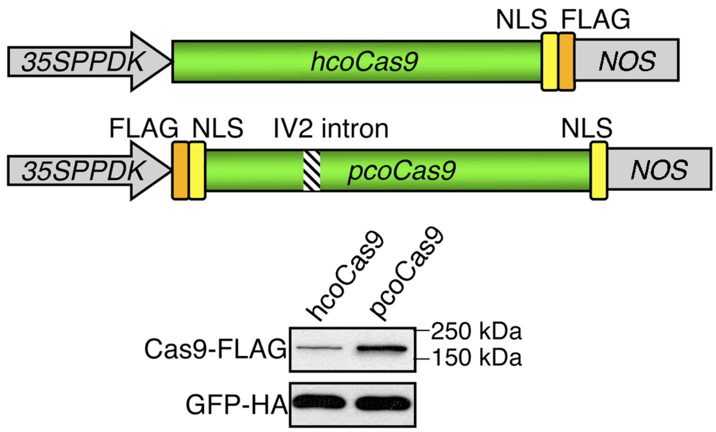
Multiplex and homologous recombination-mediated plant genome editing via guide RNA/Cas9

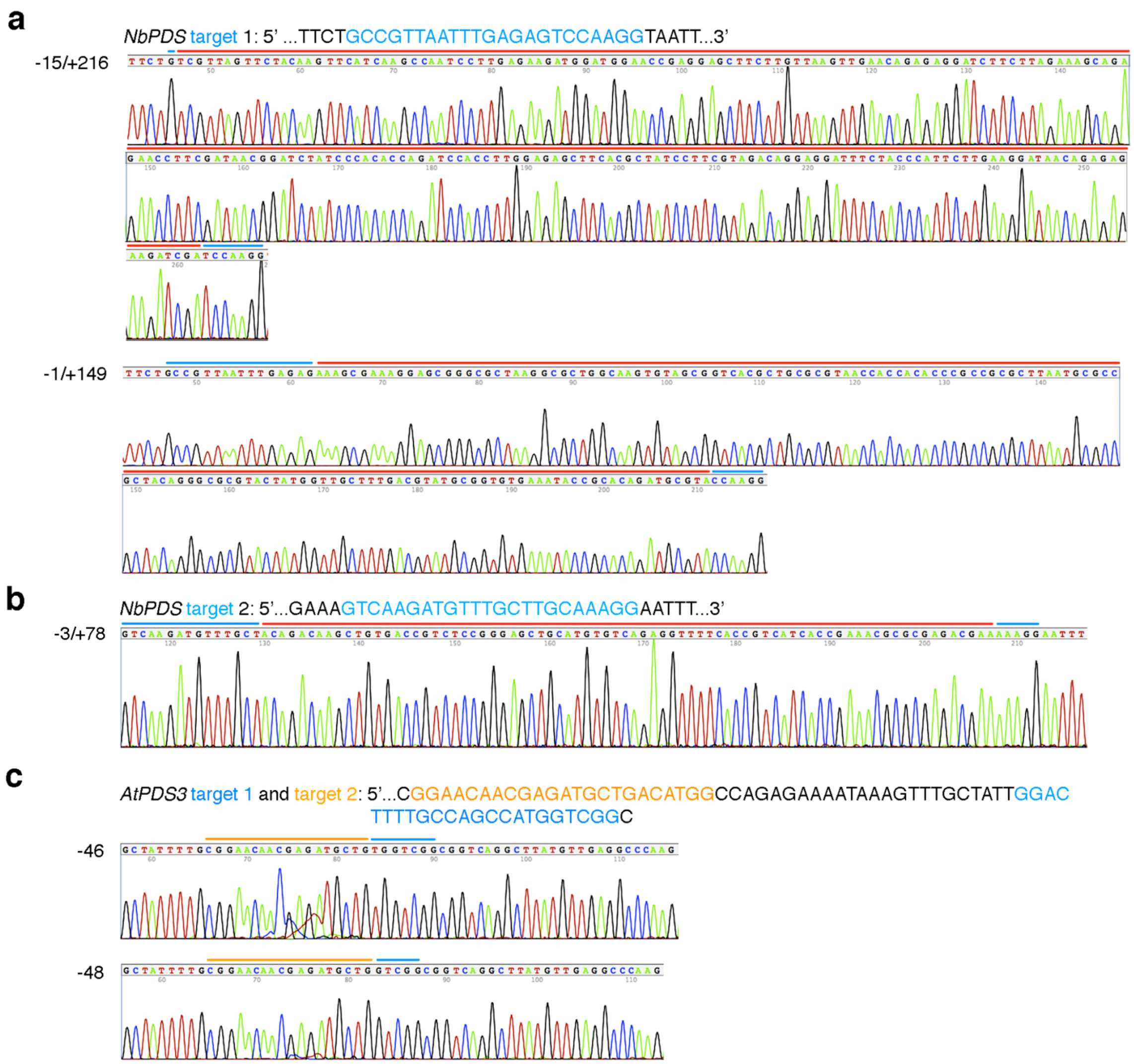
Jian-Feng Li, John Aach, Julie E. Norville, Matthew McCormack, Dandan Zhang, Jenifer Bush, George M. Church & Jen Sheen

Supplementary Information

Table of Contents



Supplementary Figure 1. The *pcoCas9* shows a higher expression level than human codon-optimized *Cas9* (*hcoCas9*) in *Arabidopsis* mesophyll protoplasts. GFP-HA served as a transfection internal control. *35SPPDK*, a hybrid constitutive promoter. NLS, the nuclear localization sequence. FLAG, double FLAG tag. *NOS*, the nopaline synthase terminator.



