https://doi.org/10.15388/vu.thesis.315
https://orcid.org/0000-0002-1432-1433

VILNIUS UNIVERSITY

Greta Bigelytė

Miniature CRISPR-Cas nucleases: from characterization to application

DOCTORAL DISSERTATION

Natural Sciences,
Biochemistry (N 004)

VILNIUS 2022

The dissertation was prepared between 2017 and 2021 at Vilnius University, Life Sciences Center, Institute of Biotechnology. The research was supported by the Research Council of Lithuania.

Academic supervisor – Dr. Giedrius Gasiūnas (Vilnius University, Natural Sciences, Biochemistry – N 004, 2017 10 01 – 2021 05 21.
Academic supervisor – Dr. Tautvydas Karvelis (Vilnius University, Natural Sciences, Biochemistry – N 004, 2021 05 19 – 2021 09 30.
Academic consultant – Prof. Dr. Virginijus Šikšnys (Vilnius University, Natural Sciences, Biochemistry – N 004).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – Prof. Dr. Edita Sužiedelienė (Vilnius University, Natural Sciences, Biochemistry – N 004).
Members:
Dr. Darius Balčiūnas (Vilnius University, Natural Sciences, Biology – N 010),
Prof. Dr. Matthias Bochtler (Institute of Biochemistry and Biophysics, Poland, Natural Sciences, Biochemistry – N 004),
Prof. Dr. Saulius Serva (Vilnius University, Natural Sciences, Biochemistry – N 004),
Dr. Remigijus Skirgaila (Thermo Fisher Scientific Baltic, Natural Sciences, Biochemistry – N 004).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 2 pm on 23 June 2022 in meeting room R 401 of the Vilnius University Life Sciences Center.
Address: Saulėtekio Ave 7, R 401, Vilnius, Lithuania
Tel. +37061424332; e-mail: greta.bigelyte@bti.vu.lt

The text of this dissertation can be accessed at the library of Vilnius University, as well as on the website of Vilnius University:
www.vu.lt/lt/naujienos/ivykiu-kalendorius

https://doi.org/10.15388/vu.thesis.315
https://orcid.org/0000-0002-1432-1433

VILNIAUS UNIVERSITETAS

Greta Bigelytė

Mažosios CRISPR-Cas nukleazės: nuo charakterizavimo iki pritaikymo

DAKTARO DISERTACIJA

Gamtos mokslai,
Biochemija (N 004)

VILNIUS 2022

Disertacija rengta 2017–2021 metais Vilniaus universiteto Gyvybės mokslų centro Biotechnologijos institute.
Mokslinius tyrimus rėmė Lietuvos mokslo taryba.

Moksliniai vadovai:
Dr. Giedrius Gasiūnas (Vilniaus universitetas, gamtos mokslai, biochemija – N 004). Nuo 2017 10 01 iki 2021 05 21.
Dr. Tautvydas Karvelis (Vilniaus universitetas, gamtos mokslai, biochemija – N 004). Nuo 2021 05 19 iki 2021 09 30).
Mokslinis konsultantas – prof. dr. Virginijus Šikšnys (Vilniaus universitetas, gamtos mokslai, biochemija – N 004).

Gynimo taryba:
Pirmininkė – prof. dr. Edita Sužiedėlienė (Vilniaus universitetas, gamtos mokslai, biochemija – N 004).
Nariai:
dr. Darius Balčiūnas (Vilniaus universitetas, gamtos mokslai, biologija – N 010);
prof. dr. Matthias Bochtler (Biochemijos ir biofizikos institutas, Lenkija, gamtos mokslai, biochemija – N 004);
prof. dr. Saulius Serva (Vilniaus universitetas, gamtos mokslai, biochemija – N 004);
dr. Remigijus Skirgaila (Thermo Fisher Scientific Baltic, gamtos mokslai, biochemija – N 004).

Disertacija ginama viešame Gynimo tarybos posėdyje 2022 m. birželio mėn. 23 d. 14.00 val. Vilniaus universiteto Gyvybės mokslų centro R-401 auditorijoje ir nuotoliniu būdu. Adresas: Saulėtekio al. 7, R 401 auditorija, Vilnius, Lietuva, tel. +37061424332 ; el. paštas greta.bigelyte@bti.vu.lt.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: *https://www.vu.lt/naujienos/ivykiu-kalendorius*

CONTENTS

ABBREVIATIONS

As	*Acidibacillus sulfuroxidans*
Au	*Aureabacteria* bacterium
AAV	adeno-associated virus
AHT	anhydrotetracycline
BH	bridge helix
bp	base pair
CARF	CRISPR-associated Rossmann fold
Cas	CRISPR-associated
Cn	*Clostridium novyi*
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DMEM	Dulbecco's Modified Eagle Medium
DNase	deoxyribonuclease
ds	double-stranded
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FLL	full length linear
gRNA	guide RNA
HDR	homology-directed recombination
HDV	hepatitis delta virus
HEPN	higher eukaryotes and prokaryotes nucleotide-binding
His	histidine
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth/Luria-Bertani
MBP	maltose-binding protein
Mi	*Micrarchaeota* archaeon
NHEJ	non-homologous end joining
nt	nucleotide
NTS	non-target strand
NUC	nuclease lobe
Nuc	nuclease domain
OC	open circle
OligoA	oligo-adenylate
PAGE	polyacrylamide gel electrophoresis
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PFM	position frequency matrix
PFS	protospacer flanking site

PI	PAM interacting domain
PMSF	phenylmethylsulfonyl fluoride
pre-crRNA	precursor CRISPR RNA
Pt	*Parageobacillus thermoglucosidasius*
REC	recognition lobe
RNase	ribonuclease
RNP	ribonucleoprotein
Ru	*Ruminococcus*
RuvC	RuvC nuclease domain
SC	supercoiled
SDS	sodium dodecyl sulfate
Sp	*Syntrophomonas palmitatica*
ss	single-stranded
TAM	transposon-associated motif
TBE	tris-borate-EDTA
TEV	tobacco etch virus
tracrRNA	*trans*-activating crRNA
Tris	tris(hydroxymethyl)aminomethane
TS	target strand
Un	uncultured archaeon
WED	wedge domain

INTRODUCTION

Prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems are found in ~85% archaea and ~40% bacteria (Makarova et al. 2020). These systems provide adaptive immunity against foreign nucleic acids by memorizing the invader through the insertion of nucleic acid fragments into the CRISPR array. CRISPR array transcription generates crRNAs that guide Cas protein effector complex for foreign nucleic acid degradation in the next round of infection (Barrangou et al. 2007; Mojica et al. 2005). Theoretically, Cas nucleases guided by RNA molecules can be directed to cleave any target sequence and this ability was utilized to adopt these nucleases as tools for genome editing (Cong et al. 2013; Deltcheva et al. 2011; Gasiunas et al. 2012; Jinek et al. 2012). However, the size of commonly used CRISPR-Cas9 and Cas12a proteins remains one of the biggest obstacles limiting cellular delivery by adeno-associated viruses – a common delivery vehicle in the clinic (Dong, Fan, and Frizzell 1996; D. Wang, Zhang, and Gao 2020). Furthermore, due to the development of next-generation base and prime editing tools based on Cas protein fusion to additional effector proteins, such as base deaminases and reverse transcriptases, respectively, Cas proteins of smaller size are highly desirable (Anzalone et al. 2019; Anzalone, Koblan, and Liu 2020; Komor et al. 2016). Therefore, recently discovered Cas12f nucleases, which are half the size of Cas9 and Cas12a, could provide an attractive alternative in these applications (Harrington et al. 2018).

The major subject of this PhD thesis – novel miniature Cas12f nucleases from recently identified class 2 type V-F CRISPR-Cas systems. **The goal** was to test whether Cas12f effectors can cleave the dsDNA and can be adopted as genome editing tools. To achieve this goal, the following **objectives** were set:

1. To test the ability of miniature Cas12f nucleases to cleave dsDNA.
2. To identify Cas12f bound guide RNA molecules.
3. Reconstitute and characterize the Cas12f complexes and their activity *in vitro*.
4. To examine the Cas12f ability to cleave genomic DNA in eukaryotic cells.

Scientific novelty and practical value:

This study is focused on the discovery and characterization of novel miniature Cas12f nucleases. With less than half the size of well-known Cas9 and Cas12a proteins, they are attractive candidates for genome editing applications providing the versatility for the delivery options, e.g., using

limited cargo vehicles such as adeno-associated viruses. Due to their small size, Cas12f proteins were predicted to be incapable of dsDNA cleavage (Shmakov et al. 2017). The first experimentally characterized Cas12f family protein (named Cas14) indirectly supported this hypothesis since it showed only ssDNA cleavage activity (Harrington et al. 2018). In this work, we showed for the first time that Cas12f proteins (including the Cas14) cleave dsDNA in the protospacer adjacent motif (PAM) dependent manner. Additionally, we demonstrated that two Cas12f1, *Acidibacillus sulfuroxidans* (As) and *Syntrophomonas palmitatica* (Sp), showed robust plasmid DNA interference activity in heterologous *E. coli*. AsCas12f1 and SpCas12f1 were selected for further biochemical characterization that revealed common features shared by Cas12f1 nucleases: i) compact size compared to other Cas12 nucleases, ii) remarkably long tracrRNAs, iii) increased temperature (45–55 °C) requirement for efficient dsDNA target binding and cleavage, iv) collateral non-specific ssDNA degradation upon target binding, v) dimerization upon binding of a single copy of guide RNA (gRNA). These unique properties bring additional flexibility to the CRISPR-Cas toolbox. Temperature-dependent dsDNA target recognition and collateral ssDNA nuclease activity may be advantageous in nucleic acid detection platforms in "one-pot" reactions combining both isothermal amplification and Cas12-based detection (Chen et al. 2018; Gootenberg et al. 2018; Joung et al. 2020). Moreover, temperature sensitivity may be used to precisely regulate activity at dsDNA targets in plants shown to tolerate or be acclimated to elevated temperatures, ultimately, reducing the potential for off-target effects and allowing activity to be controlled in a spatial-temporal fashion without the need for inducible promoters (Barone et al. 2020; Khattri, Nandy, and Srivastava 2011; Lin, Roberts, and Key 1984; Liu and Charng 2012; Nandy et al. 2019; Silva-Correia et al. 2014; N. Wang et al. 2020; Q. Wang et al. 2020). Altogether, these findings pave the way for the development of miniature Cas12f-based genome editing tools.

The major findings presented for defense in this thesis:

1. Miniature Cas12f effectors are capable to recognize and cleave dsDNA targets in a PAM-dependent manner.
2. Cas12f proteins form the complex with crRNA and tracrRNA that can be fused into a single guide (gRNA) to simplify the complex assembly.
3. Cas12f and gRNA form a 2:1 effector complex which cleaves target DNA and exhibits collateral ssDNA cleavage activity.

4. Efficient DNA binding and cleavage activities are manifested at higher temperatures, which could be implemented for nuclease activity regulation.
5. Miniature Cas12f nucleases can be successfully delivered and used to alter genomic DNA sequences in eukaryotic cells.

1. LITERATURE OVERVIEW

Some of the most common genetic disorders come from mutations in a single gene. While more than 5000 such monogenic diseases are known, their affect at least 250 million individuals globally (Doudna 2020). Mutations that cause the disease have to be corrected in order to cure genetic disorders. Hence, one of the main goals of today's personalised medicine is efficient gene manipulation through precise sequence corrections. Specific DNA sequence alterations could sufficiently disrupt or restore the function of an essential gene. Up to date, the main strategy of such gene editing remains the recruitment of intracellular DNA repair mechanisms to the intended DNA cleavage sites. However, the challenge for this is the site-specific DNA targeting. Well-known enzymes, such as zinc-finger nucleases (ZFN) or transcription activator-like effector nucleases (TALEN), rely on protein to DNA interactions (Brouns et al. 2008; Garneau et al. 2010). Thus, with every change of the target sequence, a new protein construct is needed. Meanwhile, CRISPR-Cas nucleases are guided by short RNA sequences that recognize the target DNA through Watson-Crick base pairing (Gasiunas et al. 2012; Jinek et al. 2012). Hence, they became perfect candidates for sequence-specific gene editing tools.

Due to the easy programmability, CRISPR-Cas effectors can be guided to cleave DNA at any site of need. Following introduced specific double-strand breaks of Cas nucleases and repair mechanisms in the cells, DNA sequence modifications are established with non-homologous end joining (NHEJ) or homology-directed recombination (HDR) (Figure 1.1) (Ciccia and Elledge 2010; Wyman and Kanaar 2006; Yeh, Richardson, and Corn 2019). Imprecise NHEJ base repair introduces deletions or short insertions of several nucleotides at the cleavage site (Lieber 2010; van Overbeek et al. 2016). The most common result of this repair mechanism – frameshift mutation causing premature stop codon emergence following with complete gene inactivation (Doench et al. 2014; Shalem et al. 2014; Wang et al. 2014). The introduction of two DSB at the same time in a cell can lead to bigger deletions, inversions, or even chromosomal translocations (Kosicki, Tomberg, and Bradley 2018; Maddalo et al. 2014; Mani and Chinnaiyan 2010). If a DNA repair template is present, gene replacement, point mutations, precise insertion, deletion, or integration of gene-sized DNA fragments can be achieved through HDR repair mechanism (Jasin and Rothstein 2013; Rouet, Smih, and Jasin 1994).

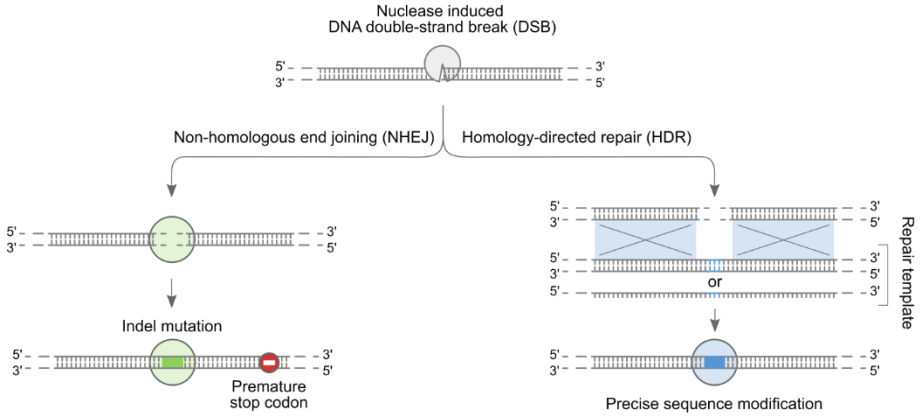


Figure 1.1. Schematic representation of DNA double-strand break (DSB) and its repair mechanisms. Induced by nuclease cleavage, DSB is typically repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR) mechanisms. The first one is error-prone, resulting from DSB ends processing and re-joining by endogenous repair machinery. This repair pathway usually results in short random indels, which may lead to premature stop codon formation and complete gene inactivation. Contrary to NHEJ, HDR can produce precise sequence modifications with homologous single-stranded or double-stranded DNA repair templates.

CRISPR-Cas nucleases were successfully applied for gene manipulations in variety of different heterologous hosts. However, it faces limitations, which encourage further research. Hence, this literature overview presents the latest CRISPR-Cas related information, including detailed biochemical characteristics of Cas12 proteins, as well as a variety of their application possibilities.

1.1. CRISPR-Cas

The ongoing molecular arms race between prokaryotes and infecting phages resulted in the rise of numerous elaborate defense mechanisms. Hence, a considerable fraction of prokaryotic genomes (up to 10%) is dedicated to their own defense (Koonin, Makarova, and Wolf 2017). One of these mechanisms – adaptive immunity system called CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated). Since discovered just a few decades ago, CRISPR-Cas became one of the most exciting scientific topics in microbiology (Barrangou et al. 2007).

In brief, the CRISPR-Cas locus consists of a CRISPR array and adjacent genes coding Cas proteins (Figure 1.2). While Cas proteins proceed a set of enzymatic activities, the CRISPR array stores the memory of the past

infections in unique spacer sequences integrated between CRISPR identical repeats (Barrangou et al. 2007; Mojica et al. 2005; Pourcel, Salvignol, and Vergnaud 2005). Therefore, upon reinfection CRISPR array provides guide RNAs for Cas nucleases to degrade specific foreign nucleic acids.

Overall, the CRISPR-Cas mechanism can be divided into three major phases: adaptation, expression, and interference (Figure 1.2). In the first phase, foreign DNA fragments (commonly referred to as protospacers) are selected, processed, and integrated by adaptation module (usually by Cas1-Cas2 complex found in most, but not all CRISPR-Cas systems) into the host's CRISPR array (Jackson et al. 2017; Makarova et al. 2020). Spacer acquisition preferentially occurs at the start of the CRISPR array, which keeps a chronological record of previous encounters (Barrangou et al. 2007; Pourcel et al. 2005). In the second phase, guide RNA molecules are produced. First, CRISPR array is transcribed into a long precursor CRISPR RNA (pre-crRNA) that is further processed by Cas proteins or, in some cases, cellular RNases into mature crRNA molecules (Deltcheva et al. 2011; Hochstrasser and Doudna 2015). Each crRNA contains a single spacer (or guide) sequence flanked by fragments of the CRISPR repeat sequence. For some types of CRISPR-Cas systems for crRNA maturation *trans*-activating CRISPR RNA (tracrRNA) with partial complementarity to repeat sequence is needed (Deltcheva et al. 2011; Karvelis et al. 2013). Such crRNA or crRNA and tracrRNA duplexes are later recognized and bound by Cas proteins. In the last, interference phase, active complexes of RNA and Cas are formed, following Cas nucleases guided by mature crRNAs to cleave complementary sequences of the foreign nucleic acids. Diverse CRISPR-Cas effector complexes can cleave DNA, RNA, or both DNA and RNA targets during the interference phase (Abudayyeh et al. 2016; Gasiunas et al. 2012; Jinek et al. 2012; Tamulaitis, Venclovas, and Siksnys 2017). Furthermore, in some types of CRISPR-Cas systems, complementary base pairing with a specific target unleashes so-called *trans* or collateral activity, resulting in non-specific non-target nucleic acid degradation (Chen et al. 2018; Kazlauskiene et al. 2017; Niewoehner et al. 2017). This also may lead to cell death limiting viral replication and providing protection to the overall bacterial population.

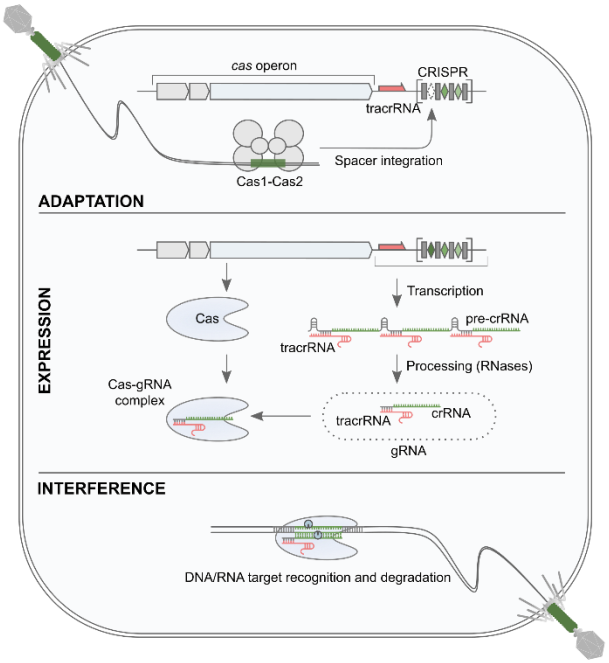


Figure 1.2. CRISPR-Cas adaptive immunity. The mechanism of CRISPR-Cas response to foreign nucleic acid hazard can be divided into three phases: adaptation, expression, and interference. In the first phase, foreign genetic elements are acquired by the Cas1-Cas2 adaptation module and integrated into the CRISPR array. Second – Cas proteins and RNAs from the CRISPR array and in some cases tracrRNA are expressed. CRISPR array transcription results into pre-crRNA, which is further processed by Cas proteins or other RNases. The mature crRNA, tracrRNA (in some cases) and Cas proteins form effector complexes. In the final stage, named interference, the effector complex recognizes and cleaves foreign genetic element complementary to crRNA, leading to the so-called immunity effect. crRNA – CRISPR RNA, pre-crRNA – precursor CRISPR RNA, tracrRNA (pink) – *trans*-activating CRISPR RNA, gRNA – guide RNA, which could be just crRNA or crRNA and tracrRNA duplex. Green diamonds in the CRISPR array and crRNA sequence elements represent spacers.

1.2. Classification

According to the most recent classification, CRISPR-Cas systems can be divided into 2 classes, 6 types and 33 subtypes (Makarova et al. 2020). The main difference between these two classes is the complexity of the effector proteins participating in the interference phase (Figure 1.3A). While in class 2 systems only single effector protein is needed, in class 1, multiple Cas proteins are required to form a complex capable of binding and cleaving or degrading target DNA or RNA molecules (Figure 1.3A) (Koonin, Makarova,

and Zhang 2017; Shmakov et al. 2017). However, more differences between CRISPR-Cas systems classes and types can be seen when comparing all functional modules found in CRISPR-Cas loci (Figure 1.3B) (Makarova et al. 2020; Makarova, Wolf, and Koonin 2013). These functional modules are based on the phase of the CRISPR-Cas mechanism where the indicated proteins participate. Starting with adaptation module, it comprises Cas1 – the main enzyme for new spacers integration found in most of known CRISPR-Cas systems, and other proteins, specific to the type (such as Cas2, Cas4, Csn2, and reverse transcriptase (RT)) (Amitai and Sorek 2016; Barrangou et al. 2007; Heler et al. 2015; Koonin et al. 2017; Plagens et al. 2012; Wei, Terns, and Terns 2015; Wright and Doudna 2016; Zhang, Kasciukovic, and White 2012).

Meanwhile, expression module's proteins are responsible for the pre-crRNA processing stage (Figure 1.3B). While in class 1 Cas6 is in control of crRNA maturation in all subtypes, this function in class 2 seems more dispersed (Özcan et al. 2019; Sternberg, Haurwitz, and Doudna 2012). RNA processing in type II CRISPR-Cas systems is catalyzed by ancillary bacterial RNase III enzyme while many type V and VI CRISPR-Cas large effector proteins contain separate catalytic sites responsible for crRNA maturation (Chylinski et al. 2014; Deltcheva et al. 2011; East-Seletsky et al. 2016; Fonfara et al. 2016; Liu, Li, et al. 2017).

The most extensively characterized is the interference phase. Target recognition and nucleic acid cleavage are performed by the interference module's Cas proteins (Figure 1.3B). To be specific, Cas3, Cas5-Cas8, Cas10, and Cas11 representing class 1 form effector complexes differently depending on the subtype. In class 2 type II, V, and VI systems the interference is carried out respectively by single Cas9, Cas12, or Cas13 protein (Makarova et al. 2020; Shmakov et al. 2017). Theoretically, the interference phase is based on sequence complementarity between the Cas bound crRNA and the target sequence which could result in cleavage of the CRISPR array in the host genome itself. Consequently, nearly all characterized CRISPR-Cas systems have evolved supplementary sequence origin mechanism which involves recognition of a short sequence, called the protospacer adjacent motif (PAM) or the protospacer flanking site (PFS), if Cas nuclease target is RNA molecule as for Cas13 (East-Seletsky et al. 2016; Nishimasu et al. 2017; Sternberg et al. 2014; Yamano et al. 2017). PAM recognition acts as the first step towards target recognition and cleavage resulting in local DNA unwinding following its cleavage by effector protein (Anders et al. 2014; Sternberg et al. 2014; Szczelkun et al. 2014).

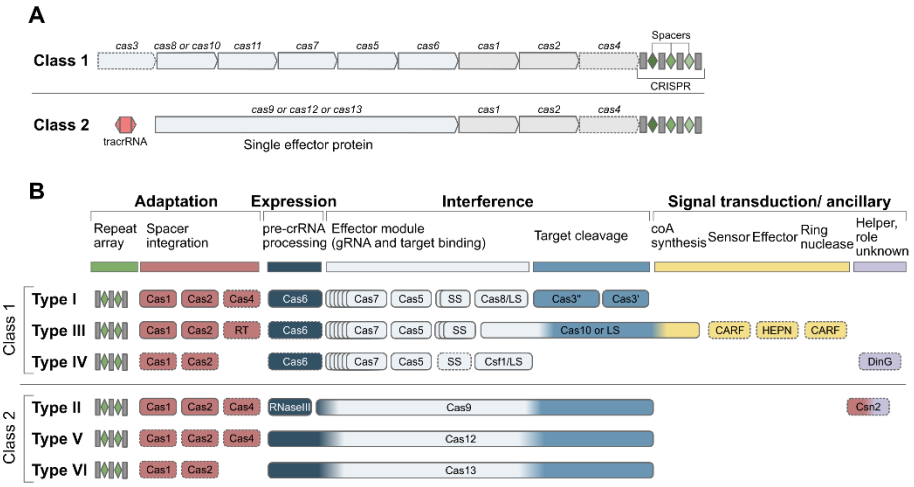


Figure 1.3. Current classification of CRISPR-Cas systems. (**A**) Schematic representation of the organization of class 1 and 2 CRISPR-Cas loci. The main difference separating these two classes is the effector module complexity (genes marked in light blue). Class 1 CRISPR-Cas effector consists of multiple Cas proteins, which together with crRNA binds and processes target nucleic acids. However, in class 2 systems functionally analogous effector compose of single, multidomain Cas protein. The other Cas proteins (light grey) are mainly involved in the adaptation process. CRISPR array is shown in grey rectangles of repeat sequences interrupted by spacer sequences marked in green diamonds. In some of class 2 CRISPR-Cas systems, for active effector complex to form additional tracrRNA (*trans*-activating CRISPR RNA) is needed, which is shown here in bright pink color. (**B**) Further classification of class 1 and 2 into separate 6 CRISPR-Cas types provides their generic loci with assigned functions for each element. Colored gene elements correspond to the same color-marked functions. Protein names follow current nomenclature and doted outlines represent dispensable (and/or missing, in some subtypes and variants) components. Cas9, Cas10, Cas12, and Cas13 shown in three colors reflect their contribution to different stages of the CRISPR-Cas mechanism. The CRISPR-associated Rossmann fold (CARF) and higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domain proteins are the most common sensors and effectors, respectively, in the type III CRISPR-Cas ancillary modules. Ring nucleases cleave cyclic oligoA produced by Cas10 and thus control the indiscriminate RNase activity of the HEPN domain of Csm6 (or Csx1). LS – large subunit, SS – small subunit. Adapted from (Makarova et al. 2020).

Lastly, all CRISPR-linked subsidiary genes are assigned to the signal transduction module. While little is known, Cas10 of type III was shown to synthesize cyclic oligoA, which activates Csm6 (or Csx1) RNase activity in higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domain by binding to CRISPR-associated Rossmann fold (CARF domain) (Kazlauskiene et al. 2017; Niewoehner et al. 2017).

1.2.1.Class 2 CRISPR-Cas systems

Regarding the simplicity offered by the single protein effector complex, class 2 systems are the most widely adapted CRISPR-Cas systems in biotechnological applications. Class 2 includes 3 types and 17 subtypes (Makarova et al. 2020) (Figure 1.4). Hence, the differences in Cas protein size in this class and their multifunctionality noting their DNA and RNA nuclease activities dictate numerous possible applications.

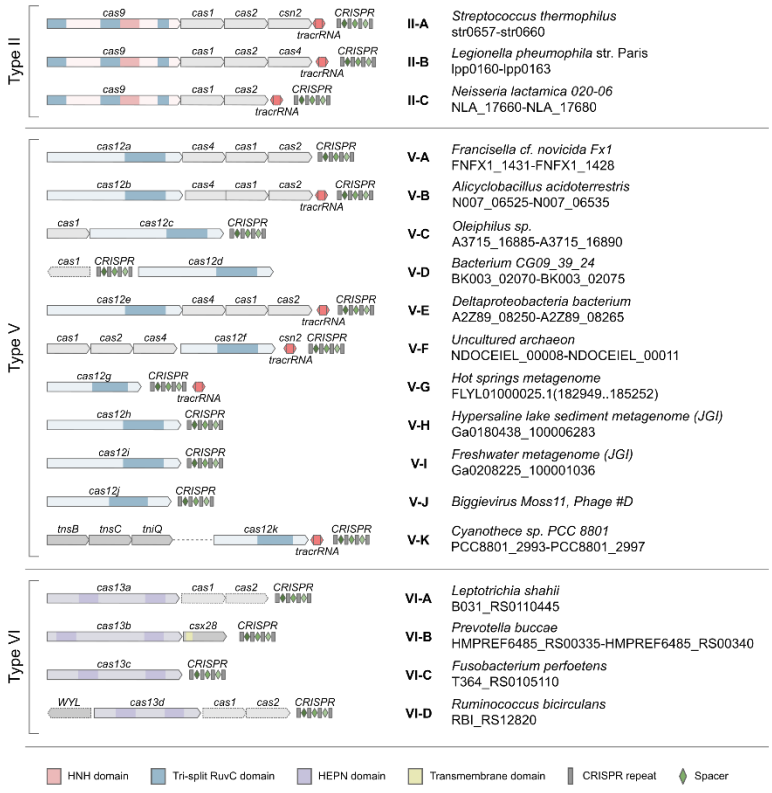


Figure 1.4. Current class 2 CRISPR-Cas classification. This figure schematically represents CRISPR-Cas loci for each class 2 subtype with its distinct variant. The column on the right indicates the organism and the corresponding gene range of the chosen variant. Homologous *cas9*, *cas12*, and *cas13* genes are color-coded: light pink, blue and purple, respectively. All other genes are shown in grey since they are not the main object of this study. Dashed lines represent the inconsistent appearance of the marked gene in the specific subtype. Characteristic nuclease domains, tracrRNA (*trans*-activating CRISPR RNA), and CRISPR array are shown as indicated in the legend at the bottom of this figure. Adapted from (Makarova et al. 2020).

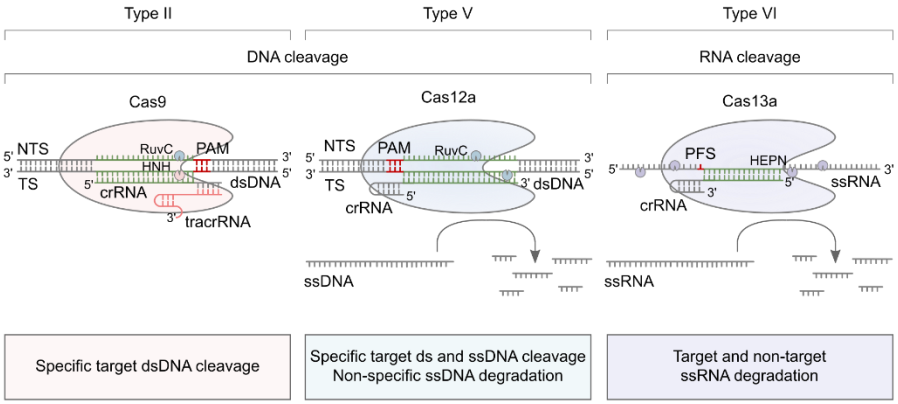


Figure 1.5. Schematic representation of CRISPR-Cas class 2 CRISPR Cas effector complexes. Type II Cas9 nuclease forms effector complex with tracrRNA (pink) and crRNA (grey with the spacer marked in green) which recognizes dsDNA target (green) proximal to a protospacer adjacent motif PAM (red) sequence. HNH (darker pink) and RuvC (darker blue) nuclease domains cleave both dsDNA strands forming blunt ends. Type V Cas12a nuclease forms effector complex with crRNA (grey with the spacer marked in green) which recognizes dsDNA target (green) proximal to a PAM (red) sequence. However, Cas12a can cleave targets containing ssDNA molecules as well. In both cases, correct base pairing with crRNA spacer sequence activates RuvC nuclease domain (darker blue), resulting in specific ssDNA and dsDNA cleavage. Furthermore, *cis* cleavage of target DNA induces collateral activity on non-target ssDNA. The last class 2 nuclease protein – Cas13 (here Cas13a) represents type VI CRISPR Cas systems. Effector complex of Cas13 nuclease and its crRNA (grey with the spacer marked in green) *cis* cleaves target (green) and protospacer flanking site (PFS) containing ssRNA molecules by two HEPN (darker purple) nuclease domains. As Cas12a, Cas13a also executes *trans* cleavage activity, but on non-target ssRNA molecules. TS – target strand, NTS – non-target strand.

Characteristic class 2 effectors – Cas9, Cas12, and Cas13, fundamentally differ from each other by the domain organization (Makarova et al. 2020). Cas9 contains two nuclease domains, a RuvC-like nuclease domain with the HNH nuclease inserted inside (Chen and Doudna 2017). Each of them is responsible for the cleavage of one strand of the target DNA at the same site (Gasiunas et al. 2012; Jinek et al. 2012). Guided by a dual RNA substrate, comprised of crRNA and tracrRNA molecules, Cas9 introduces blunt double-stranded breaks (Figure 1.5). On the contrary, Cas12 cleaves both DNA strands using single RuvC nuclease domain and forming staggered DNA ends at cleavage site. Furthermore, the most characterized of type V, Cas12a proteins require just a single crRNA molecule to fulfil this activity (Fonfara et al. 2016; Zetsche et al. 2015). Moreover, both Cas9 and Cas12 proteins require PAM sequences adjacent to the dsDNA target, although its position

differs. Cas9 proteins recognize and interact with PAM sequences at the 3'end of the target sequence while Cas12 – 5' end (Figure 1.5) (Abudayyeh et al. 2016; Gasiunas et al. 2012; Jinek et al. 2012; Zetsche et al. 2015). Interestingly, Cas12 efficient cleavage of the specific target also induces robust nonspecific ssDNA degradation, which was seen with almost all of the characterized Cas12 variants (Chen et al. 2018; Yan et al. 2019). Furthermore, some of Cas12 also contains ribonuclease domain that is responsible for pre-crRNA processing into mature crRNA (Fonfara et al. 2016). Consequently, Cas12 excels Cas9 in the diversity of identified subtypes, enzymatic activities, and the sizes of Cas proteins themselves (Figure 1.4 and 1.5).

Lastly, the specifically cleaving only RNA molecules, Cas13 hold two HEPN (higher eukaryotes and prokaryotes nucleotide-binding) domains (Shmakov et al. 2015). While Cas9 and Cas12 target invading specific DNA sequences, Cas13 potentially targets transcripts of foreign DNA genomes, which also induces collateral RNase activity resulting in infected bacteria dormancy (Figure 1.5) (Abudayyeh et al. 2016; East-Seletsky et al. 2016; Meeske, Nakandakari-Higa, and Marraffini 2019; Smargon et al. 2017).

1.3. Type V CRISPR-Cas systems

1.3.1.Cas12 diversity

Type V currently is the most diverse type of class 2 CRISPR-Cas systems (Figure 1.4). The type V CRISPR-Cas systems are further subdivided into A-K subtypes (Table 1.1) (Makarova et al. 2020). While most effectors of these subtypes can cleave dsDNA, some are capable for ssDNA or ssRNA cleavage (Chen et al. 2018; Harrington et al. 2018; S.-Y. Li et al. 2018; Yan et al. 2019).

The first discovered nuclease from type V CRISPR-Cas systems was Cas12a (formerly known as Cpf1; V-A subtype) from *Francisella tularensis*, *Acidaminococcus,* and *Lachnospiraceae* (Zetsche et al. 2015). Cas12a is a single RNA-guided endonuclease lacking tracrRNA (contrary to well-known Cas9 and other Cas12 nucleases) which utilizes a T-rich protospacer-adjacent motif (PAM) and cleaves DNA via a staggered DNA double-stranded break (Fonfara et al. 2016; Zetsche et al. 2015). Firstly, it was assumed that to cleave both, non-target and complementary to guide RNA (gRNA) target, strands Cas12a uses RuvC and another putative nuclease domain (Yamano et al. 2016). However, later it was clarified RuvC being the single and only nuclease domain needed for dsDNA cleavage (Swarts, van der Oost, and Jinek 2017). Thus, it was determined as the defining characteristic of the type V CRISPR-Cas systems (Makarova et al. 2015, 2020). Moreover, it was demonstrated that

Table 1.1 Properties of type V Cas12 effectors (adapted from (Tong et al. 2020)).

Subtype	Signature protein	Length (aa)	PAM	tracrRNA	Processing pre-crRNA	Target substrates	*Trans*-cleavage	References
V-A	Cas12a (Cpf1)	1,200-1,500	(T)TTV[1]	-	+	dsDNA/ssDNA	ssDNA	(Zetsche et al. 2015)
V-B	Cas12b (C2c1)	~1,300	TTN	tracrRNA	-	dsDNA/ssDNA	ssDNA	(Shmakov et al. 2015)
V-C	Cas12c (C2c3)	1,200-1,300	TG/TN	tracrRNA	+	dsDNA/ssDNA	ssDNA	(Shmakov et al. 2015)
V-D	Cas12d (CasY)	~1,200	TA/TG	tracrRNA	?	dsDNA/ssDNA	ssDNA	(Burstein et al. 2017)
V-E	Cas12e (CasX)	~1,000	TTCN	tracrRNA	?	dsDNA	ssDNA	(Burstein et al. 2017)
V-F	Cas12f (Cas14)	400-700	-	tracrRNA	-	ssDNA	ssDNA	(Harrington et al. 2018)
V-G	Cas12g	~800	-	tracrRNA	-	ssRNA	ssRNA/ssDNA	(Yan et al. 2019)
V-H	Cas12h	~900	RTR[2]	-	-	dsDNA/ssDNA	ssDNA	(Yan et al. 2019)
V-I	Cas12i	~1,100	TTN	-	+	dsDNA/ssDNA	ssDNA	(Yan et al. 2019)
V-J	Cas12j (CasΦ)	~750	TBN[3]	-	+	dsDNA/ssDNA	ssDNA	(Pausch et al. 2020)
V-K	Cas12k (C2c5)	~650	GTN	tracrRNA	-	dsDNA	-	(Strecker, Ladha, et al. 2019)

[1]V = A, C and G; [2]R = A and G; [3]B = C, G and T.

Cas12a is active in human cells and can be adopted for genome editing (Zetsche et al. 2015).

Up to date, 10 more type V subtypes were identified and characterized: V-B and V-C (previously named C2c1 and C2c3 respectively) (Shmakov et al. 2015), V-D and V-E (firstly identified as CasY and CasX respectively) (Burstein et al. 2017), V-F (previously identified as Cas14) (Harrington et al. 2018), V-G, V-H and V-I (Yan et al. 2019), V-J (also known as CasΦ) (Pausch et al. 2020) and finally, V-K (Strecker, Ladha, et al. 2019) (Table 1.1).

1.3.2.Cas12 enzymatic activities

Most of the characterized Cas12 are RNA-guided dsDNA nucleases. While the plethora of them is identified up to date, the most characterized in detail are Cas12a effector proteins. After the binary complex of the effector protein and its guide RNA (gRNA) is formed, it specifically recognizes the 5' T-rich PAM sequences and promotes DNA unwinding following Watson-Crick base pairing of target strand with the guide sequence of gRNA (Stella, Alcón, and Montoya 2017). Simultaneously non-target strand of target DNA is displaced and forms an R-loop structure. Both strands are cleaved by a single RuvC nuclease domain at PAM distal site forming staggered ends with 5'overhangs (Stella et al. 2017; Swarts et al. 2017; Zetsche et al. 2015). Although Cas12a initially cleaves DNA target at a single site, it also catalyzes trimming of the cleaved target ends *in vitro*, following production of heterogeneous cleavage products (Jones et al. 2021; Stephenson, Raper, and Suo 2018).

Not just Cas12a, but Cas12b-e and Cas12h-j were proven biochemically to cleave dsDNA targets (Liu et al. 2019; Pausch et al. 2020; Shmakov et al. 2015; Yan et al. 2019). However, the activity diverges significantly in some cases. For example, the cleavage efficiency of Cas12i is different for each strand of the dsDNA target, where the non-target (non-complementary to gRNA) strand is cleaved fast, but very low efficiency of complementary strand cleavage is observed (Yan et al. 2019). Potentially in respect of this characteristic, Cas12i could be used as a nickase in large scale gene fragment manipulation, for specific base editing in combination with well-characterized deaminases (Gaudelli et al. 2017; Komor et al. 2016; Nishida et al. 2016).

Most type V CRISPR effectors can also cleave ssDNA in the PAM-independent manner if it contains complementary to the guide RNA target sequence. According to Harrington et al., it was thought that Cas12f – effector protein of type V-F CRISPR-Cas systems, in complex with gRNA can cleave only ssDNA (Harrington et al. 2018, 2018; S.-Y. Li et al. 2018).

Even further, most of the Cas12 display two distinct DNA cleavage modes: *cis* – specific activity against ssDNA and dsDNA targets, and *trans* – non-target ssDNA cleavage. Following the *cis* cleavage of target DNA, the RuvC domain is still active and can access ssDNA substrates performing *trans* cleavage, also known as collateral activity (Harrington et al. 2018; S.-Y. Li et al. 2018). Sequence independent DNase activity is triggered by gRNA hybridization to a complementary target DNA strand (Swarts and Jinek 2019).

One of the exceptions – the type V-K CRISPR Cas system. Type V-K locus is associated with Tn7-like transposons (Figure 1.4) (Strecker, Ladha, et al. 2019). However, the RuvC nuclease site of Cas12k is naturally inactivated and Cas12k acts as an RNA-directed recognition tool for following transposition events. According to performed structural and biochemical experimentation, RNA-directed target selection by Cas12k primes TnsC polymerization and DNA remodelling, while following TnsB interaction with TnsC trigger its disassembly and catalyze site-specific transposon insertion (Querques et al. 2021; Xiao, Wang, et al. 2021). Nonetheless, Cas12k replacement with Cas9 completely abolished transposition, indicating Cas12k and transposase complex specific activation effect.

Another exception is V-G, identified recently by Yan et al (Yan et al. 2019). Even though Cas12g, the signature protein of these systems, contains a RuvC-like endonuclease domain, no dsDNA or ssDNA cleavage activity was detected neither *in vivo* nor *in vitro*. Nonetheless, in the presence of its crRNA and tracrRNA, cleavage of target ssRNA was observed. Additionally, *trans* cleavage activity on non-specific ssRNA was also seen for Cas12g (Yan et al. 2019). Even though Cas13, the effector protein of type VI CRISPR-Cas system, also proceeds target RNA cleavage, this activity is dependent on two HEPN domains (Liu, Li, et al. 2017). In the Cas12g case, inactivation of the RuvC domain completely abolished its activity on RNA targets, proving to be the key component for RNA and DNA cleavage for type V CRISPR-Cas effectors (Yan et al. 2019). Hence, about 700 aa Cas12g could become a potential Cas13 smaller alternative.

1.3.3.DNA cleavage model

In general, Cas12 proteins have bilobed structures containing N–terminal recognition (REC) and C-terminal nuclease (NUC) lobes connected by the wedge (WED) domain (Dong et al. 2016; Gao et al. 2016; Liu et al. 2019; Stella et al. 2017; Swarts et al. 2017; Yamano et al. 2016; H. Yang et al. 2016). REC lobe is mainly composed of 2, sometimes 3, REC domains, however, the NUC lobe is much more diverse: RuvC nuclease domain for DNA cleavage,

PI – PAM interacting domain, NUC – nuclease domain, which is important for DNA positioning and, finally, bridge helix (BH) domain (Figure 1.6A) (Swarts and Jinek 2019). REC lobe stabilizes gRNA and target DNA hybrid in R-loop formation, while domains in the NUC lobe are responsible for PAM recognition and DNA cleavage (Gao et al. 2016; Liu et al. 2019; Stella et al. 2017; Swarts and Jinek 2019).

Since most type V systems effectors cleave DNA in *cis* and *trans* modes, the full DNA cleavage cycle of Cas12a effectors was proposed recently (Figure 1.6B) (Swarts and Jinek 2019). The first key finding was supported by the obtained structure of Cas12a bound to a gRNA and PAM-lacking ssDNA target. Together with supplementary biochemical experiments, it revealed that gRNA and target DNA strand duplex formation induced conformational rearrangement allosterically activating the RuvC domain (Figure 1.6B) (Swarts and Jinek 2019). In this scheme, PAM binding does not participate in the catalytic activation of Cas12a. Nevertheless, PAM is crucial for target recognition, DNA strand separation, and wherefore base pairing of the target strand complementary to the seed segment of the gRNA following complete R-loop formation (Jeon et al. 2018, 2018; Strohkendl et al. 2018; Zetsche et al. 2015). The second structure of Cas12a complex with gRNA and dsDNA target presented the order in which the DNA strands are cleaved with the first non-target strand (NTS) cleavage and, following that, target strand (TS) cleavage (Figure 1.6B) (Swarts and Jinek 2019). gRNA and DNA TS hybridization lead to REC domain displacement, resulting in exposed RuvC nuclease center and, therefore, allosteric activation of Cas12a nuclease. Meanwhile, NTS is oriented toward RuvC active site by PI domain and positively charged structurally conserved NTS-binding groove (Stella et al. 2018, 2017; Swarts and Jinek 2019; Swarts et al. 2017). NTS cleavage enables further unwinding of dsDNA target PAM-distal end by fraying, which results in TS displacement closer to RuvC nuclease site (Figure 1.6B) (Swarts and Jinek 2019). Furthermore, after target DNA cleavage, the PAM-distal end is released, while the PAM-proximal end remains bound to Cas12a and gRNA complex (Singh et al. 2018; Swarts and Jinek 2019). With first product exclusion, the RuvC domain is exposed for *trans* cleavage activity on any ssDNA (Figure 1.6B) (Jeon et al. 2018; Singh et al. 2018; Swarts and Jinek 2019). The same RuvC nuclease domain performs both, *cis* and *trans* DNA cleavage (Chen et al. 2018; S.-Y. Li et al. 2018). With PAM-proximal end release, the new cycle begins, starting with gRNA (crRNA, in Cas12a case) processing and effector complex assembly (Figure 1.6B) (Fonfara et al. 2016; Zetsche et al. 2017). Nevertheless, Cas12a is seen as a single turnover enzyme

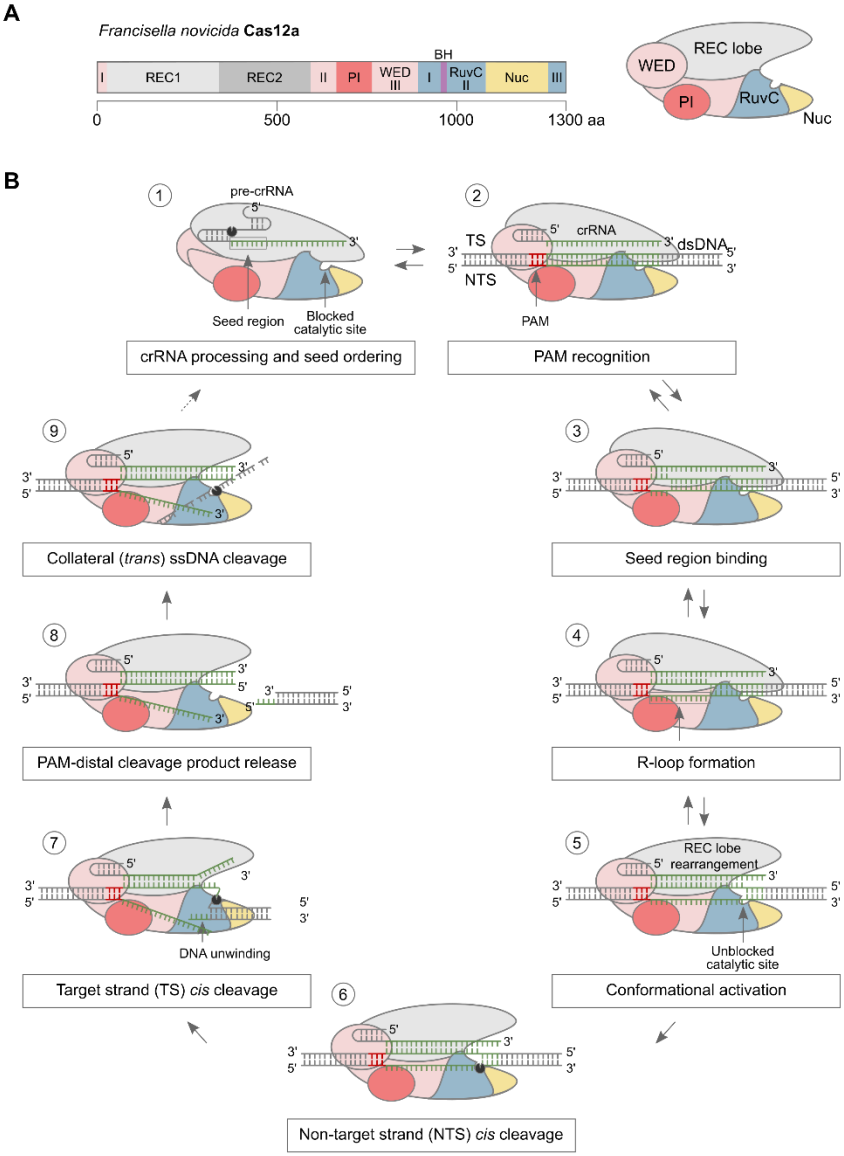


Figure 1.6. Cas12a domain organization and DNA cleavage model. (**A**) Schematic diagram of the domain organization of Cas12a effector from *Francisella novicida* type V-A CRISPR-Cas system. Here BH – bridge helix, Nuc – nuclease, PI – PAM interacting, REC – recognition, WED – wedge, and RuvC – RuvC nuclease domain. (**B**) Schematic model of Cas12a mediated nuclease activities. 1 – Cas12a processes the 5' end of pre-crRNA to obtain the matured crRNA. crRNA and Cas12a form an effector complex with pre-ordered 1-5 nt of the crRNA seed sequence. 2 – dsDNA target recognition starts with PAM sequence detection by the WED and PI domains. If correct PAM is present, DNA unwinding begins. 3 – the R-loop formation begins with DNA target strand (TS) hybridization to the pre-ordered crRNA seed segment. 4

– the continuation of dsDNA target unwinding results in complete R-loop formation. 5 – R-loop and crRNA-TS hybrid induces conformational changes within Cas12a REC lobe, which results in allosteric unblocking of RuvC nuclease catalytic site. 6 – now completely displaced non-target strand (NTS) is positioned towards RuvC catalytic site, where the first *cis* cleavage is performed. 7 – further unwinding of PAM distal DNA duplex allows TS to enter the RuvC catalytic site, where second *cis* cleavage is performed. 8 – while Cas12a-crRNA keeps bound to DNA target PAM-proximal end, the PAM-distal end is released. 9 – Cas12a-crRNA and PAM-proximal DNA complex preserve catalytically activated conformation, resulting in *trans* cleavage of non-specific ssDNA. Under PAM-proximal end release, the new cycle begins. Adapted and modified from (Swarts and Jinek 2019).

for *cis* cleavage, due to substrate dissociation rate being orders of magnitude slower than DNA cleavage, and multi turnover for *trans* activity with rapid non-target ssDNA degradation (Chen et al. 2018; Singh et al. 2018; Strohkendl et al. 2018).

Both, dsDNA or PAM lacking ssDNA target, binding to Cas12a and its gRNA complex equally initiate cleavage of non-specific ssDNA (Chen et al. 2018). Following this, PAM recognition is not required for the allosteric activation of the RuvC nuclease domain. Though PAM is obligatory for cleavage of fully duplexed dsDNA targets, it is irrelevant if guide pre-unwound dsDNA is used (Swarts and Jinek 2019). Additionally, Cas12a is considered to be more specific than Cas9, with reference to lower tolerance for mismatches between TS and seed sequence of gRNA, which could relate to single RuvC domain usage for both, *cis* and *trans* DNase activities (Strohkendl et al. 2018; Swarts and Jinek 2019).

Cis and *trans* DNase (or RNase) activities are validated for quite a few Cas12a orthologs, including most of all other type V effectors (except for naturally inactive Cas12k) (Chen et al. 2018; Harrington et al. 2018, 2020; Liu, Chen, et al. 2017; Pausch et al. 2020; Yan et al. 2019). Thus, this mechanism could be proposed as versatile throughout most type V CRISPR-Cas systems.

1.4. Origins of class 2 CRISPR-Cas systems

The origin of the CRISPR-Cas systems always raised strong interest in the CRISPR community. With the increasing variety of discovered novel subtypes, the almost complete scenario of CRISPR-Cas evolution was presented (Makarova et al. 2020). In most cases, effector proteins seem to originate from different mobile genetic elements. Nevertheless, the adaptive module seems to be repeatedly recruited into different types of CRISPR-Cas systems (Koonin and Makarova 2017, 2019; Koonin et al. 2017). Cas1, the

integrase responsible for the spacer incorporation into CRISPR arrays, together with CRISPR repeats themselves seems to originate from casposons – transposons, which employ Cas1 homologs as transposases (Figure 1.7A) (Koonin and Krupovic 2015; Krupovic et al. 2014; Krupovic, Béguin, and Koonin 2017). Some of the known casposons encode nucleases homologous to Cas4, adaptation module element seen in some CRISPR-Cas subtypes, however, none of them encode broadly found Cas2 (Krupovic et al. 2014; Krupovic and Koonin 2016). The ancestor of Cas2 is unclear, however, it is thought that Cas2 could supposedly originate from a yet unknown toxin-antitoxin system (Figure 1.7A) (Makarova et al. 2020; Shmakov et al. 2015). Finally, the ancestral CRISPR repeats and the leader sequence usually observed next to it could evolve from either TRs – terminal repeats, or duplicated target sites (TSD) of the casposon (Figure 1.7A) (Koonin and Makarova 2019; Krupovic et al. 2017).

Regardless, the current and most progressed subject – the evolution of effector proteins of class 2 CRISPR-Cas systems. Effectors of type II and V systems seem to have evolved from TnpB nucleases on multiple independent occasions (Makarova et al. 2020; Shmakov et al. 2017). TnpB nucleases are found encoded in the IS200/IS605 transposons (IS – insertion sequences). The *tnpB* genes are among the most abundant genes in bacterial and archaeal genomes and are frequently encoded together with transposases by autonomous transposons, but in some cases – as single transposon element (Bao and Jurka 2013). Regardless of its high distribution, TnpB is not essential but is thought to participate in the regulation of transposase-mediated "peel and paste" transposition (Pasternak et al. 2013). In most cases the similarity between TnpB and type II and V effector proteins cover the N-terminal RuvC-like domain and the arginine-rich motif, thought to be needed for efficient nucleic acid binding (Chylinski et al. 2014).

Regarding representative HNH nuclease domain inserted into the RuvC like domain, type II systems evolved from a diverse variety of TnpB, which are now denoted IscB (ISC – insertion sequences Cas9-like) (Figure 1.7B) (Kapitonov, Makarova, and Koonin 2015). This family of TnpB-like proteins is particularly abundant in *Cyanobacteria* (Chylinski et al. 2014). According to Kapitonov, Makarova & Koonin, ISC transposons evolved from IS200/IS605 family transposons via possible insertion of a mobile group II intron encoding the HNH nuclease domain (Kapitonov et al. 2015).

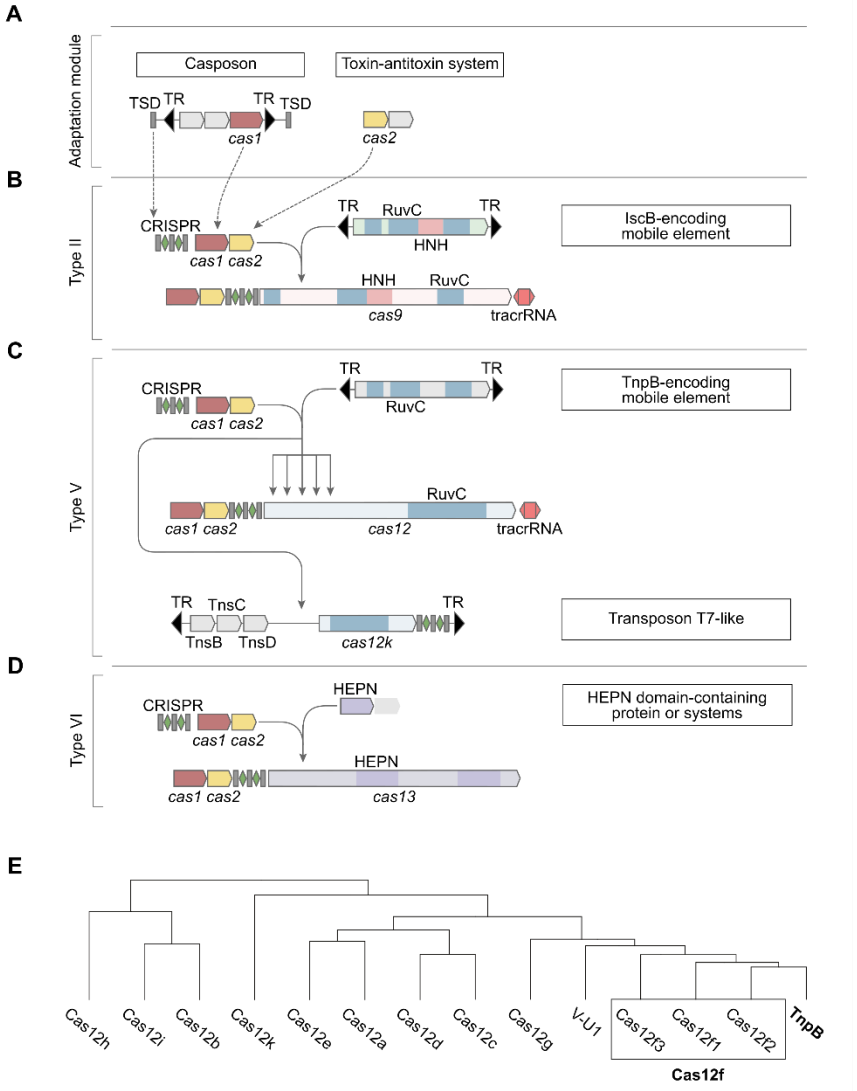


Figure 1.7. The origins of class 2 CRISPR-Cas systems. This figure shows a hypothetical scenario of the origin of class 2 CRISPR-Cas systems. (**A**) In this scenario, the adaptation module composed of Cas1 and CRISPR repeats derived from a casposon, and Cas2 potentially derived from the toxin-antitoxin system. All three types appear to have evolved independently: type II and different subtypes of type V by the recruitment of distinct TnpB nucleases ((**B**) IscB (insertion sequences Cas9-like) for type II and (**C**) TnpB for type V) that are encoded by IS200/IS605-like transposable elements, while (**D**) type VI likely originated from an RNA cleaving, HEPN domain-containing abortive infection or toxin-antitoxin system. (**C**) However, one of type V, subtype K, seems to be recruited by mobile genetic elements and lost their interference capacity along with the original defense function. Homologous genes are color-coded and identified by a family name following the previous

classification (Figure 1.3). The multiforking arrows denote events that have been inferred to have occurred on multiple, independent occasions during the evolution of CRISPR–Cas systems. TR – terminal repeats, tracrRNA – *trans*-activating CRISPR RNA, TSD – target site duplication, the likely source of ancestral CRISPR repeats. (**E**) Type V effectors sequences comparison shows Cas12f being the most related to TnpB proteins. U1 – uncharacterized subtype yet. Adapted from (Makarova et al. 2020).

