https://doi.org/10.15388/vu.thesis.544
https://orcid.org/0009-0008-2847-4386

VILNIUS UNIVERSITY

Tomas Urbaitis

Identification and Characterization of Novel CRISPR-Cas Nucleases

DOCTORAL DISSERTATION

Natural Sciences,
Biochemistry (N 004)

VILNIUS 2023

The dissertation was prepared between 2018 and 2022 at LLC „Caszyme“ together with Vilnius University, Life Sciences Center. The research was supported by Research Council of Lithuania (ES structural funds, DSP, l.r. 300949339)

Academic supervisor – Dr. Giedrius Gasiūnas (Vilnius University, Natural Sciences, Biochemistry – N 004).

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

Chairman – Prof. Dr. Edita Sužiedelienė (Vilnius University, Natural Sciences, Biochemistry – N 004).
Members:
Dr. Marc Güell (Pompeu Fabra University, Spain, Natural Sciences, Biochemistry – N 004);
Dr. Stephen Jones (Vilniaus University, Natural Sciences, Biochemistry – N 004);
Prof. Dr. Rolandas Meškys (Vilniaus University, Natural Sciences, Biochemistry – N 004);
Dr. Rapolas Žilionis (Vilniaus University, Natural Sciences, Biochemistry – N 004).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 2 pm on 27 September 2023 in meeting room R 401 of the Vilnius University Life Sciences Center.
Address: Saulėtekio Ave 7, R 401, Vilnius, Lithuania
Tel. +37062281505; e-mail: tomas.urbaitis@caszyme.com
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.544
https://orcid.org/0009-0008-2847-4386

VILNIAUS UNIVERSITETAS

Tomas Urbaitis

Naujų CRISPR-Cas nukleazių identifikavimas ir charakterizavimas

DAKTARO DISERTACIJA

Gamtos mokslai,
Biochemija (N 004)

VILNIUS 2023

Disertacija rengta 2018–2022 metais UAB „Caszyme“ kartu su Vilniaus universiteto Gyvybės mokslų centru.
Mokslinius tyrimus rėmė Lietuvos mokslo taryba (ES strukt. F., DSP, l.r. 300949339)

Mokslinis vadovas – dr. Giedrius Gasiūnas (Vilniaus universitetas, gamtos mokslai, biochemija – N 004).

Gynimo taryba:
Pirmininkė - prof. dr. Edita Sužiedėlienė (Vilniaus universitetas, gamtos mokslai, biochemija – N 004).
Nariai:
Dr. Marc Güell (Pompeu Fabra universitetas, Ispanija, gamtos mokslai, biochemija – N 004);
Dr. Stephen Jones (Vilniaus University, gamtos mokslai, biochemija – N 004);
Prof. dr. Rolandas Meškys (Vilniaus University, gamtos mokslai, biochemija – N 004);
Dr. Rapolas Žilionis (Vilniaus University, gamtos mokslai, biochemija – N 004).

Disertacija ginama viešame Gynimo tarybos posėdyje 2023 m. rugsėjo mėn. 27 d. 14 val. Vilniaus universiteto Gyvybės mokslų centro R-401 auditorijoje ir nuotoliniu būdu. Adresas: Saulėtekio al. 7, R 401 auditorija, Vilnius, Lietuva, tel. +37062281505 ; el. paštas tomas.urbaitis@caszyme.com

