SCIENTIFIC REP⚙RTS

OPEN

CRISPR-Cas9 Editing in Maize: Systematic Evaluation of Off-target Activity and Its Relevance in Crop Improvement

Joshua Young[1], Gina Zastrow-Hayes[1], Stéphane Deschamps[1], Sergei Svitashev[1], Mindaugas Zaremba[2], Ananta Acharya[1], Sushmitha Paulraj[1], Brooke Peterson-Burch[1], Chris Schwartz[1], Vesna Djukanovic[1], Brian Lenderts[1], Lanie Feigenbutz[1], Lijuan Wang[1], Clara Alarcon[1], Virginijus Siksnys[2], Gregory May[1], N. Doane Chilcoat[1] & Sandeep Kumar[1]

CRISPR-Cas9 enabled genome engineering has great potential for improving agriculture productivity, but the possibility of unintended off-target edits has evoked some concerns. Here we employ a three-step strategy to investigate Cas9 nuclease specificity in a complex plant genome. Our approach pairs computational prediction with genome-wide biochemical off-target detection followed by validation in maize plants. Our results reveal high frequency (up to 90%) on-target editing with no evidence of off-target cleavage activity when guide RNAs were bioinformatically predicted to be specific. Predictable off-target edits were observed but only with a promiscuous guide RNA intentionally designed to validate our approach. Off-target editing can be minimized by designing guide RNAs that are different from other genomic locations by at least three mismatches in combination with at least one mismatch occurring in the PAM proximal region. With well-designed guides, genetic variation from Cas9 off-target cleavage in plants is negligible, and much less than inherent variation.

Humans have been practicing crop improvement for many millennia. Since the advent of the Green Revolution in the 1960's, agriculture has used modern technology to improve the yield and nutritional quality of crops[1]. Conventional breeding, however, is unlikely to keep up with the world's increasing food demand[2]. Genome editing approaches utilizing site-directed endonucleases (SDNs) capable of making chromosomal double-strand breaks (DSBs)[3–7] can help overcome the limitations of conventional breeding and accelerate development of improved crops. By harnessing natural cellular DNA repair processes, DSBs can be used to introduce targeted disease resistance, genome edits to improve agronomic traits (e.g., plant grain yield, nutritional content)[8–12] or speed-up domestication, ultimately, helping to accelerate the development of new beneficial plant varieties for society[13]. These edits include small insertion or deletion (indel) modifications that can knockout the expression of a gene (SDN1), or correction of the DSB with a homologous template DNA resulting in the precise alteration of a sequence (SDN2), or the targeted insertion of a new DNA sequence (SDN3)[14,15].

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9[6] has emerged as a robust and versatile DSB tool with broad applications for gene discovery, trait development and expedited breeding in crop species. Cas9 endonuclease and guide RNA can be delivered into plant cells as DNA, RNA or ribonucleoprotein (RNP) to cleave target DNA sequence(s) in the genome. However, in addition to the intended target (on-target) site, Cas9 can potentially cause off-target DSBs at genomic locations with significant sequence similarity to that of the intended target sequence[16,17] resulting in the possibility of off-target edits. To mitigate the potential for off-target editing, a variety of approaches have been developed. These include protein engineering[18–20], genome editing using ribonucleoprotein (RNP)[21–24], biochemical and cellular assays to empirically assess target specificity[25–29], and computational identification of Cas9 targets with low potential for off-target editing[30–34].

[1]Corteva Agriscience™, Agriculture Division of DowDuPont™, Johnston, IA, 50131, USA. [2]Institute of Biotechnology, Vilnius University, Vilnius, LT, 10257, Lithuania. Joshua Young, Gina Zastrow-Hayes and Stéphane Deschamps contributed equally. Correspondence and requests for materials should be addressed to S.K. (email: sandeep. kumar@corteva.com)

1
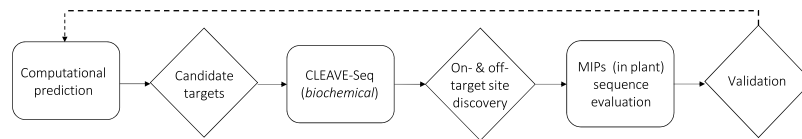
Figure 1. Overview of biochemical off-target site identification and in plant validation workflow.

Here we explored the specificity of *Streptococcus pyogenes* Cas9 (Cas9) endonuclease in the complex crop genome of maize (*Zea mays* L.) using a comprehensive approach that included: (1) *in silico* computational prediction of an off-target site portfolio; (2) biochemical confirmation of off-target cutting activity and (3) surveillance of candidate off-target sites in a cellular context. The effect of off-target editing as a function of Cas9 and guide RNA delivery was assessed by three different methods: DNA-free (using RNPs and particle gun (PG)), and DNA-based delivery using *Agrobacterium* or PG.

Our results show that bioinformatic selection of unique target sites can be used as a reliable tool to mitigate the potential for off-target editing in crop plants with reference genomes. Potential off-target sites were identified with genome-wide biochemical assay, CLEAVE-Seq, that provides increased sensitivity over similar methods[26,28]. The biochemically identified sites were subsequently monitored using molecular inversion probes (MIPs) in maize plants subjected to Cas9 editing. Computationally unique targets demonstrated no evidence of off-target cutting in a cellular context, while up to ~90% of all the observed alleles had on-target activity. At a limited number of genomic sites, we also report that inherent genetic variation in the genotype used in this study far exceeded potential genetic changes generated by CRISPR-Cas9 genome editing techniques. Therefore, without a targeted approach such as one described here, whole genome sequencing may not be a practical way to differentiate CRISPR-Cas9 off-target effects from inherent background variation in plants.
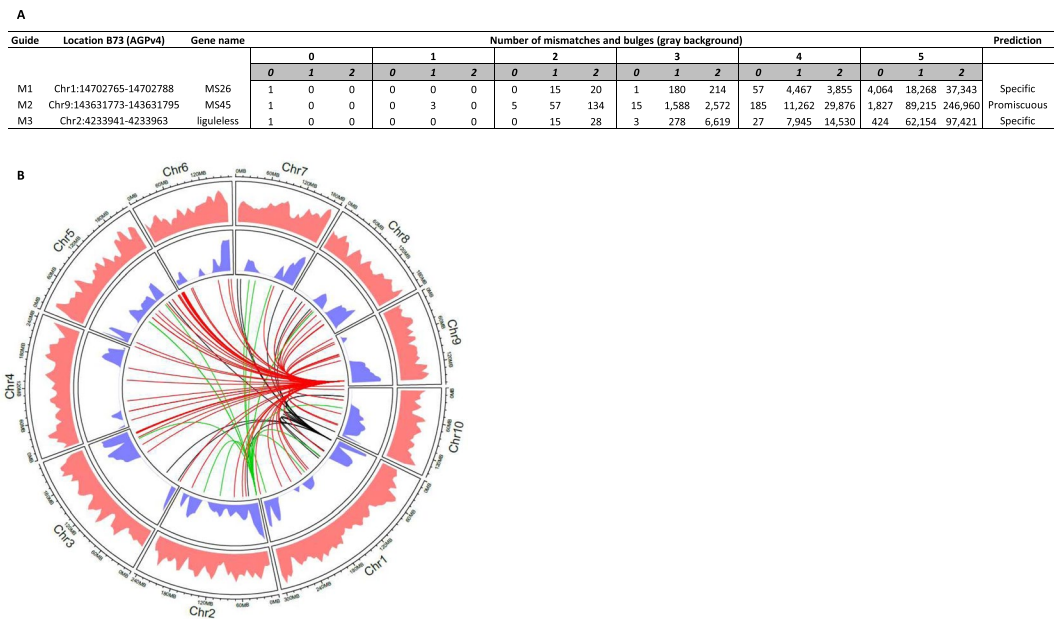
To our knowledge, this is the first comprehensive study of CRISPR-Cas9 specificity in plants highlighting prediction and validation of unintended genome editing, and their relevance in the background of innate genetic variation.