In contrast, the effectors of each subtype of type V, with distinct insertions in RuvC coding regions and unrelated N-terminal regions, seem to evolve independently from different TnpB proteins (Figure 1.7C) (Shmakov et al. 2015). With a broad range of Cas12 effectors from size similar to ones of TnpB proteins to double or triple in size as seen for subtypes V-A and B, the newly discovered miniature Cas12f proteins could potentially represent intermediate stages for independent emergence of bigger and more complex type V CRISPR-Cas variants (Shmakov et al. 2017). Supplementary, Cas12f of all type V CRISPR effectors was seen to be the most closely related to TnpB proteins by sequence (Figure 1.7E) (Makarova et al. 2020). Different subtype V-F variants seem to originate from different groups of TnpB, however, due to their highly significant sequence similarity, currently they are classified within a single subtype. Hence, type V-F detailed characterization would contribute significantly to the more specific evolutionary pathway of CRISPR-Cas systems explanation.

Additionally, the new evolutionary pathway is seen with inactivated Cas effectors hijacked by mobile genetic elements as seen in subtype V-K (Figure 1.7C) (Faure et al. 2019). Here, Cas12k acts as a target recognition and delivery factor for site-specific transposition (Querques et al. 2021; Saito et al. 2021; Strecker, Ladha, et al. 2019; Xiao, Wang, et al. 2021).

Since less is known about Cas13, the effector of type VI CRISPR-Cas systems, it is thought to originate from the HEPN-containing component of an abortive infection (Abi), toxin-antitoxin system (Figure 1.7D) (Makarova et al. 2020). It was shown that similarly to Abi, bacteria containing Cas13a demonstrated growth arrest under phage infection (Meeske et al. 2019). However, another possible type VI evolutionary pathway could have started with the recruitment of one of already HEPN-containing Cas proteins, such as Csm6 and Csx1 found in class 1 CRISPR-Cas systems (Shmakov et al. 2017).

1.5. Cas12 based molecular tools

Cas12a and Cas12b nucleases have been successfully applied for gene editing in plenty of different mammalian, plant, and bacteria cells (Kim et al. 2016;

Kleinstiver, Pattanayak, et al. 2016; Kleinstiver, Tsai, et al. 2016; Strecker, Jones, et al. 2019; Tang et al. 2017, 2018; Tong et al. 2020; Wu et al. 2020; Zetsche et al. 2015, 2017). However, the original genome editing strategy using Cas proteins is based on double-stranded break (DSB) and cell repair mechanisms, which is not highly efficient, may cause mutations in non-target DNA sites, and result in uncontrollable indels (Figure 1.8A). Hence, genome editing technologies, bypassing the need for DSB, were created. In general, Cas12 proteins, partially or fully lacking nuclease activity, are fused to additional domains to form the base, prime, transposase editors, or transcriptional regulators. Furthermore, Cas12 nuclease activities were employed for nucleic acid detection. Thus, all these molecular tools are discussed below.

1.5.1.Base editing

Precise editing of a single nucleotide is one of the gene manipulations challenges there CRISPR also made a difference. Base editors introduce point mutations without requiring double-strand break, DNA template, or efficient HDR (Gaudelli et al. 2017; Komor et al. 2016; Komor, Badran, and Liu 2017; Nishida et al. 2016). Point mutations in the RuvC nuclease domain deliver inactivated Cas12 proteins (dCas12). Two additional protein fusions to dCas12a – rat cytosine deaminase APOBEC and uracil DNA glycosylase (UDG) inhibitor (UGI) provide efficient cytosine base editing machinery (Li et al. 2012). dCas12a recognizes and binds to target DNA resulting in R-loop formation (Stella et al. 2017; Swarts et al. 2017). Following, pushed out non-complementary ssDNA is altered by APOBEC – cytosine (C) at the 8-13 nt position downstream of the PAM is deaminated to uracil (U) and with following DNA replication C-G base pair is changed to A-T (Figure 1.8B) (Kleinstiver et al. 2019). Additionally, UGI can inhibit base excision repair by UDG increasing the mutation rate (Banno et al. 2018). Additionally, adenine base editors were also easily adapted to dCas12a for A-T reversion to G-C base (Gaudelli et al. 2020; Richter et al. 2020). Interestingly, it was shown that in comparison to dCas9, dCas12a mediated base editing gives higher editing efficiency and lower off-target activity (X. Li et al. 2018).

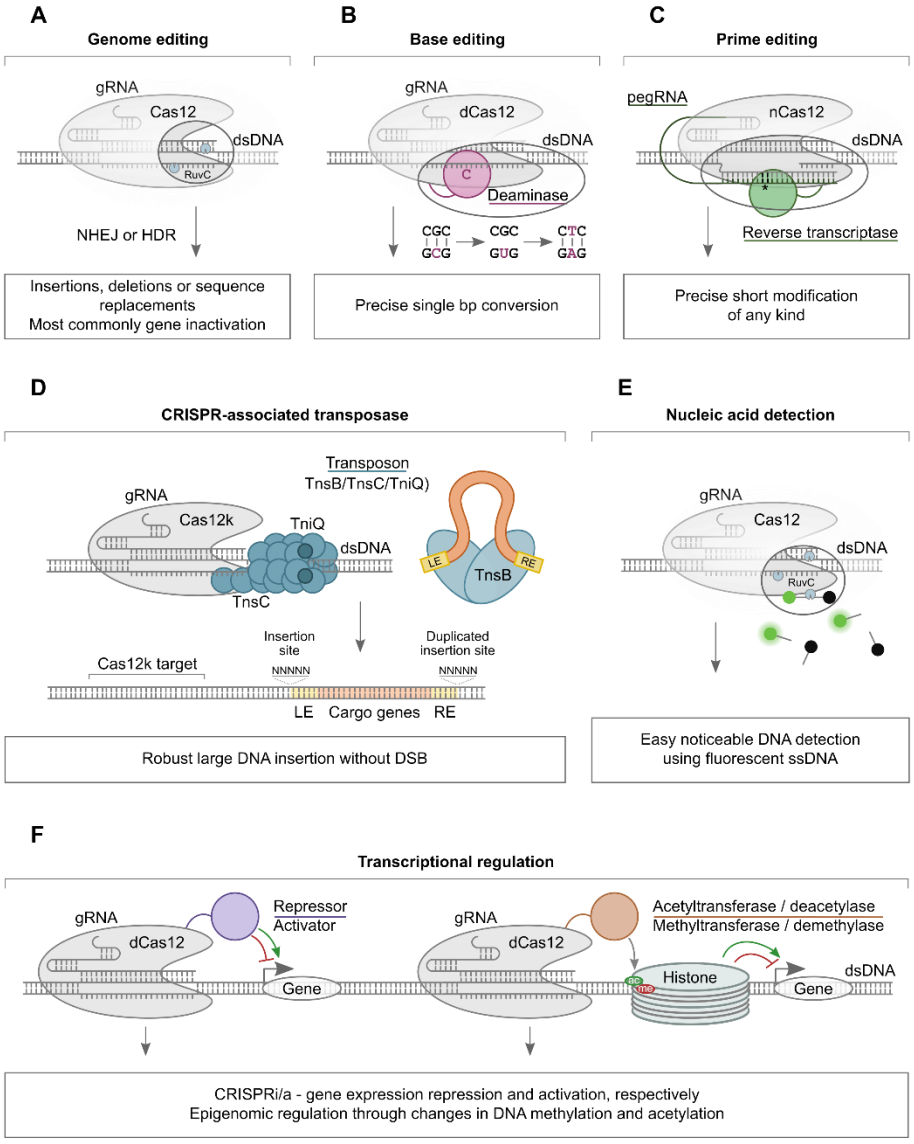


Figure 1.8. Cas12 application capabilities. (**A**) Genome editing, (**B**) base editing, (**C**) prime editing, (**D**) CRISPR-associated transposase, (**E**) nucleic acid detection, and (**F**) transcriptional gene regulation – well-validated CRISPR-Cas12 employment fields. RNA-guided Cas12 effector complexes target specific genomic sites or individual DNA molecules. (**A**) In the first case, genome editing is triggered by cellular DNA repair by non-homologous end joining (NHEJ) or homology-directed repair (HDR) after DNA cleavage by Cas12. The obtained result – random indel formation and gene inactivation in most cases. If specific single bp conversion is needed, the base editing (**B**) or prime editing (**C**) techniques are used. The latter can be used for any kind of short sequence modification. (**B**) Cas12 lacking nuclease activity (dCas12) acts as a guiding element to deliver deaminase to the area of interest.

In the process of cytidine (C) deamination, the conversion to uracil (U) and later to thymine (T) is attained. Following DNA replication, the whole base-pair swap from C-G to T-A is achieved. (**C**) Contrary, in prime editing specific pegRNA (prime editing guide RNA) with complement to target 5' end is used, which introduces sequence modification through single DNA strand break (n – nickase) and reverse transcription processes. (**D**) Recently discovered type Cas12k effector is associated with Tn7-like transposon system. As result, Cas12k and its guide RNA (gRNA) directed large DNA insertion can be achieved by transposon effectors (TnsC, TnsB, and TniQ). TnsC formed filament around the DNA recruits TnsB transposase carrying cargo gene of interest, following complete DNA insertion with transposition site duplication. CRISPR-associated transposase mechanism scheme was adapted from (Querques et al. 2021). (**E**) A unique characteristic of Cas12 to *trans* cleave ssDNA is used for nucleic acid detection. Here, fluorescent ssDNA provides an easily detectable signal even at the attomolar level of target DNA. (**F**) And lastly, dCas12 ability to deliver different active domains to specific DNA sites is also used for gene expression regulation at the transcription stage. Different repressors or activators (left scheme) may reduce or increase the transcription by blocking or attracting RNA polymerase to the gene promoter site. dCas12 fused with enzymes that can modify epigenetic markers of DNA or histones (right scheme) can also be used to change gene expression status.

1.5.2.Prime editing

Even though base editors can introduce transition point mutations (C-G to A-T or A-T to G-C), transversion point mutations (C-G to G-C or T-A to A-T), precise insertions or precise deletions cannot. Furthermore, deamination can occur in multiple positions if target nucleotides surround the position of interest. Thence, new, even more precise, prime editing technology was presented recently. Prime editors consist of nicking Cas protein, engineered reverse transcriptase, and specific long prime editing guide RNA (pegRNA) (Figure 1.8C) (Anzalone et al. 2019). pegRNA codes the target site for complex delivery at one end and the desired edit on the other end. After cleavage of one target DNA strand, the obtained ssDNA cleavage product acts as a primer for reverse transcription using the extension in the pegRNA as a template. The attained modified DNA strand is incorporated back into the target site by cellular repair machinery and later used as a new template for second strand modification.

Although the first generation prime editing machinery was created using Cas9 nickases (with inactivated HNH nuclease domain, but active RuvC domain), the potential of Cas12 application in this field remains strong (Anzalone et al. 2019, 2020). The initial concern regarding prime editing was related with the production of Cas12 nickase. Here, a single RuvC nuclease domain is used for cleavage of both DNA strands. However, a recently published Cas12a_M3 variant with K1013G and R1014G point mutations

cleaves only the target strand (complementary to gRNA target sequence) (Paul et al. 2021). Future Cas12 application for prime editing technologies is highly expected.

1.5.3.CRISPR-associated transposases

While point or short sequence manipulations are covered by previously introduced methods, long sequence targeted integration in living cells has been a challenge to solve. Natural CRISPR-associated transposases and engineered Cas-domain fused transposase chimeric systems can integrate big genomic cargos *in vitro* and into bacterial genomes (Chen and Wang 2019; Klompe et al. 2019; Strecker, Ladha, et al. 2019). One of them is the type V-K CRISPR-Cas system associated with Tn7-like transposon (CAST). Cas12k effector with a naturally inactivated RuvC domain serves as target recognition module for RNA-guided transposition performed by transposase effectors – transposase TnsB, AAA+ ATPase TnsC, and zinc-finger TniQ (Strecker, Ladha, et al. 2019). According to the recent structural publications regarding this type of system, TnsC forms filament structure on DNA that is capped by TniQ, while Cas12k binds specific DNA targets in a guide RNA-dependent way (Park et al. 2021; Querques et al. 2021; Xiao, Wang, et al. 2021). Together with supplementary *in vivo* and biochemical experimentations the further mechanism was suggested where guide RNA-directed target selection by Cas12k initiates TnsC polymerization and DNA remodelling, which recruits TnsB to catalyze transposon insertion (Figure 1.8D) (Querques et al. 2021). TniQ is thought to act as a regulatory factor that restricts TnsC polymerization, while TnsB interacts directly with TnsC filament to induce their disassembly followed by transposon insertion. After TncB initiated TnsC disassembly, a single turn helix hexamer of TnsC is left on target DNA, which possibly could act as a molecular ruler defining the spacing between Cas12k binding and transposon insertion (about 60-66 bp downstream of Cas12k PAM sequence) (Querques et al. 2021). Site-specific transposon-assisted genome engineering (STAGE), based on CAST and created by Chen et al., was already successfully adapted for rapid and efficient gene disruption in *Pseudomonas aeruginosa* and *Klebsiella pneumoniae* (Chen et al. 2021). However, more examples of CAST applications can be expected in near future.

1.5.4. Nucleic acid detection

Despite Cas12's wide applicability in the gene editing field, these proteins have been also successfully implemented for DNA detection platforms (Chen et al. 2018; Gootenberg et al. 2018). Taking advantage of *trans*-cleavage activity, DETECTR (DNA endonuclease-targeted CRISPR-based *trans* reporter) was developed as a DNA detection and diagnostic platform (Chen et al. 2018). With implemented isothermal pre-amplification, DETECTOR was shown to accurately detect clinically relevant types of human papillomavirus. Here, the quenched fluorescent ssDNA provides a quick and easily noticeable reaction. If target DNA is present in the mixture the cleavage by Cas12 of these targets activates collateral activity following cleavage of ssDNA probes to release detectable fluorescent molecules (Figure 1.8E). This detection system has been used for discrimination of single nucleotide polymorphisms (Teng et al. 2019), genotyping (Harrington et al. 2018), and even virus detection, such as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Broughton et al. 2020).

1.5.5. Transcriptional regulation

Cas proteins can be used efficiently for transcriptional gene manipulations as well. Guided by gRNA, dCas12 can bind to the target site without proceeding any alterations of DNA in that site. Firstly, dCas12 can inhibit gene transcription by competing with RNA polymerase for promoter binding or by blocking polymerase migration through gene coding region, which is known as CRISPR interference (CRISPRi) (Kim et al. 2017; Qi et al. 2013; Tang et al. 2017; Zhang et al. 2017). Also, CRISPR could act as recruiting machinery for repressor domains like Krüppel associated box (KRAB) to promoter regions in mammalian cells to inhibit expression (Gilbert et al. 2013). Moreover, it can deliver transcriptional activators, or RNA polymerase subunits to a specific site where it could potentially promote sufficient gene expression. This transcriptional activation is called CRISPRa (Kleinstiver et al. 2019; Tak et al. 2017). Together, CRISPRi and CRISPRa are useful molecular tools for gene expression regulation (Figure 1.8F left scheme). Even further, Cas12a RNase activity provides convenient multifunctionality characteristics – Cas12a can process a CRISPR array for multiple target locus (Fonfara et al. 2016; Swarts et al. 2017). This was successfully achieved with dCas12a and four targets locus CRISPR array which efficiently inhibited all four genes in similar to that of repressing the separated genes (Zhang et al.

2017). Potentially different crRNA for the same target gene for further increased repression rate could also be incorporated.

dCas12a has also been successfully used for epigenetic editing of endogenous genes by fusing to histone or DNA modifying enzymes, which could influence chromatin organization and gene expression (Figure 1.8F right scheme). For instance, histone acetyltransferase core domain of p300 fused to catalytically inactive dCas12a from *Lachnospiraceae bacterium* ND2006 successfully modified histone H3 lysine27 and increased target gene expression in different types of human cells (Zhang et al. 2018). Fusions to DNA methyltransferase DNMT3A, which allows site-specific methylation of the CpG islands around the dCas12 target sequence, could result in repression of nearby genes, as seen with dCas9 (Amabile et al. 2016; McDonald et al. 2016; Vojta et al. 2016). dCas9 fusion to the catalytic domain of TET1 DNA demethylase leads to upregulation of gene expression, which potentially could also be adapted for dCas12 (Choudhury et al. 2016; Morita et al. 2016; Xu et al. 2016).

1.6. Limitations

Even though adaptation for genetic manipulations with Cas nucleases seems to have a plethora of possibilities, the disadvantages, such as target space limitations (related to specific PAM preferences), cleavage specificity, genome editing efficiency, restrict this process (Knott and Doudna 2018). The first one was successfully reduced by creating huge libraries of Cas orthologs with a variety of PAM sequences encompassing all possible sequence combinations (Gasiunas et al. 2020). However, the path of modification of already well-characterized Cas proteins was also taken. In this case, the work was mostly done on *Streptococcus pyogenes* Cas9 reaching even near PAMless SpRY variant which recognizes NRN (R = A or G) and to a lesser extent NYN (Y = T or C) PAMs (Anders, Bargsten, and Jinek 2016; Hu et al. 2018; Kleinstiver et al. 2015; Kleinstiver, Pattanayak, et al. 2016; Nishimasu et al. 2018; Walton et al. 2020). While less is done in broadening recognizable PAM sequence combinations with Cas12 proteins (Gao et al. 2017; Tóth et al. 2020; Tran et al. 2021), engineered variants with much higher efficiency of genome editing in human cells and reduce in off-target cleavage events were successfully obtained (Kleinstiver et al. 2019). The incomplete ability to discriminate on-target sequences from similar off-target sequences throughout the genome limits further applications *in vivo* (Fu et al. 2013). As a result, to validate the off-target activities and editing efficiency, robust methods for

targeting outcomes were created (Jones et al. 2021; Kim et al. 2019, 2020; Kim and Kim 2018; Tsai 2018).

Additionally, CRISPR-Cas application in the clinical environment may be affected by the potential for pre-existing antibodies against CRISPR machinery elements, which may cause inflammation, the immunogenicity of bacterially derived proteins, or unknown effects of genome editing outcomes in a long run (Doudna 2020). While none is known regarding the last concern, the theoretical immunogenicity of CRISPR-Cas proteins could be handled by reducing treatment to a one-time event or using different effectors. Pre-existing antibodies for some of the known Cas9 have been already detected in humans exposed to pathogenic bacteria with CRISPR systems, though it is unknown if the concentrations are sufficient to provoke an immune response (Charlesworth et al. 2019; Wagner et al. 2019).

The major obstacle left – efficient and specific delivery of CRISPR-Cas machinery to the cells of interest. The theoretical number of different options can be used, such as electroporation, transfection, direct injection, or using viral vectors, to deliver the DNA encoding effector complex, or RNAs encoding gRNA and mRNA of Cas protein, or preassembled RNPs to cells (Glass et al. 2018). However, most of the methods of delivery that are acceptable in laboratory conditions are not applicable in clinical settings, with viral adeno-associated virus (AAV) vector being the most used clinical delivery vehicle for gene therapy to date (D. Wang et al. 2020). AAV vectors are the only delivery vehicle approved to introduce CRISPR-base genome editing therapeutics into the human body. The reasons for AAV attractiveness are related to reduced risk of genomic integration, ease of delivery of its single-stranded DNA vector genome to various tissues and cell types, and clinically manageable immunogenicity (Doudna 2020). Nevertheless, AAV, with a diameter of ~25 nM, can only pack 4.7 kb DNA fragments when *Acidaminococcus sp*. BV3L6 Cas12a on its own occupies 3.9 kb (Dong et al. 1996). Following the limited packing size of these vectors, the large size of Cas nucleases together with their guide RNA, including additional functional domains or template sequence for HDR in some cases, become the great challenge of application, to begin with. Strategies like using a second AAV vector for guide RNAs and DNA templates can be used (Lau and Suh 2017; Ryu et al. 2018; D. Wang et al. 2020; Y. Yang et al. 2016). Also, a split AAV method was created where large Cas protein can be separated by split inteins. However, both of the AAVs with different parts of CRISPR machinery must enter the same cell, which reduces the overall efficiency (Chew et al. 2016; Fine et al. 2015; Li et al. 2008; Tornabene et al. 2019; Truong et al. 2015). Nonetheless, the main strategy is searching for minimal CRISPR-Cas systems

or/and smaller Cas proteins (Harrington et al. 2018; Kim et al. 2017; Pausch et al. 2020). Hence, novel miniature Cas nucleases have been characterized and described in this doctoral thesis.

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Chemicals

All chemicals used in this study were purchased from Sigma-Aldrich, Roth, Fluka, and Thermo Fisher Scientific Baltics of the highest purity grade available. Radiolabelled nucleotides were purchased from Perkin Elmer.

2.1.2. Commercial proteins and kits

T4 Polynucleotide kinase, FastAP Thermosensitive Alkaline Phosphatase, Phusion, DreamTaq and T4 DNA polymerases, T4 DNA ligase, "Fast Digest" restriction enzymes, Proteinase K, RiboLock RNase inhibitor, RNase A, and DNase I used in this study were obtained from Thermo Fisher Scientific.

"Rapid DNA Ligation Kit", "GeneJET PCR Purification Kit", "GeneJET Plasmid Miniprep Kit", "GeneJET Gel Extraction Kit", "T7 High Yield Transcription Kit" and "GeneJET RNA Cleanup and Concentration Kit" were purchased from Thermo Fisher Scientific Baltics, "TruSeq Small RNA Library Preparation Kit" – Illumina, "Synergy 2.0 Plant DNA Extraction Kit" – Ops Diagnostics, "Zymoclean Gel DNA Recovery Kit" – Zymo Research, "Monarch PCR Purification Column" – New England Biolabs.

All these products were used according to the manufacturer's instructions.

2.1.3. Bacterial strains

E. coli strain BL21(DE3) [F– ompT gal dcm lon hNDSB(rB- mB-) λ(DE3)] was used for Un1Cas12f1 expression.

E. coli strain Arctic Express (DE3) [F– ompT gal dcm tetr hsdS(rB- mB-) λ(DE3) endA Hte [cpn10 cpn60 Gentr] was used for SpCas12f1 and AsCas12f1 expression and plasmid interference assay.

E. coli strain DH5α [F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–] was used for the cloning procedures and target DNA multiplication.

E. coli bacteria were cultivated in LB medium (1 % peptone, 0.5 % yeast extract, 0.5 % NaCl in water, pH 7.0 (25 °C)) or on plated LB agar, which also contained 1.5 % agar-agar.

2.1.4.Cell lines

HEK293T – adherent cell line derived from human embryonic kidney cells grown in tissue culture and expressing SV40 Large T-antigen (ATCC catalog number CRL-3216). The cell line was cultivated using Dulbecco's Modified Eagle Medium (DMEM) supplied with 10% fetal bovine serum, penicillin (100 U/ml), and streptomycin (100 μg/ml) (Thermo Fisher Scientific).

PH1V69 – DuPont Pioneer inbred line used in plant genome editing assay. All plants used were greenhouse-grown by DuPont Pioneer (Corteva).

2.1.5.Proteins and nucleic acids

Listed below appendices provide information about:

- Appendix 1 – Cas12f proteins used in this study.
- Appendix 2 – their sequences.
- Appendix 3 – different constructed plasmids with Benchling links.
- Appendix 4 – Cas12f guide RNA molecules.
- Appendix 5 – oligonucleotides used for cleavage and binding assays.
- Appendix 6 – primers used in assessing human and maize cell editing.
- Appendix 7 – gRNA targets used in human and maize cell editing.

2.1.6.Buffers and other solutions

Detecting Cas12f dsDNA cleavage and PAM recognition:

Lysis buffer: 20 mM phosphate (pH 7.0 at 25 °C), 0.5 M NaCl, 2 mM PMSF 5% (v/v) glycerol.

Reaction buffer: 10 mM Tris-HCl (pH 7.5 at 37 °C), 100 mM NaCl, 1 mM DTT, and 10 mM $MgCl_2$.

Expression and purification of Cas12f1-RNA complexes and proteins

Un1Cas12f1

Loading buffer: 20 mM Tris-HCl (pH 8.0 at 25 °C), 1.5 M NaCl, 5 mM 2-mercaptoethanol, 10 mM imidazole, 2 mM PMSF, 5% (v/v) glycerol.

Elution buffer: 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.5 M NaCl, 5 mM 2-mercaptoethanol.

Storage buffer: 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.5 M NaCl, 2 mM DTT, 50% (v/v) glycerol.

SpCas12f1 and AsCas12f1 RNP complexes

Loading buffer: 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.25 M NaCl, 5 mM 2-mercaptoethanol, 25 mM imidazole, 2 mM PMSF, 5% (v/v) glycerol.

Elution buffer: 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.25 M NaCl, 5 mM 2-mercaptoethanol and 5% (v/v) glycerol.

Storage buffer: 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.25 M NaCl, 2 mM DTT and 50% (v/v) glycerol.

SpCas12f1 and AsCas12f1

Loading buffer: 20 mM Tris-HCl (pH 8.0 at 25 °C), 1.5 M NaCl, 5 mM 2-mercaptoethanol, 25 mM imidazole, 2 mM PMSF, 5% (v/v) glycerol.

Elution buffer: 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.5 M NaCl and 5 mM 2-mercaptoethanol.

Storage buffer: 20 mM Tris-HCl (pH 8.0 at 25 °C), 500 mM NaCl, 2 mM DTT and 50% (v/v) glycerol.

RNA purification from RNP complexes

Proteinase K reaction buffer: 10 mM Tris-HCl (pH 7.5 at 37 °C), 1 mM EDTA, 1 mM DTT, 100 mM NaCl and 5 mM $MgCl_2$.

DNA cleavage assays

Cas12f1-gRNA complex assembly buffer: 10 mM Tris-HCl (pH 7.5 at 37 °C), 100 mM NaCl, 1 mM EDTA, 1 mM DTT.

Plasmid DNA and M13 cleavage:

Un1Cas12f1 reaction buffer: 2.5 mM Tris-HCl (pH 7.5 at 37 °C), 25 mM NaCl, 0.25 mM DTT and 10 mM $MgCl_2$.

SpCas12f1 reaction buffer: 10 mM Tris-HCl (pH 7.5 at 37 °C), 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$.

AsCas12f1 reaction buffer: 10 mM Tris-HCl (pH 7.5 at 37 °C), 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$.

3× loading dye solution: 0.01% Bromophenol Blue, 0.03% SDS and 75 mM EDTA in 50% (v/v) glycerol.

Oligonucleotides cleavage:

Un1Cas12f1 reaction buffer: 5 mM Tris-HCl (pH 7.5 at 37 °C), 50 mM NaCl, 0.5 mM DTT and 5 mM $MgCl_2$.

SpCas12f1 reaction buffer: 10 mM Tris-HCl (pH 7.5 at 37 °C), 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$.

AsCas12f1 reaction buffer: 10 mM Tris-HCl (pH 7.5 at 37 °C), 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$.

Loading dye solution: 95% (v/v) formamide, 0.01% Bromophenol Blue and 25 mM EDTA.

DNA binding assay

Cas12f1-gRNA complex assembly buffer: 10 mM Tris-HCl (pH 7.5 at 37 °C), 100 mM NaCl, 1 mM EDTA, 1 mM DTT.

Binding buffer: 40 mM Tris-HAc (pH 8.4 at 25 °C), 1 mM EDTA, 0.1 mg/ml BSA, 10% (v/v) glycerol and 5 mM $Mg(C_2H_3O_2)_2$.

Molecular weight measurements by mass photometry

Cas12f1-gRNA complex assembly buffer: 10 mM Tris-HCl (pH 7.5 at 37 °C), 100 mM NaCl, 1 mM EDTA, 1 mM DTT.

Binding buffer: 40 mM Tris-HAc (pH 8.4 at 25 °C), 5 mM $Mg(C_2H_3O_2)_2$.

2.2. Methods

2.2.1. Engineering CRISPR-Cas12f systems for PAM identification assay

Selected CRISPR-Cas12f systems (Appendix 1) were modified to target the 7 bp randomized PAM library described previously (Karvelis et al. 2015). This was accomplished by replacing the native CRISPR array with three repeat:spacer:repeat units that encoded a spacer (33-39 nt depending on the average spacer length observed in the respective Cas12f system) capable of complementing to a sequence (anti-sense strand) immediately 3' of the region of PAM randomization. The engineered CRISPR-Cas12f systems were then synthesized (GenScript) and cloned into a modified pETDuet-1 (MilliporeSigma) or pACYC184 (New England Biolabs) plasmid. For the CRISPR-Cas12f1 system initially named Cas14a1 (Harrington et al. 2018) and renamed here as Un1Cas12f1, the pLBH531_MBP-Cas14a1 plasmid (gift from Jennifer Doudna, Addgene plasmid #112500) was used. The sequences of the Cas12f proteins are listed in Appendix 2 and links to the plasmid sequences encoding the Cas12f CRISPR systems engineered to target the PAM library are provided in Appendix 3.

2.2.2. Detecting Cas12f dsDNA cleavage and PAM recognition

Plasmid DNA targets were cleaved with Cas12f ribonucleoprotein (RNP) complexes produced from the modified locus or by combining *E. coli* lysate containing Un1Cas12f1 (Cas14a1) protein with T7 transcribed gRNA (20 nt spacer). First, *E. coli* DH5α or ArcticExpress (DE3) cells were transformed with CRISPR-Cas12f encoding plasmids (pACYC or pETDuet-1 and pLBH531, respectively) and cultures grown in LB broth (30 ml) supplemented with either chloramphenicol (25 µg/ml) (pACYC plasmids) or ampicillin (100

μg/ml) (pETDuet-1 and pLBH531 plasmids). Next, for plasmids with a T7 promoter (pETDuet-1 and pLBH531 plasmids), expression was induced with 0.5 mM IPTG when cultures reached OD_{600} of 0.5 and incubated overnight at 16 °C. Cells (from 10 ml) were collected by centrifugation and re-suspended in 1 ml of lysis buffer (20 mM phosphate (pH 7.0 at 25 °C), 0.5 M NaCl, 5% (v/v) glycerol)) supplemented with 10 μl PMSF (final conc. 2 mM) and lysed by sonication. Cell debris was removed by centrifugation. 10 μl of the obtained supernatant containing RNPs was used directly in the digestion experiments. For Un1Cas12f1, 20 μl of clarified supernatant was combined with 1 μl of RiboLock RNase Inhibitor (Thermo Fisher Scientific) and 2 μg of gRNA and allowed to complex with the clarified lysate as described below.

Cas12f RNP complexes were used to cleave either the 7 bp randomized PAM library or a plasmid containing a fixed PAM and gRNA target. Briefly, 10 μl of Cas12f-gRNA RNP containing lysate was mixed with 1 μg of PAM library or 1 μg of the plasmid containing a single PAM and gRNA target in 100 μl of reaction buffer (10 mM Tris-HCl (pH 7.5 at 37 °C), 100 mM NaCl, 1 mM DTT, and 10 mM $MgCl_2$). After 1 h incubation at 37 °C, DNA ends were repaired by adding 1 μl of T4 DNA polymerase (Thermo Fisher Scientific) and 1 μl of 10 mM dNTP mix (Thermo Fisher Scientific) and incubating the reaction for 20 min at 11 °C. The reaction was then inactivated by heating it up to 75 °C for 10 min and 3'-dA overhangs added by incubating the reaction mixture with 1 μl of DreamTaq polymerase (Thermo Fisher Scientific) and 1 μl of 10 mM dATP (Thermo Fisher Scientific) for 30 min at 72 °C. Additionally, RNA was removed by incubation for 15 min at 37 °C with 1 μl of RNase A (Thermo Fisher Scientific). Following purification with a GeneJET PCR Purification column (Thermo Fisher Scientific), the end-repaired cleavage products (100 ng) were ligated with a double-stranded DNA adapter containing a 3'-dT overhang (100 ng) for 1 h at 22 °C using T4 DNA ligase (Thermo Fisher Scientific). After ligation, cleavage products were PCR amplified appending sequences required for deep sequencing and subjected to Illumina sequencing (Karvelis et al. 2015; Karvelis, Young, and Siksnys 2019).

Double-stranded DNA target cleavage was evaluated by examining the unique junction generated by target cleavage and adapter ligation in deep sequence reads. This was accomplished by first generating a collection of sequences representing all possible outcomes of dsDNA cleavage and adapter ligation within the target region. For example, cleavage and adapter ligation at just after the 21st position of the target would produce the following sequence (5'-CCGCTCTTCCGATCTGCCGGCGACGTTGGGTCAACT-3') where the adapter and target sequences comprise 5'-

CCGCTCTTCCGATCT-3' and 5'-GCCGGCGACGTTGGGTCAACT-3', respectively. The frequency of the resulting sequences was then tabulated using a custom Perl script (provided at https://github.com/cortevaCRISPR/Cas12f-InformaticsTools.git) and compared to negative controls (experiments setup without functional Cas12f complexes) to identify target cleavage.

Evidence of PAM recognition was evaluated as described previously (Karvelis et al. 2015, 2019). Briefly, the sequence of the protospacer adapter ligation exhibiting an elevated frequency in the previous step was used in combination with a 10 bp sequence 5' of the 7 bp PAM region to identify reads that supported dsDNA cleavage. Once identified, the intervening 7 bp PAM sequence was isolated by trimming away the 5' and 3' flanking sequences using a custom Perl script (provided at https://github.com/cortevaCRISPR/Cas12f-InformaticsTools.git) and the frequency of the extracted PAM sequences normalized to the original PAM library to account for inherent biases using the following formula.

Normalized Frequency = (Treatment Frequency)/((Control Frequency)/(Average Control Frequency))

Following normalization, a position frequency matrix (PFM) (Stormo 2013) was calculated and compared to negative controls (experiments setup without functional Cas12f complexes) to look for biases in nucleotide composition as a function of PAM position. Biases were considered significant and indicative of PAM recognition if they deviated by more than 2.5-fold from the negative control. Analyses were limited to the top 10% most frequent PAMs to reduce the impact of background noise resulting from non-specific cleavage coming from other components in the *E. coli* cell lysate mixtures.

2.2.3.Modifications of CRISPR-Cas12f1 for guide RNA pull-down

Plasmid-borne SpCas12f1 and AsCas12f1 CRISPR systems described earlier were engineered to also encode a 10× histidine (His) and maltose-binding protein (MBP) tag fused to the N-terminus of the *cas12f1* gene. Additionally, a sequence encoding the tobacco etch virus (TEV) protease recognition sequence (ENLYFQS) was also included. The SpCas12f1 plasmid was digested with XagI and NcoI restriction enzymes (New England Biolabs) and the backbone was isolated by agarose gel purification (Qiagen). Next, a synthesized DNA fragment (Genscript) containing a 5' XagI restriction site, T7 promoter, lac operator, and ribozyme binding sequence in addition to the sequence encoding the 10×His:MBP:TEV tag followed by an inverted BbsI

site incorporating a sequence that upon digestion would yield a compatible NcoI overhang was digested with XagI and BbsI and column purified (Qiagen). The two purified fragments were then joined using T4 DNA ligase (New England Biolabs), transformed into One Shot TOP10 *E. coli* cells (Invitrogen), and constructs confirmed by Sanger sequencing. For, AsCas12f1 a similar strategy was used except XagI and AvaI restriction enzymes (New England Biolabs) were used. Links to the plasmid sequences (pMBP-SpCas12f1 and pMBP-AsCas12f1) are provided in Appendix 3.

2.2.4.Cas12f1 expression and purification

Un1Cas12f1

Un1Cas12f1 protein was expressed in *E. coli* BL21(DE3) strain from the pLBH531_MBP-Cas14a1 plasmid (gift from Jennifer Doudna, Addgene plasmid #112500). Un1Cas12f1^{D326A} and Un1Cas12f1^{D510A} expression plasmids were engineered from pLBH531 using Phusion Site-Directed Mutagenesis Kit (Thermo Fisher Scientific). *E. coli* cells were grown in LB broth supplemented with ampicillin (100 μg/ml) at 37 °C. After culturing to an OD_{600} of 0.6-0.8, the temperature was decreased to 16 °C, and expression was induced with 0.5 mM IPTG. After 16 h cells were pelleted, resuspended in loading buffer (20 mM Tris-HCl (pH 8.0 at 25 °C), 1.5 M NaCl, 5 mM 2-mercaptoethanol, 10 mM imidazole, 2 mM PMSF, 5% (v/v) glycerol) and disrupted by sonication. Cell debris was removed by centrifugation. The supernatant was loaded on the Ni^{2+}-charged HiTrap chelating HP column (GE Healthcare) and eluted with a linear gradient of increasing imidazole concentration (from 10 to 500 mM) in 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.5 M NaCl, 5 mM 2-mercaptoethanol buffer. The fractions containing Un1Cas12f1 variants were pooled and subsequently loaded on the HiTrap heparin HP column (GE Healthcare) for elution using a linear gradient of increasing NaCl concentration (from 0.1 to 1.5 M). The fractions containing the protein of interest were pooled and the 10×His:MBP:TEV tag was cleaved by overnight incubation with TEV protease at 4 °C. To remove the cleaved 10×His:MBP:TEV tag and TEV protease, reaction mixtures were loaded onto a HiTrap heparin HP 5 column (GE Healthcare) for elution using a linear gradient of increasing NaCl concentration (from 0.1 to 1.5 M). Next, the elution from the HiTrap heparin column was loaded on an MBPTrap column (GE Healthcare) and the Un1Cas12f1 proteins were collected in the flow-through. The collected fractions with Un1Cas12f1 were then dialyzed against 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.5 M NaCl, 2 mM DTT and 50% (v/v)

glycerol buffer and stored at -20 °C. The sequences of the Un1Cas12f1 proteins are listed in Appendix 2.

SpCas12f1 and AsCas12f1

To obtain Cas12f1-RNA complexes, pMBP-SpCas12f1, and pMBP-AsCas12f1 plasmid-borne CRISPR systems encoding both nuclease and guide RNA (gRNA) were transformed into *E. coli* cells (Arctic Express (DE3)). Cultures were grown in LB broth supplemented with ampicillin (100 μg/ml) and gentamicin (10 μg/ml) at 37 °C to an OD_{600} of 0.6-0.8. At which point, the temperature was decreased to 16 °C, and expression was induced with 0.5 mM IPTG. After 16 h, cells were pelleted, resuspended in loading buffer (20 mM Tris-HCl (pH 8.0 at 25 °C), 0.25 mM NaCl, 5 mM 2-mercaptoethanol, 25 mM imidazole, 2 mM PMSF, 5% (v/v) glycerol) and disrupted by sonication. After removing cell debris by centrifugation, the supernatant was loaded on Ni^{2+}-charged HiTrap chelating HP column (GE Healthcare) and eluted with a linear gradient of increasing imidazole concentration (from 25 to 500 mM) in 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.25 mM NaCl, 5 mM 2-mercaptoethanol and 5% (v/v) glycerol. The fractions with Cas12f1-RNA complexes were then dialyzed against 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.25 mM NaCl, 2 mM DTT and 50% (v/v) glycerol and stored at -20 °C.

SpCas12f1 and AsCas12f1 proteins (without gRNA) were also expressed and purified using pMBP-SpCas12f1 and pMBP-AsCas12f1. For experimentation requiring dead (d) or nuclease inactivated Cas12f1 protein, pMBP-SpCas12f1 and pMBP-AsCas12f1 were further modified introducing D228A and D225A encoding codons into SpCas12f1 and AsCas12f1 genes, respectively, using Phusion Site-Directed Mutagenesis Kit (Thermo Fisher Scientific). *E. coli* cells were grown in LB broth supplemented with ampicillin (100 μg/ml) and gentamicin (10 μg/ml) at 37 °C. After culturing to an OD_{600} of 0.6-0.8, the temperature was decreased to 16 °C, and protein expression was induced with 0.5 mM IPTG. After 16 h, cells were pelleted, resuspended in loading buffer (20 mM Tris-HCl (pH 8.0 at 25 °C), 1.5 M NaCl, 5 mM 2-mercaptoethanol, 25 mM imidazole, 2 mM PMSF, 5% (v/v) glycerol) and disrupted by sonication. Cell debris was removed by centrifugation. The supernatant was loaded on Ni^{2+}-charged HiTrap chelating HP column (GE Healthcare) and eluted with a linear gradient of increasing imidazole concentration (from 25 to 500 mM) in 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.5 M NaCl, and 5 mM 2-mercaptoethanol. The fractions containing Cas12f1 protein were pooled and subsequently loaded on the HiTrap heparin HP column (GE Healthcare). A linear gradient of increasing NaCl concentration (from 0.2 to 1.0 M) was used for elution. The fractions containing the protein of interest were pooled and the 10×His:MBP:TEV tag was cleaved by

incubating overnight with TEV protease at 4 °C. To remove the cleaved 10×His:MBP:TEV tag and TEV protease, reaction mixtures were loaded onto a HiTrap heparin HP 5 column (GE Healthcare), a linear gradient of increasing NaCl concentration (from 0.2 to 1.0 M) was used for elution. The collected fractions with Cas12f1 were then dialyzed against 20 mM Tris-HCl (pH 8.0 at 25 °C), 0.5 M NaCl, 2 mM DTT and 50% (v/v) glycerol and stored at -20 °C. The sequences of the Cas12f1 proteins are listed in Appendix 2.

2.2.5.RNA purification from Cas12f1-RNA complex

To isolate Cas12f1 bound RNA species, SpCas12f1 and AsCas12f1 ribonucleoprotein (RNP) complexes (250 µl) were incubated with 5 µl (20 mg/ml) of Proteinase K (Thermo Fisher Scientific) for 45 min at 37 °C in 1 ml of 10 mM Tris-HCl (pH 7.5 at 37 °C), 1 mM EDTA, 1 mM DTT, 100 mM NaCl and 5 mM $MgCl_2$ buffer. Furthermore, DNA was removed by incubation for 45 min at 37 °C with 10 µl of DNase I (Thermo Fisher Scientific). The RNA was purified using a GeneJET PCR Purification column (Thermo Fisher Scientific) and eluted in nuclease-free water. RNA concentration and purity were measured by NanoDrop spectrophotometer and RNA integrity was visualized by separating reaction products on TBE-Urea (8 M) 15% denaturing polyacrylamide gel with 0.5×TBE (Tris-borate-EDTA) buffer (Thermo Fisher Scientific) and staining with SYBR Gold (Thermo Fisher Scientific).

2.2.6.RNA sequencing and analysis

Purified RNA was prepared for sequencing using a TruSeq Small RNA Library Preparation Kit (Illumina) according to the manufacturer's instruction except an expanded size selection was performed allowing RNA species ~30-300 nt in length to be captured. After library preparation, 150 nt paired-end sequencing was performed on a MiSeq System (Illumina). The resulting data was post-processed trimming to a Phred quality score of 13, adapters hard-clipped using Cutadapt v2.10, and mapped to the reference using Bowtie2 v2.4.2 (Langmead and Salzberg 2012). Coverage data was then viewed in IGV (Robinson et al. 2011) and crRNA and tracrRNA species were identified from the resulting read pileups.

2.2.7.RNA synthesis

Templates for T7 transcription of Cas12f1 single gRNAs were generated by PCR using overlapping oligonucleotides, altogether, containing a T7 promoter

followed by the gRNA sequence. RNAs were produced by *in vitro* transcription using TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific) and purified using GeneJET RNA Cleanup and Concentration Kit (Thermo Fisher Scientific). Sequences of the gRNAs used in this study are available in Appendix 4.

2.2.8.Cas12f1-gRNA complex assembly for *in vitro* DNA cleavage

1 μM of purified Cas12f1 protein was combined with its corresponding guide RNA (gRNA) in 1:1 molar ratio in complex assembly buffer (10 mM Tris-HCl (pH 7.5 at 37 °C), 100 mM NaCl, 1 mM EDTA, 1 mM DTT) and allowed to incubate at 37 °C for 30 min.

2.2.9.DNA substrate generation

Complementary oligonucleotides (Metabion) containing target and PAM sequences were annealed and cloned into pUC18 plasmid over HindIII (Thermo Fisher Scientific) and EcoRI (Thermo Fisher Scientific) restriction sites. The links to the plasmid sequences are provided in Appendix 3.

The 5'-ends of oligonucleotides were first radiolabelled using T4 PNK (Thermo Fisher Scientific) and [γ-^{33}P]ATP (Un1Cas12f1) or [γ-^{32}P]ATP (SpCas12f1 and AsCas12f1) (PerkinElmer). Then DNA substrates were generated by annealing two oligonucleotides with complementary sequences of whom one already had a radioactive label introduced at the 5'-end. Annealing was performed at 95 °C following slow cooling to room temperature. The sequences of the oligo duplexes are provided in Appendix 5.

2.2.10. DNA cleavage assays

Reaction mixtures of 3 nM plasmid DNA, 100 nM Cas12f1 RNP complex in 2.5 mM Tris-HCl (pH 7.5 at 37 °C), 25 mM NaCl, 0.25 mM DTT and 10 mM MgCl2 buffer for Un1Cas12f1 or 10 mM Tris-HCl (pH 7.5 at 37 °C), 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$ and 200 or 100 mM NaCl buffers for SpCas12f1 and AsCas12f1, respectively. The mixtures were incubated at 46 °C (Un1Cas12f1), 45 °C (SpCas12f1 and AsCas12f1) or as specified. The reaction was initiated by addition of Cas12f1 RNP complexes and was quenched at timed intervals (30 min for Un1Cas12f1 or 60 min for SpCas12f1 and AsCas12f1 if not indicated differently) by mixing with 3× loading dye solution (0.01% Bromophenol Blue, 0.03% SDS and 75 mM EDTA in 50% (v/v) glycerol). Reaction products were analysed by agarose gel electrophoresis and ethidium bromide staining.

Reactions with oligoduplexes or ssDNA oligonucleotides were typically carried-out by mixing labelled DNA samples with Cas12f1 RNP complex and incubating at 46 °C (Un1Cas12f1) or 45 °C (SpCas12f1 and AsCas12f1). Reaction mixtures contained 1 nM labelled duplex, 100 nM Cas12f1 RNP complex, and 5 mM Tris-HCl (pH 7.5 at 37 °C), 50 mM NaCl, 0.5 mM DTT and 5 mM $MgCl_2$ buffer for Un1Cas12f1 or 10 mM Tris-HCl (pH 7.5 at 37 °C), 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$ and 200 or 100 mM NaCl buffers for SpCas12f1 and AsCas12f1, respectively, in a 100 µl final volume. Aliquots of 6 µl were removed from the reaction mixture at timed intervals (0, 1, 2, 5, 10, 15 and 30 min for Un1Cas12f1, 0, 5, 15, 30 and 60 min for SpCas12f1 or 0, 1, 5, 15 and 30 min for AsCas12f1) and quenched with 10 µl of a loading dye (95% (v/v) formamide, 0.01% Bromophenol Blue and 25 mM EDTA). Reaction products were analysed by denaturing gel electrophoresis (20% polyacrylamide containing 8.5 M urea in 0.5× TBE buffer) which were dried and visualized by phosphor imaging.

2.2.11. Collateral activity assay

M13 ssDNA cleavage reactions were initiated by mixing M13 ssDNA (New England Biolabs) with or without DNA activator and Cas12f1 RNP complex at 46 °C (Un1Cas12f1) or 45 °C (SpCas12f1 and AsCas12f1). 10 mM Tris-HCl (pH 7.5 at 37 °C), 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$ and 200 or 100 mM NaCl buffers were used respectively for SpCas12f1 and AsCas12f1, while 2.5 mM Tris-HCl (pH 7.5 at 37 °C), 0.25 mM DTT, 10 mM $MgCl_2$ and 25 mM NaCl - for Un1Cas12f1. The final reaction mixture consisted of 3 nM M13 ssDNA (5 nM for experiments with Un1Cas12f1), 100 nM ssDNA or dsDNA activator or no activator and 100 nM Cas12f1 RNP. After initiating the reaction by adding Cas12f1 RNP, the samples were collected at timed intervals (0, 5, 15, 30, 60 (and 90 min for Un1Cas12f1) min) by mixing with 3× loading dye solution (0.01% Bromophenol Blue, 0.03% SDS and 75 mM EDTA in 50% (v/v) glycerol). Reaction products were separated on an agarose gel and stained with SYBR Gold (Thermo Fisher Scientific). The sequences of the activators are listed in Appendix 5.

2.2.12. DNA binding assay

Binding assays were performed by incubating different amounts of Cas12f1 RNP complexes (0, 10, 50, 100 and 250 nM) with 1 nM of ^{32}P-5′-labeled ss or dsDNA substrates (Appendix 5) in binding buffer (40 mM Tris-HAc (pH 8.4 at 25 °C), 1 mM EDTA, 0.1 mg/ml BSA, 10% (v/v) glycerol and 5 mM

$Mg(C_2H_3O_2)_2$). All reactions were incubated for 30 min at room temperature (or as indicated) before electrophoresis on a native 8% (w/v) PAGE. Electrophoresis was carried out at room temperature for 3 h at 110V using 40 mM Tris-HAc (pH 8.4 at 25 °C), 1 mM EDTA, and 5 mM $Mg(C_2H_3O_2)_2$ as the running buffer. Gels were dried and visualized by phosphor imaging.

2.2.13. Molecular weight measurements by mass photometry

Measurements were performed on a OneMP mass photometer (Refeyn Ltd.). To prepare the measurements, coverslips (No. 1.5 H, 24 × 50 mm, Marienfeld) were cleaned by sequential sonication for 5 min in Milli-Q-water, isopropanol, and Milli-Q-water. Coverslips were then dried using a clean stream of nitrogen. Measurement stock solutions of Cas12f1 RNP complex were prepared freshly before each measurement by mixing Cas12f1 protein (1 μM) and gRNA (500 nM) in complex assembly buffer (10 mM Tris-HCl (pH 7.5 at 37 °C), 100 mM NaCl, 1 mM EDTA, 1 mM DTT) followed by incubation at 37 °C for 30 min. To prepare measurement stock solutions for pure Cas12f1 protein, gRNA, and DNA samples the respective stock solutions were diluted to 500 nM concentration in complex assembly buffer and incubated for 30 min at 37 °C. For DNA binding experiments, 200 nM Cas12f1 RNP complex and 25 nM DNA (Appendix 5) were mixed in binding buffer (40 mM Tris-HAc (pH 8.4 at 25 °C), 5 mM $Mg(C_2H_3O_2)_2$) and incubated for 30 min at 45 °C. After incubation, all samples were diluted by 1:10 in the respective sample buffer just before the measurement. Before the measurements, a cleaned coverslip was mounted onto the mass photometer and a gasket (CultureWell™ Reusable Gasket, Grace Bio-Labs) was placed on top. A gasket well was filled with 10 μl of the corresponding sample buffer, 10 μl of the diluted sample were added and the adsorption of biomolecules was monitored for 120 s using the AcquireMP software (Refeyn Ltd, Version 2.3.0). For converting the measured ratiometric contrast into molecular mass, Un1Cas12f1 and its oligomers ranging from 60 to 250 kDa (monomer to tetramer) were used for calibration. All mass photometry movies were analyzed using the DiscoverMP (Refeyn Ltd, Version 2.3.0). All samples were measured in triplicates.

2.2.14. Plasmid interference assay

Plasmid interference assays were performed in *E. coli* Arctic Express (DE3) strain bearing Cas12f systems (plasmids encoding CRISPR-Cas12f systems are listed in Appendix 3). For Un1Cas12f1, *E. coli* BL21 (DE3) strain was

transformed with pGB53 plasmid, which was engineered from the pLBH545_Tet-Cas14a1_Locus plasmid (gift from Jennifer Doudna, Addgene plasmid #112501) by removing tracrRNA and CRISPR array with Bsp1407I and AvrII and adding gRNA encoding sequence with T7 promoter, HDV ribozyme, and terminator sequences. The cells were grown at 37 °C to OD_{600} of ~0.5 and electroporated with 100 ng of low copy number pSC101 target plasmids obtained by cloning oligo duplexes over EcoRI and XhoI or EcoRI and NheI restriction sites into pTHSSe_1 (gift from Christopher Voigt, Addgene plasmid #109233) or pSG4K5 (gift from Xiao Wang, Addgene plasmid #74492) plasmids, respectively (the links to the plasmid sequences are provided in Appendix 3). The co-transformed cells were further diluted by serial 10× fold dilutions and grown at 37 °C for 16-20 h on plates containing inducer and antibiotics. For Un1Cas12f1 – AHT (50 ng/ml), IPTG (0.5 mM), chloramphenicol (30 μg/ml) and carbenicillin (100 μg/ml); Cas12f2 from *Micrarchaeota* archaeon (Mi1) previously named Cas14b4 (Harrington et al. 2018) – IPTG (0.5 mM), gentamycin (10 μg/ml), carbenicillin (100 μg/ml) and chloramphenicol (30 μg/ml); for all other Cas12f proteins – IPTG (0.5 mM), gentamycin (10 μg/ml), carbenicillin (100 μg/ml) and kanamycin (50 μg/ml) were used.

2.2.15. Human cell culture and transfection

HEK293T cells were purchased from ATCC (catalog number CRL-3216) and cultivated using Dulbecco's Modified Eagle Medium (DMEM) supplied with 10% fetal bovine serum, penicillin (100 U/ml), and streptomycin (100 μg/ml) (Thermo Fisher Scientific). Cells were first seeded in a 24-well plate at a density of $1.4*10^5$ cells/well. After approximately one day of growth, a transfection mixture was prepared by diluting 1 μg of plasmid, encoding nuclease and its gRNA (listed in Appendix 3) in 100 μl serum-free DMEM and 2 μl of TurboFect transfection reagent added (Thermo Fisher Scientific). After a 15 min incubation at room temperature, the transfection mixture was then added dropwise to each well containing the prepared cells. Transfected cells were then grown for 72 h at 37 °C and 5% CO_2.

2.2.16. *Zea mays* transformation

0.6 μM (average size) gold particles were first coated with SpCas12f1 expression cassettes (Appendix 3) using TransIT-2020, pelleted by centrifugation, washed with ethanol, and resuspended using sonication. 10 ul of the DNA-linked gold particles were then loaded onto a microcarrier and

allowed to air dry. Using a PDS-1000/He Gun (Bio-Rad), particles were next bombarded into 9–10-day old immature maize embryos (genotype PH1V69) with a 425 lb/in^2 rupture disc. For transient assays, a gene encoding a yellow fluorescent protein, ZsYELLOW1 N1 (Hoerster et al. 2020), was also delivered to aid in the selection of evenly transformed embryos three days after transformation. To produce T0 plants, post-bombardment culture, selection, and plant regeneration were performed using methods described previously (Gordon-Kamm et al. 2002) except *bbm* and *wus2* genes were expressed with non-constitutive promoters, maize phospholipid transferase protein (Zm-PLTP), and maize auxin-inducible (Zm-Axig1) promoters, respectively (Lowe et al. 2018).

2.2.17. Human and *Zea mays* cells genome editing assay

Transfected HEK293T cells were collected by trypsinization, and their genomic DNA was extracted using QuickExtract solution (Lucigen). For transient *Zea mays* assays, immature embryos were harvested 3 days post transformation, lyophilized, finely ground, and their total DNA extracted using Synergy 2.0 Plant DNA Extraction Kit (Ops Diagnostics). In the case of regenerated plants, two leaf punches were taken from V2 or V3 leaves, wet tissue ground, and DNA extracted using PB buffer (Qiagen) and glass fiber 96-well microfilter plate (Agilent). PCR was then performed in two rounds to amplify the DNA region surrounding each target site and add on the sequences required for Illumina sequencing and indexing (Karvelis et al. 2015; Svitashev et al. 2015). Briefly, 1-4 µl of DNA (10-200 ng) was used in a primary PCR with primers specific to the genomic locus that were 5' tailed with Illumina sequences in a final volume of 20-50 µl (Appendix 6). To ensure a balanced read composition within the initial cycles of sequencing, a mixture of four forward primers was used (see F1-F4 in Appendix 6). Each of these primers was identical except for a 6 nt region immediately 3' of the Illumina sequencing primer binding site (Appendix 6). Primary PCR was followed by a second round of PCR using 1 µl of the initial reaction as template and primers specific to the Illumina sequences added in the primary PCR that also encoded the remaining sequences needed for Illumina bridge amplification, sequencing, and data deconvolution (in a 20-50 µl final volume). All primers and targets can be found in Appendixes 6 and 7, respectively. Both rounds of PCR were allowed to proceed for 20-cycles and were carried out using NEBNext Q5 Hot Start HiFi PCR Master Mix (New England Biolabs), Phusion High-Fidelity PCR Master Mix with GC Buffer (ThermoFisher Scientific), or Platinum SuperFi II Master Mix with Green Dye (ThermoFisher

Scientific) according to the manufacturer's instruction. After PCR, 5-10 µl were separated on a 1-2% agarose gel, stained with RedSafe (iNtRON) or Ethidium Bromide (Sigma), and visualized relative to DNA molecular weight standards to be the correct size. DNA was then purified using a Monarch PCR purification column (New England Biolabs) or Zymoclean Gel DNA Recovery Kit (Zymo Research), combined in an equimolar fashion, and sequenced on a MiSeq System (Illumina) with custom sequence primers, one for the amplicon and the second for the index (Appendix 6). Sequences were trimmed to a Phred quality score of 13 and evaluated using a custom-script[35] for detection of insertion or deletion (indel) mutations that occurred within the expected cut-site. To be considered as true evidence of double-strand break repair, indel types were grouped, counted, and required to be at least 30 times greater in frequency than that found in the negative controls. The frequency of mutant reads was calculated by dividing the total number of mutant reads by the total number of wild-type reads. Mutant reads were visualized by aligning them against the wild-type reference highlighting differences in contrasting colors. The percentage of edited plants was calculated by dividing the number of plants with the specific mutation by the total number of plants with targeted modification.