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

CONTENTS

CONTENTS 5
ABBREVIATIONS 8
INTRODUCTION 10
1. LITERATURE OVERVIEW 13
1.1. CRISPR-Cas 14
1.2. Classification 17
1.3. Class 2 CRISPR-Cas systems 19
1.4. Type II CRISPR-Cas systems 21
1.5. Type V CRISPR-Cas systems 26
1.5.1. Structural and organizational diversity of type V CRISPR-Cas systems 26
1.5.2. Functional diversity of type V CRISPR-Cas systems 30
1.5.3. dsDNA cleavage model of Cas12 enzymes 32
1.6. CRISPR-Cas-based nucleic acid detection 35
1.7. CRISPR-Cas-based genome engineering 35
1.7.1. Genome editing with CRISPR-Cas nucleases 36
1.7.2. Base editors 37
1.7.3. Prime editors 38
1.8. Limitations 39
2. MATERIALS AND METHODS 43
2.1. Materials 43
2.1.1. Chemicals 43
2.1.2. Commercial proteins, dyes and kits 43
2.1.3. Bacterial strains 43
2.1.4. Buffers 44
2.1.5. Proteins and nucleic acids 45
2.2. Methods 45
2.2.1. Identification and phylogeny of Cas9 orthologs 45

2.2.2. Engineering Cas9 single guide RNA solutions 46
2.2.3. Computational analysis of Cas9 tracrRNAs..... 47
2.2.4. Production of single guide RNAs 47
2.2.5. PAM library cleavage using *in vitro* translated Cas9s 48
2.2.6. Capture and sequencing of cleaved library fragments 48
2.2.7. Identification of PAM preferences..... 49
2.2.8. Cas9 expression and purification 49
2.2.9. Computational analysis of Cas9 PAM interacting domains..... 50
2.2.10. Evaluation of protospacer cleavage patterns 50
2.2.11. *In vitro* Cas9 cleavage assays for determining optimal buffer, temperature and spacer length..... 51
2.2.12. Cas9 protein thermal stability..... 52
2.2.13. Identification of CRISPR-Cas12l..... 52
2.2.14. Engineering CRISPR-Cas12l systems to target a PAM library ... 54
2.2.15. Detecting Cas12l dsDNA cleavage and PAM recognition..... 54
2.2.16. Expression and purification of Cas12l proteins..... 55
2.2.17. Cas12l-sgRNA complex assembly for *in vitro* cleavage..... 55
2.2.18. Cas12l DNA substrate generation 56
2.2.19. Cas12l DNA substrate cleavage assays..... 56
2.2.20. M13 ssDNA cleavage assays 57
2.2.21. Fluorophore quencher-labelled reporter assays..... 57
2.2.22. Plasmid interference assay 59
3. RESULTS AND DISCUSSION..... 60
3.1. Characterization of type II CRISPR-Cas9 orthologs 60
3.1.1. Selection of Cas9 orthologs 60
3.1.2. Cas9 guide RNA determination and analysis..... 61
3.1.3. Cas9 ortholog PAM requirements..... 62
3.1.4. Cas9 PAM interacting domain phylogeny 65
3.1.5. Cas9 biochemical activity 66
3.2. Characterization of type V CRISPR-Cas nucleases..... 70

3.2.1. Identification of CRISPR-Cas12l systems 70
3.2.2. Cas12l PAM determination 72
3.2.3. Cas12l guide RNA identity 73
3.2.4. Biochemical characterization of Cas12l dsDNA cleavage activity 74
3.2.5. Cas12l collateral nucleic acid cleavage 78
3.2.6. Asp2Cas12l activity in *E. coli* 81
3.3. Final remarks 82
3.4. Current progress in the field 84
CONCLUSIONS 86
LIST OF PUBLICATIONS 87
CONFERENCE PRESENTATIONS 88
REFERENCES 89
APPENDICES 116
SANTRAUKA 122
SANTRUMPOS 122
ĮVADAS 124
METODAI 127
REZULTATAI IR JŲ APTARIMAS 141
IŠVADOS 166
ACKNOWLEDGEMENTS 167
CURRICULUM VITAE 168

