Synthetic chimeric nucleases function for efficient genome editing

R.M. Liu¹, L.L. Liang¹, E. Freed¹, H. Chang², E. Oh¹, Z.Y. Liu², A. Garst³, C.A. Eckert¹,⁴ & R.T. Gill¹,⁵*

CRISPR–Cas systems have revolutionized genome editing across a broad range of biotechnological endeavors. Many CRISPR-Cas nucleases have been identified and engineered for improved capabilities. Given the modular structure of such enzymes, we hypothesized that engineering chimeric sequences would generate non-natural variants that span the kinetic parameter landscape, and thus provide for the rapid selection of nucleases fit for a particular editing system. Here, we design a chimeric Cas12a-type library with approximately 560 synthetic chimeras, and select several functional variants. We demonstrate that certain nuclease domains can be recombined across distantly related nuclease templates to produce variants that function in bacteria, yeast, and human cell lines. We further characterize selected chimeric nucleases and find that they have different protospacer adjacent motif (PAM) preferences and the M44 chimera has higher specificity relative to wild-type (WT) sequences. This demonstration opens up the possibility of generating nuclease sequences with implications across biotechnology.

¹Renewable and Sustainable Energy Institute (RASEI), University of Colorado, Boulder, CO, USA. ²Department of Biochemistry, University of Colorado, Boulder, CO, USA. ³Inscripta, Inc., Boulder, CO, USA. ⁴National Renewable Energy Laboratory, Golden, CO, USA. ⁵NNF-Center for Biosustainability, Danish Technical University, Lyngby, Denmark. *email: rtg@biosustain.dtu.dk
CRISPR-Cas (clustered regularly interspaced short palindromic repeats) driven genome editing and engineering has dramatically impacted biology and biotechnology in general\textsuperscript{12}. CRISPR-Cas editing systems require a polynucleotide guided nuclease, a guide polynucleotide (i.e., a guide RNA (gRNA)) that directs by homology the nuclease to cut a specific region of the genome, and optionally, a donor DNA cassette that can be used to repair the cut dsDNA and thereby incorporate programmable edits at the site of interest. The earliest demonstrations and applications of CRISPR-Cas editing used Cas9 nucleases and associated gRNA\textsuperscript{3-5}. These systems have been used for gene editing in a broad range of species encompassing bacteria to higher order mammalian systems\textsuperscript{4,5,7-13}. It is well established, however, that key editing parameters such as protospacer adjacent motif (PAM) specificity, editing efficiency, and off-target rates, among others, are species, loci, and nuclease dependent. As such, there is intense interest in identifying and rapidly characterizing nuclease systems that can be exploited to broaden and improve overall editing capabilities\textsuperscript{14}.

For this end, Cas12a-type nucleases have emerged as suitable alternatives to Cas9 nucleases, where several nucleases (e.g., Acidaminococcus sp. (AsCas12a) and Lachnospiraceae bacterium (LbCas12a)) have now been shown to display comparable genome-editing capability to Cas9 while providing different PAM preferences\textsuperscript{15-19}. We employ the term "Cas12a-type" or "like" throughout in recognition of consistent changes in CRISPR-Cas evolutionary classification and naming schemes\textsuperscript{14,20}. The structures of AsCas12a/LbCas12a contain a bi-lobed architecture consisting of an α-helical recognition (REC) lobe and a nuclease (NUC) lobe with a positively charged channel between them that binds the crRNA-DNA hybrid\textsuperscript{15,21} (Fig. 1a). The REC lobe consists of the REC1 and REC2 domains, and the NUC lobe consists of the RuvC domain and three additional domains, which are referred to as the WED, PI, and Nuc domains, respectively\textsuperscript{15,21} (Fig. 1a). The WED domain is assembled from three regions (WED-I, WED-II, and WED-III) in the Cas12a sequence (Fig. 1a). The REC lobe (REC1 and REC2) is located between the WED-I and WED-II regions, and the PI domain is inserted between the WED-II and WED-III regions (Fig. 1a). The RuvC domain contains the three motifs (RuvC-I, RuvC-II, and RuvC-III) (Fig. 1a). The bridge helix (BH) is located between the RuvC-I and RuvC-II motifs and connects the REC and NUC lobes, whereas the Nuc domain is inserted between the RuvC-II and RuvC-III motifs (Fig. 1a).

The canonical Cas12a protein–RNA complex recognizes a T-rich PAM and leads to a staggered DNA double-stranded break\textsuperscript{19,22} (Fig. 1a). The Cas12a-like nuclease interacts with the pseudoknot structure formed by the 5’-handle of crRNA\textsuperscript{22,23}. A guide RNA segment, composed of a seed region and the 3’ terminus, possesses complementary binding sequences with the target DNA sequences. Cas12a-like nucleases characterized to date have been shown to work with a single gRNA and to process gRNA arrays\textsuperscript{24}. In addition, Kim et al. compared the ratio of total off-target with on-target modification for AsCas12a and LbCas12a, and found that both orthologs show lower off-target activity than that previously observed with SpCas9\textsuperscript{24}. While Cas12a and Cas9 like nuclease systems have proven highly impactful, neither system has been shown to function as predictably as is desired to enable the full range of applications envisioned\textsuperscript{14}.

A range of efforts have attempted to engineer improved CRISPR editing systems, which included engineering of the PAM specificity, stability, and sequence of the gRNA and/or the nuclease\textsuperscript{25-32}. For example, chemical modifications of CRISPR–Cas9 gRNA expected to increase gRNA stability did indeed lead to 3.8-fold higher indel frequencies in human cells\textsuperscript{27}.

