does a good job of placing these data in the context of other studies in this rapidly evolving field. They successfully avoid making too-broad generalizations. Their experiments involve multiple species and target different genes in different tissues but are essentially an N of 1. The DNA repair processes likely vary considerably between these circumstances and therefore only comparisons within experiments are robust.

Weak points
1. The paper is presented (dressed up) as a series of experiments conducted to test their toolkit. In reality, for the most part, it describes observations gleaned from experiments on multiple species targeting different genes as technologies for genome editing and their expertise evolved. The resulting data is none the worse for this but it would be better and more honest to present it as such as well as make the experimental designs more comprehensible.

We submitted our manuscript for consideration as a ‘Large-Scale Biology’ article, because we felt we had ‘a new resource of significant value to the broader plant science community.’ Specifically, we wanted to provide the plant science community with a diverse set of reagents that make it easier to carry out many different gene-editing approaches. As such, we have validated a range of applications using many different plant species (including both monocots and dicots). We have stated our purpose in the last paragraph of the Introduction to make our objective clear to the reader.

2. They seem to have put more thought into the design of the toolkit than interpretation of the transgenics. Analysis of leaves of T0 plants will not represent the events in the germline unless editing occurred very early on. Many (most?) plants will be chimeric and different events may occur in different germline lineages. Their data are consistent with this. Analysis of more T1 plants is needed before they can make claims of transmission efficiencies. There is only weak evidence for heritability. Because they did not segregate out the SSN in the offspring, editing events could have occurred in the progeny; this is supported by their observation of a T1 showing 4 alleles indicating that the TALEN was still active, which makes it impossible to be certain that the mutations were transmitted through the gametes. While the 63-bp deletion was almost certainly not induced de novo in multiple T1s, a sequence-specific effect due to microhomology cannot be excluded. That said, there is no reason to suspect that changes in the sequence would not be inherited (actually, it would remarkable if they were not); therefore, demonstrating the heritability of the edits is of minor importance.

In the revised manuscript, we added additional data on the heritability of gene edits for two of the experiments in Medicago. Specifically, we present analyses of progeny from two to four T0 plants in each experiment. As described in detail in the revised manuscript, all T0 plants gave rise to T1 progeny with the same types of mutations observed in the T0. Further, in some plants, the transgene had segregated away, suggesting that the mutations were stable, germline events and not the result of de novo mutagenesis by the nuclease.

3. Also, from their data it looks like there is a possible correlation with the editing efficiency for certain gRNAs and the position of that gRNA in the polycistronic array. This should be addressed and/or more data should be shown to show that there is no correlation.

To address this point, we conducted additional experiments using a polycistronic array of eight gRNAs. We
assessed frequencies of mutagenesis for polycistronic arrays processed by three different mechanisms: Csy4, tRNA-processing enzymes, and ribozymes. Consistent with our previous data, the highest frequencies of mutagenesis were obtained with polycistronic messages processed by Csy4 or tRNA-processing enzymes. Both constructs resulted in mutagenesis frequencies approximately two-fold higher than constructs expressing the eight gRNAs individually from U6 promoters. We also tested the effect of the position of the gRNA within the array. Typically, the first two gRNAs in the array resulted in higher frequencies of mutagenesis relative to the last two gRNAs. The last two gRNAs were nonetheless comparable in mutagenesis frequencies to gRNAs expressed from the U6 promoter. This data is discussed in the manuscript and presented in a new figure (Figure 4).

4. The paper is totally focused on DNA-mediated delivery of editing reagents. In the Discussion, they should at least acknowledge the movement towards the use of RNPs when feasible.

   The reviewer is correct in that our vectors are limited to DNA delivery. The delivery of RNPs for gene editing cannot be accomplished using our current vector system.

5. Also, there are multiple sets of modules developed by others that use the Golden Gate assembly strategy but differ in the overhangs generated by the Type II endonucleases. They should state in the Discussion the compatibility or lack thereof with other sets of modules.

   Our vector system is distinct and not compatible with other published Golden Gate assembly strategies.

Specific points:

6. In the web tool, the ‘Cas9 Only T-DNA’ vector selection page does not include the AtCas9 nuclease or the AtCas9 H840A nickase.

   This has been corrected.

7. Comments on the toolkit: Currently there seems to be no flexibility in terminators in the module.

8. Why are there different terminators in modules A, B and C?

   Because repeated sequences can lead to recombination in bacteria and often lead to gene silencing in plants, we designed our vector system so that users can make constructs with different (non-redundant) transcriptional terminators. Module A contains the sequence specific nuclease of choice (i.e. either TALENs or Cas9), which uses an HSP transcriptional terminator. The polycistronic gRNA transcription units in Module B are terminated with the 35S terminator, and the RuBisCo small subunit terminator from *Pisum sativum* is used for most genes in Module C. Finally, we would like to point out the flexibility of our vector system: restriction sites are provided allowing users to add additional transcriptional terminators of choice (see Supplemental Fig. 1).

