scientific reports

OPEN

Knock-in and precise nucleotide substitution using near-PAMless engineered Cas9 variants in *Dictyostelium discoideum*

Yuu Asano[1,2], Kensuke Yamashita[1,2], Aoi Hasegawa[1], Takanori Ogasawara[1], Hoshie Iriki[1] & Tetsuya Muramoto [1][✉]

The powerful genome editing tool *Streptococcus pyogenes* Cas9 (SpCas9) requires the trinucleotide NGG as a protospacer adjacent motif (PAM). The PAM requirement is limitation for precise genome editing such as single amino-acid substitutions and knock-ins at specific genomic loci since it occurs in narrow editing window. Recently, SpCas9 variants (i.e., xCas9 3.7, SpCas9-NG, and SpRY) were developed that recognise the NG dinucleotide or almost any other PAM sequences in human cell lines. In this study, we evaluated these variants in *Dictyostelium discoideum*. In the context of targeted mutagenesis at an NG PAM site, we found that SpCas9-NG and SpRY were more efficient than xCas9 3.7. In the context of NA, NT, NG, and NC PAM sites, the editing efficiency of SpRY was approximately 60% at NR (R = A and G) but less than 22% at NY (Y = T and C). We successfully used SpRY to generate knock-ins at specific gene loci using donor DNA flanked by 60 bp homology arms. In addition, we achieved point mutations with efficiencies as high as 97.7%. This work provides tools that will significantly expand the gene loci that can be targeted for knock-out, knock-in, and precise point mutation in *D. discoideum*.

Under optimal nutrient conditions, the social amoeba *D. discoideum* grows as a single-cell organism; however, in response to starvation, about 100,000 cells aggregate to form a multicellular fruiting body that consists of stalk and spore cells. It possesses a rather small haploid genome (~ 34 Mb) that contains a wide range of homologous genes involved in cell motility, signal transduction, chemotaxis, phagocytosis, and multicellular formation[1,2]. The high efficiency of gene manipulations in this organism, including knock-out and knock-in by homologous recombination, RNA interference (RNAi), and overexpression, has enabled us to study the functions of these genes in detail[3–6].

Recently, genome editing using the clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein 9 (Cas9) has been developed as a toolbox for functional analysis in a variety of organisms[7–9]. In *D. discoideum*, all-in-one vectors capable of expressing Cas9 endonuclease and a chimeric single-guide RNA (sgRNA) that recognises target sequences have allowed highly efficient gene disruption[10,11]. Although one issue with CRISPR is off-target effects, these vectors are not stably maintained in cells, minimising the frequency of unwanted editing. Despite the transient expression of the CRISPR components, more than half of the cells were edited[11], indicating that this system is both highly efficient and relatively specific. Furthermore, using the double-nicking method with Cas9 nickase, the probability of off-target effects can be further reduced, and target-specific genome deletions larger than 1 kb have been successfully introduced in *D. discoideum*[12]. Recruitment of Cas9 to target DNA is programmed by sgRNA, but Cas9 also requires a PAM flanking the target site[13–15]. NGG is the canonical PAM sequence for SpCas9, the endonuclease most widely used in CRISPR. The PAM requirement significantly decreases the number of targetable sites, especially in the AT-rich *D. discoideum* genome. Homology-directed repair (HDR) mediated gene knock-in and precise nucleotide substitution, which are genome editing applications that require high-resolution targeting, are even more strongly affected by the PAM requirement because these types of editing must generally occur in a narrow window around the target sequence.

To overcome these limitations on targeting range, several Cas9 variants with distinct PAM sequences have been developed, including SpCas9-VQR, SpCas9-VRER, and SpCas9-EQR[16]. In addition, Cas9 orthologs derived from other species with different PAM sequences, such as *Staphylococcus aureus* Cas9 (SaCas9), SaCas9 variant

---

[1]Department of Biology, Faculty of Science, Toho University, 2-2-1 Miyama, Funabashi, Chiba 274-8510, Japan. [2]These authors contributed equally: Yuu Asano and Kensuke Yamashita. [✉]email: tetsuya.muramoto@sci.toho-u.ac.jp
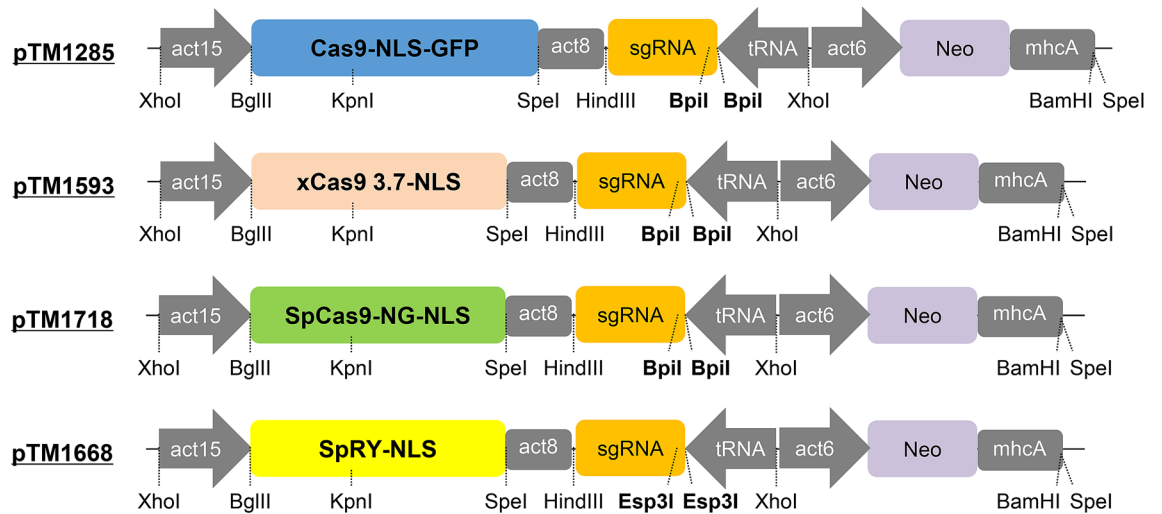
Figure 1. Diagrams of all-in-one Cas9 and sgRNA expression vectors used for gene manipulation in *D. discoideum*. act15, act15 promoter; act8, act8 terminator; tRNA, isoleucine tRNA; act6, act6 promoter; neo, neomycin resistance gene.

(SaCas9-KKH), *Streptococcus canis* (ScCas9), *Streptococcus thermophilus 1* (St1Cas9), *Streptococcus thermophilus 3* (St3Cas9), and *Campylobacter jejuni* Cas9 (CjCas9), have also been characterised[17–22]. Furthermore, the newly engineered SpCas9 variants, xCas9 3.7 and SpCas9-NG can recognise the non-canonical NGN PAM in human cells[23,24], and have been successfully tested in rice, *Arabidopsis*, and mice[25–28]. Comparison of xCas9 3.7 and SpCas9-NG in mammalian cells and plants revealed that SpCas9-NG has a higher editing efficiency than xCas9 3.7 at NG PAMs[24,29,30]. More recently, a structure-based engineering approach was used to generate SpRY, which can edit almost all PAMs (NRN > NYN) in mammalian cells[29]. However, it remains unclear whether SpRY functions effectively in other model organisms, including *D. discoideum*. A demonstration that SpCas9 variants such as xCas9 3.7, SpCas9-NG, and SpRY could be used in *D. discoideum* would expand the potential for gene knockouts, gene knock-ins, and precise base substitutions in regions that do not contain canonical NGG PAMs.