3. RESULTS

CRISPR-Cas are the most widely examined microbial defense systems in the past decade. They protect their hosts against foreign nucleic acids (NA) by utilizing small guide RNAs (gRNAs) transcribed from a CRISPR locus (Koonin et al. 2017; Mohanraju et al. 2016). gRNAs together with Cas proteins form effector complexes which are directed to silence invading foreign RNA or DNA (Jackson et al. 2017; Jiang and Doudna 2017; Koonin et al. 2017). Cas9 and Cas12 proteins have been shown to specifically cleave invading double-stranded (ds) DNA, single-stranded (ss) DNA, and ssRNA (Chen et al. 2018; Gasiunas et al. 2012; Jinek et al. 2012; Ma et al. 2015; Yan et al. 2019; Zhang et al. 2015). Though these endonucleases have been adopted as robust genome editing and transcriptome manipulation tools, the sizes of most of Cas9 and Cas12 of about 1000 to 1500 aa provide constraints on cellular delivery using AAV that may limit certain applications (Lino et al. 2018; Wu, Yang, and Colosi 2010). This work reports 10 new exceptionally compact (422-603 amino acids) CRISPR-Cas12f nucleases with detailed *in vitro* examination, demonstration of activity in *E. coli*, and in human and maize cells. Thus, the presented results confirm the potential of their usage as new genome editing tools.

3.1. Cas12f – PAM dependent dsDNA nucleases

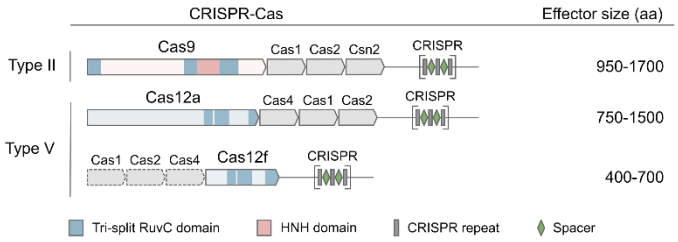


Figure 3.1. Schematic representation of CRISPR-Cas loci and effector proteins for type II and V systems. Cas9 – *Streptococcus pyogenes* SF370 (NC_002737.2), Cas12a – *Acidaminococcus sp*. BV3L6 (NZ_AWUR01000016.1), Cas12f – uncultured archaeon (KU_516197.1), and *Syntrophomonas palmitatica* (NZ_BBCE01000017.1). Tri-split RuvC domains of effector proteins are shown in blue, HNH domain of Cas9 is in pink. Grey rectangles and green diamonds represent CRISPR repeats and spacers, respectively.

With the expansion of type V CRISPR-Cas systems collection, the smallest Cas12f nucleases were discovered (Shmakov et al. 2017). However, it was thought to be an intermediate between ancestor and more complex and

larger type V CRISPR-Cas systems. Despite their miniature size, shared structural similarities of Cas12f with other Cas12 proteins inspired to assay for possible dsDNA cleavage activity (Figure 3.1).

3.1.1.Cas12f PAM sequence characterization

Cas12f2 protein from *Micrarchaeota* archaeon (Mi1) was chosen as the first experimental subject (Appendix 1). To begin with, the CRISPR-Mi1Cas12f2 locus with a modified CRISPR array, capable of targeting a randomized PAM plasmid library (Karvelis et al. 2015), was synthesized. A PAM characterization assay (Karvelis et al. 2015, 2019) was adapted to test the ability of the Mi1Cas12f2 to recognize and cleave a dsDNA target *in vitro* (Figure 3.2A). *E. coli* lysate containing Mi1Cas12f2 protein and gRNAs expressed from the reengineered locus was mixed with the PAM library. Further, DNA breaks were captured by double-stranded adapter ligation, enriched by PCR, and deep sequenced as described previously (Karvelis et al. 2015, 2019). To identify DNA cleavage, regions in the protospacer sequence were tested for elevated frequencies of adapter ligation relative to negative controls (*E. coli* lysate lacking Mi1Cas12f2). A 2.5-fold increase in the number of adapter-ligated sequences was recovered after the 21st protospacer position 3' of the randomized PAM (Figure 3.2B). The recovery of a T-rich sequence (5'-TTAT-3') immediately after 5' of the gRNA target only in the Mi1Cas12f2 treated sample was observed (Figure 3.2C-D). Additional examination for obtained results was performed: a plasmid containing a target adjacent to the identified 5'-TTAT-3' PAM sequence was subjected to cell lysate cleavage experiments and a higher copy number DNA expression plasmid equipped with an inducible T7 promoter for Mi1Cas12f2 locus was utilized. Sequencing of the target plasmid cleavage products confirmed cleavage at the 21st position (Figure 3.2E). Additionally, reactions using deletion variants further confirmed that Mi1Cas12f2 was the sole endonuclease required for the observed dsDNA target recognition and cleavage activity (Figure 3.2E).

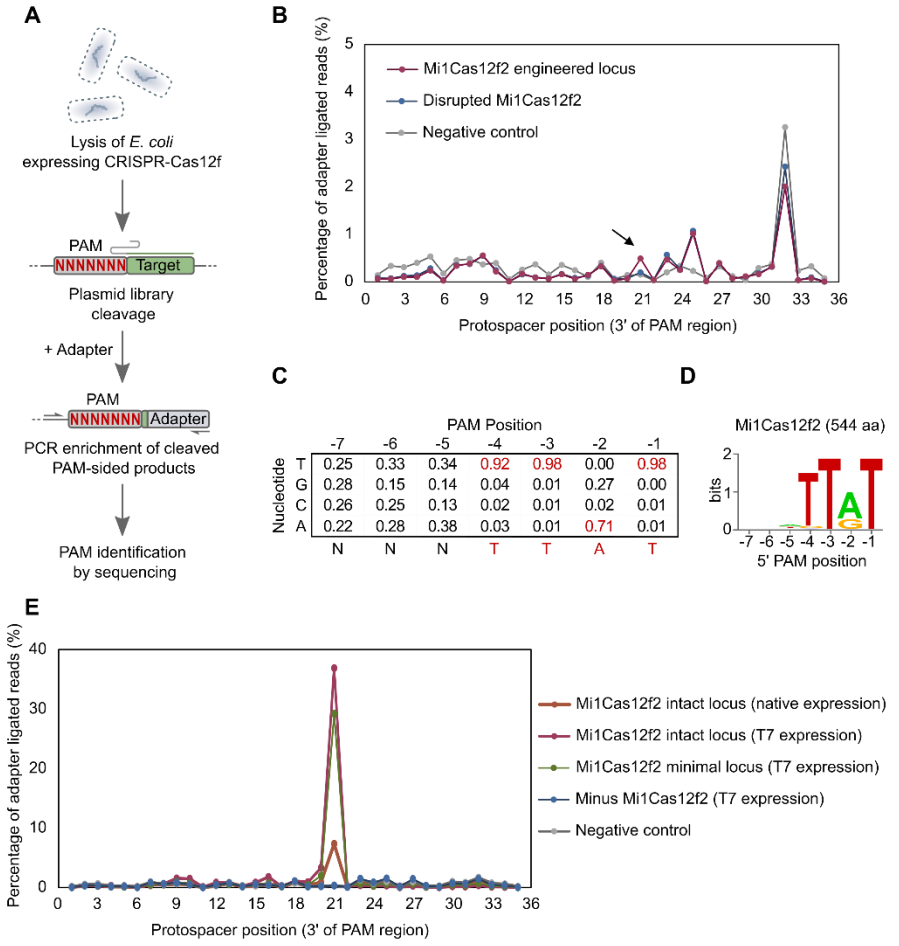


Figure 3.2. dsDNA cleavage and PAM requirements for Mi1Cas12f2 effector protein. (**A**) Workflow of the biochemical approach used to detect dsDNA cleavage and examine PAM recognition. *E. coli* cells were transformed with a plasmid carrying CRISPR-Cas12f loci engineered to target the PAM library, allowed to express, and then lysed. The resulting lysate containing Cas12f RNP complexes was used to assay for target cleavage and PAM recognition. (**B**) Relative to the negative controls, the Mi1Cas12f2 locus engineered to target a PAM library (36 nt spacers) produced a spike in the recovery of protospacer fragments ligated to an adapter just after the 21st position 3' of the PAM region. (**C**) PAM sequences that supported cleavage generated a position frequency matrix (PFM) exhibiting preferences for T and A bp 5' of the gRNA target. As a reference, a PFM at the same position in the control lysate was also calculated. (**D**) WebLogo of the PAM sequence recovered for Mi1Cas12f2 protein. (**E**) dsDNA plasmids containing a PAM and gRNA target showed an even greater enrichment in the recovery of adapters ligated just after the 21st position, especially for reactions where expression was enhanced with a T7 promoter. Experiments deleting *cas1*, *cas2*, and *cas4* genes (Mi1Cas12f2 minimal locus) and the *cas12f2* gene itself (minus Mi1Cas12f2) confirmed that Mi1Cas12f2 was the only protein required for the observed dsDNA target recognition and cleavage.

3.1.2.Diversity of Cas12f specific PAM sequences

To further explore the DNA cleavage requirements of miniature CRISPR-Cas effectors, we sought to evaluate the dsDNA cleavage activity of additional nine Cas12f proteins. All of them contain a conserved C-terminal tri-split RuvC domain, which coincides with other Cas12 nucleases. However, in some of the selected CRISPR-Cas systems, the adaptation module is absent (Koonin et al. 2017; Shmakov et al. 2017).

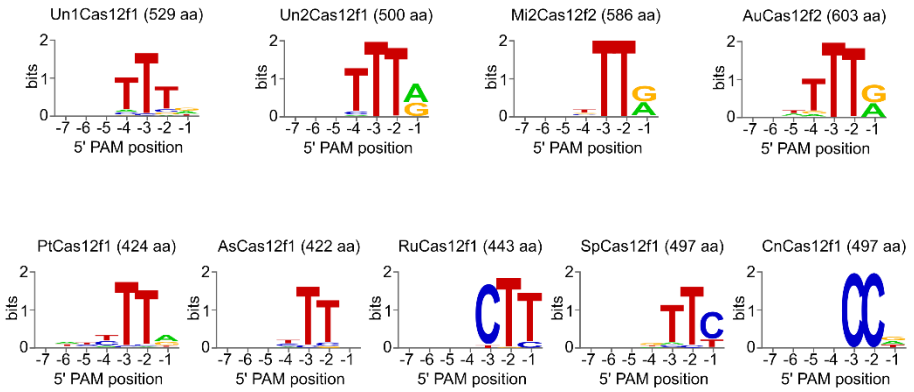


Figure 3.3. dsDNA cleavage and PAM requirements for Cas12f effector proteins. WebLogos of the PAM sequences recovered for each Cas12f protein. *E. coli* cells were transformed with a plasmid carrying CRISPR-Cas12f loci engineered to target the PAM library, allowed to express, and then lysed. The resulting lysate containing Cas12f RNP complexes was used to assay for target cleavage and PAM recognition. For Un1Cas12f1 (Cas14a1), *E. coli* lysate expressing the nuclease was mixed with *in vitro* transcribed gRNA.

Two Cas12f1 proteins from uncultured archaeon (Un1 and Un2) and two Cas12f2 proteins from *Micrarchaeota* archaeon (Mi2) and *Aureabacteria* bacterium (Au) and even smaller five additional Cas12f1, ranging in size between 422-497 amino acids, primarily from bacteria lineages, in particular *Clostridia* and *Bacilli*, were selected (Appendix 1). Apart from Un1Cas12f1, synthesized expression plasmids enclosed minimal locus of *cas12f* gene, a sequence encoding a putative tracrRNA between the nuclease gene and CRISPR repeats, and a CRISPR array modified to target the PAM library. Then, similarly to Mi1Cas12f2 experimentation, *E. coli* lysate from cells expressing the Cas12f nuclease and guide RNAs was combined with a randomized PAM library (Figure 3.2A). Obtained cleavage products were captured and analyzed. For Un1Cas12f1, *E. coli* lysate containing protein was supplemented with an *in vitro* transcribed PAM library targeting single gRNA

and then assayed for dsDNA target recognition and cleavage as described above. All produced cleavage around the 21-24 bp positions in a 5' PAM-dependent manner and altogether expanded miniature Cas nucleases PAM diversity to encompass not only T-rich but also C-rich motifs (Figure 3.3).

3.2. Un1Cas12f1 – RNA programmable ss and dsDNA nuclease

It was previously demonstrated, that Un1Cas12f1 cleaves just ssDNA containing complementary to gRNA target sequence (Harrington et al. 2018). However, the PAM characterization assay described above proved Un1Cas12f1 nuclease activity against dsDNA as well. Considering that specific gRNA for this nuclease was already described (Harrington et al. 2018), a set of programmable dsDNA cleavage experiments was performed. Under optimized reaction conditions, supercoiled (SC) plasmid DNA containing a target sequence flanked by an Un1Cas12f1 PAM (5'-TTTA-3') was completely converted to a linear form (FLL) indicating the formation of a dsDNA break (Figure 3.4A). In contrast, any DNA targets missing PAM, or the specific target site were unaffected. Additionally, alanine substitution of conserved RuvC active site (detected in most of characterized type V effectors as essential for cutting activity) residues (D326A and D510A) (Harrington et al. 2018) abolished dsDNA cleavage activity (Figure 3.4B).

The type of dsDNA break generated by Un1Cas12f1 was examined next. Run-off sequencing showed cleavage predominantly centered around positions 20-24 bp in respect to the PAM sequence (Figure 3.4C). Un1Cas12f1 produces staggered dsDNA breaks with 5' overhangs, possibly favorable for genome knock-in assays. Alike to plasmid DNA cleavage, the cleavage pattern observed on synthetic double-stranded oligodeoxynucleotides generated a 5' staggered cut pattern (Figure 3.4D). Though, a less strictly defined cleavage position was seen. While Un1Cas12f1 cleavage of dsDNA requires an additional PAM (5'-TTTA-3') sequence, it is unnecessary for ssDNA cleavage. As shown in Figure 3.4E, ssDNA oligodeoxynucleotides with or without PAM, but containing target sequences, were cleaved in the same manner.

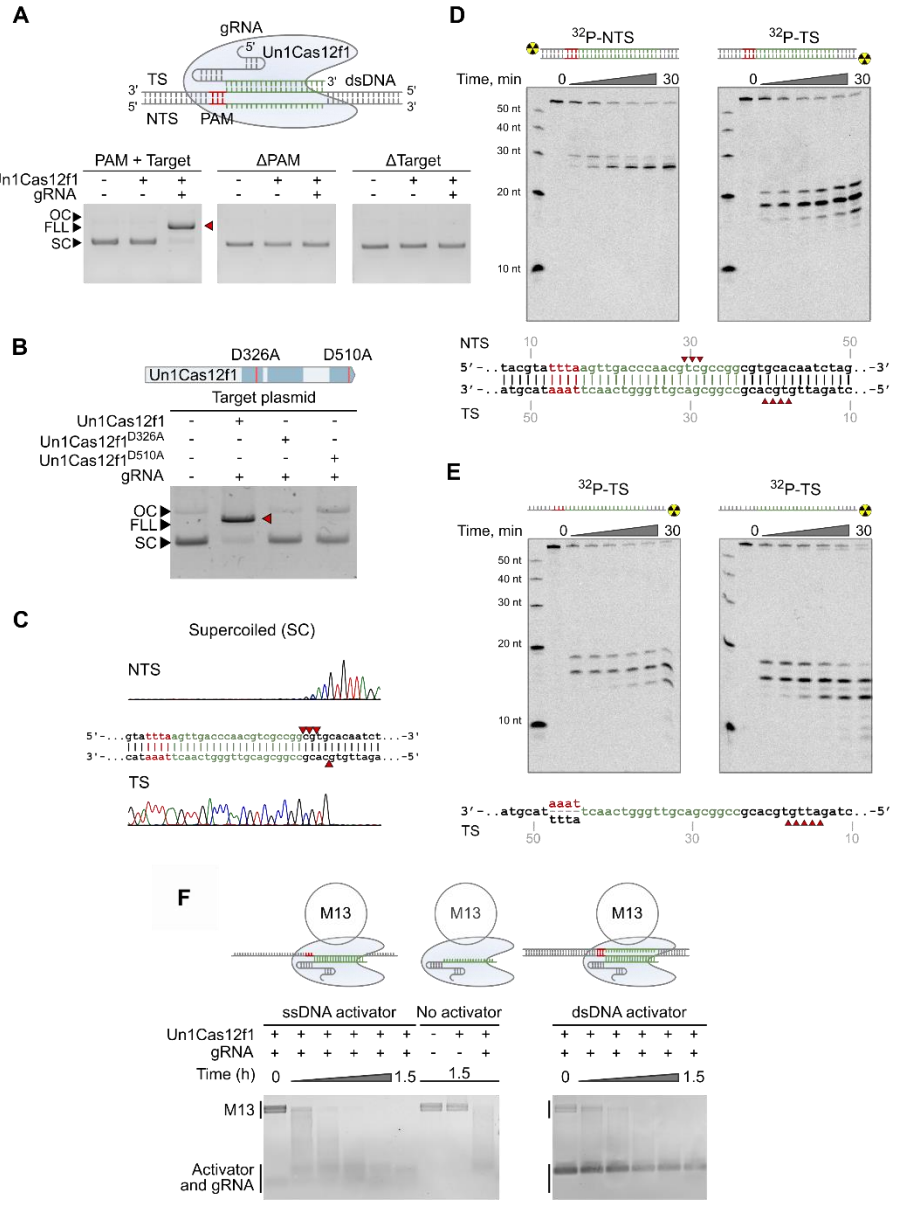


Figure 3.4. Un1Cas12f1 RNP complex is a PAM-dependent dsDNA endonuclease. (**A**) Un1Cas12f1 RNP complexes cleave plasmid DNA targets in a PAM-dependent manner (left panel) requiring both PAM (center panel) and gRNA recognition (right panel). (**B**) Alanine substitution of two conserved RuvC active site residues completely abolishes Un1Cas12f1 DNA cleavage activity. (**C**) Run-off sequencing of Un1Cas12f1 pre-cleaved plasmid DNA indicates that cleavage is centered around positions 20-24 bp 3' of the PAM resulting in 5' overhangs. (**D**) Purified Un1Cas12f1 RNP complexes cleaved radiolabelled dsDNA oligo duplexes containing a gRNA target (marked in green) in a PAM-dependent (red) manner

generating a staggered cleavage pattern. (**E**) ssDNA substrates complementary to the gRNA spacer sequence were also cleaved albeit in the PAM-independent manner. (**F**) Collateral non-specific M13 ssDNA degradation activity by Un1Cas12f1 is triggered by ssDNA or PAM-containing dsDNA targets. As observed for Cas12a (Chen et al. 2018), slight nuclease activity against non-specific ssDNA can be also seen without any DNA activator for Un1Cas12f1. Un1Cas12f1 RNP complexes were assembled using a gRNA with a 20 nt length spacer. SC, OC, and FLL stand for supercoiled, open-circle, and full-length linearization, respectively. NTS and TS represent non-target strand and target strand, respectively.

Finally, a feature shared by most other type V family members (Chen et al. 2018; Yan et al. 2019) – collateral activity on non-specific ssDNA was tested (Figure 3.4F). For this, two types of activators were used. First, the ability of Un1Cas12f1 to indiscriminately degrade single-stranded M13 DNA in the presence of an ssDNA target without a PAM was confirmed. Then, a dsDNA target containing a 5' PAM and gRNA target for Un1Cas12f1 was also tested for its ability to trigger non-selective ssDNA degradation. As shown in Figure 3.4F, the *trans*-acting DNase activity of Un1Cas12f1 was activated by both ssDNA and dsDNA targets, similarly to Cas12a (Chen et al. 2018). However, in the absence of a target, the Un1Cas12f1 RNP complex non-selectively also degraded single-stranded M13 DNA, but at much slower rates (Figure 3.4F).

3.3. Cas12f activity in *E. coli*

Next, it was tested if CRISPR-Cas12f mediated plasmid DNA interference can be performed in heterologous *E. coli* hosts. Ten systems were programmed to target and cleave invading dsDNA. For this *E. coli* plasmid DNA interference assay (Burstein et al. 2017; Sapranauskas et al. 2011), low copy number target plasmid DNA and modified minimal Cas12f CRISPR locus containing expression plasmids were used (Figure 3.5A). To assess transformation efficiency, each experiment was serially diluted by 10× and compared with controls (experiments performed with a target sequence lacking plasmid DNA). A previous study showed that Un1Cas12f1 is incapable of depleting PAM plasmid libraries in a heterologous *E. coli* host and failed to detect PAM-dependent dsDNA cleavage (Harrington et al. 2018). Accordingly, almost all of the selected Cas12f nucleases interfered with plasmid DNA transformation as evidenced by the similar recovery of resistant colonies compared to controls (Figure 3.5B). In contrast, compact CRISPR-associated nucleases from *Acidibacillus sulfuroxidans* (As) (422 aa) and

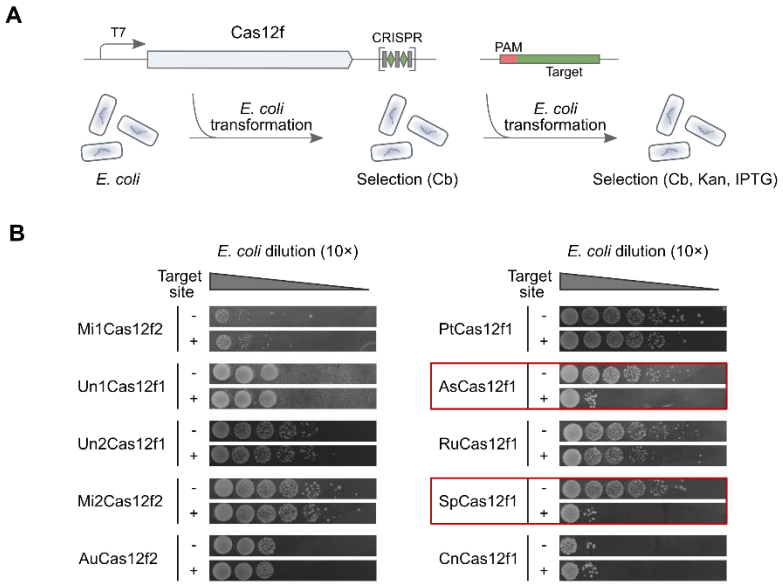


Figure 3.5. Cas12f mediated plasmid DNA interference in *E. coli*. (**A**) Overview of plasmid DNA interference in *E. coli* experimental flow. Cells bearing a minimal CRISPR-Cas12f locus were transformed with a low copy number plasmid DNA containing the target sequence. The engineered CRISPR loci contained 33-39 nt spacers except for CRISPR-Un1Cas12f1 where the CRISPR locus was replaced with a T7 expressed gRNA (20 nt spacer). (**B**) To assess transformation efficiency, each experiment was serially diluted by 10× and compared with controls (experiments performed with a plasmid that does not contain a target site). Red boxes indicate Cas12f variants that showed visually detectable levels of DNA interference activity.

Syntrophomonas palmitatica (Sp) (497 aa), both produced notable levels of plasmid interference. Slight interference was also observable for nucleases from *P. thermoglucosidasius* (Pt) and *Ruminococcus* sp. (Ru).

3.4. SpCas12f1 and AsCas12f1 – new miniature DNA nucleases

Since two of the selected Cas12f effectors (*Acidibacillus sulfuroxidans* (As) and *Syntrophomonas palmitatica* (Sp)) (Figure 3.6A) showed efficient nuclease activity in heterologous, *E. coli*, host, we decided to examine their genome editing capabilities starting with the identification of the key molecular and biochemical requirements for DNA cleavage.

3.4.1.Cas12f1 binds crRNA and tracrRNA molecules

CRISPR-Cas function as gRNA programmable DNA nucleases, where small RNA molecules transcribed from the CRISPR locus acts as foreign DNA

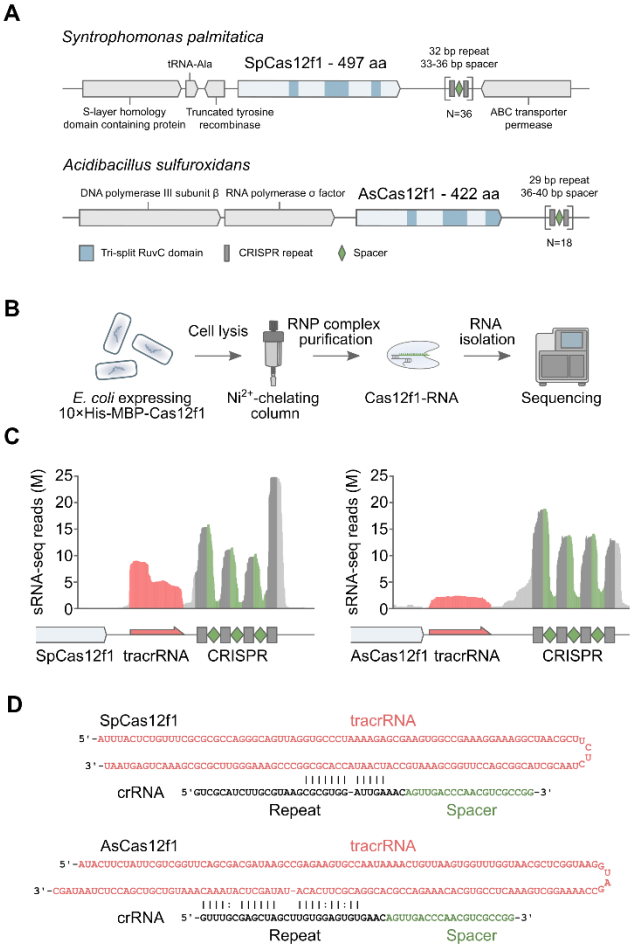


Figure 3.6. SpCas12f1 and AsCas12f1 CRISPR-Cas loci and effector complex components. (**A**) Schematic representation of native CRISPR-Cas loci encoding SpCas12f1 and AsCas12f1 effector proteins. (**B**) Workflow of the biochemical approach used to isolate and identify Cas12f1 RNP-bound RNA molecules. (**C**) Obtained small RNA sequences analysis indicated enrichment of predicted tracrRNA and CRISPR array. (**D**) *In silico* prediction of base pairing between identified tracrRNA and crRNA sequences.

recognition factors (Jackson et al. 2017; Jiang and Doudna 2017; Koonin et al. 2017). The gRNA required for nuclease activity was experimentally characterized by sequencing Cas12f1 bound RNA species. Plasmids bearing CRISPR-Cas12f1 systems were modified to include a sequence that encoded a 10×His-MBP tag at the N-terminus of each Cas12f1 nuclease. Tagged Cas12f1 ribonucleoprotein (RNP) complexes expressed in *E. coli* were pulled down from cellular lysates, the RNA was extracted and subsequently sequenced (Figure 3.6B). Two detected highly enriched RNA species included

40-50 nucleotide (nt) long CRISPR RNAs (crRNAs) comprising part of the repeat followed by the spacer, and a long (153 and 169 nt for SpCas12f1 and AsCas12f1, respectively) *trans*-activating RNA (tracrRNA) encoded between the *cas12f1* gene and the CRISPR array (Figure 3.6C and D). Both tracrRNAs contained an anti-repeat region capable of base pairing with the CRISPR repeat suggesting that the crRNA and tracrRNA may form a partial duplex (Figure 3.6D).

To further simplify the Cas and RNA complex, the single guide RNA (named as gRNA further in the text) was achieved by linking crRNA and tracrRNA through four nucleotide 5'-GAAA-3' linker for both nucleases. Interestingly, while SpCas12f1 and AsCas12f1 proteins are the most compact (<500 aa) class 2 CRISPR-Cas nucleases characterized to date, their gRNA length noticeably exceeds that identified for other class 2 effectors (Figure 3.7) (Cong et al. 2013; Harrington et al. 2018, 2020; Liu et al. 2019; Pausch et al. 2020; Ran et al. 2015; Strecker, Jones, et al. 2019; Takeda et al. 2021; Teng et al. 2018; Yan et al. 2019; Zetsche et al. 2015).

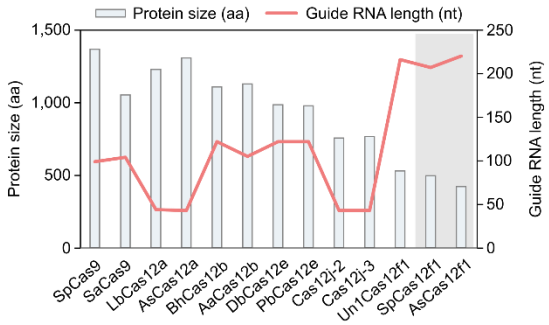


Figure 3.7. Size comparison of Cas effector proteins and respective guide RNA length. SpCas12f1 and AsCas12f1 nucleases highlighted in the grey area.

3.4.2.Cas12f1 optimal reaction conditions

Next, the biochemical properties of SpCas12f1 and AsCas12f1 proteins were assessed. All RNP complexes were assembled by mixing Cas12f1 protein with an engineered respective single gRNA, obtained by linking the identified crRNA and tracrRNA sequences through a 5'-GAAA-3' linker (Appendix 4). Initially, three different effects on plasmid DNA cleavage *in vitro* were tested. Firstly, as seen in Figure 3.8 left graphs, ambient temperature increase resulted in up to a two-fold higher percentage of final cleavage product. While active at broad range of temperature, SpCas12f1 (Figure 3.8A) and AsCas12f1 (Figure 3.8B) favour 45-55 °C. Secondly, the outcome of variation in salt concentration in the reaction mixture was assessed. Even though, similarly to

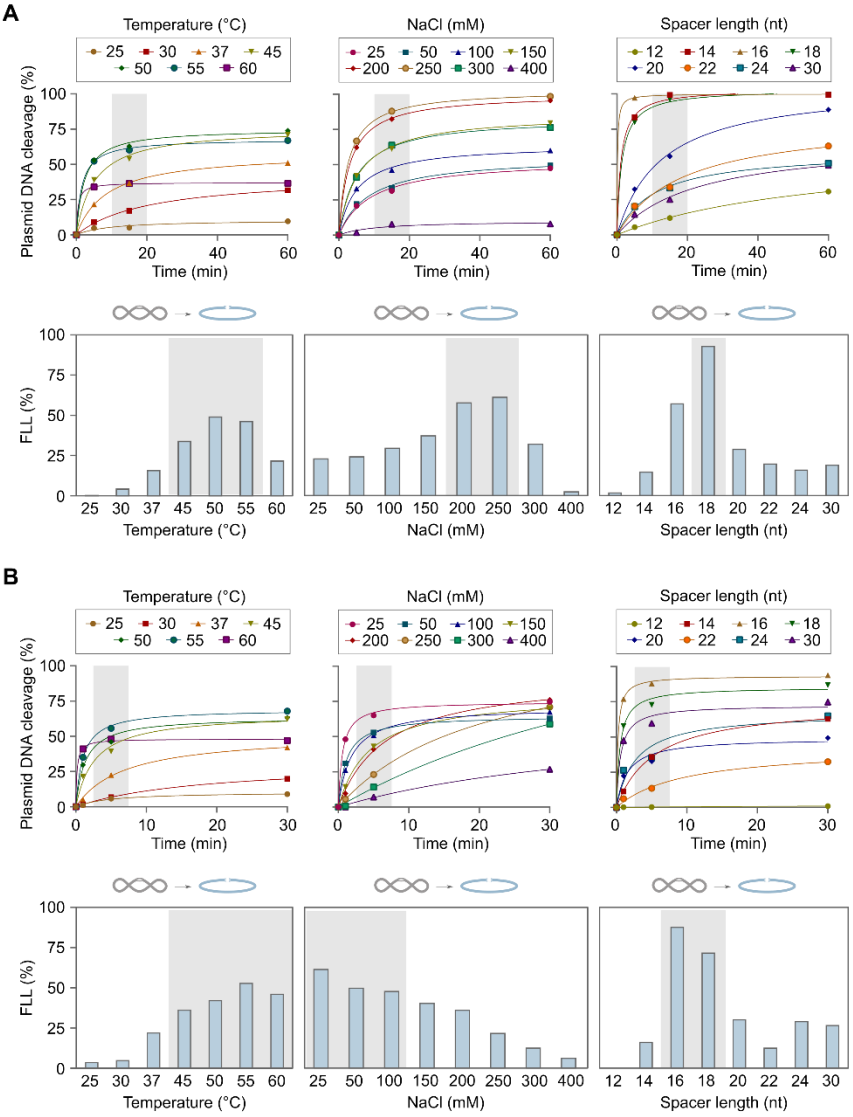


Figure 3.8. SpCas12f1 and AsCas12f1 optimal plasmid DNA cleavage conditions *in vitro*. SpCas12f1 (**A**) and AsCas12f1 (**B**) RNP plasmid DNA cleavage was assayed by independently varying temperature (with 100 mM NaCl), NaCl concentration (at 45 °C temperature) or gRNA spacer length (at optimal temperature and salt concentration conditions for both proteins). In the line graphs, grey areas represent the time point used in the histograms to compare the efficiency of full-length linear (FLL) DNA cleavage under different conditions. The grey areas in the histograms represent the optimal biochemical conditions for Cas12f1 cleavage activity. Cas12f1 RNP complexes were assembled using 20 nt spacer gRNAs.

Un1Cas12f1, AsCas12f1 showed preference for lower (25-100 mM, NaCl), SpCas12f1 exceptionally requires higher (200-250 mM, NaCl) salt

concentrations (Figure 3.8). Lastly, different optimal spacer lengths for Cas proteins were shown. While Un1Cas12f1 functions best with 20 nt spacer gRNA, both SpCas12f1 and AsCas12f1 required spacer lengths of at least 16 nt for effective cleavage of both DNA strands (Figure 3.8). Furthermore, after additionally testing cleavage of two more targets for each nuclease, the most effective were 18 nt length spacer gRNA containing RNP complexes.

3.4.3.Cas12f1 efficiently cleaves dsDNA

Further, additional requirements for dsDNA target cleavage were tested. Under optimal reaction conditions for each Cas12f1, *in vitro* preassembled RNP complexes can cleave both, supercoiled and linear, specific PAM containing dsDNA targets (Figure 3.9A and B for SpCas12f1 and AsCas12f1, respectively). However, the absence of gRNA or alanine substitutions of the conserved residues in the RuvC active site completely abolished plasmid DNA cleavage (Figure 3.9). As result, it confirmed gRNA and RuvC domain involvement in Cas12f1 nuclease activity.

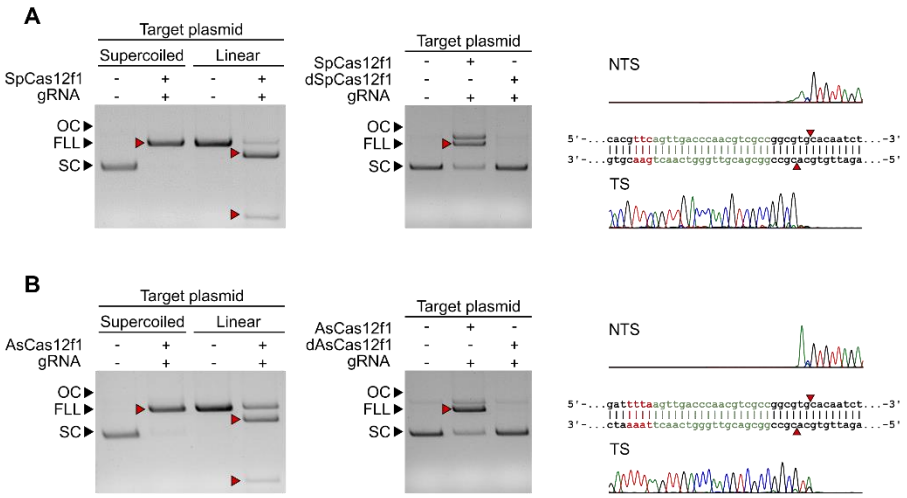


Figure 3.9. Cas12f1 cleavage of dsDNA targets. (**A**) SpCas12f1 and (**B**) AsCas12f1 RNP complexes efficiently cleave supercoiled and linear plasmid dsDNA targets resulting in a double-stranded break centered around positions 22-24 bp 3' from the PAM. However, alanine substitution of conserved RuvC active site residues (dSpCas12f1 – D228A, dAsCas12f1 – D225A) completely abolishes DNA cleavage activity for both Cas12f1 nucleases. Cas12f1 RNP complexes were assembled using gRNAs with 18 nt spacers. SC, OC, and FLL stand for supercoiled, open-circle, and full-length linear, respectively. NTS and TS represent non-target strand and target strand, respectively. The PAM is represented in red color, while the target sequence is shown in green.

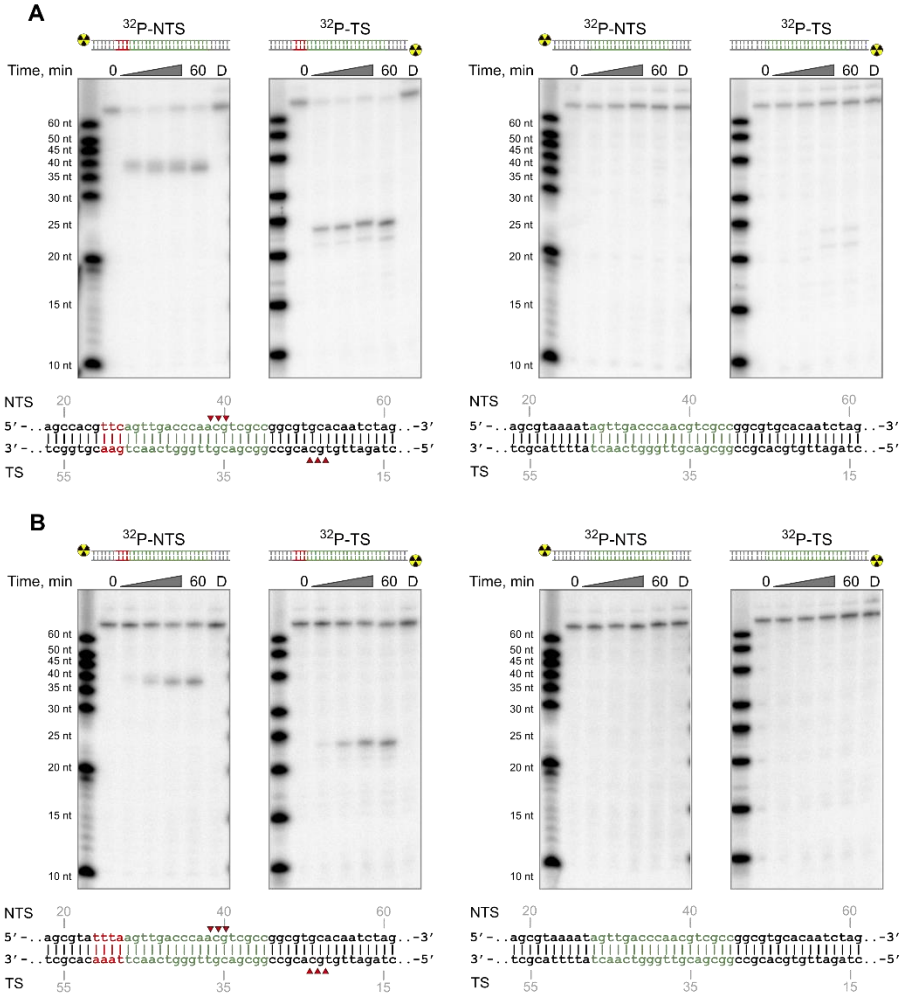


Figure 3.10. Oligo duplex cleavage by Cas12f1 RNP complexes. Cleavage of radiolabelled dsDNA oligo duplexes by *in vitro* assembled SpCas12f1 (**A**) and AsCas12f1 (**B**) RNP complexes. Both Cas12f1 RNP complexes required target sequence (marked in green) and PAM (red) to efficiently cleave dsDNA. Cas12f1 RNP complexes were assembled using 18 nt spacer gRNAs. NTS and TS represent non-target strand and target strand, respectively, D – catalytically dead (d) Cas12f1 RNP complex (dSpCas12f1 – D228A, dAsCas12f1 – D225A), which was incubated with the DNA substrate for 60 min.

Subsequent run-off sequencing of the cleavage products revealed that cutting occurred between 22 and 24 bp downstream of the 5' PAM similar to other type V effectors (Liu et al. 2019; Yan et al. 2019; Zetsche et al. 2015)(Figure 3.9).

Next, PAM requirement was checked performing reactions with shorter synthetic double-stranded oligodeoxynucleotide substrates. As shown in

Figure 3.10, products of successful DNA cleavage are visible only in the presence of a 5' PAM for both Cas12f1 nucleases. Distinct to plasmid DNA cleavage (Figure 3.9) reaction yielded 5' overhang products with a similar cut site at the target strand, but closer position to the PAM sequence at the non-target strand (Figure 3.10). Former could be a result of differences in dsDNA unwinding and R-loop formation processes.

3.4.4.Cas12f1 cleaves ssDNA in a PAM-independent manner

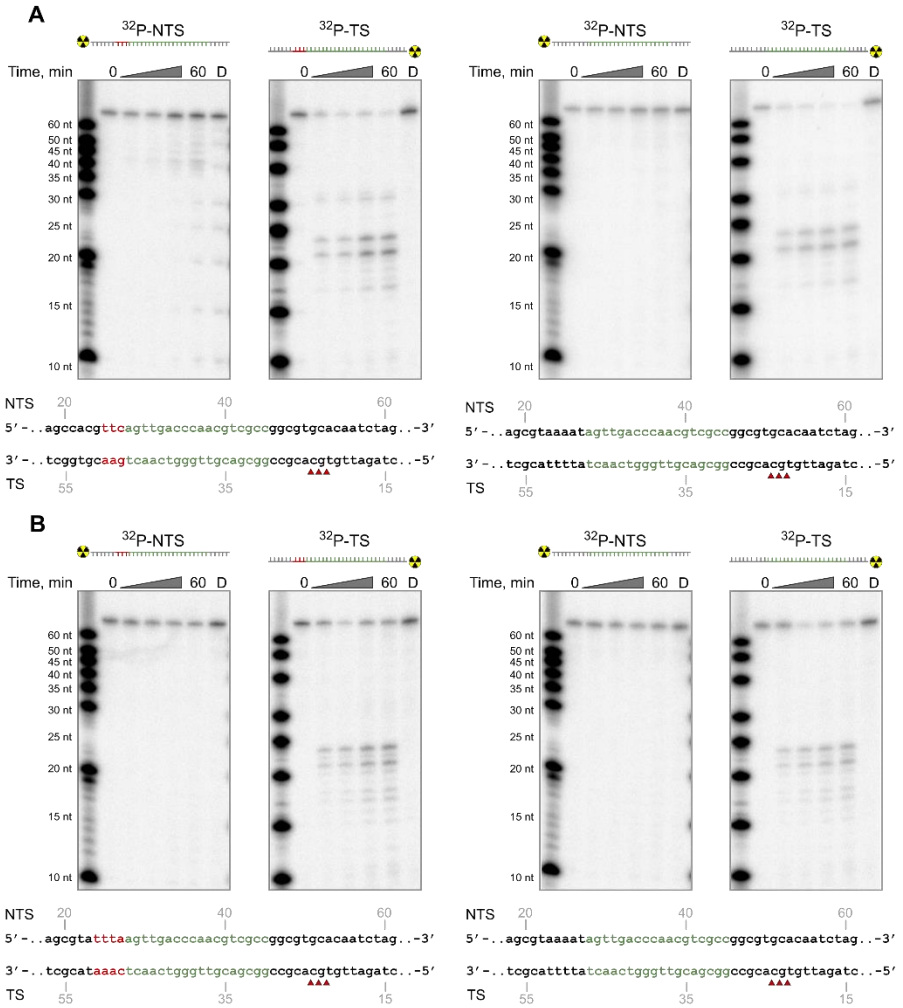


Figure 3.11. ssDNA cleavage by Cas12f1 RNP complexes. Cleavage of radiolabelled ssDNA by *in vitro* assembled SpCas12f1 (**A**) and AsCas12f1 (**B**) RNP complexes. Substrates containing complementary to gRNA sequence (marked in green) were efficiently cleaved in a PAM (red) independent manner. Cas12f1 RNP complexes were assembled using 18 nt spacer gRNAs. NTS and TS represent non-target strand and target strand, respectively, D - dCas12f1 RNP complex (dSpCas12f1

– D228A, dAsCas12f1 – D225A), which was incubated with DNA substrate for 60 min.

While Cas12f1 dsDNA cleavage requires additional PAM sequence, similar dependence on ssDNA was tested. As previously shown in Harrington et al., 2018 and Figure 3.4E, Un1Cas12f1 cleaves ssDNA in PAM independent manner, which completely overlaps with the resulting cleavage for both, AsCas12f1 and SpCas12f1, nucleases (Figure 3.11). Though less strictly defined, partially coinciding to oligo duplex cleavage position was seen.

3.4.5. *Cis*-activated Cas12f1 *trans*-cleaves non-specific ssDNA

Next, Cas12f1 nuclease activity on non-specific ssDNA was examined. Triggered by DNA target binding, SpCas12f1 and AsCas12f1 showed collateral nuclease activity that manifested as indiscriminate degradation of ssDNA (Figure 3.12). As seen in Figures 3.4F and 3.12, the *trans*-acting DNase activity of Cas12f1 nucleases can be activated by PAM and gRNA target containing dsDNA or PAM lacking ssDNA. While it was also shown to be present for Cas12a, this feature could be broadly shared by more type V family members. Oddly, Cas12g was reported to initially target ssRNA and then indiscriminately degrade both ssDNA and ssRNA (Chen et al. 2018; Yan et al. 2019). However, in the absence of a target, as seen for Un1Cas12f1 RNP complex at Figure 3.4F, AsCas12f1 and SpCas12f1 also non-selectively degraded single-stranded M13 DNA at much slower rates (Figure 3.12). As result, Cas12f1 possesses *cis*, *trans*-activated, and non-specific single-stranded DNA nuclease activity.

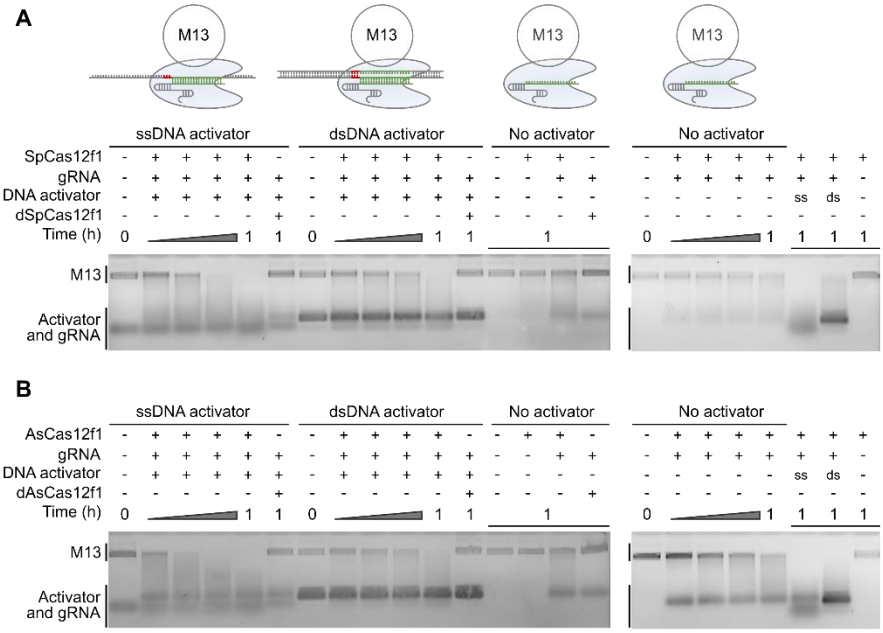


Figure 3.12. Collateral ssDNA cleavage activity of Cas12f1 RNP complexes. Collateral non-specific M13 ssDNA degradation activity by SpCas12f1 (**A**) and AsCas12f1 (**B**) is triggered by ssDNA or PAM-containing dsDNA targets. As observed for Cas12a (Chen et al. 2018), slight nuclease activity against non-specific ssDNA can be seen without any DNA activator for SpCas12f1 and AsCas12f1. dCas12f1 RNP complexes (dSpCas12f1 – D228A and dAsCas12f1 – D225A) did not degrade ssDNA. Cas12f1 and dCas12f1 RNP complexes were assembled using gRNAs with 18 nt long spacers.

3.5. SpCas12f1 and AsCas12f1 DNA binding activity

3.5.1. Cas12f1 requires a higher temperature for dsDNA binding *in vitro*

Optimal reaction temperatures of 45–55 °C for dsDNA and ssDNA cleavage by SpCas12f1 and AsCas12f1 nucleases prompted us to also investigate the effect of temperature on DNA target binding. According to electrophoretic mobility shift assays, both nucleases robustly associate with the target ssDNA strand at room temperature (Figure 3.13A and C). However, binding activity is barely seen when using dsDNA targets. PAM-dependent dsDNA-binding affinity increased significantly after incubation at 37 and 45 °C, consistent with cleavage results (Figure 3.13B and D).

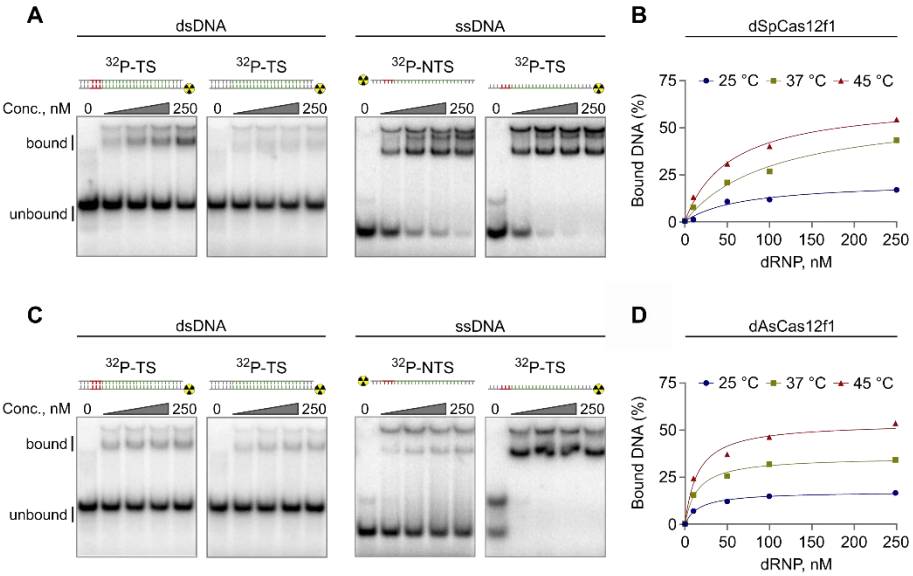


Figure 3.13. DNA binding activity of Cas12f1 RNP complexes. dsDNA and ssDNA binding by dSpCas12f1 (D228A) (**A**) and dAsCas12f1 (D225A) (**C**) RNP complexes. dSpCas12f1 and dAsCas12f1 both preferentially bind PAM containing dsDNA targets. For ssDNA, dAsCas12f1 binds more strongly to the TS (**C**), while dSpCas12f1 binds both TS and NTS with similar affinity (**A**). Different amounts of dSpCas12f1 and dAsCas12f1 RNP complexes were pre-incubated with 1 nM of ^{32}P-5′-labeled ds or ssDNA substrates at room temperature. (**B** and **D**) Temperature-dependent dsDNA binding by dSpCas12f1 (D228A) and dAsCas12f1 (D225A) RNP complexes, respectively. Different amounts of dSpCas12f1 and dAsCas12f1 RNP complexes were pre-incubated with 1 nM of ^{32}P-5′-labeled dsDNA substrates at the indicated temperatures. Samples were analyzed by non-denaturing PAGE (polyacrylamide gel electrophoresis) at room temperature. Schematic representation of the DNA substrates is shown above the corresponding gel (PAM shown in red color, target in green). NTS and TS represent non-target strand and target strand, respectively. dCas12f1 RNP complexes were assembled using gRNAs with 18 nt long spacers.

3.5.2. Active ternary Cas12f1 complex

Recently, it was demonstrated that the active Un1Cas12f1 RNP complex consists of two Un1Cas12f1 nucleases and one of each - its specific gRNA and target DNA (Takeda et al. 2021; Xiao, Li, et al. 2021). Considering SpCas12f1 and AsCas12f1 structural and functional similarities to Un1Cas12f1, RNP component ratios were decided to be measured using mass photometry (Young et al. 2018). Apo-form – binary nuclease complex with its gRNA, and ternary complex – nuclease and gRNA bound to a dsDNA target, were evaluated for SpCas12f1 and AsCas12f1 (Figure 3.14A and B,

respectively). Omitting gRNA and dsDNA target resulted in both nucleases as predominantly monomers, although, a smaller fraction of homodimers was observed. The addition of gRNA contributed to predominant species occurring at 179 and 168 kDa for SpCas12f1 and AsCas12f1, respectively, corresponding to a 2:1 Cas12f1:gRNA complexes (Figure 3.14). Following inclusion of target dsDNA, the ternary SpCas12f1 and AsCas12f1:gRNA:dsDNA complexes showed a 2:1:1 stoichiometry. In agreement with the previous Un1Cas12f1 structural assessment (Takeda et al. 2021; Xiao, Li, et al. 2021), catalytically competent Cas12f1 complexes consist of two nuclease subunits bound to one gRNA and a single DNA molecule (Figure 3.14).

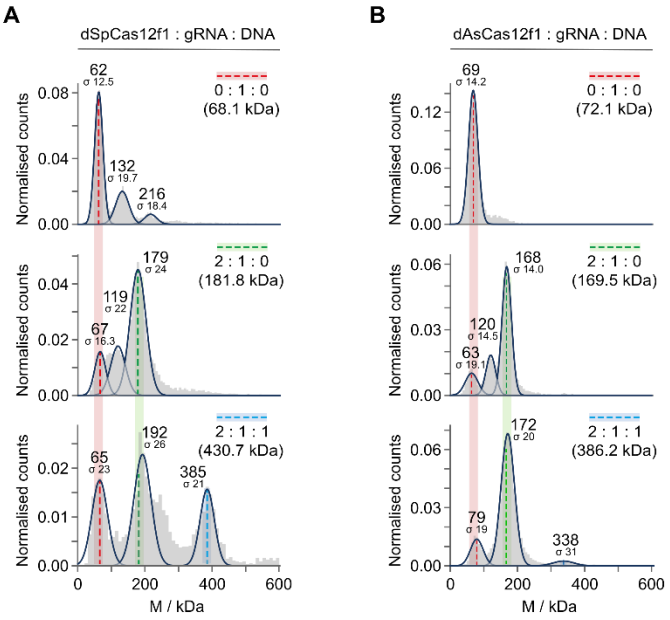


Figure 3.14. Molecular weight measurements of Cas12f1 protein and RNP complexes using mass photometry. Molecular mass distributions obtained for dSpCas12f1 (D228A) (**A**) and dAsCas12f1 (D225A) (**B**). Colored dashed lines indicate the observed molecular weights for the different, components: red – gRNA, green – dCas12f1-gRNA binary complex, blue – dCas12f1-gRNA-DNA ternary complex. Theoretical masses of the main species are shown in brackets for the given stoichiometries, while validated ones are assigned to each identified peak. dCas12f1 RNP complexes were assembled using gRNAs with 18 nt long spacers.

3.6. Cas12f1 – DNA manipulation tools for eukaryotes

3.6.1. SpCas12f1 cleaves genomic DNA in human cells

Since both nucleases were active in *E. coli* (Figure 3.5), we decided to further evaluate the genome editing potential of SpCas12f1 and AsCas12f1. For this assay human HEK293T cells, that prefer 37 °C, were selected. Initially, three target sites, in *VEGFA* and *DNMT1* genes, bearing optimal PAM sequences for SpCas12f1 (5'-TTC-3') and AsCas12f1 (5'-YTTN-3') were chosen. Expression plasmids encoding SpCas12f1 or AsCas12f1 nuclease and gRNA for each specific target site were transfected. Afterward, in 72 h period, cells were harvested and lysed, obtained genomic DNA PCR specific amplified, and sequenced for the presence of mutations at each of the selected target sites (Figure 3.15A). While AsCas12f1 nuclease activity in HEK293T was undetectable, all three SpCas12f1 target sites tested showed evidence of DNA double-strand break repair and ranged in frequency from 0.1-3.6% (Figure 3.15B). In agreement with *in vitro* cleavage results (Figure 3.9A), attained mutations also were centered around 22-24 nt from the PAM sequence, which even exceeds the actual target site (Figure 3.15C).

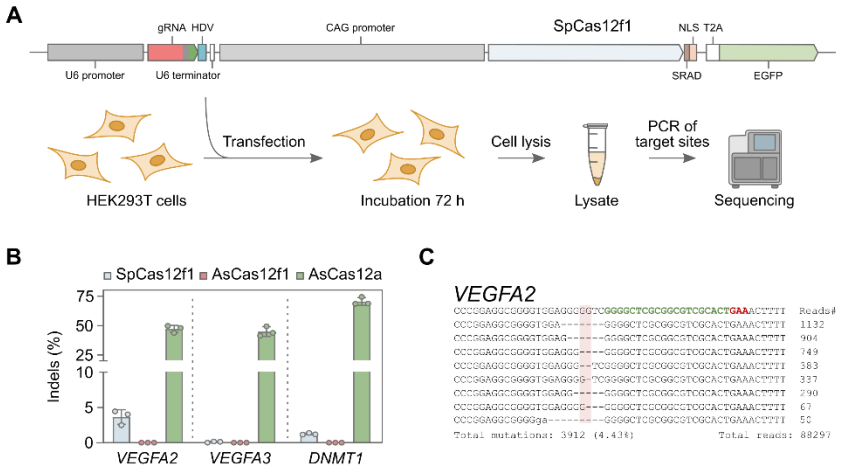


Figure 3.15. SpCas12f1 activity in human (HEK293T) cells. (**A)** Overview of HEK293T genome editing experimental flow. Cells were transfected with a DNA expression construct encoding Cas12f1 or AsCas12a and its corresponding gRNA. (**B)** Frequency of indels obtained three days after HEK293T transfection. As a control, experiments were also conducted with *Acidaminococcus sp*. BV3L6 (As) Cas12a. Bars represent mean values with ± SD (standard deviation) error bars and dots represent data for n=3 independent biological replicates. (**C**) Amplicon sequencing data yielded indels for VEGFA2 target sites where PAM is marked in red and gRNA target sequence in green. The expected position of cleavage is shaded in red.