Supplementary Figure 2. DNA sequencing evidence for long DNA insertions or deletions during gRNA/pcoCas9-mediated genome editing in protoplasts. (**a**) Long DNA insertions are coupled with DNA deletions in the *NbPDS* target site 1. (**b**) Long DNA insertion is coupled with DNA deletion in the *NbPDS* target site 2. (**c**) Long genomic deletions are induced between two juxtaposed target sites in *AtPDS3*. The red bar on top of the chromatogram marks DNA insertions. The bar in blue or orange on top of the chromatogram marks the remaining target sequence after DNA deletion.

a

```
AtRACK1b     1 ATGGCTGAAGGACTCGTGTTGAAAGGCACAATGTGTGCCCACACTGATATGGTCACCGCC
AtRACK1c     1 ATGGCCGAGGGACTCGTATTGAAGGGCATTATGCGCGCCCACACCGACATTGTCACGGCC

AtRACK1b    61 ATTGCTACACCGGTCGATAACTCCGACGTGATTGTAACTTCGTCGCGTGACAAATCAATC
AtRACK1c    61 ATCGCTACGCCGATCGACAATTCCGACATCATCGTCACAGCGTCGCGTGACAAATCCATC
                    Candidate 1
AtRACK1b   121 ATCCTCTGGAAACTCACAAAGGAAGACAAGTCATACGGTGTTGCTCAGCGGAGGATGACT
AtRACK1c   121 ATCCTCTGGAAACTCACAAAGGACGATAAGTCTTACGGTGTTGCTCAGCGTAGGCTCACA

AtRACK1b   181 GGTCACTCTCACTTCGTTCAAGACGTTGTTCTCTCCTCCGATGGACAATTCGCTCTCTCC
AtRACK1c   181 GGTCACTCTCACTTCGTGGAAGATGTTGTTCTCTCATCGGACGGTCAGTTTGCACTCTCC

AtRACK1b   241 GGATCTTGGGACGGTGAGCTTCGTCTCTGGGATCTCGCTACCGGAGAATCTACTCGTCGT
AtRACK1c   241 GGAAGCTGGGACGGTGAGCTCCGTCTCTGGGATCTTGCCACGGGAGAAACAACTCGTCGA

AtRACK1b   301 TTCGTTGGTCACACGAAAGATGTTCTCTCTGTTGCTTTCTCTACCGATAACCGTCAGATC
AtRACK1c   301 TTCGTTGGTCATACGAAAGATGTGCTCTCTGTTGCATTCTCTACTGATAACCGGCAGATC
                                   Candidate 2
AtRACK1b   361 GTCTCTGCTTCTCGTGATCGTACGATTAAGCTTTGGAACACGCTTGGTGAGTGTAAGTAT
AtRACK1c   361 GTGTCTGCTTCTCGTGATCGTACGATTAAGCTTTGGAACACACTTGGTGAGTGTAAGTAT
                                                   Candidate 3
AtRACK1b   421 ACCATCTCTGAAGCTGATGGTCACAAGGAATGGGTTAGTTGTGTTAGGTTTAGTCCTAAT
AtRACK1c   421 ACCATCTCTGAAGGTGATGGTCACAAGGAATGGGTTAGTTGTGTTAGGTTTAGTCCTAAT
                         Candidate 4 & 5
AtRACK1b   481 ACTCTTGTGCCAACTATTGTATCTGCTTCTTGGGATAAAACTGTGAAGGTCTGGAATCTT
AtRACK1c   481 ACTCTTGTACCAACTATTGTATCTGCTTCTTGGGATAAAACTGTGAAAGTTTGGAATCTC

AtRACK1b   541 CAGAATTGTAAGCTTAGGAACACTCTTGCTGGTCACTCTGGTTACTTGAACACTGTGGCT
AtRACK1c   541 CAGAACTGTAAGCTGAGGAACTCTCTTGTTGGTCACTCTGGTTACCTCAACACTGTTGCT

AtRACK1b   601 GTGTCACCTGATGGTTCGTTGTGTGCCAGTGGTGGCAAAGATGGTGTGATCTTGCTTTGG
AtRACK1c   601 GTCTCGCCTGATGGTTCGCTATGCGCCAGTGGTGGGAAAGATGGTGTTATCTTGTTGTGG

AtRACK1b   661 GATTTGGCTGAAGGGAAGAAGTTGTATTCTCTTGAAGCTGGCTCTATTATTCATTCACTT
AtRACK1c   661 GATTTGGCTGAAGGAAAGAAGCTTTACTCGCTTGAGGCGGGTTCGATTATTCACTCGCTT

AtRACK1b   721 TGCTTCAGTCCCAATAGGTACTGGTTGTGTGCAGCAACAGAGAATAGCATTAGGATTTGG
AtRACK1c   721 TGCTTCAGTCCTAACAGATACTGGTTGTGTGCTGCTACTGAGAATAGCATTAGGATTTGG

AtRACK1b   781 GATCTCGAGAGTAAGTCTGTTGTTGAGGATTTGAAGGTTGATCTTAAGGCTGAGGCTGAA
AtRACK1c   781 GATCTTGAGAGCAAGTCTGTTGTTGAGGACTTGAAGGTTGATCTCAAGTCTGAGGCAGAG
                                                              Candidate 6
AtRACK1b   841 AAGACTGATGGTTCTACTGGAATCGGAAACAAGACCAAGGTGATCTACTGCACAAGCTTG
AtRACK1c   841 AAGAATGAAGGTGGTGTTGGAACTGGTAACCAGAAGAAGGTTATCTACTGCACAAGCTTG
                   Candidate 7
AtRACK1b   901 AACTGGAGTGCAGATGGAAACACTTTGTTCAGTGGATACACTGATGGAGTTATCAGGGTT
AtRACK1c   901 AACTGGAGTGCAGATGGAAGCACATTGTTCAGTGGTTACACTGATGGAGTTGTCAGGGTC

AtRACK1b   961 TGGGGTATTGGTCGTTAC
AtRACK1c   961 TGGGGTATTGGTCGTTAC
```

b

	Target candidates for ***AtRACK1b/c*** (seed sequence underlined)	Off-targets of seed(12)NGG				Potential off-targets of seed(12)NAG			
		N=A	N=T	N=G	N=C	N=A	N=T	N=G	N=C
1	CATCCTCTGGAAACTCACAAAGG	At5g61810						At5g15920	
2	GTGATCGTACGATTAAGCTTTGG								
3	AGGAATGGGTTAGTTGTGTTAGG							At5g28490	
4	CAACTATTGTATCTGCTTCTTGG							At1g45332	
5	AACTATTGTATCTGCTTCTTGGG						At2g13860		At2g13830
6	TCTACTGCACAAGCTTGAACTGG	At5g41790				At3g10160			
7	AGCTTGAACTGGAGTGCAGATGG								

c

Align CDS for multigenes → Select $N_{21}GG$ → Make sure no intron in $N_{21}GG$ by SeqViewer from TAIR tools → Blast seed(12)NGG or seed(12)NAG → Select $N_{21}GG$ with no hit in other exons