In this study, we generated transient all-in-one vectors to evaluate the editing efficiencies of SpCas9, xCas9 3.7, SpCas9-NG, and SpRY at various PAM sites in *D. discoideum*. All of the Cas9 variants could edit target regions harbouring NG PAM sequences. In addition, we showed that SpRY had robust nuclease activity at almost all PAMs. Moreover, SpRY improved editing resolution at precise positions for gene knock-in and nucleotide substitution.
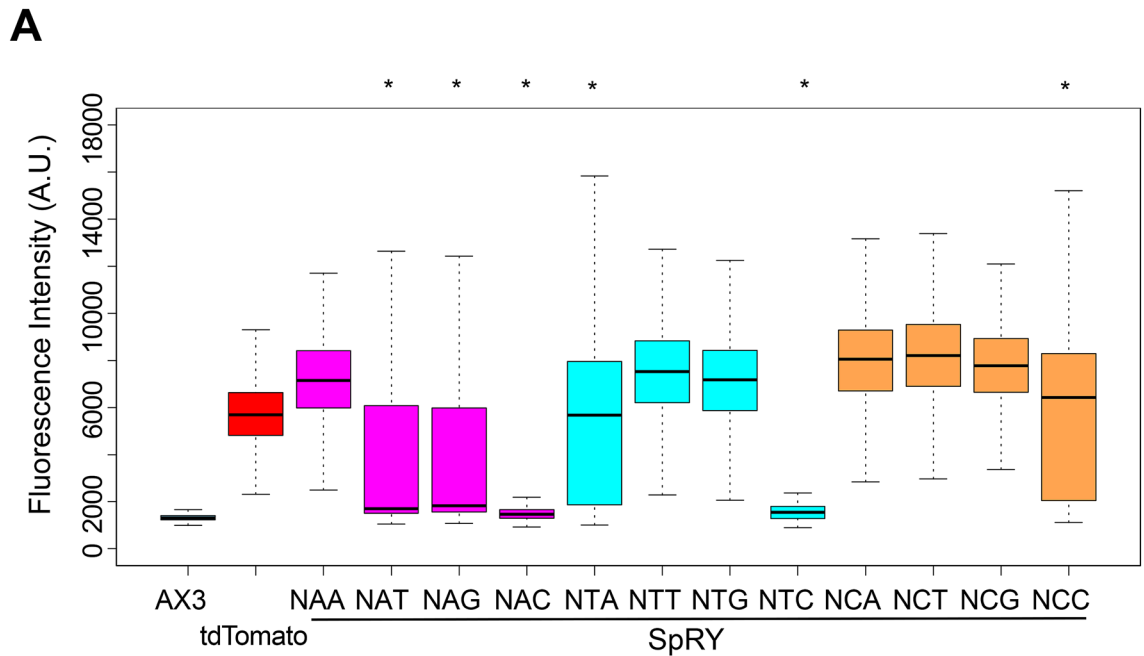
Results

Cas9 variant-mediated genome editing in tdTomato knock-in cells. We constructed codon-optimised versions of Cas9 variants xCas9 3.7, SpCas9-NG, and SpRY, and then evaluated their editing efficiencies at various PAM sequences in *D. discoideum* (Fig. 1). In these experiments, we targeted the tdTomato gene, which was knocked in at the *act5* locus (Fig. 2A). Mutations led to a loss of fluorescence; thus, mutation efficiency could be measured by monitoring the percentage of non-fluorescent cells. Although tdTomato is not an endogenous gene in *D. discoideum*, the efficiencies we observed were almost identical to those of endogenous genes[11]. Four sgRNA sequences recognising different targets harbouring NGG, NGA, NGC or NGT PAMs were designed against the tdTomato gene (Table S1), and then inserted into all-in-one vectors containing various Cas9 variants (Tables S2 and S3). Classical SpCas9 edited DNA efficiently using a canonical NGG PAM, but only a few percent of cells lost red fluorescence when using non-canonical NGA, NGT, and NGC PAMs (Figs. S1 and S2). The initially generated all-in-one vector with classical SpCas9, pTM1285, contained a GFP sequence followed by Cas9. We also found that vectors without GFP, pTM1599 and pTM1644, could induce genome editing as efficiently as pTM1285 (Figs. S2 and S3).

In cells expressing a Cas9 variant and an sgRNA targeting the canonical NGG PAM, we detected loss of red fluorescence, as in cells expressing classical SpCas9. The mutation efficiencies of SpCas9-NG and SpRY were 97.8% and 81.3%, respectively, whereas that of xCas9 3.7 was only 14.5% (Fig. 2B,C). Loss of fluorescence due to xCas9 3.7-mediated editing at non-canonical PAMs was observed in less than 1% of cells, suggesting that xCas9 3.7 is not suitable for genome editing at NG PAM sequences in *D. discoideum*. On the other hand, SpCas9-NG and SpRY had efficiencies greater than 25%, except for SpCas9-NG at NGC, but none had higher efficiencies at these sequences than at the canonical NGG PAM. SpRY was more than 25% efficient with all NG PAMs, whereas SpCas9-NG exhibited higher or lower efficiencies at different PAMs.