In addition, Gao et al. performed structure-guided mutagenesis of Cas12a and screened to identify variants with an increased range of recognized PAM sequences. The engineered AsCas12a recognized TYCV and TATV PAMs in addition to the canonical TTTT, with enhanced activities in vitro and in human cells\textsuperscript{29}. Using the crystal structures of Cas9\textsuperscript{22,33}, Slaymaker et al. performed rational engineering of the DNA binding region to attempt to decrease binding under the hypothesis that this would result in a lower off-target editing rates\textsuperscript{25}. Indeed, the authors reported eSpCas9(1.0) and eSpCas9(1.1) mutants that decreased 50% cleavage with off-target sites (<0.2% indel) relative to wild-type (WT) Cas9 using 20-nucleotide RNA guides\textsuperscript{25}. Strohikendl et al. reported that Cas12a binding to target DNA is rate limiting for cleavage, and suggested ways to engineer nucleases focused on altering such binding (kon)\textsuperscript{15}.

Here, we demonstrate a platform for construction of a library of synthetic nucleases that span the kinetic space encompassing overall editing considerations. In our design process we specifically note that even though Cas12a-like nucleases exhibit considerable overall sequence diversity, they also retain several conserved regions that we hypothesize may be recombined in a modular fashion from one Cas12a template to another to produce active chimeric nucleases. We hypothesize that such nucleases would exhibit altered characteristics that would allow for the discovery of sequences with desired editing characteristics. To test this hypothesis, we design and build, in Escherichia coli, a Cas12a-type nuclease library with 560 mutants that combined up to six conserved regions from a diverse starting pool of Cas12a-like nucleases (Supplementary Fig. 1 and Supplementary Table 1). We then select and screen for functional Cas12a-type chimeras, identifying several dozen that show basal editing. To then demonstrate the use of this strategy as a platform for rapid generation of nucleases with altered characteristics, we characterize several of the most active chimeras and demonstrate altered PAM preferences and on- vs off-targeting. The best performing chimera, M44, has higher specificity than all other tested nucleases. Finally, we demonstrate that such nucleases retain activity in yeast (Saccharomyces cerevisiae) and human cells (HEK293T), thus opening up the use of this strategy to rapidly build and select for synthetic nucleases with desired characteristics across a broad range of applications.

Results

Structure-guided design of Cas12a-type chimera library. A broad genetic and functional diversity of CRISPR-Cas immune systems have been identified across bacteria and archaea\textsuperscript{14}. Here, we sought to demonstrate a platform for rapid generation of a diverse set of nuclease sequences that would expand desired activities for genome-editing applications. We reasoned that instead of relying on isolation and characterization of Cas12a-like nucleases, the generation of chimeric sequences from known orthologs could more effectively expand the search space for the selection of novel functions\textsuperscript{34,35}. In addition, the module-like nature of CRISPR nucleases would facilitate domain recombination strategies for chimera generation\textsuperscript{36}. To this end, we selected nine Cas12a-type gene sequences (Supplementary Fig. 1 and Supplementary Table 1) as our initial starting pool that spanned a broad sequence space (560 variants) and tested their editing efficiencies (CFU/μg) of all the nucleases were at or higher than 10\textsuperscript{5} (Fig. 1b).

We employed a structure-guided design approach to build our chimeric nuclease library. We first aligned our nine Cas12a-like nucleases’ sequences with the AsCas12a and LbCas12a sequences.
Isolation of functional Cas12a-type chimeras. Functional chimeras were selected using both a galK based growth selection and galK colorimetric screen (Fig. 2a, Supplementary Figs. 2 and 3) and observed a relatively low level (33.8–42.99%) of sequence identity (Supplementary Table 2), thus limiting homology driven strategies and constraining the mutational space we could explore via chimeragenesis. Therefore, we had to identify modules based on known structures from AsCas12a/LbCas12a where we could design junction points for crossover. Based on the sequence alignment results, we identified six crossover points for chimeragenesis (Fig. 1c, Supplementary Fig. 2). The swapped regions are shown by different colors; the colors are the same as in (b). Source Data are available in the Source Data File.

Fig. 1 Cas12a-type chimera library construction. a The Cas12a-like protein structure analysis based on AsCas12a (PDB:5B43). b The editing and transformation efficiencies of the Cas12a-like nuclease used in this study. The Cas12a-like nuclease, used in this study, are SD_Cas12a (Succinivibrio dextrinosolvens), CT_Cas12a (Candidatus Methanoplasma termitum), TX_Cas12a (Thiomicrospira sp. X55), CA_Cas12a (Candidatus Methanomethylophilus alvus), PC_Cas12a (Porphyromonas crevioricanis), FB_Cas12a (Flavobacterium branchiophilum), CR_Cas12a (Candidatus Roizmannbacteria bacterium GW2011_GWA2_37_7), SC_Cas12a (synthetic construct of AsCas12a), and MAD7. The editing efficiency, as determined by galK inactivation assay, was shown in blue, and the transformation efficiency was shown in green. c The domains were separated by the six crossover points, which were WED-I&REC1, REC2, WED-II&PI, WED-III, RuvC-I&BH&RUVC-II, and Nuc&RuvC-III. The swapped regions are shown by different colors; the colors are the same as in (b). Source Data are available in the Source Data File.
natural nuclease sequences, in our experience the results from a single editing assay are not sufficient to understand the functional capabilities of an editing system.