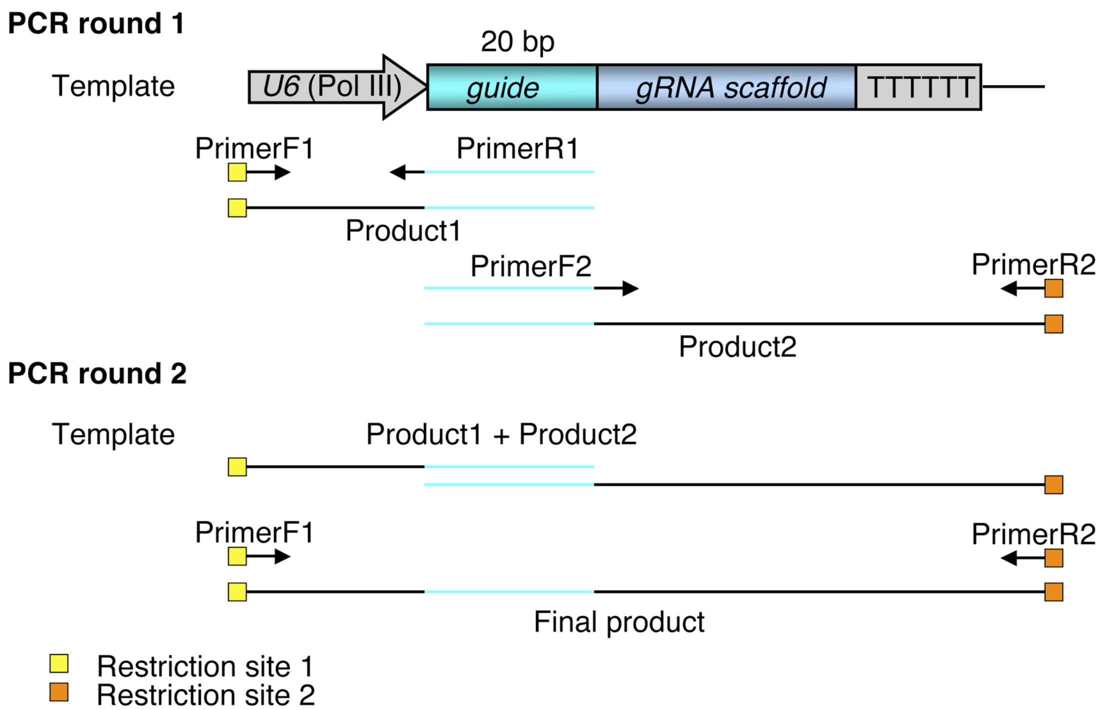
Supplementary Figure 3. Manual design for specific gRNA target sites in multiple homologous target genes. (**a**) Select $N_{21}GG$ candidates based on the coding sequence (CDS) alignment of target genes. Sequence alignment greatly facilitated the identification of 23-bp identical sequences ending with GG in homologous target genes. Note that one can also align the non-coding strands of target genes to identify more $N_{21}GG$ candidates. (**b**) Blast the 12-nt seed sequence plus 3' NGG or NAG to evaluate off-targets. A perfect match between the seed(12)NGG and a site in exon of other *Arabidopsis* gene means the $N_{21}GG$ candidate is not specific, while a perfect match between the seed(12)NAG and a site in exon of other *Arabidopsis* gene suggests the $N_{21}GG$ candidate may have a potential off-target. The $N_{21}GG$ candidates without any hit of the seed(12)NGG or seed(12)NAG in other exon are considered highly specific. Selection priority is given to the highly specific $N_{21}GG$ candidate(s) starting with a G due to optimal transcription by the RNA polymerase III. (**c**) Flow chart to summarize the manual design procedure.

NbPDS

Cas9 : gRNA : HDR template = 5 : 3 : 2   14.2% (22/155) mutated by NHEJ

```
                                                    PAM
WT                5'…TTCTGCCGTTAATTTGAGAGT-CCAAGGTAATT…3'
+1                5'…TTCTGCCGTTAATTTGAGAGTACCAAGGTAATT…3'
+1                5'…TTCTGCCGTTAATTTGAGAGTTCCAAGGTAATT…3'
-100              5'…TTCTGC---------------------------…3'
-19               5'…TTCTGCCG--------------------TAATT…3'
-19               5'…TT--------------------CCAAGGTAATT…3'
-19               5'…TTCTGCCGT--------------------AATT…3'
-12     (×2)      5'…TTCTGCCGT-------------CCAAGGTAATT…3'
-10               5'…TTCTGCCGTT----------T-CCAAGGTAATT…3'
-8      (×2)      5'…TTCTGCCGTTAAT---------CCAAGGTAATT…3'
-6                5'…TTCTGCCGTTAATTT-------CCAAGGTAATT…3'
-5                5'…TTCTGCCGTTAATTTGA------CAAGGTAATT…3'
-3      (×3)      5'…TTCTGCCGTTAATTTGAG----CCAAGGTAATT…3'
-3                5'…TTCTGCCGTTAATTTGA---T-CCAAGGTAATT…3'
-3                5'…TTCTGCCGTTAATTTGAGA----CAAGGTAATT…3'
-3                5'…TTCTGCCGTTAATTTGAGAG----AAGGTAATT…3'
-1      (×2)      5'…TTCTGCCGTTAATTTGAGA-T-CCAAGGTAATT…3'
-1                5'…TTCTGCCGTTAATTTGAGAG--CCAAGGTAATT…3'
```

Supplementary Figure 4. Targeted mutagenesis by NHEJ on the *NbPDS* target site in the presence of the HDR template. The mutation rate was calculated based on the NHEJ mutant/total (WT+NHEJ+HDR) alleles of randomly selected clonal amplicons of the target locus. Blue shadow marks the target sequence recognized by cognate gRNA. PAM, the protospacer adjacent motif. DNA insertions and deletions are shown in red as upper case letters and dashes, respectively.