Because SpRY is a near-PAMless Cas9 variant (NRN > NYN)[29], we next assessed SpRY editing activity while targeting non-canonical NHN PAM sites (where H is A, C, or T). For this purpose, we selected 12 targets harbouring NHN PAMs, all of which could target the tdTomato gene (Table S1). Half of the targets (6/12) exhibited highly efficient editing (25.7–87.6%) with SpRY, and two of them (NAC and NTC PAMs) were as or more efficient than the canonical NGG PAM (Fig. 3). Editing efficiency at NR (where R is A or G) was greater than
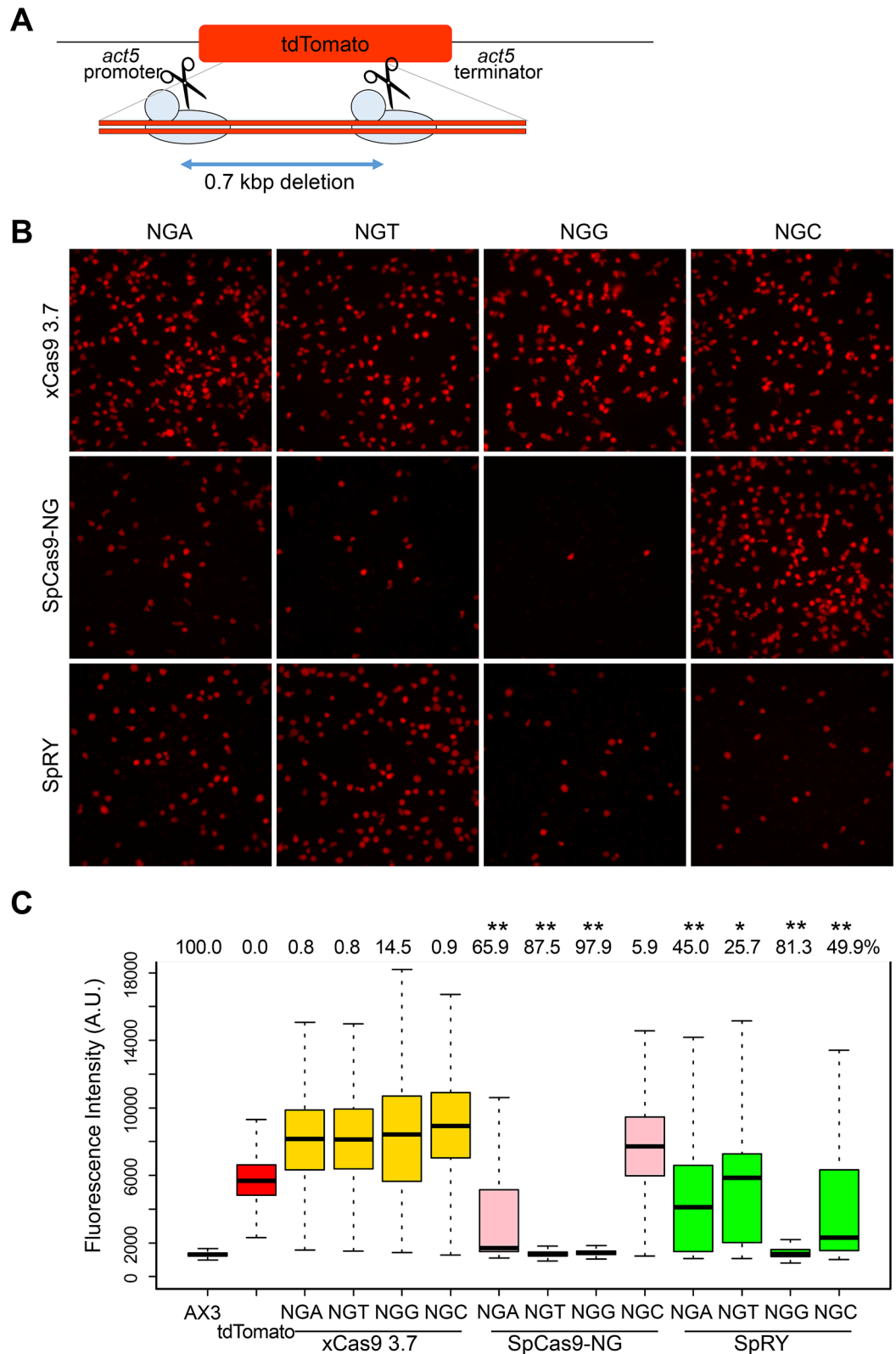
Figure 2. Gene mutagenesis by xCas9 3.7, SpCas9-NG, and SpRY in *D. discoideum*. (**A**) Schematic diagram of gene targeting in tdTomato knock-in cells. Because tdTomato contains tandem repeats, a 700 bp deletion can be generated with a single sgRNA. (**B**) Fluorescence observation of tdTomato in mutated cells. Four sgRNAs harbouring NGA, NGT, NGG or NGC PAMs were introduced into tdTomato-expressing cells along with the indicated Cas9 variants. Red fluorescence images of representative areas are shown. (**C**) Box plot represents loss of red fluorescence induced by Cas9 variants. AX3 indicates the auto-fluorescence level without tdTomato, and tdTomato indicates the expression level of the parental strain. Numbers above the graph show knockout efficiencies. *$P = 0.064$ and **$P < 0.001$; ANOVA followed by Tukey's post hoc test. n = 3 biological replicates.

Figure 3. Targeting of non-canonical NHN PAMs in *D. discoideum* using SpRY. (**A**) Box plot represents fluorescence intensity with different sgRNAs in targets harbouring NHN PAMs. The target sites covered all twelve NHN PAM combinations. *$P < 0.001$; ANOVA followed by Tukey's post hoc test. (**B**) Proportion of cells in which SpRY induced loss of fluorescence.

50% on average, whereas efficiencies at NT and NC were 33.8% and 9.2%, respectively, supporting the previous observation that SpRY is capable of efficiently targeting the majority of target sites with NRN PAMs[29]. Given that we observed highly efficient editing at almost all PAMs, including NGN PAMs, SpRY is likely to allow editing of the majority of genomic DNA in *D. discoideum*, with no restriction due to a requirement for NGG sequences.

SpRY-mediated knock-in of tag-sequences. In cases of homologous recombination-mediated knock-in in *D. discoideum*, it is necessary to introduce a drug resistance cassette into the genome to allow isolation of transformants. It is challenging to insert a tag sequence inside a gene or immediately after the start codon without perturbation by the drug resistance cassette. It is possible to excise a loxP-flanked drug resistance cassette with a Cre protein, leaving the loxP sequence behind in the recombination region[31]. Overexpression of recombinant proteins with N-terminal tags or nucleotide substitutions is often used for functional analysis, but overproduction occasionally results in ectopic expression or dominant-negative effects. Because the nearly PAM-less Cas9 variant SpRY is functional in *D. discoideum*, and it is not necessary to introduce a drug resistance cassette into the genome, the tag sequence can be knocked in at any location, even in the middle of a gene. We selected the cAR1 gene as a target to investigate whether we could knock-in the GFP sequence immediately after the start codon. Design of target sequences within 30 bp of the start codon yielded 16 candidate sgRNAs with low off-target potential and high specificity for the desired region. Due to the low GC content within this region, these candidates did not include canonical NGG PAM targets. From the candidates, we selected target sequences

A
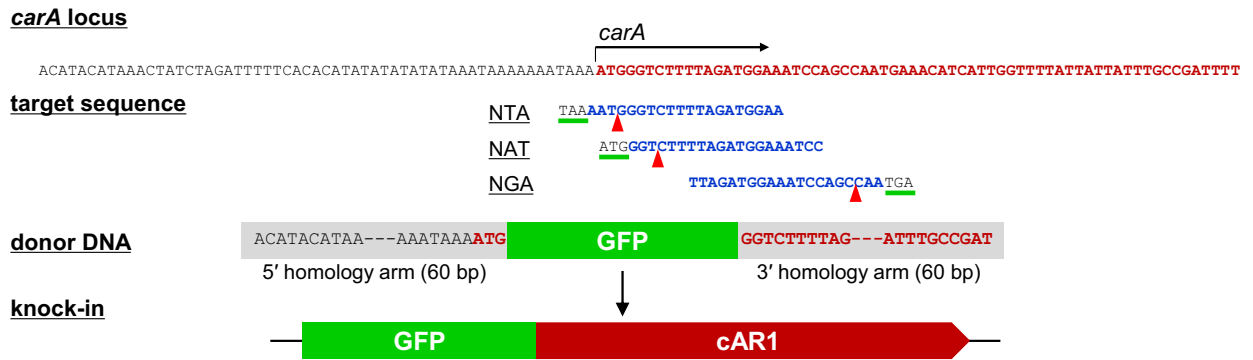
B
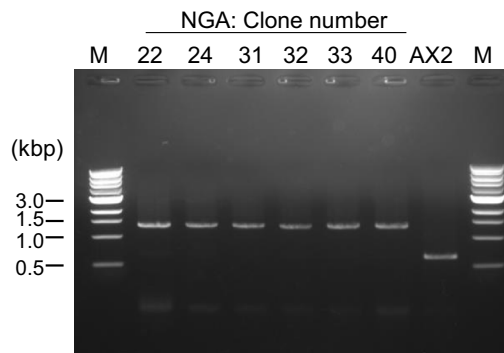
C

Cas9	PAM	Knock-in Efficiency (%)	No. analysed
SpCas9	NTA	0.0	39
	NAT	0.0	32
	NGA	0.0	37
SpCas9-NG	NTA	0.0	34
	NAT	0.0	32
	NGA	3.9	51
SpRY	NTA	0.0	121
	NAT	52.7	93
	NGA	41.0	39