ABBREVIATIONS

Asp	*Armatimonadetes* species
AAV	adeno-associated virus
BH	bridge helix
bp	base pair
Cas	CRISPR-associated
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
cryo-EM	cryogenic electron microscopy
CTD	C-terminal domain
DNase	deoxyribonuclease
ds	double-stranded
DSB	double-stranded break
DTT	dithiothreitol
FLL	full length linear
gRNA	guide RNA
HDR	homology-directed recombination
His	histidine
HMM	hidden Markov models
IPTG	isopropyl-β-D-1-thiogalactopyranoside
IVT	in vitro translation
LB	Lysogeny broth/Luria-Bertani
MBP	maltose binding protein
NHEJ	non-homologous end-joining
nt	nucleotide
NTS	non-target strand
NUC	nuclease lobe
OC	open circle
PAGE	polyacrilamide gel electrophoresis
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PFM	position frequency matrix
PFS	protospacer flanking site
PI	PAM-interacting
PMSF	phenylmethylsulfonyl fluoride
pre-crRNA	precursor crRNA
REC	recognition lobe
RNase	ribonuclease
RNP	ribonucleoprotein

SC	supercoiled
SDS	sodium dodecyl sulfate
sgRNA	single-guide RNA
ss	single-stranded
SUMO	small ubiquitin-like modifier
tracrRNA	trans-activating CRISPR RNA
Tris	tris(hydroxymethyl)aminomethane
TS	target strand

INTRODUCTION

Prokaryotic CRISPR (clustered regularly interspaced palindromic repeats)-Cas (CRISPR-associated) systems provide adaptive immunity to bacteria and archaea against foreign nucleic acids in the form of viruses and mobile genome elements (Barrangou et al., 2007). This defence is conferred by capturing fragments of invading DNA and integrating them into a CRISPR array and subsequently transcribing the array in order to produce crRNAs that guide Cas proteins to cleave complementary nucleic acids during repeated infections (Barrangou et al., 2007; Mojica et al., 2005). By changing the guide RNA sequences the Cas nucleases can theoretically be directed to target any nucleic acid sequence, a feature which was utilized to adapt these enzymes for genome editing (Cong et al., 2013; Gasiunas et al., 2012; Jinek et al., 2012). However, the actual targeting space of Cas nucleases is limited by the requirement of the recognition of a short nucleic acid sequence near the target, termed the PAM (protospacer adjacent motif), which is inherent to each individual Cas nuclease (Mojica et al., 2009; Shah et al., 2013). Furthermore, the large size of most commonly used genome editors CRISPR-Cas9 and Cas12a make the delivery of these systems into cells a challenge, particularly using the commonplace adeno-associated virus delivery method (Wilbie et al., 2019). Considering that CRISPR-Cas are abundant in prokaryotes with ~85% of archaea and ~40% of bacteria harboring these systems, this natural diversity could be harnessed in search of novel CRISPR-Cas nucleases with diverse structural and biochemical features (Makarova et al., 2020).

The major subject of this PhD thesis – novel type II and type V CRISPR-Cas nucleases. **The goal** was to identify and characterize new enzymes exhibiting biochemical and structural features beneficial for gene editing purposes, such as recognition of novel PAM sequences, robust cleavage activity and reduced size. To achieve this goal, the following **objectives** were set:

1. To identify and test new type II CRISPR-Cas nucleases for dsDNA cleavage activity and determine their guide RNA and PAM requirements.
2. Reconstitute and characterize the type II Cas nuclease complexes *in vitro.*
3. To identify and test new type V CRISPR-Cas nucleases for dsDNA cleavage activity and determine their guide RNA and PAM requirements.
4. Reconstitute and characterize the type V Cas nucleases *in vitro.*

Scientific novelty and practical value:
This study is focused on the identification and characterization of novel type II and type V CRISPR-Cas nucleases.