Results

We designed a three-step approach to evaluate the specificity of CRISPR-Cas9 editing activity. This included the computational prediction of target specificity, the biochemical capture and identification of genomic sequences susceptible to Cas9 induced DSBs, and off-target site validation in plants (Fig. 1). First, Cas-OFFinder[32] was utilized to predict the specificity of targets. Next, a new biochemical method, CLEAVE-Seq, was used for the biochemical discovery of candidate off-target sites as the method avoids potential complex steps associated with other biochemical methods[25,27,35] and enhances on- and off-target discovery. Finally, MIPs analysis was performed in plants to examine off-target sites identified computationally and confirmed biochemically[36]. MIPs was chosen due to its scalability for throughput and multiplexable analysis capability. Thus, permitting many genomic loci to be monitored simultaneously for off-target non-homologous end-joining (NHEJ) mutations. A similar two-step strategy, Verification of *In Vivo* Off-targets (VIVO), has recently been used to identify and evaluate off-target cutting in the mouse genome[37].

Target sites and computational predictions. Three targets sites spread across different chromosomes in the maize genome (Fig. 2) were selected for analyses. Each target was chosen to be within a gene non-essential for embryonic cell proliferation and plant regeneration. Guide RNA design and prediction was accomplished using Cas-OFFinder[32] set to search for all potential off-target sites with up to 5 mismatches and 2 bulges between the guide RNA and DNA target sequence. Guide RNAs M1 and M2 were designed to target the male sterile 26 (*Ms26*)[38,39] and 45 (*Ms45*) genes[40,41], respectively. The M3 guide RNA was designed to target the liguleless 1 (*Lig1*) gene[42]. Only targets adjacent to an appropriate protospacer adjacent motif (PAM) for Cas9 (NAG or NGG) were considered. M1 and M3 were designated as specific based on their lack of homology with other sequences in the maize B73 reference genome, AGPv4[43]. For example, the most closely related sequences to M1 and M3 targets were different by at least 2 mismatches and 1 bulge. Additionally, these guides had no predicted off-target site containing a perfect match within the PAM proximal seed region (10 bases 5′ of the PAM)[44,45] of their respective target site (Supplementary Tables 1 and 2). In contrast to M1 and M3 guides, M2 guide had multiple matching genomic sites (Fig. 2) with 1 or 2 mismatches and at least 1 bulge, without any mismatch or bulge in the PAM proximal seed region (Fig. 2A and Supplementary Table 3). The M2 guide, designated as promiscuous, was intentionally selected for its potential of inducing off-target edits, which would also serve as a positive control for method validation.

Genome-wide identification of off-target sites. The genomic specificity of Cas9 was examined biochemically using CLEAVE-Seq at 37 °C. Previous experiments in maize demonstrated that the modifications made here (described in the Methods section) increased the number of Illumina reads mapping to various cut sites biochemically, both on-target and off-target, by approximately 10-fold over the original SITE-Seq protocol (Supplementary Fig. 1). Specifically, a phosphatase treatment step prior to RNP cleavage was added to reduce adapter ligation to random ends generated by physical shearing of genomic DNA templates during DNA extraction and pipetting. An additional step of DNA release through NotI cleavage after biotinylated adapter ligation and biotin selection, followed by exonuclease treatment and second strand synthesis with a semi-random primer (see Methods) was added to the protocol. This step further improves specificity of the final PCR amplification step

A

Guide	Location B73 (AGPv4)	Gene name	Number of mismatches and bulges (gray background)																	Prediction
			0			1			2			3			4			5			
			0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	
M1	Chr1:14702765-14702788	MS26	1	0	0	0	0	0	0	15	20	1	180	214	57	4,467	3,855	4,064	18,268	37,343	Specific
M2	Chr9:143631773-143631795	MS45	1	0	0	0	3	0	5	57	134	15	1,588	2,572	185	11,262	29,876	1,827	89,215	246,960	Promiscuous
M3	Chr2:4233941-4233963	liguleless	1	0	0	0	0	0	0	15	28	3	278	6,619	27	7,945	14,530	424	62,154	97,421	Specific

Figure 2. Computational prediction of guide and off-target sites. (**A**) Summary table indicating number of genomic sites with up to 5 mismatches and 1 bulge. (**B**) Circos plot representing the guides used for this study and their potential off targets. Layers outermost to innermost: B73 AGPv4 chromosomes, repeats (pink), genes (blue), guides and their computationally predicted off-sites up to 2 mismatches and 1 bulge. All lines originate from the on-site. Black: M1, Red: M2, Green: M3.

prior to DNA sequencing, replacing PCR amplification performed directly off streptavidin bead-bound DNA in the original protocol. Two independent CLEAVE-Seq replicates were performed for each guide.

Analysis was performed by scanning chromosomally mapped CLEAVE-Seq data for discontinuities in sequence read coverage within a $+/-2$ bp window of the expected cut site for Cas9 (3 bp 5′ of the PAM) similar to that described in Cameron, *et al.*[28]. Since Cas9 has been reported to generate off-target cutting in eukaryotic cells in sequences with up to 5 mismatches and tolerate bulges between guide RNA and DNA target recognition[25,46–48], we limited our search for CLEAVE-Seq signatures to protospacer targets within 5 mismatches and 2 bulges (a total of 7 differences) from the on-target protospacer sequence (Fig. 2A). To account for off-targets generated by mismatches in PAM recognition, protospacer targets adjacent to a 3′ prime NAG were also considered in addition to the canonical 3′ NGG PAM. Next, this collection of computationally predicted sites, permitting up to 8 differences from the on-target site, were searched for the presence of cleavage signatures.

CLEAVE-Seq was first performed for the promiscuous M2 guide RNA to optimize the target discovery process for the two-remaining guide RNAs computationally predicted to be specific. As observed previously with other biochemical based methods[25,26,28], many cleaved genomic target sites were detected (3,052 in total (Supplementary Table 4). This feature of off-target detection methods is perhaps due soley to their biochemical nature. Being devoid of a cellular, nuclear, and chromosomal context, we speculate that Cas9 and guide RNA under high persistent concentrations are capable of cleaving target sites that are less likely to be edited in cellular context[49]. Next, the average number of reads (normalized based on CLEAVE-Seq library read depth) identified at the target cut-site were compared with the number of mismatches and bulges relative to the on-target site. For example, 227 CLEAVE-Seq reads (normalized) were recovered at the on-target site, M2-1 (Table 1). A relatively high read count (>40) was observed for the sites M2-2, M2–4 and M2-6, which all carried 1-2 mismatches and 1-2 bulges (Table 1). None of the mismatches or bulges in these sites were within the PAM-proximal seed region. In comparison, the rest of the sites, all with significantly lower read counts, contained 2 or more mismatches and 1 or more bulge, but with the difference that at least 1 mismatch was located within the seed region (Table 1) or PAM (Supplementary Table 4). This observation confirms that the binding interaction between guide RNA and target sequence is more sensitive to mismatches within the PAM proximal seed sequence[50]. Interestingly, off-target M2-6 resulted in more reads than the on-site. This could be the result of more robust biochemical cleavage efficiency at this target under the digest conditions used or the result of enhanced recovery for this location by CLEAVE-Seq. Finally, low numbers of CLEAVE-Seq reads were observed from two of the computationally predicted sites with 1–2 mismatches and 1 bulge (M2-3, M2-5 in Table 1 & site 4 in Supplementary Table 5). In each of these three cases, each site had either a bulge or a mismatch in the seed region.