D

Figure 4. SpRY-mediated tag knock-in. (**A**) Schematic illustration of GFP knock-in at the cAR1 gene. Target sites in *carA* are in blue, and PAM sequences are underlined in green. Red arrowheads indicate sites of predicted cleavage by SpRY. Donor DNA was amplified by PCR using GFP primers flanked with homology arms (Table S4). (**B**) PCR amplification using primers flanking the knock-in site. AX2 presents the parental strain, and the rest of the lanes represent six individual mutants. An unprocessed image of gel shows that no other bands were present on the gel. (**C**) Summary of knock-in efficiencies with the indicated targets (NTA, NAT or NGA PAMs). (**D**) Sequencing results from the knock-in region. Sequences of GFP-cAR1 (top) and individual PCR positive mutants are shown. GFP sequences are in green, and target sequences are in blue. PAM sequences are underlined in green.

with NTA, NAT, and NGA PAMs, and then generated all-in-one vectors containing SpCas9, SpCas9-NG, and SpRY, respectively (Table S3). To generate donor DNA containing the GFP sequence and the two homologous fragments (5′ and 3′ arms), we amplified the fragment by PCR using GFP primers flanked with 60 bp homology arms (Fig. 4A). Next, we simultaneously introduced the all-in-one vector and donor DNA into the AX2 strain to induce transient expression of the Cas9 and sgRNAs. We then isolated single clones and extracted the genomic DNA. Because the knocked-in clones yielded bands ~ 700 bp larger than the unedited clones, we could calculate the knock-in frequency by PCR (Fig. 4B). About 50% of the clones expressing SpRY with NAT and NGA PAMs contained knock-ins vs only a few percent of clones expressing SpCas9 and SpCas9-NG with NTA, NAT, and NGA PAMs (Fig. 4C). Sequence analysis confirmed that the correct knock-ins were present in more than 60% of

the clones with NGA and NAT PAMs (Fig. 4D). Based on these results, we succeeded in generating a knock-in strain with only a tag sequence at an arbitrary part of the gene without introducing any extra sequences, such as drug resistance cassettes.

Homologous recombination for gene knock-out or knock-in in *D. discoideum* requires homology arms of 500 bp or longer[32]. In this study, we showed that CRISPR/Cas9 could efficiently generate a knock-in strain using homology arms of just 60 bp. However, the relationship between the length of the homology arms and knock-in efficiency remained unknown. Therefore, we generated donor DNA with 30 bp and 90 bp homology arms and compared the knock-in efficiencies. The 90 bp homology arms were roughly half as efficient as the 60 bp arms, whereas the knock-in efficiency of the 30 bp arms was significantly lower, indicating that 60 bp arms were sufficient (Fig. S4).
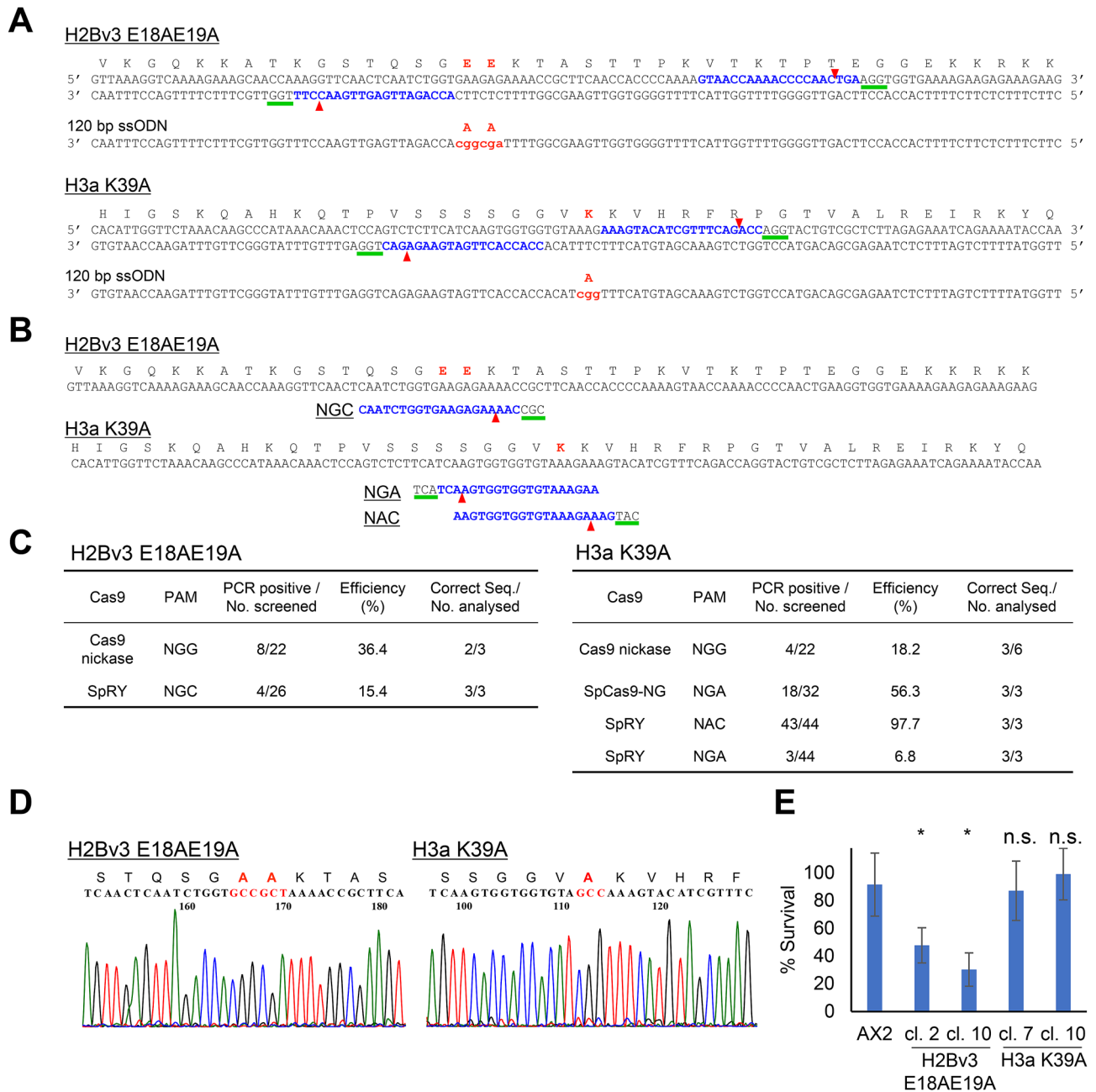
Overexpressed cAR1-GFP (a C-terminal GFP fusion of cAR1) is localised to the cell surface[33]. In GFP-cAR1 knock-in cells, localisation to the plasma membrane was very weak, even in cells in the aggregation stage (Fig. S5). This was because GFP was knocked in at the N-terminus, which is required for efficient translocation to the surface membrane. Formation of aggregation streams in submerged condition or on agar were not observed, and subsequent fruiting body formation was significantly delayed. However, unlike the cAR1 knockouts[34], the mutants formed fruiting bodies within 2 days after starvation (Fig. S5).