Supplementary Figure 5. Diagram of the rapid assembly of a new gRNA construct by overlapping PCR. The custom guide sequence is introduced into a new gRNA through PCR primers, PrimerR1 and PrimerF2, in the first round of PCR. The PrimerR1 is in a format (from 5' to 3') of "reverse complement of the 5'-most N_{20} of a gRNA target site + AATCACTACTTCGTCTCT". PrimerF2 is in a format (from 5' to 3') of "the 5'-most N_{20} of a gRNA target site + GTTTTAGAGCTAGAAATAGC". PrimerF1 is in a format (from 5' to 3') of "NNN + Restriction site + AGAAATCTCAAAATTCCG" and PrimerR2 is in a format (from 5' to 3') of "NNN + Restriction site + TAATGCCAACTTTGTACA", where NNN represents necessary 5' sequence for a given restriction site to ensure efficient restriction digestion of PCR products.

Supplementary Table 1. Primers used in this study

Primer name	**Sequence (5' to 3')**	**Restriction site (RS)**	**Usage**
hcoCas-F hcoCas-R	CGA**CCATGG**ACAAGAAGTACTCCATT CGA**AGGCCT**CACCTTCCTCTTCTTCTTGGG	*Nco*I *Stu*I	PCR the *hcoCas9*
pcoCas-F1 pcoCasN-IV-R pcoCasN-IV-F pcoCasC-IV-R pcoCasC-IV-F pcoCas-R1	CGA**GGATCC**ATGGATTACAAGGATGATGAT AGGTAGAAGCAGAAACTTACCTCCTCGAAGTTCCAAGG CCTTGGAACTTCGAGGAGGTAAGTTTCTGCTTCTACCT GAAGCTCCCTTATCAACAACCTGCACATCAACAAATTTTG CAAAATTTGTTGATGTGCAGGTTGTTGATAAGGGAGCTTC CGA**CTGCAG**TCACTTCTTCTTCTTAGCCTGTCC	*BamH*I *Pst*I	Insert the potato *IV2* intron into *pcoCas9* by overlapping PCR
U6promoter-F U6-gRNA-AtPDS3-R	CGA**GAGCTC**AGAAATCTCAAAATTCCGGCA TCC**CCCGGG**TAATGCCAACTTTGTACAAGAAAGCTGGGT CTAGAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGT TGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGC TCTAAAACACCATGGCTGGCAAAAGTCCAATCACTACTT CGTCTCTAAC	*Sac*I *Sma*I	PCR the U6-gRNA chimera for *AtPDS3* target 1
U6-PacI-F gRNA-PDS-2R gRNA-PDS-2F gRNA-PacI-R	CGA**TTAATTAA**AGAAATCTCAAAATTCCG TGTCAGCATCTCGTTGTTCCAATCACTACTTCGTCTCT GGAACAACGAGATGCTGACAGTTTTAGAGCTAGAAATAGC CGA**TTAATTAA**TAATGCCAACTTTGTACA	*Pac*I *Pac*I	PCR the U6-gRNA chimera for *AtPDS3* target 2
U6-AscI-F gRNA-AscI-R	CGA**GGCGCGCC**AGAAATCTCAAAATTCCG CGA**GGCGCGCC**TAATGCCAACTTTGTACA	*Asc*I *Asc*I	
gRNA-FLS-R gRNA-FLS-F	AAATCGCTTACGTGAGCAACAATCACTACTTCGTCTCT GTTGCTCACGTAAGCGATTTGTTTTAGAGCTAGAAATAGC		For creating gRNA for *AtFLS2*
gRNA-RAC-R gRNA-RAC-F	AAGCTTAATCGTACGATCACAATCACTACTTCGTCTCT GTGATCGTACGATTAAGCTTGTTTTAGAGCTAGAAATAGC		For creating gRNA for *AtRACK1b,c*
gRNA-Nb1-R gRNA-Nb1-F	TGGACTCTCAAATTAACGGCAATCACTACTTCGTCTCT GCCGTTAATTTGAGAGTCCAGTTTTAGAGCTAGAAATAGC		For creating gRNA for *NbPDS* target1
gRNA-Nb2-R gRNA-Nb2-F	TTGCAAGCAAACATCTTGACAATCACTACTTCGTCTCT GTCAAGATGTTTGCTTGCAAGTTTTAGAGCTAGAAATAGC		For creating gRNA for *NbPDS* target2
pcoCas-F2 pcoCas-R2	CGA**TCTAGA**ATGGATTACAAGGATGATGAT CGA**GCGGCCGC**TCACTTCTTCTTCTTAGCCTGTCC	*Xba*I *Not*I	PCR *pcoCas9* to gain new RS sites
AtPDS-gDNA-F AtPDS-gDNA-R	CGA**GGATCC**GTTGTTGCTGTTGGATTTACG CGA**AGGCCT**CACAACAACCACATGGACTAG	*Bam*HI *Stu*I	PCR *AtPDS3* target region
FLS2-gDNA-F FLS2-gDNA-R	CGA**GGATCC**GCAGCACCGATTGGGTCGCTT CGA**AGGCCT**CTGTTCATGAAAAATAAGAGT	*Bam*HI *Stu*I	PCR *AtFLS2* target region
RK1a-gDNA-F RK1a-gDNA-R	CGA**GGATCC**CCTCCGATGGACAATTCGCG CGA**AGGCCT**CAGTTCGAAAGGTTCCACACT	*Bam*HI *Stu*I	PCR *AtRACK1a* target region
RK1b-gDNA-F RK1b-gDNA-R	CGA**GGATCC**CCTCCGATGGACAATTCGCTC CGA**AGGCCT**CAATTCTGAAGATTCCAGACC	*Bam*HI *Stu*I	PCR *AtRACK1b* target region
RK1c-gDNA-F RK1c-gDNA-R	CGA**GGATCC**CATCGGACGGTCAGTTTGCAC CGA**AGGCCT**CAGTTCTGGAGATTCCAAACT	*Bam*HI *Stu*I	PCR *AtRACK1c* target region
NbPDS-gDNA-F NbPDS-gDNA-R	CGA**GGATCC**ATGCCCCAAATTGGACTTGTT CGA**AGGCCT**CTTGGAGTACGAATCCTTAAC	*Bam*HI *Stu*I	PCR *NbPDS* target region
NbPDS-HDR-F NbPDS-HDR-R NbPDS-Mut-F NbPDS-HDR-R	CGA**GGATCC**TCTTTCAACTTCAACACAACA CGA**AGGCCT**TTATGCCCCATGGAGTCGCTA GCCGTTAATTTGAGAGT**CCTAGG**TAATTCAGCTTATCTTT AAAGATAAGCTGAATTA**CCTAGG**ACTCTCAAATTAACGGC	*Bam*HI *Stu*I *Avr*II *Avr*II	Create *Avr*II in HDR template for *NbPDS*