Next, CLEAVE-Seq was performed for the two guides computationally predicted to be specific, M1 and M3. Similar to the CLEAVE-Seq data obtained from guide M2, high on-target normalized read counts were detected for both M1 (Table 2, M1-1) and M3 (Table 3, M3-1). No other M1 or M3 sites were detected with more than 19 normalized reads, except for the M1-2 site where 93 CLEAVE-Seq reads were recovered. This target had 2 mismatches and 2 bulges with 1 mismatch being within the PAM proximal seed sequence of the protospacer (Table 2).

Site	Computational prediction		CLEAVE-Seq reads	In plant validation (% Mutant Alleles)			Sequence																							Genomic location
	Mismatch	Bulge		Agro	DNA-PG	RNP-PG	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	-1	-2	
M2-1	0	0	227	99.0%	95.9%	33.3%	G	G	C	C	G	A	G	G	T	C	G	A	C	T	A	C	C	G	G	C	*C*	*G*	*G*	Chr9:143631773-143631795
M2-2	1	1	48	33.1%	89.3%	0.0%	■										G												*A*	Chr3:200475375-200475397
M2-3	1	1	12	0.0%	0.0%	0.0%				A												■							*G*	Chr6:166308906-166308927
M2-4	2	1	111	1.2%	7.4%	0.0%	■					C			G														*A*	Chr10:5286571-5286593
M2-5	2	1	1	0.0%	0.0%	0.0%			T					■								C							*A*	Chr6:157449386-157449407
M2-6	2	2	529	1.1%	1.2%	0.0%	T		■	■			G																*A*	Chr5:37860732-37860754
M2-7	3	1	4	0.0%	0.0%	0.0%	■						C		G			G											*A*	Chr1:5942859-5942881
M2-8	3	1	2	0.0%	0.0%	0.0%				A		c											A	C					*G*	Chr10:39633639-39633661
M2-9	3	2	11	0.0%	0.0%	0.0%	■							C	G									T					*G*	Chr2:80845032-80845054
M2-10	3	2	11	0.0%	0.0%	0.0%	A		ga	T			A																*G*	Chr7:173232468-173232490
M2-11	4	1	3	0.0%	0.0%	0.0%		A											A			G	A						*G*	Chr1:168558804-168558826
M2-12	4	1	16	0.0%	0.0%	0.0%		A											A			A	A						*G*	Chr10:84440789-84440811
M2-13	4	1	9	0.0%	0.0%	0.0%		A					A						A				A						*G*	Chr2:197687616-197687638
M2-14	4	1	1	0.0%	0.0%	0.0%		A											A			G	A						*G*	Chr2:45333382-45333404
M2-15	4	1	1	0.0%	0.0%	0.0%	A					C							C				C						*C*	Chr4:233088295-233088317
M2-16	4	1	2	0.0%	0.0%	0.0%			t		G			A			T			T									*T*	Chr5:6293703-6293726
M2-17	4	1	3	0.0%	0.0%	0.0%	■	T						C		C							A						*G*	Chr5:82350877-82350899
M2-18	4	2	3	0.0%	0.0%	0.0%	A		ga	G											G	A							*G*	Chr3:168788394-168788416
M2-19	5	0	3	0.0%	0.0%	0.0%		A	A				G		C			G											*T*	Chr1:124111388-124111410
M2-20	5	1	5	0.0%	0.0%	0.0%	A	■			T			A						T	T								*C*	Chr1:247722181-247722203
M2-21	5	1	1	0.0%	0.0%	0.0%		■	A	A				G		C			G										*T*	Chr8:78781894-78781916

Table 1. CLEAVE-seq data and validation of M2 sites in plants. On-target site M2-1 is shown on the top. Percent mutant allele is number of alleles with mutation/total number of alleles observed. DNA and RNA bulges are shown as gray and black boxes, respectively. Lower case indicates additional nucleotides in RNA. Mismatch is shown as a letter indicating the nucleotide, absence of a nucleotide indicates no mismatch. PAM is italicized.

Validation of potential off-target sites in plants.

Following the computational and biochemical identification of potential off-target sites for M2, we validated off-target cleavage in maize using three different plant Cas9-guide RNA delivery methods: DNA-free (using RNP and particle gun (PG), and DNA-based using *Agrobacterium* or PG. First, 21 genomic targets sites, including the on-target site, were selected from a wide range of CLEAVE-Seq read count conditions and examined in a cellular context for evidence of DNA cleavage and repair. Reasoning that genomic targets with a higher CLEAVE-Seq read coverage are more likely to be cut, all sites with greater than or equal to 40 normalized reads originating from the cut-site were selected. Additionally, targets were selected to come from all mismatch (1-5) and bulge (1-2) categories (Supplementary Table 6). To validate our target discovery process, an additional 75 targets not identified by CLEAVE-Seq but computationally predicted with different mismatch classifications, were also selected for characterization in maize plants (Supplementary Table 6). MIPs analysis was then performed on all 96 targets in ~300 T_0 plants exposed to Cas9 and the M2 guide RNA using three different plant delivery methods, *Agrobacterium*, particle gun (PG)-mediated DNA or RNP (Supplementary Table 7). High efficiency on-target mutagenesis (>95% of the total number of alleles) was observed in the T_0 maize plants obtained using *Agrobacterium* or PG-mediated DNA delivery (Table 1, M2-1 site). Targeted mutagenesis was also obtained using PG-mediated RNP delivery, although the frequency of alleles mutated using this method was much lower (33% of the total alleles analysed). No off-target activity was observed in any of the T_0 plants using the RNP delivery.

As expected, M2 off-target sites identified by CLEAVE-Seq were also mutated with *Agrobacterium* and PG-mediated DNA delivery, though at varying frequencies (Table 1). High-frequency off-target activity was observed at site M2-2 with 33% and 89% mutant alleles obtained using *Agrobacterium* and PG-mediated DNA delivery methods, respectively. Low-frequency off-target mutation (1.1–7.4% of total alleles analysed) was also detected at sites M2-4 and M2-6. No activity was observed for the other 17 candidate off-target sites identified with CLEAVE-Seq. Notably, only off-target sites with a normalized CLEAVE-Seq read count greater than 40 were modified in plants (Table 1). Moreover, analysis of the 75 computationally predicted M2 targets not detected by CLEAVE-Seq also did not produce evidence of cutting activity in plants.

Since *Agrobacterium*-mediated DNA delivery provided the highest frequency of on-target site editing with the lowest propensity for off-target cleavage, validation of potential off-targets for M1 and M3 guide RNAs in plants was performed only with this method. Additionally, given that off-targets demonstrated to be cleaved bio-chemically were modified in plants, MIPs analysis was conducted only for those sites identified by CLEAVE-Seq. Off-target sites for M1 and M3 guide RNAs were selected from a wide range of read counts with a preference being given to those with higher normalized read coverage. Similar to guide RNA M2, MIPs analysis revealed robust on-target activity for M1 and M3 guide RNAs with mutation frequencies of 61% and 94%, respectively, of the total alleles analysed in ~50 T_0 plants (Tables 2 and 3, respectively; Supplementary Table 7). In contrast to guide RNA M2, M1 and M3 guide RNAs produced no detectable off-target activity.