We set out to evaluate biochemically the largest set of putative Cas9 nucleases to date and identified functional Cas9 proteins with novel G-, C-, A-, and T-rich PAM recognition of varying compositions, in principle expanding the sequence space targetable by Cas9. In one particular instance, Mga Cas9, identified in this study, was applied for single nucleotide polymorphism (SNP) detection due to its PAM sequence overlapping with the relevant SNP (Balderston et al., 2021). This facilitated more sensitive discrimination than when the SNP was located in the protospacer region, illustrating the advantage provided by diverse PAM recognition. While the majority of orthologs were observed to recognize PAMs longer than the 2 bp, like that of the canonical *Streptococcus pyogenes* (Spy) Cas9, this may be beneficial for genome editing applications as orthologs with longer PAM recognition (≥3 bp) may afford higher specificity (C. M. Lee et al., 2016; Müller et al., 2016). Further biochemical evaluation of DNA cleavage activity of the Cas9 nucleases described here exposed additional differences among orthologs. These included a wide range of temperature dependencies. Of particular interest was Cme2 Cas9, which was only robustly active from ~30 °C to 55 °C suggesting the possibility of temperature-controlled DNA search and modification (F. Richter et al., 2016; Zhuo et al., 2021). Additionally, the DNA cleavage activity at different temperatures for Nsa and Ssa Cas9s suggested they could be harnessed for use in thermo- or psychrophiles, respectively. Furthermore, we characterized orthologous Cas9 nucleases with different and potentially advantageous properties compared to those generally prescribed to Spy Cas9. These included variation in the cleavage product DNA termini resulting from target cleavage, which might facilitate distinct DNA repair outcomes (Y.-W. Fu et al., 2021), as well as a preference for a longer tract of gRNA and DNA target-site homology. Lastly, some smaller Cas9 orthologs of around 1100 amino acid length, like Nsa Cas9 and Tsp Cas9 were observed to maintain robust DNA cleavage activity and thus may be favored in applications where the delivery of larger Cas9 proteins is challenging (E. Kim et al., 2017).

During the second part of this study we identified a new family of type V CRISPR-Cas nuclease. The new Cas12l enzymes are ~860 amino acid long which is significantly less than other commonly used Cas9 and Cas12a proteins and only surpassed in compactness by type V-F and type V-J systems (Bigelyte et al., 2021; Pausch et al., 2020). Moreover, the collection of Cas12l

nucleases identified here were shown to cleave dsDNA in the presence of a 5' C-rich PAM. While CRISPR-Cas9 nucleases have been determined to recognize a C-rich PAM, including here in this study as well, this feature in a Cas12 effector makes them a desirable addition to the type V family providing an important counterbalance to the 5' T-rich PAM recognition typified by most other Cas12 enzymes. Additionally, Cas12l are shown to degrade ssDNA and ssRNA upon target recognition, a feature that was adapted for nucleic acid detection systems using Cas12 proteins (Gootenberg et al., 2018; J. Joung et al., 2020). With Cas12l exhibiting increased activity at elevated reaction temperatures around 50°C, they might be beneficial in isothermal applications combining nucleic acid amplification and detection (J. Joung et al., 2020). Finally, Asp2Cas12l mediated interference against plasmid DNA in a heterologous *E. coli* host. This suggests that this new type of effector may be harnessed for editing in other cell types as well.

The major findings presented for defense in this thesis:

1. 79 phylogenetically distinct Cas9 orthologs exhibit dsDNA cleavage activity, facilitated by diverse PAM and guide RNA requirements. .
2. Cas9 orthologs exhibit varied biochemical properties, including temperature dependencies, spacer length preferences and dsDNA cleavage patterns.
3. A new family of type V nucleases, Cas12l, are capable of dsDNA cleavage in a C-rich PAM and dual guide RNA dependent manner.
4. Cas12l exhibit collateral ssDNA and ssRNA cleavage activity upon target recognition.
5. Asp2Cas12l mediates interference against invading plasmid DNA in *E. coli.*