3.6.2. Efficient DNA cleavage by SpCas12f1 in plants

While SpCas12f1 successfully showed DNA cleavage activity in human cells, the obtained DNA mutation frequency reached the highest of just 3.6% (Figure 3.15B). Thereafter, reasoning that the ambient 37 °C temperature used for culturing HEK293T cells may limit SpCas12f1 DNA target binding and cleavage, genome editing was assayed in *Zea mays* (maize) cells. The following experiments were performed in collaboration with Joshua K. Young and his team at Corteva company.

Zea mays cells have been shown to tolerate short periods of higher temperatures bursts for up to 45 °C – optimal SpCas12f1 dsDNA target binding and cleavage temperature *in vitro* (Figure 3.8A and 3.13A) (Barone et al. 2020; D. Wang et al. 2020). First, targets were selected in two agronomically relevant genes, *male sterile 26* (*ms26*) and *waxy* (Djukanovic et al. 2013; Fan et al. 2009), and expression plasmids of nuclease and respective gRNA for each target were transformed using the biolistic method (Figure 3.16A). Next, 24 h after, immature embryos were incubated at 45 °C or 37 °C, for 4 h once per day for a total of three days, while as control, embryos subject to 28 °C were maintained at this temperature for the duration of the experiment (Figure 3.16A). Lastly, embryos were harvested, and target regions were deep sequenced. For both *ms26* and *waxy* sites, only in the treatments incubated at 45 °C evidence of targeted mutagenesis were seen (Figure 3.16B).

To further examine heat treatment impact, the transformation was repeated with or without either one or three 4 h 45 °C incubations. As observed just three days after transformation, analysis of generated T0 plants showed the presence of DNA sequence alterations only after at least one 45 °C heat treatment (Figure 3.16C). Here, in both, *waxy* and *ms26*, cases the percent of SpCas12f1 transformed plants, containing target mutation, rose significantly with an increase in the number of heat treatments. Similarly, to genome editing in human cells results, SpCas12f1 targeted alterations consisted predominately of deletions that originated near or spanned the expected cut-site (Figure 3.16D). Analogous experiments conducted using *Streptococcus pyogenes* (Sp) Cas9 and gRNAs programmed to target regions overlapping with SpCas12f1 *ms26* and *waxy* sites verified comparable SpCas12f1 genome editing potential. When averaged across both targets, the editing efficiencies of SpCas12f1 using three 4 h heat treatments were half to two-thirds of those produced with a constitutively active SpCas9 (Figure 3.16C).

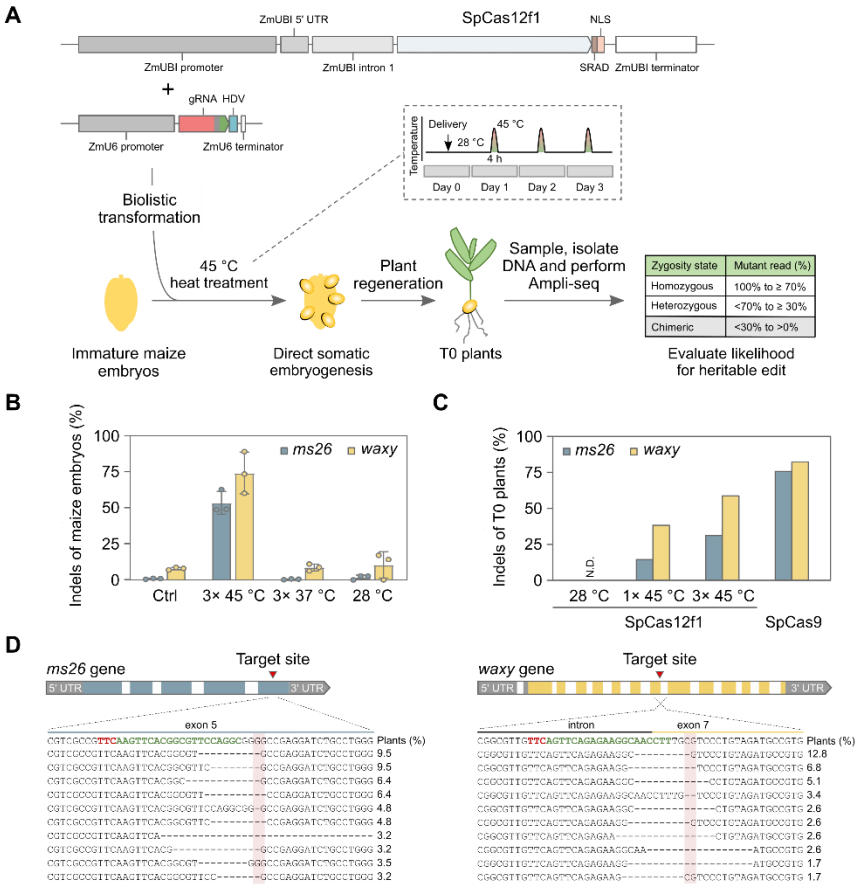


Figure 3.16. SpCas12f1 activity in maize (*Zea mays*) cells. **(A)** Overview of *Zea mays* genome editing assay. Expression constructs were delivered using biolistic transformation and short 4 h heat treatments applied one, two, or three days after transformation. T0 plants were evaluated for the likelihood of heritable edit. **(B)** Indel frequencies recovered three days after transformation at 28 °C or with the application of three consecutive 4 h long heat treatments at either 37 °C or 45 °C. Control (Ctrl) experiments were performed by omitting the gRNA expression construct from transformation. Bars represent mean values with ± SD (standard deviation) error bars and dots represent data for n=3 independent biological replicates. **(C)** Percentage of T0 plants predicted to contain a heritable *ms26* and *waxy* targeted mutation. SpCas12f1 experiments were performed at 28 °C, with one (1× 45 °C) or three (3× 45 °C) 4 h 45 °C treatments. SpCas9 experiments were performed at 28 °C. **(D)** Alignment of the ten most abundant targeted indel mutations recovered in 3× 45 °C SpCas12f1 edited T0 plants. The PAM is shown in red and the protospacer target is in green. The expected cut-site is indicated with red shading.

3.7. Final remarks

Class 2 CRISPR systems are exceptionally diverse, nevertheless, all share the property of having a single effector protein (Makarova et al. 2020). However, the size of these nucleases ranges greatly: from >1000 amino acids (aa) for Cas9/Cas12a to as small as 400-600 aa for Cas12f (Makarova et al. 2020). For *in vivo* genome editing applications, compact RNA-guided nucleases are desirable and would streamline cellular delivery approaches (D. Wang et al. 2020). This work presents biochemical characteristics of miniature Cas12f1 proteins and their possible usage as new genome editing tools in eukaryotic cells.

Cas12f1 is able to cleave ssDNA and dsDNA targets if a 5' PAM sequence is present in the vicinity of the guide RNA target. Furthermore, Cas12f1 nucleases share a set of common features: 1) compact size compared to other Cas12 nucleases; 2) remarkably long tracrRNA; 3) preferential dsDNA target binding and cleavage at higher temperatures (45-55 °C); 4) non-target ssDNA degradation upon *cis* target cleavage; 5) dimerization upon binding of a single copy of gRNA. Hence, temperature-dependent dsDNA target recognition and collateral ssDNA nuclease activity may be advantageous in nucleic acid detection platforms when being simplified as "one-pot" reactions by combining both isothermal amplification and Cas12-based detection (Chen et al. 2018; Gootenberg et al. 2017, 2018; Joung et al. 2020). Moreover, temperature sensitivity may be used to precisely regulate activity at dsDNA targets as shown in maize cells. Finally, the small size and self-dimerization of Cas12f1 enzymes provide an advantage for viral-based delivery since the exceptionally compact transcript size of the nuclease would help to overcome viral genome packaging constraints.

3.8. Current progress in the field

Despite its broad range of DNA cleavage activities *in vitro*, naturally occurring Un1Cas12f1 did not show DNA cleavage activity in heterologous *E. coli* host (Figure 3.5) (Harrington et al. 2018). However, Un1Cas12f1 and its gRNA modifications established by Xu et al. and Kim et al. broke through this limitation. Point mutations in Un1Cas12f1 DNA binding pocket (D326A/D510A – inactivated RuvC domain and additional D143R/T147R/K330R/E528R mutations) increased gene activation in mammalian cells up to a thousand times (in comparison to wild type Un1Cas12f1 activity with slightly modified gRNA) (Xu et al. 2021). Furthermore, Un1Cas12f1 containing D143R, T147R, and E151A

replacements demonstrated the best gene editing, while using already modified gRNA – repeat:anti-repeat duplex truncation (Figure 3.17A) (Xu et al. 2021). On the other hand, Kim et al. focused on extensive gRNA modifications while comparing Un1Cas12f1 gene editing efficiency in mammalian cells (Kim et al. 2021). The final optimal gRNA molecule contained 20 nt truncation of tracrRNA 5' end, the addition of a 3'- poly-uridinylate sequence (U4AU4) to the spacer, partial truncation of putative tracrRNA stem-loop, and large truncation of repeat:anti-repeat duplex (Figure 3.17A) (Kim et al. 2021).

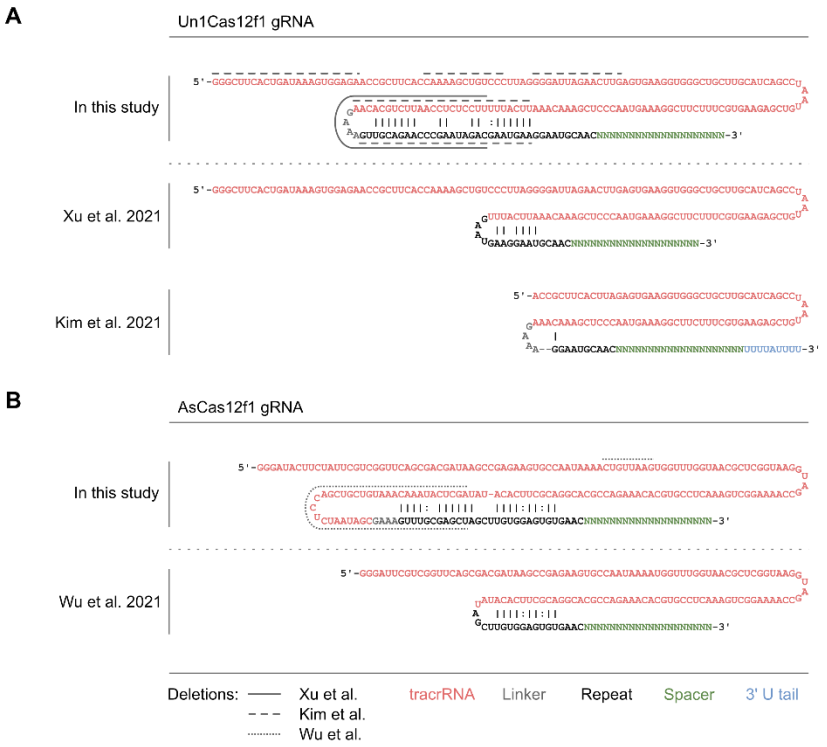


Figure 3.17. Un1Cas12f1 and AsCas12f1 gRNAs used in this and competing studies. (**A)** illustrates Un1Cas12f1 gRNAs, while (**B**) – AsCas12f1 gRNAs. In every case, the upper gRNA scheme shows in this study used exact construct and depicts modification zones where alterations were made in comparison to specific gRNAs obtained by the identified research group. Accordingly, the lower gRNA scheme presents the most optimal one obtained by the specific group with also marked alteration sites.

Moreover, Wu et al. performed AsCas12f1 biochemical characterization, very similar to this study, also confirming its activities described here (Wu et

al. 2021). The reported results do not contradict this study. Interestingly, selected optimizations of AsCas12f1 gRNA ensured its activity in a heterologous host (Wu et al. 2021). These modifications cover partial truncation of presumed tracrRNA stem-loop and significantly shorter repeat:anti-repeat duplex (Figure 3.17B) (Wu et al. 2021). Thus, consistently successful modifications described in the CRISPR field will inspire future work on Cas12f1 improvement as a new genome manipulation tool.

Furthermore, articles presenting Cas12f1 structures were also published. Two independent groups attained cryo-electron microscopy structures of Un1Cas12f1-gRNA-dsDNA and Un1Cas12f1-gRNA complexes (Takeda et al. 2021; Xiao, Li, et al. 2021). In agreement with results obtained and described in this study (3.5.2 subsection), it was shown that Un1Cas12f1 homodimer binds one gRNA and one PAM sequence containing dsDNA target molecules. Interestingly, each Un1Cas12f1 monomer adopts a different conformation, where only one RuvC nuclease domain in this dimer can cleave the DNA target (Takeda et al. 2021; Xiao, Li, et al. 2021).

On the other side, Cas12f is thought to serve as a transitional stage, between TnpB proteins found in transposons and bigger and more complex type V CRISPR-Cas systems (Harrington et al. 2018; Makarova et al. 2020; Shmakov et al. 2017). Accordingly, IscB proteins, also found in transposons, are seen as the main candidates of type II, Cas9, ancestors (Makarova et al. 2020). Recently, Altae-Tran et al. and Karvelis et al characterized both, IscB and TnpB (Altae-Tran et al. 2021; Karvelis et al. 2021). Firstly, the key additional component for the nuclease activity of IscB and TnpB was found. RNA-seq showed both types of proteins bound to specific guide RNA molecules. ωRNA (ω – OMEGA (Obligate Mobile Element Guided Activity)) coding regions were found in close proximity to *iscB* genes with a variable sequence at the 5’ end, while reRNA (re – right element) coding sequence overlaps with TnpB gene 3’end completely with additional variable 16 nt length sequence at the 3’ end (Altae-Tran et al. 2021; Karvelis et al. 2021). These variable segments act as spacers in guide RNA molecules allowing to redirect enzyme to any DNA target of interest. Furthermore, IscB and TnpB proteins together in complexes with respective RNA can cleave double-stranded DNA targets. Identically to Cas9 and Cas12 specificity to PAM (protospacer adjacent motif), IscB and TnpB proteins recognize short TAM (transposon-associated motif) sequences near the target sequence (Altae-Tran et al. 2021; Karvelis et al. 2021). Interestingly, these sequences correlate with the transposition sites, providing more information on the original function of these proteins as part of the transposon element. Lastly, both nucleases were shown to cleave DNA target sites in human cells with up to ~20% efficiency

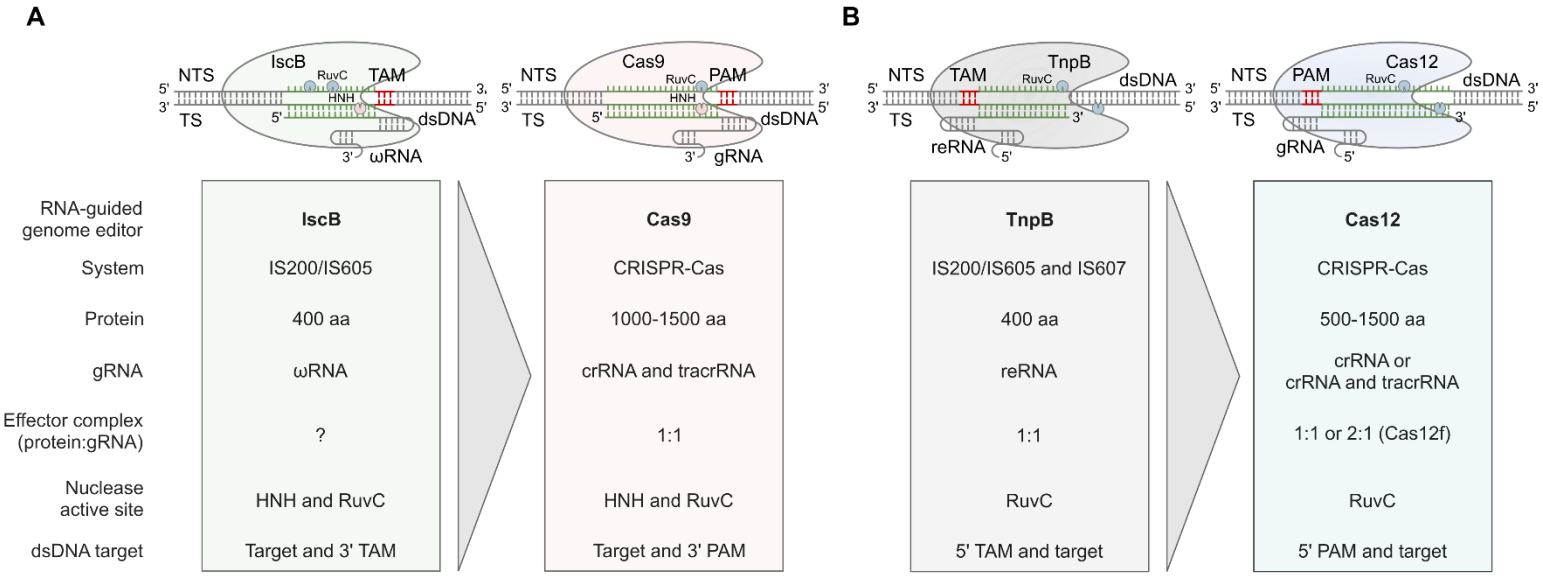


Figure 3.18. Comparison of biochemical features of Cas9 and Cas12 nucleases with their ancestor, IscB and TnpB, nucleases, respectively. (**A**) IscB and Cas9, and (**B**) TnpB and Cas12 – RNA-guided nucleases. Here, the differences and similarities of these proteins are presented. HNH and RuvC – nuclease domains, ωRNA – OMEGA (Obligate Mobile Element Guided Activity) RNA, reRNA – transposon right end RNA, crRNA – CRISPR RNA, tracrRNA – *trans*-activating RNA, gRNA – guide RNA, in this case, it could be both, just crRNA or crRNA and tracrRNA hybrid, PAM – protospacer adjacent motif, TAM – transposon adjacent motif. Adapted from (Karvelis et al. 2021) with the addition of IscB information adapted from (Altae-Tran et al. 2021)

with *Deinococcus radiodurans* TnpB proving their potential for future applications in the genome editing field (Altae-Tran et al. 2021; Karvelis et al. 2021). A detailed comparison of IscB to Cas9 and TnpB to Cas12 is displayed in Figure 3.18.

With the growing collection of characterized CRISPR-Cas nucleases and, now, their ancestor proteins we can expect an even greater impetus in genome editing and other gene manipulations fields. Here, Cas12f, as well as TnpB and IscB, provide DNA nuclease activity with important ease of programmability without any shortage on delivery cargo limitations.

CONCLUSIONS

1. Miniature Cas12f nucleases cleave dsDNA targets containing T or C-rich 5’ PAM sequences.
2. *Acidibacillus sulfuroxidans* (As) and *Syntrophomonas palmitatica* (Sp) Cas12f1 proteins bind crRNA and tracrRNA molecules, which can be engineered into a single guide RNA (gRNA).
3. Cas12f1:gRNA complexes show 2:1 stoichiometry and require increased temperature (45-55 °C) for optimal target DNA binding, cleavage, and collateral ssDNA degradation *in vitro*.
4. SpCas12f1 nucleases can be adapted to cleave genomic DNA in human and maize cells.

LIST OF PUBLICATIONS

The thesis is based on the following original publications:

1. Karvelis, T.[*], **Bigelyte, G.[*]**, Young, J.[*], Hou, Z., Zedaveinyte, R., Budre, K., Paulraj, S., Djukanovic, V., Gasior, S., Silanskas, A., Venclovas, Č., Siksnys, V. PAM recognition by miniature CRISPR-Cas12f1 nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acid Research* 2020 May; 48(9): 5016-5023 (IF = 16,971, Q1) doi: 10.1093/nar/gkaa208.

2. **Bigelyte, G.[*]**, Young, J.[*], Karvelis, T.[*], Budre, K., Zedaveinyte, R., Djukanovic, V., Ginkel, E., Paulraj, S., Gasior, S., Jones, S., Feigenbutz, L., Clair, G., Barone, P., Bohn, J., Acharya, A., Zastrow-Hayes, G., Henkel-Heinecke, S., Silanskas, A., Seidel, R., Siksnys, V. Miniature type V-F CRISPR-Cas nucleases enable targeted DNA modification in cells. *Nature Communications* 2021 Dec; 12(1): 6191 (IF = 14,919, Q1) doi: 10.1038/s41467-021-26469-4.

[*] – these authors contributed equally.

CONFERENCE PRESENTATIONS

Oral presentations

1. **Bigelyte, G.**, Šilanskas, A., Karvelis, T., Gasiūnas, G. *Naujų CRISPR-Cas9 baltymų PAM sekos charakterizavimas.* 10-oji jaunųjų mokslininkų konferencija „Bioateitis: gamtos ir gyvybės mokslų perspektyvos". Vilnius, Lithuania, 2017 12 07.
2. **Bigelyte, G.**, Karvelis, T., Young, J., Hou, Z., Žedaveinytė, R., Budrė, K., Pauraj, S., Djukanovic, V., Stephen, G., Šilanskas, A., Venclovas, Č., Šikšnys, V. *Cas12f – mažosios CRISPR-Cas nukleazės*. 13-oji jaunųjų mokslininkų konferencija „Bioateitis: gamtos ir gyvybės mokslų perspektyvos". Vilnius, Lithuania, 2020 12 04 (diploma for the best oral presentation).

Poster presentations

1. Karvelis, T., Gasiunas, G., Young, J., **Bigelyte, G.**, Silanskas, A., Lin, H., King, M., Cigan, M., Siksnys, V. *Pipeline for rapid identification and characterization of novel Cas9 orthologs.* CRISPR: from biology to technology and novel therapeutics. Cell Symposia. Sitges, Spain, 2017 10 22-24.
2. **Bigelyte, G.**, Karvelis, T., Gasiunas, G., Young, J., Silanskas, A., Lin, H., King, M., Cigan, M., Siksnys, V. *Pipeline for rapid identification and characterization of novel Cas9 orthologs.* VitaScientia. Vilnius, Lithuania, 2018-01-03.
3. Karvelis, T., Gasiunas, G., Young, J., **Bigelyte, G.**, Urbaitis, T., Silanskas, A., Lin, H., King, M., Cigan, M., Siksnys, V. *A rapid method to interrogate the protospacer adjacent motif biodiversity presented by Class2 CRISPR proteins.* CRISPR2018. Vilnius, Lithuania, 2018 06 20-23.
4. Robb, G., Gasiunas, G., Young, J., Karvelis, T., Wang, P., Urbaitis, T., Paulraj, S., **Bigelyte, G.**, Silanskas, A., Curcuru, J., Mabuchi, M., Fuchs, R., Schildkraut, E., Siksnys, V. *Harnessing the diversity of Cas9 orthologs for genome editing.* Keystone Symposia, Genome Engineering: From Mechanisms to Therapies (B4). Victoria, Canada, 2019.02.19-23.
5. **Bigelyte, G.**, Karvelis, T., Gasiunas, G., Young, J., Urbaitis, T., Silanskas, A., Lin, H., King, M., Cigan, M., Siksnys, V. *Characterization of novel Cas9 proteins.* Conference FEBS3+. Riga, Latvia, 2019.06.17-19.
6. Gasiunas, G., Young, J., Karvelis, T., Wang, P., Urbaitis, T., Jasnauskaite, M., Grusyte, M., Paulraj, S., **Bigelyte, G.**, Silanskas, A., Curcuru, J.,

Mabuchi, M., Fuchs, R., Schildkraut, E., Robb, G., Siksnys, V. *Harnessing the diversity of Cas9 orthologs for genome editing.* CRISPR2019. Quebec, Canada, 2019.06.17-19.

7. **Bigelyte, G.**, Karvelis, T., Young, J., Hou, Z., Zedaveinyte, R., Budre, K., Silanskas, A., Venclovas, Č., Siksnys, V. *Novel miniature CRISPR-Cas14 activity against DNA.* The Coins 2020. Vilnius, Lithuania, 2020 02 25-27.
8. **Bigelyte, G.**, Karvelis, T., Young, J., Hou, Z., Zedaveinyte, R., Budre, K., Silanskas, A., Venclovas, Č., Siksnys, V. *Programmable miniature CRISPR-Cas14 cuts ss and dsDNA.* VitaScientia. Vilnius, Lithuania, 2020 01 03 (diploma for the best poster presentation).
9. Karvelis, T., **Bigelyte, G.**, Young, J., Budre, K., Zedaveinyte, R., Paulraj, S., Henkel-Heinecke, S., Silanskas, E., Seidel, R., Siksnys, V. *Programmable double-stranded DNA target cleavage by miniature CRISPR-Cas12f nucleases.* CRISPR 2021. Paris, France, 2021 06 01-10.
10. Young, J., Dukanovic, V., **Bigelyte, G.**, Karvelis, T., Barone, P., Gasior, S., Ginkel, L., Budre, K., Zedaveinyte, R., Silanskas, A., Siksnys, V. *Programmable double-stranded DNA target cleavage by miniature CRISPR-Cas12f nucleases.* CRISPR 2021. Paris, France, 2021 06 01-10.

APPENDICES

Appendix 1. Cas12f proteins used in this study.

Name	Previous name	Size (aa)	Molecular mass (kDa)	Protein id (NCBI)	Organism	Scaffold accession (NCBI)
Mi1Cas12f2	Cas14b4*	544	63.9	OIO21000.1	*Candidatus Micrarchaeota* archaeon	MK005740.1
Un1Cas12f1	Cas14a1*	529	61.5	QBM01166.1	uncultured archaeon	MK005734
Un2Cas12f1	Cas14a3*	500	58.4	QBM01093.1	uncultured archaeon	MK005732
Mi2Cas12f2	Cas14b17**	586	69.4	RLG21245.1	*Candidatus Micrarchaeota* archaeon	QMVG01000004.1
AuCas12f2	Cas14b18**	603	69.9	RJP56748.1	*Candidatus Aureabacteria* bacterium	QZJZ01000091.1
PtCas12f1[+]	C2c10[++]	424	49.5	WP_064552366.1	*Parageobacillus thermoglucosidasius*	NZ_LXMA01000038.1
AsCas12f1[+]	C2c10[++]	422	48.7	WP_109431741.1	*Acidibacillus sulfuroxidans*	NZ_MPDK01000047.1
RuCas12f1[+]	C2c10[++]	440	51.2	WP_117896622.1	Unclassified *Ruminococcus*	NZ_QTWX01000005.1
SpCas12f1[+]	C2c10[++]	497	56.9	WP_054696859.1	*Syntrophomonas palmitatica*	NZ_BBCE01000017.1
CnCas12f1[+]	C2c10[++]	497	58.5	WP_120361969.1	*Clostridium novyi*	NZ_CP029458.1

* Identified by (Harrington et al. 2018).

** Identified by BLAST alignments against the NCBI NR database. Numbering is continued from that described in (Harrington et al. 2018).

[+] Type V-U3 nuclease identified by PSI-BLAST.

[++] Originally denoted name by (Shmakov et al. 2017).

Appendix 2. Sequences of the Cas12f proteins used in this study.

Un1Cas12f1

MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFRQLQK
QAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSDVCYTRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEISNHNSDF
IIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIGEKSAWMLNLSIDVPKIDKGVDPSI
IGGIDVGVKSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKVGTVQMENLESMK
RKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENADYNAALNISNPKLKSTKEEP*

10×His:MBP:TEV:Un1Cas12f1

MKSSHHHHHHHHHHGSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKL
YPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLV
DLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVA
LKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEENLYFQSNAMAKNTITKTLKLRIVRP
YNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEI
FIKGKGIANASSVEHYLSDVCYTRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPIPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDK
YRPWEKFDFEQVQKSPKPISLLLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKICEKSAWMLNLSIDVPKIDKGVDPSIIGGIDVGVKSPLVCAIN
NAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRGFWPY
AEMQNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENADYNAALNISNPKLKSTKEEP*

SpCas12f1

MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCAEAGNMCLRDLYNYFSMPKEDRISSKDLYNAMYHKTKLLHPELPGKVANQIVNHAKDVWKRNAKLIYR
NQISMPTYKITTAPIRLQNNIYKLIKNKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYTQGQLQIKKDHQGKWYCIIPYTFPT
HETVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVSIQNQIKHSGDARKGHGRKRALKPIETLSEKEKNFRDTINHRYANRIVEAAIKQGC
GTIQIENLEGIADTTGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYGSRTICINCRHVQVSGDVCEECGGIVKK
ENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKECGHIQASGNTCEVCGSTNILKPKKIRKAK*

10×His:MBP:TEV:SpCas12f1

MKSSHHHHHHHHHHGSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKL
YPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLV

DLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVA
LKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEENLYFQSNAMGESVKAIKLKILDMFL
DPECTKQDDNWRKDLSTMSRFCAEAGNMCLRDLYNYFSMPKEDRISSKDLYNAMYHKTKLLHPELPGKVANQIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRL
QNNIYKLIKNKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYTQGQLQIKKDHQGKWYCIIPYTFPTHETVLDPDKVMGVDLGV
AKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVSIQNQIKHSGDARKGHGRKRALKPIETLSEKEKNFRDTINHRYANRIVEAAIKQGCGTIQIENLEGIADTTGS
KFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYGSRTICINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPY
IDQIIMEKCLELGIPYRSITCKECGHIQASGNTCEVCGSTNILKPKKIRKAK*

AsCas12f1

MIKVYRYEIVKPLDLDWKEFGTILRQLQQETRFALNKATQLAWEWMGFSSDYKDNHGEYPKSKDILGYTNVHGYAYHTIKTKAYRLNSGNLSQTIKRATDRFKAYQK
EILRGDMSIPSYKRDIPLDLIKENISVNRMNHGDYIASLSLLSNPAKQEMNVKRKISVIIIVRGAGKTIMDRILSGEYQVSASQIIHDDRKNKWYLNISYDFEPQTR
VLDLNKIMGIDLGVAVAVYMAFQHTPARYKLEGGEIENFRRQVESRRISMLRQGKYAGGARGGHGRDKRIKPIEQLRDKIANFRDTTNHRYSRYIVDMAIKEGCGTI
QMEDLTNIRDIGSRFLQNWTYYDLQQKIIYKAEEAGIKVIKIDPQYTSQRCSECGNIDSGNRIGQAIFKCRACGYEANADYNAARNIAIPNIDKIIAESIK*

10×His:MBP:TEV:AsCas12f1

MKSSHHHHHHHHHHGSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKL
YPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLV
DLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVA
LKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEENLYFQSNAMIKVYRYEIVKPLDLDW
KEFGTILRQLQQETRFALNKATQLAWEWMGFSSDYKDNHGEYPKSKDILGYTNVHGYAYHTIKTKAYRLNSGNLSQTIKRATDRFKAYQKEILRGDMSIPSYKRDIP
LDLIKENISVNRMNHGDYIASLSLLSNPAKQEMNVKRKISVIIIVRGAGKTIMDRILSGEYQVSASQIIHDDRKNKWYLNISYDFEPQTRVLDLNKIMGIDLGVAVA
VYMAFQHTPARYKLEGGEIENFRRQVESRRISMLRQGKYAGGARGGHGRDKRIKPIEQLRDKIANFRDTTNHRYSRYIVDMAIKEGCGTIQMEDLTNIRDIGSRFLQ
NWTYYDLQQKIIYKAEEAGIKVIKIDPQYTSQRCSECGNIDSGNRIGQAIFKCRACGYEANADYNAARNIAIPNIDKIIAESIK*

Mi1Cas12f2

MNMSKTTISVKLKIIDLSSEKKEFLDNYFNEYAKATTFCQLRIRRLLRNTHWLGKKEKSSKKWIFESGICDLCGENKELVNEDRNSGEPAKICKRCYNGRYGNQMIR
KLFVSTKKREVQENMDIRRVAKLNNTHYHRIPEEAFDMIKAADTAEKRRKKNVEYDKKRQMEFIEMFNDEKKRAARPKKPNERETRYVHISKLESPSKGYTLNGIKR
KIDGMGKKIERAEKGLSRKKIFGYQGNRIKLDSNWVRFDLAESEITIPSLFKEMKLRITGPTNVHSKSGQIYFAEWFERINKQPNNYCYLIRKTSSNGKYEYYLQYT
YEAEVEANKEYAGCLGVDIGCSKLAAAVYYDSKNKKAQKPIEIFTNPIKKIKMRREKLIKLLSRVKVRHRRRKLMQLSKTEPIIDYTCHKTARKIVEMANTAKAFIS
MENLETGIKQKQQARETKKQKFYRNMFLFRKLSKLIEYKALLKGIKIVYVKPDYTSQTCSSCGADKEKTERPSQAIFRCLNPTCRYYQRDINADFNAAVNIAKKALN
NTEVVTTLL*

Un2Cas12f1

MEVQKTVMKTLSLRILRPLYSQEIEKEIKEEKERRKQAGGTGELDGGFYKKLEKKHSEMFSFDRLNLLLNQLQREIAKVYNHAISELYIATIAQGNKSNKHYISSIV
YNRAYGYFYNAYIALGICSKVEANFRSNELLTQQSALPTAKSDNFPIVLHKQKGAEGEDGGFRISTEGSDLIFEIPIPFYEYNGENRKEPYKWVKKGGQKPVLKLIL
STFRRQRNKGWAKDEGTDAEIRKVTEGKYQVSQIEINRGKKLGEHQKWFANFSIEQPIYERKPNRSIVGGLDVGIRSPLVCAINNSFSRYSVDSNDVFKFSKQVFAF
RRRLLSKNSLKRKGHGAAHKLEPITEMTEKNDKFRKKIIERWAKEVTNFFVKNQVGIVQIEDLSTMKDREDHFFNQYLRGFWPYYQMQTLIENKLKEYGIEVKRVQA
KYTSQLCSNPNCRYWNNYFNFEYRKVNKFPKFKCEKCNLEISADYNAARNLSTPDIEKFVAKATKGINLPEK*

Mi2Cas12f2

MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHI
CKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRV
KRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRN
LETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRK
KFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDENN
RKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDK*

AuCas12f2

MKSFKLKLLPTDEQNVLLNEVFCKWASLCTRMASKGHDKERLAPPDSSGNYFNKTQLNQVNTDVTDHMGALEESASQKERAVEKVKRRLKLISDMLSEPNLRDVSQQ
KPTTFRPLEWVKEGLLKTKYHTVHYWQKECDKLTKQKERMEKTIEKIKKGKITFKPTKMSLHQNCFSLSFGKGTFSMRPFSDTKRGINLDMLTAPIQPAIGKNDGKS
SLRSKEFIARNIENYIIFSIHSQLFGLSRSEELLLNAKKEELVAKRDAMLKKKSDSLSKKIKELEKIVGRKITDSERSEIMSQGGKLSSEKFSEDNSYLKTLKVLAK
DIIGREELFRLKKYPIVIRKPLNERKKLKNLKPDEWEYYLQLSYDELEKKEFTPKTIMGIDRGLKHILAIAIYDPVQNKFVKNMLIPNPILGWKWKLRKIKRSIQHM
ERRIRAQQNAHVPENQLKKRLKSIENKIDYYYHNVSRQILNLAHDFKSAIVVEDLQNMKQHGRKKSKGLRGLNYALSNFDYGKIMGLVKYKAESENVPLLTVLPAGT
SQNCAYCLLYGKEQGNYVRNNVNSKIGKCKLHGEIDADINAARTIAICYHKNINEPKPYGERKTFKRK*

PtCas12f1

MKYTKVMRYQIIKPLNAEWDELGMVLRDIQKETRAALNKTIQLCWEYQGFSADYKQIHGQYPKPKDVLGYTSMHGYAYDRLKNEFSKIASSNLSQTIKRAVDKWNSD
LKEILRGDRSIPNFRKDCPIDIVKQSTKIQKCNDGYVLSLGLINREYKNELGRKNGVFDVLIKANDKTQQTILERIINGDYTYTASQIINHKNKWFINLTYQFETKE
TALDPNNVMGVDLGIVYPVYIAFNNSLHRYHIKGGEIERFRRQVEKRKRELLNQGKYCGDGRKGHGYATRTKSIESISDKIARFRDTCNHKYSRFIVDMALKHNCGI
IQMEDLTGISKESTFLKNWTYYDLQQKIEYKAREAGIQVIKIEPQYTSQRCSKCGYIDKENRQEQATFKCIECGFKTNADYNAARNIAIPNIDKIIRKTLKMQ*

RuCas12f1

MTLLVKVVKIHLISEQFDKAGNRIDYEEVNKILWELQKQTREAKNKTVQLLWEWNNFSSDYVKASGIYPKAKDIFGYSSVHGQANKELRTKLALNSSNLSTTTMDVC
KNFNTYKKEVWKGKRSVPSYKSDQPLDLHKDSIKLIYENNEFYVRLALLKKAEFAKYGFKDGFRFKMQVKDNSTKTILERCFDEVYKINASKLLYDQKKKWKLNLS
YSFDNKNISELDKEKILGVDVGVNCPLVASVFGDRDRFIIKGGEIEKFRKSVEARRRSMLEQTKYCGDGRIGHGRKKRTEPALNIGDKIARFRDTTNHKYSRALIEY
AVKKGCGTIQMEKLTGITSKSDRFLKDWTYYDLQTKIENKAKEVGINVVYIAPKYTSQRCSKCGYIHKDNRPNQAKFRCLECDFESNADYNASQNIGIKNIDKIIEK
DLQKQESEVQVNENK*

CnCas12f1

MITVRKIKLTIMGDKDTRNSQYKWIRDEQYNQYRALNMGMTYLAVNDILYMNESGLEIRTIKDLKDCEKDIDKNKKEIEKLTARLEKEQNKKNSSSEKLDEIKYKIS
LVENKIEDYKLKIVELNKILEETQKERMDIQKEFKEKYVDDLYQVLDKIPFKHLDNKSLVTQRIKADIKSDKSNGLLKGERSIRNYKRNFPLMTRGRDLKFKYDDND
DIEIKWMEGIKFKVILGNRIKNSLELRHTLHKVIEGKYKICDSSLQFDKNNNLILNLTLDIPIDIVNKKVSGRVVGVDLGLKIPAYCALNDVEYIKKSIGRIDDFLK
VRTQMQSRRRRLQIAIQSAKGGKGRVNKLQALERFAEKEKNFAKTYNHFLSSNIVKFAVSNQAEQINMELLSLKETQNKSILRNWSYYQLQTMIEYKAQREGIKVKY
IDPYHTSQTCSKCGNYEEGQRESQADFICKKCGYKVNADYNAARNIAMSNKYITKKEESKYYKIKESMV*

Appendix 3. Plasmids used in this study.

Plasmid name	Description	Figures	Link
PV424	Mi1Cas12f2 engineered and intact locus (native expression)	3.2B, E	https://benchling.com/s/seq-uDaLUdexDYQQSz7QDMXF
PV477	Disrupted Mi1Cas12f2 (native expression)	3.2B	https://benchling.com/s/seq-y7Fh5ryHUNNX7R3sDYMa
R-652	Mi1Cas12f2 intact locus (T7 expression)	3.2E	https://benchling.com/s/seq-DNGdXS8DCheZt5Nv6kMu
R-656	Mi1Cas12f2 minimal locus (T7 expression)	3.2E	https://benchling.com/s/seq-kUL4zwpBZDDJIwIwwLBT
R-658	Minus Mi1Cas12f2 (T7 expression)	3.2E	https://benchling.com/s/seq-A2E4WYBe0vLNbHiWQ2py
pLBH531	10×His-MBP-Un1Cas12f1 (Cas14a1) expression	3.4	https://www.addgene.org/112500/
pLBH545	Un1Cas12f1 (Cas14a1) locus (tetracycline inducible expression)	-	https://www.addgene.org/112501/
pGB53	Un1Cas12f1 and gRNA expression (pLBH545-based; T7 and tetracycline inducible expression)	3.5	https://benchling.com/s/seq-QoHpAbpI97JLSnhzh8Mn

pGB49	10×His-MBP- Un1Cas12f1 D326A expression (pLBH531-based)	3.4B	https://benchling.com/s/seq-qpUDPxM6IBhTJFXpNVUJ
pGB50	10×His-MBP- Un1Cas12f1 D510A expression (pLBH531-based)	3.4B	https://benchling.com/s/seq-W1miOvRPGZ44fgZnV0Nn
pUn2Cas12f1-pETduet-1	Un2Cas12f1 intact locus (T7 expression)	3.3, 3.5	https://benchling.com/s/seq-6PdzgqR721Th21GAzyM7
pMi2Cas12f2-pETduet-1	Mi2Cas12f2 intact locus (T7 expression)	3.3, 3.5	https://benchling.com/s/seq-QJzxQ4p6OonoQ8ZszpF7
pAuCas12f2-pETduet-1	AuCas12f2 intact locus (T7 expression)	3.3, 3.5	https://benchling.com/s/seq-6lJN8dk7ZTmbimkqw7Dg
pPtCas12f1-pETduet-1	PtCas12f1 intact locus (T7 expression)	3.3, 3.5	https://benchling.com/s/seq-qaU17VDPHblLUii1K5QZ
pAsCas12f1-pETduet-1	AsCas12f1 intact locus (T7 expression)	3.3, 3.5, 3.6B-C	https://benchling.com/s/seq-JWWCYPn66yxJl12WMBMX
pRuCas12f1-pETduet-1	RuCas12f1 intact locus (T7 expression)	3.3, 3.5	https://benchling.com/s/seq-XFA0y65xFCT57R719yVV
pSpCas12f1-pETduet-1	SpCas12f1 intact locus (T7 expression)	3.3, 3.5, 3.6B-C	https://benchling.com/s/seq-7LSUEFWIvEQk2AMpvsGA
pCnCas12f1-pETduet-1	CnCas12f1 intact locus (T7 expression)	3.3, 3.5	https://benchling.com/s/seq-92ITTuoZVYrJfN1hYgX1

pTZ57	7N PAM plasmid library	3.2, 3.3	https://benchling.com/s/seq-nu2IvfXbn7smVQ7T6MYi
pGB33	Mi1Cas12f2 target plasmid (pUC18-based)	3.2E	https://benchling.com/s/seq-iYcV6jflHOPbUxMdCGs9
pGB40	Un1Cas12f1 target plasmid (pUC18-based)	3.4A-C	https://benchling.com/s/seq-XGplLg5diY1G7BwBDHXn
pGB41	Un1Cas12f1 ΔPAM (target plasmid) (pUC18-based)	3.4A	https://benchling.com/s/seq-QS7mA2qo4e3JRlBvKsXb
pGB42	Un1Cas12f1 non-target plasmid (pUC18-based)	3.4A	https://benchling.com/s/seq-rXjk6jlGlmPUSbb2GeT9
pTHSSe_1	Un1Cas12f1 and Mi1Cas12f2 non-target plasmid (pSC101 ori)	3.5	https://www.addgene.org/109233/
pGB43	Un1Cas12f1 target plasmid (pTHSSe_1-based)	3.5	https://benchling.com/s/seq-Xt674hLUcgdIn2BO0tJ1
pKP17	Mi1Cas12f2 target plasmid (pTHSSe_1-based)	3.5	https://benchling.com/s/seq-g4SqeTyluv6C5FEwfmfS
pSG4K5	Un2Cas12f1, Mi2Cas12f2, AuCas12f2, PtCas12f1, AsCas12f1, RuCas12f1, SpCas12f1 and CnCas12f1 non-target plasmid (pSC101 ori)	3.5	https://www.addgene.org/74492/

pKP8	Un2Cas12f1, Mi2Cas12f2, AuCas12f2, PtCas12f1 and AsCas12f1 target plasmid (pSG4K5-based)	3.5	https://benchling.com/s/seq-P3Bzoe526M7vh0Cv3mtD
pKP9	RuCas12f1 target plasmid (pSG4K5-based)	3.5	https://benchling.com/s/seq-8JkUE7ndI46oxuapcPWS
pKP10	SpCas12f1 target plasmid (pSG4K5-based)	3.5	https://benchling.com/s/seq-dIsHC0DVhRF4z8bLWdbw
pKP11	CnCas12f1 target plasmid (pSG4K5-based)	3.5	https://benchling.com/s/seq-6t7yhm7Vs6FwSj2N2qOu
pMBP-SpCas12f1	SpCas12f1 expression	3.8-3.12A	https://benchling.com/s/seq-a5RJPUjSE8qamkW8plaq
pMBP-AsCas12f1	AsCas12f1 expression	3.8- 3.12B	https://benchling.com/s/seq-aT9kTNS2MI3XM29YkUKz
pGB-070	SpCas12f1 D228A expression	3.9-3.12A, 3.13A-B, 3.14A	https://benchling.com/s/seq-JpMFdgoZnZHqMlVmlaAS
pGB-069	AsCas12f1 D225A expression	3.9-3.12B, 3.13C-D, 3.14B	https://benchling.com/s/seq-kWfpv4nv5Qt6tvqxVKJh
pKP14	Target plasmid for SpCas12f1	3.8-3.9A	https://benchling.com/s/seq-K1h6yATOwXhuZS3U9TpR
pKP12	Target plasmid for AsCas12f1	3.8-3.9B	https://benchling.com/s/seq-zddimtDpD66mX5kb0vuz

pRZ-101-AsCas12a-NT	Plasmid for AsCas12a and non-targeting gRNA expression in human cells	3.15B	https://benchling.com/s/seq-Jn2n10uAKcuIcQwPcyMw
pRZ-162-AsCas12a-VEGFA2	Plasmid for AsCas12a and VEGFA2 gRNA expression in human cells	3.15B	https://benchling.com/s/seq-qQYLPhsJNBC87KWS4yOr
pRZ-102-AsCas12a-VEGFA3	Plasmid for AsCas12a and VEGFA3 gRNA expression in human cells	3.15B	https://benchling.com/s/seq-1oXgHfDSFcWBXbd0gmiF
pRZ-104-AsCas12a-DNMT1	Plasmid for AsCas12a and DNMT1 gRNA expression in human cells	3.15B	https://benchling.com/s/seq-OMPNs5you5aCgpJFqMhG
pRZ-105-AsCas12f1-NT	Plasmid for AsCas12f1 and non-targeting gRNA expression in human cells	3.15B	https://benchling.com/s/seq-vJLmdTxj6Yqx9uBfCohs
pRZ-107-AsCas12f1-VEFGA2	Plasmid for AsCas12f1 and VEGFA2 gRNA expression in human cells	3.15B	https://benchling.com/s/seq-i8ijGcGmJOZUDzNTLk7Q
pRZ-106-AsCas12f1-VEGFA3	Plasmid for AsCas12f1 and VEGFA3 gRNA expression in human cells	3.15B	https://benchling.com/s/seq-OoicdRiNYQjmlbBfPC5M
pRZ-108-AsCas12f1-DNMT1	Plasmid for AsCas12f1 and DNMT1 gRNA expression in human cells	3.15B	https://benchling.com/s/seq-w2sr4NYCFEPkdVMjvvwC
pRZ-109-SpCas12f1-NT	Plasmid for SpCas12f1 and non-targeting gRNA expression in human cells	3.15B	https://benchling.com/s/seq-fsEF6XOnwpnY0iwQ2370
pRZ-111-SpCas12f1-VEGFA2	Plasmid for SpCas12f1 and VEGFA2 gRNA expression in human cells	3.15B	https://benchling.com/s/seq-2kh3jwkJBUrUKs3E43EX

pRZ-110-SpCas12f1-VEGFA3	Plasmid for SpCas12f1 and VEGFA3 gRNA expression in human cells	3.15B	https://benchling.com/s/seq-7wgKEHkf2IKsBOgmjbZs
pRZ-112-SpCas12f1-DNMT1	Plasmid for SpCas12f1 and DNMT1 gRNA expression in human cells	3.15B	https://benchling.com/s/seq-YPXSi87VxSNYN3nbCZLm
RV039055	Plasmid for SpCas12f1 protein expression in *Zea mays*	3.16	https://benchling.com/s/seq-XdKgJhCh8cyRmREsgArD
RV008870_ms26	Plasmid for SpCas12f1 ms26 target gRNA expression in *Zea mays*	3.16	https://benchling.com/s/seq-KrjiLkg1lu0VlLvzmNCE
RV008870_waxy	Plasmid for SpCas12f1 waxy target gRNA expression in *Zea mays*	3.16	https://benchling.com/s/seq-keKAvQu27WPH6p2SyzL4
RV035712	Plasmid for SpCas9 protein expression in *Zea mays*	3.16C	https://benchling.com/s/seq-6ORzhzv2gwL0yyGJNvtc
RV008870-Cas9_ms26	Plasmid for SpCas9 ms26 target gRNA expression in *Zea mays*	3.16C	https://benchling.com/s/seq-m7Z4MaRuVT8rFZfev76l
RV022942	Plasmid for SpCas9 waxy target gRNA expression in *Zea mays*	3.16C	https://benchling.com/s/seq-hVMIqViehRMxIlDLYOtC

Appendix 4. gRNAs used in this study.

Spacer, nt	Sequence 5'-3' (Linker, Target)	Figures
	Un1Cas12f1	
20	GGGCUUCACUGAUAAAGUGGAGAACCGCUUCACCAAAAGCUGUCCCUUAGGGGAUUAGAACUUGAGUGAAGGUGGGCUGC UUGCAUCAGCCUAAUGUCGAGAAGUGCUUUCUUCGGAAAGUAACCCUCGAAACAAAUUCAUUUUUCCUCUCCAAUUCUGC ACAAGAAAGUUGCAGAACCCGAAUAGACGAAUGAAGGAAUGCAACAGUUGACCCAACGUCGCCGG	3.4
	SpCas12f1	
18	GGGAUUUACUCUGUUUCGCGCGCCAGGGCAGUUAGGUGCCCUAAAAGAGCGAAGUGGCCGAAAGGAAAGGCUAACGCUUC UCUAACGCUACGGCGACCUUGGCGAAAUGCCAUCAAUACCACGCGGCCCGAAAGGGUUCGCGCGAAACUGAGUAAUGAAA GUCGCAUCUUGCGUAAGCGCGUGGAUUGAAACAGUUGACCCAACGUCGCC	3.8-3.12A, 3.13A-B, 3.14A
12, 14, 16, 20, 22, 24, 30	GGGAUUUACUCUGUUUCGCGCGCCAGGGCAGUUAGGUGCCCUAAAAGAGCGAAGUGGCCGAAAGGAAAGGCUAACGCUUC UCUAACGCUACGGCGACCUUGGCGAAAUGCCAUCAAUACCACGCGGCCCGAAAGGGUUCGCGCGAAACUGAGUAAUGAAA GUCGCAUCUUGCGUAAGCGCGUGGAUUGAAACAGUUGACCCAACGUCGCCGGCGUGCACAAU	3.8A
	AsCas12f1	
18	GGGAUACUUCUAUUCGUCGGUUCAGCGACGAUAAGCCGAGAAGUGCCAAUAAAACUGUUAAGUGGUUUGGUAACGCUCGG UAAGGUAGCCAAAAGGCUGAAACUCCGUGCACAAAGACCGCACGGACGCUUCACAUAUAGCUCAUAAACAAAUGUCGUCG ACCUCUAAUAGCGAAAGUUUGCGAGCUAGCUUGUGGAGUGUGAACAGUUGACCCAACGUCGCC	3.8- 3.12B, 3.13C-D, 3.14B
12, 14, 16, 20, 22, 24, 30	GGGAUACUUCUAUUCGUCGGUUCAGCGACGAUAAGCCGAGAAGUGCCAAUAAAACUGUUAAGUGGUUUGGUAACGCUCGG UAAGGUAGCCAAAAGGCUGAAACUCCGUGCACAAAGACCGCACGGACGCUUCACAUAUAGCUCAUAAACAAAUGUCGUCG ACCUCUAAUAGCGAAAGUUUGCGAGCUAGCUUGUGGAGUGUGAACAGUUGACCCAACGUCGCCGGCGUGCACAAU	3.8B

Appendix 5. Oligonucleotides and DNA activators used for different cleavage assays in this study.

Description	Sequence 5'-3' (PAM, Target)	Figures
	Un1Cas12f1	
Target (forward)	GCACCTTACGTATTTAAGTTGACCCAACGTCGCCGGCGTGCACAATCTAGATGCATCAGC	3.4D
Target (reverse)	GCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTTAAATACGTAAGGTGC	3.4D-E
Control target (reverse)	GCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTATTTTACGTAAGGTGC	3.4E
Forward strand marker	GCACCTTACGTATTTAAGTTGACCCAACGTCGCCGGCGTGCACAATCTAG	3.4
	GCACCTTACGTATTTAAGTTGACCCAACGTCGCCGGCGTG	
	GCACCTTACGTATTTAAGTTGACCCAACGT	
	GCACCTTACGTATTTAAGTT	
	GCACCTTACG	
Reverse strand marker	GCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTTAAATA	3.4
	GCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTC	
	GCTGATGCATCTAGATTGTGCACGCCGGCG	

	GCTGATGCATCTAGATTGTG	3.4
	GCTGATGCAT	
ssDNA activator	GCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTTAAATACGTAAGGTGC	3.4F
dsDNA activator	GCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTAGTATTTAAGTTGACCCAACGTC GCCGGCGTGCACAATCTAGATGCATAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGC TCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTC ACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGG CCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGC	3.4F
	SpCas12f1	
Target (forward)	GCACCTTACTGCAAGGTAGCCACGTTCAGTTGACCCAACGTCGCCGGCGTGCACAATCTAGATGCATCAGCTGC	3.10-3.11A, 3.13A
Target (reverse)	GCAGCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTGAACGTGGCTACCTTGCAGTAAGGTGC	3.10-3.11A, 3.13A
ssDNA activator	GCAGCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTGAACGTGGCTACCTTGCAGTAAGGTGC	3.12A
dsDNA activator	GCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCAATTCCA CGTTCAGTTGACCCAACGTCGCCGGCGTGCACAATCTAGATGCTAGGACTCTAGAGGATCCCCGGGTACCGAGC TCGAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACG AGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCAC TGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGT TTGCGTATTGGGCGCTCTTCCGCTTCCTCGC	3.12A

dsDNA used for mass photometry assay	AAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGG CCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTT TCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCAATTCCACGTTCAGTTG ACCCAACGTCGCCGGCGTGCACAATCTAGATGCTAGGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGT AATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGC ATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTG	3.14A
	AsCas12f1	
Target (forward)	GCACCTTACTGCAAGGTAGCGTATTTAAGTTGACCCAACGTCGCCGGCGTGCACAATCTAGATGCATCAGCTGC	3.10-3.11B, 3.13B
Target (reverse)	GCAGCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTTAAATACGCTACCTTGCAGTAAGGTGC	3.10-3.11B, 3.13B
ssDNA activator	GCAGCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTTAAATACGCTACCTTGCAGTAAGGTGC	3.12B
dsDNA activator	GCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCAATTCGA TTTTAAGTTGACCCAACGTCGCCGGCGTGCACAATCTAGATGCTAGGACTCTAGAGGATCCCCGGGTACCGAGC TCGAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACG AGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCAC TGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGT TTGCGTATTGGGCGCTCTTCCGCTTCCTCGC	3.12B
dsDNA used for mass photometry assay	AAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGG CCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTT	3.14B

	TCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTAGTATTTAAGTTGACCCAACGTCGCCGGCGTGC ACAATCTAGATGCATAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCC ACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTG	
DNA control and markers for SpCas12f1 and AsCas12f1		
Control target (forward)	GCACCTTACTGCAAGGTAGCGTAAAATAGTTGACCCAACGTCGCCGGCGTGCACAATCTAGATGCATCAGCTGC	3.10, 3.11, 3.13A,C
Control target (reverse)	GCAGCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTATTTTACGCTACCTTGCAGTAAGGTGC	3.10, 3.11, 3.13A,C
Forward strand marker	GCACCTTACTGCAAGGTAGCGTATTTAAGTTGACCCAACGTCGCCGGCGTGCACAATCTA	3.10, 3.11
	GCACCTTACTGCAAGGTAGCGTATTTAAGTTGACCCAACGTCGCCGGCGT	
	GCACCTTACTGCAAGGTAGCGTATTTAAGTTGACCCAACGTCGCC	
	GCACCTTACTGCAAGGTAGCGTATTTAAGTTGACCCAACG	
	GCACCTTACTGCAAGGTAGCGTATTTAAGTTGACC	
	GCACCTTACTGCAAGGTAGCGTATTTAAGT	
	GCACCTTACTGCAAGGTAGC	
	GCACCTTACT	

Reverse strand marker	GCAGCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTTAAATACGCTACC	3.10, 3.11
	GCAGCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGGGTCAACTTAA	
	GCAGCTGATGCATCTAGATTGTGCACGCCGGCGACGTTGG	
	GCAGCTGATGCATCTAGATTGTGCACGCCG	
	GCAGCTGATGCATCTAGATTGTGCA	
	GCAGCTGATGCATCTAGATT	
	GCAGCTGATGCATCT	
	GCAGCTGATG	

Appendix 6. Primers used in assessing human and maize cell editing.

The blue font represents the sequence required for Illumina sequencing, the green font shows the additional sequence included to ensure balanced read composition in initial rounds of Illumina sequencing and the black font indicates the sequence responsible for annealing to the genomic locus.