Veriﬁcation of functional Cas12a-type variants. We further veriﬁed the activity of selected Cas12a-type chimeras using several additional editing assays and the MAD7 nuclease for comparison. We speciﬁcally selected a few additional inactivating mutations positioned in the galK and lacZ genes that could be used to measure nuclease-mediated cell killing (gRNA directed cutting only) (online methods, Fig. 3a, b) and nuclease-mediated editing via color screening (Fig. 3c and Supplementary Fig. 6). Mutants M44 and M21 (which is identical to M44 except for a single G218A substitution), displayed the same cutting efﬁciency (100%) with WT control using four of six designed gRNAs targeted on galK and lacZ genes (Fig. 3b). However, not every gRNA elicits cleavage although the gRNAs were targeted in the same gene. Low activity could result from either failure to form a functional Cas12a–gRNA complex or inability to recognize targets in vivo39–41. Then, the four gRNAs with high cutting efﬁciency were used for editing constructs as a representative sample and tested the inactivation editing at different position of galK and lacZ genes. We observed that the editing efﬁciencies of MAD7, M44, and M21 were not as high as their cutting efﬁciencies. Using galK2 and lacZ1 gRNAs, the editing efﬁciencies of M21 and M44 were 12.5% and 38% higher than MAD7. Using galK1 and lacZ2, the editing efﬁciency of MAD7 was 7% and 60% higher than M44 (mutants with highest editing efﬁciency) (Fig. 3c). These results indicated that there were more factors affecting the recombineering using Cas12a-type chimeras. Notably, the transformation efﬁciency for all the Cas12a-type chimeras were 2–5-fold higher than MAD7 (Supplementary Fig. 7). These results suggested the nuclease-mediated cell killing of Cas12a-type chimeras were not as strong as MAD7, which could be an important feature when working with species that are more diﬃcult to transform or attempting to construct large mutant libraries.

After testing different targets at the same position on the genome, we next tested how the editing capability of chimeras changed as a function of targeting diﬀerent positions in the genome. We speciﬁcally targeted ﬁve distinct “safe sites”, which are non-essential sites in E. coli BW25113 genome chosen for integration of heterologous genes with minimal predicted side eﬀects42,43. We deleted the WT galK gene, and then integrated the galK gene with a strong constitutive promoter J23119 into the ﬁve chosen safe sites (Supplementary Fig. 8a). Similar to the prior results, editing efﬁciency of M44, M21, and M38 was conﬁrmed albeit with consistently lower efﬁciency when compared with MAD7 (Supplementary Fig. 8b, c). We also noted a positional dependency conserved across all testing nuclease, with higher editing efﬁciency observed closer to the origin of replication (Supplementary Fig. 8b, c).

Increased expression increases editing efﬁciency. The editing efﬁciency of the characterized chimeras spanned a range of 5–95% depending on the speciﬁc PAM targeted (galk, lacZ) or the speciﬁc loci in the genome targeted. This broad range was consistent with our hypothesis that chimeric sequences would not only be functional but would provide a range of kinetic capabilities that could be used as starting points in screens or selections for desired performance (e.g., on- vs off-targeting). The core supposition here is that chimeric sequences are less stable
than WT sequences since they have not been selected for function in nature. Lowered stability could affect function broadly, including altered on-/off-targeting and cleavage kinetic constants, or overall concentrations due to increased degradation in vivo.44. Including altered on-/off-targeting and cleavage kinetic constants, than WT sequences since they have not been selected for function in nature.

To investigate the above supposition, we sought to evaluate the chimeric Cas12a-like proteins through the process of CRISPR editing, including on-targeting binding, cutting, and editing associated with recombineering proteins. Thus, we first developed a dCas12a (or Cas12a with greatly reduced activity) protein binding assay that allowed the on-target and off-target status to be monitored by antibiotic selections in E. coli (Fig. 3d). We constructed a three plasmid system that expresses dCas12a (or Cas12a with greatly reduced activity) using a single crRNA (with J23119 promoter) targeting the galK or lacZ gene and a homology arm (HM) containing a gene-inactivating mutation. For the cutting, there were no lambda red proteins or homology arm in the system.

The genome-editing test with different gRNAs for chimera library variants in E. coli. a Editing (cutting) efficiency test using gRNA targeting galK or lacZ genes. We constructed a two plasmid system for genome editing; one plasmid expresses a Cas protein as well as lambda red proteins (exo, bet, and gam)67; a second plasmid expresses a single crRNA (with J23119 promoter) targeting the galK or lacZ gene and a homology arm (HM) containing a gene-inactivating mutation. For the cutting, there were no lambda red proteins or homology arm in the system. b Cutting efficiency of chimeric Cas12a-like proteins using six different gRNA plasmids (online methods). The gRNA plasmids galK1, galK2, and galK3 targeted different positions in the galK gene with homology arm (HM) which makes the cells sensitive to metronidazole. e, f The cutting efficiency of chimeric Cas12a-like nucleases with different arabinose induction times using different gRNA. e galK1, f galK2. g The arabinose inducible system for chimeric Cas12a-like proteins. We constructed a three plasmid system for genome editing; one plasmid expresses a Cas protein as well as lambda red proteins (exo, bet, and gam) using a temperature-inducible promoter (pL); a third plasmid expresses a single crRNA (with J23119 promoter) targeting the galK gene with homology arm (HM) containing a galK-inactivating mutation as a template for recombineering. h, i The editing efficiency of chimeric Cas12a-like nucleases with different arabinose induction times using different gRNA. h galK1, i galK2. Source Data are available in the Source Data File.