Next, computational predictions and data from CLEAVE-Seq and MIPs analyses were examined for trends that could be utilized to avoid off-target cleavage in plants. For guide RNAs M1, M2, and M3, CLEAVE-Seq identified targets with a normalized read count of greater than 40 yielded evidence of off-target cleavage and repair in plants (Tables 1–3), which suggests that targets with a higher read count tend to be more prone to modification in plants. Overall, this feature of CLEAVE-Seq translated well, accurately predicting 6 of the 7 targets showing editing in plants. Moreover, none of the targets modified in plants contained mismatches or bulges in the PAM proximal region of the protospacer and fewer than a combination of 5 mismatches and/or bulges in the PAM distal region of the protospacer. Taken together, this indicates that off-target editing can be significantly reduced in

Site	Computational prediction		CLEAVE-Seq reads	In plant validation (Agro)	Sequence																								Genomic location
	Mismatch	Bulge		% Mutant Alleles	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	-1	-2	
M1-1	0	0	201	61.3%	G	C	A	C	G	T	A	C	G	T	C	A	C	C	A	T	C	C	C	G	C	*C*	*G*	*G*	Chr1:14702765-14702788
M1-2	2	2	93	0.0%						A	▪	▪													T	*A*			Chr10:138875022-138875045
M1-3	3	1	7	0.0%						A	a					T						A				*G*			Chr10:120497792-120497768
M1-4	3	1	5	0.0%						A	a					T						A				*G*			Chr10:120753156-120753132
M1-5	3	1	3	0.0%						C				▪						A	G					*G*			Chr4:38687989-38688011
M1-6	4	1	4	0.0%	▪		T		A										C		A					*C*			Chr4:170037576-170037599
M1-7	4	1	19	0.0%	C		▪			T	A													T		*G*			Chr4:53255942-53255919
M1-8	4	1	15	0.0%			G	C		a			C		G										*T*			Chr9:14541330-14541354
M1-9	5	1	5	0.0%	G	C		C			▪			G											*G*			Chr1:3497051-3497072
M1-10	5	1	4	0.0%		▪		G		G	T				A	C										*C*			Chr2:114192761-114192783
M1-11	5	1	4	0.0%	a		G			A	A					A	G								*G*			Chr6:32905376-32905399
M1-12	5	1	6	0.0%				▪	T	T	A		T	T												*C*			Chr7:135823779-135823801
M1-13	5	1	8	0.0%	a		T		G		C	A			A										*T*			Chr7:26447133-26447156
M1-14	5	1	9	0.0%	G	T		A	▪								T		A						*C*			Chr8:174780615-174780640
M1-15	5	1	4	0.0%	C			G	▪		A	A									A				*G*			Chr8:53709884-53709861
M1-16	5	2	5	0.0%		G	G		gc	G				G							A				*C*			Chr2:208170984-208170961

Table 2. CLEAVE-seq data and validation of M1 sites in plants. On-target site M1-1 is shown on the top. Percent mutant allele is number of alleles with mutation/total number of alleles observed. DNA and RNA bulges are shown as gray and black boxes, respectively. Lower case indicates additional nucleotides in RNA. Mismatch is shown as a letter indicating the nucleotide, absence of a nucleotide indicates no mismatch. PAM is italicized.

Site	Computational prediction		CLEAVE-Seq reads	In plant validation (Agro)	Sequence																							Genomic location
	Mismatch	Bulge		% Mutant Alleles	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	-1	-2	
M3-1	0	0	164	94.3%	G	C	G	G	A	G	A	C	T	A	A	G	T	G	G	C	T	G	T	A	*G*	*G*	*G*	Chr2:4233941-4233963
M3-2	3	1	14	0.0%	T							A			a								C		*T*			Chr3:211837283-211837306
M3-3	3	2	7	0.0%		▪	▪			G	G					G									*C*			Chr1:135533537-135533562
M3-4	3	2	11	0.0%		▪	▪			G	A					G									*C*			Chr1:135560571-135560593
M3-5	3	2	19	0.0%	A	▪	▪				A				A										*T*			Chr4:155729295-155729317
M3-6	4	2	7	0.0%	A	▪	▪	T			A											T		*G*			Chr2:234759654-234759632
M3-7	5	1	6	0.0%						G	A							C	A	C	a			*G*			Chr3:74618205-74618227

Table 3. CLEAVE-seq data and validation of M3 sites in plants. On-target site M3-1 is shown on the top. Percent mutant allele is number of alleles with mutation/total number of alleles observed. DNA and RNA bulges are shown as gray and black boxes, respectively. Lower case in gray box indicates additional nucleotide(s) in RNA. Mismatch is shown as a letter indicating the nucleotide, absence of a nucleotide indicates no mismatch. PAM is italicized.

plants by computationally selecting unique Cas9 targets that are different from other genomic locations by at least a combination of 5 mismatches and/or bulges in the PAM distal region or at least three differences in the PAM distal region with at least one additional discrepancy in the PAM proximal region (Tables 1–3).

Natural variation in control plants. To evaluate polymorphism in the Hi-II maize genotype used in this study, we obtained MIPs data for 19 M1, 96 M2, and 8 M3 sites obtained from 500 non-transformed control plants. The MIPs data when compared to B73 AGPv4 revealed single nucleotide variations (SNVs) in 50% of the M2 sites analysed. Similarly, SNVs were observed in 12% and 16% of M3 and M1 MIPS sites, respectively (Table 4). In total, 228 variant nucleotides were observed out of 8033 nucleotides sequenced from 96 M2 sites.

Discussion

CRISPR-Cas based genome editing is very precise[51,52] compared to other crop improvement technologies such as traditional or mutational breeding. However, unintended off-target cleavage and editing can occur at locations within the genome that share significant sequence similarity to the intended target site[17,50,53]. In this report we assessed CRISPR-Cas9 specificity in maize, a crop plant with a complex genome similar in size to the human genome. Three different genes, *Ms45*, *Ms26*, and *Lig1*, each residing on one of the three different chromosomes, were targeted for cleavage with three different Cas9-guide RNA delivery methods (DNA-free (using RNP and PG), and DNA-based delivery using *Agrobacterium* or, PG). Two specific and one promiscuous guide RNAs were used to develop our methodologies. Biochemically cleavable genomic off-target sites were subsequently identified using a new sensitivity enhanced biochemical method, CLEAVE-Seq. Next, we validated our biochemical cleavage method, first by examining a portfolio of targets not detected by CLEAVE-Seq but computationally predicted as potential off-target sites. Finally, the relevance of biochemically identified targets to cleavage in T_0 plants was established. In all, none of the computationally predicted targets that were not detected by CLEAVE-Seq yielded sequence alterations in plants, while 6 of the 7 (86%) targets identified by CLEAVE-Seq with a read coverage greater than 40 were modified. RNP delivery with no off-target cutting being observed with the promiscuous M2 guide, provided highest specificity,

Guide	No of Sites	Nucleotides sequenced	Variant sites	Variant nucleotides
M1	19	1078	3 (16%)	4
M2	96	8033	48 (50%)	228
M3	8	480	1 (12%)	2

Table 4. Inherent variation in Hi-II genotype. Same "number of sites" (Column 2) were MIPS analyzed in T_0 plants CRISPR/Cas9-treated plants. "Nucleotides sequenced" indicate total sequence coverage calculated by multiplying sequence length of MIPS assay to number of sites analyzed. "Variant sites" show the total number of sites observed with at least one nucleotide variation. "Variant nucleotides" indicate total number of nucleotide variation observed.

which is consistent with previous reports[21–24]. This enhancement to specificity, however, came with a lower on-target mutation frequency. Also, DNA-PG due to its propensity to deliver more copies of the expression cassettes resulting in higher cellular concentrations of Cas9-gRNA, could be a better approach to evaluate potential off-target activity. Overall, the delivery of Cas9 and guide RNA by *Agrobacterium* provided the best balance between on-target editing and off-target cleavage for future applications aimed at plant improvement.