9. Ln 72: The list of Cas9 applications could include site specific editing by deamination.

   We have added this to the text.

10. Ln 179: While the 7:1 ratio is not significantly different from a 3:1 ratio, it is also consistent with some gametes being generated from nuclei lacking editing events.

   It is formally true that some gametes could have been generated from nuclei lacking a gene-editing event. However, the mutations recovered in the T1 were the same as those observed in the T0. This is consistent with an editing event that occurred in one allele early in development. The floral tissue derived from cells having undergone this event then segregated in a normal Mendelian fashion.

11. Ln 184: The conclusion that 'heritable mutations can be induced at high frequency' is overstating the data.

   We removed the words ‘at high frequency’ so as not to overstate the data.

12. Ln 201: They should state the function of the wild type and mutant phenotype of the tomato *ARF8A* gene.

   We added a sentence clarifying function of the *ARF8A* gene.

13. Ln 235: It is interesting that most insertions mapped either to the tomato genome or to the vector, covering almost every region of the vector sequence (Supplemental Figure 7). Is there any evidence for micro-homology driving this? What are the consequences for NHEJ-mediated gene targeting?
We did not observe any evidence for microhomology-mediated repair. No particular insertion predominated at the break site, which would be suggestive of repair due to microhomology. As such, we do not have any data that informs NHEJ-mediated gene targeting.

14. Ln 247 et seq.: This paragraph is confusing and needs to be rewritten. Need to clearly state which genes were targeted in which species with how many gRNAs. Currently, seems to be hiding the diversity within this experiment.

We edited this paragraph to add additional clarity.

15. Ln 267: Only T0s were analyzed. There are no data on heritability.

In the revised manuscript, we provide additional data on the heritability of mutations induced in the T0 plants. As described in detail in the text, we analyzed seven T1 progeny from each of four T0 plants and show transmission of most mutations.

16. Ln 276: It is true that 44 plants had mutations in at least one of the three loci but they over sell the data. Looking at supplemental data 2 and supplemental figure 11, only NRC53 had a high mutation rate (44/46 plants with editing in either or both gRNAs), for NRC54, only 8/46 plants had editing at one or both gRNAs, for NRC55: only 11/46 plants had editing at one or both gRNAs (Table 1). It is confusing why there are so many SNPs in the NRC53 alleles? Some comment on this observation should be made.

As stated in the text, there are two, duplicate NRC53 genes in close proximity. These genes differ by a number of SNPs, and yet they are targeted by the same gRNA. We show sequence data for all mutants recovered in both copies of the NRC53 gene, including the SNPs that distinguish these genes. One of the NRC54 guides (gRNA3) was less effective in mutagenesis, because it has two mismatches with its target. The NRC55 gRNA was at the last position in the array, and we now present new data (Figure 4) that shows gRNAs at the last position are less effective in mutagenesis.

Ln 295: 58 kb seems a very small region to span a QTL. What are the LOD 1 and 2 sizes of this QTL?

We no longer refer to this region as a QTL. Rather, we say candidate genes were revealed by GWAS studies in this region of the genome as having a potential role in N2 fixation.

Ln 306: The question about the position in the polycistronic array goes for the second Medicago experiment: only 3/46 plants with 58 kb deletion (between gRNAs mpc-6 and mpc-1), and 2/48 had 27 kb deletion (between mpc1 and mpc4). What is the order of the gRNAs on this vector (mpc1→6?)?

We now specify the order of the gRNAs in the vector (mpc1→6); however, we do not consider 3/46 and 2/48 to be statistically different mutagenesis frequencies. As mentioned above, we carried out a new experiment to specifically test the impact of position of the gRNAs in the array on mutagenesis.

Ln 310: This plant may be chimeric rather than heterozygous.

We now acknowledge this possibility.

Ln 313: The only data on heritability is based on 12 T1 segregants from a single T0 plant.

We now show heritability from an additional T0 plant and one of the 12 T1 plants in the T2 generation. Some of the T2 plants are transgene free and homozygous for the mutation. Figure 7 has been revised to include this additional data.

Author actions necessary for acceptance. The authors should revise the paper taking the above comments into account. In particular their analysis of the transgenic plants seems surprisingly naïve and needs to be rethought and rewritten.

As indicated in our responses above, we believe we have addressed the majority of this reviewer’s concerns.