1. LITERATURE OVERVIEW

Ever since the establishment of the central dogma of molecular biology scientists have striven to develop technologies that would facilitate the analysis, manipulation and modification of the genome, which is crucial not only for fundamental understanding of biological processes, but also for therapeutic applications. Advancements in genome sequencing technology have enabled the rapid development of genetic disease diagnostics. More than 6000 genetic disease phenotypes are caused by mutations in a single gene (Condò, 2022). Treatment via precise DNA sequence alterations to disrupt a faulty gene or restore normal function is one of the leading goals of modern personalized medicine. The most commonly applied method of gene editing is the recruitment of intracellular DNA repair mechanisms to a DNA cleavage site (Porteus, 2014). However, the means of introducing such a site at a specific DNA sequence is the main challenge of genome editing and the area where most of the innovations in the field have been developed.

In general, the precise editing of genetic information requires a molecular tool comprised of two parts: a DNA (or RNA) binding module that is responsible for target recognition and localization, and an effector module that mediates DNA (or RNA) cleavage. An early form of such tools were meganucleases, or homing endonucleases, that recognize DNA sequences which are sufficiently long so as not to occur at the genome more than once stochastically (Paques & Duchateau, 2007). While it was demonstrated that the I-SceI homing endonuclease could be used to modify human somatic cells via homology-directed recombination, it initially could only target its native recognition sequence and redirecting it to new genomic sites was challenging and time-consuming (Choulika et al., 1995; Thyme et al., 2014). This was improved upon with the development of chimeric proteins harboring a Zinc-finger or transcription activator-like effector DNA binding domains fused to a nuclease domain of restriction enzyme FokI (J. K. Joung & Sander, 2012; Urnov et al., 2010). These systems rely on specific protein – DNA interactions for target recognition and can be redirected to novel DNA sequences by the mutagenesis of the DNA binding domains. However, protein engineering is a labor- and time-intensive process. Meanwhile, a nuclease from an adaptive bacterial immunity system CRISPR-Cas was shown to recognize a DNA target through Watson-Crick base pairing by using a RNA guide (Gasiunas et al., 2012; Jinek et al., 2012). Cas nucleases became the perfect candidates for sequence-specific genome editing tools.

Various Cas enzymes have since been adapted for genome editing in a wide range of cells and organisms (Bigelyte et al., 2021; Endo et al., 2016; W. Jiang

et al., 2013; Jinek et al., 2013; H. Kim et al., 2016; Tsuchida et al., 2022; Zetsche, Gootenberg, et al., 2015). Furthermore, their applicability is not limited to the generation of a double-stranded DNA break to induce DNA repair. Nuclease deficient Cas enzymes fused to various effectors like transcriptional activators, repressors, epigenetic modifiers, deaminases, reverse transcriptases have been used to regulate gene expression, enable specific nucleobase editing and priming insertions and deletions based on a RNA template, respectively (Anzalone et al., 2019; Gilbert et al., 2014; Hilton et al., 2015; Komor et al., 2016). Nevertheless, these tools are not without their limitations. This literature overview thus aims to present the research performed to facilitate the understanding of CRISPR-Cas systems and develop them as genome editing reagents.

1.1. CRISPR-Cas

The most abundant biological entities in nature are viruses, outnumbering cells by a factor of 10 in most environments (Aziz et al., 2015). Viruses and mobile genomic elements have been found in association with virtually every studied cellular organism (Koonin & Dolja, 2013). Therefore, driven by a constant host – parasite coevolution, cellular organisms have developed diverse defense systems protecting against invasive genomic information (Koonin et al., 2016). One of these defense mechanisms is the CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats - CRISPR-associated) adaptive immunity system found in bacteria and archaea (Barrangou et al., 2007). Since its discovery it has become one of the flagship fields in life sciences.

In general, the CRISPR-Cas system is a genetic locus comprised of a CRISPR array and adjacent CRISPR-associated proteins (Figure 1.1). The CRISPR array stores information on past phage infection events in the form of short sequences, termed spacers, found integrated between identical repeating palindromic sequences (Barrangou et al., 2007; Mojica et al., 2005). The CRISPR associated, or Cas, proteins exhibit various enzymatic activities mediating the adaptive immunity, the main of which is the ability to use the spacer sequences transcribed from the CRISPR array as guide RNAs for the site-specific degradation of invading nucleic acids (Gasiunas et al., 2012).