The results from the M2-6 site in this report present an intriguing case study. CLEAVE-Seq read coverage for this site far exceeded on-target (M2-1) site. However, unexpectedly low mutation frequency was observed in T_0 plants. Interestingly, the M2-6 target site was mapped to an intergenic region while all other target sites exhibiting more robust mutation frequencies were located within genic regions (M2-1: *Ms45* and M2-2: uncharacterized gene LOC100217080). This finding is consistent with previous report where indel frequencies were significantly reduced in the absence of gene transcription[54].

Our data confirms previous reports demonstrating high-specificity of CRISPR-Cas9-mediated genome editing in plants[50,51,55–57]. Expected off-target activity in plants was observed for the promiscuous M2 guide RNA at sites predicted bioinformatically and identified biochemically. All sites showing evidence of cutting in plants were observed to contain mismatches and/or bulges only in the PAM distal region (nucleotides 11–20) outside of the PAM proximal region (nucleotides 1–10) reported to be the seed region necessary for establishing R-loop formation[44,58,59]. This hyper-sensitivity to mismatches in the seed region of the target sequence has previously been reported in plants[50] and documented in biochemical studies examining Cas9 specificity during R-loop formation[58]. Taken together, our data fully support the notion of computationally predicting specific target sites with at least a 3-nt mismatch, bulge, or mismatch bulge combination with at least one difference being in the seed sequence of the protospacer to ameliorate off-target Cas9 activity[50]. Interestingly, these results are inconsistent with Cas9 specificity reports in mammalian cells[25–27]. One major difference between mammalian and plant cells are the temperature at which tissue cultures are maintained. In the case of human cells, this is 37 °C while for maize the temperature is significantly lower, 28 °C. This difference may have significant effects on Cas9 cleavage activity and specificity[60].

Inherent genetic diversity in maize is extensive[61]. Maize high type II (Hi-II) line[62] was generated by crossing two partially inbred lines (Hi-II Parents A and B) selected from a cross between A188 and B73 which demonstrated a greatly improved Type-II tissue culture response. MIPs analysis of 96 M2 targets revealed variation in 50% of the sites among 500 non-treated plants screened in this study. A portion of the genetic variation observed could be attributed to the parental origins of the Hi-II line but also may be the result of *de novo* spontaneous mutations over many generations. Genetic variation due to spontaneous mutations during multiple generations of breeding is well established in plants including major food crops[63,64]. Additionally, all conventional crop improvement methods including crosses between different genotypes, varieties, and species, chemical or irradiation mutagenesis, and plant tissue culture inherently produces random genetic variation[50,63–71]. Taken together, our study indicates that with appropriately designed guide RNAs, genetic changes observed through CRISPR-Cas9 off-target editing are negligible and much less than the naturally occurring diversity in plants.

The work described herein illustrates that CRISPR-Cas9 is remarkably specific and efficient at generating on-target genome edits. Robust bioinformatics tools alleviate the potential for off-target cleavage by identifying targets that are unique within the genome by at least combination of 3 mismatches and/or bulges with at least 1 difference within the PAM proximal seed region. Furthermore, biochemical identification of off-target sites using CLEAVE-Seq followed by MIPs validation in plants complements computational guide RNA predictions, in particular when the intended target site is not unique and has highly similar sequences elsewhere in the genome. While CRISPR-Cas9 has the potential to generate off-target cutting in genomic sites that are substantially similar to the target site, off-target edits are likely to be negligible in the background of existing natural variation and continuous unintended changes being generated during the plant breeding process. Finally, regardless of the breeding method, standard practices of commercial crop development include advancement of candidate lines following extensive agronomic evaluations specific for a given crop. This has proven to be an effective tool to eliminate plants with undesirable characteristics resulting in crops with a history of safe use. Therefore, concerns related to specificity of CRISPR-Cas9 technology in crop improvement have little relevance.

Methods

Computational prediction of target sites and guide selection.
Target site selection was based on the number of potential off-target sites observed at 0, 1, 2, 3, 4, and 5 nt mismatches from the guide RNA protospacer target permitting up to 2 bulges between guide RNA spacer and protospacer target sequence. Mismatch and bulge assessments were performed using Cas-OFFinder[32] against the B73 (one of the Hi-II parent line) reference AGPv4[43] using a NRG protospacer adjacent motif (PAM). We used a command line/custom version

of Cas-OFFinder to include reference AGPv4. Two target sites were selected to be specific (have low off-target potential) and one target site was chosen to be promiscuous (have high off-target potential) based on the number of homologous targets identified with up to a combination of 2 mismatches and bulges (Fig. 2A,B).

Cas9 protein and guide RNA molecules.   Recombinant *Cas9* protein containing a C-terminal 6X His was expressed and purified from *E. coli* as described previously in Karvelis, T. *et al.*[72]. Single guide RNAs (guide RNAs) were generated by T7 *in vitro* transcription using AmpliScribe™ T7-Flash™ kit (Epicentre, USA) according to the manufacturer's recommendations. Products were purified using NucAway™ Spin Columns (Invitrogen, Life Technologies Inc., USA) followed by ethanol precipitation.