Supplementary Sequences

>*pcoCas9*
(GenBank ID: KF264451. The potato *IV2* intron is marked in orange)
ATGGATTACAAGGATGATGATGATAAGGATTACAAGGATGATGATGATAAGATGGCT
CCAAAGAAGAAGAGAAAGGTTGGAATCCACGGAGTTCCAGCTGCTGATAAGAAGTA
CTCTATCGGACTTGACATCGGAACCAACTCTGTTGGATGGGCTGTTATCACCGATG
AGTACAAGGTTCCATCTAAGAAGTTCAAGGTTCTTGGAAACACCGATAGACACTCTA
TCAAGAAGAACCTTATCGGTGCTCTTCTTTTCGATTCTGGAGAGACCGCTGAGGCT
ACCAGATTGAAGAGAACCGCTAGAAGAAGATACACCAGAAGAAAGAACAGAATCTG
CTACCTTCAGGAAATCTTCTCTAACGAGATGGCTAAGGTTGATGATTCTTTCTTCCA
CAGACTTGAGGAGTCTTTCCTTGTTGAGGAGGATAAGAAGCACGAGAGACACCCAA
TCTTCGGAAACATCGTTGATGAGGTTGCTTACCACGAGAAGTACCCAACCATCTAC
CACCTTAGAAAGAAGTTGGTTGATTCTACCGATAAGGCTGATCTTAGACTTATCTAC
CTTGCTCTTGCTCACATGATCAAGTTCAGAGGACACTTCCTTATCGAGGGAGACCTT
AACCCAGATAACTCTGATGTTGATAAGTTGTTCATCCAGCTTGTTCAGACCTACAAC
CAGCTTTTCGAGGAGAACCCAATCAACGCTTCTGGAGTTGATGCTAAGGCTATCCT
TTCTGCTAGACTTTCTAAGTCTCGTAGACTTGAGAACCTTATCGCTCAGCTTCCAGG
AGAGAAGAAGAACGGACTTTTCGGAAACCTTATCGCTCTTTCTCTTGGACTTACCCC
AAACTTCAAGTCTAACTTCGATCTTGCTGAGGATGCTAAGTTGCAGCTTTCTAAGGA
TACCTACGATGATGATCTTGATAACCTTCTTGCTCAGATCGGAGATCAGTACGCTGA
TCTTTTCCTTGCTGCTAAGAACCTTTCTGATGCTATCCTTCTTTCTGACATCCTTAGA
GTTAACACCGAGATCACCAAGGCTCCACTTTCTGCTTCTATGATCAAGAGATACGAT
GAGCACCACCAGGATCTTACCCTTTTGAAGGCTCTTGTTAGACAGCAGCTTCCAGA
GAAGTACAAGGAAATCTTCTTCGATCAGTCTAAGAACGGATACGCTGGATACATCG
ATGGAGGAGCTTCTCAGGAGGAGTTCTACAAGTTCATCAAGCCAATCCTTGAGAAG
ATGGATGGAACCGAGGAGCTTCTTGTTAAGTTGAACAGAGAGGATCTTCTTAGAAA
GCAGAGAACCTTCGATAACGGATCTATCCCACACCAGATCCACCTTGGAGAGCTTC
ACGCTATCCTTCGTAGACAGGAGGATTTCTACCCATTCTTGAAGGATAACAGAGAG
AAGATCGAGAAGATCCTTACCTTCAGAATCCCATACTACGTTGGACCACTTGCTAGA
GGAAACTCTCGTTTCGCTTGGATGACCAGAAAGTCTGAGGAGACCATCACCCCTTG
GAACTTCGAGGAGGTAAGTTTCTGCTTCTACCTTTGATATATATATAATAATTATCAT
TAATTAGTAGTAATATAATATTTCAAATATTTTTTTCAAAATAAAAGAATGTAGTATAT
AGCAATTGCTTTTCTGTAGTTTATAAGTGTGTATATTTTAATTTATAACTTTTCTAATA
TATGACCAAAATTTGTTGATGTGCAGGTTGTTGATAAGGGAGCTTCTGCTCAGTCTT
TCATCGAGAGAATGACCAACTTCGATAAGAACCTTCCAAACGAGAAGGTTCTTCCAA
AGCACTCTCTTCTTTACGAGTACTTCACCGTTTACAACGAGCTTACCAAGGTTAAGT
ACGTTACCGAGGGAATGAGAAAGCCAGCTTTCCTTTCTGGAGAGCAGAAGAAGGCT
ATCGTTGATCTTCTTTTCAAGACCAACAGAAAGGTTACCGTTAAGCAGTTGAAGGAG
GATTACTTCAAGAAGATCGAGTGCTTCGATTCTGTTGAAATCTCTGGAGTTGAGGAT
AGATTCAACGCTTCTCTTGGAACCTACCACGATCTTTTGAAGATCATCAAGGATAAG
GATTTCCTTGATAACGAGGAGAACGAGGACATCCTTGAGGACATCGTTCTTACCCT
TACCCTTTTCGAGGATAGAGAGATGATCGAGGAGAGACTCAAGACCTACGCTCACC
TTTTCGATGATAAGGTTATGAAGCAGTTGAAGAGAAGAAGATACACCGGATGGGGT
AGACTTTCTCGTAAGTTGATCAACGGAATCAGAGATAAGCAGTCTGGAAAGACCAT
CCTTGATTTCTTGAAGTCTGATGGATTCGCTAACAGAAACTTCATGCAGCTTATCCA
CGATGATTCTCTTACCTTCAAGGAGGACATCCAGAAGGCTCAGGTTTCTGGACAGG
GAGATTCTCTTCACGAGCACATCGCTAACCTTGCTGGATCTCCAGCTATCAAGAAG
GGAATCCTTCAGACCGTTAAGGTTGTTGATGAGCTTGTTAAGGTTATGGGTAGACA
CAAGCCAGAGAACATCGTTATCGAGATGGCTAGAGAGAACCAGACCACCCAGAAG
GGACAGAAGAACTCTCGTGAGAGAATGAAGAGAATCGAGGAGGGAATCAAGGAGC