Description	Designation within set	Sequence 5'-3'
Primary PCR VEGFA3 (SpCas12f1, AsCas12f1, AsCas12a)	NGS-F1	ATCGGGAAGCTGAAGATCTGCGTTTGGGAGGTCAGAAATA
	NGS-F2	ATCGGGAAGCTGAAGGCTCTAGTTTGGGAGGTCAGAAATA
	NGS-F3	ATCGGGAAGCTGAAGCGAGATGTTTGGGAGGTCAGAAATA
	NGS-F4	ATCGGGAAGCTGAAGTAGACGGTTTGGGAGGTCAGAAATA
	NGS-R	ATCCGACGGTAGTGTGACGTCCTCACTCTCGAAG
Primary PCR DNMT1 (SpCas12f1)	NGS-F1	ATCGGGAAGCTGAAGATCTGCCTCAAGTGAGCAGCTGAG
	NGS-F2	ATCGGGAAGCTGAAGGCTCTACTCAAGTGAGCAGCTGAG
	NGS-F3	ATCGGGAAGCTGAAGCGAGATCTCAAGTGAGCAGCTGAG
	NGS-F4	ATCGGGAAGCTGAAGTAGACGNCTCAAGTGAGCAGCTGAG
	NGS-R	ATCCGACGGTAGTGTGGTGAGGATTGAGTGAGTT
Primary PCR DNMT1 (AsCas12f1, AsCas12a)	NGS-F1	ATCGGGAAGCTGAAGATCTGCTGAACACTCCTCAAACGGTC
	NGS-F2	ATCGGGAAGCTGAAGGCTCTATGAACACTCCTCAAACGGTC

	NGS-F3	ATCGGGAAGCTGAAGCGAGATTGAACACTCCTCAAACGGTC
	NGS-F4	ATCGGGAAGCTGAAGTAGACGTGAACACTCCTCAAACGGTC
	NGS-R	ATCCGACGGTAGTGTGCCCTCACTCCTGCTCGGT
Primary PCR VEGFA2 (SpCas12f1)	NGS-F1	ATCGGGAAGCTGAAGATCTGCAGCTACCACCTCCTCCCCGGC
	NGS-F2	ATCGGGAAGCTGAAGGCTCTAAGCTACCACCTCCTCCCCGGC
	NGS-F3	ATCGGGAAGCTGAAGCGAGATAGCTACCACCTCCTCCCCGGC
	NGS-F4	ATCGGGAAGCTGAAGTAGACGAGCTACCACCTCCTCCCCGGC
	NGS-R	ATCCGACGGTAGTGTGCGGCTCCTCCGAAGCGAGAAC
Primary PCR VEGFA2 (AsCas12f1 and AsCas12a)	NGS-F1	ATCGGGAAGCTGAAGATCTGCGGGCGTGCGAGCAGCGAAAG
	NGS-F2	ATCGGGAAGCTGAAGGCTCTAGGGCGTGCGAGCAGCGAAAG
	NGS-F3	ATCGGGAAGCTGAAGCGAGATGGGCGTGCGAGCAGCGAAAG
	NGS-F4	ATCGGGAAGCTGAAGTAGACGGGGCGTGCGAGCAGCGAAAG
	NGS-R	ATCCGACGGTAGTGTGTGTCCGTCAGCGCGACTGGT
Primary PCR *ms26* (SpCas12f1 and SpCas9	NGS-F1	ATCGGGAAGCTGAAGATCTGCCGACGCGGCGAGCTTCCG
	NGS-F2	ATCGGGAAGCTGAAGGCTCTACGACGCGGCGAGCTTCCG

	NGS-F3	ATCGGGAAGCTGAAGCGAGATCGACGCGGCGAGCTTCCG
	NGS-F4	ATCGGGAAGCTGAAGTAGACGCGACGCGGCGAGCTTCCG
	NGS-R	ATCCGACGGTAGTGTTCATGCGGTACTGCACCGGGT
Primary PCR *waxy* (SpCas12f1 and SpCas9)	NGS-F1	ATCGGGAAGCTGAAGATCTGCGACGTCGTGTTCGTCTGCAAC
	NGS-F2	ATCGGGAAGCTGAAGGCTCTAGACGTCGTGTTCGTCTGCAAC
	NGS-F3	ATCGGGAAGCTGAAGCGAGATGACGTCGTGTTCGTCTGCAAC
	NGS-F4	ATCGGGAAGCTGAAGTAGACGGACGTCGTGTTCGTCTGCAAC
	NGS-R	ATCCGACGGTAGTGTTCTGGTAGGTAGTACGTGAAGATGGT
Universal secondary PCR forward primer		AATGATACGGCGACCACCGAGATCTACACATACGAGATCCGTAATCGGGAAGCTGAAG
Universal secondary PCR reverse primer ("N" indicates variable 8 bp index sequence for sample deconvolution)		CAAGCAGAAGACGGCATACGAGATNNNNNNNNACACGCACGATCCGACGGTAGTGT
Amplicon sequencing primer		CATACGAGATCCGTAATCGGGAAGCTGAAG
Index sequencing primer		ACACTACCGTCGGATCGTGCGTGT

Appendix 7. gRNA targets used in human and maize cell editing.

Protein	Target Name	Target sequence (5'-3')
	For genome editing in human cells	
AsCas12a	Non-target (NT)	AGTTGACCCAACGTCGCCGGCGT
	VEGFA3	GCCAGAGCCGGGGTGTGCAGACG
	VEGFA2	AGTGCGACGCCGCGAGCCCCGAC
	DNMT1	GCTCAGCAGGCACCTGCCTCAGC
AsCas12f1	Non-target (NT)	AGTTGACCCAACGTCGCCGG
	VEGFA3	GCCAGAGCCGGGGTGTGCAG
	VEGFA2	GGGGTGACCGCCGGAGCGCG
	DNMT1	GCTCAGCAGGCACCTGCCTC
SpCas12f1	Non-target (NT)	AGTTGACCCAACGTCGCCGG
	VEGFA3	CCTCTTTAGCCAGAGCCGGG
	VEGFA2	AGTGCGACGCCGCGAGCCCC
	DNMT1	CCTCACTCCTGCTCGGTGAA

For genome editing in maize cells		
SpCas12f1	*ms26*	AAGTTCACGGCGTTCCAGGC
	waxy	AGTTCAGAGAAGGCAACCTT
SpCas9	*ms26*	AAGTTCACGGCGTTCCAGGC
	waxy	GGCATCTACAGGGACGCAA

REFERENCES

1. Abudayyeh, Omar O., Jonathan S. Gootenberg, Silvana Konermann, Julia Joung, Ian M. Slaymaker, David B. T. Cox, Sergey Shmakov, Kira S. Makarova, Ekaterina Semenova, Leonid Minakhin, Konstantin Severinov, Aviv Regev, Eric S. Lander, Eugene V. Koonin, and Feng Zhang. 2016. 'C2c2 Is a Single-Component Programmable RNA-Guided RNA-Targeting CRISPR Effector'. *Science* 353(6299). doi: 10.1126/science.aaf5573.
2. Altae-Tran, Han, Soumya Kannan, F. Esra Demircioglu, Rachel Oshiro, Suchita P. Nety, Luke J. McKay, Mensur Dlakić, William P. Inskeep, Kira S. Makarova, Rhiannon K. Macrae, Eugene V. Koonin, and Feng Zhang. 2021. 'The Widespread IS200/605 Transposon Family Encodes Diverse Programmable RNA-Guided Endonucleases'. *Science*. doi: 10.1126/science.abj6856.
3. Amabile, Angelo, Alessandro Migliara, Paola Capasso, Mauro Biffi, Davide Cittaro, Luigi Naldini, and Angelo Lombardo. 2016. 'Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing'. *Cell* 167(1):219-232.e14. doi: 10.1016/j.cell.2016.09.006.
4. Amitai, Gil, and Rotem Sorek. 2016. 'CRISPR–Cas Adaptation: Insights into the Mechanism of Action'. *Nature Reviews Microbiology* 14(2):67–76. doi: 10.1038/nrmicro.2015.14.
5. Anders, Carolin, Katja Bargsten, and Martin Jinek. 2016. 'Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9'. *Molecular Cell* 61(6):895–902. doi: 10.1016/j.molcel.2016.02.020.
6. Anders, Carolin, Ole Niewoehner, Alessia Duerst, and Martin Jinek. 2014. 'Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease'. *Nature* 513(7519):569–73. doi: 10.1038/nature13579.
7. Anzalone, Andrew V., Luke W. Koblan, and David R. Liu. 2020. 'Genome Editing with CRISPR–Cas Nucleases, Base Editors, Transposases and Prime Editors'. *Nature Biotechnology* 38(7):824–44. doi: 10.1038/s41587-020-0561-9.
8. Anzalone, Andrew V., Peyton B. Randolph, Jessie R. Davis, Alexander A. Sousa, Luke W. Koblan, Jonathan M. Levy, Peter J. Chen, Christopher Wilson, Gregory A. Newby, Aditya Raguram, and David R. Liu. 2019. 'Search-and-Replace Genome Editing without Double-Strand Breaks or Donor DNA'. *Nature* 576(7785):149–57. doi: 10.1038/s41586-019-1711-4.
9. Banno, Satomi, Keiji Nishida, Takayuki Arazoe, Hitoshi Mitsunobu, and Akihiko Kondo. 2018. 'Deaminase-Mediated Multiplex Genome Editing in Escherichia Coli'. *Nature Microbiology* 3(4):423–29. doi: 10.1038/s41564-017-0102-6.
10. Bao, Weidong, and Jerzy Jurka. 2013. 'Homologues of Bacterial TnpB_IS605 Are Widespread in Diverse Eukaryotic Transposable Elements'. *Mobile DNA* 4:12. doi: 10.1186/1759-8753-4-12.
11. Barone, Pierluigi, Emily Wu, Brian Lenderts, Ajith Anand, William Gordon-Kamm, Sergei Svitashev, and Sandeep Kumar. 2020. 'Efficient Gene Targeting in Maize Using Inducible CRISPR-Cas9 and Marker-Free Donor Template'. *Molecular Plant* 13(8):1219–27. doi: 10.1016/j.molp.2020.06.008.
12. Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. 2007. 'CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes'. *Science* 315(5819):1709–12. doi: 10.1126/science.1138140.

13. Broughton, James P., Xianding Deng, Guixia Yu, Clare L. Fasching, Venice Servellita, Jasmeet Singh, Xin Miao, Jessica A. Streithorst, Andrea Granados, Alicia Sotomayor-Gonzalez, Kelsey Zorn, Allan Gopez, Elaine Hsu, Wei Gu, Steve Miller, Chao-Yang Pan, Hugo Guevara, Debra A. Wadford, Janice S. Chen, and Charles Y. Chiu. 2020. 'CRISPR–Cas12-Based Detection of SARS-CoV-2'. *Nature Biotechnology* 38(7):870–74. doi: 10.1038/s41587-020-0513-4.
14. Brouns, S. J. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. H. Slijkhuis, A. P. L. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, and J. van der Oost. 2008. 'Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes'. *Science* 321(5891):960–64. doi: 10.1126/science.1159689.
15. Burstein, David, Lucas B. Harrington, Steven C. Strutt, Alexander J. Probst, Karthik Anantharaman, Brian C. Thomas, Jennifer A. Doudna, and Jillian F. Banfield. 2017. 'New CRISPR-Cas Systems from Uncultivated Microbes.' *Nature* 542(7640):237–41. doi: 10.1038/nature21059.
16. Charlesworth, Carsten T., Priyanka S. Deshpande, Daniel P. Dever, Joab Camarena, Viktor T. Lemgart, M. Kyle Cromer, Christopher A. Vakulskas, Michael A. Collingwood, Liyang Zhang, Nicole M. Bode, Mark A. Behlke, Beruh Dejene, Brandon Cieniewicz, Rosa Romano, Benjamin J. Lesch, Natalia Gomez-Ospina, Sruthi Mantri, Mara Pavel-Dinu, Kenneth I. Weinberg, and Matthew H. Porteus. 2019. 'Identification of Preexisting Adaptive Immunity to Cas9 Proteins in Humans'. *Nature Medicine* 25(2):249–54. doi: 10.1038/s41591-018-0326-x.
17. Chen, Janice S., and Jennifer A. Doudna. 2017. 'The Chemistry of Cas9 and Its CRISPR Colleagues'. *Nature Reviews Chemistry* 1(10):1–15. doi: 10.1038/s41570-017-0078.
18. Chen, Janice S., Enbo Ma, Lucas B. Harrington, Maria Da Costa, Xinran Tian, Joel M. Palefsky, and Jennifer A. Doudna. 2018. 'CRISPR-Cas12a Target Binding Unleashes Indiscriminate Single-Stranded DNase Activity'. *Science* 360(6387):436–39. doi: 10.1126/science.aar6245.
19. Chen, Sway P., and Harris H. Wang. 2019. 'An Engineered Cas-Transposon System for Programmable and Site-Directed DNA Transpositions'. *The CRISPR Journal* 2(6):376–94. doi: 10.1089/crispr.2019.0030.
20. Chen, Weizhong, Ze-Hui Ren, Na Tang, Guoshi Chai, Hongyuan Zhang, Yifei Zhang, Jiacheng Ma, Zhaowei Wu, Xia Shen, Xingxu Huang, Guan-Zheng Luo, and Quanjiang Ji. 2021. 'Targeted Genetic Screening in Bacteria with a Cas12k-Guided Transposase'. *Cell Reports* 36(9):109635. doi: 10.1016/j.celrep.2021.109635.
21. Chew, Wei Leong, Mohammadsharif Tabebordbar, Jason K. W. Cheng, Prashant Mali, Elizabeth Y. Wu, Alex H. M. Ng, Kexian Zhu, Amy J. Wagers, and George M. Church. 2016. 'A Multifunctional AAV-CRISPR-Cas9 and Its Host Response'. *Nature Methods* 13(10):868–74. doi: 10.1038/nmeth.3993.
22. Choudhury, Samrat Roy, Yi Cui, Katarzyna Lubecka, Barbara Stefanska, and Joseph Irudayaraj. 2016. 'CRISPR-DCas9 Mediated TET1 Targeting for Selective DNA Demethylation at BRCA1 Promoter'. *Oncotarget* 7(29):46545–56. doi: 10.18632/oncotarget.10234.
23. Chylinski, Krzysztof, Kira S. Makarova, Emmanuelle Charpentier, and Eugene V. Koonin. 2014. 'Classification and Evolution of Type II CRISPR-Cas Systems'. *Nucleic Acids Research* 42(10):6091–6105. doi: 10.1093/nar/gku241.

24. Ciccia, Alberto, and Stephen J. Elledge. 2010. 'The DNA Damage Response: Making It Safe to Play with Knives'. *Molecular Cell* 40(2):179–204. doi: 10.1016/j.molcel.2010.09.019.
25. Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang. 2013. 'Multiplex Genome Engineering Using CRISPR/Cas Systems'. *Science* 339(6121):819–23. doi: 10.1126/science.1231143.
26. Deltcheva, Elitza, Krzysztof Chylinski, Cynthia M. Sharma, Karine Gonzales, Yanjie Chao, Zaid A. Pirzada, Maria R. Eckert, Jörg Vogel, and Emmanuelle Charpentier. 2011. 'CRISPR RNA Maturation by Trans-Encoded Small RNA and Host Factor RNase III'. *Nature* 471(7340):602–7. doi: 10.1038/nature09886.
27. Djukanovic, Vesna, Jeff Smith, Keith Lowe, Meizhu Yang, Huirong Gao, Spencer Jones, Michael G. Nicholson, Ande West, Janel Lape, Dennis Bidney, Saverio Carl Falco, Derek Jantz, and Leszek Alexander Lyznik. 2013. 'Male-Sterile Maize Plants Produced by Targeted Mutagenesis of the Cytochrome P450-like Gene (*MS26*) Using a Re-Designed I- *Cre* I Homing Endonuclease'. *The Plant Journal* 76(5):888–99. doi: 10.1111/tpj.12335.
28. Doench, John G., Ella Hartenian, Daniel B. Graham, Zuzana Tothova, Mudra Hegde, Ian Smith, Meagan Sullender, Benjamin L. Ebert, Ramnik J. Xavier, and David E. Root. 2014. 'Rational Design of Highly Active SgRNAs for CRISPR-Cas9–Mediated Gene Inactivation'. *Nature Biotechnology* 32(12):1262–67. doi: 10.1038/nbt.3026.
29. Dong, De, Kuan Ren, Xiaolin Qiu, Jianlin Zheng, Minghui Guo, Xiaoyu Guan, Hongnan Liu, Ningning Li, Bailing Zhang, Daijun Yang, Chuang Ma, Shuo Wang, Dan Wu, Yunfeng Ma, Shilong Fan, Jiawei Wang, Ning Gao, and Zhiwei Huang. 2016. 'The Crystal Structure of Cpf1 in Complex with CRISPR RNA'. *Nature* 532(7600):522–26. doi: 10.1038/nature17944.
30. Dong, J. Y., P. D. Fan, and R. A. Frizzell. 1996. 'Quantitative Analysis of the Packaging Capacity of Recombinant Adeno-Associated Virus'. *Human Gene Therapy* 7(17):2101–12. doi: 10.1089/hum.1996.7.17-2101.
31. Doudna, Jennifer A. 2020. 'The Promise and Challenge of Therapeutic Genome Editing'. *Nature* 578(7794):229–36. doi: 10.1038/s41586-020-1978-5.
32. East-Seletsky, Alexandra, Mitchell R. O'Connell, Spencer C. Knight, David Burstein, Jamie H. D. Cate, Robert Tjian, and Jennifer A. Doudna. 2016. 'Two Distinct RNase Activities of CRISPR-C2c2 Enable Guide-RNA Processing and RNA Detection'. *Nature* 538(7624):270–73. doi: 10.1038/nature19802.
33. Fan, Longjiang, Jiandong Bao, Yu Wang, Jianqiang Yao, Yijie Gui, Weiming Hu, Jinqing Zhu, Mengqian Zeng, Yu Li, and Yunbi Xu. 2009. 'Post-Domestication Selection in the Maize Starch Pathway'. *PLOS ONE* 4(10):e7612. doi: 10.1371/journal.pone.0007612.
34. Faure, Guilhem, Sergey A. Shmakov, Winston X. Yan, David R. Cheng, David A. Scott, Joseph E. Peters, Kira S. Makarova, and Eugene V. Koonin. 2019. 'CRISPR–Cas in Mobile Genetic Elements: Counter-Defence and Beyond'. *Nature Reviews Microbiology* 17(8):513–25. doi: 10.1038/s41579-019-0204-7.
35. Fine, Eli J., Caleb M. Appleton, Douglas E. White, Matthew T. Brown, Harshavardhan Deshmukh, Melissa L. Kemp, and Gang Bao. 2015. 'Trans-Spliced Cas9 Allows Cleavage of HBB and CCR5 Genes in Human Cells Using Compact Expression Cassettes'. *Scientific Reports* 5:10777. doi: 10.1038/srep10777.

36. Fonfara, Ines, Hagen Richter, Majda Bratovič, Anaïs Le Rhun, and Emmanuelle Charpentier. 2016. 'The CRISPR-Associated DNA-Cleaving Enzyme Cpf1 Also Processes Precursor CRISPR RNA'. *Nature* 532(7600):517–21. doi: 10.1038/nature17945.
37. Fu, Yanfang, Jennifer A. Foden, Cyd Khayter, Morgan L. Maeder, Deepak Reyon, J. Keith Joung, and Jeffry D. Sander. 2013. 'High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells'. *Nature Biotechnology* 31(9):822–26. doi: 10.1038/nbt.2623.
38. Gao, Linyi, David B. T. Cox, Winston X. Yan, John C. Manteiga, Martin W. Schneider, Takashi Yamano, Hiroshi Nishimasu, Osamu Nureki, Nicola Crosetto, and Feng Zhang. 2017. 'Engineered Cpf1 Variants with Altered PAM Specificities'. *Nature Biotechnology* 35(8):789–92. doi: 10.1038/nbt.3900.
39. Gao, Pu, Hui Yang, Kanagalaghatta R. Rajashankar, Zhiwei Huang, and Dinshaw J. Patel. 2016. 'Type V CRISPR-Cas Cpf1 Endonuclease Employs a Unique Mechanism for CrRNA-Mediated Target DNA Recognition'. *Cell Research* 26(8):901–13. doi: 10.1038/cr.2016.88.
40. Garneau, Josiane E., Marie-Ève Dupuis, Manuela Villion, Dennis A. Romero, Rodolphe Barrangou, Patrick Boyaval, Christophe Fremaux, Philippe Horvath, Alfonso H. Magadán, and Sylvain Moineau. 2010. 'The CRISPR/Cas Bacterial Immune System Cleaves Bacteriophage and Plasmid DNA'. *Nature* 468(7320):67–71. doi: 10.1038/nature09523.
41. Gasiunas, Giedrius, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. 2012. 'Cas9-CrRNA Ribonucleoprotein Complex Mediates Specific DNA Cleavage for Adaptive Immunity in Bacteria.' *Proceedings of the National Academy of Sciences* 109(39):2579–86. doi: 10.1073/pnas.1208507109.
42. Gasiunas, Giedrius, Joshua K. Young, Tautvydas Karvelis, Darius Kazlauskas, Tomas Urbaitis, Monika Jasnauskaite, Mantvyda M. Grusyte, Sushmitha Paulraj, Po-Hao Wang, Zhenglin Hou, Shane K. Dooley, Mark Cigan, Clara Alarcon, N. Doane Chilcoat, Greta Bigelyte, Jennifer L. Curcuru, Megumu Mabuchi, Zhiyi Sun, Ryan T. Fuchs, Ezra Schildkraut, Peter R. Weigele, William E. Jack, G. Brett Robb, Česlovas Venclovas, and Virginijus Siksnys. 2020. 'A Catalogue of Biochemically Diverse CRISPR-Cas9 Orthologs'. *Nature Communications* 11(1):5512. doi: 10.1038/s41467-020-19344-1.
43. Gaudelli, Nicole M., Alexis C. Komor, Holly A. Rees, Michael S. Packer, Ahmed H. Badran, David I. Bryson, and David R. Liu. 2017. 'Programmable Base Editing of A•T to G•C in Genomic DNA without DNA Cleavage'. *Nature* 551(7681):464–71. doi: 10.1038/nature24644.
44. Gaudelli, Nicole M., Dieter K. Lam, Holly A. Rees, Noris M. Solá-Esteves, Luis A. Barrera, David A. Born, Aaron Edwards, Jason M. Gehrke, Seung-Joo Lee, Alexander J. Liquori, Ryan Murray, Michael S. Packer, Conrad Rinaldi, Ian M. Slaymaker, Jonathan Yen, Lauren E. Young, and Giuseppe Ciaramella. 2020. 'Directed Evolution of Adenine Base Editors with Increased Activity and Therapeutic Application'. *Nature Biotechnology* 38(7):892–900. doi: 10.1038/s41587-020-0491-6.
45. Gilbert, Luke A., Matthew H. Larson, Leonardo Morsut, Zairan Liu, Gloria A. Brar, Sandra E. Torres, Noam Stern-Ginossar, Onn Brandman, Evan H. Whitehead, Jennifer A. Doudna, Wendell A. Lim, Jonathan S. Weissman, and Lei S. Qi. 2013. 'CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes'. *Cell* 154(2):442–51. doi: 10.1016/j.cell.2013.06.044.

46. Glass, Zachary, Matthew Lee, Yamin Li, and Qiaobing Xu. 2018. 'Engineering the Delivery System for CRISPR-Based Genome Editing'. *Trends in Biotechnology* 36(2):173–85. doi: 10.1016/j.tibtech.2017.11.006.
47. Gootenberg, Jonathan S., Omar O. Abudayyeh, Max J. Kellner, Julia Joung, James J. Collins, and Feng Zhang. 2018. 'Multiplexed and Portable Nucleic Acid Detection Platform with Cas13, Cas12a, and Csm6'. *Science* 360(6387):439–44. doi: 10.1126/science.aaq0179.
48. Gootenberg, Jonathan S., Omar O. Abudayyeh, Jeong Wook Lee, Patrick Essletzbichler, Aaron J. Dy, Julia Joung, Vanessa Verdine, Nina Donghia, Nichole M. Daringer, Catherine A. Freije, Cameron Myhrvold, Roby P. Bhattacharyya, Jonathan Livny, Aviv Regev, Eugene V. Koonin, Deborah T. Hung, Pardis C. Sabeti, James J. Collins, and Feng Zhang. 2017. 'Nucleic Acid Detection with CRISPR-Cas13a/C2c2'. *Science* 356(6336):438–42. doi: 10.1126/science.aam9321.
49. Gordon-Kamm, William, Brian P. Dilkes, Keith Lowe, George Hoerster, Xifan Sun, Margit Ross, Laura Church, Chris Bunde, Jeff Farrell, Patrea Hill, Sheila Maddock, Jane Snyder, Louisa Sykes, Zhongsen Li, Young-min Woo, Dennis Bidney, and Brian A. Larkins. 2002. 'Stimulation of the Cell Cycle and Maize Transformation by Disruption of the Plant Retinoblastoma Pathway'. *Proceedings of the National Academy of Sciences* 99(18):11975–80. doi: 10.1073/pnas.142409899.
50. Harrington, Lucas B., David Burstein, Janice S. Chen, David Paez-Espino, Enbo Ma, Isaac P. Witte, Joshua C. Cofsky, Nikos C. Kyrpides, Jillian F. Banfield, and Jennifer A. Doudna. 2018. 'Programmed DNA Destruction by Miniature CRISPR-Cas14 Enzymes'. *Science* 362(6416):839–42. doi: 10.1126/science.aav4294.
51. Harrington, Lucas B., Enbo Ma, Janice S. Chen, Isaac P. Witte, Dov Gertz, David Paez-Espino, Basem Al-Shayeb, Nikos C. Kyrpides, David Burstein, Jillian F. Banfield, and Jennifer A. Doudna. 2020. 'A ScoutRNA Is Required for Some Type V CRISPR-Cas Systems'. *Molecular Cell* S109727652030424X. doi: 10.1016/j.molcel.2020.06.022.
52. Heler, Robert, Poulami Samai, Joshua W. Modell, Catherine Weiner, Gregory W. Goldberg, David Bikard, and Luciano A. Marraffini. 2015. 'Cas9 Specifies Functional Viral Targets during CRISPR–Cas Adaptation'. *Nature* 519(7542):199–202. doi: 10.1038/nature14245.
53. Hochstrasser, Megan L., and Jennifer A. Doudna. 2015. 'Cutting It Close: CRISPR-Associated Endoribonuclease Structure and Function'. *Trends in Biochemical Sciences* 40(1):58–66. doi: 10.1016/j.tibs.2014.10.007.
54. Hoerster, George, Ning Wang, Larisa Ryan, Emily Wu, Ajith Anand, Kevin McBride, Keith Lowe, Todd Jones, and Bill Gordon-Kamm. 2020. 'Use of Non-Integrating Zm-Wus2 Vectors to Enhance Maize Transformation'. *In Vitro Cellular & Developmental Biology - Plant* 56(3):265–79. doi: 10.1007/s11627-019-10042-2.
55. Hu, Johnny H., Shannon M. Miller, Maarten H. Geurts, Weixin Tang, Liwei Chen, Ning Sun, Christina M. Zeina, Xue Gao, Holly A. Rees, Zhi Lin, and David R. Liu. 2018. 'Evolved Cas9 Variants with Broad PAM Compatibility and High DNA Specificity'. *Nature* 556(7699):57–63. doi: 10.1038/nature26155.
56. Jackson, Simon A., Rebecca E. McKenzie, Robert D. Fagerlund, Sebastian N. Kieper, Peter C. Fineran, and Stan J. J. Brouns. 2017. 'CRISPR-Cas: Adapting to Change'. *Science* 356(6333). doi: 10.1126/science.aal5056.

57. Jasin, M., and R. Rothstein. 2013. ‘Repair of Strand Breaks by Homologous Recombination’. *Cold Spring Harbor Perspectives in Biology* 5(11):a012740–a012740. doi: 10.1101/cshperspect.a012740.
58. Jeon, Yongmoon, You Hee Choi, Yunsu Jang, Jihyeon Yu, Jiyoung Goo, Gyejun Lee, You Kyeong Jeong, Seung Hwan Lee, In-San Kim, Jin-Soo Kim, Cherlhyun Jeong, Sanghwa Lee, and Sangsu Bae. 2018. ‘Direct Observation of DNA Target Searching and Cleavage by CRISPR-Cas12a’. *Nature Communications* 9(1):1–11. doi: 10.1038/s41467-018-05245-x.
59. Jiang, Fuguo, and Jennifer A. Doudna. 2017. ‘CRISPR–Cas9 Structures and Mechanisms’. *Annual Review of Biophysics* 46(1):505–29. doi: 10.1146/annurev-biophys-062215-010822.
60. Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. 2012. ‘A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity’. *Science* 337(6096):816–21. doi: 10.1126/science.1225829.
61. Jones, Stephen K., John A. Hawkins, Nicole V. Johnson, Cheulhee Jung, Kuang Hu, James R. Rybarski, Janice S. Chen, Jennifer A. Doudna, William H. Press, and Ilya J. Finkelstein. 2021. ‘Massively Parallel Kinetic Profiling of Natural and Engineered CRISPR Nucleases’. *Nature Biotechnology* 39(1):84–93. doi: 10.1038/s41587-020-0646-5.
62. Joung, Julia, Alim Ladha, Makoto Saito, Nam-Gyun Kim, Ann E. Woolley, Michael Segel, Robert P. J. Barretto, Amardeep Ranu, Rhiannon K. Macrae, Guilhem Faure, Eleonora I. Ioannidi, Rohan N. Krajeski, Robert Bruneau, Meei-Li W. Huang, Xu G. Yu, Jonathan Z. Li, Bruce D. Walker, Deborah T. Hung, Alexander L. Greninger, Keith R. Jerome, Jonathan S. Gootenberg, Omar O. Abudayyeh, and Feng Zhang. 2020. ‘Detection of SARS-CoV-2 with SHERLOCK One-Pot Testing’. *The New England Journal of Medicine* 383(15):1492–94. doi: 10.1056/NEJMc2026172.
63. Kapitonov, Vladimir V., Kira S. Makarova, and Eugene V. Koonin. 2015. ‘ISC, a Novel Group of Bacterial and Archaeal DNA Transposons That Encode Cas9 Homologs’. *Journal of Bacteriology* 198(5):797–807. doi: 10.1128/JB.00783-15.
64. Karvelis, Tautvydas, Gytis Druteika, Greta Bigelyte, Karolina Budre, Rimante Zedaveinyte, Arunas Silanskas, Darius Kazlauskas, Česlovas Venclovas, and Virginijus Siksnys. 2021. ‘Transposon-Associated TnpB Is a Programmable RNA-Guided DNA Endonuclease’. *Nature* 1–5. doi: 10.1038/s41586-021-04058-1.
65. Karvelis, Tautvydas, Giedrius Gasiunas, Algirdas Miksys, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. 2013. ‘CrRNA and TracrRNA Guide Cas9-Mediated DNA Interference in *Streptococcus Thermophilus*’. *RNA Biology* 10(5):841–51. doi: 10.4161/rna.24203.
66. Karvelis, Tautvydas, Giedrius Gasiunas, Joshua Young, Greta Bigelyte, Arunas Silanskas, Mark Cigan, and Virginijus Siksnys. 2015. ‘Rapid Characterization of CRISPR-Cas9 Protospacer Adjacent Motif Sequence Elements’. *Genome Biology* 16(1):253. doi: 10.1186/s13059-015-0818-7.
67. Karvelis, Tautvydas, Joshua K. Young, and Virginijus Siksnys. 2019. ‘A Pipeline for Characterization of Novel Cas9 Orthologs’. Pp. 219–40 in *Methods in Enzymology*. Vol. 616. Elsevier Inc.
68. Kazlauskiene, Migle, Georgij Kostiuk, Česlovas Venclovas, Gintautas Tamulaitis, and Virginijus Siksnys. 2017. ‘A Cyclic Oligonucleotide Signaling

Pathway in Type III CRISPR-Cas Systems'. *Science (New York, N.Y.)* 357(6351):605–9. doi: 10.1126/science.aao0100.
69. Khattri, Abhilasha, Soumen Nandy, and Vibha Srivastava. 2011. 'Heat-Inducible Cre-Lox System for Marker Excision in Transgenic Rice'. *Journal of Biosciences* 36(1):37–42. doi: 10.1007/s12038-011-9010-8.
70. Kim, Daesik, and Jin-Soo Kim. 2018. 'DIG-Seq: A Genome-Wide CRISPR off-Target Profiling Method Using Chromatin DNA'. *Genome Research* 28(12):1894–1900. doi: 10.1101/gr.236620.118.
71. Kim, Daesik, Jungeun Kim, Junho K. Hur, Kyung Wook Been, Sun-heui Yoon, and Jin-Soo Kim. 2016. 'Genome-Wide Analysis Reveals Specificities of Cpf1 Endonucleases in Human Cells'. *Nature Biotechnology* 34(8):863–68. doi: 10.1038/nbt.3609.
72. Kim, Daesik, Kevin Luk, Scot A. Wolfe, and Jin-Soo Kim. 2019. 'Evaluating and Enhancing Target Specificity of Gene-Editing Nucleases and Deaminases'. *Annual Review of Biochemistry* 88(1):191–220. doi: 10.1146/annurev-biochem-013118-111730.
73. Kim, Do Yon, Jeong Mi Lee, Su Bin Moon, Hyun Jung Chin, Seyeon Park, Youjung Lim, Daesik Kim, Taeyoung Koo, Jeong-Heon Ko, and Yong-Sam Kim. 2021. 'Efficient CRISPR Editing with a Hypercompact Cas12f1 and Engineered Guide RNAs Delivered by Adeno-Associated Virus'. *Nature Biotechnology*. doi: 10.1038/s41587-021-01009-z.
74. Kim, Nahye, Hui Kwon Kim, Sungtae Lee, Jung Hwa Seo, Jae Woo Choi, Jinman Park, Seonwoo Min, Sungroh Yoon, Sung-Rae Cho, and Hyongbum Henry Kim. 2020. 'Prediction of the Sequence-Specific Cleavage Activity of Cas9 Variants'. *Nature Biotechnology* 38(11):1328–36. doi: 10.1038/s41587-020-0537-9.
75. Kim, Seong Keun, Haseong Kim, Woo-Chan Ahn, Kwang-Hyun Park, Eui-Jeon Woo, Dae-Hee Lee, and Seung-Goo Lee. 2017. 'Efficient Transcriptional Gene Repression by Type V-A CRISPR-Cpf1 from Eubacterium Eligens'. *ACS Synthetic Biology* 6(7):1273–82. doi: 10.1021/acssynbio.6b00368.
76. Kleinstiver, Benjamin P., Vikram Pattanayak, Michelle S. Prew, Shengdar Q. Tsai, Nhu T. Nguyen, Zongli Zheng, and J. Keith Joung. 2016. 'High-Fidelity CRISPR–Cas9 Nucleases with No Detectable Genome-Wide off-Target Effects'. *Nature* 529(7587):490–95. doi: 10.1038/nature16526.
77. Kleinstiver, Benjamin P., Michelle S. Prew, Shengdar Q. Tsai, Ved V. Topkar, Nhu T. Nguyen, Zongli Zheng, Andrew P. W. Gonzales, Zhuyun Li, Randall T. Peterson, Jing-Ruey Joanna Yeh, Martin J. Aryee, and J. Keith Joung. 2015. 'Engineered CRISPR-Cas9 Nucleases with Altered PAM Specificities'. *Nature* 523(7561):481–85. doi: 10.1038/nature14592.
78. Kleinstiver, Benjamin P., Alexander A. Sousa, Russell T. Walton, Y. Esther Tak, Jonathan Y. Hsu, Kendell Clement, Moira M. Welch, Joy E. Horng, Jose Malagon-Lopez, Irene Scarfò, Marcela V. Maus, Luca Pinello, Martin J. Aryee, and J. Keith Joung. 2019. 'Engineered CRISPR-Cas12a Variants with Increased Activities and Improved Targeting Ranges for Gene, Epigenetic and Base Editing'. *Nature Biotechnology* 37(3):276–82. doi: 10.1038/s41587-018-0011-0.
79. Kleinstiver, Benjamin P., Shengdar Q. Tsai, Michelle S. Prew, Nhu T. Nguyen, Moira M. Welch, Jose M. Lopez, Zachary R. McCaw, Martin J. Aryee, and J. Keith Joung. 2016. 'Genome-Wide Specificities of CRISPR-Cas Cpf1 Nucleases in Human Cells'. *Nature Biotechnology* 34(8):869–74. doi: 10.1038/nbt.3620.

80. Klompe, Sanne E., Phuc L. H. Vo, Tyler S. Halpin-Healy, and Samuel H. Sternberg. 2019. 'Transposon-Encoded CRISPR–Cas Systems Direct RNA-Guided DNA Integration'. *Nature* 571(7764):219–25. doi: 10.1038/s41586-019-1323-z.
81. Knott, Gavin J., and Jennifer A. Doudna. 2018. 'CRISPR-Cas Guides the Future of Genetic Engineering'. *Science* 361(6405):866–69. doi: 10.1126/science.aat5011.
82. Komor, Alexis C., Ahmed H. Badran, and David R. Liu. 2017. 'CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes'. *Cell* 168(1–2):20–36. doi: 10.1016/j.cell.2016.10.044.
83. Komor, Alexis C., Yongjoo B. Kim, Michael S. Packer, John A. Zuris, and David R. Liu. 2016. 'Programmable Editing of a Target Base in Genomic DNA without Double-Stranded DNA Cleavage'. *Nature* 533(7603):420–24. doi: 10.1038/nature17946.
84. Koonin, Eugene V., and Mart Krupovic. 2015. 'Evolution of Adaptive Immunity from Transposable Elements Combined with Innate Immune Systems'. *Nature Reviews. Genetics* 16(3):184–92. doi: 10.1038/nrg3859.
85. Koonin, Eugene V., and Kira S. Makarova. 2017. 'Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back'. *Genome Biology and Evolution* 9(10):2812–25. doi: 10.1093/gbe/evx192.
86. Koonin, Eugene V., and Kira S. Makarova. 2019. 'Origins and Evolution of CRISPR-Cas Systems'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374(1772):20180087. doi: 10.1098/rstb.2018.0087.
87. Koonin, Eugene V., Kira S. Makarova, and Yuri I. Wolf. 2017. 'Evolutionary Genomics of Defense Systems in Archaea and Bacteria'. *Annual Review of Microbiology* 71(1):233–61. doi: 10.1146/annurev-micro-090816-093830.
88. Koonin, Eugene V, Kira S. Makarova, and Feng Zhang. 2017. 'Diversity, Classification and Evolution of CRISPR-Cas Systems'. *Current Opinion in Microbiology* 37:67–78. doi: 10.1016/j.mib.2017.05.008.
89. Kosicki, Michael, Kärt Tomberg, and Allan Bradley. 2018. 'Repair of Double-Strand Breaks Induced by CRISPR-Cas9 Leads to Large Deletions and Complex Rearrangements'. *Nature Biotechnology* 36(8):765–71. doi: 10.1038/nbt.4192.
90. Krupovic, Mart, Pierre Béguin, and Eugene V. Koonin. 2017. 'Casposons: Mobile Genetic Elements That Gave Rise to the CRISPR-Cas Adaptation Machinery'. *Current Opinion in Microbiology* 38:36–43. doi: 10.1016/j.mib.2017.04.004.
91. Krupovic, Mart, and Eugene V. Koonin. 2016. 'Self-Synthesizing Transposons: Unexpected Key Players in the Evolution of Viruses and Defense Systems'. *Current Opinion in Microbiology* 31:25–33. doi: 10.1016/j.mib.2016.01.006.
92. Krupovic, Mart, Kira S. Makarova, Patrick Forterre, David Prangishvili, and Eugene V. Koonin. 2014. 'Casposons: A New Superfamily of Self-Synthesizing DNA Transposons at the Origin of Prokaryotic CRISPR-Cas Immunity'. *BMC Biology* 12(1):36. doi: 10.1186/1741-7007-12-36.
93. Langmead, Ben, and Steven L. Salzberg. 2012. 'Fast Gapped-Read Alignment with Bowtie 2'. *Nature Methods* 9(4):357–59. doi: 10.1038/nmeth.1923.
94. Lau, Cia-Hin, and Yousin Suh. 2017. 'In Vivo Genome Editing in Animals Using AAV-CRISPR System: Applications to Translational Research of Human Disease'. *F1000Research* 6:2153. doi: 10.12688/f1000research.11243.1.

95. Li, Juan, Wunchang Sun, Bing Wang, Xiao Xiao, and Xiang-Qin Liu. 2008. 'Protein Trans-Splicing as a Means for Viral Vector-Mediated in Vivo Gene Therapy'. *Human Gene Therapy* 19(9):958–64. doi: 10.1089/hum.2008.009.
96. Li, Ming, Shivender M. D. Shandilya, Michael A. Carpenter, Anurag Rathore, William L. Brown, Angela L. Perkins, Daniel A. Harki, Jonathan Solberg, Derek J. Hook, Krishan K. Pandey, Michael A. Parniak, Jeffrey R. Johnson, Nevan J. Krogan, Mohan Somasundaran, Akbar Ali, Celia A. Schiffer, and Reuben S. Harris. 2012. 'First-in-Class Small Molecule Inhibitors of the Single-Strand DNA Cytosine Deaminase APOBEC3G'. *ACS Chemical Biology* 7(3):506–17. doi: 10.1021/cb200440y.
97. Li, Shi-Yuan, Qiu-Xiang Cheng, Jia-Kun Liu, Xiao-Qun Nie, Guo-Ping Zhao, and Jin Wang. 2018. 'CRISPR-Cas12a Has Both Cis- and Trans-Cleavage Activities on Single-Stranded DNA'. *Cell Research* 28(4):491–93. doi: 10.1038/s41422-018-0022-x.
98. Li, Xiaosa, Ying Wang, Yajing Liu, Bei Yang, Xiao Wang, Jia Wei, Zongyang Lu, Yuxi Zhang, Jing Wu, Xingxu Huang, Li Yang, and Jia Chen. 2018. 'Base Editing with a Cpf1-Cytidine Deaminase Fusion'. *Nature Biotechnology* 36(4):324–27. doi: 10.1038/nbt.4102.
99. Lieber, Michael R. 2010. 'The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway'. *Annual Review of Biochemistry* 79(1):181–211. doi: 10.1146/annurev.biochem.052308.093131.
100. Lin, C. Y., J. K. Roberts, and J. L. Key. 1984. 'Acquisition of Thermotolerance in Soybean Seedlings : Synthesis and Accumulation of Heat Shock Proteins and Their Cellular Localization'. *Plant Physiology* 74(1):152–60. doi: 10.1104/pp.74.1.152.
101. Lino, Christopher A., Jason C. Harper, James P. Carney, and Jerilyn A. Timlin. 2018. 'Delivering CRISPR: A Review of the Challenges and Approaches'. *Drug Delivery* 25(1):1234–57. doi: 10.1080/10717544.2018.1474964.
102. Liu, Hsiang-chin, and Yee-yung Charng. 2012. 'Acquired Thermotolerance Independent of Heat Shock Factor A1 (HsfA1), the Master Regulator of the Heat Stress Response'. *Plant Signaling & Behavior* 7(5):547–50. doi: 10.4161/psb.19803.
103. Liu, Jun-Jie, Natalia Orlova, Benjamin L. Oakes, Enbo Ma, Hannah B. Spinner, Katherine L. M. Baney, Jonathan Chuck, Dan Tan, Gavin J. Knott, Lucas B. Harrington, Basem Al-Shayeb, Alexander Wagner, Julian Brötzmann, Brett T. Staahl, Kian L. Taylor, John Desmarais, Eva Nogales, and Jennifer A. Doudna. 2019. 'CasX Enzymes Comprise a Distinct Family of RNA-Guided Genome Editors'. *Nature* 566(7743):218–23. doi: 10.1038/s41586-019-0908-x.
104. Liu, Liang, Peng Chen, Min Wang, Xueyan Li, Jiuyu Wang, Maolu Yin, and Yanli Wang. 2017. 'C2c1-SgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism'. *Molecular Cell* 65(2):310–22. doi: 10.1016/j.molcel.2016.11.040.
105. Liu, Liang, Xueyan Li, Jiuyu Wang, Min Wang, Peng Chen, Maolu Yin, Jiazhi Li, Gang Sheng, and Yanli Wang. 2017. 'Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities'. *Cell* 168(1–2):121-134.e12. doi: 10.1016/j.cell.2016.12.031.
106. Lowe, Keith, Mauricio La Rota, George Hoerster, Craig Hastings, Ning Wang, Mark Chamberlin, Emily Wu, Todd Jones, and William Gordon-Kamm. 2018. 'Rapid Genotype "Independent" Zea Mays L. (Maize) Transformation via Direct

Somatic Embryogenesis'. *In Vitro Cellular & Developmental Biology - Plant* 54(3):240–52. doi: 10.1007/s11627-018-9905-2.
107. Ma, Enbo, Lucas B. B. Harrington, Mitchell R. O'Connell, Kaihong Zhou, Jennifer A. A. Doudna, Mitchell R. O'Connell, Kaihong Zhou, and Jennifer A. A. Doudna. 2015. 'Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes'. *Molecular Cell* 60(3):398–407. doi: 10.1016/j.molcel.2015.10.030.
108. Maddalo, Danilo, Eusebio Manchado, Carla P. Concepcion, Ciro Bonetti, Joana A. Vidigal, Yoon-Chi Han, Paul Ogrodowski, Alessandra Crippa, Natasha Rekhtman, Elisa de Stanchina, Scott W. Lowe, and Andrea Ventura. 2014. 'In Vivo Engineering of Oncogenic Chromosomal Rearrangements with the CRISPR/Cas9 System'. *Nature* 516(7531):423–27. doi: 10.1038/nature13902.
109. Makarova, Kira S., Yuri I. Wolf, Omer S. Alkhnbashi, Fabrizio Costa, Shiraz A. Shah, Sita J. Saunders, Rodolphe Barrangou, Stan J. J. Brouns, Emmanuelle Charpentier, Daniel H. Haft, Philippe Horvath, Sylvain Moineau, Francisco J. M. Mojica, Rebecca M. Terns, Michael P. Terns, Malcolm F. White, Alexander F. Yakunin, Roger A. Garrett, John van der Oost, Rolf Backofen, and Eugene V. Koonin. 2015. 'An Updated Evolutionary Classification of CRISPR–Cas Systems'. *Nature Reviews Microbiology* 13(11):722–36. doi: 10.1038/nrmicro3569.
110. Makarova, Kira S., Yuri I. Wolf, Jaime Iranzo, Sergey A. Shmakov, Omer S. Alkhnbashi, Stan J. J. Brouns, Emmanuelle Charpentier, David Cheng, Daniel H. Haft, Philippe Horvath, Sylvain Moineau, Francisco J. M. Mojica, David Scott, Shiraz A. Shah, Virginijus Siksnys, Michael P. Terns, Česlovas Venclovas, Malcolm F. White, Alexander F. Yakunin, Winston Yan, Feng Zhang, Roger A. Garrett, Rolf Backofen, John van der Oost, Rodolphe Barrangou, and Eugene V. Koonin. 2020. 'Evolutionary Classification of CRISPR–Cas Systems: A Burst of Class 2 and Derived Variants'. *Nature Reviews Microbiology* 18(2):67–83. doi: 10.1038/s41579-019-0299-x.
111. Makarova, Kira S., Yuri I. Wolf, and Eugene V. Koonin. 2013. 'The Basic Building Blocks and Evolution of CRISPR–Cas Systems'. *Biochemical Society Transactions* 41(6):1392–1400. doi: 10.1042/BST20130038.
112. Mani, Ram-Shankar, and Arul M. Chinnaiyan. 2010. 'Triggers for Genomic Rearrangements: Insights into Genomic, Cellular and Environmental Influences'. *Nature Reviews. Genetics* 11(12):819–29. doi: 10.1038/nrg2883.
113. McDonald, James I., Hamza Celik, Lisa E. Rois, Gregory Fishberger, Tolison Fowler, Ryan Rees, Ashley Kramer, Andrew Martens, John R. Edwards, and Grant A. Challen. 2016. 'Reprogrammable CRISPR/Cas9-Based System for Inducing Site-Specific DNA Methylation'. *Biology Open* 5(6):866–74. doi: 10.1242/bio.019067.
114. Meeske, Alexander J., Sandra Nakandakari-Higa, and Luciano A. Marraffini. 2019. 'Cas13-Induced Cellular Dormancy Prevents the Rise of CRISPR-Resistant Bacteriophage'. *Nature* 570(7760):241–45. doi: 10.1038/s41586-019-1257-5.
115. Mohanraju, Parthana, Kira S. Makarova, Bernd Zetsche, Feng Zhang, Eugene V Koonin, and John Van der Oost. 2016. 'Diverse Evolutionary Roots and Mechanistic Variations of the CRISPR-Cas Systems'. *Science* 353(6299):aad5147. doi: 10.1126/science.aad5147.
116. Mojica, Francisco J. M., Cesar Diez-Villasenor, Jesus Garcia-Martinez, and Elena Soria. 2005. 'Intervening Sequences of Regularly Spaced Prokaryotic

Repeats Derive from Foreign Genetic Elements'. *Journal of Molecular Evolution* 60(2):174–82. doi: 10.1007/s00239-004-0046-3.
117. Morita, Sumiyo, Hirofumi Noguchi, Takuro Horii, Kazuhiko Nakabayashi, Mika Kimura, Kohji Okamura, Atsuhiko Sakai, Hideyuki Nakashima, Kenichiro Hata, Kinichi Nakashima, and Izuho Hatada. 2016. 'Targeted DNA Demethylation in Vivo Using DCas9-Peptide Repeat and ScFv-TET1 Catalytic Domain Fusions'. *Nature Biotechnology* 34(10):1060–65. doi: 10.1038/nbt.3658.
118. Nandy, Soumen, Bhuvan Pathak, Shan Zhao, and Vibha Srivastava. 2019. 'Heat-Shock-Inducible CRISPR/Cas9 System Generates Heritable Mutations in Rice'. *Plant Direct* 3(5):e00145. doi: 10.1002/pld3.145.
119. Niewoehner, Ole, Carmela Garcia-Doval, Jakob T. Rostøl, Christian Berk, Frank Schwede, Laurent Bigler, Jonathan Hall, Luciano A. Marraffini, and Martin Jinek. 2017. 'Type III CRISPR-Cas Systems Produce Cyclic Oligoadenylate Second Messengers'. *Nature* 548(7669):543–48. doi: 10.1038/nature23467.
120. Nishida, K., T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K. Y. Hara, Z. Shimatani, and A. Kondo. 2016. 'Targeted Nucleotide Editing Using Hybrid Prokaryotic and Vertebrate Adaptive Immune Systems'. *Science* 353(6305):aaf8729–aaf8729. doi: 10.1126/science.aaf8729.
121. Nishimasu, Hiroshi, Xi Shi, Soh Ishiguro, Linyi Gao, Seiichi Hirano, Sae Okazaki, Taichi Noda, Omar O. Abudayyeh, Jonathan S. Gootenberg, Hideto Mori, Seiya Oura, Benjamin Holmes, Mamoru Tanaka, Motoaki Seki, Hisato Hirano, Hiroyuki Aburatani, Ryuichiro Ishitani, Masahito Ikawa, Nozomu Yachie, Feng Zhang, and Osamu Nureki. 2018. 'Engineered CRISPR-Cas9 Nuclease with Expanded Targeting Space'. *Science* 361(6408):1259–62. doi: 10.1126/science.aas9129.
122. Nishimasu, Hiroshi, Takashi Yamano, Linyi Gao, Feng Zhang, Ryuichiro Ishitani, and Osamu Nureki. 2017. 'Structural Basis for the Altered PAM Recognition by Engineered CRISPR-Cpf1'. *Molecular Cell* 67(1):139-147.e2. doi: 10.1016/j.molcel.2017.04.019.
123. van Overbeek, Megan, Daniel Capurso, Matthew M. Carter, Matthew S. Thompson, Elizabeth Frias, Carsten Russ, John S. Reece-Hoyes, Christopher Nye, Scott Gradia, Bastien Vidal, Jiashun Zheng, Gregory R. Hoffman, Christopher K. Fuller, and Andrew P. May. 2016. 'DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks'. *Molecular Cell* 63(4):633–46. doi: 10.1016/j.molcel.2016.06.037.
124. Özcan, Ahsen, Patrick Pausch, Andreas Linden, Alexander Wulf, Karola Schühle, Johann Heider, Henning Urlaub, Thomas Heimerl, Gert Bange, and Lennart Randau. 2019. 'Type IV CRISPR RNA Processing and Effector Complex Formation in Aromatoleum Aromaticum'. *Nature Microbiology* 4(1):89–96. doi: 10.1038/s41564-018-0274-8.
125. Park, Jung-Un, Amy Wei-Lun Tsai, Eshan Mehrotra, Michael T. Petassi, Shan-Chi Hsieh, Ailong Ke, Joseph E. Peters, and Elizabeth H. Kellogg. 2021. 'Structural Basis for Target Site Selection in RNA-Guided DNA Transposition Systems'. *Science (New York, N.Y.)* 373(6556):768–74. doi: 10.1126/science.abi8976.
126. Pasternak, Cécile, Rémi Dulermo, Bao Ton-Hoang, Robert Debuchy, Patricia Siguier, Geneviève Coste, Michael Chandler, and Suzanne Sommer. 2013. 'ISDra2 Transposition in Deinococcus Radiodurans Is Downregulated by TnpB'. *Molecular Microbiology* 88(2):443–55. doi: 10.1111/mmi.12194.

127. Paul, Bijoya, Loïc Chaubet, Dideke Emma Verver, and Guillermo Montoya. 2021. 'Mechanics of CRISPR-Cas12a and Engineered Variants on λ-DNA'. *Nucleic Acids Research* gkab1272. doi: 10.1093/nar/gkab1272.
128. Pausch, Patrick, Basem Al-Shayeb, Ezra Bisom-Rapp, Connor A. Tsuchida, Zheng Li, Brady F. Cress, Gavin J. Knott, Steven E. Jacobsen, Jillian F. Banfield, and Jennifer A. Doudna. 2020. 'CRISPR-CasΦ from Huge Phages Is a Hypercompact Genome Editor'. *Science* 369(6501):333–37. doi: 10.1126/science.abb1400.
129. Plagens, A., B. Tjaden, A. Hagemann, L. Randau, and R. Hensel. 2012. 'Characterization of the CRISPR/Cas Subtype I-A System of the Hyperthermophilic Crenarchaeon Thermoproteus Tenax'. *Journal of Bacteriology* 194(10):2491–2500. doi: 10.1128/JB.00206-12.
130. Pourcel, C., G. Salvignol, and G. Vergnaud. 2005. 'CRISPR Elements in Yersinia Pestis Acquire New Repeats by Preferential Uptake of Bacteriophage DNA, and Provide Additional Tools for Evolutionary Studies'. *Microbiology (Reading, England)* 151(Pt 3):653–63. doi: 10.1099/mic.0.27437-0.
131. Qi, Lei S., Matthew H. Larson, Luke A. Gilbert, Jennifer A. Doudna, Jonathan S. Weissman, Adam P. Arkin, and Wendell A. Lim. 2013. 'Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression'. *Cell* 152(5):1173–83. doi: 10.1016/j.cell.2013.02.022.
132. Querques, Irma, Michael Schmitz, Seraina Oberli, Christelle Chanez, and Martin Jinek. 2021. 'Target Site Selection and Remodelling by Type V CRISPR-Transposon Systems'. *Nature* 1–6. doi: 10.1038/s41586-021-04030-z.
133. Ran, F. Ann, Le Cong, Winston X. Yan, David A. Scott, Jonathan S. Gootenberg, Andrea J. Kriz, Bernd Zetsche, Ophir Shalem, Xuebing Wu, Kira S. Makarova, Eugene V. Koonin, Phillip A. Sharp, and Feng Zhang. 2015. 'In Vivo Genome Editing Using Staphylococcus Aureus Cas9'. *Nature* 520(7546):186–91. doi: 10.1038/nature14299.
134. Richter, Michelle F., Kevin T. Zhao, Elliot Eton, Audrone Lapinaite, Gregory A. Newby, Benjamin W. Thuronyi, Christopher Wilson, Luke W. Koblan, Jing Zeng, Daniel E. Bauer, Jennifer A. Doudna, and David R. Liu. 2020. 'Phage-Assisted Evolution of an Adenine Base Editor with Improved Cas Domain Compatibility and Activity'. *Nature Biotechnology* 38(7):883–91. doi: 10.1038/s41587-020-0453-z.
135. Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov. 2011. 'Integrative Genomics Viewer'. *Nature Biotechnology* 29(1):24–26. doi: 10.1038/nbt.1754.
136. Rouet, P., F. Smih, and M. Jasin. 1994. 'Introduction of Double-Strand Breaks into the Genome of Mouse Cells by Expression of a Rare-Cutting Endonuclease.' *Molecular and Cellular Biology* 14(12):8096–8106. doi: 10.1128/MCB.14.12.8096.
137. Ryu, Seuk-Min, Taeyoung Koo, Kyoungmi Kim, Kayeong Lim, Gayoung Baek, Sang-Tae Kim, Heon Seok Kim, Da-Eun Kim, Hyunji Lee, Eugene Chung, and Jin-Soo Kim. 2018. 'Adenine Base Editing in Mouse Embryos and an Adult Mouse Model of Duchenne Muscular Dystrophy'. *Nature Biotechnology* 36(6):536–39. doi: 10.1038/nbt.4148.
138. Saito, Makoto, Alim Ladha, Jonathan Strecker, Guilhem Faure, Edwin Neumann, Han Altae-Tran, Rhiannon K. Macrae, and Feng Zhang. 2021. 'Dual Modes of CRISPR-Associated Transposon Homing'. *Cell* S0092867421002919. doi: 10.1016/j.cell.2021.03.006.

139. Sapranauskas, Rimantas, Giedrius Gasiunas, Christophe Fremaux, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. 2011. 'The Streptococcus Thermophilus CRISPR/Cas System Provides Immunity in Escherichia Coli.' *Nucleic Acids Research* 39(21):9275–82.
140. Shalem, Ophir, Neville E. Sanjana, Ella Hartenian, Xi Shi, David A. Scott, Tarjei S. Mikkelsen, Dirk Heckl, Benjamin L. Ebert, David E. Root, John G. Doench, and Feng Zhang. 2014. 'Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells'. *Science* 343(6166):84–87. doi: 10.1126/science.1247005.
141. Shmakov, Sergey, Omar O. Abudayyeh, Kira S. Makarova, Yuri I. Wolf, Jonathan S. Gootenberg, Ekaterina Semenova, Leonid Minakhin, Julia Joung, Silvana Konermann, Konstantin Severinov, Feng Zhang, and Eugene V. Koonin. 2015. 'Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems'. *Molecular Cell* 60(3):385–97. doi: 10.1016/j.molcel.2015.10.008.
142. Shmakov, Sergey, Aaron Smargon, David Scott, David Cox, Neena Pyzocha, Winston Yan, Omar O. Abudayyeh, Jonathan S. Gootenberg, Kira S. Makarova, Yuri I. Wolf, Konstantin Severinov, Feng Zhang, and Eugene V. Koonin. 2017. 'Diversity and Evolution of Class 2 CRISPR–Cas Systems'. *Nature Reviews Microbiology* 15(3):169–82. doi: 10.1038/nrmicro.2016.184.
143. Silva-Correia, Joana, Sara Freitas, Rui M. Tavares, Teresa Lino-Neto, and Herlânder Azevedo. 2014. 'Phenotypic Analysis of the Arabidopsis Heat Stress Response during Germination and Early Seedling Development'. *Plant Methods* 10(1):7. doi: 10.1186/1746-4811-10-7.
144. Singh, Digvijay, John Mallon, Anustup Poddar, Yanbo Wang, Ramreddy Tippana, Olivia Yang, Scott Bailey, and Taekjip Ha. 2018. 'Real-Time Observation of DNA Target Interrogation and Product Release by the RNA-Guided Endonuclease CRISPR Cpf1 (Cas12a)'. *Proceedings of the National Academy of Sciences of the United States of America* 115(21):5444–49. doi: 10.1073/pnas.1718686115.
145. Smargon, Aaron A., David B. T. Cox, Neena K. Pyzocha, Kaijie Zheng, Ian M. Slaymaker, Jonathan S. Gootenberg, Omar A. Abudayyeh, Patrick Essletzbichler, Sergey Shmakov, Kira S. Makarova, Eugene V. Koonin, and Feng Zhang. 2017. 'Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28'. *Molecular Cell* 65(4):618-630.e7. doi: 10.1016/j.molcel.2016.12.023.
146. Stella, Stefano, Pablo Alcón, and Guillermo Montoya. 2017. 'Structure of the Cpf1 Endonuclease R-Loop Complex after Target DNA Cleavage'. *Nature* 546(7659):559–63. doi: 10.1038/nature22398.
147. Stella, Stefano, Pablo Mesa, Johannes Thomsen, Bijoya Paul, Pablo Alcón, Simon B. Jensen, Bhargav Saligram, Matias E. Moses, Nikos S. Hatzakis, and Guillermo Montoya. 2018. 'Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity'. *Cell* 175(7):1856-1871.e21. doi: 10.1016/j.cell.2018.10.045.
148. Stephenson, Anthony A., Austin T. Raper, and Zucai Suo. 2018. 'Bidirectional Degradation of DNA Cleavage Products Catalyzed by CRISPR/Cas9'. *Journal of the American Chemical Society* 140(10):3743–50. doi: 10.1021/jacs.7b13050.
149. Sternberg, S. H., R. E. Haurwitz, and J. A. Doudna. 2012. 'Mechanism of Substrate Selection by a Highly Specific CRISPR Endoribonuclease'. *RNA* 18(4):661–72. doi: 10.1261/rna.030882.111.