CLEAVE-Seq.   On-site and off-site detection was performed using a new biochemical method called CLEAVE-Seq (Supplementary Fig. 3A). Briefly, 3 μg of high-molecular weight genomic DNA was treated with 1U FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific, USA) in 1X FastAP Buffer in final 20 μl volume at 37 °C for 1 hour, followed by 20 min at 80 °C. The treatment of genomic DNA with a phosphatase prior to Cas9 digestion facilitates removal of 5′-phosphates from random fragmented DNA ends to prevent their ligation to the adapter during the adapter ligation step below. After phosphatase treatment, DNA (20 μl) was combined with a previously assembled RNP (20 μl). For RNP assembly, Cas9 (final concentration 2 μM) and guide RNA (final concentration 4 μM) were incubated at 37 °C for 1 hour in Assembly Buffer (10 mM Tris-HCl (pH7.5 at 25 °C), 100 mM NaCl, 1 mM DTT, 1 mM EDTA) and later adding 1 M MgCl2 to the final concentration of 16 mM. After incubation of DNA with RNP for 1 hour at 37 °C, followed by 20 min at 80 °C, Adapters 1 (Upper: 5′-Biotin-AGTTACGCAACCGAGACGCGGCCGCsGsTsGsACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3 ′, where "s" stands for PTO modification; Lower: 5′ AGATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCC GGGCGTCTCGGTTGCddC-3′) were ligated to the Cas9-digested DNA by adding 40 μl of Ligation Mix (5 mM Tris-HCl (pH8.0 at 25 °C), 10 mM Mg-acetate, 20 mM DTT, 1 mM ATP, 7.5% PEG 4000 (50 w/v) 0.5 U/μl T4 DNA ligase (Thermo Scientific), 2 μM Adapters 1) final 80 μl volume and incubating at 22 °C for 1 hour in, followed by 80 °C for 20 min. After adapter ligation, nick removal was performed by adding 20 μl of Nick Removal Mix (5X Taq Buffer with (NH4)2SO4 (Thermo Scientific), 1 mM dNTPs, 20 mM EDTA, 0.5 U/μl DreamTaq DNA polymerase (Thermo Fisher Scientific) and in final 100 μl volume incubating at 72 °C for 15 min. The reaction was stopped by adding 5 μl 0.5 M EDTA. After shearing the DNA with a M220 Focused-ultrasonicator (Covaris, USA) using microTUBE AFA Fiber Pre-Slit Snap-Cap tubes (Covaris, USA) and followed shearing conditions: 50 W peak incident power, 20% duty factor, 200 cycles per burst, 200 s duration at 20 °C. Residual adapters were removed using a MagJET NGS Cleanup and Size Selection kit (Thermo Fisher Scientific), following the manufacturer's instructions. DNA was resuspended in 50 μl Elution Buffer (Thermo Fisher Scientific) and purified using 25 μl Dynabeads MyOne Streptavidin C1 beads (Thermo Fisher Scientific), following the manufacturer's instructions. Bead-bound DNA then was resuspended in 25 μl 1X FastDigest Buffer (Thermo Scientific) and cleaved with 1 μl of FastDigest NotI (Thermo Fisher Scientific) at 37 °C for 30 min. Supernatant was collected using a magnetic stand and digested with Lambda Exonuclease (Thermo Fisher Scientific) by adding 95 μl of Exonuclease Mix (67 mM glycine-KOH (pH9.4 at 25 °C), 0.01% Triton X-100, 0.2 U/μl Lambda Exonuclease) and incubating at 37 °C for 1 hour in final 120 μl volume, followed by 20 min at 80 °C. The resulting single-stranded DNA was purified using an EpiJET Bisulfite Conversion kit column (Thermo Fisher Scientific), following the manufacturer's instructions, and eluted in a total of 20 μl water. Second strand synthesis then was performed first by adding Second Strand Synthesis primer (5′-CCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNN-3′) to a final concentration of 135 nM in 1X T4 DNA Polymerase buffer (Thermo Fisher Scientific) additionally supplemented with 4 mM Mgacetate and 8.8% (w/v) PEG8000, incubating for 1 min at 98 °C in final 45.5 μl volume, then slowly cooling down to 25 °C by setting the ramp speed to 10% (0.5 °C/second) followed by 30 min at 25 °C before adding 3.5 μl dNTP mix (10 mM each) and 5U T4 DNA polymerase and incubating at 25 °C for 15 min. Double-stranded DNA was purified using the Purification Module with Magnetic Beads (Lexogen, Austria), following manufacturer's instruction. To avoid under- or overamplification of the DNA library the optimal number of cycles was determined via qPCR as described in QuantSeq/SENSE for Illumina Kit (Lexogen). DNA fragments were amplified and barcoded using the PCR Add-on Kit for Illumina and i7 Index Plate for QuantSeq/SENSE for Illumina (Lexogen, Austria) with the following cycling steps: 98 °C for 2 min, followed by 98 °C for 10 s; 65 °C for 20 s; 72 °C for 30 s (number of cycles determined via qPCR), and hold at 4 °C. The amplified DNA was purified and size-selected using the Purification Module with Magnetic Beads (Lexogen), and quantified using a Qubit fluorometer prior to sequencing on an Illumina HiSeq 2500 sequencer.

CLEAVE-Seq analysis.   Analyses to identify cleaved genomic targets were performed as described in Supplementary Fig. 3B.

Physical mapping of CLEAVE-Seq data.   Paired end read FASTQ files were pre-processed with Skewer v0.2.2[73] to remove Illumina universal adapters and enforce a minimum surviving read length of 35 bp. Then, bowtie 2-build v2.3.2[74] was used to index the reference sequences with default parameters and Bowtie 2 was used to perform non-discordant (for paired input), end to end alignments of the reads to the genomic reference.

Generation of normalized read coverage profiles.   Normalized read coverage profiles were built for each sample starting with the negative control (a CLEAVE-Seq reaction assembled in the absence of a guide RNA) followed by the treated experimental samples. Coverage counts were calculated for reads initiating within a $+/- 2$ bp window surrounding the expected site of cleavage (3 bp 5′ of the protospacer adjacent motif (PAM)) of each computationally predicted target site using BEDTools[75]. The total number of reads originating from the cut-site was then adjusted by dividing it by a normalizer (established by dividing the sample's read depth by the lowest read depth present in the dataset). Normalized read coverage profiles between replicates were averaged.

Filtering for biochemically cleaved sites. To identify biochemically cleaved target sites, normalized read coverage was compared with the same location in the negative control and used to ascertain a targets validity. Accounting for potential biases and read duplication generated during PCR construction of CLEAVE-Seq Illumina compatible libraries (Supplementary Fig. 2A step M) and those introduced by Illumina sequencing, genuine biochemically cleaved sites were defined as having a normalized read coverage of at least 5 with at least 5-fold excess coverage over the control.

CLEAVE-Seq false negative and positive rates. The rate of false discovery and missed identification of genuinely cleaved target sites were calculated by spiking artificial read coverage into a control dataset. In total, 500 computationally predicted targets were randomly selected and artificial paired end reads (150 bp in length) initiating within the expected window of cleavage were generated using the art_illumina tool[76]. A read coverage of 25 was spiked into the 500 randomly selected targets. Simulated reads were then spiked into the control sample's FASTQ file and mapped back to the reference genome. After physical mapping, CLEAVE-Seq analysis was performed as described above using an independent control as a comparator and false positive and negative rates were calculated for M1, M2, and M3. Importantly, for all three targets, the false negative rate was 0% illustrating that all targets spiked into the control dataset were recovered. The false positive rates for M1, M2, and M3 were 1.4% (7 additional targets identified (+7 targets)), 19% (+95 targets), and 8.4% (+42 targets), respectively.

Plant material. Publicly available maize (*Zea mays* L.) hybrid high type II (Hi-II) line was obtained from internal Corteva sources.

Plasmids and reagents used for plant transformation. Cas9 and guide RNA vector construction was previously described in Svitashev *et al.*[77]. Plasmids containing cell division promoting transcription factors (maize ovule developmental protein 2 (ODP2) and maize Wuschel (WUS)), selectable and visible marker MOPAT-DSRED (a translational fusion of the bialaphos resistance gene, phosphinothricin-N-acetyl-transferase, and the red fluorescent protein DSRED) were previously described in Ananiev, E.V. *et al.*[78]. Ribonucleotprotein (RNP) complex formation was performed as described in Svitashev S. *et al.*[24]. Briefly, Cas9 protein and guide RNA molecules were mixed (in a 1:2 molar ratio, respectively) in 1x NEB Buffer 3 with 1 μl of RNA inhibitor (Ribo Guard™, Epicentre, USA) in a total volume of 20 μl and incubated at room temperature for at least 15 min.

Maize transformation. Biolistic-mediated delivery of plasmid vectors containing Ubiquitin promoter-regulated Cas9, maize U6 polymerase III promoter-regulated gRNA, Ubiquitin promoter-regulated ODP2, maize IN2 promoter-regulated WUS, and Ubiquitin promoter-regulated selectable and visible marker, *MOPAT-DSRED* fusion, to maize immature embryos was performed as previously described in Svitashev *et al.*[77]. The particle delivery of matrix comprising the RNPs complemented with plasmids containing the Ubiquitin promoter-regulated ODP2, maize IN2 promoter-regulated WUS, and Ubiquitin promoter-regulated selectable and visible marker, *MOPAT-DSRED* fusion, were delivered into maize embryo cells as described in Svitashev, S. *et al.*[24]. Post-bombardment culture, selection, and plant regeneration were performed as previously described in Gordon-Kamm, W. *et al.*[79]. *A. tumefaciens* transformation vectors containing Ubiquitin promoter-regulated Cas9, maize U6 polymerase III promoter-regulated gRNA, and visible marker, *MOPAT-DSRED* fusion, were introduced into maize followed the detailed protocol previously described in Gao, H.R. *et al.*[80]. Regenerated plantlets were moved to soil, where they were sampled (7 mm leaf punch per plant) and grown to maturity in greenhouse conditions. In total 390 plants were analyzed using MIPs and each treatment contained 26–117 plants.