Fig. 3 The genome-editing test with different gRNAs for chimera library variants in E. coli. a Editing (cutting) efficiency test using gRNA targeting galK or lacZ genes. We constructed a two plasmid system for genome editing; one plasmid expresses a Cas protein as well as lambda red proteins (exo, bet, and gam)67; a second plasmid expresses a single crRNA (with J23119 promoter) targeting the galK or lacZ gene and a homology arm (HM) containing a gene-inactivating mutation. For the cutting, there were no lambda red proteins or homology arm in the system. b Cutting efficiency of chimeric Cas12a-like proteins using six different gRNA plasmids (online methods). The gRNA plasmids galK1, galK2, and galK3 targeted different positions in the galK gene. The gRNA plasmids lacZ1, lacZ2, and lacZ3 targeted different positions in the lacZ gene. c The editing efficiency of chimera library variants with different gRNAs. The gRNAs used in the test were galK1, galK2, lacZ1, and lacZ2. Editing efficiency was determined by color screening (red/white for GalK or blue/white for LacZ). d The dCas12a (or Cas12a with reduced activity) protein binding assay. We constructed a three plasmid system; one plasmid expresses dCas12a (or Cas12a with reduced activity) using an arabinose inducible promoter (pBAD); a second plasmid expresses a single crRNA (with J23119 promoter) targeting the kanR gene; a third plasmid expresses the kanamycin resistance protein (encoded by kanR gene) using a constitutive promoter containing a fully complementary (on-target) crRNA binding site as well as a nitroreductase (encoded by nfsI gene) which makes the cells sensitive to metronidazole. e, f The cutting efficiency of chimeric Cas12a-like nucleases with different arabinose induction times using different gRNA. e galK1, f galK2. g The arabinose inducible system for chimeric Cas12a-like proteins. We constructed a three plasmid system for genome editing; one plasmid expresses a Cas12a-like protein using an arabinose inducible promoter; a second plasmid expresses lambda red proteins (exo, bet, and gam) using a temperature-inducible promoter (pL); a third plasmid expresses a single crRNA (with J23119 promoter) targeting the galK gene with homology arm (HM) containing a galK-inactivating mutation as a template for recombineering. h, i The editing efficiency of chimeric Cas12a-like nucleases with different arabinose induction times using different gRNA. h galK1, i galK2. Source Data are available in the Source Data File.
gene, and a kanamycin resistance protein (encoded by *kanR* gene) using a constitutive promoter containing a fully complementary (on-target) crRNA binding site as well as a nitroreductase (encoded by *nfiI* gene) which conferred the cells sensitive to metronidazole45 (Fig. 3d and Supplementary Fig. 9). We introduced two tested on-target crRNA (galK1 and galK2) binding sites in the upstream of *kanR* gene individually. Using dCas12a:crRNA as a transcriptional repressor, the cells cannot grow with unexpressed kanamycin, and expressed nitroreductase also repress the cell growth with metronidazole (Fig. 3d).

The decreased cell grown under antibiotic selection means the dCas12a:crRNA repressed the transcription of kanamycin resistance protein. However, repression levels of chimeras were 5–60% lower than the WT MAD7 nucleuse (Supplementary Fig. 10). Recent publications described that DNA target binding by CRISPR-Cas12a is rate limiting for DNA cleavage, and the maximal rate constant (k\text{max}) for the targeted DNA binding of (AsCas12a) is 0.13 ± 0.01 s\(^{-1}\) which is orders of magnitude slower than the DNA cleavage19. In light of our DNA binding assay data, we hypothesized that the chimeric Cas12a-like proteins were less stable than the WT Cas12a under the same expression level, and increased chimera degradation may limit the overall DNA binding rate.

To test this hypothesis, we introduced the Cas12a-like proteins into three plasmid system to replace the dCas12a, and tested the cutting efficiency with the different expression level of Cas12a proteins by controlling the induction time. Prolonged induction increases the Cas12a expression level, which resulted in increased cutting efficiency of chimeric Cas12a-like proteins (Fig. 3c, f). In addition, we also tested the genome editing under this inducible system (Fig. 3g). The results showed that the editing efficiency of chimeric Cas12a-type mutants were further improved with longer induction time, and the editing efficiency of M44 and M21 were similar to the WT Cas12a after 2 h induction (Fig. 3b, i). These studies support the hypothesis that the studied chimeric nucleases are less stable than WT sequences and that this effect can be mitigated by increased expression. Again, these data demonstrate that chimeric sequences do indeed provide for the development of editing systems that span a broad kinetic landscape.

**Altered PAM preferences and off-target editing rates.** Prior efforts have engineered CRISPR nucleases to achieve reduced off-targeting and altered PAM preferences by targeted mutagenesis of the modules responsible for such activities25–29. The thought in such studies was to reduce or alter overall DNA binding such that off-target events would not be energetically favorable enough to allow for cutting. We utilized similar rationale in our chimera strategy, and thus expected that the functional chimeras identified may have altered PAM preferences and/or off-target rates.