150. Sternberg, Samuel H., Sy Redding, Martin Jinek, Eric C. Greene, and Jennifer A. Doudna. 2014. 'DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9'. *Nature* 507(7490):62–67. doi: 10.1038/nature13011.
151. Stormo, Gary D. 2013. 'Modeling the Specificity of Protein-DNA Interactions'. *Quantitative Biology* 1(2):115–30. doi: 10.1007/s40484-013-0012-4.
152. Strecker, Jonathan, Sara Jones, Balwina Koopal, Jonathan Schmid-Burgk, Bernd Zetsche, Linyi Gao, Kira S. Makarova, Eugene V. Koonin, and Feng Zhang. 2019. 'Engineering of CRISPR-Cas12b for Human Genome Editing'. *Nature Communications* 10(1):212. doi: 10.1038/s41467-018-08224-4.
153. Strecker, Jonathan, Alim Ladha, Zachary Gardner, Jonathan L. Schmid-Burgk, Kira S. Makarova, Eugene V. Koonin, and Feng Zhang. 2019. 'RNA-Guided DNA Insertion with CRISPR-Associated Transposases'. *Science* 365(6448):48–53. doi: 10.1126/science.aax9181.
154. Strohkendl, Isabel, Fatema A. Saifuddin, James R. Rybarski, Ilya J. Finkelstein, and Rick Russell. 2018. 'Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a'. *Molecular Cell* 71(5):816-824.e3. doi: 10.1016/j.molcel.2018.06.043.
155. Svitashev, Sergei, Joshua K. Young, Christine Schwartz, Huirong Gao, S. Carl Falco, and A. Mark Cigan. 2015. 'Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA'. *Plant Physiology* 169(2):931–45. doi: 10.1104/pp.15.00793.
156. Swarts, Daan C., and Martin Jinek. 2019. 'Mechanistic Insights into the Cis- and Trans-Acting DNase Activities of Cas12a'. *Molecular Cell* 73(3):589-600.e4. doi: 10.1016/j.molcel.2018.11.021.
157. Swarts, Daan C., John van der Oost, and Martin Jinek. 2017. 'Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a'. *Molecular Cell* 66(2):221-233.e4. doi: 10.1016/j.molcel.2017.03.016.
158. Szczelkun, Mark D., Maria S. Tikhomirova, Tomas Sinkunas, Giedrius Gasiunas, Tautvydas Karvelis, Patrizia Pschera, Virginijus Siksnys, and Ralf Seidel. 2014. 'Direct Observation of R-Loop Formation by Single RNA-Guided Cas9 and Cascade Effector Complexes'. *Proceedings of the National Academy of Sciences* 111(27):9798–9803. doi: 10.1073/pnas.1402597111.
159. Tak, Y. Esther, Benjamin P. Kleinstiver, James K. Nuñez, Jonathan Y. Hsu, Joy E. Horng, Jingyi Gong, Jonathan S. Weissman, and J. Keith Joung. 2017. 'Inducible and Multiplex Gene Regulation Using CRISPR-Cpf1-Based Transcription Factors'. *Nature Methods* 14(12):1163–66. doi: 10.1038/nmeth.4483.
160. Takeda, Satoru N., Ryoya Nakagawa, Sae Okazaki, Hisato Hirano, Kan Kobayashi, Tsukasa Kusakizako, Tomohiro Nishizawa, Keitaro Yamashita, Hiroshi Nishimasu, and Osamu Nureki. 2021. 'Structure of the Miniature Type V-F CRISPR-Cas Effector Enzyme'. *Molecular Cell* 81(3):558-570.e3. doi: 10.1016/j.molcel.2020.11.035.
161. Tamulaitis, Gintautas, Česlovas Venclovas, and Virginijus Siksnys. 2017. 'Type III CRISPR-Cas Immunity: Major Differences Brushed Aside'. *Trends in Microbiology* 25(1):49–61. doi: 10.1016/j.tim.2016.09.012.
162. Tang, Xu, Guanqing Liu, Jianping Zhou, Qiurong Ren, Qi You, Li Tian, Xuhui Xin, Zhaohui Zhong, Binglin Liu, Xuelian Zheng, Dengwei Zhang, Aimee Malzahn, Zhiyun Gong, Yiping Qi, Tao Zhang, and Yong Zhang. 2018. 'A Large-Scale Whole-Genome Sequencing Analysis Reveals Highly Specific Genome Editing by Both Cas9 and Cpf1 (Cas12a) Nucleases in Rice'. *Genome Biology* 19(1):84. doi: 10.1186/s13059-018-1458-5.

163. Tang, Xu, Levi G. Lowder, Tao Zhang, Aimee A. Malzahn, Xuelian Zheng, Daniel F. Voytas, Zhaohui Zhong, Yiyi Chen, Qiurong Ren, Qian Li, Elida R. Kirkland, Yong Zhang, and Yiping Qi. 2017. 'A CRISPR-Cpf1 System for Efficient Genome Editing and Transcriptional Repression in Plants'. *Nature Plants* 3:17018. doi: 10.1038/nplants.2017.18.
164. Teng, Fei, Tongtong Cui, Guihai Feng, Lu Guo, Kai Xu, Qingqin Gao, Tianda Li, Jing Li, Qi Zhou, and Wei Li. 2018. 'Repurposing CRISPR-Cas12b for Mammalian Genome Engineering'. *Cell Discovery* 4(1):63. doi: 10.1038/s41421-018-0069-3.
165. Teng, Fei, Lu Guo, Tongtong Cui, Xin-Ge Wang, Kai Xu, Qingqin Gao, Qi Zhou, and Wei Li. 2019. 'CDetection: CRISPR-Cas12b-Based DNA Detection with Sub-Attomolar Sensitivity and Single-Base Specificity'. *Genome Biology* 20(1):132. doi: 10.1186/s13059-019-1742-z.
166. Tong, Baisong, Huina Dong, Yali Cui, Pingtao Jiang, Zhaoxia Jin, and Dawei Zhang. 2020. 'The Versatile Type V CRISPR Effectors and Their Application Prospects'. *Frontiers in Cell and Developmental Biology* 8:622103. doi: 10.3389/fcell.2020.622103.
167. Tornabene, Patrizia, Ivana Trapani, Renato Minopoli, Miriam Centrulo, Mariangela Lupo, Sonia de Simone, Paola Tiberi, Fabio Dell'Aquila, Elena Marrocco, Carolina Iodice, Antonella Iuliano, Carlo Gesualdo, Settimio Rossi, Laura Giaquinto, Silvia Albert, Carel B. Hoyng, Elena Polishchuk, Frans P. M. Cremers, Enrico M. Surace, Francesca Simonelli, Maria A. De Matteis, Roman Polishchuk, and Alberto Auricchio. 2019. 'Intein-Mediated Protein Trans-Splicing Expands Adeno-Associated Virus Transfer Capacity in the Retina'. *Science Translational Medicine* 11(492):eaav4523. doi: 10.1126/scitranslmed.aav4523.
168. Tóth, Eszter, Éva Varga, Péter István Kulcsár, Virág Kocsis-Jutka, Sarah Laura Krausz, Antal Nyeste, Zsombor Welker, Krisztina Huszár, Zoltán Ligeti, András Tálas, and Ervin Welker. 2020. 'Improved LbCas12a Variants with Altered PAM Specificities Further Broaden the Genome Targeting Range of Cas12a Nucleases'. *Nucleic Acids Research* 48(7):3722–33. doi: 10.1093/nar/gkaa110.
169. Tran, Mai H., Hajeung Park, Christopher L. Nobles, Pabalu Karunadharma, Li Pan, Guocai Zhong, Haimin Wang, Wenhui He, Tianling Ou, Gogce Crynen, Kelly Sheptack, Ian Stiskin, Huihui Mou, and Michael Farzan. 2021. 'A More Efficient CRISPR-Cas12a Variant Derived from Lachnospiraceae Bacterium MA2020'. *Molecular Therapy. Nucleic Acids* 24:40–53. doi: 10.1016/j.omtn.2021.02.012.
170. Truong, Dong-Jiunn Jeffery, Karin Kühner, Ralf Kühn, Stanislas Werfel, Stefan Engelhardt, Wolfgang Wurst, and Oskar Ortiz. 2015. 'Development of an Intein-Mediated Split-Cas9 System for Gene Therapy'. *Nucleic Acids Research* 43(13):6450–58. doi: 10.1093/nar/gkv601.
171. Tsai, Shengdar Q. 2018. 'Discovering the Genome-Wide Activity of CRISPR-Cas Nucleases'. *ACS Chemical Biology* 13(2):305–8. doi: 10.1021/acschembio.7b00847.
172. Vojta, Aleksandar, Paula Dobrinić, Vanja Tadić, Luka Bočkor, Petra Korać, Boris Julg, Marija Klasić, and Vlatka Zoldoš. 2016. 'Repurposing the CRISPR-Cas9 System for Targeted DNA Methylation'. *Nucleic Acids Research* 44(12):5615–28. doi: 10.1093/nar/gkw159.
173. Wagner, Dimitrios L., Leila Amini, Desiree J. Wendering, Lisa-Marie Burkhardt, Levent Akyüz, Petra Reinke, Hans-Dieter Volk, and Michael

Schmueck-Henneresse. 2019. 'High Prevalence of Streptococcus Pyogenes Cas9-Reactive T Cells within the Adult Human Population'. *Nature Medicine* 25(2):242–48. doi: 10.1038/s41591-018-0204-6.
174. Walton, Russell T., Kathleen A. Christie, Madelynn N. Whittaker, and Benjamin P. Kleinstiver. 2020. 'Unconstrained Genome Targeting with Near-PAMless Engineered CRISPR-Cas9 Variants'. *Science* 368(6488):290–96. doi: 10.1126/science.aba8853.
175. Wang, Dan, Feng Zhang, and Guangping Gao. 2020. 'CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors'. *Cell* 181(1):136–50. doi: 10.1016/j.cell.2020.03.023.
176. Wang, Ning, Maren Arling, George Hoerster, Larisa Ryan, Emily Wu, Keith Lowe, William Gordon-Kamm, Todd J. Jones, N. Doane Chilcoat, and Ajith Anand. 2020. 'An Efficient Gene Excision System in Maize'. *Frontiers in Plant Science* 11:1298. doi: 10.3389/fpls.2020.01298.
177. Wang, Qiongqiong, Muna Alariqi, Fuqiu Wang, Bo Li, Xiao Ding, Hangping Rui, Yajun Li, Zhongping Xu, Lei Qin, Lin Sun, Jianying Li, Jiawei Zou, Keith Lindsey, Xianlong Zhang, and Shuangxia Jin. 2020. 'The Application of a Heat-Inducible CRISPR/Cas12b (C2c1) Genome Editing System in Tetraploid Cotton (G. Hirsutum) Plants'. *Plant Biotechnology Journal* 18(12):2436–43. doi: 10.1111/pbi.13417.
178. Wang, Tim, Jenny J. Wei, David M. Sabatini, and Eric S. Lander. 2014. 'Genetic Screens in Human Cells Using the CRISPR-Cas9 System'. *Science* 343(6166):80–84. doi: 10.1126/science.1246981.
179. Wei, Yunzhou, Rebecca M. Terns, and Michael P. Terns. 2015. 'Cas9 Function and Host Genome Sampling in Type II-A CRISPR–Cas Adaptation'. *Genes & Development* 29(4):356–61. doi: 10.1101/gad.257550.114.
180. Wright, Addison V., and Jennifer A. Doudna. 2016. 'Protecting Genome Integrity during CRISPR Immune Adaptation'. *Nature Structural & Molecular Biology* 23(10):876–83. doi: 10.1038/nsmb.3289.
181. Wu, Yaokang, Taichi Chen, Yanfeng Liu, Rongzhen Tian, Xueqin Lv, Jianghua Li, Guocheng Du, Jian Chen, Rodrigo Ledesma-Amaro, and Long Liu. 2020. 'Design of a Programmable Biosensor-CRISPRi Genetic Circuits for Dynamic and Autonomous Dual-Control of Metabolic Flux in Bacillus Subtilis'. *Nucleic Acids Research* 48(2):996–1009. doi: 10.1093/nar/gkz1123.
182. Wu, Zhaowei, Yifei Zhang, Haopeng Yu, Deng Pan, Yujue Wang, Yannan Wang, Fan Li, Chang Liu, Hao Nan, Weizhong Chen, and Quanjiang Ji. 2021. 'Programmed Genome Editing by a Miniature CRISPR-Cas12f Nuclease'. *Nature Chemical Biology* 17(11):1132–38. doi: 10.1038/s41589-021-00868-6.
183. Wu, Zhijian, Hongyan Yang, and Peter Colosi. 2010. 'Effect of Genome Size on AAV Vector Packaging.' *Molecular Therapy : The Journal of the American Society of Gene Therapy* 18(1):80–86. doi: 10.1038/mt.2009.255.
184. Wyman, Claire, and Roland Kanaar. 2006. 'DNA Double-Strand Break Repair: All's Well That Ends Well'. *Annual Review of Genetics* 40:363–83. doi: 10.1146/annurev.genet.40.110405.090451.
185. Xiao, Renjian, Zhuang Li, Shukun Wang, Ruijie Han, and Leifu Chang. 2021. 'Structural Basis for Substrate Recognition and Cleavage by the Dimerization-Dependent CRISPR-Cas12f Nuclease'. *Nucleic Acids Research* 49(7):4120–28. doi: 10.1093/nar/gkab179.
186. Xiao, Renjian, Shukun Wang, Ruijie Han, Zhuang Li, Clinton Gabel, Indranil Arun Mukherjee, and Leifu Chang. 2021. 'Structural Basis of Target DNA

Recognition by CRISPR-Cas12k for RNA-Guided DNA Transposition'. *Molecular Cell* 81(21):4457-4466.e5. doi: 10.1016/j.molcel.2021.07.043.
187. Xu, Xiaoshu, Augustine Chemparathy, Leiping Zeng, Hannah R. Kempton, Stephen Shang, Muneaki Nakamura, and Lei S. Qi. 2021. 'Engineered Miniature CRISPR-Cas System for Mammalian Genome Regulation and Editing'. *Molecular Cell* 81(20):4333-4345.e4. doi: 10.1016/j.molcel.2021.08.008.
188. Xu, Xingxing, Yonghui Tao, Xiaobo Gao, Lei Zhang, Xufang Li, Weiguo Zou, Kangcheng Ruan, Feng Wang, Guo-Liang Xu, and Ronggui Hu. 2016. 'A CRISPR-Based Approach for Targeted DNA Demethylation'. *Cell Discovery* 2:16009. doi: 10.1038/celldisc.2016.9.
189. Yamano, Takashi, Hiroshi Nishimasu, Bernd Zetsche, Hisato Hirano, Ian M. Slaymaker, Yinqing Li, Iana Fedorova, Takanori Nakane, Kira S. Makarova, Eugene V. Koonin, Ryuichiro Ishitani, Feng Zhang, and Osamu Nureki. 2016. 'Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA'. *Cell* 165(4):949–62. doi: 10.1016/j.cell.2016.04.003.
190. Yamano, Takashi, Bernd Zetsche, Ryuichiro Ishitani, Feng Zhang, Hiroshi Nishimasu, and Osamu Nureki. 2017. 'Structural Basis for the Canonical and Non-Canonical PAM Recognition by CRISPR-Cpf1'. *Molecular Cell* 67(4):633-645.e3. doi: 10.1016/j.molcel.2017.06.035.
191. Yan, Winston X., Pratyusha Hunnewell, Lauren E. Alfonse, Jason M. Carte, Elise Keston-Smith, Shanmugapriya Sothiselvam, Anthony J. Garrity, Shaorong Chong, Kira S. Makarova, Eugene V. Koonin, David R. Cheng, and David A. Scott. 2019. 'Functionally Diverse Type V CRISPR-Cas Systems'. *Science* 363(6422):88–91. doi: 10.1126/science.aav7271.
192. Yang, Hui, Pu Gao, Kanagalaghatta R. Rajashankar, and Dinshaw J. Patel. 2016. 'PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease'. *Cell* 167(7):1814-1828.e12. doi: 10.1016/j.cell.2016.11.053.
193. Yang, Yang, Lili Wang, Peter Bell, Deirdre McMenamin, Zhenning He, John White, Hongwei Yu, Chenyu Xu, Hiroki Morizono, Kiran Musunuru, Mark L. Batshaw, and James M. Wilson. 2016. 'A Dual AAV System Enables the Cas9-Mediated Correction of a Metabolic Liver Disease in Newborn Mice'. *Nature Biotechnology* 34(3):334–38. doi: 10.1038/nbt.3469.
194. Yeh, Charles D., Christopher D. Richardson, and Jacob E. Corn. 2019. 'Advances in Genome Editing through Control of DNA Repair Pathways'. *Nature Cell Biology* 21(12):1468–78. doi: 10.1038/s41556-019-0425-z.
195. Young, Gavin, Nikolas Hundt, Daniel Cole, Adam Fineberg, Joanna Andrecka, Andrew Tyler, Anna Olerinyova, Ayla Ansari, Erik G. Marklund, Miranda P. Collier, Shane A. Chandler, Olga Tkachenko, Joel Allen, Max Crispin, Neil Billington, Yasuharu Takagi, James R. Sellers, Cédric Eichmann, Philipp Selenko, Lukas Frey, Roland Riek, Martin R. Galpin, Weston B. Struwe, Justin L. P. Benesch, and Philipp Kukura. 2018. 'Quantitative Mass Imaging of Single Biological Macromolecules'. *Science* 360(6387):423–27. doi: 10.1126/science.aar5839.
196. Zetsche, Bernd, Jonathan S. Gootenberg, Omar O. Abudayyeh, Ian M. Slaymaker, Kira S. Makarova, Patrick Essletzbichler, Sara E. Volz, Julia Joung, John van der Oost, Aviv Regev, Eugene V. Koonin, and Feng Zhang. 2015. 'Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System'. *Cell* 163(3):759–71. doi: 10.1016/j.cell.2015.09.038.
197. Zetsche, Bernd, Matthias Heidenreich, Prarthana Mohanraju, Iana Fedorova, Jeroen Kneppers, Ellen M. DeGennaro, Nerges Winblad, Sourav R. Choudhury,

Omar O. Abudayyeh, Jonathan S. Gootenberg, Wen Y. Wu, David A. Scott, Konstantin Severinov, John van der Oost, and Feng Zhang. 2017. 'Multiplex Gene Editing by CRISPR-Cpf1 Using a Single CrRNA Array'. *Nature Biotechnology* 35(1):31–34. doi: 10.1038/nbt.3737.

198. Zhang, Jing, Taciana Kasciukovic, and Malcolm F. White. 2012. 'The CRISPR Associated Protein Cas4 Is a 5′ to 3′ DNA Exonuclease with an Iron-Sulfur Cluster' edited by M. G. Marinus. *PLoS ONE* 7(10):e47232. doi: 10.1371/journal.pone.0047232.
199. Zhang, Xiaochun, Jingman Wang, Qiuxiang Cheng, Xuan Zheng, Guoping Zhao, and Jin Wang. 2017. 'Multiplex Gene Regulation by CRISPR-DdCpf1'. *Cell Discovery* 3:17018. doi: 10.1038/celldisc.2017.18.
200. Zhang, Xin, Wei Wang, Lin Shan, Le Han, Shufeng Ma, Yan Zhang, Bingtao Hao, Ying Lin, and Zhili Rong. 2018. 'Gene Activation in Human Cells Using CRISPR/Cpf1-P300 and CRISPR/Cpf1-SunTag Systems'. *Protein & Cell* 9(4):380–83. doi: 10.1007/s13238-017-0491-6.
201. Zhang, Yan, Rakhi Rajan, H. Steven Seifert, Alfonso Mondragón, and Erik J. Sontheimer. 2015. 'DNase H Activity of Neisseria Meningitidis Cas9'. *Molecular Cell* 60(2):242–55. doi: 10.1016/j.molcel.2015.09.020.

SANTRAUKA

SANTRUMPOS

As	*Acidibacillus sulfuroxidans*
Au	*Aureabacteria* bacterium
AAV	adeno-asocijuotas virusas
AHT	anhidrotetraciklinas
bp	bazių pora
Cas	angl. *CRISPR associated*
Cn	*Clostridium novyi*
CRISPR	angl. *clustered regularly interspaced short palindromic repeats*
crRNR	CRISPR RNR
dgDNR	dvigrandinė DNR
DMEM	angl. *Dulbecco's Modified Eagle Medium*
DNazė	deoksiribonukleazė
DTT	ditiotreitolis
EDTA	etilendiamintetraacto rūgštis
FLL	angl. *full length linear*, pilno ilgio linijinis fragmentas
gRNR	angl. *guide RNA*, nukreipiančioji RNR
HDV	hepatito delta virusas
His	histidinas
IPTG	izopropilo β-D-1-tiogalaktopiranozidas
LB	Luria-Bertani terpė
MBP	maltozę surišantis baltymas
Mi	*Micrarchaeota* archaeon
nt	nukleotidas
NTG	ne-taikinio grandinė
OC	angl. *open circle*, viengrandinį trūkį turinti plazmidė
PAGE	poliakrilamido gelio elektroforezė
PAM	angl. *protospacer adjacent motif*
PGR	polimerazės grandininė reakcija
PFM	angl. *position frequency matrix*, padėties dažnio matrica
PMSF	fenilmetilsulfonilfluoridas
Pt	*Parageobacillus thermoglucosidasius*
RNase	ribonunukleazė
RNP	ribonukleoproteinas
Ru	*Ruminococcus*
RuvC	RuvC nukleazės domenas

SC	angl. *supercoiled*, superspiralizuota plazmidė
SDS	natrio dodecilsulfatas
Sp	*Syntrophomonas palmitatica*
TAM	angl. *transposon-associated motif*
TBE	tris-boratinis-EDTA
TEV	angl. *tobacco etch virus*
tracrRNR	*trans*-aktyvuojanti CRISPR RNR
Tris	tris(hidroksimetil)aminometanas
TG	taikinio grandinė
Un	angl. *uncultured archaeon*
vgDNR	viengrandinė DNR

ĮVADAS

Prokariotinės CRISPR-Cas – angl. *clustered regularly interspaced short palindromic repeats and CRISPR associated*, sistemos aptinkamos ~ 85 % archėjų ir ~ 40 % bakterijų (Makarova et al. 2020). Šios sistemos suteikia adaptyvią apsaugą prieš svetimas nukleorūgštis saugodamos informaciją apie ankstesnes infekcijas DNR sekų forma CRISPR regione, kuris vėliau panaudojamas nukreipiančiųjų RNR ekspresijai. Pasikartojančios infekcijos atveju šios RNR nukreipia Cas nukleazes svetimų nukleorūgščių degradacijai (Barrangou et al. 2007; Mojica et al. 2005). RNR molekulėmis programuojamos Cas nukleazės gali būti nukreiptos perkirpti bet kurį DNR taikinį, todėl šios nukleazės yra pritaikomos genomo redagavimo srityje (Cong et al. 2013; Deltcheva et al. 2011; Gasiunas et al. 2012; Jinek et al. 2012). Tačiau dažniausiai naudojamų Cas9 ir Cas12a nukleazių dydis išlieka vienu iš pagrindinių veiksnių, ribojančių jų pernešimą į tikslines ląsteles, pavyzdžiui, naudojant adeno-asocijuotus virusus (Dong et al. 1996; D. Wang et al. 2020). Be to, sukūrus naujos kartos bazių ir „prime“ redagavimo įrankius, pagrįstus Cas baltymų suliejimu su papildomais efektoriniais baltymais, tokiais kaip deaminazės ir atvirkštinės transkriptazės, padidėjo mažesnių Cas baltymų poreikis (Anzalone et al. 2019, 2020; Komor et al. 2016; Tang et al. 2017). Atitinkamai, neseniai atrastos Cas12f nukleazės, kurios yra perpus mažesnės nei Cas9 ir Cas12a baltymai, tapo patrauklia alternatyva (Harrington et al. 2018).

Pagrindinis šios disertacijos objektas – neseniai atrastų 2 klasės V-F tipo CRISPR-Cas sistemų mažosios Cas12f nukleazės. **Tikslas** buvo patikrinti ar šios nukleazės geba kirpti dgDNR taikinius bei gali būti pritaikytos kaip nauji genomo redagavimo įrankiai. Tuo tikslu buvo iškelti šie **uždavinia**i:

1. Patikrinti mažųjų Cas12f nukleazių gebėjimą karpyti dgDNR taikinius.
2. Identifikuoti Cas12f surišamas nukreipiančiąsias RNR molekules.
3. Atkurti ir charakterizuoti Cas12f kompleksus ir jų aktyvumą *in vitro*.
4. Ištirti Cas12f gebėjimą kirpti genominę DNR eukariotinėse ląstelėse.

Mokslinis naujumas ir praktinė vertė:

Šis tyrimas yra skirtas naujų mažųjų Cas12f nukleazių atradimui ir apibūdinimui. Bene perpus mažesnės už gerai žinomus Cas12a ir Cas9 baltymus, Cas12f nukleazės tapo patrauklia alternatyva genomo redagavimui palengvinant efektorinių kompleksų supakavimą ir pernešimą į tikslines ląsteles tokiomis platformomis, kaip, pavyzdžiui, adeno-asocijuoti virusai. Buvo manoma, kad dėl savo dydžio Cas12f baltymai nėra pajėgūs kirpti

dgDNR taikinius (Shmakov et al. 2017). Pirmasis eksperimentiškai apibūdintas Cas12f šeimos baltymas (pavadintas Cas14) netiesiogiai patvirtino šią hipotezę, kadangi buvo nustatytas tik vgDNR kirpimas (Harrington et al. 2018). Šiame darbe pirmą kartą pademonstravome nuo PAM (angl. *protospacer adjacent motif*) priklausomą dgDNR kirpimo aktyvumą dešimčiai Cas12f nukleazių (įskaitant ir Cas14). Taip pat dvi, *Acidibacillus sulfuroxidans* (As) ir *Syntrophomonas palmitatica* (Sp), Cas12f1 nukleazės sėkmingai ribojo plazmidžių transformaciją į *E. coli* ląsteles. Šios nukleazės buvo atrinktos tolimesniam biocheminiam charakterizavimui, kuris atskleidė bendrus Cas12f1 nukleazių bruožus: i) lyginant su kitomis Cas12 nukleazėmis būdingas kompaktiškas dydis, ii) nepaprastai ilgos tracrRNR, iii) efektyvus dgDNR taikinio surišimas ir kirpimas esant aukštesnei temperatūrai (45–55 °C), iv) kolateralinis aktyvumas prieš nespecifinę vgDNR, v) Cas12f1 dimerizacija surišus vieną gRNR molekulę. Šios unikalios savybės praplečia CRISPR-Cas molekulinių įrankių pritaikymo galimybes. Nuo temperatūros priklausomas dgDNR taikinių atpažinimas ir papildomas vgDNR skaidymas gali būti pritaikytas nukleorūgščių aptikimo platformose, supaprastinant sistemos derinimą su izotermine amplifikacija (Chen et al. 2018; Gootenberg et al. 2018; Joung et al. 2020). Be to, jautrumas temperatūrai gali būti naudojamas reguliuoti dgDNR taikinių kirpimo aktyvumą augaluose, kurie, kaip buvo įrodyta, toleruoja arba gali būti aklimatizuojami aukštesnėse temperatūrose, (Barone et al. 2020; Khattri et al. 2011; Lin et al. 1984; Liu and Charng 2012; Nandy et al. 2019; Silva-Correia et al. 2014; N. Wang et al. 2020; Q. Wang et al. 2020). Tokiu būdu sumažinamas nespecifinis kirpimas ir galima procesą kontroliuoti tiek erdvėje, tiek ir laike. Apibendrinant, šios savybės atveria kelią miniatiūrinių Cas12f pagrindu veikiančių genomo redagavimo įrankių kūrimui.

Ginamieji disertacijos teiginiai:

1. Mažosios Cas12f nukleazės gali atpažinti ir vykdyti nuo PAM priklausomą dgDNR taikinių kirpimą.
2. Siekiant supaprastinti Cas12f sistemą, surišamos crRNR ir tracrRNR molekulės gali būti sujungtos į vieną nukreipiančiąją gRNR.
3. Cas12f ir gRNA 2:1 santykiu formuojamas efektorinis kompleksas kerpa taikinį turinčias DNR molekules ir pasižymi kolateraliniu aktyvumu prieš nespecifines vgDNR.
4. Abejais atvejais efektyvesnis DNR kirpimas *in vitro* stebimas esant aukštesnei aplinkos temperatūrai, kas galėtų būti pritaikoma nukleazių aktyvumo reguliavimui.

5. Mažosios Cas12f nukleazės gali būti sėkmingai pernešamos ir pritaikomos genominių DNR sekų modifikavimui eukariotinėse ląstelėse.

MEDŽIAGOS

Cheminės medžiagos

Visos šiame tyrime naudotos cheminės medžiagos yra aukščiausios grynumo klasės ir buvo įsigytos iš Sigma-Aldrich, Roth, Fluka ir Thermo Fisher Scientific Baltics. Radioaktyviai pažymėti nukleotidai buvo įsigyti iš Perkin Elmer.

Komerciniai baltymai ir rinkiniai

T4 polinukleotidkinazė, FastAP termiškai jautri šarminė fosfatazė, Phusion, DreamTaq ir T4 DNR polimerazė, T4 DNR ligazė, „Fast Digest" restrikcijos fermentai, proteinazė K, RiboLock RNazių inhibitorius, RNazė A ir DNazė I, naudotos šiame tyrime, buvo gautos iš Thermo Fisher Scientific.

„Rapid DNA Ligation Kit", „GeneJET PCR Purification Kit", „GeneJET Plasmid Miniprep Kit", „GeneJET Gel Extraction Kit", „T7 High Yield Transcription Kit" ir „GeneJET RNA Cleanup and Concentration Kit" buvo įsigyti iš Thermo Fisher Scientific Baltics, „TruSeq Small RNA Library Preparation Kit" – Illumina, „Synergy 2.0 Plant DNA Extraction Kit" – Ops Diagnostics, „Zymoclean Gel DNA Recovery Kit" – Zymo Research, „Monarch PCR Purification Column" – New England Biolabs.

Visi šie produktai buvo naudojami pagal gamintojo instrukcijas.

Bakterijų kamienai

E. coli BL21(DE3) kamienas [F– ompT gal dcm lon hNDSB(rB- mB-) λ(DE3)] panaudotas Un1Cas12f1 raiškai.

E. coli Arctic Express (DE3) kamienas [F– ompT gal dcm tetr hsdS(rB-mB-) λ(DE3) endA Hte [cpn10 cpn60 Gentr] panaudotas SpCas12f1 ir AsCas12f1 raiškai ir transformacijos ribojimo eksperimentams.

E. coli DH5α kamienas [F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–] panaudotas konstruktų klonavimo ir taikininių plazmidžių padauginimo procedūroms.

E. coli bakterijos buvo kultivuojamos skystoje LB terpėje (1 % peptono, 0.5 % mielių ekstrakto, 0.5 % NaCl vandenyje, pH 7,0 (25 °C)) arba ant agarizuotos LB terpės (1,5 % agaro) petri lėkštelėse.

Ląstelių linijos

HEK293T – ląstelių linija, gauta iš žmogaus embrioninių inkstų ląstelių, išaugintų audinių kultūroje ir ekspresuojanti SV40 didelį T-antigeną (ATCC

katalogo numeris CRL-3216). Ląstelių linija buvo kultivuojama naudojant DMEM terpę, papildant 10 % galvijų vaisiaus serumu, penicilinu (100 U/ml) ir streptomicinu (100 µg/ml) (Thermo Fisher Scientific).

PH1V69 – DuPont Pioneer išvesta linija, naudojama augalų genomo redagavimo tyrimuose. Visi naudojami augalai buvo auginami DuPont Pioneer (Corteva) šiltnamiuose.

Baltymai ir nukleorūgštys

Toliau pateiktuose prieduose pateikiama informacija apie:

- Appendix 1 – šiame darbe tirti Cas12f baltymai.
- Appendix 2 – Cas12f baltymų sekos.
- Appendix 3 – plazmidiniai konstruktai su Benchling nuorodomis.
- Appendix 4 – Cas12f nukreipiančiosios RNR molekulės.
- Appendix 5 – oligonukleotidai naudoti kirpimo ir surišimo tyrimuose.
- Appendix 6 – pradmenys, naudoti įvertinant žmogaus ir kukurūzų ląstelių redagavimą.
- Appendix 7 – gRNR taikiniai, naudojami žmogaus ir kukurūzų ląstelių redagavimui.

Buferiniai tirpalai

PAM nustatymo metodas:

Lizės buferis: 20 mM fosfato (pH 7,0 (25 °C)), 0,5 M NaCl, 2 mM PMSF 5 % (v/v) glicerolio.

Reakcijos buferis: 10 mM Tris-HCl (pH 7,5 (37 °C)), 100 mM NaCl, 1 mM DTT, ir 10 mM $MgCl_2$.

Cas12f1-RNR kompleksų ir Cas12f1 baltymų gryninimas:

Un1Cas12f1

Užnešimo buferis: 20 mM Tris-HCl (pH 8,0 (25 °C)), 1,5 M NaCl, 5 mM 2-merkaptoetanolio, 10 mM imidazolo, 2 mM PMSF, 5 % (v/v) glicerolio.

Eliucijos buferis: 20 mM Tris-HCl (pH 8,0 (25 °C)), 0,5 M NaCl, 5 mM 2-merkaptoetanolio.

Baltymų saugojimo buferis: 20 mM Tris-HCl (pH 8,0 (25 °C)), 0,5 M NaCl, 2 mM DTT, 50 % (v/v) glicerolio.

SpCas12f1 ir AsCas12f1 RNP kompleksai

Užnešimo buferis: 20 mM Tris-HCl (pH 8,0 (25 °C)), 0,25 M NaCl, 5 mM 2-merkaptoetanolio, 25 mM imidazolo, 2 mM PMSF, 5 % (v/v) glicerolio.

Eliucijos buferis: 20 mM Tris-HCl (pH 8,0 (25 °C)), 0,25 M NaCl, 5 mM 2-merkaptoetanolio ir 5 % (v/v) glicerolio.

Baltymų saugojimo buferis: 20 mM Tris-HCl (pH 8.0 (25 °C)), 0.25 M NaCl, 2 mM DTT ir 50 % (v/v) glicerolio.

SpCas12f1 ir AsCas12f1

Užnešimo buferis: 20 mM Tris-HCl (pH 8,0 (25 °C)), 1,5 M NaCl, 5 mM 2-merkaptoetanolio, 25 mM imidazolo, 2 mM PMSF, 5 % (v/v) glicerolio.

Eliucijos buferis: 20 mM Tris-HCl (pH 8,0 (25 °C)), 0,5 M NaCl ir 5 mM 2-merkaptoetanolio.

Baltymų saugojimo buferis: 20 mM Tris-HCl (pH 8,0 (25 °C)), 500 mM NaCl, 2 mM DTT ir 50 % (v/v) glicerolio.

RNR gryninimas iš RNP kompleksų:

Proteinazės K reakcijos buferis: 10 mM Tris-HCl (pH 7,5 (37 °C)), 1 mM EDTA, 1 mM DTT, 100 mM NaCl ir 5 mM $MgCl_2$.

Kompleksų surinkimas:

Cas12f1-gRNR kompleksų surinkimo buferis: 10 mM Tris-HCl (pH 7,5 (37 °C)), 100 mM NaCl, 1 mM EDTA, 1 mM DTT.

Plazmidinės DNR ir M13 kirpimas

Un1Cas12f1 reakcijos buferis: 2,5 mM Tris-HCl (pH 7,5 (37 °C)), 25 mM NaCl, 0,25 mM DTT ir 10 mM $MgCl_2$.

SpCas12f1 reakcijos buferis: 10 mM Tris-HCl (pH 7,5 (37 °C)), 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$.

AsCas12f1 reakcijos buferis: 10 mM Tris-HCl (pH 7,5 (37 °C)), 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$.

3× reakcijos stabdymo ir mėginio užnešimo buferis: 0,01 % bromfenolio mėlio, 0,03 % SDS ir 75 mM EDTA in 50 % (v/v) glicerolio.

Oligonukleotidų kirpimas

Un1Cas12f1 reakcijos buferis: 5 mM Tris-HCl (pH 7,5 (37 °C)), 50 mM NaCl, 0,5 mM DTT ir 5 mM $MgCl_2$.

SpCas12f1 reakcijos buferis: 10 mM Tris-HCl (pH 7,5 (37 °C)), 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$.

AsCas12f1 reakcijos buferis: 10 mM Tris-HCl (pH 7,5 (37 °C)), 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$.

Reakcijos stabdymo ir mėginio užnešimo buferis: 95 % (v/v) formamido, 0,01 % bromfenolio mėlio ir 25 mM EDTA.

DNR surišimas:
Cas12f1-gRNR komplekso surinkimo buferis: 10 mM Tris-HCl (pH 7,5 (37 °C)), 100 mM NaCl, 1 mM EDTA, 1 mM DTT.
Surišimo buferis: 40 mM Tris-HAc (pH 8,4 (25 °C)), 1 mM EDTA, 0,1 mg/ml BSA, 10 % (v/v) glicerolio ir 5 mM $Mg(C_2H_3O_2)_2$.

Masių fotometrija molekulinės masės nustatymui:
Cas12f1-gRNR komplekso surinkimo buferis: 10 mM Tris-HCl (pH 7,5 (37 °C)), 100 mM NaCl, 1 mM EDTA, 1 mM DTT.
Surišimo buferis: 40 mM Tris-HAc (pH 8,4 (25 °C)), 5 mM $Mg(C_2H_3O_2)_2$.

METODAI

CRISPR-Cas12f sistemų konstruktai PAM sekų nustatymui

Pasirinktos CRISPR-Cas12f sistemos (Appendix 1) buvo modifikuotos taip, kad būtų pritaikytos anksčiau aprašytai 7 bp ilgio atsitiktinių PAM sekų plazmidžių bibliotekai kirpti (Karvelis et al. 2015). Tai buvo pasiekta pakeičiant natyvų CRISPR regioną trimis pasikartojimo seka:skirtukas: pasikartojimo seka motyvais, kurie koduoja skirtuką (33–39 nt, priklausomai nuo vidutinio skirtuko ilgio, stebimo atitinkamoje Cas12f sistemoje) komplementarų PAM bibliotekos taikiniui esančiam iškart už PAM sekų fragmento. Tada sukurtos CRISPR-Cas12f sistemos buvo susintetintos (GenScript) ir klonuotos į modifikuotą pETDuet-1 (MilliporeSigma) arba pACYC184 (New England Biolabs) plazmidę. CRISPR-Cas12f1 sistemai, kuri iš pradžių buvo pavadinta Cas14a1 (Harrington et al. 2018), o čia pervadinta į Un1Cas12f1, buvo naudojama pLBH531_MBP-Cas14a1 plazmidė (Jennifer Doudna dovana, Addgene plazmidė Nr. 112500). Cas12f baltymų sekos išvardytos 2 priede (Appendix 2), o nuorodos į plazmidžių sekas, koduojančias į PAM biblioteką nukreiptas Cas12f CRISPR sistemas, pateiktos 3 priede (Appendix 3).

Cas12f dgDNR taikinio kirpimo ir PAM sekos nustatymas

Plazmidinės DNR taikiniai buvo kerpami naudojant Cas12f ribonukleoproteino (RNP) kompleksus, gautus iš modifikuoto lokuso arba sujungiant *E. coli* lizatą, turintį Un1Cas12f1 (Cas14a1) baltymą su T7 transkribuota gRNR (20 nt taikinys). Pirmiausia *E. coli* DH5α arba ArcticExpress (DE3) ląstelės buvo transformuotos CRISPR-Cas12f koduojančiomis plazmidėmis (atitinkamai pACYC arba pETDuet-1 ir pLBH531) ir atrinktos kultūros augintos LB terpėje (30 ml) su chloramfenikoliu (25 μg/ml) (pACYC plazmidės) arba ampicilinu

(100 µg/ml) (pETDuet-1 ir pLBH531 plazmidės). Toliau, plazmidėms su T7 promotoriumi (pETDuet-1 ir pLBH531 plazmidės) raiška indukuota 0,5 mM IPTG, kai kultūros pasiekė OD_{600} 0,5 ir inkubuojamos per naktį 16 °C temperatūroje. Ląstelės (nuo 10 ml) buvo surinktos centrifuguojant ir pakartotinai suspenduotos 1 ml lizavimo buferio (20 mM fosfato (pH 7,0 25 °C temperatūroje), 0,5 M NaCl, 5 % (v/v) glicerolio), papildyto 10 µl PMSF (galutinė koncentracija 2 mM) ir lizuojama ultragarsu. Ląstelių liekanos buvo pašalintos centrifuguojant. 10 µl gauto supernatanto, kuriame yra RNP, buvo tiesiogiai panaudota DNR karpymo eksperimentams. Un1Cas12f1 atveju 20 µl supernatanto buvo sumaišyta su 1 µl RiboLock RNazių inhibitoriumi (Thermo Fisher Scientific) ir 2 µg gRNR.

Cas12f RNP kompleksai buvo naudojami 7 bp ilgio atsitiktinių PAM sekų plazmidžių bibliotekai arba plazmidei, turinčiai fiksuotą PAM ir gRNR taikinį, karpymams. Trumpai tariant, 10 µl Cas12f-gRNR RNP turinčio lizato buvo sumaišyta su 1 µg PAM bibliotekos arba 1 µg plazmidės, turinčios konkretų PAM ir gRNR taikinį, 100 µl reakcijos buferyje (10 mM Tris-HCl (pH 7,5 37 °C temperatūroje)), 100 mM NaCl, 1 mM DTT ir 10 mM $MgCl_2$). Po 1 valandos inkubacijos 37 °C temperatūroje DNR galai buvo pataisyti, pridedant 1 µl T4 DNR polimerazės (Thermo Fisher Scientific) ir 1 µl 10 mM dNTP mišinio (Thermo Fisher Scientific) ir inkubuojant reakciją 20 minučių 11 °C temperatūroje. Tada reakcija buvo stabdoma mėginį kaitinant iki 75 °C 10 min., o 3’-dA iškyšos pridedamos inkubuojant reakcijos mišinį su 1 µl DreamTaq polimerazės (Thermo Fisher Scientific) ir 1 µl 10 mM dATP (Thermo Fisher Scientific) 30 min. 72 °C temperatūroje. Be to, RNR buvo pašalinta inkubuojant 15 minučių 37 °C temperatūroje su 1 µl RNazės A (Thermo Fisher Scientific). Po gryninimo naudojant GeneJET PGR gryninimo kolonėlę (Thermo Fisher Scientific), kirpimo produkto (100 ng) galai buvo sulieti su dvigrandiniu DNR adapteriu, turinčiu 3’-dT iškyšą (100 ng) 1 valandą 22 °C temperatūroje naudojant T4 DNR ligazę (Thermo Fisher Scientific). Po ligavimo atliktas tikslinių sričių PGR padauginimas, pridedant sekas, reikalingas Illumina sekoskaitai atlikti (Karvelis et al. 2015, 2019).

Dvigrandiniai DNR taikinio trūkiai buvo įvertinti gautų sekų pagalba tiriant unikalią jungtį tarp taikinio sekos ir adapterio. Tai buvo atlikta pirmiausia sukūrus sekų rinkinį, atspindintį visus galimus dgDNR trūkių ir adapterio ligavimo rezultatus tiksliniame regione. Pavyzdžiui, trūkis ir adapterio ligavimas iškart po 21-osios taikinio padėties sudarytų tokią seką (5’-CCGCTCTTCCGATCTGCCGGCGACGTTGGGTCAACT-3’), kur adapterio ir tikslinės sekos fragmentus sudaro 5’-CCGCTCTTCCGATCT-3’ ir 5’-GCCGCTCTTCCGCGACGACTTGGGCGACGCTT-3’ sekos atitinkamai. Tada gautų sekų dažnis buvo sudarytas naudojant pasirinktinį

Perl scenarijų (pateiktą adresu https://github.com/cortevaCRISPR/Cas12f-InformmaticsTools.git) ir lyginamas su neigiamomis kontrolėmis (eksperimentų be funkcinių Cas12f kompleksų).

PAM atpažinimas buvo įvertintas taip, kaip aprašyta anksčiau (Karvelis et al. 2015, 2019). Taikinio adapterio ligavimo seka, pasižyminti padidintu dažniu ankstesniame etape, buvo naudojama kartu su 10 bp seka 5' iš 7 bp PAM srities, kad būtų galima nustatyti sekas, lemiančias dgDNR kirpimą. 7 bp PAM seka buvo išskirta apkarpant 5' ir 3' greta esančias sekas, naudojant pasirinktinį Perl scenarijų(pateiktą https://github.com/cortevaCRISPR/Cas12f-InformaticsTools.git) ir dažnį. Išskirtos PAM sekos normalizuotos pagal pradinę PAM biblioteką, kad būtų atsižvelgta į paklaidas, naudojant šią formulę.

Normalizuotas dažnis = (kirpimo dažnis)/((kontrolinis dažnis)/(vidutinis kontrolinis dažnis))

Po normalizavimo buvo apskaičiuota padėties dažnio matrica (PFM) (Stormo 2013) ir palyginta su neigiama kontrole (eksperimentai be funkcinių Cas12f kompleksų). Tik daugiau nei 2,5 karto didesni nukrypimai nuo neigiamos kontrolės buvo laikomi reikšmingais ir rodančiais PAM atpažinimą. Siekiant sumažinti foninį triukšmą, atsirandantį dėl nespecifinio kirpimo bei kitų *E. coli* ląstelių lizato komponentų, šiame darbe nagrinėti tik 10 % dažniausiai pasitaikančių PAM sekų.

CRISPR-Cas12f1 modifikacijos, skirtos nukreipiančiosios RNR gavimui

Anksčiau aprašytos, SpCas12f1 ir AsCas12f1 CRISPR sistemas koduojančios, plazmidės buvo modifikuotos taip, kad koduotų 10× histidino (His) ir maltozę surišančio baltymo (MBP) žymą, sujungtą su *cas12f1* geno N-galu. Be to, buvo įtraukta seka, koduojanti tabako viruso (TEV) proteazės atpažinimo seką (ENLYFQS), skirtą žymens pašalinimui tolimesniuose žingsniuose. SpCas12f1 plazmidė buvo sukirpta XagI ir NcoI restrikcijos fermentais (New England Biolabs) ir reikalingas fragmentas išgrynintas naudojant agarozės gelį (Qiagen). Susintetintas DNR fragmentas (Genscript), kuriame yra 5' XagI restrikcijos vieta, T7 promotorius, lac operatorius ir ribozimo surišimo seka, kartu su seka, koduojančia 10 × His:MBP:TEV žymą, po kurios seka apversta BbsI kirpimo sritis, kuri po sukirpimo susidarytų sritis suderinamas su NcoI. Šis fragmentas buvo sukirptas XagI ir BbsI ir išgrynintas kolonėle (Qiagen). Du išgryninti fragmentai buvo sujungti naudojant T4 DNR ligazę (New England Biolabs), transformuoti į One Shot TOP10 *E. coli* ląsteles (Invitrogen), teigiami klonai atrinkti atlikus Sangerio sekoskaitą. AsCas12f1 buvo naudojama panaši strategija, išskyrus XagI ir AvaI restrikcijos fermentus (New England Biolabs). Nuorodos į plazmidžių

sekas (pMBP-SpCas12f1 ir pMBP-AsCas12f1) pateiktos 3 priede (Appendix 3).

Cas12f1 raiška ir gryninimas

Un1Cas12f1

Un1Cas12f1 baltymas buvo ekspresuojamas *E. coli* BL21(DE3) kamiene iš pLBH531_MBP-Cas14a1 plazmidės (Jennifer Doudna dovana, Addgene plazmidė Nr. 112500). Un1Cas12f1^{D326A} ir Un1Cas12f1^{D510A} ekspresijos plazmidės buvo sukurtos iš pLBH531, naudojant Phusion Site-Directed Mutagenesis Kit (Thermo Fisher Scientific). *E. coli* ląstelės buvo auginamos LB terpėje, papildytoje ampicilinu (100 µg/ml), 37 °C temperatūroje. Po kultivavimo iki OD_{600} 0,6-0,8, temperatūra buvo sumažinta iki 16 °C ir raiška indukuota su 0,5 mM IPTG. Po 16 valandų ląstelės buvo surinktos centrifuguojant, pakartotinai suspenduotos užnešimo buferyje (20 mM Tris-HCl (pH 8,0 25 °C temperatūroje), 1,5 M NaCl, 5 mM 2-merkaptoetanolio, 10 mM imidazolo, 2 mM PMSF, 5 % (v/v) glicerolis) ir suardomos ultragarsu. Ląstelių liekanos buvo pašalintos centrifuguojant. Supernatantas buvo užneštas į Ni^{2+} HiTrap chelatinę HP kolonėlę (GE Healthcare) ir eliuojamas linijiniu didėjančios imidazolo koncentracijos gradientu (nuo 10 iki 500 mM) 20 mM Tris-HCl (pH 8,0 esant 25 °C), 0,5 M NaCl, 5 mM 2-merkaptoetanolio buferyje. Frakcijos, kuriose buvo Un1Cas12f1 baltymų, buvo sujungtos ir po to pernešamos į HiTrap heparino HP kolonėlę (GE Healthcare), kad būtų galima eliuoti naudojant tiesinį didėjančios NaCl koncentracijos gradientą (nuo 0,1 iki 1,5 M). Frakcijos, turinčios tikslinį baltymą, buvo sujungtos ir 10×His:MBP:TEV žyma buvo pašalinta inkubuojant per naktį su TEV proteaze 4 °C temperatūroje. Norėdami pašalinti 10×His:MBP:TEV žymą ir TEV proteazę iš gryninamo mėginio, reakcijos mišiniai buvo užnešti į HiTrap heparino HP 5 kolonėlę (GE Healthcare) ir plauti naudojant tiesinį didėjančios NaCl koncentracijos gradientą (nuo 0,1 iki 1,5 M). Toliau gautas tirpalas buvo užneštas į MBPTrap kolonėlę (GE Healthcare) ir Un1Cas12f1 baltymai surinkti peršokimo frakcijoje. Surinkta frakcija su Un1Cas12f1 buvo dializuojamos 20 mM Tris-HCl (pH 8,0 25 °C temperatūroje), 0,5 M NaCl, 2 mM DTT ir 50 % (v/v) glicerolio saugojimo buferyje ir laikomos -20 °C temperatūroje. Un1Cas12f1 baltymų sekos pasiekiamos 2 priede (Appendix 2).

SpCas12f1 ir AsCas12f1

Norint gauti Cas12f1-RNR kompleksus, *E. coli* ląstelės (Arctic Express (DE3)) buvo transformuotos pMBP-SpCas12f1 ir pMBP-AsCas12f1 CRISPR sistemų plazmidėmis, koduojančiomis ir nukleazę, ir nukreipiančiąją RNR

(gRNR). Atrinktos ląstelių kultūros buvo auginamos LB terpėje, papildytoje ampicilinu (100 µg/ml) ir gentamicinu (10 µg/ml), 37 °C temperatūroje iki OD_{600} 0,6-0,8. Tuo metu temperatūra buvo sumažinta iki 16 °C ir tikslinių baltymų sintezė indukuota su 0,5 mM IPTG. Po 16 valandų ląstelės buvo surinktos centrifuguojant, pakartotinai suspenduotos užnešimo buferyje (20 mM Tris-HCl (pH 8,0 esant 25 °C), 0,25 mM NaCl, 5 mM 2-merkaptoetanolio, 25 mM imidazolo, 2 mM PMSF, 5 % v/v) glicerolis) ir suardomos ultragarsu. Pašalinus ląstelių liekanas centrifuguojant, supernatantas buvo užneštas į Ni^{2+} HiTrap chelatinę HP kolonėlę (GE Healthcare) ir eliuojamas linijiniu didėjančios imidazolo koncentracijos gradientu (nuo 25 iki 500 mM) 20 mM Tris-HCl (pH 8,0 25 °C), 0,25 mM NaCl, 5 mM 2-merkaptoetanolio ir 5 % (v/v) glicerolio buferyje. Frakcijos su Cas12f1-RNR kompleksais buvo dializuojamos 20 mM Tris-HCl (pH 8,0 25 °C temperatūroje), 0,25 mM NaCl, 2 mM DTT ir 50 % (v/v) glicerolio saugojimo buferyje ir laikomos -20 °C temperatūroje.

SpCas12f1 ir AsCas12f1 baltymai (be gRNR) taip pat buvo ekspresuojami ir išgryninti naudojant pMBP-SpCas12f1 ir pMBP-AsCas12f1. Eksperimentams, kuriems reikalingas inaktyvuotas (d) Cas12f1 baltymas, pMBP-SpCas12f1 ir pMBP-AsCas12f1 buvo toliau modifikuoti, įvedant D228A ir D225A modifikacijas į, atitinkamai, SpCas12f1 ir AsCas12f1genus, naudojant Phusion Site-Directed Mutagenesis Kit (Thermo Fisher Scientific). *E. coli* ląstelės buvo auginamos LB terpėje, papildytoje ampicilinu (100 µg/ml) ir gentamicinu (10 µg/ml) 37 °C temperatūroje. Po kultivavimo iki OD_{600} 0,6-0,8, temperatūra buvo sumažinta iki 16 °C ir baltymų ekspresija buvo indukuota naudojant 0,5 mM IPTG. Po 16 valandų ląstelės buvo surinktos centrifuguojant, pakartotinai suspenduotos užmešimo buferyje (20 mM Tris-HCl (pH 8,0 25 °C temperatūroje), 1,5 M NaCl, 5 mM 2-merkaptoetanolio, 25 mM imidazolo, 2 mM PMSF, 5 % v/v) glicerolis) ir suardomos ultragarsu. Ląstelių liekanos buvo pašalintos centrifuguojant. Supernatantas užneštas į Ni^{2+} HiTrap chelatinę HP kolonėlę (GE Healthcare) ir eliuojamas linijiniu didėjančios imidazolo koncentracijos gradientu (nuo 25 iki 500 mM) 20 mM Tris-HCl (pH 8,0 esant 25 °C), 0,5 M NaCl ir 5 mM 2-merkaptoetanolio buferyje. Frakcijos, kuriose yra Cas12f1 baltymo, buvo apjungtos ir vėliau užneštos į HiTrap heparino HP kolonėlę (GE Healthcare). Eliucijai buvo naudojamas tiesinis didėjančios NaCl koncentracijos gradientas (nuo 0,2 iki 1,0 M). Frakcijos, turinčios tikslinį baltymą, buvo apjungtos ir 10×His:MBP:TEV žyma buvo suskaldyta per naktį inkubuojant su TEV proteaze 4 °C temperatūroje. Norint pašalinti 10×His:MBP:TEV žymą ir TEV proteazę, reakcijos mišiniai buvo užnešti į HiTrap heparino HP 5 kolonėlę (GE Healthcare) ir plauti naudojant tiesinį didėjančios NaCl koncentracijos

gradientą (nuo 0,2 iki 1,0 M). Tada surinktos frakcijos su Cas12f1 buvo dializuojamos 20 mM Tris-HCl (pH 8,0 25 °C temperatūroje), 0,5 M NaCl, 2 mM DTT ir 50 % (v/v) glicerolio saugojimo buferyje ir laikomos -20 °C temperatūroje. Cas12f1 baltymų sekos pateiktos 2 priede (Appendix 2).

RNR išgryninimas iš Cas12f1-RNR komplekso

Norint išskirti su Cas12f1 surištas RNR, SpCas12f1 ir AsCas12f1 ribonukleoproteinų (RNP) kompleksai (250 µl) buvo inkubuojami su 5 µl (20 mg/ml) proteinazės K (Thermo Fisher Scientific) 45 min. 37 °C temperatūroje 1 ml 10 mM Tris-HCl (pH 7,5 esant 37 °C), 1 mM EDTA, 1 mM DTT, 100 mM NaCl ir 5 mM $MgCl_2$ buferio. DNR buvo pašalinta inkubuojant 45 min. 37 °C temperatūroje su 10 µl DNazės I (Thermo Fisher Scientific). RNR buvo išgryninta naudojant GeneJET PGR gryninimo kolonėlę (Thermo Fisher Scientific). RNR koncentracija ir grynumas buvo matuojami NanoDrop spektrofotometru, o RNR fragmentų profiliai buvo vizualizuotas atskiriant reakcijos produktus ant TBE-karbamido (8 M) 15 % denatūruojančio poliakrilamido gelio su 0,5×TBE (Tris-borate-EDTA) buferiu (Thermo Fisher Scientific) ir dažant su SYBR Gold (Thermo Fisher Scientific).

RNR sekoskaita ir analizė

Išgryninta RNR buvo paruošta sekos nustatymui naudojant „TruSeq Small RNA Library Preparation Kit“ (Illumina) pagal gamintojo instrukcijas, išskyrus tai, kad buvo atlikta išplėstinis dydžio parinkimas, leidžiantis užfiksuoti ~ 30–300 nt ilgio RNR rūšis. Paruošus biblioteką, 150 nt suporuotų galų sekoskaita buvo atlikta naudojant MiSeq sistemą (Illumina). Gauti duomenys buvo apdoroti pasitelkiant Phred 13 vertės kokybės balą, adapterių sekos apkarpytos su Cutadapt v2.10 ir susieti naudojant Bowtie2 v2.4.2 (Langmead and Salzberg 2012). Tada aprėpties duomenys buvo peržiūrėti naudojant IGV (Robinson et al. 2011) ir nustatytos crRNR ir tracrRNR sekos.