Reviewer #2:

1. The manuscript of Cermak et al. comprises numerous data sets, which are interesting in themselves, however cannot be considered novel or conclusive enough to justify publication alone. Upon reading, it appears as though
In the final paragraph of the Introduction, we acknowledge that the purpose of this study is to describe and provide to the plant science community a comprehensive toolkit for plant gene editing. None of the experiments were leftover from previous studies. Rather, they were conducted to test all of our vectors to give confidence to users that our system is robust and effective.

2. Although a variety of clones are described that can ultimately be used for genome engineering in plants, a vast majority of plant-specific CRISPR/Cas tools have been published previously. Several new aspects are addressed and documented with appropriate experimental data; nevertheless, the data depicted can be considered as anecdotal, rather than reflecting a complete, reliable analysis required. Therefore, it is not clear as to how this article would successfully meet the criteria required for a resource paper of *The Plant Cell*.

In advance of publication, we have already distributed our toolkit to over 30 laboratories worldwide. We believe this interest by the community speaks to the value of having a single resource that can be used to perform many, diverse genome engineering applications. We have bolstered our data in responses to specific reviewer comments, and believe the manuscript is much improved as a consequence.

3. The question of which approach is most efficient to express multiple sgRNAs in plants is certainly an interesting aspect. However, it is apparent that there is only one comparative experiment that has been performed, as shown in Figure 3. The authors only tested constructs with two sgRNAs, which is unfortunately not sufficient, as it does not show a complete and relevant analysis. A dataset showing the analysis of simultaneous expression with a larger number of sgRNAs, such as five or six, would be required in order to adequately draw final conclusions regarding efficiencies. As a result, I do not regard the data presented in this manuscript to be conclusive.

Our response to Reviewer #1 addresses this question: We conducted additional experiments using a polycistronic array of eight gRNAs. We assessed frequencies of mutagenesis for polycistronic arrays processed by three different mechanisms: Csy4, tRNA-processing enzymes, and ribozymes. Consistent with our previous data, the highest frequencies of mutagenesis were obtained with polycistronic messages processed by Csy4 or tRNA-processing enzymes. Both constructs resulted in mutagenesis frequencies approximately two-fold higher than constructs expressing the eight gRNAs individually from U6 promoters. We also tested the effect of the position of the gRNA within the array. Typically, the first two gRNAs in the array resulted in higher frequencies of mutagenesis relative to the last two gRNAs. The last two gRNAs were nonetheless comparable in mutagenesis frequencies to gRNAs expressed from the U6 promoter. This data is discussed in the manuscript and presented in a new figure (Figure 4).

4. The GT experiments whereby a nickase is used are also notable, although too preliminary for publication (Figure 9). In order to obtain final conclusions regarding the nature of the repair reaction of the new kind of lesion, the author should show via Southern blotting and sequencing that both ends of the targeting vector are indeed integrated by HR and that no further rearrangement of the target locus is occurring. Moreover, the approach shown is one which would only be valuable for those within the plant community, especially if it is the case that fertile mutants can be regenerated, which the authors do not show. Due to the fact that the nickase approach does not permit higher GT frequencies, it is not clear to general users as to how such an approach would be beneficial.

In the revised manuscript, we provide DNA sequence analysis demonstrating that gene targeting, stimulated by nickases, occurred in tobacco and was precise. We have added this data to Figure 9. Due to time constraints, we were not able to generate whole plants with targeted modifications, and so we were not able to complete Southern analysis. However, we showed previously that when a nuclease was used (which in the current experiment resulted in similar GT frequencies as the paired nickases), we were able to recover plants with targeted modifications that were heritable and verified by Southern blotting (Baltes et al. 2014, *Plant Cell* 26: 151-63; Cermak et al. 2015, *Genome Biol.* 16:232). Finally, we moved the gene targeting data in wheat – which was not corroborated by molecular analyses – to a supplemental figure.

Reviewer #3:
The data from this paper are from the following subheadings:

1. Direct and modular assembly of genome engineering constructs

   The authors describe plasmid vectors, and how different pieces of DNA can fit into these vectors to create a desired construct for use in genome editing in plants.

2. Targeted mutagenesis of genes in Medicago truncatula using direct vectors.

   Two Medicago genes related to phosphate regulation previously described in Arabidopsis were mutated using TALEN. One of the mutated genes transmitted to the next generation, while the other mutated gene appeared to have not transmitted. This shows that TALEN can cause a site-specific mutation, which in itself is not novel, but this may be a first report of TALEN-induced mutation in Medicago.

3. Expressing multiple gRNAs by Csy4, tRNA-processing enzymes and ribozymes.

   This section tested the ability to create mutations in two sites in the tomato ARF8A gene. The original procedure uses two separate Pol III transcription units of gRNA, but more recently, various reports have described posttranscriptional processing of a single Pol II transcript. The data show that in tomato protoplasts, the Cys4 and tRNA system are about 2 fold better than the original system of expressing separate Pol III transcription units of gRNA (while the ribozyme system was 2 fold less efficient than the original system). I think this section, as in the previous section, is just a demonstration that their vectors work, rather than to claim novelty of the Csy4 and tRNA systems for generating multiple gRNAs.