TTGGATCTCAAATCTTGAAGGAGCACCCAGTTGAGAACACCCAGCTTCAGAACGAG
AAGTTGTACCTTTACTACCTTCAGAACGGAAGAGATATGTACGTTGATCAGGAGCTT
GACATCAACAGACTTTCTGATTACGATGTTGATCACATCGTTCCACAGTCTTTCTTG
AAGGATGATTCTATCGATAACAAGGTTCTTACCCGTTCTGATAAGAACAGAGGAAAG
TCTGATAACGTTCCATCTGAGGAGGTTGTTAAGAAGATGAAGAACTACTGGAGACA
GCTTCTTAACGCTAAGTTGATCACCCAGAGAAAGTTCGATAACCTTACCAAGGCTGA
GAGAGGAGGACTTTCTGAGCTTGATAAGGCTGGATTCATCAAGAGACAGCTTGTTG
AGACCAGACAGATCACCAAGCACGTTGCTCAGATCCTTGATTCTCGTATGAACACC
AAGTACGATGAGAACGATAAGTTGATCAGAGAGGTTAAGGTTATCACCTTGAAGTCT
AAGTTGGTTTCTGATTTCAGAAAGGATTTCCAGTTCTACAAGGTTAGAGAGATCAAC
AACTACCACCACGCTCACGATGCTTACCTTAACGCTGTTGTTGGAACCGCTCTTATC
AAGAAGTACCCAAAGTTGGAGTCTGAGTTCGTTTACGGAGATTACAAGGTTTACGAT
GTTAGAAAGATGATCGCTAAGTCTGAGCAGGAGATCGGAAAGGCTACCGCTAAGTA
CTTCTTCTACTCTAACATCATGAACTTCTTCAAGACCGAGATCACCCTTGCTAACGG
AGAGATCAGAAAGAGACCACTTATCGAGACCAACGGAGAGACCGGAGAGATCGTT
TGGGATAAGGGAAGAGATTTCGCTACCGTTAGAAAGGTTCTTTCTATGCCACAGGT
TAACATCGTTAAGAAAACCGAGGTTCAGACCGGAGGATTCTCTAAGGAGTCTATCC
TTCCAAAGAGAAACTCTGATAAGTTGATCGCTAGAAAGAAGGATTGGGACCCAAAG
AAGTACGGAGGATTCGATTCTCCAACCGTTGCTTACTCTGTTCTTGTTGTTGCTAAG
GTTGAGAAGGGAAAGTCTAAGAAGTTGAAGTCTGTTAAGGAGCTTCTTGGAATCAC
CATCATGGAGCGTTCTTCTTTCGAGAAGAACCCAATCGATTTCCTTGAGGCTAAGG
GATACAAGGAGGTTAAGAAGGATCTTATCATCAAGTTGCCAAAGTACTCTCTTTTCG
AGCTTGAGAACGGAAGAAAGAGAATGCTTGCTTCTGCTGGAGAGCTTCAGAAGGG
AAACGAGCTTGCTCTTCCATCTAAGTACGTTAACTTCCTTTACCTTGCTTCTCACTAC
GAGAAGTTGAAGGGATCTCCAGAGGATAACGAGCAGAAGCAGCTTTTCGTTGAGCA
GCACAAGCACTACCTTGATGAGATCATCGAGCAAATCTCTGAGTTCTCTAAGAGAG
TTATCCTTGCTGATGCTAACCTTGATAAGGTTCTTTCTGCTTACAACAAGCACAGAG
ATAAGCCAATCAGAGAGCAGGCTGAGAACATCATCCACCTTTTCACCCTTACCAAC
CTTGGTGCTCCAGCTGCTTTCAAGTACTTCGATACCACCATCGATAGAAAAAGATAC
ACCTCTACCAAGGAGGTTCTTGATGCTACCCTTATCCACCAGTCTATCACCGGACTT
TACGAGACCAGAATCGATCTTTCTCAGCTTGGAGGAGATAAGAGACCAGCTGCTAC
CAAGAAGGCTGGACAGGCTAAGAAGAAGAAGTGA

>*U6:gRNA*
(GenBank ID: KF264452. The *U6* promoter and the 20-nt guide sequence for the *AtPDS3* target 1 are colored in red and blue, respectively)
AGAAATCTCAAAATTCCGGCAGAACAATTTTGAATCTCGATCCGTAGAAACGAGACG
GTCATTGTTTTAGTTCCACCACGATTATATTTGAAATTTACGTGAGTGTGAGTGAGA
CTTGCATAAGAAAATAAAATCTTTAGTTGGGAAAAAATTCAATAATATAAATGGGCTT
GAGAAGGAAGCGAGGGATAGGCCTTTTTCTAAAATAGGCCCATTTAAGCTATTAAC
AATCTTCAAAAGTACCACAGCGCTTAGGTAAAGAAAGCAGCTGAGTTTATATATGGT
TAGAGACGAAGTAGTGATTGGACTTTTGCCAGCCATGGTGTTTTAGAGCTAGAAATA
GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT
GCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTA

Supplementary Database

Bioinformatically identified gRNA target sites in exons of *Arabidopsis* nuclear genes