Accurate single-base substitution mediated by SpRY. For relatively short genes such as histones, accurate single-base substitution within the genome can be achieved by homologous recombination[35,36]. However, this approach is often challenging because it has low efficiency, generating only one positive out of hundreds of clones. To determine whether it would be possible to generate highly efficient single-base substitutions using the CRISPR/Cas9 system, we used Cas9 variants and Cas9 nickase. As targets, we selected H2Bv3 E18/E19 and H3a K39 (equivalent to K36 in mammals). Histone H2Bv3 E18/E19 is ADP-ribosylated in response to DNA double-strand breaks, and H3a K39 is methylated in association with active transcription at euchromatin[37,38]. We targeted the two histone sites with pairs of sgRNAs harbouring NGG PAM to generate double nicking mediated single-base substitution (Fig. 5A). All-in-one vectors expressing Cas9 nickase and sgRNAs were introduced into AX2 strains along with a single-stranded oligonucleotide (ssODN) containing a nucleotide substitution that would convert the corresponding amino-acid residue to alanine. Mutation detection PCR using primers with substituted nucleotides at the 3′ ends revealed that 18.2–36.4% of the independent clones yielded positive bands (Fig. 5C). In the case of these histones, we were able to design a pair of target sequences with canonical NGG PAMs around the nucleotide of interest. However, it is not always the case that two targets with NGG PAMs are located near a specific locus. Therefore, we used the Cas9 variants SpCas9-NG and SpRY to investigate whether a single-base substitution could be introduced into various targets harbouring non-canonical PAMs (Fig. 5A,B). PCR-positive base substitutions were present in 6.8–97.7% of the independent clones. In particular, base substitutions occurred in almost all (97.7%) of the clones expressing SpRY and sgRNA with the NAC PAM. To further evaluate the accuracy of single-base substitution, we analysed the nucleotide sequences; substitutions were present in 50–66% and 100% of PCR-positive clones expressing Cas9 nickase and Cas9 variants, respectively (Fig. 5C,D). E18 and E19 of H2Bv3 are ADP-ribosylated in response to DNA double-strand-breaks (DSBs)[37]. Hence, we investigated whether cells harbouring disruptions in histone modification sites were sensitive to the DSB-inducing reagent phleomycin. Cells harbouring H2Bv3 E18A E19A were more sensitive than AX2 to phleomycin, whereas cells harbouring a point mutation in H3a K39, a mark of active transcription, were insensitive (Fig. 5E).

Discussion

The ability to perform gene modification in regions that are not accessible by conventional SpCas9 expands the toolbox for editing the AT-rich genome of *D. discoideum*. We successfully used the Cas9 variants xCas9 3.7, SpCas9-NG, and SpRY to perform genome editing at various PAM sites. In particular, with SpRY, half of the targets harbouring NHN PAMs were editable with high efficiency. As previously shown for CRISPR/Cas9-mediated genome editing in *D. discoideum*, it is relatively easy to design target sequences within a gene of interest that contains an NGG PAM[11,12,39]. In fact, we can design target sequences with high specificity for all 283 of the known kinase genes. On the other hand, the PAM requirement is a severe barrier for applications other than gene knock-out, such as nucleotide substitution, targeting of a narrow genetic region, and tiling of regulatory elements[40–42]. For example, if we wished to design target sequences with NGG PAMs that cleave 25 bp before and after the start codon, we would not be able to target 251/283 (88.7%) of the kinase genes. Thus, SpCas9-NG and SpRY represent useful alternatives to SpCas9, and could be used to generate a variety of mutations that require precise DNA breaks at positions of interest not accessible by NGG PAMs.

The editing efficiencies of the three Cas9 variants were similar to those reported in other organisms, and the low editing efficiency of xCas9 3.7 at NG PAMs was consistent with results obtained in plants[24,29,30,43]. However, stable transgenic rice lines revealed that xCas9 3.7 functioned efficiently, with mutation rates of approximately 80% at NG PAM sites[25]. It is unclear whether the differences in editing efficiencies are due to stable versus transient expression or differences in other components. Recently, new SpCas9 variants that recognise NRNN PAMs have been engineered via the continuous evolution strategies used to generate xCas9 3.7[44]. Because xCas9 3.7 has high fidelity[45], future improvement is likely to make xCas9 a more practical choice for targeting non-NGG sites. Although SpCas9-NG and SpRY are presumed to generate more off-target effects due to the relaxed PAM requirements, SpCas9-NG had similar or even higher editing specificity in mammalian cells and rice[24,26]. SpRY was more prone to off-target editing than SpCas9, but these off-target effects could be eliminated by combining high-fidelity variants[29,46,47]. In *D. discoideum*, CRISPR/Cas9-mediated genome editing is achieved through transient expression of the all-in-one vector. Consequently, off-target effects are minimised, and elimination of the

6

A

H2Bv3 E18AE19A

```
      V K G Q K K A T K G S T Q S G E E K T A S T T P K V T K T P T E G G E K K R K K
5' GTTAAAGGTCAAAAGAAAGCAACCAAAGGTTCAACTCAATCTGGTGAAGAGAAAACCGCTTCAACCACCCCAAAAGTAACCAAAACCCCAACTGAAGGTGGTGAAAAGAAGAGAAAGAAG 3'
3' CAATTTCCAGTTTTCTTTCGTTGGTTTCCAAGTTGAGTTAGACCACTTCTCTTTTGGCGAAGTTGGTGGGGTTTTCATTGGTTTTGGGGTTGACTTCCACCACTTTTCTTCTCTTTCTTC 5'
```

120 bp ssODN
```
                                                                          A A
3' CAATTTCCAGTTTTCTTTCGTTGGTTTCCAAGTTGAGTTAGACCAcggcgaTTTTGGCGAAGTTGGTGGGGTTTTCATTGGTTTTGGGGTTGACTTCCACCACTTTTCTTCTCTTTCTTC 5'
```

H3a K39A

```
      H I G S K Q A H K Q T P V S S S S G G V K K V H R F R P G T V A L R E I R K Y Q
5' CACATTGGTTCTAAACAAGCCCATAAACAAACTCCAGTCTCTTCATCAAGTGGTGGTGTAAAGAAAGTACATCGTTTCAGACCAGGTACTGTCGCTCTTAGAGAAATCAGAAAATACCAA 3'
3' GTGTAACCAAGATTTGTTCGGGTATTTGTTTGAGGTCAGAGAAGTAGTTCACCACCACATTTCTTTCATGTAGCAAAGTCTGGTCCATGACAGCGAGAATCTCTTTAGTCTTTTATGGTT 5'
```

120 bp ssODN
```
                                                                  A
3' GTGTAACCAAGATTTGTTCGGGTATTTGTTTGAGGTCAGAGAAGTAGTTCACCACCACATcggTTTCATGTAGCAAAGTCTGGTCCATGACAGCGAGAATCTCTTTAGTCTTTTATGGTT 5'
```

B

H2Bv3 E18AE19A

```
   V K G Q K K A T K G S T Q S G E E K T A S T T P K V T K T P T E G G E K K R K K
GTTAAAGGTCAAAAGAAAGCAACCAAAGGTTCAACTCAATCTGGTGAAGAGAAAACCGCTTCAACCACCCCAAAAGTAACCAAAACCCCAACTGAAGGTGGTGAAAAGAAGAGAAAGAAG
                                    NGC CAATCTGGTGAAGAGAAAACCGC
```

H3a K39A

```
   H I G S K Q A H K Q T P V S S S S G G V K K V H R F R P G T V A L R E I R K Y Q
CACATTGGTTCTAAACAAGCCCATAAACAAACTCCAGTCTCTTCATCAAGTGGTGGTGTAAAGAAAGTACATCGTTTCAGACCAGGTACTGTCGCTCTTAGAGAAATCAGAAAATACCAA
                                    NGA TCATCAAGTGGTGGTGTAAAGAA
                                    NAC AAGTGGTGGTGTAAAGAAAGTAC
```