4. Multiplexed mutagenesis in tomato, wheat, and barley protoplasts.

   This section tested specifically the Csy4 system on making deletions in one, three, and six genes in barley, wheat and tomato protoplasts, respectively. They work.

5. Heritable, multiplexed mutagenesis in Medicago truncatula.

   This section tested generating deletions within 4 Medicago genes at the same time. Not just a protoplast demonstration, but transmission of these mutated genes was not described.

6. Targeted deletion of 58 kb in Medicago using the tRNA and Csy4 multiplexing systems

   Here, 27 kb and 58 kb deletions in the genome were generated and germlinal transmission confirmed in one line.

7. Enhancing targeted mutagenesis with TREX2.

   This section tested the inclusion of TREX2 for generating mutations in tomato and barley protoplasts. TREX2 is a 3’ repair exonuclease previously reported effective for enhancing the frequency of resecting double stranded breaks in human and plant cells including rice. The difference here appears to be the use of TREX2 in combination with Cas9 in tomato and barley protoplasts. The data show longer length deletions and a modest 1.5 to 2.5 fold increase in mutation frequency.

8. Gene targeting with Cas9 nickases.

   This section tested the use of Cas9 nickases, which has been reported to reduce off-target activity because it requires the recognition of two target sequences, one on each side of the double-strand break to cause the nicking of each DNA strand at the same location. Here, the authors used, as previously reported, geminviral replicons to amplify the copy number of the DNA introduced by Agro-infiltration into tobacco leaves or microparticle bombardment into wheat scutella. They show that new DNA can recombine with genomic DNA to replace or insert into the genome target. However, the frequency for this gene targeting appears no more efficient than the use of the gRNA by itself. The assumption is that this approach should be better, as it should be less prone to generating off-target mutations, although there is no experimental data, just logic, to support this supposition.

   We acknowledge that we did not check for off-target mutations using Cas9 nickases. Such studies have been conducted in other systems. Cas9 nickases have been shown to reduce off-target activity up to 1,500-fold in mammalian cell lines (Ran et al. 2013, Cell 154: 1380-89), and to significantly suppress off-target mutations in rice plants (Mikami et al. 2016, Plant Cell Physiol. 57: 1058-68).

10. Overall, the paper is mainly about a toolkit for genome editing that entails a large set of vectors, and various pieces of DNA that can fit into these vectors to create a desired construct for use in genome editing in plants. With different plasmid backbones, promoters, terminators, and genes encoding sequence-specific interaction proteins, there are hundreds of different DNA constructs. A website is provided that enables one to select among the many combinations of vectors/inserts to create the genome editing construct, but it not clear how these reagents will be provided to the practitioner. For example, will these reagents be deposited into a public resource center, similar to
what we can order with Arabidopsis mutants? If these reagents are available, and at minimal cost, one still has to take into consideration whether it would be more convenient to order them delivered as separate constructs, or would it be just as convenient (timewise) in this day and age to re-create them from existing vector systems? Once the DNA sequence is known, making constructs these days is not that difficult to do. Given what has already appeared in the published literature, including the latest improvements to the original system, it is unlikely that researchers would need a large collection of constructs and parts. Most likely, one would choose the most appropriate system needed to generate the desired mutations or gene targeting events. For that, it could be as easy as modifying an existing vector system. Nonetheless, I guess requesting a few pieces of DNA from this research group could also help.

We have submitted all the plasmids reported in the manuscript (236 plasmids total) to the non-profit clone repository – Addgene. Anyone can obtain any number of these plasmids for a nominal cost.

11. As for the experimental data in this paper, the demonstration that the system works is convincing. However, aside from this demonstration, which reinforces what we already know or expected to be true, there is not much novelty to this paper, unless one considers transformation in Medicago or barley protoplasts as breakthroughs in plant biotechnology.

We agree with the reviewer that if this was a ‘Research’ article, more novelty would be required. However, and as stated above, our paper describes a resource – a gene editing toolkit that we wish to make available to the plant science community. We tested all of our vectors to give confidence to users that our system is both effective and efficient. Further, we provide new data to guide best practices, particularly for multiplexed gene editing.

TPC2016-00922-LSB1 2nd Editorial decision – acceptance pending April 25, 2017

We are pleased to inform you that your paper entitled "A multi-purpose toolkit to enable advanced genome engineering in plants" has been accepted for publication in *The Plant Cell*, pending a final minor editorial review by journal staff.

Final acceptance from Science Editor May 16, 2017