Target site monitoring in plants with Molecular Inversion Probes. Molecular inversion probes (MIP)[81,82] assays were designed by analyzing a 100 nt window surrounding the target sites of interest identified via *in silico* and biochemical methods. Targeting arms flanking the region of interest were selected based on the following assay criteria: arm length of 17–28 nt, distance between 5′ and 3′ targeting arms of 1–70 nt and predicted melting temperature of 68–72 °C. Following design, targeting arms for each assay were linked by a common backbone sequence 30–50 nt in length and ordered as individual oligos with a 5′ phosphorylation. The individual 250 μM MIPs oligos were pooled in equal volumes to generate a 250 uM assay pool.

MIPs targeting and sequencing pool creation was accomplished via a four-step process: hybridization, circularization, exonuclease digestion and indexing/amplification. Briefly, hybridization reactions were prepared by combining 250 ng of DNA with 1.25 μl ampligase buffer (Epicentre), 0.5 μl 1 M blocking oligo, a volume of MIPs assay pool that resulted in a DNA:MIPs ratio of 500:1 to 5000:1 depending on panel size, and water to a final reaction volume of 12.5 μl. Reactions were denatured for 10 min at 95 °C followed by 30 min incubation at 60 °C in a thermocycler with heated lid. Following incubation hybridized MIPs were recircularized by addition of 0.2 μl of 10x Ampligase buffer, 1 ul 2 U/μl HF Phusion polymerase (New England Biolabs), 0.25 ul 100 U/μl Ampligase enzyme (Epicentre) and 0.55 μl 0.25 mM dNTP mix (New England Biolabs) to the completed hybridization reaction, while the reaction was maintained at 60 °C. The final circularization reaction was mixed gently, sealed and incubated at 60 °C for 16–18 hours. Following circularization, incubation reactions were collected by centrifugation, incubated for 1 min at 37 °C and stored at 4 °C until exonuclease digestion.

Exonuclease digestion to remove linear genomic DNA and un-circularized probes was performed by adding 1 μl of 20 U/ul Exo I and 1 μl of 100 U/μl Exo III (New England Biolabs, USA) to the circularized MIP reaction from the previous step. Reactions were incubated in a thermocycler for 15 min at 37 °C followed by 2 min inactivation at 95 °C. Following digestion, targeted sequences were indexed and amplified by adding 12.5 μl of 2X iProof Master mix (Biorad), 0.125 μl 100 μM universal backbone forward primer, 0.125 μl 100 μM indexed backbone reverse primer, and 9.8 μl water. Reactions were denatured at 98 °C for 2 min and amplified by 25 cycles of

98 °C for 10 seconds, 60 °C for 30 seconds, 72 °C for 60 seconds. Resulting indexed amplicons were pooled and purified by a 1:1 Ampure XP cleanup according to manufacturer's recommendations (Beckman Coulter Inc., USA). Purified amplicon pools were sequenced via manufactures recommendations with custom primers to the backbone sequence on Illumina MiSeq sequencers, generating 100 nt paired end reads.

Analysis of targeted loci sequence.   Sequencing reads were deconvoluted into sample bins by index sequence. Per sample reads were analyzed by identifying reads that belong to a specific MIPs assay via the 5′ and 3′ targeting arm. Reads were aligned via Bowtie v2 to the wildtype reference used in design of the assay arms. Differences between the reference sequences were identified by mismatches in alignment and reported via SAM Tools.

Data Availability

The data supporting the findings of this study are available within the paper and its Supplementary Information Files. Sequencing data have been deposited in the National Center for Biotechnology Information Sequence Read Archive database under accession code PRJNA526862.

References

1. Glenn, K. C. *et al.* Bringing New Plant Varieties to Market: Plant Breeding and Selection Practices Advance Beneficial Characteristics while Minimizing Unintended Changes. *Crop Sci* **57**, 2906–2921 (2017).
2. Parry, M. A. & Hawkesford, M. J. An integrated approach to crop genetic improvement. *J Integr Plant Biol* **54**, 250–259 (2012).
3. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA* **93**, 1156–1160 (1996).
4. Epinat, J. C. *et al.* A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res* **31**, 2952–2962 (2003).
5. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757–761 (2010).
6. Jinek, M. *et al.* A programmable dual-RNA–guided DNA Endonuclease in adaptive bacterial immunity. *Science.* **337** (2012).
7. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* **109**, E2579–2586 (2012).
8. Yin, K., Gao, C. & Qiu, J. L. Progress and prospects in plant genome editing. *Nat Plants* **3**, 17107 (2017).
9. Zhang, Z. *et al.* Simultaneous Editing of Two Copies of Gh14-3-3d Confers Enhanced Transgene-Clean Plant Defense Against Verticillium dahliae in Allotetraploid Upland Cotton. *Front Plant Sci* **9**, 842 (2018).
10. Wang, Y. *et al.* Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol* **32**, 947–951 (2014).
11. Lu, K. *et al.* Blocking amino acid transporter OsAAP3 improves grain yield by promoting outgrowth buds and increasing tiller number in rice. *Plant Biotechnol J* **16**, 1710–1722 (2018).
12. Borrelli, V. M. G., Brambilla, V., Rogowsky, P., Marocco, A. & Lanubile, A. The Enhancement of Plant Disease Resistance Using CRISPR/Cas9 Technology. **9** (2018).
13. Lemmon, Z. H. *et al.* Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat Plants* **4**, 766–770 (2018).
14. C. L. I. Site Directed Nucleases (SDN) for targeted genome modification. https://croplife-r9qnrxt3qxgjra4.netdna-ssl.com/wp-content/uploads/2015/01/CLI-SDN-Definitions-Position-Paper.pdf (2018).
15. Podevin, N., Davies, H. V., Hartung, F., Nogue, F. & Casacuberta, J. M. Site-directed nucleases: a paradigm shift in predictable, knowledge-based plant breeding. *Trends Biotechnol* **31**, 375–383 (2013).
16. Zheng, T. *et al.* Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. *Sci Rep* **7**, 40638 (2017).
17. Zhang, Q. *et al.* Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in Arabidopsis and its prevention. *Plant Mol Biol* (2018).
18. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
19. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
20. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
21. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **24**, 1012–1019 (2014).
22. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014).
23. Liang, X. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* **208**, 44–53 (2015).
24. Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K. & Mark Cigan, A. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nat Commun* **7**, 13274 (2016).
25. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187–197 (2015).
26. Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods* **12**, 237–243, 231 p following 243 (2015).
27. Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive *in vitro* screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods* **14**, 607–614 (2017).
28. Cameron, P. *et al.* Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods* **14**, 600–606 (2017).
29. Yan, W. X. *et al.* BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat Commun* **8**, 15058 (2017).
30. Xiao, A. *et al.* CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* **30**, 1180–1182 (2014).
31. Park, J., Bae, S. & Kim, J. S. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics* **31**, 4014–4016 (2015).
32. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
33. Alkan, F., Wenzel, A., Anthon, C., Havgaard, J. H. & Gorodkin, J. CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters. *Genome Biol* **19**, 177 (2018).
34. Abadi, S., Yan, W. X., Amar, D. & Mayrose, I. A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLoS Comput Biol* **13**, e1005807 (2017).
35. Martin, F., Sanchez-Hernandez, S., Gutierrez-Guerrero, A., Pinedo-Gomez, J. & Benabdellah, K. Biased and Unbiased Methods for the Detection of Off-Target Cleavage by CRISPR/Cas9: An Overview. *Int J Mol Sci* **17** (2016).
36. Zhang, J. *et al.* A molecular inversion probe-based next-generation sequencing panel to detect germline mutations in Chinese early-onset colorectal cancer patients. *Oncotarget* **8**, 24533–24547 (2017).