We first tested the PAM preferences of three of the selected chimeras. To elucidate functional PAM sequences, we developed a high-throughput in vivo screen with two features: applicability across PAM-dependent CRISPR-Cas systems and the generation of a distinct signal for functional PAMs. More recent efforts have developed high-throughput experimental screens to determine functional PAMs based on the depletion of a target plasmid or on the introduction of a double-stranded break in vitro46–48. Here, we modified the Cas12a binding assay described above to generate a comprehensive screen to elucidate the complete landscape of functional PAM sequences (Supplementary Fig. 11a). We constructed the reporter plasmid containing *KanR* gene encoding for kanamycin resistance and the functional protospacer with NNNN PAM library. We then transformed the chimeric Cas12a-like proteins and two of equivalent gRNA plasmids individually into the *E. coli* MG1655. One gRNA design is targeted on the *KanR* gene, and another gRNA is a non-targeting control. We collected the cells grown on kanamycin media using different gRNA plasmids, and amplified the region of the PAM library from the reported plasmid for the high-throughput sequencing.

The PAM enrichment score revealed that PAM preferences appeared to differ among the chimeric Cas12a-like proteins tested (Fig. 4, Supplementary Fig. 11b–h, 12). Interestingly, the TTTC PAM is still the top one for the tested chimeric Cas12a-like and WT Cas12a proteins (expect TX_Cas12a) (Fig. 4, Supplementary Fig. 12). Furthermore, we observed that CITT PAM with the lowest enrichment score for all known PAMs (Fig. 4, Supplementary Fig. 12), which may be explained by weak and strong PAMs all eliciting irreversible DNA damage and infrequent escape49–51. To confirm these high-throughput observations, we tested several of the previously unreported PAMs individually, and the results also revealed PAM specificity among different chimeric Cas12a-like proteins (Supplementary Fig. 13).

Off-target mutations observed at frequencies greater than desired is still a major concern when applying CRISPR systems to biomedical and clinical application24,48,52,53. Several prior studies have engineered altered off-target rates by site-directed and random mutagenesis on CRISPR nucleases to decrease non-specific interactions with target DNA25,26,54,55. Thus, we expected that our chimera strategy may affect the non-specific interactions with the target DNA site. We carried out an off-targeting library test in which we assessed the effects of systematically mismatching various positions within gRNAs. We designed nine of such off-target cassettes, including 3 each with substitutions, insertions, or deletions in different positions (Fig. 5a). We observed considerable potential for off-target activity in both LbCas12a and MAD7, only a single potential off-target for AsCas12a, and no off-targeting for the chimeric M44 (color screening assay) (Fig. 5a, Supplementary Fig. 14a–c). While these data were compelling, we further expanded these studies with a comprehensive and sensitive genome-wide off-target assay (CIRCLE-seq)36,57. Using *E. coli* MG1655 genome as targeting DNA, we tested 2 gRNAs (galK1 and lacZ2) using AsCas12a, LbCas12a, MAD7, and M44 (Fig. 5b, c). The target sequence is shown at the top of the figure (Fig. 5b, c) and off-target sequences are shown below. The read counts for each sequence are shown to the right and represent a measure of cleavage efficiency at a given site. For each gRNA, the target sequence is shown at the top of the figure (Fig. 5b, c) and off-target sequences are shown below (Fig. 5b, c). The read counts for each sequence are shown to the right and represent a measure of cleavage efficiency at a given site. M44 showed the same or less off-target activity when compared with AsCas12a and LbCas12a, and less off-target activity than MAD7. For the first guide, M44, AsCas12a, and LbCas12a all had 0% off-target activity, while MAD7 had 0.8% off-target activity (Fig. 5b). For the second guide, M44, AsCas12a, and LbCas12a, had 0, 7.8, and 23.7%, off-target activity, respectively, while MAD7 had more off-target activity than on-target (Fig. 5c). These results generally confirmed the results obtained using our designer off-target assay (Fig. 5b, c), MAD7 and LbCas12a had substantially higher off-targeting relative to AsCas12a and the chimeric M44.

**Chimeric nucleases enable genome editing in eukaryotic cell.** Targeted genome editing with the hope of treating and curing diseases has always been a major goal in the field of biology. Thus, we tested the chimeric Cas12a-like nucleases for genome editing in mammalian cells. We constructed a plasmid expressing the M44 chimeric nuclease (with T7 promoter), a single crRNA (with
Fig. 4 The specificity detection of chimeric Cas12a-type variants. a-f The enrichment score for two rounds of PAM scans. The enrichment score is the frequency change (log2) of each PAM using different gRNA plasmids (on-targeting and non-targeting gRNAs) (online method). a AsCas12a, b LbCas12a, c TX_Cas12a, d MAD7, e M44, and f M21.

Fig. 5 The off-target assay for chimeric Cas12a-type variants. a The individual off-target assay. We designed nine different off-target spacers, of which three were substitutions, three were deletions, and three were insertions. b, c Genome-wide off-target analysis was done using the CIRCLE-seq method56,57. b gRNA targeting the galK1 site and c gRNA targeting the lacZ2 site. Positions with mismatches to the target sequences, i.e., off-target sites, are highlighted in color. CIRCLE-seq read counts are shown to the right of the on- and off-target sequences and represent a measure of cleavage efficiency at a given site. The on/off-target reads shown in the figure were higher than 10. Source Data are available in the Source Data File.
U6 promoter), and GFP as a transfection control (Fig. 6a). Transfected cells were isolated by FACs (Fig. 6b), cell lysate harvested 72 h post-transfection, and indel detection was performed using T7E1 assay52,58,59 (Fig. 6b, c). The results demonstrated that this chimeric nuclease is fully functional in mammalian gene editing experiments, although it has a lower editing efficiency than MAD7 (Fig. 6d) and AsCas12a60.