RNR sintezė

Cas12f1 gRNR transkripcijos matricos buvo paruoštos naudojant PGR su persidengiančiais oligonukleotidais su T7 promotoriumi. RNR buvo paruoštos *in vitro* transkripcijos būdu naudojant „TranscriptAid T7 High Yield Transcription Kit“ („Thermo Fisher Scientific“) ir išgrynintos naudojant „GeneJET RNA Cleanup and Concentration Kit“ („Thermo Fisher Scientific“). Šiame tyrime naudotų gRNR sekos pateikiamos 4 priede (Appendix 4).

Cas12f1-gRNR komplekso surinkimas

1 μM išgryninto Cas12f1 baltymo buvo inkubuojama su atitinkama nukreipiančiąja RNR (gRNR) 1:1 moliniu santykiu komplekso surinkimo buferyje (10 mM Tris-HCl (pH 7,5 esant 37 °C), 100 mM NaCl, 1 mM EDTA, 1 mM DTT) ir 37 °C temperatūroje 30 min.

DNR substratų paruošimas

Komplementarūs oligonukleotidai (Metabion), turintys taikinio ir PAM sekas, buvo sujungti ir klonuoti į pUC18 plazmidę per HindIII (Thermo Fisher Scientific) ir EcoRI (Thermo Fisher Scientific) restrikcijos sritis. Nuorodos į plazmidžių sekas pateiktos 3 priede (Appendix 3).

Oligonukleotidų 5' galai pirmiausia buvo radioaktyviai pažymėti naudojant T4 PNK (Thermo Fisher Scientific) ir [γ-^{33}P]ATP (Un1Cas12f1) arba [γ-^{32}P]ATP (SpCas12f1 ir AsCas12f1) (PerkinElmer). dgDNR substratai buvo sukurti hibridizuojant du komplementarius oligonukleotidus, iš kurių vienas jau turi radioaktyvią žymę, įvestą 5' gale. Oligonukleotidų hibridizacija atlikta mėginį kaitinant iki 95 °C, po to lėtai vėsinant iki kambario temperatūros. Oligodupleksų sekos pateiktos 5 priede (Appendix 5).

DNR nukleazinio aktyvumo tyrimai

Reakcijos mišiniai sudaryti iš 3 nM plazmidinės DNR ir 100 nM Cas12f1 RNP komplekso 2,5 mM Tris-HCl (pH 7,5 esant 37 °C), 25 mM NaCl, 0,25 mM DTT ir 10 mM $MgCl_2$ buferyje, skirto Un1Cas12f1 arba 10 mM Tris-HCl (pH 7,5 esant 37 °C), 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$ ir 200 arba 100 mM NaCl buferiuose, skirtuose atitinkamai SpCas12f1 ir AsCas12f1. Mišiniai buvo inkubuojami 46 °C (Un1Cas12f1), 45 °C (SpCas12f1 ir AsCas12f1) arba kaip nurodyta. Reakcija buvo inicijuota pridedant Cas12f1 RNP komplekso ir stabdoma skirtingais laiko intervalais (30 min. Un1Cas12f1 arba 60 min. SpCas12f1 ir AsCas12f1, jei nenurodyta kitaip), sumaišant su 3× reakcijos stabdymo ir užnešimo dažu (0,01 % bromfenolio mėlio, 0,03 % SDS ir 75 mM EDTA 50 % (v/v) glicerolio). Reakcijos produktai buvo analizuojami atliekant agarozės gelio elektroforezę ir dažant etidžio bromidu.

Reakcijos su oligodupleksais arba vgDNR oligonukleotidais paprastai buvo atliekamos sumaišant pažymėtus DNR mėginius su Cas12f1 RNP kompleksu ir inkubuojant 46 °C (Un1Cas12f1) arba 45 °C temperatūroje (SpCas12f1 ir AsCas12f1). Reakcijos mišinius sudaro 1 nM pažymėto substrato, 100 nM Cas12f1 RNP komplekso, ir 5 mM Tris-HCl (pH 7,5 esant 37 °C), 50 mM NaCl, 0,5 mM DTT ir 5 mM $MgCl_2$ buferis jei naudojamas Un1Cas12f1 arba 10 mM Tris-HCl (pH 7,5 esant 37 °C), 1 mM EDTA, 1 mM

DTT, 10 mM $MgCl_2$ ir 200 arba 100 mM NaCl buferiai skirti, atitinkamai, SpCas12f1 ir AsCas12f1. Bendras reakcijos tūris – 100 µl. 6 µl mėginai buvo paimami iš reakcijos mišinio nustatytais laiko intervalais (0, 1, 2, 5, 10, 15 ir 30 min. Un1Cas12f1, 0, 5, 15, 30 ir 60 min. SpCas12f1 arba 0, 1, 5, 15 ir 30 min. AsCas12f1) ir reakcija stabdoma su 10 µl paruošto dažo (95 % (v/v) formamido, 0,01 % bromfenolio mėlio ir 25 mM EDTA). Reakcijos produktai analizuojami denatūruojančio gelio elektroforezės būdu (20 % poliakrilamido, 8,5 M karbamido 0,5 × TBE buferyje).

Kolateralinio aktyvumo tyrimas

M13 vgDNR degradavimo reakcijos buvo pradėtos sumaišant M13 vgDNR (New England Biolabs) su DNR aktyvatoriumi ir Cas12f1 RNP kompleksu 46 °C (Un1Cas12f1) arba 45 °C (SpCas12f1 ir AsCas12f1) temperatūrose. 10 mM Tris-HCl (pH 7,5 esant 37 °C), 1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$ ir 200 arba 100 mM NaCl buferiai buvo naudojami atitinkamai su SpCas12f1 ir AsCas12f1, o 2,5 mM Tris-HCl (pH 7,5 esant 37 °C), 0,25 mM DTT, 10 mM $MgCl_2$ ir 25 mM NaCl su Un1Cas12f1. Galutinį reakcijos mišinį sudarė 3 nM M13 vgDNR (5 nM eksperimentams su Un1Cas12f1), 100 nM vgDNR arba dgDNR aktyvatoriaus (arba be aktyvatoriaus) ir 100 nM Cas12f1 RNP. Pradėjus reakciją, pridedant Cas12f1 RNP, mėginiai buvo paimti nustatytais laiko intervalais (0, 5, 15, 30, 60 min. (ir 90 min. Un1Cas12f1)), sumaišant su 3× reakcijos stabdymo ir užnešimo dažo buferiu (0,01 % bromfenolio mėlio, 0,03 % SDS ir 75 mM EDTA 50 % (v/v) glicerolyje). Reakcijos produktai buvo atskirti atliekant elektroforezę agarozės gelyje ir nudažyti SYBR Gold (Thermo Fisher Scientific). DNR aktyvatorių sekos pateiktos 5 priede (Appendix 5).

DNR surišimo tyrimas

Surišimo tyrimai buvo atlikti inkubuojant skirtingus Cas12f1 RNP kompleksų kiekius (0, 10, 50, 100 ir 250 nM) su 1 nM ^{32}P-5′ pažymėtais vg arba dgDNR substratais (Appendix 5) surišimo buferyje (40 mM Tris-HAc (pH 8,4 esant 25 °C), 1 mM EDTA, 0,1 mg/ml BSA, 10 % (v/v) glicerolio ir 5 mM $Mg(C_2H_3O_2)_2$). Visos reakcijos buvo inkubuojamos 30 minučių kambario temperatūroje (arba kaip nurodyta) prieš elektroforezę (natyvios sąlygos, 8 % (w/v) PAGE). Elektroforezė buvo atliekama kambario temperatūroje 3 val., esant 110 V įtampai, naudojant 40 mM Tris-HAc (pH 8,4 25 °C temperatūroje), 1 mM EDTA ir 5 mM $Mg(C_2H_3O_2)_2$ buferį.

Masių fotometrija pagrįstas molekulinių masių matavimai

Matavimai buvo atlikti OneMP masės fotometru (Refeyn Ltd.). Norint paruošti matavimus, dengiamieji stikleliai (Nr. 1,5 H, 24 × 50 mm, Marienfeld) buvo valomi nuosekliu ultragarsu 5 minutes Milli-Q-vandenyje, izopropanolyje ir Milli-Q-vandenyje. Tada dangteliai buvo džiovinami švaria azoto srove. Cas12f1 RNP komplekso pradiniai matavimo tirpalai buvo paruošti prieš kiekvieną matavimą, sumaišant Cas12f1 baltymą (1 μM) ir gRNR (500 nM) komplekso surinkimo buferyje (10 mM Tris-HCl (pH 7,5 37 °C temperatūroje), 100 mM NaCl, 1 mM EDTA, 1 mM DTT), po to inkubuojama 37 °C temperatūroje 30 min. Norint paruošti gryno Cas12f1 baltymo, gRNR ir DNR mėginių matavimo pradinius tirpalus, atitinkami pradiniai tirpalai buvo atskiesti iki 500 nM koncentracijos kompleksiniame surinkimo buferyje ir inkubuojami 30 minučių 37 °C temperatūroje. DNR surišimo eksperimentams atlikti 200 nM Cas12f1 RNP komplekso ir 25 nM DNR (Appendix 5) buvo sumaišyti surišimo buferyje (40 mM Tris-HAc (pH 8,4 esant 25 °C), 5 mM $Mg(C_2H_3O_2)_2$) ir inkubuoti 30 min. 45 °C temperatūroje. Po inkubacijos visi mėginiai buvo atskiesti 1:10 atitinkamame mėginio buferyje prieš pat matavimą. Prieš matavimus ant masės fotometro buvo uždėtas išvalytas dengiamasis stiklelis, o ant viršaus uždėta tarpinė (CultureWell™ Reusable Gasket, Grace Bio-Labs). Šulinėlis buvo užpildytas 10 μl atitinkamo mėginio buferio, pridėta 10 μl praskiesto mėginio ir 120 s buvo stebima biomolekulių adsorbcija naudojant AcquireMP programinę įrangą (Refeyn Ltd, 2.3.0 versija). Norint konvertuoti išmatuotą ratiometrinį kontrastą į molekulinę masę, kalibravimui buvo naudojamas Un1Cas12f1 ir jo oligomerai nuo 60 (monomeras) iki 250 kDa (tetrameras). Visi masės fotometrijos gauti rezultatai buvo analizuojami naudojant DiscoverMP (Refeyn Ltd, 2.3.0 versija). Visi mėginiai buvo matuojami mažiausiai tris kartus.

Transformacijos ribojimo tyrimai

Transformacijos ribojimo eksperimentai buvo atlikti naudojant *E. coli* Arctic Express (DE3) ląsteles ekspresuojančias CRISPR-Cas12f sistemas (plazmidės, koduojančios CRISRP-Cas12f sistemas, išvardytos 3 priede (Appendix 3). Un1Cas12f1 atveju *E. coli* BL21 (DE3) kamienas buvo transformuotas pGB53 plazmide, kuri buvo sukurta iš pLBH545_Tet-Cas14a1_Locus plazmidės (Jennifer Doudna dovana, Addgene plazmidė Nr. 112501), pašalinant tracrRNR seką ir CRISPR regioną su Bsp1407I ir AvrII restriktazėmis. Šis fragmentas buvo pakeistas gRNR su papildomu T7 promotoriumi, HDV ribozimu ir terminatoriaus seka. Ląstelės buvo auginamos 37 °C temperatūroje iki ~0,5 OD_{600} ir elektroporuojamos su 100 ng

mažo kopijų skaičiaus pSC101 taikininių plazmidžių, gautų klonuojant oligodupleksus per EcoRI ir XhoI arba EcoRI ir NheI restrikcijos vietas į, atitinkamai, pTHSSe_1 (dovana iš Christopher Voigt, Addgene plazmidė Nr. 109233) arba pSG4K5 (dovana iš Xiao Wang, Addgene plazmidė Nr. 74492) plazmides (nuorodos į plazmidžių sekas pateiktos 3 priede (Appendix 3)). Atlikus ko-transformaciją ląstelės buvo toliau skiedžiamos serijiniais 10 kartų skiedimais ir auginamos 37 °C temperatūroje 16–20 valandų ant agarizuotos LB terpės, papildytos induktoriumi ir antibiotikais. Un1Cas12f1 – AHT (50 ng/ml), IPTG (0,5 mM), chloramfenikolis (30 μg/ml) ir karbenicilinas (100 μg/ml); Cas12f2 iš *Micrarchaeota* archaeon (Mi1) – IPTG (0,5 mM), gentamicinas (10 μg/ml), chloramfenikolis (30 μg/ml) ir karbenicilinas (100 μg/ml); visiems kitiems Cas12f baltymams – IPTG (0,5 mM), gentamicinas (10 μg/ml), karbenicilinas (100 μg/ml) ir kanamycinas (50 μg/ml).

Žmogaus ląstelių kultūra ir transfekcija

HEK293T ląstelės buvo įsigytos iš ATCC (katalogo numeris CRL-3216) ir kultivuojamos naudojant Dulbecco modifikuotą erelio terpę (DMEM), papildytą 10 % galvijų vaisiaus serumu, penicilinu (100 V/ml) ir streptomicinu (100 μg/ml) (Thermo Fisher Scientific). Pirmiausia, ląstelės buvo užsėtos į 24 šulinėlių plokštelę 1,4 x 105 ląstelės/šulinėlyje tankiu. Praėjus maždaug vienai augimo dienai, paruoštas transfekcijos mišinys: 1 μg nukleazę ir jos gRNR koduojanti plazmidė (Appendix 3) sumaišyta su 100 μl DMEM (be serumo) ir 2 μl TurboFect transfekcijos reagentu (Thermo Fisher Scientific). Po 15 minučių inkubacijos kambario temperatūroje transfekcijos mišinys buvo lašinamas į kiekvieną šulinėlį, kuriame buvo paruoštos ląstelės. Tada transfekuotos ląstelės buvo auginamos 72 valandas 37 °C temperatūroje, esant 5 % CO_2.

***Zea mays* transformacija**

0,6 μM (vidutinio dydžio) aukso dalelės pirmiausia buvo padengtos SpCas12f1 ekspresijos kasetėmis (Appendix 3), naudojant TransIT-2020, tada surinktos centrifuguojant, plaunamos etanoliu ir pakartotinai suspenduotos naudojant ultragarsą. Tada 10 μl su DNR susietų aukso dalelių buvo įdėta į mikronešiklį ir paliktos išdžiūti. Naudojant PDS-1000/He Gun (Bio-Rad), dalelės buvo bombarduojamos į 9–10 dienų nesubrendusius kukurūzų embrionus (genotipas PH1V69) su 425 lb/in2 plyšimo disku. Laikiniesiems tyrimams taip pat buvo pristatytas genas, koduojantis geltoną fluorescencinį baltymą ZsYELLOW1 N1 (Hoerster et al. 2020), kad būtų lengviau atrinkti tolygiai transformuotus embrionus praėjus trims dienoms po transformacijos. Norint išauginti T0 augalus, kultivavimas po bombardavimo, selekcija ir

augalų regeneracija buvo atliekami naudojant anksčiau aprašytus metodus (Gordon-Kamm et al. 2002), išskyrus *bbm* ir *wus2* genus, kurie buvo ekspresuojami su pakeistais promotoriais. Atitinkamai naudoti kukurūzų fosfolipidų transferazės baltymo (Zm- PLTP) ir kukurūzų auksinu indukuojami (Zm-Axig1) promotoriai (Lowe et al. 2018).

Žmogaus ir *Zea mays* ląstelių genomo redagavimas

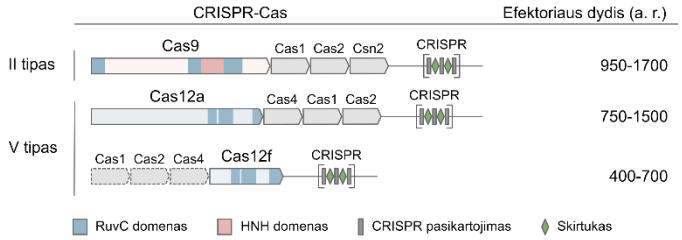
Transfekuotos HEK293T ląstelės buvo surinktos tripsinizuojant, o jų genominė DNR buvo ekstrahuota naudojant QuickExtract tirpalą (Lucigen). Laikiniems *Zea mays* tyrimams, nesubrendę embrionai buvo paimti praėjus 3 dienoms po transformacijos, liofilizuoti, smulkiai sumalti ir jų bendra DNR ekstrahuota naudojant „Synergy 2.0 Plant DNA Extraction Kit“ („Ops Diagnostics“). Regeneruotų augalų atveju iš V2 arba V3 lapų buvo paimti du lapų štampai, drėgnas audinys sumaltas ir DNR ekstrahuota naudojant PB buferį (Qiagen) ir stiklo pluošto 96 šulinėlių mikrofiltro plokštelę (Agilent). Tada PGR buvo atlikta dviem etapais, siekiant padauginti kiekvieną tikslinę DNR sritį ir pridėti sekas, reikalingas Illumina sekos nustatymui ir indeksavimui (Karvelis et al. 2015; Svitashev et al. 2015). Trumpai, 1–4 µl DNR (10–200 ng) buvo panaudoti atliekant pirminį PGR 20–50 µl tūryje. Čia naudojami pradmenys turi specifiniam genomo lokusui komplementarias sekas ir papildomas Illumina sekoskaitai reikalingas 5’ uodegas (Appendix 6). Buvo naudojamas keturių pirminių pradmenų mišinys (žr. F1-F4 Appendix 6). Šie pradmenys identiški, išskyrus 6 nt padėtį, esančią 3’ nuo Illumina sekvenavimui reikalingos srities (Appendix 6). Po pirminės PGR sekė antrasis PGR, naudojant 1 µl pradinės reakcijos ir pradmenis, komplementarius jau pridėtoms Illumina sekoms, taip pat su likusiomis sekoskaitai reikalingomis sekomis (20-50 µl reakcijos tūris). Visus pradmenis ir taikinius galima rasti Appendix 6 ir 7 lentelėse. Abu PGR etapai buvo atlikti 20 ciklų naudojant NEBNext Q5 Hot Start HiFi PCR Master Mix (New England Biolabs), Phusion High-Fidelity PCR pagrindinį mišinį su GC buferiu (Thermo Fisher Scientific) arba Platinum SuperFi II Master Mix su žaliuoju dažu (Thermo Fisher Scientific) pagal gamintojo instrukcijas. Po PGR 5–10 µl buvo atskirti 1–2 % agarozės gelyje, nudažyti RedSafe (iNtRON) arba etidžio bromidu (Sigma) ir vizualizuoti kartu su tinkamo dydžio DNR molekulinės masės standartais. Atrinkta, tinkamo dydžio DNR buvo išgryninta naudojant Monarch PGR gryninimo kolonėlę (New England Biolabs) arba Zymoclean Gel DNA Recovery Kit (Zymo Research), sujungta ekvimoliariškai ir sekvenuota naudojant MiSeq System (Illumina) su pasirinktiniais sekos pradmenimis. Sekos buvo apkarpytos iki 13 Phred kokybės balo ir įvertintos naudojant pasirinktinį scenarijų nr. 35, siekiant aptikti delecija ir insercijas,

įvykusias numatomoje kirpimo vietoje. Delecijų/insercijų tipai buvo sugrupuoti, suskaičiuoti ir reikalaujama, kad jų dažnis būtų bent 30 kartų didesnis nei nustatytas neigiamuose kontroliniuose mėginiuose. Mutantų sekų dažnis buvo apskaičiuotas padalijus bendrą mutantų sekų skaičių iš bendro laukinio tipo sekų skaičiaus. Redaguotų augalų procentas buvo apskaičiuotas padalijus augalų, turinčių specifinę mutaciją, skaičių iš bendro tikslinių modifikacijų augalų skaičiaus.

CRISPR-Cas – pastarąjį dešimtmetį plačiausiai tyrinėjamos prokariotų apsaugos sistemos. Šios sistemos apsaugo savo šeimininkus nuo svetimų nukleorūgščių naudodamos mažas nukreipiančiąsias RNR (gRNR) molekules, transkribuotas nuo CRISPR regiono (Koonin et al. 2017; Koonin et al. 2017; Mohanraju et al. 2016). gRNR kartu su Cas baltymais sudaro efektorinius kompleksus, gebančius kirpti svetimas nukleorūgštis (Jackson et al. 2017; Jiang and Doudna 2017; Koonin et al. 2017). Cas9 ir Cas12 geba kirpti dvigrandinius (dg) DNR, viengrandinius (vg) DNR ir vgRNR taikinius (Chen et al. 2018; Gasiunas et al. 2012; Jinek et al. 2012; Ma et al. 2015; Yan et al. 2019; Zhang et al. 2015). Nors šios endonukleazės yra pritaikytos kaip patikimi genomo redagavimo ir transkripto manipuliavimo įrankiai, daugumos Cas9 ir Cas12 dydžiai riboja jų pernešimą į tikslines ląsteles (Lino et al. 2018; Wu et al. 2010). Šiame darbe pristatoma 10 naujų išskirtinai kompaktiškų (422-603 aminorūgščių dydžio) CRISPR-Cas12f nukleazių. Atliktas išsamus *in vitro* tyrimas, parodytas aktyvumas *E. coli* ir žmogaus bei kukurūzų ląstelėse. Taigi pateikti rezultatai patvirtina jų, kaip naujų genomo redagavimo įrankių, panaudojimo galimybes.

Cas12f – nuo PAM priklausomos dgDNR nukleazės

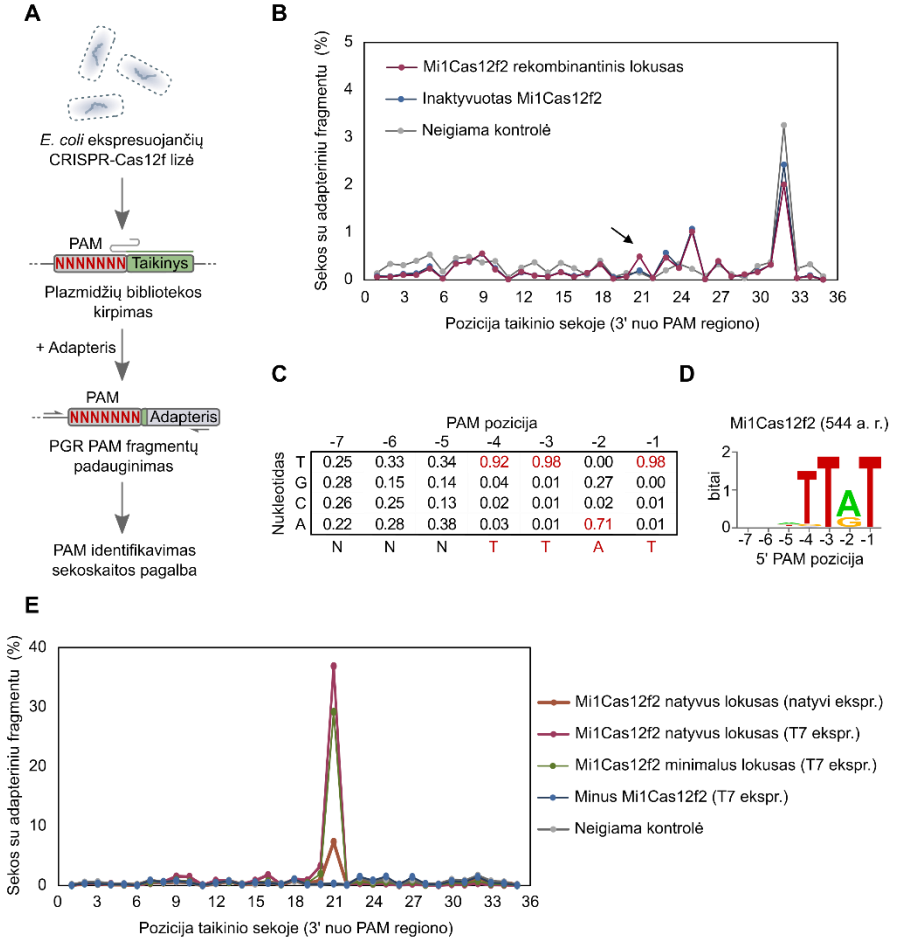
Didėjant charakterizuotų V tipo CRISPR-Cas sistemų įvairovei, atrasta mažiausia žinoma Cas12f nukleazė (Shmakov et al. 2017). Tačiau ji buvo laikoma tarpininku tarp evoliucinio protėvio ir sudėtingesnių bei didesnių V tipo CRISPR-Cas sistemų. Nepaisant mažo dydžio, Cas12f struktūriškai panašūs į kitus Cas12 baltymus. Todėl buvo nuspręsta patikrinti, ar šios nukleazės taip pat gali kirpti dgDNR taikinius (1 pav.).



1 pav. II ir V tipo CRISPR-Cas sistemų lokusai ir jų efektoriniai baltymai. Cas9 – *Streptococcus pyogenes* SF370 (NC_002737.2), Cas12a – *Acidaminococcus sp*. BV3L6 (NZ_AWUR01000016.1), Cas12f – uncultured archaeon (KU_516197.1) ir *Syntrophomonas palmitatica* (NZ_BBCE01000017.1). Efektorinių baltymų RuvC domenas pažymėtas mėlyna spalva, o Cas9 baltyme aptinkamas HNH domenas – rožine. Pilki stačiakampiai ir žali rombai žymi, atitinkamai, CRISPR pakartojimus ir skirtukus.

Cas12f atpažįstamos PAM sekos charakterizavimas

Pirmuoju eksperimento objektu pasirinktas Cas12f2 baltymas iš *Micrarchaeota* archaeon (Mi1) (Appendix 1). Pirmiausia buvo susintetintas CRISPR-Mi1Cas12f2 lokusas su modifikuotu CRISPR regionu, galinčiu nukreipti prieš atsitiktinių PAM sekų plazmidžių biblioteką (Karvelis et al. 2015). PAM nustatymo metodas (Karvelis et al. 2015, 2019) buvo pritaikytas *in vitro* patikrinti Mi1Cas12f2 gebėjimą atpažinti ir perkirpti dgDNR taikinius (2 A pav.). *E. coli*, ekspresuojančių Mi1Cas12f2 baltymo ir gRNR, lizatas buvo sumaišytas su PAM biblioteka. Toliau, DNR trūkiai fiksuoti dvigrandinio adapterio priligavimu, praturtinti PGR pagalba ir identifikuoti sekoskaitos pagalba (Karvelis et al. 2015, 2019). Siekiant identifikuoti DNR kirpimo sritis taikinio sekoje, patikrintas padidėjęs adapterio ligavimo dažnis (neigiama kontrolė – *E. coli* lizatas, kuriame nėra Mi1Cas12f2). 2,5 karto adapterio liguotų sekų dažnio padidėjimas fiksuotas 21 3’ galo taikinio pozicijoje (2 B pav.). Pastebėtas T turtingos sekos (5’-TTAT-3’) atkūrimas iškart po 5’ gRNR taikinio sekos tik su Mi1Cas12f2 paveiktame mėginyje (2 C-D pav.). Buvo atliktas papildomas gautų rezultatų tyrimas: naudotos plazmidės, kuriose yra taikinys su šalia esančia nustatyta 5’-TTAT-3’ PAM seka, bei ląstelių lizatai su didesnio kopijų skaičiaus ir T7 promotoriumi indukuojama Mi1Cas12f2 sistema. Šio eksperimento sekoskaita patvirtino dgDNR kirpimą 21 nuo PAM sekos taikinio pozicijoje (2 E pav.). Kontroliniai eksperimentai naudojant delecijas turinčius variantus patvirtino Mi1Cas12f2 esant vieninteliu reikalingu Cas baltymu dgDNR taikinio atpažinimui ir jo kirpimui (2 E pav.).

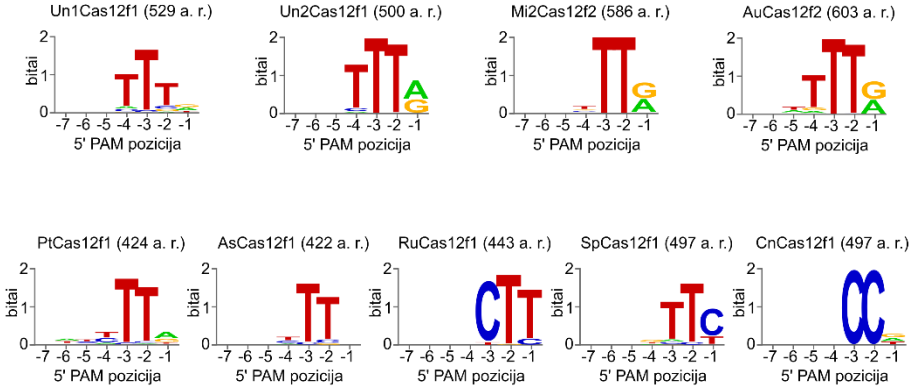


2 pav. Mi1Cas12f1 nukleazės dgDNR karpymas ir atpažįstamos PAM sekos nustatymas. (**A**) Biocheminio metodo, naudoto aptikti dgDNR kirpimą ir nustatyti atpažįstamas PAM sekas, darbo eiga. *E. coli* ląstelės buvo transformuotos plazmidėmis, koduojančiomis CRISPR-Cas12f sistemas, nukreiptas į PAM plazmidžių biblioteką. Inicijuota sistemos ekspresija ir atlikta ląstelių lizė. Gautas lizatas, turintis Cas12f RNP kompleksų, buvo naudojamas dgDNR taikinio kirpimo ir atpažįstamos PAM sekos nustatymui. (**B**) Stebimas nežymus adapterių prisijungusio prie taikinio sekos sekų kiekio pokytis 21-oje pozicijoje nuo PAM sekos. Mi1Cas12f2 CRISPR regionas buvo modifikuotas panaudojant 36 nt ilgio skirtuką prieš PAM plazmidžių biblioteką. (**C**) Remiantis padėties dažnio matrica (PFM – angl. *position frequency matrix*), Mi1Cas12f2 atpažįsta T ir A bp PAM seką, esančią taikinio 5' gale. Tų pačių pozicijų PFM buvo apskaičiuota ir kontroliniams mėginiams. (**D**) Mi1Cas12f1 baltymo atpažintos PAM sekos WebLogo. (**E**) Naudojant dgDNR plazmides, turinčias nustatytą PAM seką ir gRNR taikinį, gautas dažnesnis adapterių sėkmingas surišimas iškart po 21-osios padėties, ypač reakcijose, kuriose ekspresija buvo sustiprinta naudojant T7 promotorių. Kontroliniai eksperimentai atlikti pašalinus cas1, *cas2* ir *cas4* genus (Mi1Cas12f2 minimalus lokusas) arba pašalinus

patį *cas12f2* geną (minus Mi1Cas12f2) patvirtino, jog sėkmingam taikinio atpažinimui ir kirpimui reikalingas tik Mi1Cas12f2 baltymas.

Cas12f atpažįstamų PAM sekų įvairovė

Siekdami detaliau charakterizuoti mažųjų CRISPR-Cas baltymų vykdomą DNR kirpimą nusprendėme įtraukti papildomus 9 Cas12f efektorinius baltymus. Kaip ir kituose Cas12 baltymuose, šiuose naujuose C gale aptinkamas konservatyvus RuvC domenas. Tačiau kai kuriose pasirinktose CRISPR-Cas sistemose trūksta adaptacijos modulio (Koonin et al. 2017; Shmakov et al. 2017).



3 pav. Cas12f baltymų atpažįstamos PAM sekos kerpant dgDNR taikinius. PAM sekų WebLogo gauti konkrečiam Cas12f variantui. *E. coli* ląstelės buvo transformuotos plazmidėmis, koduojančiomis CRISPR-Cas12f sistemas, nukreiptas į PAM biblioteką. Inicijuota sistemos ekspresija ir atlikta ląstelių lizė. Gautas lizatas, turintis Cas12f RNP kompleksų, buvo naudojamas dgDNR taikinio kirpimo ir atpažįstamos PAM sekos nustatymui. Un1Cas12f1 (Cas14a1) *E. coli* lizatas, ekspresuojantis nukleazę, buvo sumaišytas su *in vitro* transkribuota nukreipiančiąja RNR.

Atrinkti du Cas12f1 baltymai iš nežinomų archėjų (Un1 ir Un2) ir du Cas12f2 baltymai iš *Micrarchaeota* archaeon (Mi2) ir *Aureabacteria* bacterium (Au) bei dar mažesni penki papildomi Cas12f1, kurių dydis svyruoja nuo 422 iki 497 aminorūgščių, pagrinde iš bakterijų linijų, pavyzdžiui, iš *Clostridia* ir *Bacilli* (Appendix 1). Visų, išskyrus Un1Cas12f1, variantų atveju susintetintos ekspresijos plazmidės apėmė minimalų CRISPR-Cas sistemos lokusą, su *cas12f* koduojančia seka, spėjamą tracrRNR koduojančia seka, bei CRISPR regionu, modifikuotą PAM nustatymo metodui įvykdyti. Tada, kaip ir eksperimente su Mi1Cas12f2, *E. coli* lizatai iš ląstelių, ekspresuojančių Cas12f nukleazę ir jų nukreipiančiąsias RNR, buvo maišomi

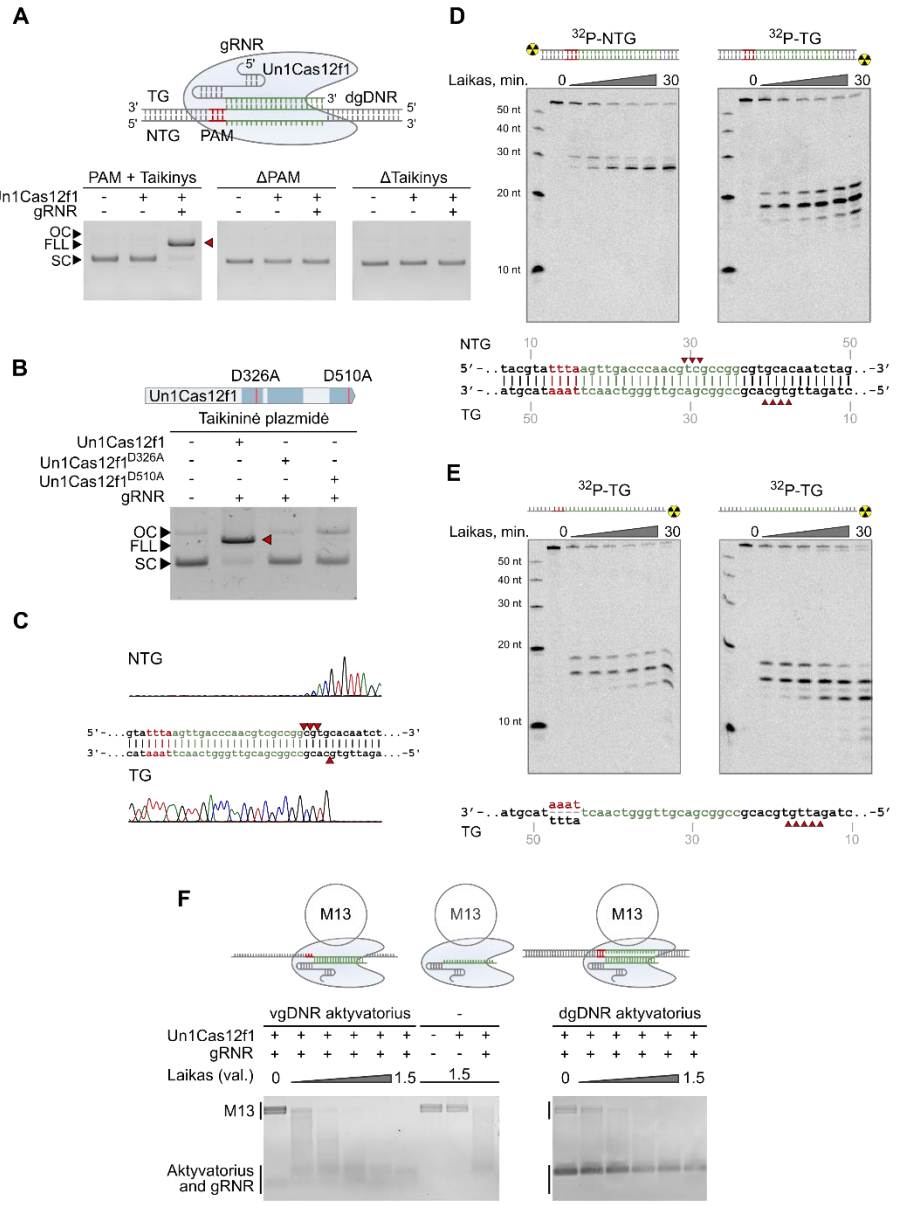
su atsitiktinių PAM sekų plazmidžių biblioteka (2 A pav.). Kirpimo produktai fiksuoti ir atlikta analizė. Un1Cas12f1 atveju *E. coli* lizatas turintis baltymas buvo papildytas *in vitro* transkribuota gRNR, toliau tęsiant eksperimentą tomis pačiomis sąlygomis kaip ir kitų Cas12f baltymų atveju. Visi tirti Cas12f baltymai atliko dgDNR kirpimą 21-24 bp atstumu nuo T ir C turtingų 5' PAM sekų (3 pav.).

Un1Cas12f1 – RNR programuojama vg ir dgDNR nukleazė

Anksčiau buvo manoma, kad Un1Cas12f1, dar žinoma kaip Cas14a1, gali kirpti tik komplementarius gRNR vgDNR taikinius (Harrington et al. 2018). Tačiau aukščiau aprašytas PAM sekos nustatymo tyrimas taip pat parodė Un1Cas12f1 nukleazės aktyvumą prieš dgDNR. Atsižvelgiant į tai, kad specifinė šios nukleazės gRNR jau buvo aprašyta (Harrington et al. 2018), atlikti įvairūs nukleazinio aktyvumo eksperimentai su dgDNR taikiniais. Optimaliomis reakcijos sąlygomis, superspiralizuota plazmidinė DNR, turinti gRNR taikinio ir Un1Cas12f1 PAM (5'-TTTA-3') sekas, buvo pilnai perkirpta. Fiksuojamas dvigrandinis DNR trūkis ir linijinio produkto susidarymas (4 A pav.). Priešingai, DNR taikiniai, kuriems trūksta PAM arba taikinio sekos, nebuvo paveikti. Be to, konservatyvių RuvC aktyviojo centro aminorūgščių (aptinkamos daugumoje būdingų V tipo efektorių) (D326A ir D510A) pakeitimas alaninu (Harrington et al. 2018) panaikino Un1Cas12f1 nukleazinį aktyvumą (4 B pav.).

Toliau, sekoskaitos pagalba, buvo nagrinėjamas Un1Cas12f1 dgDNR kirpimo profilis. DNR grandinės yra kerpamos 20-24 bp atstumu nuo PAM sekos ir formuojami 5' lipnūs galai, galimai naudingi „knock-in" genų modifikacijoms atlikti (4 C pav.). Tačiau stebima mažiau griežtai apibrėžta skilimo padėtis. Nors dgDNR skilimui Un1Cas12f1 reikalinga papildoma PAM (5'-TTTA-3') seka, vgDNR kirpimui ji nebūtina. Formuojamas vienodas vgDNR oligodeoksinukleotidų, su ar be PAM sekos, kirpimo produktų profilis (4 E pav.).

Galiausiai, tirtas daugeliui kitų V tipo šeimos narių būdingas kolateralinis aktyvumas prieš nespecifinius vgDNR substratus (4 F pav.) (Chen et al. 2018; Yan et al. 2019). Tam buvo naudojami dviejų tipų DNR aktyvatoriai. Pirma, buvo patvirtintas Un1Cas12f1 gebėjimas degraduoti vienos grandinės M13 DNR, esant specifiniam vgDNR taikiniui be PAM. Tada taip pat buvo patikrintas dgDNR taikinys, turintis ir 5' PAM ir gRNR komplementarų taikinį. Remiantis 4 F pav. rezultatu, Un1Cas12f1 *trans* vgDNazės aktyvumas matomas su vgDNR, ir dgDNR specifiniais substratais. Vis dėl to ir nesant taikiniui, Un1Cas12f1 RNP kompleksas neselektyviai, bet žymiai lėčiau degradavo nespecifinę vgDNR (4 F pav.).

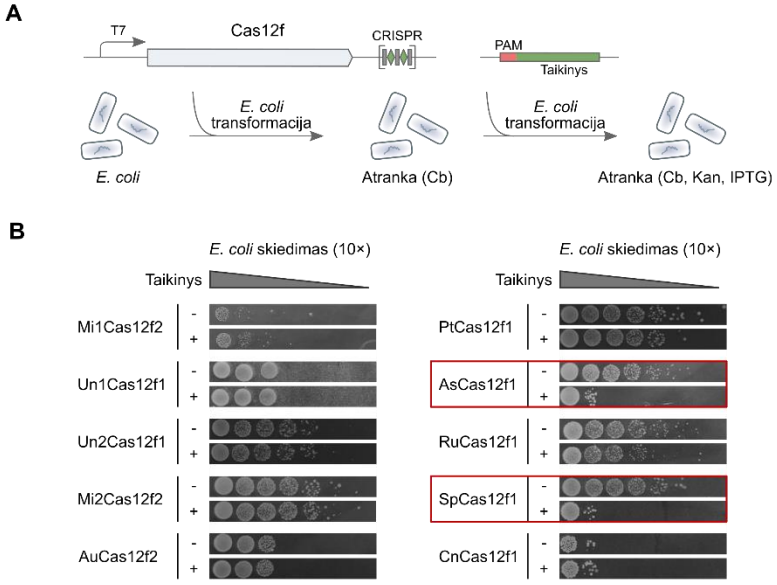


4 pav. Un1Cas12f1 RNP kompleksas yra nuo PAM priklausoma dgDNR endonukleazė. (**A**) Un1Cas12f1 RNP komplekso atpažįstamas dvigubas taikinys sudarytas iš PAM ir gRNR komplementarios sekos. (**B**) Dviejų konservatyvių RuvC aktyviojo centro aminorūgščių pakeitimas alaninu visiškai panaikina DNR nukleazinį Un1Cas12f1 aktyvumą. (**C**) Un1Cas12f1 RNP komplekso perkirptos plazmidinės DNR sekoskaita. Remiantis ja, kirpimas vyksta ties 20-24 bp nuo 3' PAM sekos ir formuojami lipnūs galai. (**D**) Išgryninti Un1Cas12f1 RNP kompleksai kerpa radioaktyviai pažymėtus dgDNR oligodupleksus, turinčius gRNR komplementarų

taikinį (pažymėtą žalia spalva) ir šalia jo esančią PAM seką (raudona spalva). (**E**) Tuo tarpu vgDNR taikinių kirpimas vykdomas nepriklausomai nuo PAM sekos. (**F**) Nespecifinis M13 vgDNR degradavimas stebimas aktyvuojant Un1Cas12f1 taikinio seką turinčiais vg ir dgDNR substratais. Kaip pastebėta Cas12a atveju (Chen et al. 2018), nedidelį nukleazės aktyvumą prieš nespecifinę vgDNR galima stebėti ir nenaudojant DNR aktyvatoriaus. Un1Cas12f1 RNP kompleksai buvo surinkti naudojant gRNR su 20 nt ilgio taikinio seka. SC – superspiralizuota, FLL – linijinė, OC – viengrandinį trūkį turinti plazmidinė forma. NTG ir TG yra atitinkamai, netaikinio ir taikinio (komplementari gRNR sekai) grandinės.

Cas12f aktyvumas *E. coli* ląstelėse

Toliau buvo tiriamas CRISPR-Cas12f sistemų aktyvumas heterologiniame šeimininke atliekant transformacijos ribojimo eksperimentus *E. coli* ląstelėse. Dešimt sistemų buvo užprogramuotos taip, kad būtų nukreiptos prieš naudojamą plazmidinę dgDNR. Šiam *E. coli* transformacijos ribojimo eksperimentui naudota (Burstein et al. 2017; Sapranauskas et al. 2011) mažo kopijų skaičiaus tikslinės plazmidinės DNR ir ekspresijos konstruktai su modifikuotu minimaliu Cas12f CRISPR lokusu (5 A pav.). Norint įvertinti transformacijos efektyvumą, kiekvieno eksperimento metu buvo atlikti 10-ties kartų serijiniai skiedimai, rezultatus lyginant su kontroliniais mėginiais (eksperimentai atlikti su tikslinės sekos neturinčia plazmidine DNR). Ankstesni tyrimai parodė, kad Un1Cas12f1 neveikia heterologiniame *E. coli* šeimininke ir todėl negalėjo detektuoti nuo PAM priklausomo dgDNR kirpimo (Harrington et al. 2018). Atitinkamai, beveik visos šiame darbe atrinktos CRISPR-Cas sistemos neribojo *E. coli* transformacijos DNR substratu (5 B pav.). Tačiau, bene mažiausios iš atrinktų, Cas12f nukleazės iš *Acidibacillus sulfuroxidans* (As) (422 a. r.) ir *Syntrophomonas palmitatica* (Sp) (497 a. r.), sukėlė reikšmingą transformacijos slopinimą. Nežymus poveikis buvo nustatytas ir su *P. thermoglucosidasius* (Pt) bei *Ruminococcus* sp. (Ru) nukleazėmis. Tolimesnių darbų metu, AsCas12f1 ir SpCas12f1 nukleazės buvo pasirinktos detalesniam jų biocheminiam charakterizavimui.



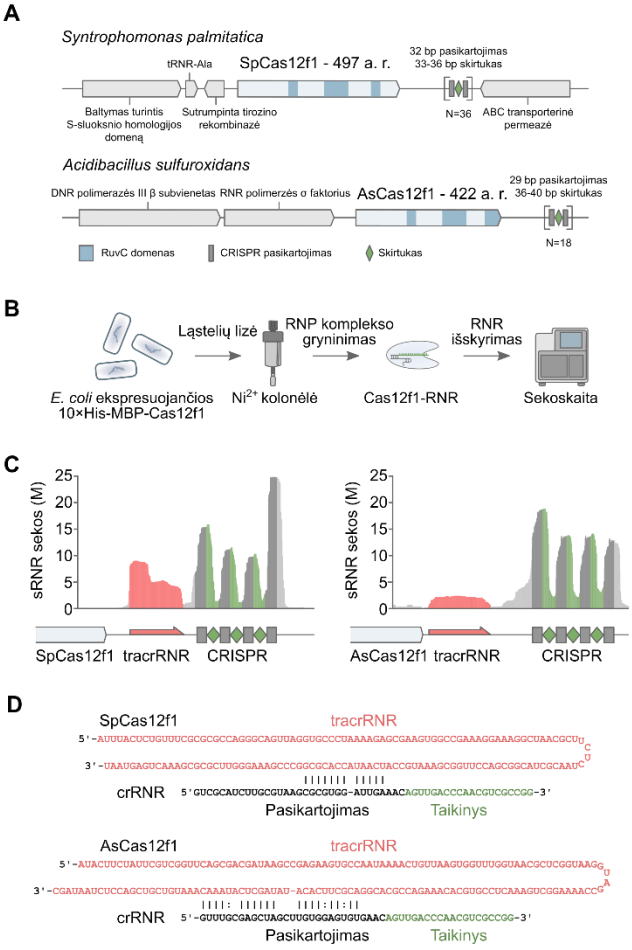
5 pav. Cas12f sistemų sukeliamas plazmidžių transformacijos ribojimas *E. coli* ląstelėse. (**A**) Transformacijos ribojimo į *E. coli* ląsteles eksperimento apžvalga. *E. coli* ląstelės, ekspresuojančios minimalų CRISPR-Cas12f lokusą, buvo transformuotos mažo kopijų skaičiaus taikinio seką turinčia plazmidine DNR. CRISPR regionai konstruoti naudojant 33-39 nt ilgio skirtukus, išskyrus CRISPR-Un1Cas12f1 sistemą, kur CRISPR regionas pakeistas T7 promotoriumi indukuojama gRNR seka (20 nt skirtukas/taikinys). (**B**) Norint įvertinti transformacijos efektyvumą, kiekvieno eksperimento metu buvo atlikti 10-ties kartų serijiniai skiedimai (kontroliniai eksperimentai atlikti su analogiška plazmide, neturinčia taikinio sekos). Raudonais rėmeliais atskirti didžiausiu poveikiu transformacijai pasižymintys variantai.

SpCas12f1 ir AsCas12f1 – naujos mažosios DNR nukleazės

Du iš atrinktų Cas12f efektorių (*Acidibacillus sulfuroxidans* (As) and *Syntrophomonas palmitatica* (Sp)) (6 A pav.) geba kirpti dgDNR taikinius ir heterologinėje, *E. coli*, sistemoje. Nusprendėme patikrinti jų potencialą būti pritaikytiems genomo redagavimo metoduose pradedant nuo nukleaziniam aktyvumui būtinų molekulinių komponentų ir optimalių sąlygų identifikavimo.

Cas12f1 suriša crRNR ir tracrRNR molekules

CRISPR-Cas veikia kaip gRNR programuojamos DNR nukleazės. Šios mažos RNR molekulės, transkribuotos nuo CRISPR lokuso, padeda atpažinti DNR taikinius (Jackson et al. 2017; Jiang and Doudna 2017; Koonin et al. 2017).

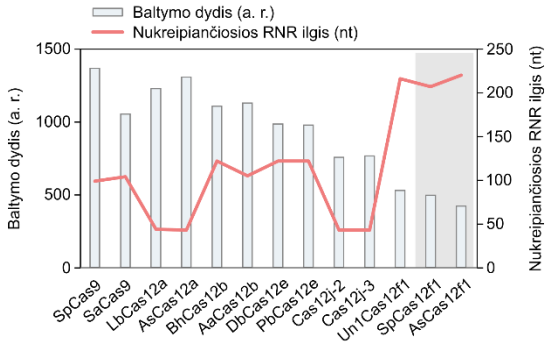


6 pav. SpCas12f1 ir AsCas12f1 CRISPR-Cas lokusai ir efektorinio komplekso komponentai. (**A**) Natyvių SpCas12f1 ir AsCas12f1 CRISPR-Cas lokusų schemos. (**B**) Biocheminio metodo, naudojamo Cas12f1 RNP surištoms RNR molekulėms išskirti ir identifikuoti, eiga. (**C**) Mažųjų RNR sekoskaitos analizė parodė numatomas tracrRNR ir CRISPR regionų sekas. (**D**) *In silico* nustatytos tracrRNR ir crRNR dupleksų bazių poros.

Nukleazės aktyvumui reikalinga gRNR buvo eksperimentiškai identifikuota atlikus Cas12f1 surištų RNR sekoskaitą. CRISPR-Cas12f1 ekspresuojantys plazmidiniai konstruktai buvo modifikuoti įtraukiant 10×His-MBP žymę prie kiekvienos Cas12f1 nukleazės geno sekos N galo. Pažymėti Cas12f1 ribonukleoproteino (RNP) kompleksai buvo sintetinti *E. coli*, išgryninti iš šių ląstelių lizatų, o išskyrus jų surištą RNR buvo atlikta sekoskaita (6 B pav.). Aptiktos dvi praturtintos RNR rūšys: 40-50 nt ilgio CRISPR RNR (crRNR), apimančios dalį CRISPR regiono pasikartojimo sekos, bei dalį skirtuko sekos, ir ilga (atitinkamai 153 nt SpCas12f1 ir 169 nt AsCas12f1) *trans*-aktyvuojanti

RNR (tracrRNR), užkoduota tarp *cas12f1* geno ir CRISPR regiono (6 C-D pav.). Abejose tracrRNR aptikta komplementari CRISPR regiono pasikartojimui seka, kuri galėtų dalyvauti crRNR ir tracrRNR duplekso susiformavime (6 D pav.).

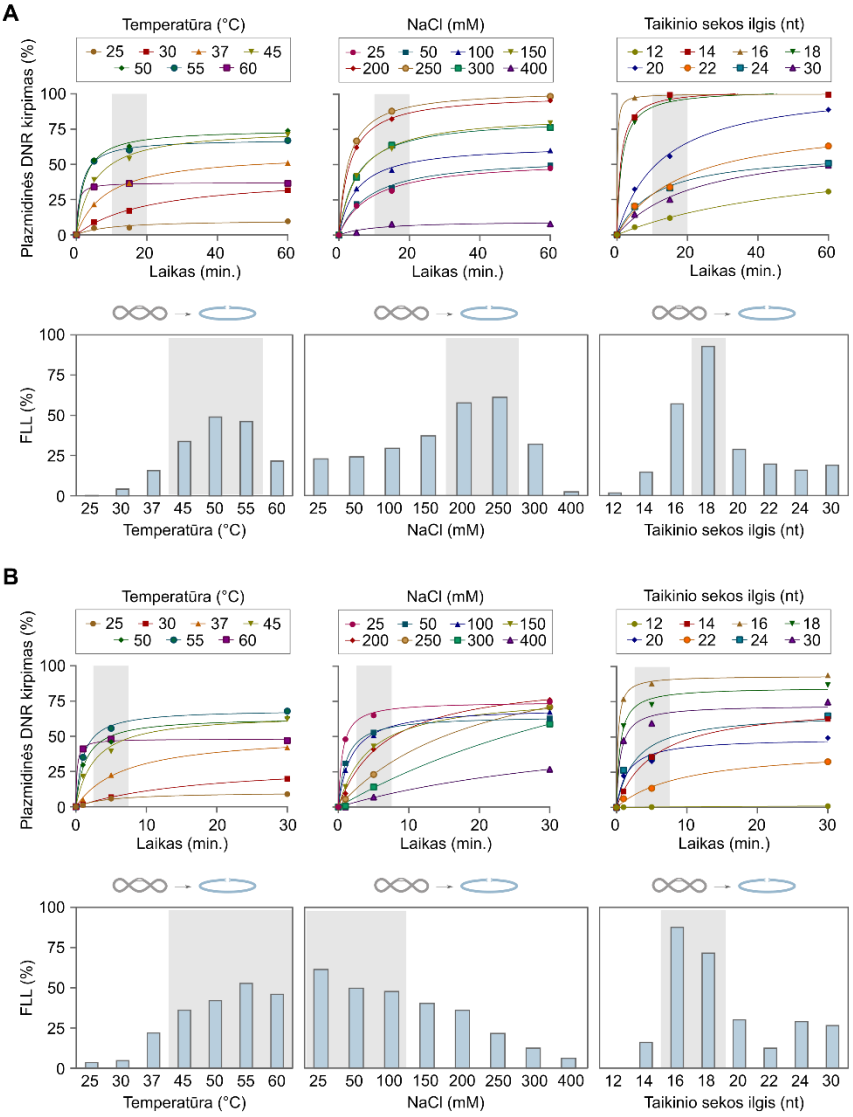
Siekiant dar labiau supaprastinti Cas ir RNR kompleksą, sukonstruota nukreipiančioji RNR (toliau tekste pavadinta gRNR) buvo gauta suliejant atitinkamas crRNR ir tracrRNR sekas per keturių nukleotidų 5'-GAAA-3' jungtukus. Įdomu tai, kad nors SpCas12f1 ir AsCas12f1 baltymai yra kompaktiškiausios iki šiol apibūdintos (<500 a. r.) 2 klasės CRISPR-Cas nukleazės, jų gRNR ilgis pastebimai viršija kitų 2 klasės efektorių gRNR ilgius (7 pav.) (Cong et al. 2013; Harrington et al. 2018, 2020; Liu et al. 2019; Pausch et al. 2020; Ran et al. 2015; Strecker, Jones, et al. 2019; Takeda et al. 2021; Teng et al. 2018; Yan et al. 2019; Zetsche et al. 2015).



7 pav. Cas efektorinių baltymų dydžio ir gRNR ilgių palyginimas. SpCas12f1 ir AsCas12f1 nukleazės pažymėtos pilku fonu.

Cas12f1 optimalios kirpimo sąlygos

Tada buvo įvertintos SpCas12f1 ir AsCas12f1 baltymų biocheminės savybės. Visi RNP kompleksai buvo surinkti sumaišant Cas12f1 baltymą su gauta gRNR suliejus nustatytas crRNR ir tracrRNR sekas per 5'-GAAA-3' jungtį (Appendix 4). Iš pradžių buvo išbandyti trys skirtingi poveikiai plazmidinės DNR kirpimui *in vitro*. Pirma, kaip matyti 8 pav. kairėje esančiuose grafikuose, aplinkos temperatūros padidėjimas lėmė iki dviejų kartų didesnį galutinio produkto kiekį. Nors SpCas12f1 (8 A pav.) ir AsCas12f1 (8 B pav.) yra aktyvūs esant plačiam temperatūrų diapazonui, optimaliausia temperatūra yra 45–55 °C. Antra, įvertintas druskos koncentracijos reakcijos mišinyje poveikis. Nors, kaip ir Un1Cas12f1, AsCas12f1 pirmenybę teikia mažesnei (25-100 mM, NaCl), o SpCas12f1 pasižymi didesniu aktyvumo aukštesnėse

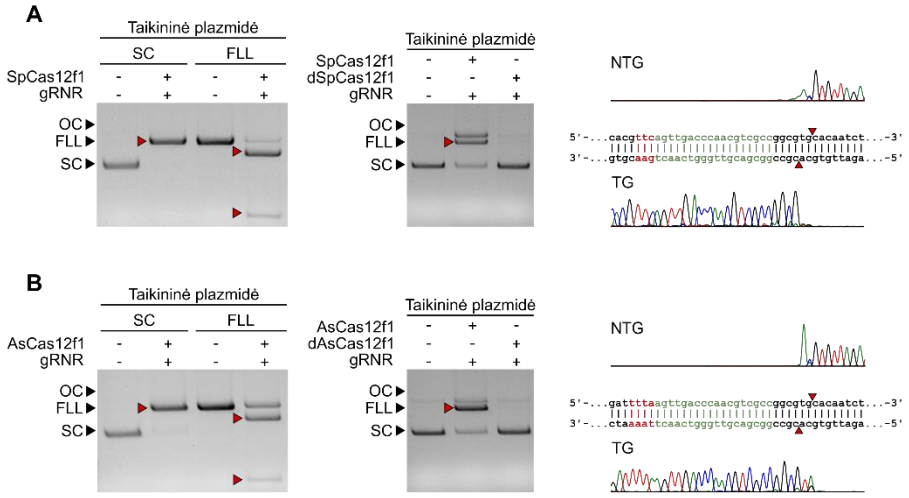


8 pav. Optimalių sąlygų SpCas12f1 ir AsCas12f1 plazmidinės DNR kirpimui nustatymas *in vitro*. SpCas12f1 (**A**) ir AsCas12f1 (**B**) RNP kompleksų plazmidinės DNR kirpimas atliktas nepriklausomai keičiant temperatūrą (su 100 mM NaCl koncentracija), NaCl koncentraciją (esant 45 °C temperatūrai) arba gRNR taikinio sekos ilgį (esant optimaliai temperatūrai ir druskos koncentracijoms). Linijinėse diagramose pilkos sritys nurodo laiko taško vertes, naudojamas histogramose, siekiant palyginti skirtingose sąlygose gautus linijinio (FLL) DNR produkto kiekius. Pilkos sritys histogramose nurodo atrinktas optimaliausias Cas12f1 nukleazinio aktyvumo biochemines sąlygas. Cas12f1 RNP kompleksai buvo surinkti naudojant gRNR su 20 nt ilgio taikinio sekomis.