C

H2Bv3 E18AE19A

Cas9	PAM	PCR positive / No. screened	Efficiency (%)	Correct Seq./ No. analysed
Cas9 nickase	NGG	8/22	36.4	2/3
SpRY	NGC	4/26	15.4	3/3

H3a K39A

Cas9	PAM	PCR positive / No. screened	Efficiency (%)	Correct Seq./ No. analysed
Cas9 nickase	NGG	4/22	18.2	3/6
SpCas9-NG	NGA	18/32	56.3	3/3
SpRY	NAC	43/44	97.7	3/3
SpRY	NGA	3/44	6.8	3/3

D

H2Bv3 E18AE19A

```
  S  T  Q  S  G  A  A  K  T  A  S
TCAACTCAATCTGGTGCCGCTAAAACCGCTTCA
        160        170        180
```

H3a K39A

```
  S  S  G  G  V  A  K  V  H  R  F
TCAAGTGGTGGTGTAGCCAAAGTACATCGTTTC
      100        110        120
```

E

% Survival (y-axis: 0, 20, 40, 60, 80, 100)

Categories: AX2, cl. 2, cl. 10 (H2Bv3 E18AE19A), cl. 7, cl. 10 (H3a K39A)

Significance: *, *, n.s., n.s.

Figure 5. Precise nucleotide substitution mediated by double nicking and SpRY. (**A**) A pair of sgRNAs and ssODN for double-nicking. Point mutations are indicated by red letters. (**B**) Target sequences used for SpRY. Three targets with different PAMs are shown. (**C**) Summary of point mutation frequencies with the indicated CRISPR/Cas9 systems. (**D**) Sequencing chromatogram of each cell line. Mutated nucleotides and amino acids are shown in red. (**E**) Cell viability in response to phleomycin. Results are expressed as the percentage of point-mutated clones that survived. Error bars indicate S.D. of three biological replicates. *$P < 0.001$; n.s., not significant; Student's t-test. n = 4 biological replicates.

vectors can be confirmed by re-exposure to G418 selection. However, unintended editing remains a significant problem, and future studies should seek to further minimise this effect.

Importantly, we only tested a subset of PAMs. Because differences in 4 and 5 nt PAM sequences affect the editing efficiency[45], more targets must be tested to accurately determine efficiencies in *D. discoideum*. The lack of GFP knock-in at cAR1 with the NTA PAM, which had an efficiency of ~ 30% in tdTomato editing using SpRY, could be due to the fourth nucleotide of the PAM sequence. Although fine-tuning of PAM sequences requires further analysis, one effective method for generating mutants would be to design multiple targets for genome editing.

Previously, several 5′- or 3′-terminal knock-in mutants were successfully generated via homologous recombination in *D. discoideum*[48–50]. To date, however, no study has reported knock-in of GFP alone at the 5′-end of a gene. Even when GFP has been integrated at the 5′-end of a gene, the drug resistance cassette and *act6* promoter

for induction of target gene expression were introduced as extra sequences, and induction of expression could not be achieved under the control of the endogenous promoter[48]. Using the CRISPR-based knock-in method described here, it is possible to introduce the tag sequence anywhere in the gene of interest, allowing for analysis of tagged proteins expressed at approximately the same levels as the endogenous target proteins. This would make it possible to introduce tags into a gene that cannot be analysed with overexpression vectors. In addition, it would facilitate more accurate measurements of the number of intracellular molecules in live-imaging studies, such as Fluorescence Resonance Energy Transfer (FRET) experiments in G proteins[51] and GFP-tagging experiments to monitor protein expression levels[52]. For nucleotide substitution, base editors have frequently been used in various organisms as an alternative tool to HDR-mediated precise nucleotide substitution[53–56]. Although we have not used cytidine base editor (CBE) and adenine base editor (ABE) in *D. discoideum*, in this study we achieved HDR-mediated nucleotide substitution with efficiency of up to 97.7% using SpRY. Therefore, generation of amino-acid substitutions with SpRY could replace the traditional approach in which a gene with a point mutation is overexpressed in *D. discoideum*.

Our results show that the Cas9 variants used in this study substantially expand the utility of knock-in at a precise position within a target gene in *D. discoideum*. The versatile toolboxes developed here significantly expand the potential genome editing, making the majority of the *D. discoideum* editable.

Methods

Plasmid constructs. Coding sequences of xCas9 3.7, SpCas9-NG and SpRY followed by an NLS were codon-optimised for *D. discoideum* using the Codon Optimization Tool from Integrated DNA Technologies (IDT). The resultant 4.2 kb DNA sequences were divided into five fragments and synthesised as gBlocks (IDT) with 15 nt flanking sequences allowing In-Fusion HD cloning (Takara Bio). The first two and second three fragments were cloned into pBlueScriptII, and the sequences were confirmed by Sanger sequencing. The two fragments were joined using the KpnI site within Cas9 to obtain a full-length Cas9 fragment. The BglII- and SpeI-flanked Cas9 variants encoding xCas9 3.7 and SpCas9-NG were cloned into pTM1285 by replacing the Cas9-NLS-GFP sequence. The SpRY coding sequence was cloned into pTM1416 by replacing Cas9-NLS-GFP. pTM1416 was constructed by converting the two BpiI sites of the tRNA–sgRNA cassette in pTM1285 to two Esp3I sites by inserting a pair of annealed oligonucleotides, 5′-AGCAGGAGACGGGCGTCTCG-3′ and 5′-AAACCGAGACGCCCGTCTCC-3′. We also constructed the SpCas9-NG vector, which contained an tRNA–sgRNA cassette with two Esp3I sites (pTM1719), by replacing the two BpiI sites in pTM1718. To construct SpCas9-NLS vectors without GFP, we eliminated the GFP sequence in pTM1285 and pTM1416 by digestion with ClaI and SpeI, followed by blunting and self-ligation to yield pTM1599 and pTM1644, respectively (Table S2). Predicted nucleotide sequences for the all-in-one vectors are shown in Fig. S6. These plasmids will be made available to all researchers via NBRP Nenkin and other stock centres.

Design and cloning of sgRNAs. sgRNAs for SpCas9 were designed by identifying the NGG PAM sequence using CRISPR RGEN[57]. The freely available stand-alone Python scripts Cas-designer, Cas-OFFinder-bulge, and Cas-OFFinder Binaery ver. 2.4 were used to design 15 targeting sequences (NNN PAM with the exception of NGG PAM)[58]. Genome sequence (dicty_2.7) was downloaded via the Ensembl genome browser and converted to a 2-bit file that was smaller than the corresponding FASTA format. Sequences for target genes were saved in FASTA format. Output data were filtered as follows: GC content 30–75%, out-of-frame scores > 66, and mismatch number "1,0,0". In the case of duplicated genes such as tdTomato or *carA*, mismatch numbers "2,0,0" was used. The 20 nt targeting sequences without PAMs were synthesised as pairs of oligonucleotides with overhang sequences AGCA (forward) and AAAC (reverse) (Table S1). After mixing the two oligonucleotides (10 μM each) in annealing buffer (40 mM Tris–HCl pH 8.0, 20 mM MgCl₂, 50 mM NaCl), the mixture was heated at 95 °C for 5 min and cooled to 25 °C (1 °C/min). The annealed oligonucleotides (1.0 μl) were then ligated into an all-in-one vector (25 ng) via a Golden Gate digestion/ligation reaction using 70 U T4 DNA ligase and 1.5 U BpiI (Thermo) or Esp3I (NEB) in 4.0 μl reactions. The reactions were placed in a thermal cycler and subjected to eight cycles of 37 °C for 5 min and 16 °C for 17 min. After the Golden Gate reaction, additional BpiI- or Esp3I-digestion was performed at 37 °C for 60 min to remove all unligated all-in-one vectors. Transformed colonies were analysed for insertion of target sequence via colony PCR using forward target oligonucleotide and tracr-Rv (Tables S1 and S4). PCR primers were designed to amplify target sequence and tracrRNA within the sgRNAs such that the correct PCR band was ~ 120 bp. When further validation of the insert was necessary, sequencing analysis was performed using NeoUp (Table S4). For Cas9 nickase, a pair of target sequences with PAMs facing outward were selected and synthesised with overhangs appropriate for one-step Golden Gate as follows (Table S5). A pair of annealed oligonucleotides (0.26 μl each) prepared as described above, was ligated into pTM1544 (40 ng) via the Golden Gate reaction using 140 U T4 DNA ligase and 3.0 U BpiI (Thermo) in 8.0 μl reactions. Correct assembly of dual-sgRNA was confirmed via colony PCR using GFPdown and Reverse (+ direction) primers for the first target, and NeoUp2 and Forward (− direction) primers for the second target (Tables S4 and S5).