The compressed text file includes 1,932,067 specific gRNA target sites bioinformatically identified in exons of *Arabidopsis* nuclear genes (see Supplementary Methods for detailed information regarding the database generation). Comprehensive Information is provided for each gRNA target site, including the chromosome number (column 1), the gene identification number (column 2), the start position on chromosome (column 3), the ending position on chromosome (column 4), DNA strand (column 5), gRNA target sequence (column 6), number of matches with 1 mismatch in the 12-nt seed sequence (column 7), number of matches in the seed sequence followed by NAG instead of NGG (column 8). Columns 9 and 10 suggest other types of genome matches which can be completely ignored based on current knowledge of gRNA/Cas9 DNA binding requirements. All gRNA target sequences listed in column 6 are specific to the corresponding genes. The information provided in columns 7 and 8 is only to inform the users about “potential” off-targets based on a more recent study[1] on the CRISPR/Cas specificity in *E. coli*, which indicated that some Cas9 activity also occurred at sites using an NAG instead of NGG as the PAM, and that some single-base mismatches in the seed sequence of gRNA were tolerated for Cas9 activity. Therefore, the priority of target selection may be given to those target sequences with minimal matches in columns 7 and 8.

Due to the existence of different transcript isoforms for some genes, some gRNA target sites are multiply presented. Therefore, only 1,466,718 gRNA target sites are unique, which cover >99% (26,942 out of 27,206) of *Arabidopsis* nuclear protein-coding genes.

Supplementary Methods

Plant growth. Wild-type *Arabidopsis thaliana* Columbia-0 plants were grown on Jiffy 7 soil (Jiffy Group) in a plant growth room with conditions maintained at 65% humidity and 75 μmol m^{-2} s^{-1} light intensity under photoperiods of 12 hr light at 23°C and 12 hr dark at 20°C[2]. Tobacco (*Nicotiana benthamiana*) plants were grown on Fafard soil (Fafard) under the same conditions as *Arabidopsis*.

Plasmid construction. Routine molecular cloning procedures were followed for plasmid construction and primers used are listed in the Supplementary Table 1. To express *hcoCas9* in protoplasts, PCR products of *hcoCas9* using the plasmid *p414-TEF-Cas9*[10] as template were digested by *Nco*I and *Stu*I, and were inserted into the same digested HBT-FLAG vector, which contains the hybrid constitutive promoter *35SPPDK* upstream of the *Nco*I site and an in-frame double FLAG tag coding sequence and the *NOS* terminator downstream of the *Stu*I site. To express *pcoCas9* (GenBank ID: KF264451) in protoplasts, *pcoCas9* encoding the same amino acids as *FLAG-NLS-SpCas9-NLS*[3] with *Arabidopsis* favored codons was synthesized in Genscript. The 189-bp *IV2* intron from potato was introduced into *pcoCas9* through overlapping PCR. PCR products were digested by *Bam*HI and *Pst*I and were inserted into the same digested HBT-FLAG vector (the vector-carrying double FLAG tag coding sequence was eliminated after *Pst*I digestion). To express gRNA in protoplasts, *Arabidopsis U6* polymerase III promoter was cloned using a long oligo as the reverse primer that included a gRNA aiming for the *AtPDS3* target 1 (GenBank ID: KF264452). PCR products were digested by *Sac*I and *Sma*I and were inserted into the *Sac*I and *Eco*RV digested pUC119-RCS vector. A second *U6* promoter-driven gRNA aiming for the *AtPDS3* target 2 was created by the overlapping PCR method illustrated in the Supplementary Figure 5 and was inserted into the *Pac*I site of the pUC119-RCS plasmid containing the expression cassette for the first *AtPDS3*-targeting gRNA. The gRNAs for other target sites were individually assembled with overlapping PCR (**Supplementary Fig. 5**) and were inserted into the *Asc*I site of the pUC119-RCS vector. To co-express *pcoCas9* and gRNA *in planta* through Agrobacteria, PCR products of *pcoCas9* with *IV2* intron were digested with *Xba*I and *Not*I and were inserted into the same digested pAN vector containing the constitutive *35S* promoter and the *NOS* terminator. The *35S:pcoCas9:NOS* cassette was cut out by *Sac*I and *Eco*RV and was inserted into the same digested pUC119-RCS vector. The *35S:pcoCas9:NOS* cassette was again cut out by *I-Ceu*I and *Asc*I and was inserted into the same digested pFGC-RCS binary vector. The gRNA expression cassette for the *AtPDS3* target 1 or *NbPDS* taget 1 was cut out from the pUC119-RCS-based plasmid by *Asc*I and was inserted into the same digested pFGC-RCS binary plasmid containing the *35S:pcoCas9:NOS* cassette. The resultant binary plasmids were introduced into *Agrobacterium tumefaciens* GV3101 cells through electroporation. To provide the DNA donor for the gRNA/pcoCas9-mediated HDR in *NbPDS*, a 648-bp genomic DNA spanning the *NbPDS* target site 1 was first cloned into the HBT-FLAG vector through *Bam*HI and *Stu*I sites. The *NbPDS* target site 1 in the plasmid was then mutated to possess an *Avr*II site. The resulting plasmid served as the PCR template to generate the 648-bp gPCR products, which were subsequently used as HDR templates. All recombinant plasmids were subjected to DNA sequencing for sequence verification.