The CRISPR-Cas mediated adaptive immunity process can be divided into three parts: adaptation, expression, and interference (Figure 1.1). In the first phase, adaptation, a foreign DNA fragment (prespacer) is captured by the Cas1-Cas2 integrase complex (and in some cases, auxiliary Cas proteins) and integrated into the CRISPR array as a new spacer. Cas1 and Cas2 are highly

conserved across all CRISPR systems and assemble to form a heterohexamer $Cas1_4$-$Cas2_2$ (Nuñez et al., 2015). The Cas1-Cas2 complex splays the ends of the bound DNA to create two 3' -OH ends needed for the nucleophilic attack and integration into the CRISPR array (Nuñez et al., 2015; Y. Xiao et al., 2017). The integration occurs near a leader site in one end of the array thus conferring a sort of temporal dynamic to the immunity, since it was shown that interference against recently encountered invaders is enhanced (S. A. Jackson et al., 2017). During the second phase of the CRISPR immunity process, expression, guide RNAs for Cas effector proteins are produced. First, a long precursor CRISPR RNA (pre-crRNA) is transcribed from the CRISPR array. It is subsequently processed into mature CRISPR RNAs (crRNA) either by dedicated Cas proteins, CRISPR interference effectors themselves, or by cellular RNAses, depending on the type of the CRISPR system (Charpentier et al., 2015; Fonfara et al., 2016; Niewoehner et al., 2014). Mature crRNAs contain a single spacer sequence, which acts as a guide for Cas effectors, that is flanked on one side by a fragment of a CRISPR array repeat sequence (Brouns et al., 2008; Gasiunas et al., 2012). Some CRISPR systems employ an additional RNA molecule, called the trans-activating CRISPR RNA (tracrRNA) for crRNA biogenesis (and, subsequently, interference) (Deltcheva et al., 2011). In the last step - interference - active ribonucleoprotein complexes of mature crRNA and Cas effector proteins (and sometimes, tracrRNA) are formed. The crRNAs then guide the complexes to cleave complementary DNA, RNA, or both, depending on the CRISPR-Cas system (Abudayyeh et al., 2017; Gasiunas et al., 2012; Özcan et al., 2021; Tamulaitis et al., 2017; Yan et al., 2019). Additionally, in some systems target binding by the Cas-crRNA ribonucleoprotein complex initiates a *trans*-acting collateral non-specific nuclease cleavage activity (J. S. Chen, Ma, et al., 2018).

Since the CRISPR-Cas interference is based on the complementarity between crRNA and the target, theoretically, active Cas ribonucleoproteins could target their own CRISPR arrays. To distinguish between 'self' and 'non-self' and avoid autoimmunity, the target recognition by Cas effectors relies on a protospacer adjacent motif (PAM) – a short (generally 2-5 bp) sequence adjacent to the target (Figure 1.1) (Semenova et al., 2011; Westra et al., 2013). Cas proteins responsible for target binding can scan long DNA stretches for PAM sequences and, upon binding to them, unwind the adjacent dsDNA helix to allow for hybridization with the crRNA (Jones et al., 2017; Sternberg et al., 2014; Swarts & Jinek, 2019; Szczelkun et al., 2014). PAM sequences are also recognized during the acquisition of new spacers. Proteins active in the integration process, Cas1 and Cas2 (in some cases in combination with the interference complex) recognize PAM sequences to ensure that the newly

incorporated spacer can subsequently be used to direct the interference complex to the target DNA (Heler et al., 2015; C. Richter et al., 2014; Vorontsova et al., 2015). Various PAM sequences and arrangements are recognized by the different types of CRISPR-Cas systems and will be discussed more in depth in later sections.