In addition to mammalian cells, we further tested the chimeric Cas12a-type variants for genome editing in S. cerevisiae (Fig. 6c). S. cerevisiae has long been the most tractable organism for eukaryotic cell biology, owing to its genetic malleability, greatly facilitated by a preference for homologous recombination (HR) over non-homologous end joining (NHEJ) for double-stranded break (DSB) repair. To examine chimeric Cas12a-type nuclease activity in S. cerevisiae, we designed gRNAs to target the endogenous genomic negative selectable marker CAN1, which its null mutation can be selected with media containing canavanine (a toxic arginine analog)61. We found M44 edited at efficiencies >40%, which is lower than the editing efficiency of ~86% seen with MAD7 (Fig. 6f). Collectively, these results demonstrated that the chimeragenesis strategy reported can be used to rapidly develop synthetic nucleases with altered characteristics that are functional across multiple species.

**Discussion**

Structure-guided chimeragenesis is an effective way to generate synthetic protein families with broad sequence diversity while maintaining a relatively high percentage of folded and functional proteins34,35. Furthermore, the proportion of folded variants can be increased through simple solutions such as utilizing stabilized parental sequences. Large datasets are generated by characterizing these libraries, and, unlike natural protein families, these sets include both functional and nonfunctional sequences that can be queried for specific properties in high-throughput formats62. There are abundant Cas12a-like family proteins in the database. However, not all the characterized Cas12a-like proteins are efficient in the model systems, such as E. coli, yeast, and mammalian cells15, emphasizing the need to improve existing nuclease classifications. While many Cas12a-like sequences are easily identified, predicting which ones are functional and what are the preferred PAM designs and gRNA designs remains intractable. Here, we designed a Cas12a-type nuclease library with 560 mutants using domain recombination points based on the homology substructure of nine different Cas12as, and then to account for the lack of predictability we developed a selection/screening system to identify functional Cas12a-type chimeras. Interestingly, we observed that ~30% sequences of the positive
variants do not align to the WT Cas12a proteins. In addition, the PAM specificity and off-targeting characteristics were different among the characterized chimera mutants. Our approach therefore not only emphasizes the unique status of such nucleases but also the potential of this strategy for generating nucleases fit for a particular application.

Genome-editing efficiency and off-target effects are major challenges for the broad application of CRISPR systems63,64. Recently, several groups have reported that tailored substitutions abolishing nonspecific contacts between SpCas9 and the DNA substrate generate more precise RNA-guided endonucleases (eSpCas9(1.1)25, evoCas955, and SpCas9–HFI126) with increased dependency on sgRNA:target DNA pairing. Even though these variants offer improved specificity, for some sites, off-target cleavage remains a problem25,26,54,55. Here, we have discovered several chimeric Cas12a-type variants with reduced off-target activity yet well-preserved on-target activity as demonstrated in vitro and in vivo in E. coli. We further showed that the chimeric Cas12a-like nucleases facilitate genome editing in E. coli, yeast, and mammalian cells, although editing efficiency in both yeast and mammalian cells is lower with the M44 chimera than with other nucleases50 (Fig. 6d, f). While the M44 chimeric nuclease could benefit from further optimization to improve on-target editing in eukaryotic cells, there are also instances (e.g., allogenic cell therapies) in which having lower off-target activity may be a higher priority than having high on-target activity. We believe these chimeric nucleases could be beneficial in downstream applications ranging from human health to industrial biotechnology.

Engineering Cas12a-like nucleases based on structural information expands the current genome-editing toolbox29. For example, future efforts to further alter PAM preference or on/off-target specificity could entail the evaluation of a much larger collection of REC1 domains or by saturation/random mutagenesis of specific regions of chimeric Cas12a-like nucleases, among other well established directed evolution approaches. Furthermore, our strategy can be adapted to engineer additional homologous and non-homologous RNA-guided endonucleases. Finally, our screening platform could be applied for the development of chimeric Cas12a-type variants tailored to a range of specific functional objectives, such as optimization of editing at a specific loci or in a targeted cell line, among others. The vast collection of already identified nucleases in combination with a rapid approach for generating large combinations thereof opens up the potential for generation of specialized synthetic nucleases tailor made for a range of applications.

Methods

Chimeric Cas12a-type nuclease library construction. Using Cas12a-type nuclease sequences available from the NCBI database, we performed alignments (Supplementary Fig. 2) to determine homologous domains to design chimera library sequences (Table S1). We used a ~40 bp homology arm with MAD7 and its nuclease sequences available from the NCBI database, we performed alignments (Supplementary Fig. 2) to determine homologous domains to design chimera library sequences. The isolation of functional Cas12a-type mutants started with negative selection. The number of colonies that can grow on the plate with on-targeting gRNA plasmid, and \( b \) is the number of colonies that can grow on the plate with non-targeting gRNA plasmid.

Generation of heterologous plasmids. To generate the Cas12a locus for heterologous expression, the Cas12a-type DNA sequences after codon optimization was PCR amplified and cloned into pSpCas9(bbt2), pX2, pMiniK, and pYF094 using Gibson cloning kit (New England Biolabs). Sequences of all the chimera and gRNA design can be found in Supplementary Data 1.