Cell culture, transformation, and identification of transformants. Axenic strains AX2 and tdTomato knock-in cells in the AX3 background were cultured at 22 °C in HL5 medium or on SM agar plates with *Klebsiella pneumoniae* (KpGe)[59]. To obtain cells transiently expressing Cas9 and sgRNAs, 10 μg all-in-one vector was transformed into cells using H50 buffer[11,12]. For knock-in or point mutagenesis, 2.4 μg donor DNA or 2.5 μl of 10 μM ssODN were added, respectively. After electroporation, the cells were cultured in HL5 for 7–24 h and then maintained for another 1–2 days in HL5 containing 10 μg/ml G418. As the cells became rounded, they were recovered in HL5 without G418. To isolate single clones, the cells were plated on SM agar plates and incubated for 4 days until plaques formed. Genomic DNA from single clones was isolated in lysis buffer (1× PCR buffer,

0.5% NP40 and 50 µg/ml of Proteinase K), and the suspension was incubated at 56 °C for 45 min followed by 95 °C for 10 min. Cell lysate was used as a template for PCR to detect mutations mediated by CRISPR/Cas9.

Calculation of targeting efficiency.   Targeting efficiency mediated by Cas9 variants was calculated by monitoring the decrease in tdTomato fluorescence within a single cell. Equal numbers of cells (1.8×10^5 cells/cm^2) were plated on Nunc Lab-Tek II two-well chambered coverglass. To visualise individual cells, their nuclear were stained with 10 µg/ml DAPI for 25 min. Blue and red fluorescence images were acquired on an Olympus IX71 inverted fluorescence microscope with a 0.75 NA 40× objective and Orca-Flash4.0 V2 Digital CMOS camera (Hamamatsu). We measured the intensity of tdTomato by averaging the intensity within individual cells using Volocity (Perkin Elmer). The efficiency was determined by pooling the intensity data of individual cells obtained from at least three independent transformations per vector. Cells that exhibited lower red fluorescence than tdTomato knock-in cells were defined as knockouts. Targeting efficiencies of knock-ins were calculated by PCR using primers flanking the target sites; efficiencies of point mutations were calculated using primers with substituted nucleotides at the 3′ ends (Table S4).

DNA damage assay.   Growing cells were plated on 24-well culture dishes at 2.5×10^5 cells/cm^2 and exposed to 10 µg/ml phleomycin (Cayman Chemical) or mock-treated for 30 min. Cell viability was determined by plating 100 cells on three SM agar plates and counting the number of plaques after 4–6 days.

Statistical analysis.   We used a one-way ANOVA followed by Tukey's post hoc test for the comparison of knock-out frequencies between PAM groups. The mean values of the survival rates of the wild-type and mutants in the DNA damage assay were considered to be equal using Student's t-test under the null hypothesis.

References

1. Eichinger, L. *et al.* The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**, 43–57. https://doi.org/10.1038/nature03481 (2005).
2. Kessin, R. H. *Dictyostelium—Evolution, Cell Biology, and the Development of Multicellularity* (Cambridge University Press, 2001).
3. Gaudet, P., Pilcher, K. E., Fey, P. & Chisholm, R. L. Transformation of *Dictyostelium discoideum* with plasmid DNA. *Nat. Protoc.* **2**, 1317–1324. https://doi.org/10.1038/nprot.2007.179 (2007).
4. Martens, H. *et al.* RNAi in *Dictyostelium*: the role of RNA-directed RNA polymerases and double-stranded RNase. *Mol. Biol. Cell* **13**, 445–453. https://doi.org/10.1091/mbc.01-04-0211 (2002).
5. Muramoto, T. *et al.* Construction of a gamete-enriched gene pool and RNAi-mediated functional analysis in *Dictyostelium discoideum*. *Mech. Dev.* **120**, 965–975 (2003).
6. Veltman, D. M., Keizer-Gunnink, I. & Haastert, P. J. An extrachromosomal, inducible expression system for *Dictyostelium discoideum*. *Plasmid* **61**, 119–125. https://doi.org/10.1016/j.plasmid.2008.11.002 (2009).
7. Yin, K., Gao, C. & Qiu, J. L. Progress and prospects in plant genome editing. *Nat. Plants* **3**, 17107. https://doi.org/10.1038/nplants.2017.107 (2017).
8. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **169**, 559. https://doi.org/10.1016/j.cell.2017.04.005 (2017).
9. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096. https://doi.org/10.1126/science.1258096 (2014).
10. Muramoto, T., Iriki, H., Watanabe, J. & Kawata, T. Recent advances in CRISPR/Cas9-mediated genome editing in *Dictyostelium*. *Cells* **8**, 46. https://doi.org/10.3390/cells8010046 (2019).
11. Sekine, R., Kawata, T. & Muramoto, T. CRISPR/Cas9 mediated targeting of multiple genes in *Dictyostelium*. *Sci. Rep.* **8**, 8471. https://doi.org/10.1038/s41598-018-26756-z (2018).
12. Iriki, H., Kawata, T. & Muramoto, T. Generation of deletions and precise point mutations in *Dictyostelium discoideum* using the CRISPR nickase. *PLoS ONE* **14**, e0224128. https://doi.org/10.1371/journal.pone.0224128 (2019).
13. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573. https://doi.org/10.1038/nature13579 (2014).
14. 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. U. S. A.* **109**, E2579-2586. https://doi.org/10.1073/pnas.1208507109 (2012).
15. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821. https://doi.org/10.1126/science.1225829 (2012).
16. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485. https://doi.org/10.1038/nature14592 (2015).
17. Chatterjee, P., Jakimo, N. & Jacobson, J. M. Minimal PAM specificity of a highly similar SpCas9 ortholog. *Sci. Adv.* **4**, eaau0766. https://doi.org/10.1126/sciadv.aau0766 (2018).
18. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823. https://doi.org/10.1126/science.1231143 (2013).
19. Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* **8**, 14500. https://doi.org/10.1038/ncomms14500 (2017).
20. Kleinstiver, B. P. *et al.* Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298. https://doi.org/10.1038/nbt.3404 (2015).
21. Ran, F. A. *et al.* In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191. https://doi.org/10.1038/nature14299 (2015).
22. Xu, K. *et al.* Efficient genome engineering in eukaryotes using Cas9 from *Streptococcus thermophilus*. *Cell Mol. Life Sci.* **72**, 383–399. https://doi.org/10.1007/s00018-014-1679-z (2015).
23. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63. https://doi.org/10.1038/nature26155 (2018).

24. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262. https://doi.org/10.1126/science.aas9129 (2018).
25. Hua, K., Tao, X., Han, P., Wang, R. & Zhu, J. K. Genome engineering in rice using Cas9 variants that recognize NG PAM sequences. *Mol. Plant* **12**, 1003–1014. https://doi.org/10.1016/j.molp.2019.03.009 (2019).
26. Endo, M. *et al.* Genome editing in plants by engineered CRISPR-Cas9 recognizing NG PAM. *Nat. Plants* **5**, 14–17. https://doi.org/10.1038/s41477-018-0321-8 (2019).
27. Fujii, W. *et al.* Generation of genetically modified mice using SpCas9-NG engineered nuclease. *Sci. Rep.* **9**, 12878. https://doi.org/10.1038/s41598-019-49394-5 (2019).
28. Ren, B. *et al.* Cas9-NG greatly expands the targeting scope of the genome-editing toolkit by recognizing NG and other Atypical PAMs in rice. *Mol. Plant* **12**, 1015–1026. https://doi.org/10.1016/j.molp.2019.03.010 (2019).
29. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296. https://doi.org/10.1126/science.aba8853 (2020).
30. Zhong, Z. *et al.* Improving plant genome editing with high-fidelity xCas9 and non-canonical PAM-targeting Cas9-NG. *Mol. Plant* **12**, 1027–1036. https://doi.org/10.1016/j.molp.2019.03.011 (2019).
31. Faix, J., Kreppel, L., Shaulsky, G., Schleicher, M. & Kimmel, A. R. A rapid and efficient method to generate multiple gene disruptions in *Dictyostelium discoideum* using a single selectable marker and the Cre-loxP system. *Nucleic Acids Res.* **32**, e143. https://doi.org/10.1093/nar/gnh136 (2004).
32. Wiegand, S., Kruse, J., Gronemann, S. & Hammann, C. Efficient generation of gene knockout plasmids for *Dictyostelium discoideum* using one-step cloning. *Genomics* **97**, 321–325. https://doi.org/10.1016/j.ygeno.2011.02.001 (2011).
33. Xiao, Z., Zhang, N., Murphy, D. B. & Devreotes, P. N. Dynamic distribution of chemoattractant receptors in living cells during chemotaxis and persistent stimulation. *J. Cell Biol.* **139**, 365–374. https://doi.org/10.1083/jcb.139.2.365 (1997).
34. Sun, T. J. & Devreotes, P. N. Gene targeting of the aggregation stage cAMP receptor cAR1 in *Dictyostelium*. *Genes Dev.* **5**, 572–582. https://doi.org/10.1101/gad.5.4.572 (1991).
35. Hsu, D. W., Chubb, J. R., Muramoto, T., Pears, C. J. & Mahadevan, L. C. Dynamic acetylation of lysine-4-trimethylated histone H3 and H3 variant biology in a simple multicellular eukaryote. *Nucleic Acids Res.* **40**, 7247–7256. https://doi.org/10.1093/nar/gks367 (2012).
36. Muramoto, T., Muller, I., Thomas, G., Melvin, A. & Chubb, J. R. Methylation of H3K4 Is required for inheritance of active transcriptional states. *Curr. Biol.* **20**, 397–406. https://doi.org/10.1016/j.cub.2010.01.017 (2010).
37. Rakhimova, A. *et al.* Site-specific ADP-ribosylation of histone H2B in response to DNA double strand breaks. *Sci. Rep.* **7**, 43750. https://doi.org/10.1038/srep43750 (2017).
38. Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **13**, 115–126. https://doi.org/10.1038/nrm3274 (2012).
39. You, Y., Ramachandra, S. G. & Jin, T. A CRISPR-based method for testing the essentiality of a gene. *Sci. Rep.* **10**, 14779. https://doi.org/10.1038/s41598-020-71690-8 (2020).
40. Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120–123. https://doi.org/10.1038/nature13695 (2014).
41. Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192–197. https://doi.org/10.1038/nature15521 (2015).
42. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667. https://doi.org/10.1038/nbt.3235 (2015).
43. Niu, Q. *et al.* Expanding the scope of CRISPR/Cas9-mediated genome editing in plants using an xCas9 and Cas9-NG hybrid. *J. Integr. Plant Biol.* **62**, 398–402. https://doi.org/10.1111/jipb.12886 (2020).
44. Miller, S. M. *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* **38**, 471–481. https://doi.org/10.1038/s41587-020-0412-8 (2020).
45. Kim, H. K. *et al.* High-throughput analysis of the activities of xCas9, SpCas9-NG and SpCas9 at matched and mismatched target sequences in human cells. *Nat. Biomed. Eng.* **4**, 111–124. https://doi.org/10.1038/s41551-019-0505-1 (2020).
46. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495. https://doi.org/10.1038/nature16526 (2016).
47. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88. https://doi.org/10.1126/science.aad5227 (2016).
48. Kypri, E., Schmauch, C., Maniak, M. & De Lozanne, A. The BEACH protein LvsB is localized on lysosomes and postlysosomes and limits their fusion with early endosomes. *Traffic* **8**, 774–783. https://doi.org/10.1111/j.1600-0854.2007.00567.x (2007).
49. Mukai, A., Ichiraku, A. & Horikawa, K. Reliable handling of highly A/T-rich genomic DNA for efficient generation of knockin strains of *Dictyostelium discoideum*. *BMC Biotechnol.* **16**, 37. https://doi.org/10.1186/s12896-016-0267-8 (2016).
50. Rodriguez-Centeno, J., Perona, R. & Sastre, L. Dyskerin mutations present in Dyskeratosis Congenita patients increase oxidative stress and DNA damage signalling in *Dictyostelium discoideum*. *Cells* **8**, 1406. https://doi.org/10.3390/cells8111406 (2019).
51. Janetopoulos, C., Jin, T. & Devreotes, P. Receptor-mediated activation of heterotrimeric G-proteins in living cells. *Science* **291**, 2408–2411. https://doi.org/10.1126/science.1055835 (2001).
52. Muramoto, T. *et al.* Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7350–7355. https://doi.org/10.1073/pnas.1117603109 (2012).
53. Gaudelli, N. M. *et al.* Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471. https://doi.org/10.1038/nature24644 (2017).
54. Hess, G. T., Tycko, J., Yao, D. & Bassik, M. C. Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol. Cell* **68**, 26–43. https://doi.org/10.1016/j.molcel.2017.09.029 (2017).
55. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424. https://doi.org/10.1038/nature17946 (2016).
56. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729. https://doi.org/10.1126/science.aaf8729 (2016).
57. Park, J., Bae, S. & Kim, J. S. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics* **31**, 4014–4016. https://doi.org/10.1093/bioinformatics/btv537 (2015).
58. 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. https://doi.org/10.1093/bioinformatics/btu048 (2014).
59. Lima, W. C. *et al.* Genome sequencing and functional characterization of the non-pathogenic *Klebsiella pneumoniae* KpGe bacteria. *Microbes Infect.* **20**, 293–301. https://doi.org/10.1016/j.micinf.2018.04.001 (2018).

Acknowledgements

Author contributions

T.M. conceived and designed the experiments; Y.A., K.Y., A.H. and T.M. performed the experiments; T.O. performed sgRNA design; H.I. contributed materials and analysis tools; T.M. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-89546-0.

Correspondence and requests for materials should be addressed to T.M.

Reprints and permissions information is available at www.nature.com/reprints.

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