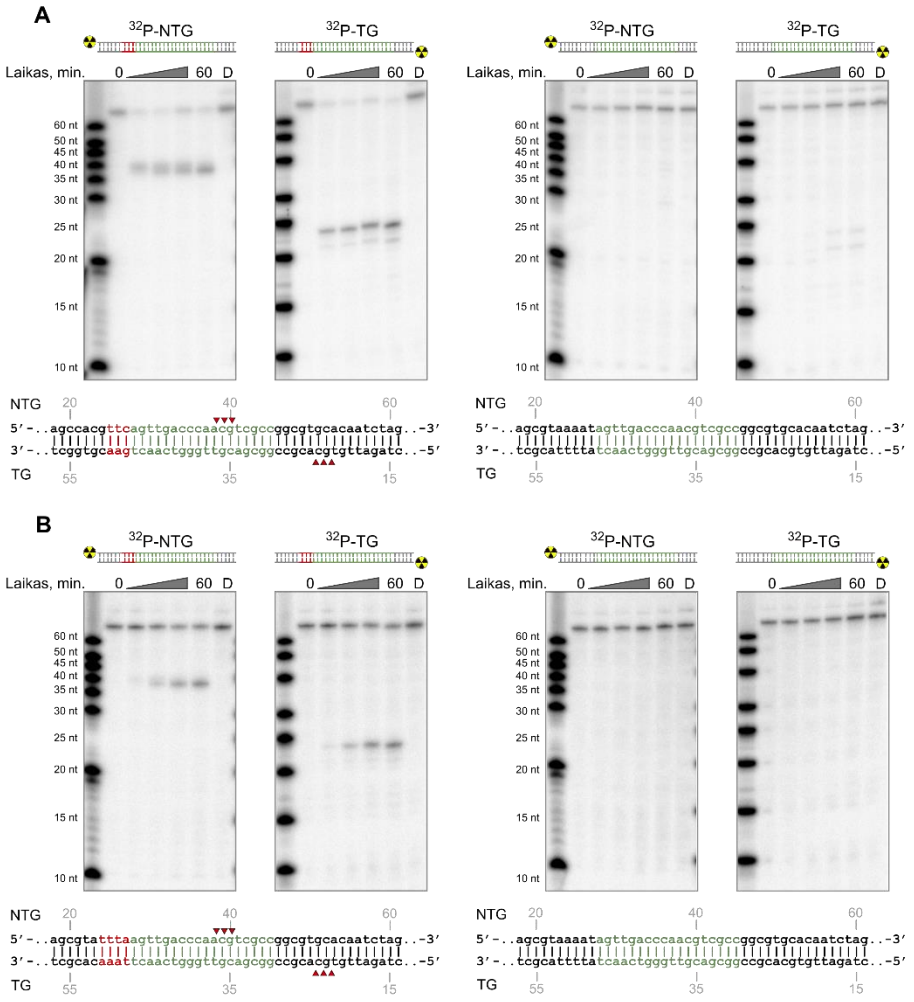
(200-250 mM, NaCl) druskos koncentracijose (8 pav.). Galiausiai, taip pat įvertintas gRNR komplementarios taikiniui sekos ilgio poveikis. Nors Un1Cas12f1 geriausiai funkcionuoja su 20 nt taikinio ilgio gRNR, tiek SpCas12f1, tiek AsCas12f1 reikėjo mažiausiai 16 nt taikinio ilgio, kad būtų sėkmingai perkirptos abi DNR grandinės (8 pav.). Be to, papildomai ištyrus dar dviejų taikinių kirpimą kiekvienai nukleazei, efektyviausi buvo RNP kompleksai su 18 nt ilgio taikinio sekos gRNR variantais.

Cas12f1 efektyviai kerpa dgDNR taikinius

Tolimesnių tyrimų metu patikrinome dgDNR taikinio kirpimo reikalavimus. Esant optimalioms, kiekvieno Cas12f1, reakcijos sąlygoms, *in vitro* iš anksto surinkti RNP kompleksai gali perkirpti tiek superspiralizuotus (SC), tiek linijinius (FLL), specifines PAM sekas turinčius dgDNR taikinius (9 pav.). gRNR nebuvimas arba RuvC aktyviojo centro aminorūgščių pakeitimas visiškai panaikino šį nukleazinį aktyvumą patvirtinant gRNR ir šio domeno svarbą (9 pav.). Kirpimo produktų sekoskaita atskleidė, kad DNR kerpama 22–24 bp atstumu nuo 5’ PAM ir šis kirpimas sutampa su kitų V tipo efektorių atliekamu (Liu et al. 2019; Yan et al. 2019; Zetsche et al. 2015)(9 pav.).



9 pav. Cas12f1 dgDNR taikinių kirpimas. (**A**) SpCas12f1 ir (**B**) AsCas12f1 RNP kompleksai efektyviai karpo superspiralizuotus (SC) ir linijinius (FLL) plazmidinius dgDNR taikinius. Kerpamos abi DNR grandinės suformuojant dvigrandinį trūkį 22–24 bp 3’ atstumu nuo PAM sekos. Tačiau RuvC aktyvaus centro aminorūgščių (dSpCas12f1 – D228A, dAsCas12f1 – D225A) pakeitimas alaninu visiškai panaikina abiejų Cas12f1 baltymų nukleazinius aktyvumus. Cas12f1 ir dCas12f1 RNP kompleksai buvo surinkti naudojant gRNR su 18 nt ilgio taikinio sekomis. SC - superspiralizuota, FLL – linijinė, OC – viengrandinį trūkį turinti plazmidinė forma. NTG ir TG yra atitinkamai, ne-taikinio ir taikinio (komplementari gRNR sekai) grandinės. PAM seka pažymėta raudona spalva, o taikinio – žalia.

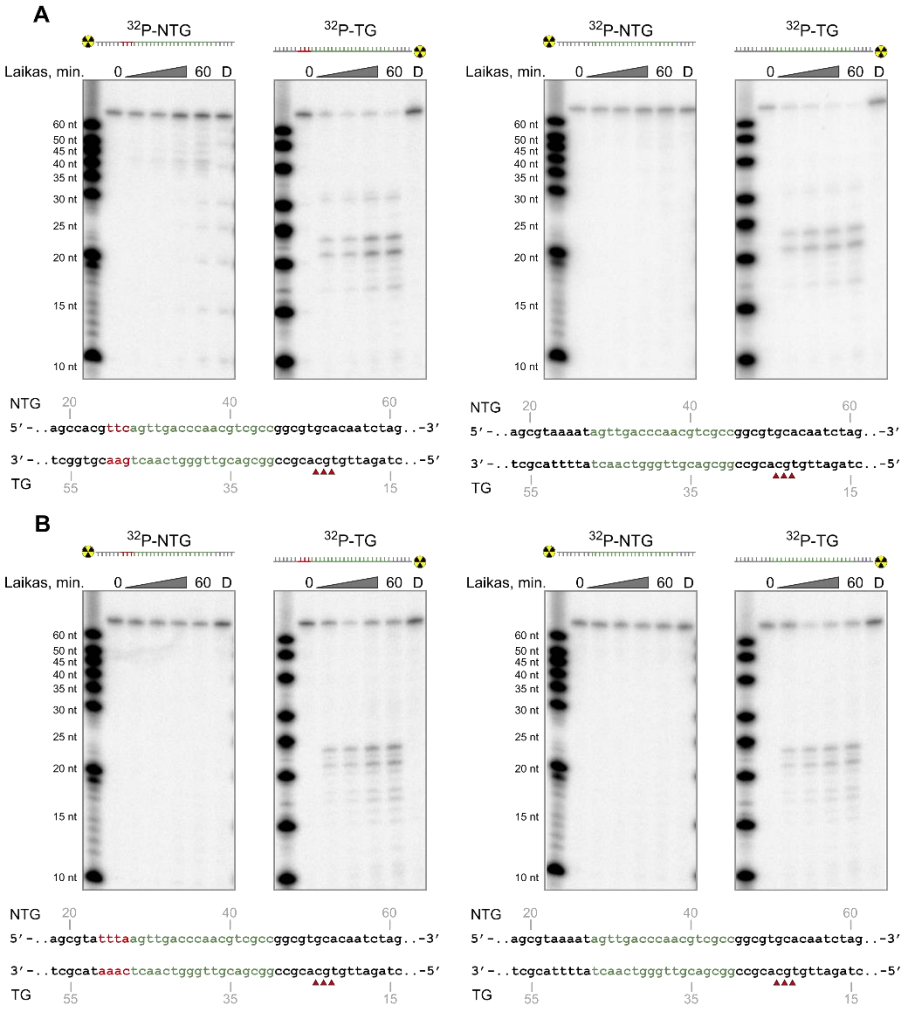


10 pav. Cas12f1 tikslinių oligodupleksų kirpimas. Radioaktyviai pažymėtų vgDNR kirpimas su *in vitro* surinktais SpCas12f1 (**A**) ir AsCas12f1 (**B**) RNP kompleksais. Efektyviam dgDNR taikinių kirpimui Cas12f1 RNP kompleksais yra būtini ir tikslinė, gRNR komplementari taikinio seka (pažymėta žalia spalva) ir papildoma PAM (pažymėta raudona spalva) seka. Cas12f1 ir dCas12f1 RNP kompleksai buvo surinkti naudojant gRNR su 18 nt ilgio taikinio sekomis. NTG ir TG yra atitinkamai, netaikinio ir taikinio (komplementari gRNR sekai) grandinės. D - dCas12f1 RNP kompleksas (dSpCas12f1 – D228A, dAsCas12f1 – D225A), kuris buvo inkubuojamas su DNR substratu 60 min.

Galiausiai, buvo nuspręsta patikrinti PAM sekos svarbą dgDNR taikinių kirpimui naudojant trumpesnius, dvigrandinius oligonukleotidinius substratus. Remiantis 10 pav., sėkmingas DNR kirpimas matomas tik esant 5’ PAM fragmentams tiek su SpCas12f1, tiek su AsCas12f1 nukleazėmis. Skirtingai nuo plazmidinės DNR kirpimo (9 pav.) čia gaunami 5’ lipnius galus

turintys produktai su panašia kirpimo pozicija gRNR komplementarioje (taikinio) grandinėje, bet artimesniu PAM sekai pjūviu ne-taikinio grandinėje (10 pav.). Šie skirtumai galėtų būti susiję su skirtingu dgDNR taikinio grandinių išsukimu ir R-kilpos formavimosi procesu.

Nuo PAM nepriklausomas Cas12f1 vgDNR taikinių kirpimas

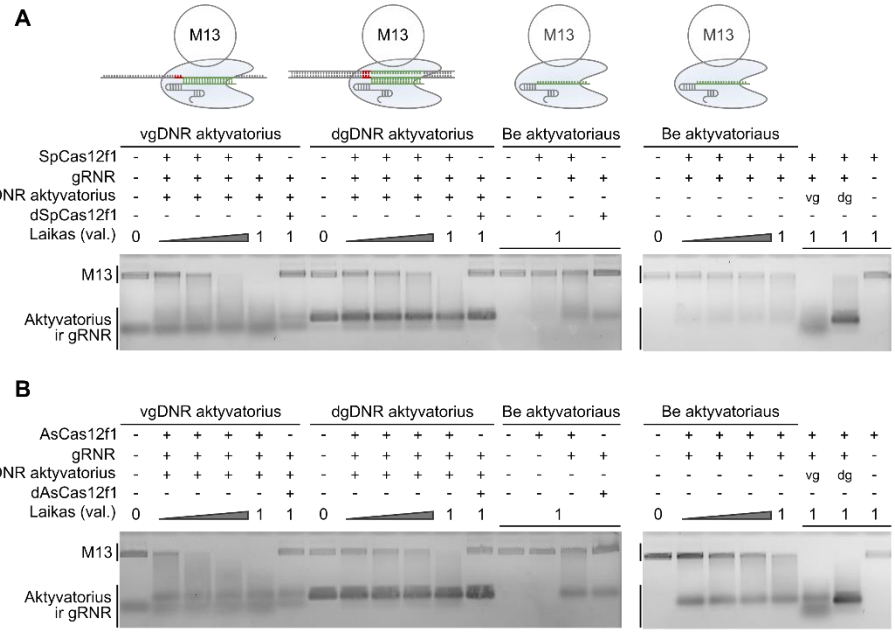


11 pav. Cas12f1 vgDNR taikinių kirpimas. Radioaktyviai pažymėtų vgDNR kirpimas su *in vitro* surinktais SpCas12f1 (**A**) ir AsCas12f1 (**B**) RNP kompleksais. Nepriklausomai nuo PAM sekos (pažymėta raudona spalva), Cas12f1 nukleazės kerpa visus vgDNR substratus, turinčius komplementarią gRNR seką (pažymėta žalia spalva). Cas12f1 ir dCas12f1 RNP kompleksai buvo surinkti naudojant gRNR su 18 nt ilgio taikinio sekomis NTG ir TG yra atitinkamai, ne-taikinio ir taikinio (komplementari gRNR sekai) grandinės. D - dCas12f1 RNP kompleksas (dSpCas12f1 – D228A, dAsCas12f1 – D225A), kuris buvo inkubuojamas su DNR substratu 60 min.

Cas12f1 dgDNR taikinių kirpimui reikalinga papildoma PAM seka. Ši priklausomybė buvo patikrinta ir su vgDNR substratais. Remiantis Harrington et al., 2018 ir 4 E pav., Un1Cas12f1 kerpa vgDNR substratus nuo PAM nepriklausomu būdu, o tai visiškai sutampa su gautu abiejų nukleazių – AsCas12f1 ir SpCas12f1 – aktyvumu (11 pav). Taip pat pastebima mažiau griežtai apibrėžta, bet iš dalies sutampanti su oligodupleksais, kirpimo padėtis.

***Cis* aktyvuotas Cas12f1 *trans* aktyvumus prieš vgDNR**

Siekiant papildyti Cas12f1 ištirtą *in vitro cis* nukleazės aktyvumą, ištirtas *trans* aktyvumas prieš nespecifinę vgDNR. Specifinio taikinio DNR surišimas inicijuoja nespecifinį vgDNR degradavimą (12 pav). Kaip matyti 4 F ir 12 pav., Cas12f1 nukleazių *trans* vgDNazės aktyvumas sukeliamas PAM ir gRNR komplemetarų taikinį turinčių dgDNR arba tik taikinį turinčių vgDNR substratų. Ši savybė buvo parodyta ir su Cas12a nukleazėms, tačiau ji galėtų būti būdinga ir daugiau V tipo šeimos narių. Įdomu, kad Cas12g, kuris specifiškai kerpa vgRNR taikinius, aktyvuotas gali hidrolizuoti tiek vgRNR, tiek vgDNR (Chen et al. 2018; Yan et al. 2019). Vis dėl to, kaip ir su Un1Cas12f1 (4 F pav.), taip ir su SpCas12f1 ir AsCas12f1 pastebimas lėtesnis vgDNR degradavimas net ir nesant jokiam DNR aktyvatoriui (12 pav.). Taigi Cas12f1 *cis* kerpa specifinius dg ir vgDNR taikinius ir *trans*-aktyvuotas – nespecifinius vgDNR substratus.

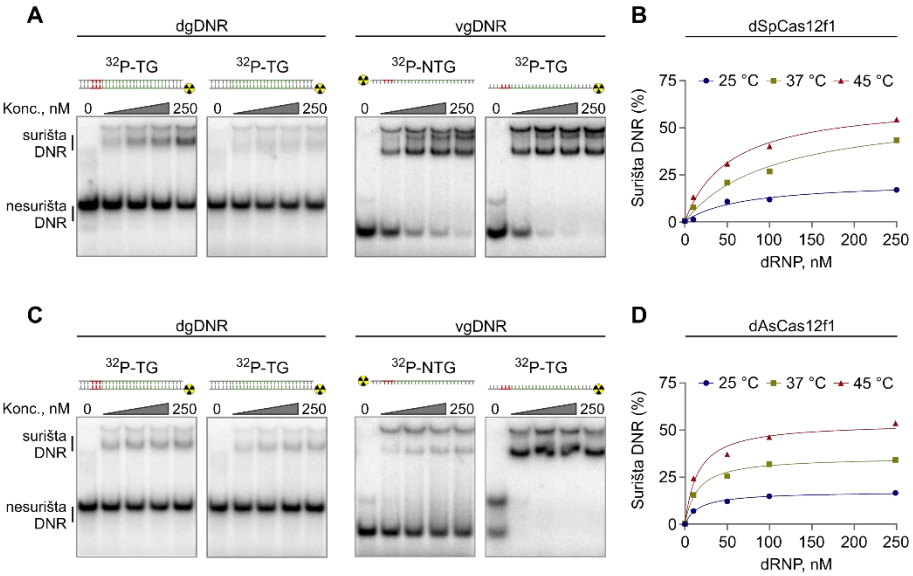


12 pav. Kolateralinis Cas12f1 RNP kompleksų aktyvumas prieš nespecifinę vgDNR. Nespecifinis M13 vgDNR degradavimas stebimas aktyvuojant SpCas12f1 (**A**) ir AsCas12f1 (**B**) taikinių sekas turinčiais vg ir dgDNR substratais. Kaip pastebėta Cas12a atveju (Chen et al. 2018), nedidelį nukleazės aktyvumą prieš nespecifinę vgDNR galima stebėti ir nenaudojant SpCas12f1 ir AsCas12f1 DNR aktyvatoriaus. dCas12f1 RNP kompleksai (dSpCas12f1 – D228A ir dAsCas12f1 – D225A) nedegradavo vgDNR. Cas12f1 ir dCas12f1 RNP kompleksai buvo surinkti naudojant gRNR su 18 nt ilgio taikinio sekomis.

SpCas12f1 ir AsCas12f1 DNR taikinio surišimas

Cas12f1 dgDNR *in vitro* surišimui reikalinga aukštesnė temperatūra

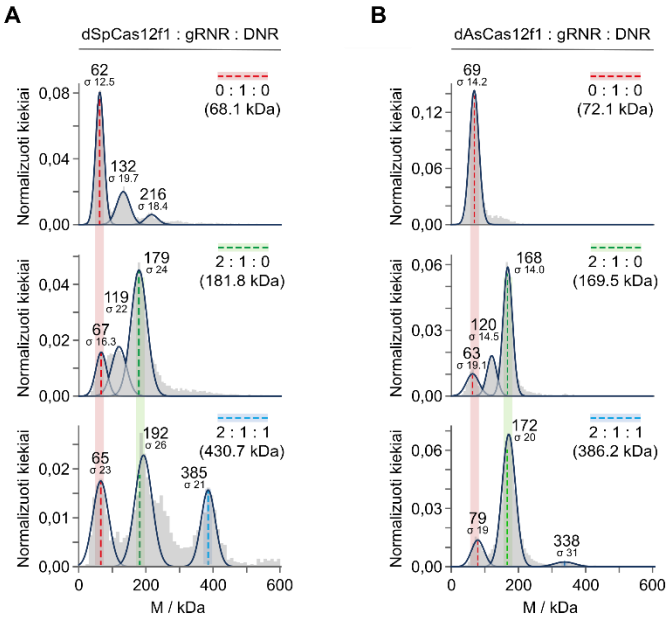
SpCas12f1 ir AsCas12f1 vg ir dgDNR taikinio kirpimui optimali 45–55 °C temperatūra paskatinio taip pat patikrinti temperatūros poveikį DNR surišimui. Remiantis mobilimo poslinkio gelių rezultatais, SpCas12f1 ir AsCas12f1 efektyviai suriša vgDNR taikinius net ir esant mažiausioms RNP koncentracijoms kambario temperatūroje (13 A ir C pav.). Tačiau naudojant dgDNR taikinius, surišimo aktyvumas vos pastebimas. Nuo PAM priklausomas afiniškumas dgDNR substratams žymiai padidėjo po inkubacijos 37 ir 45 °C temperatūrose tiek su SpCas12f1, tiek ir su AsCas12f1 nukleaze (13 B ir D pav.).



13 pav. Cas12f1 RNP kompleksų DNR surišimo aktyvumas. dgDNR ir vgDNR surišimas dSpCas12f1 (D228A) (**A**) ir dAsCas12f1 (D225A) (**C**) RNP kompleksais. Abu, dSpCas12f1 ir dAsCas12f1 suriša tik PAM turinčiu dgDNR taikinius. dAsCas12f1 stipriau suriša TG vgDNR (**C**), tačiau dSpCas12f1 panašiai riša tiek TG, tiek ir NTG (**A**). Skirtingi dSpCas12f1 ir dAsCas12f1 RNP kompleksų kiekiai buvo inkubuojami su 1 nM ^{32}P-5′ pažymėtais dg arba vgDNR substratais kambario temperatūroje. dSpCas12f1 (D228A) (**B**) ir dAsCas12f1 (D225A) (**D**) RNP kompleksų nuo temperatūros priklausomas dgDNR taikinių surišimas. Skirtingi dSpCas12f1 ir dAsCas12f1 RNP kompleksų kiekiai buvo inkubuojami su 1 nM ^{32}P-5′ pažymėtais dgDNR substratais nurodytose temperatūrose. Mėginiai buvo analizuojami naudojant nedenatūruojančią PAGE (poliakrilamido gelio elektroforezę) kambario temperatūroje. DNR substratai schematiškai pavaizduoti virš kiekvieno atitinkamo gelio (PAM pažymėtas raudona spalva, taikinys – žalia). NTG ir TG yra atitinkamai, ne-taikinio ir taikinio (komplementari gRNR sekai) grandinės. dCas12f1 RNP kompleksai buvo surinkti naudojant gRNR su 18 nt ilgio taikinio seka.

Cas12f1 formuojamas keturnaris kompleksas

Nustatyta, kad aktyvų Un1Cas12f1 RNP kompleksą sudaro dvi Un1Cas12f1 nukleazės ir po vieną – jos specifinė gRNR ir tikslinė DNR molekulės (Takeda et al. 2021; Xiao, Li, et al. 2021). Atsižvelgiant į SpCas12f1 ir AsCas12f1 struktūrinius ir funkcinius panašumus su Un1Cas12f1, buvo nuspręsta RNP komponentų santykius matuoti naudojant masių fotometrijos metodą (Young et al. 2018). Nustatytos dvi skirtingos kompleksų kombinacijos: 1) apo forma, Cas12f1 homodimeras surišęs vieną gRNR molekulę, 2) keturnaris kompleksas, kur Cas12f1 homodimeras surišęs vieną gRNR ir vieną DNR taikinio molekulę (14 A (SpCas12f1) ir B (AsCas12f1) pav.).

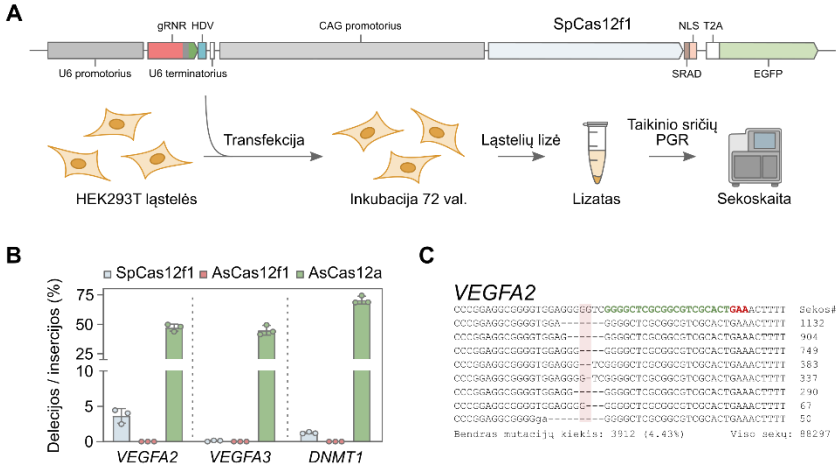


14 pav. Masės fotometrijos metodu nustatytos Cas12f1 baltymų ir RNP kompleksų molekulinės masės. Gautas dSpCas12f1 (D228A) (**A**) ir dAsCas12f1 (D225A) (**B**) molekulinės masės pasiskirstymas. Spalvotos brūkšninės linijos nurodo skirtingų komponentų molekulinių masių smailes: raudona – gRNR, žalia – dCas12f1-gRNR, mėlyna – dCas12f1-gRNR-DNR kompleksas. Teorinės masės nurodytos pateiktos skliausteliuose, o eksperimentinės – priskirtos kiekvienai nustatytai smailei. dCas12f1 RNP kompleksai buvo surinkti naudojant gRNR su 18 nt ilgio taikinio seka.

Cas12f1 – eukariotų genomų redagavimo įrankiai

SpCas12f1 kerpa žmogaus ląstelių genominę DNR

Kadangi abi nukleazės buvo aktyvios *E. coli* (5 pav.), buvo nuspręsta toliau įvertinti SpCas12f1 ir AsCas12f1 genomo redagavimo galimybes. Šiam tyrimui buvo atrinktos žmogaus HEK293T ląstelės, kurioms optimali 37 °C temperatūra. Iš pradžių buvo pasirinktos trys tikslinės vietos VEGFA ir DNMT1 genuose, turinčiuose optimalias PAM sekas - SpCas12f1 (5’-TTC-3’) ir AsCas12f1 (5’-YTTN-3’). Žmogaus ląstelės transfekuotos plazmidėmis, koduojančiomis SpCas12f1 arba AsCas12f1 nukleazę ir gRNR kiekvienai konkrečiai tikslinei vietai. Po 72 val., ląstelės buvo surinktos ir lizuojamos, išgryninta genominė DNR PGR amplifikuota ir sekvenuota siekiant nustatyti, ar kiekvienoje pasirinktoje tikslinėje vietoje buvo įvestos mutacijos (15 A pav.). Nors AsCas12f1 nukleazės aktyvumas HEK293T nebuvo nustatytas, SpCas12f1 atveju visose trijose tirtose tikslinėse vietose buvo



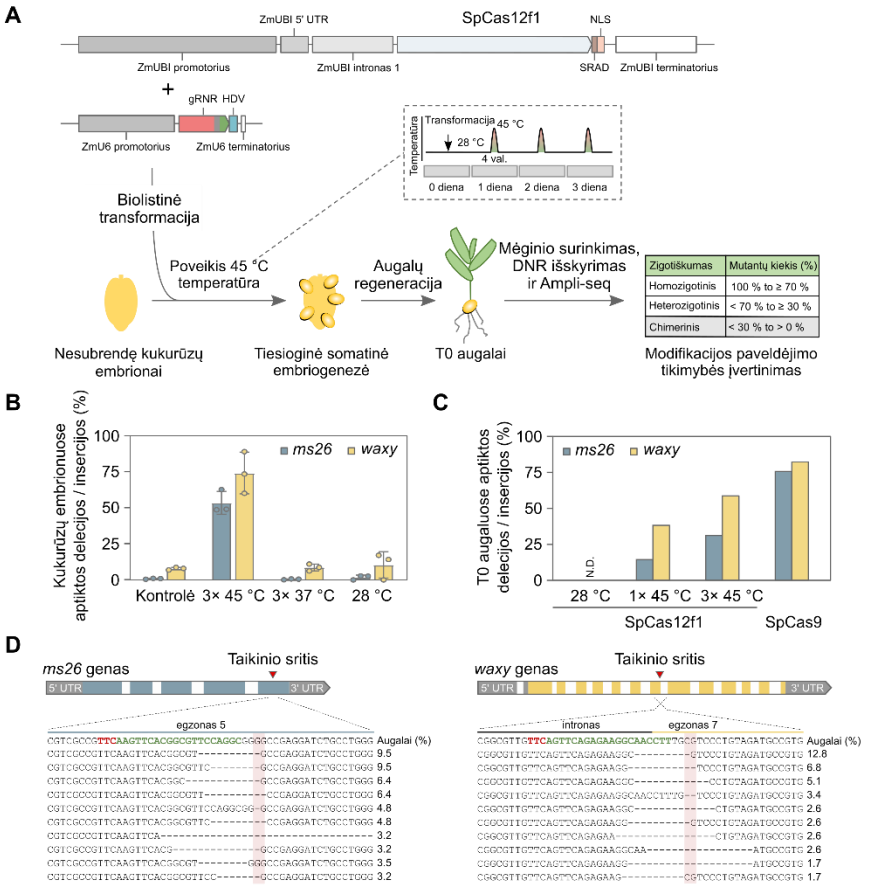
15 pav. SpCas12f1 aktyvumas žmogaus (HEK293T) ląstelėse. (**A)** HEK293T genomo redagavimo eksperimento apžvalga. Ląstelės buvo transfekuotos DNR ekspresijos konstruktais, koduojančiais Cas12f1 arba AsCas12a baltymus ir jų atitinkamą gRNR. **(B)** Delecijų/insercijų dažnis gautas praėjus trims dienoms po HEK293T transfekcijos. Kaip kontrolė, eksperimentai taip pat buvo atlikti su *Acidaminococcus sp*. BV3L6 (As) Cas12a. Stulpeliai rodo vidutines vertes su ± SN (standartinis nuokrypis), o taškai – n=3 nepriklausomų biologinių pakartojimų duomenis. (**C**) Sekoskaitos rezultatas SpCas12f1 nukleazę nukreipus į VEGFA2 taikinio sritį. Atpažįstama PAM seka pažymėta raudonai, o taikinys – žaliai. Numatoma kirpimo vieta identifikuota raudonu fono stačiakampiu.

matyti dvigrandinio DNR trūkio atstatymo požymiai (delecijos/insercijos), o jų dažnis svyravo nuo 0,1 iki 3,6 % (15 B pav.). Lyginant su plazmidinės DNR karpymais *in vitro* (9 A pav.), gautas panašus DNR trūkis 22-24 bp atstumu nuo PAM sekos (15 C pav.). Įdomu tai, kad atitinkamai kirpimas vykdomas iškart už gRNR komplementarios taikinio sekos.

Efektyvus SpCas12f1 DNR kirpimas augaluose

Nors SpCas12f1 sėkmingai parodė DNR nukleazinį aktyvumą žmogaus ląstelėse, gautas DNR mutacijų dažnis pasiekė tik 3,6 % reikšmę (15 B pav.). Remiantis tuo, kad aplinkos 37 °C temperatūra, naudojama HEK293T ląstelių kultivavimui, gali riboti SpCas12f1 DNR taikinio surišimą ir kirpimą, genomo redagavimas buvo ištirtas *Zea mays* (kukurūzų) ląstelėse. Šie eksperimentai buvo atlikti bendradarbiaujant su Joshua K. Young ir jo komanda Corteva įmonėje.

Nustatyta, kad *Zea mays* ląstelės toleruoja trumpus aukštesnės temperatūros pliūpsnius iki 45 °C, o ši yra optimali SpCas12f1 dgDNR taikinio surišimo ir kirpimo temperatūra nustatyta *in vitro* (8 A ir 13 A pav.)



16 pav. SpCas12f1 aktyvumas kukurūzų (*Zea mays*) ląstelėse. (**A**) *Zea mays* genomo redagavimo tyrimo apžvalga. Biolistinės transformacijos metodu efektorinius kompleksus koduojantys konstruktai buvo suleisti į nesubrendusius kukurūzų embrionus. Šie toliau auginti 28 °C temperatūroje, kai kuriais atvejais veikiant aukštesne temperatūra 4 val. vieną arba 3 dienas iš eilės. T0 augalai buvo įvertinti dėl paveldimo redagavimo tikimybės. (**B**) Nustatytas delecijų/insercijų dažnis praėjus trims dienoms po transformacijos laikant 28 °C temperatūroje arba kiekvieną dieną tris kartus iš eilės papildomai paveikus 4 val. 37 °C arba 45 °C temperatūra. Kontroliniai (Ctrl) eksperimentai buvo atlikti transformuojant gRNR neturinčius ekspresijos variantus. Stulpeliai rodo vidutines vertes su ± SN (standartinis nuokrypis), o taškai – n=3 nepriklausomų biologinių pakartojimų duomenis. (**C**) T0 augalų, kuriuose yra paveldima *ms26* ir *waxy* tikslinė mutacija, procentinė dalis. SpCas12f1 eksperimentai buvo atlikti 28 °C temperatūroje, naudojant vieną (1 × 45 °C) arba tris (3 × 45 °C) 4 val. 45 °C temperatūrinius šokus. Eksperimentai su SpCas9 buvo atlikti 28 °C temperatūroje. (**D**) Dešimt gausiausių tikslinių mutacijų, nustatytų mėginiuose su 3 × 45 °C 4 val. temperatūriniais pliūpsniais SpCas12f1 redaguotuose T0 augaluose. Atpažįstama PAM seka pažymėta raudonai, o taikinys – žaliai. Numatoma kirpimo vieta identifikuota raudonu fono stačiakampiu.

(Barone et al. 2020; D. Wang et al. 2020). Pirma, taikiniai buvo atrinkti dviejuose agronomiškai svarbiuose – *male sterile 26* (ms26) ir *waxy*, genuose (Djukanovic et al. 2013; Fan et al. 2009), o kiekvieno taikinio nukleazės ir atitinkamos gRNR ekspresijos plazmidės buvo transformuotos naudojant biolistinį metodą (16 A pav.). Po to, praėjus 24 valandoms, nesubrendę kukurūzų embrionai buvo inkubuojami 45 °C arba 37 °C temperatūroje 4 valandas vieną kartą per dieną iš viso tris dienas, o kaip kontrolė, dalis embrionų buvo inkubuojami 28 °C temperatūroje visą eksperimento laiką (16 A pav.). Galiausiai surinkti embrionai buvo sutrinti ir atlikta jų genomo tikslinių sričių sekoskaita. Tiek *ms26,* tiek *waxy* atveju, tikslinės mutacijos pastebimos tik 45 °C temperatūra paveiktuose mėginiuose (16 B pav.).

Norint toliau ištirti terminio šoko poveikį, transformacija buvo pakartota su viena arba trimis 45 °C inkubacijomis arba be jų. Kaip ir praėjus trims dienoms po transformacijos, susidariusių T0 augalų analizė parodė DNR sekos pakitimus tik po bent vieno terminio apdorojimo 45 °C temperatūroje (16 C pav.). Tiek *waxy*, tiek *ms26* atvejais SpCas12f1 transformuotų augalų, turinčių tikslinę mutaciją, procentas žymiai padidėjo, didėjant terminių poveikių skaičiui. Panašiai kaip ir žmogaus ląstelių genomo redagavimo rezultatuose, SpCas12f1 tikslinius pokyčius daugiausia sudarė delecijos, atsiradusios netoli numatomos kirpimo vietos (16 D pav.). Analogiški eksperimentai, atlikti naudojant *Streptococcus pyogenes* (Sp) Cas9 ir gRNR, užprogramuotą tiksliniams regionams, persidengiantiems su SpCas12f1 *ms26* ir *waxy* taikiniais, patvirtino panašų SpCas12f1 genomo redagavimo potencialą. Vertinant taikinių kirpimo vidurkį, SpCas12f1 redagavimo aktyvumas pritaikius 4 val. terminius pliūpsnius pasiekė nuo pusės iki dviejų trečdalių vertės gaunamos su SpCas9 (16 C pav.).

Apibendrinimas

2 klasės CRISPR sistemos yra išskirtinai įvairios, tačiau visoms būdinga savybė – efektoriniam kompleksui suformuoti reikalingas tik vienas Cas baltymas (Makarova et al. 2020). Šių nukleazių dydis yra labai įvairus: nuo >1000 aminorūgščių (a. r.) Cas9/Cas12a iki 400–600 a. r. Cas12f (Makarova et al. 2020). Kompaktiškos RNR valdomos nukleazės supaprastintų CRISPR-Cas efektorių pernešimą į tikslines ląsteles ir praplėstų jų pritaikymo galimybes genomo redagavimui *in vivo* (D. Wang et al. 2020). Šiame darbe charakterizuotos naujos mažosios Cas12f1 nukleazės, taip pat jas pritaikant genomo redagavimui eukariotinėse ląstelėse.

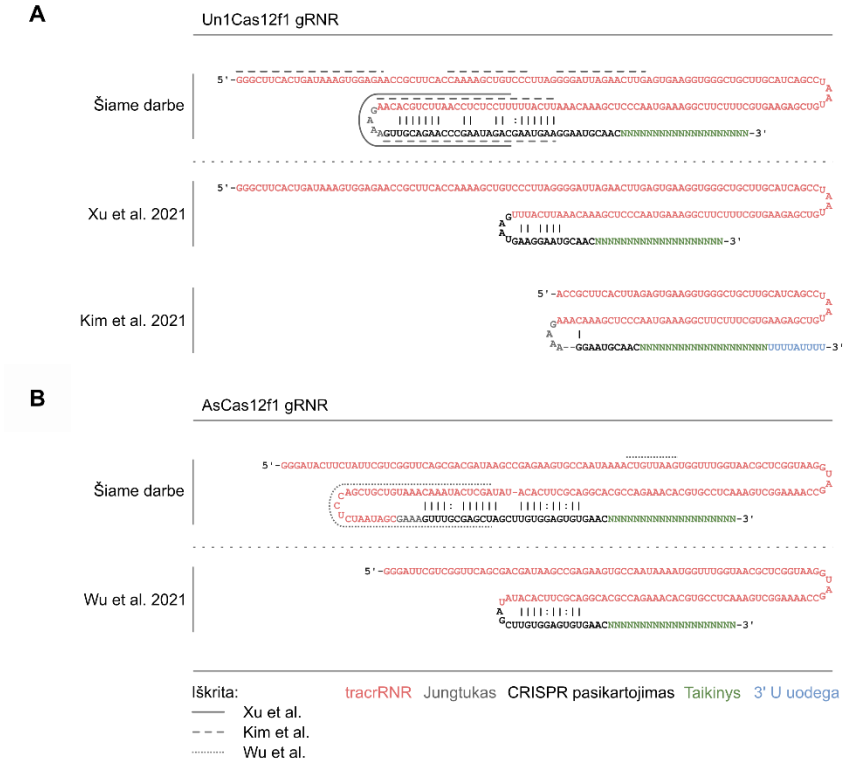
Cas12f1 kerpa vgDNR ir dgDNR taikinius, jei šalia nukreipiančiosios gRNR taikinio yra 5’ PAM seka. Cas12f1 nukleazėms būdinga: 1) kompaktiškas dydis, lyginant su kitomis Cas12 nukleazėmis; 2) nepaprastai

ilgos tracrRNR; 3) greitesnis dgDNR taikinio surišimas ir kirpimas aukštesnėje temperatūroje (45-55 °C); 4) *cis* taikinio kirpimo aktyvuotas netikslinės vgDNR degradavimas; 5) dimerizacija surišus vieną gRNR kopiją. Tokios savybės kaip efektyvus DNR kirpimas aukštose temperatūrose ir kolateralinis aktyvumas prieš nespecifines vgDNR, gali būti pritaikomi Cas12a pagrįstose nukleorūgščių aptikimo platformose, pavyzdžiui, DETECTR ar SHERLOCK (Chen et al. 2018; Gootenberg et al. 2017, 2018; Joung et al. 2020). Be to, jautrumas temperatūrai gali būti pritaikomas ir kontroliuojamam dgDNR kirpimui, kuris šiame darbe pademonstruotas su kukurūzų ląstelėmis. Galiausiai, mažas Cas12f1 fermentų dydis ir savaiminė dimerizacija suteikia pranašumą virusiniam pernešimui į ląsteles.

Pažanga šioje temoje

Nepaisant efektyvaus DNR taikinio kirpimo *in vitro*, Un1Cas12f1 neparodė DNR kirpimo aktyvumo heterologiniame *E. coli* šeimininke (5 pav.) (Harrington et al. 2018). Tačiau Xu ir kt. ir Kim ir kt. pavyko apeitį šį ribojimą atlikus tikslines Un1Cas12f1 ir jo gRNR modifikacijas. Taškinės mutacijos Un1Cas12f1 DNR surišimo kišenėje (D326A/D510A – inaktyvuotas RuvC domenas ir papildomos D143R/T147R/K330R/E528R mutacijos) padidino genų aktyvaciją žinduolių ląstelėse iki tūkstančio kartų (lyginant su laukinio tipo Un1Cas12f1 aktyvumu naudojant šiek tiek pakeistą gRNR) (Xu et al. 2021). Be to, Un1Cas12f1 su D143R, T147R ir E151A modifikacijomis ir gRNR su pasikartojimo: anti-pasikartojimo duplekso sutrumpintu variantu nulėmė didžiausią genų redagavimo dažnį (17 A pav.) (Xu et al. 2021). Kita vertus, Kim ir kt. daugiausia dėmesio skyrė gRNR modifikacijoms, kurias tarpusavyje lygino tiriant Un1Cas12f1 genų redagavimo efektyvumą žinduolių ląstelėse (Kim et al. 2021). Identifikuota optimali gRNR molekulė turi 20 nt tracrRNR 5’ galo sutrumpinimą, prie taikinio sekos pridėtą papildomą 3’-poliuridinilato seką (U4AU4), dalinis numanomos tracrRNR vidinės kilpos sutrumpinimą ir didelį pasikartojimo: anti-pasikartojimo duplekso sutrumpinimą (17 A pav.) (Kim et al. 2021).

Wu ir kt. charakterizavo AsCas12f1 nukleazę atliekant labai panašius biocheminius eksperimentus ir jų gauti rezultatai neprieštarauja gautiems šiame darbe (Wu et al. 2021). Įdomu tai, kad pasirinkta gRNR modifikacija užtikrino AsCas12f1 aktyvumą heterologiniame šeimininke (Wu et al. 2021). Šios modifikacijos apima dalinį numanomos tracrRNR vidinės kilpos sutrumpinimą ir žymiai trumpesnį pasikartojimo: anti-pasikartojimo dupleksą (17 B pav.)(Wu et al. 2021).



17 pav. Šiame ir panašiuose tyrimuose naudotos Un1Cas12f1 ir AsCas12f1 gRNR. (**A**) iliustruoja Un1Cas12f1 gRNR, o (**B**) – AsCas12f1 gRNR. Kiekvienu atveju viršutinė gRNR schema rodo šiame tyrime naudotą tikslią konstrukciją ir nurodo modifikavimo zonas, kuriose buvo atlikti pakeitimai. Šie pakeitimai lyginami su konkrečiomis gRNR, gautomis schemoje įvardintos tyrimo grupės. Atitinkamai, apatinė gRNR schema pateikia optimaliausią, konkrečios tyrimo grupės naudotą variantą.

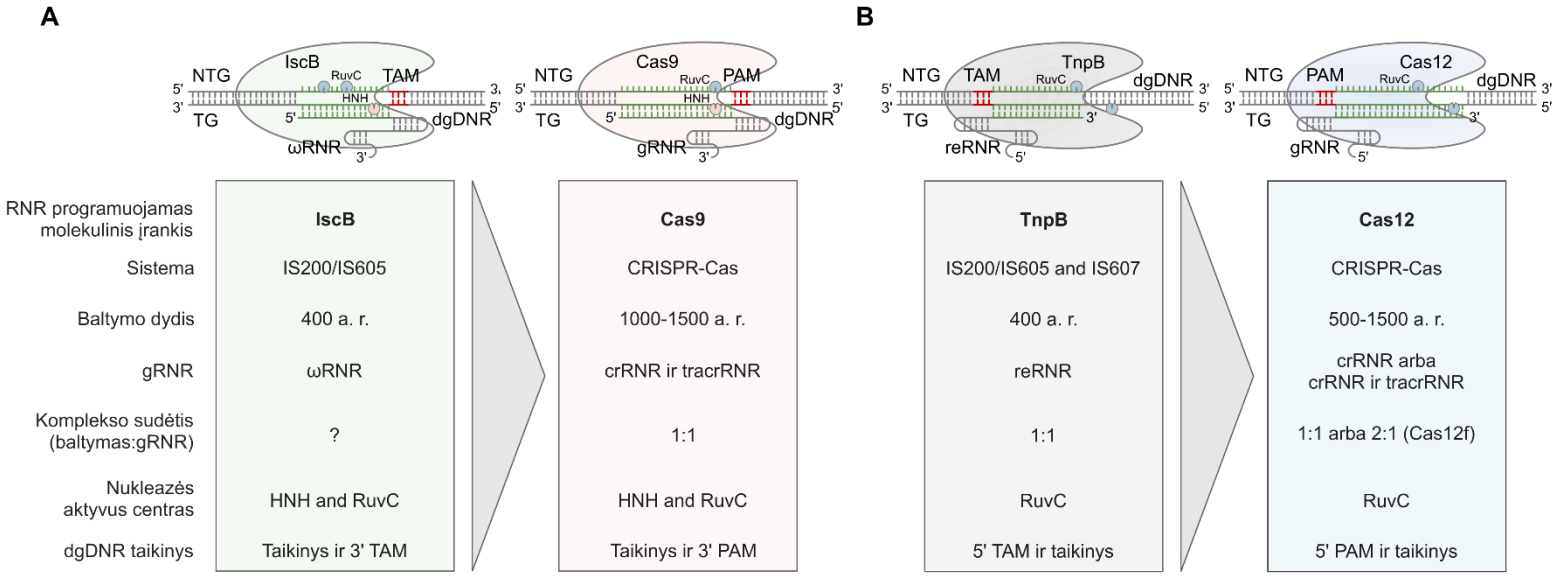
Nuosekliai aprašytos modifikacijos ir toliau įkvėps tobulinti ir sėkmingai pritaikyti Cas12f1, kaip naują genomo manipuliavimo įrankį.

Rengiant disertaciją taip pat buvo paskelbtos dar dvi Cas12f1 RNP kompleksų struktūrų publikacijos. Pastarosiose, dvi nepriklausomos grupės pristatė krioelektroninės mikroskopijos Un1Cas12f1-gRNR-dgDNR ir Un1Cas12f1-gRNR kompleksų struktūras. (Takeda et al. 2021; Xiao, Li, et al. 2021). Papildant gautus ir aprašytus šio tyrimo rezultatus, buvo parodyta, kad Un1Cas12f1 homodimeras suriša vieną gRNR ir vieną taikinio bei PAM sekas turintį dgDNR substratą. Įdomu tai, kad kiekvienas Un1Cas12f1 monomeras turi skirtingą konformaciją, kur tik vienas RuvC nukleazės domenas šiame dimere gali kirpti DNR taikinius (Takeda et al. 2021; Xiao, Li, et al. 2021).

Kita vertus, manoma, kad CRISPR-Cas12f yra tarpinės sistemos tarp TnpB baltymų, randamų transpozonuose, ir didesnių bei sudėtingesnių V tipo

CRISPR-Cas sistemų (Harrington et al. 2018; Makarova et al. 2020; Shmakov et al. 2017). Atitinkamai, IscB baltymai, taip pat randami transpozonuose, yra laikomi pagrindiniais II tipo, Cas9, protėviais (Makarova et al. 2020). Beruošiant šią disertaciją Altae-Tran et al. ir mūsų grupė apibūdinome minėtus IscB ir TnpB baltymus (Altae-Tran et al. 2021; Karvelis et al. 2021). Pirmiausia, buvo rastas pagrindinis papildomas IscB ir TnpB nukleazių aktyvumo komponentas. RNR sekoskaita parodė, jog abu šie baltymai suriša specifines nukreipiamąsias RNR molekules. ωRNR (ω – OMEGA (angl. *Obligate Mobile Element Guided Activity*)) koduojančios sritys buvo aptiktos arti IscB genų su kintama 5’ gale esančia seka (Altae-Tran et al. 2021). Tuo tarpu reRNR (re – dešinysis elementas) koduojanti seka visiškai persidengia su TnpB geno 3’ galu su taip pat kintama seka, tik šiuo atveju 16 nt ilgio ir esančia 3’ gale (Karvelis et al. 2021). Šie kintamieji segmentai veikia kaip Cas9 ir Cas12 efektorių nukreipiančiųjų RNR molekulių taikinių sekos, leidžiančios nukreipti fermentą į bet kurią dominančią DNR sritį. IscB ir TnpB baltymai, kompleksuose kartu su atitinkama RNR, gali kirpti dvigrandinius DNR taikinius. Lygiai taip pat kaip ir Cas9 ir Cas12 specifiškumas PAM (angl. *protospacer adjacent motif*), IscB ir TnpB baltymai atpažįsta trumpas TAM (angl. *transposon-associated motif*) sekas, esančias greta taikinio sekos (Altae-Tran et al. 2021; Karvelis et al. 2021). Įdomu tai, kad šios (TAM) sekos koreliuoja su transpozicijos vietomis, suteikiant daugiau informacijos apie pradinę šių baltymų, kaip transpozonų elementų, funkciją. Galiausia buvo įrodytas ir šių baltymų potencialas genomo redagavimo srityje. IscB ir TnpB sėkmingai kerpa DNR taikinius žmogaus ląstelėse – su *Deinococcus radiodurans* TnpB pasiektas net iki ∼20 % efektyvumas (Altae-Tran et al. 2021; Karvelis et al. 2021). Išsamus IscB palyginimas su Cas9 ir TnpB su Cas12 pateiktas 3.18 paveiksle.

Nuolat augant CRISPR-Cas ir į jas panašių nukleazių kolekcijai, galime tikėtis dar spartesnio postūmio genomo inžinerijos ir kitose genų manipuliacijos srityse. Cas12f kartu su TnpB ir IscB baltymais ne tik užtikrina lengvai programuojamą nukleazinį aktyvumą, tačiau ir leidžia išvengti pernešimo sistemos dydžio apribojimų.



18 pav. Cas9 ir Cas12 nukleazių biocheminių savybių palyginimas su atitinkamomis IscB ir TnpB nukleazėmis. (A) IscB ir Cas9 ir (B) TnpB ir Cas12 – RNR programuojamos nukleazės. Čia pateikiami šių baltymų skirtumai ir panašumai. HNH ir RuvC – nukleazės domenai, ωRNR – OMEGA (angl. *Obligate Mobile Element Guided Activity*) RNR, reRNR – transpozono dešiniojo galo RNR, crRNR – CRISPR RNR, tracrRNR – *trans*-aktyvuojanti RNR, gRNR – nukreipiamoji RNR, šiuo atveju crRNR arba crRNR ir tracrRNR hibridas, PAM – angl. *protospacer adjacent motif*, TAM – angl. *transposon adjacent motif*. Pritaikyta iš (Karvelis et al. 2021), papildant informacija apie IscB iš (Altae-Tran et al. 2021).

IŠVADOS

1. Mažosios Cas12f nukleazės kerpa dgDNR taikinius šalia T arba C turtingų 5’ PAM sekų.
2. *Acidibacillus sulfuroxidans* (As) ir *Syntrophomonas palmitatica* (Sp) Cas12f1 suriša crRNR ir tracrRNR molekules, kurios gali būti supaprastintos jas sujungiant į viena gRNR.
3. Cas12f1:gRNA 2:1 santykiu formuojami kompleksai suriša ir kerpa DNR taikinius bei vykdo kolateralinę vgDNR degradaciją esant aukštesnei aplinkos temperatūrai (45-55 °C) *in vitro*.
4. SpCas12f1 nukleazės gali būti pritaikytos kirpti genominius DNR taikinius žmogaus ir kukurūzų ląstelėse.

ACKNOWLEDGEMENTS

The process of earning a doctorate and writing is long and onerous and it is certainly not done without the help of others. Therefore, I would like to express my deepest gratitude to all of them.

Foremost, I am deeply indebted to my supervisors Dr. Tautvydas Karvelis and Dr. Giedrius Gasiūnas, who provided the opportunity to work in the exciting and rapidly changing CRISPR field and also for patient guidance and support throughout this and previous projects. Special thanks to Prof. Dr. Virginijus Šikšnys for ideas, consultations and support which allowed to proceed and successfully finish this project.

Furthermore, I would like to acknowledge the help of Prof. Dr. Ralf Seidel and Selgar Henkel-Heinecke (University of Leipzig) who performed mass photometry experiments, Prof. Dr. Česlovas Venclovas for valuable insights from structural modeling, Dr. Arūnas Šilanskas for performing protein purifications, Rimantė Žedaveinytė for cell transfection experiments, Karolina Budre for plasmid interference experiments, Dr. Giedrius Sasnauskas, Dr. Mindaugas Zaremba and Dr. Miglė Kazlauskienė for countless advice and practical suggestions, as well as for reagents you shared over the years. I would also like to extend my gratitude to Joshua K. Young and all the team at the DuPont Pioneer (Corteva) company for collaboration throughout this study, specifically for the successful selection of research objects, deep sequencing of the samples, and genome editing in maize cells. I am extremely grateful to all co-authors of the publications, that were involved in the work described in this thesis. Moreover, I am grateful to Prof. Dr. Saulius Serva and Dr. Remigijus Skirgaila for critical reading and comments on this doctoral thesis.

Thanks should also go to everyone in the Department of Protein-RNA Interactions. Your support is deeply appreciated. My success would not have been possible without the effort of the whole “TK team”: Dr. Tautvydas Karvelis, Karolina Budrė, Rimantė Žedaveinytė, Gytis Druteika, Ieva Lingytė, Brigita Duchovska. Thank you! Special thanks to the V337 office team, Dr. Arūnas Šilanskas and Dalia Smalakytė, whose patience, help, and support are unmeasurable.

Lastly, but most importantly, I would like to express my deepest appreciation to my family and closest friends who, although not always understood the details of this job, were always excited for me. Your encouragement is the source of my energy and motivation.

CURRICULUM VITAE

Greta Bigelytė

Department of Protein – DNA interactions, Institute of Biotechnology,
Life Sciences Center, Vilnius University
7 Saulėtekio Ave, LT- 10257 Vilnius
Phone no. +370 614 24332
E-mail: greta.bigelyte@bti.vu.lt

EDUCATION

2015-2017 Master's degree in Biochemistry, Vilnius University
2011-2015 Bachelor's degree in Biochemistry, Vilnius University

SCIENTIFIC WORK EXPERIENCE

Vilnius University, Institute of Biotechnology, Department of Protein- DNA interactions

2018-present Junior Research Scientist
2015-2018 Laboratory Assitant

PUBLISHED SCIENTIFIC RESEARCH

- **Bigelyte, G.***, Young, J.*, Karvelis, T.*, Budre, K., Zedaveinyte, R., Djukanovic, V., Ginkel, E., Paulraj, S., Gasior, S., Jones, S., Feigenbutz, L., Clair, G., Barone, P., Bohn, J., Acharya, A., Zastrow-Hayes, G., Henkel-Heinecke, S., Silanskas, A., Seidel, R., Siksnys, V. Miniature type V-F CRISPR-Cas nucleases enable targeted DNA modification in cells. *Nature Communications* **2021** Dec; 12(1): 6191; doi: 10.1038/s41467-021-26469-4.
- Karvelis, T., Druteika, G., **Bigelyte, G.**, Budre, K., Zedaveinyte, R., Silanskas, A., Kazlauskas D., Venclovas, Č., Siksnys, V. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* **2021** Oct; 599: 692–696; doi: 10.1038/s41586-021-04058-1.
- Gasiunas, G., Young, J.K., Karvelis, T., Kazlauskas, D., Urbaitis, T., Jasnauskaite, M., Grusyte M., Paulraj, S., Wang, P., Hou, Z., Dooley, S., Cigan, M., Alarcon, C., Chilcoat, D., **Bigelyte, G.**, Curcuru, J., Mabuchi, M., Sun, Z., Fuchs, R., Schildkrout, E., Weigele, P., Jack, W., Robb, B., Venclovas, Č., Siksnys, V. A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nature Communications* **2020** Nov; 11: 5512; doi: 10.1038/s41467-020-19344-1.
- Karvelis, T.*, **Bigelyte, G.***, Young, J.*, Hou, Z., Zedaveinyte, R., Budre, K., Paulraj, S., Djukanovic, V., Gasior, S., Silanskas, A., Venclovas, Č., Siksnys, V. PAM recognition by miniature CRISPR-Cas12f1 nucleases

triggers programmable double-stranded DNA target cleavage. *Nucleic Acid Research* **2020** May; 48(9): 5016-5023; doi: 10.1093/nar/gkaa208.

- Karvelis, T., Gasiunas, G., Young, J., **Bigelyte, G.**, Silanskas, A., Cigan, M., Siksnys, V. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biology* **2015** Nov; 16, 253; doi: 10.1186/s13059-015-0818-7.

[*] – these authors contributed equally.

AWARDS AND DISTINCTIONS

2021 – Vilnius University Rector's Science Prize for significant research.

2020 – Diploma for the best oral presentation in "13-oji jaunųjų mokslininkų konferencija „Bioateitis: gamtos ir gyvybės mokslų perspektyvos", Vilnius, Lithuania.

2020 – Diploma for the best poster presentation in VitaScientia. Vilnius, Lithuania.

2018 and **2021** – Lithuanian State Science and Studies Foundation scholarship for Ph.D. students.

2016 – Together with other authors received Vilnius University award for the best publication in physical, biomedical, and technology science areas ("Rapid Characterization of CRISPR-Cas9 Protospacer Adjacent Motif Sequence elements").

TEACHING AND ORGANIZATIONAL EXPERIENCE

2020 – VitaScientia Vilnius – international conference for young scientists; participation in organizational team.

2018 – CRISPR 2018 Vilnius – international conference for CRISPR-Cas systems research; participation in organizational team.

2018 and **2017** – participation in European Biotech week – lab experiments "CRISPR-Cas kinetics" for secondary school students.

2016, **2017,** and **2018** – consulting Vilnius iGEM (International Genetically Engineered Machine Competition) teams.

2014 – participation in preparation and evaluation of exercises for 27th international exact sciences olympiad "Riga-Vilnius-Tallinn-Helsinki-Vasteras-St. Petersburg".

2011 and **2012** – participation in "Pažinimas" (*eng.* knowledge) – extramural chemistry school for secondary school students, as a consultant.

MAIN RESEARCH INTEREST

CRISPR-Cas and other cell defense systems, their characterization, optimization, and appliance for genome editing.

NOTES

NOTES

NOTES

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius
El. p. info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Tiražas 15 egz.