Mesophyll protoplast isolation and transfection. Four-week-old *Arabidopsis* or tobacco plants were used for protoplast isolation by following the same procedure as previously described[2]. Briefly, leaves were cut into 1-mm strips with razor blade and were digested in 10 ml enzyme solution (1.5% Cellulase R10, 0.4% macerozyme R10, 0.4 M mannitol, 20 mM MES, pH 5.7, 20 mM KCl, 10 mM $CaCl_2$ and 0.1% BSA) for 3 hr. The slurry was shaken at 60 rpm for 3 min to facilitate cell release. After adding 10 ml W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl and 2 mM MES, pH 5.7), the mixture was filtered through a piece of Miracloth and protoplasts were pelleted by centrifugation at 1,200 rpm for 2 min in a CL2 centrifuge with swing buckets (Thermo Scientific). After resuspension in 10 ml W5 solution, protoplasts were left on ice for 30 min and then collected by centrifugation at 1,000 rpm for 30 sec in the CL2 centrifuge. The MMg solution (0.4 M mannitol, 15 $MgCl_2$ and 4 mM MES, pH 5.7) was used to resuspend protoplasts to a concentration of 2×10^5 cells per ml. DNA transfection was conducted in a 2-ml round-bottom microcentrifuge tube, where 200 μl protoplasts were well mixed with 20 μl DNA (2 μg/μl) and 220 μl polyethylene glycol (PEG) solution (40% PEG, v/v, 0.2 M mannitol and 0.1 M $CaCl_2$) and incubated for 5 min. Transfection was quenched by adding 800 μl W5 solution, and transfected protoplasts were harvested by centrifugation at 1,200 rpm for 2 min in the CL2 centrifuge. Cells were resuspended in 100 μl W5 solution and transferred into 1 ml WI solution (0.5 M mannitol, 20 mM KCl and 4 mM MES, pH 5.7) in a six-well plate pre-coated with 5% fetal calf serum. For NHEJ-mediated genome mutagenesis, plasmid DNA expressing *pcoCas9* and gRNA were used to co-transfect 200 μl protoplasts at different ratios, namely 10 μl: 10 μl (1:1) or 19 μl: 1 μl (19:1). For HDR-mediated gene replacement, 10 μl of plamid DNA expressing *pcoCas9*, 6 μl of plasmid DNA expressing gRNA and 4 μl (5 μg) of dsDNA donor were used to co-transfect 200 μl protoplasts. Additionally, 2 μl of plasmid DNA expressing *CYCD3* was co-transfected if required. Transfected protoplasts were incubated in dark at room temperature for 36 hr before mutagenesis analysis.

Detection and quantification of targeted mutagenesis. To visualize targeted mutagenesis in plant genomes, total genomic DNA (gDNA) was extracted from transfected protoplasts or agroinfiltrated leaves. For protoplasts, this was conducted by pelleting the 200 μl protoplasts, resuspending the cells in 50 μl sterile water and boiling the cells at 95°C for 10 min. For leaves, total gDNA was purified using the DNeasy Plant Mini Kit (Qiagen). PCR amplification of 200-300 bp target regions spanning individual gRNA target sequences was performed using Phusion high-fidelity DNA polymerase (New England Biolabs), where 2 μl protoplast lysates or 50 ng leaf gDNA was used as PCR template in a 50 μl PCR system for 30 amplification cycles. PCR products were digested with *Bam*HI and *Stu*I and were inserted into the same digested HBT-FLAG vector. Plasmid DNA was extracted from randomly selected colonies and was subjected to Sanger sequencing. DNA sequencing results of targeted mutants were visualized by 4Peaks program. The mutagenesis frequency was calculated as the ratio of mutated clonal amplicons vs. total sequenced clonal amplicons.

Agroinfiltration-mediated gene transfer *in planta*. A single colony of *Agrobacterium tumefaciens* GV3101 cells harboring the binary plasmid expressing *pcoCas9* and gRNA or expressing *pcoCas9* alone was grown overnight at 28°C in 5 ml LB medium with antibiotics. The next day, the culture was inoculated into 50 ml fresh LB medium containing antibiotics, 10 mM MES, pH 5.7, and 20 μM acetosyringone and was grown overnight at 28°C. The Agrobacteria cells were harvested and resuspended in the infiltration solution (10 mM $MgCl_2$, 10 mM MES, pH 5.7, and 100 μl acetosyringone) to an OD600 of 1.5 for *Arabidopsis* seedling infiltration or OD600 of 0.5 for tobacco leaf infiltration. The Agrobacteria cells were left at room temperature for 4 hr before infiltration. The two largest leaves from two-week-old *Arabidopsis* seedlings or all well-expanded leaves from five-week-old tobacco plants were subjected to agroinfiltration from the underside using a needleless 1 ml syringe. The gRNA/pcoCas9-mediated targeted mutagenesis *in planta* was examined 7 days post infiltration.

Bioinformatics identification of gRNA targets in *Arabidopsis* genome. We identified all gRNA target sites in the TAIR10 *Arabidopsis* genome according to the previously described computing method[4] with some modifications: we assumed the size of the seed sequence in a gRNA to be 12 rather than 13 nt, eliminated the requirement for the gRNA to start with a G, and filtered out all target sites whose 5'-most 20 bp contained six consecutive Ts. The resulting 2,939,334 target sites were all of the form 5'-N(8)S(12)NGG-3', where S(12) is the 12-nt seed sequence and NGG is the PAM for SpCas9. These target sites have no other occurrences of S(12)NGG in the *Arabidopsis* genome, satisfying the criterion of specificity[5]. A more recent study on the CRISPR/Cas specificity in *E. coli* indicated that some Cas9 activity also occurred at sites using an NAG instead of NGG as the PAM, and that some single-base mismatches in the seed sequence of gRNA were tolerated for Cas9 activity[1]. Therefore, for each gRNA target site identified in the *Arabidopsis* genome, we counted all genomic occurrences of S(12)NAG and sequences with a single mismatch to S(12) while followed by an NGG, in order to provide users with information regarding how many potential off-targets of these sorts exist. We then mapped this whole-genome gRNA target site database to the *Arabidopsis* exome and annotated all target sites whose expected DNA cleavage by gRNA/Cas9 is within or at the boundary of the coding sequence. Due to the existence of different transcript isoforms for the same gene, some gRNA target sites were multiply presented. In total, 1,466,718 unique gRNA target sites were identified from exons of >99% (26,942 out of 27,206) of *Arabidopsis* nuclear protein-coding genes.

Supplementary References

1. Jiang, W. *et al. Nat. Biotechnol.* **31**, 233-239 (2013).

2. Yoo, S.D. *et al. Nat. Protoc.* **2**, 1565-1572 (2007).

3. Cong, L. *et al. Science* **339**, 819-823 (2013).

4. Mali, P. *et al. Science* **339**, 823-826 (2013).

5. Jinek, M. *et al. Science* **337**, 816-821 (2012).