The isolation of functional Cas12a-type mutants. The host strain carried the plasmid expressing lambda red proteins and chimeric Cas12a-like proteins library. The strain were cultured in 30 °C and supplemented with 0.2% arabinose for inducing lambda red proteins. When OD₆₀₀ reached 0.5–0.6, the cells were induced for 15 min at 42 °C to induce chimeric Cas12a-like proteins. After chilling on ice for 15–30 min, the cells were washed twice with 20% of the initial culture volume of ddH₂O. Then, the gRNA plasmid was mixed with the cells, followed by chilling on ice for 5 min. Following electroporation, the cells were recovered in SOb medium for 3 h. Then, 1 μL of cells was plated in the M9 agar media supplemented with 2-deoxy-galactose (DOG). The isolation of functional Cas12a-type mutants directly in vivo potentially enabled the identification of Cas12a–Cas9 mutants with higher editing efficiency. The galk gene product, galactokinase, catalyzes the first step in the galactose degradation pathway, phosphorylating galactose to galactose-1-phosphate. Galactokinase also efficiently catalyzes the phosphorylation of a galactose analog, 2-deoxy-galactose (DOG). The product of this reaction cannot be further metabolized, leading to a toxic build up of 2-deoxy-galactose-1-phosphate38. Thus, strains with galk inactivation can grow in the media supplementation with 2-DOG and background following negative selection is reduced and no colony screening is necessary. The selected Cas12a-type mutants were verified using the above competent cell preparation and transformation method. After 3 h recovery, 1 μL of cells was plated in the MacConkey agar. The color screening method based on the galk inactivation to evaluate the editing efficiency of CRISPR–Cas9 was same as the previous studies27.

Cas12a PAM screen. PAM plasmid libraries were constructed using synthesized oligonucleotides (IDT) containing the designed NNNN PAM library. The dsDNA product was assembled into a linearized plasmid (containing kanR gene) using Gibson cloning (New England Biolabs). The gRNA library was transformed into MG1655 with the plasmid expressing chimeric Cas12a-like proteins using the electroporation method. We then transformed two equivalent gRNA plasmids individually into the E. coli MG1655. One gRNA design is targeted on the library sites, and another gRNA plasmid is non-targeting control. We collected the cells grown on kanamycin media using different gRNA plasmids, and amplified the GalE gene from the gRNA plasmid. The Ligation and transformation were conducted using high-frequency transformation and the assembled plasmid was sequenced using negative selection. The enrichment score of PAM and accompanying sequence logo for one of two library replicates are shown in PAM screening revealed the PAM specificity were different between different chimeric Cas12a-like proteins. The prepared cDNA libraries were sequenced on a MiSeq with a single-end 300 cycle kit (illumina). Indels were mapped using a Python implementation of the Geneious 6.0.3 Read Mapper.

Nucleosome-mediated cell killing assay. We constructed a two plasmid system for genome editing, which expresses a Cas12a-like protein and a single cRNA (with J23119 promoter) targeting the galk or lacZ gene. For each experiment, we transformed equal amounts of non-targeting and on-targeting (e.g., galkKI) gRNA plasmids. The cutting efficiency was calculated as following:

\[
\text{Cutting efficiency} = \left(1 - \frac{b}{a}\right) \times 100\%
\]

The same amount of culture was plated in two LB agar plates with chloramphenicol and ampicillin, and the number of colonies that can grow on the plate with on-targeting gRNA plasmid, and \( b \) is the number of colonies that can grow on the plate with non-targeting gRNA plasmid.

Nucleosome-mediated cell killing assay. We constructed a two plasmid system for genome editing, which expresses a Cas12a-like protein and a single cRNA (with J23119 promoter) targeting the galk or lacZ gene. For each experiment, we transformed equal amounts of non-targeting and on-targeting (e.g., galkKI) gRNA plasmids. The cutting efficiency was calculated as following:

\[
\text{Cutting efficiency} = \left(1 - \frac{b}{a}\right) \times 100\%
\]

The same amount of culture was plated in two LB agar plates with chloramphenicol and ampicillin, and the number of colonies that can grow on the plate with on-targeting gRNA plasmid, and \( b \) is the number of colonies that can grow on the plate with non-targeting gRNA plasmid.
relative band intensities. Indel percentage was determined by the formula, $100 \times \frac{G}{T7E1 \text{ assay}}$.

Supplementary Figs. 7, 8, 10, 11, 13 are provided as a Source Data le for review by the authors upon reasonable request. The source data underlying Figs. 1, 2, 3, 5, 6 and 10 are available in the Nature Research Reporting Summary linked to this article.