37. Akcakaya, P. *et al. In vivo* CRISPR editing with no detectable genome-wide off-target mutations. *Nature* **561**, 416–419 (2018).

38. Chaubal, R. *et al.* Two male-sterile mutants of Zea Mays (Poaceae) with an extra cell division in the anther wall. *American Journal of Botany* **87**, 1193–1201 (2000).

39. Djukanovic, V. *et al.* Male-sterile maize plants produced by targeted mutagenesis of the cytochrome P450-like gene (MS26) using a re-designed I-CreI homing endonuclease. *Plant J* **76**, 888–899 (2013).

40. Dobritsa, A. A. *et al.* LAP3, a novel plant protein required for pollen development, is essential for proper exine formation. *Sex Plant Reprod* **22**, 167–177 (2009).

41. Cigan, A. M. *et al.* Targeted mutagenesis of a conserved anther-expressed P450 gene confers male sterility in monocots. *Plant Biotechnol J* **15**, 379–389 (2017).

42. Becraft, P. W., Bongard-Pierce, D. K., Sylvester, A. W., Poethig, R. S. & Freeling, M. The liguleless-1 gene acts tissue specifically in maize leaf development. *Dev Biol* **141**, 220–232 (1990).

43. Jiao, Y. *et al.* Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).

44. Jiang, F. & Doudna, J. A. CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys* **46**, 505–529 (2017).

45. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci USA* **108**, 10098–10103 (2011).

46. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822–826 (2013).

47. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–832 (2013).

48. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839–843 (2013).

49. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32**, 670–676 (2014).

50. Tang, X. *et al.* A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biol* **19**, 84 (2018).

51. Wang, P. *et al.* High efficient multisites genome editing in allotetraploid cotton (Gossypium hirsutum) using CRISPR/Cas9 system. *Plant Biotechnol J* **16**, 137–150 (2018).

52. Feng, Z. *et al.* Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc Natl Acad Sci USA* **111**, 4632–4637 (2014).

53. Wolt, J. D., Wang, K., Sashital, D. & Lawrence-Dill, C. J. Achieving Plant CRISPR Targeting that Limits Off-Target Effects. *Plant Genome* **9** (2016).

54. Clarke, R. *et al.* Enhanced Bacterial Immunity and Mammalian Genome Editing via RNA-Polymerase-Mediated Dislodging of Cas9 from Double-Strand DNA Breaks. *Mol Cell* **71**, 42–55 e48 (2018).

55. Lee, K. *et al.* Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnol J* (2018).

56. Li, X. *et al.* Lycopene Is Enriched in Tomato Fruit by CRISPR/Cas9-Mediated Multiplex Genome Editing. *Front Plant Sci* **9**, 559 (2018).

57. Feng, C. *et al.* High-efficiency genome editing using a dmc1 promoter-controlled CRISPR/Cas9 system in maize. *Plant Biotechnol J* (2018).

58. Zeng, Y. *et al.* The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic Acids Res* **46**, 350–361 (2018).

59. Xu, X., Duan, D. & Chen, S. J. CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: from physical mechanism to off-target assessment. *Sci Rep* **7**, 143 (2017).

60. LeBlanc, C. *et al.* Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. *Plant J* **93**, 377–386 (2018).

61. Sun, S. *et al.* Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat Genet* **50**, 1289–1295 (2018).

62. Armstrong, C. L., Green, C. E. & Phillips, R. L. Development and availability of germplasm with high Type II culture formation response. *Maize Genetics Cooperaion Newsletter* **65**, 92–93 (1991).

63. Yang, N. *et al.* Contributions of Zea mays subspecies mexicana haplotypes to modern maize. *Nat Commun* **8**, 1874 (2017).

64. Anderson, J. E. *et al.* Genomic variation and DNA repair associated with soybean transgenesis: a comparison to cultivars and mutagenized plants. *Bmc Biotechnol* **16**, 41 (2016).

65. Phillips, R. L., Kaeppler, S. M. & Olhoft, P. Genetic instability of plant tissue cultures: breakdown of normal controls. *Proc Natl Acad Sci USA* **91**, 5222–5226 (1994).

66. Wei, F. J. *et al.* Somaclonal variation does not preclude the use of rice transformants for genetic screening. *Plant J* **85**, 648–659 (2016).

67. Hase, Y., Satoh, K., Kitamura, S. & Oono, Y. Physiological status of plant tissue affects the frequency and types of mutations induced by carbon-ion irradiation in Arabidopsis. *Sci Rep* **8**, 1394 (2018).

68. Batista, R., Saibo, N., Lourenco, T. & Oliveira, M. M. Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. *Proc Natl Acad Sci USA* **105**, 3640–3645 (2008).

69. Bednarek, P. T., Orlowska, R., Koebner, R. M. & Zimny, J. Quantification of the tissue-culture induced variation in barley (Hordeum vulgare L.). *Bmc Plant Biol* **7**, 10 (2007).

70. Szurman-Zubrzycka, M. E. *et al.* HorTILLUS-A Rich and Renewable Source of Induced Mutations for Forward/Reverse Genetics and Pre-breeding Programs in Barley (Hordeum vulgare L.). *Front Plant Sci* **9**, 216 (2018).

71. Ossowski, S. *et al.* The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. *Science* **327**, 92–94 (2010).

72. Karvelis, T. *et al.* Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol* **16**, 253 (2015).

73. Jiang, H., Lei, R., Ding, S. W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *Bmc Bioinformatics* **15**, 182 (2014).

74. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

75. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

76. Huang, W., Li, L., Myers, J. R. & Marth, G. T. ART: a next-generation sequencing read simulator. *Bioinformatics* **28**, 593–594 (2012).

77. Svitashev, S. *et al.* Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA. *Plant Physiol* **169**, 931–945 (2015).

78. Ananiev, E. V. *et al.* Artificial chromosome formation in maize (Zea mays L.). *Chromosoma* **118**, 157–177 (2009).

79. Gordon-Kamm, W. *et al.* Stimulation of the cell cycle and maize transformation by disruption of the plant retinoblastoma pathway. *P Natl Acad Sci USA* **99**, 11975–11980 (2002).

80. Gao, H. R. *et al.* Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant J* **61**, 176–187 (2010).

81. Hiatt, J. B., Pritchard, C. C., Salipante, S. J., O'Roak, B. J. & Shendure, J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res* **23**, 843–854 (2013).

82. Turner, E. H., Lee, C., Ng, S. B., Nickerson, D. A. & Shendure, J. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods* **6**, 315–316 (2009).

Acknowledgements

Author Contributions

G.M., N.D.C., S.K., J.Y., S.D., G.Z.-H. and S.S. designed the research. J.Y., V.D., S.D., M.Z. and V.S. designed and performed CLEAVE-Seq experiments. G.Z.-H. designed and executed MIPS analysis. S.S., C.S., B.L., L.F. and L.W. were involved in vector construction, plant transformation and DNA analysis of transgenic plants. S.P., B.P.-B. and A.A. performed bioinformatics and data analysis. S.K., J.Y., S.D., N.D.C., G.Z.-H., S.D., S.S., A.A., C.A. and G.M. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-43141-6.

Competing Interests: Some of the authors are employees of Corteva Agriscience™, Agriculture Division of DowDuPont™.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.