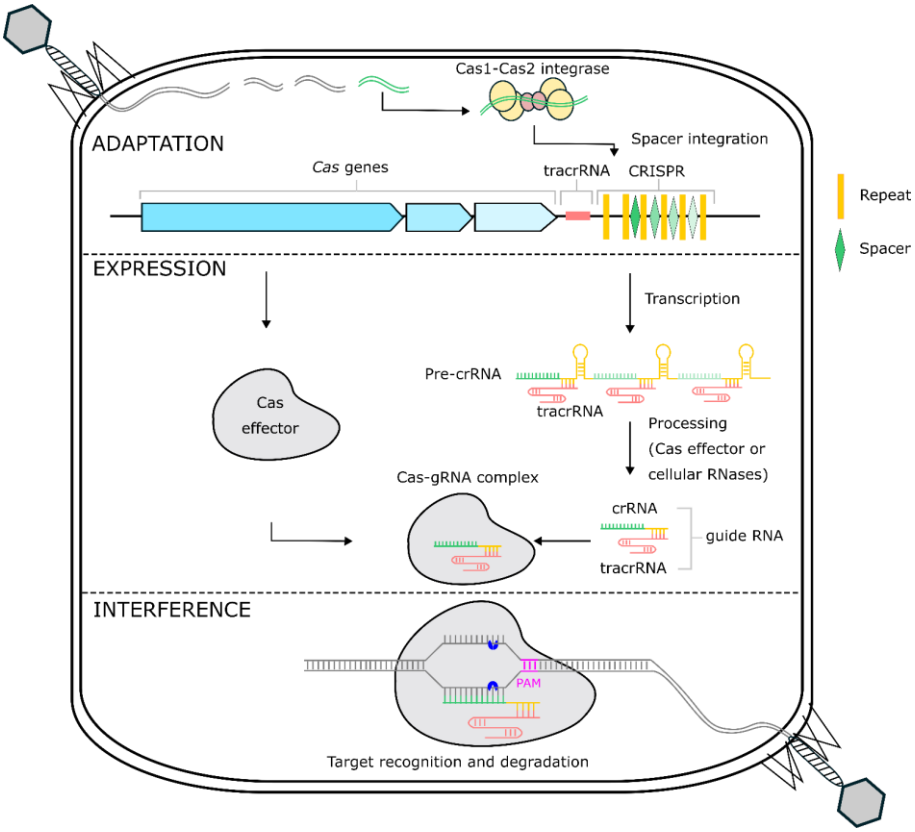


Figure 1.1. CRISPR-Cas adaptive immunity. The process of CRISPR-Cas response to invasive nucleic acids is separated into three stages. In the first stage, adaptation, fragments of foreign genetic material are captured by the Cas1-Cas2 integrase complex and integrated into the CRISPR array as spacers. In the second phase, expression, Cas nuclease effectors and auxiliary Cas proteins are expressed from the *cas* genes and the CRISPR array is transcribed as a long pre-crRNA transcript, along with the tracrRNA, if present in a particular CRISPR-Cas system. The pre-crRNA transcript is further processed either by the Cas effector or endogenous RNases to generate mature crRNAs. The crRNAs, tracrRNA and Cas effectors assemble to form Cas-guide RNA complexes. In the last stage, interference, the Cas-gRNA complex binds and cleaves invading nucleic acids at target sites complementary to the spacer sequence derived from the transcribed CRISPR array and adjacent a PAM sequence. tracrRNA (red) – trans activating CRISPR RNA, pre-crRNA – precursor crRNA, crRNA – CRISPR RNA, gRNA – guide RNA, PAM (pink) – protospacer adjacent motif. CRISPR repeats and fragments derived from them are shown in yellow, spacers and fragments derived from them – in green.