1. Adli, M. The CRISPR tool kit for genome editing and beyond. Nat. Commun. 9, 1911 (2018).
2. Wu, W. Y., Lebbink, J. H. G., Kanaar, R., Geijsen, N. & van der Oost, J. Genome editing by natural and engineered CRISPR-associated nucleases. Nat. Chem. Biol. 14, 642–651 (2018).
3. Liu, J. et al. Multiplex gene editing using CRISPR-Cas systems. Science 339, 819–823 (2013).
4. Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823–826 (2013).
5. Mali, P., Easelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. Nat. Methods 10, 957–963 (2013).
6. Garst, A. D. et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. Nat. Biotechnol. 35, 48–55 (2017).
7. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442–451 (2013).
8. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346, 1258096 (2014).
9. Huo, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262–1278 (2014).
10. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Rev. Genet. 32, 357–371 (2015).
11.Tsai, S. Q. & Joung, J. K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. Nat. Rev. Genet. 17, 300 (2016).
12. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. 37, 67–78 (2017).
13. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771 (2015).
14. Fonfarà, L., Richter, H., Bratovic, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursors CRISPR RNA. Nature 532, 517–521 (2016).
15. Hur, J. K. et al. Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. Nat. Biotechnol. 34, 807–808 (2016).
16. Kim, Y. et al. Generation of knockout mice by Cpf1-mediated gene targeting. Nat. Biotechnol. 34, 808–810 (2016).
17. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR-Cas systems. Nat. Rev. Microbiol. 15, 169–182 (2016).
18. Zetsche, B. et al. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. Nat. Biotechnol. 35, 31 (2016).
19. Yamano, T. et al. Crystal structure of Cpf1 in complex with guide DNA and target DNA. Cell 165, 949–962 (2016).
20. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. Nat. Rev. Genet. 15, 259–310 (2014).
21. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. Science 339, 823–826 (2013).
22. Meyer, M. M., Hochrein, L. & Arnold, F. H. Structure-guided SCHEMA recombination generates diverse chimeric chimerichordoplasms. Proc. Natl Acad. Sci. USA 114, E2624–E2633 (2017).
23. Meyer, M. M., Hochrein, L. & Arnold, F. H. Structure-guided SCHEMA recombination of distantly related beta-lactamases. Protein Eng. Des. Sel. 19, 563–570 (2006).
24. Voigt, C. A., Mayo, S. L., Arnold, F. H. & Wang, Z. G. Computational method to reduce the search space for directed protein evolution. Proc. Natl Acad. Sci. USA 98, 3778–3783 (2001).
25. Liu, L. et al. Iterative genome engineering of Escherichia coli for 3-hydroxypropionic acid production. Metab. Eng. 47, 303–313 (2018).
26. Alper, M. D. & Ames, B. N. Positive selection of mutants with deletions of the FNR promoter. J. Bacteriol. 182, 259–266 (1995).
43. Egbert, R. G. et al. A versatile platform strain for high-fidelity multiplex genome editing. *Nucleic Acids Res.* 47, 3244–3256 (2019).
44. Farasat, I. & Salis, H. M. A biophysical model of CRISPR-Cas9 activity for rational design of genome editing and gene regulation. *PLoS Comput. Biol.* 12, e1004724 (2016).
45. Nillius, D., Müller, J. & Müller, N. Nitroreductase (GlnR1) increases susceptibility of *Giardia lamblia* and *Escherichia coli* to nitro drugs. *J. Antimicrob. Chemother.* 66, 1029–1035 (2011).
46. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31, 233–239 (2013).
47. Karvelis, T. et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol.* 16, 253 (2015).
48. Pattanayak, V. et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 31, 839 (2013).
49. Gomaa, A. A. et al. Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *MiBio* 5, e00928–13 (2014).
50. Jiang, W. et al. Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genet.* 9, e1003844 (2013).
51. Vercoe, R. B. et al. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet.* 9, e1003454 (2013).
52. Kleinstiver, B. P. et al. Genome-wide specificities of CRISPR-Cas9 nucleases in human cells. *Nat. Biotechnol.* 34, 869 (2016).
53. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822–826 (2013).
54. Chen, J. S. et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550, 407–410 (2017).
55. Casini, A. et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* https://doi.org/10.1038/nbt.4066 (2018).
56. 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).
57. Lazzarotto, C. R. et al. Defining CRISPR-Cas9 genome-wide nuclease activities with CIRCLE-seq. *Nat. Protoc.* 13, 2615–2642 (2018).
58. Li, B., Zeng, C. & Dong, Y. Design and assessment of engineered CRISPR-Cpf1 and its use for genome editing. *Nat. Protoc.* 13, 899–914 (2018).
59. Kleinstiver, B. P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523, 481–485 (2015).
60. Li, B. et al. Engineering CRISPR-Cpf1 crRNAs and mRNAs to maximize genome editing efficiency. *Nat. Biomed.Eng.* 1, 0066 (2017).
61. DiCarlo, J. E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41, 4336–4343 (2013).
62. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 16, 379–394 (2015).
63. Bosley, K. S. et al. CRISPR germline engineering—the community speaks. *Nat. Biotechnol.* 33, 478 (2015).
64. Bolukbasi, M. F., Gupta, A. & Wolfe, S. A. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nat. Methods* 13, 41–50 (2016).
65. Liu, R. et al. Multiplex navigation of global regulatory networks (MINR) in yeast for improved ethanol tolerance and production. *Metab. Eng.* 51, 50–58 (2019).
66. Gietz, R. D. & Schiestl, R. H. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 35–37 (2007).
67. Bassalo, M. C. et al. Deep scanning lysine metabolism in *Escherichia coli*. *Mol. Syst. Biol.* 14, e8371 (2018).

**Acknowledgements**

The project was sponsored by the US Department of Energy (Grant DE-SC0018368).

**Author contributions**

R.L. and R.T.G. developed the concept. R.L., L.L., C.E. and A.D.G. all aided in the design of experiments. The experiments of *E. coli* were done by R.L., L.L., H.C. and F.E. The experiments of yeast were done by L.L. and O.E. The experiments of mammalian cells were done by R.L., F.E. and Z.L.

**Competing interests**

The authors declare the following competing interests: R.T.G. and A.G. have financial interests in Inscripta, Inc.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-13500-y.

**Correspondence**

requests for materials should be addressed to R.T.G.

**Peer review information**

*Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information**

is available at http://www.nature.com/reprints

**Publisher’s note**

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

© The Author(s) 2019