1.2. Classification

CRISPR-Cas systems show a remarkable diversity of Cas protein sequences, gene compositions and architecture of genomic loci. Currently, CRISPR-Cas systems are divided into 2 classes, 6 types and at least 33 subtypes (Makarova et al., 2020). The separation into classes is based on the nature of the effector module responsible for the interference phase – class 1 CRISPR-Cas systems employ multiple Cas proteins to assemble a functional crRNA-binding complex capable of processing the target, whereas in class 2 systems only a single multidomain effector nuclease is needed (Figure 1.2A) (Shmakov et al., 2017). Further subdivision into types and subtypes is based on the identities of the functional modules of CRISPR systems (Figure 1.2B). These modules are divided into four categories related to their role in the three phases of CRISPR-Cas immunity process (Makarova et al., 2020).

The adaptation module includes genes encoding enzymes involved in spacer integration. Cas1 and Cas2 genes are conserved across all systems, whereas additional genes encoding for auxiliary proteins Cas4 and/or Csn2 are found in types I, II and IV, and reverse transcriptase genes are found in RNA-targeting type III and VI systems (Ka et al., 2018; Kieper et al., 2018; Shiimori et al., 2018; Toro et al., 2017, 2019).

The second – expression module is responsible for pre-crRNA processing. Most class I systems employ Cas6 for processing (Niewoehner et al., 2014). In type II systems this is carried out by bacterial housekeeping RNaseIII while in type V and VI systems pre-crRNA processing is handled by the main effector nuclease itself (Charpentier et al., 2015; East-Seletsky et al., 2016; Fonfara et al., 2016).

The interference module encompasses enzymes responsible for guide RNA and target binding and nucleic acid cleavage. As mentioned before, in class 1 CRISPR-Cas systems (encompassing types I, III and IV) it is carried out by large complexes made up of multiple Cas proteins in different combinations, depending on type and subtype. Across all class 1 types the backbone of the effector complex is formed by Cas5 and multiple copies of Cas7, which facilitate crRNA and target binding (T. Y. Liu & Doudna, 2020). In type I and IV systems Cas6 is also physically associated to the complex through the 3' hairpin on the crRNA (Chowdhury et al., 2017; Özcan et al., 2019). Type I and III complexes also contain several copies of Cas11, referred to as small subunit (R. N. Jackson & Wiedenheft, 2015; Venclovas, 2016). Proteins comprising the rest of the effector complexes are more diverse across different

class 1 types and define to which type the system is attributed to. Large subunits Cas8, Cas10 and Csf1 (Type I, III, IV systems, respectively) occupy a similar position in the crystal structures of effector complexes but are divergent by sequence and structure (T. Y. Liu & Doudna, 2020; Makarova et al., 2020). Also, in type III systems Cas10 (also called Csm1) is fused to HD nuclease domain with ssDNAse activity, whereas type I systems recruit additional Cas3 nucleases to mediate target cleavage (He et al., 2020; Kazlauskiene et al., 2016). In some cases, for example in subtype III-E systems, the role of Cas10 is instead fulfilled by a Cas7-11 fusion protein (Özcan et al., 2021). In the case of class 2 CRISPR systems the interference process is carried out by single large multidomain proteins with distinct nuclease domains differentiating their types - Cas9 (type II), Cas12 (type V) or Cas13 (type VI).

The fourth module is the ancillary feature module, with CRISPR-linked genes thought to have roles in CRISPR-Cas mediated immunity. One of these processes that has been characterized is a cyclic oligoadenylate (cOA) signal transduction pathway. A unique feature of type III Cas10 is the synthesis of cOAs via two polymerase-like Palm domains found in the enzyme (Kazlauskiene et al., 2017). These messenger molecules act in a signal transduction pathway whereupon they are bound by the CARF (CRISPR associated Rossmann fold) domain in Csm6 which then initiates non-specific RNA degradation by its HEPN (higher eukaryotes and prokaryotes nucleotide-binding) domain (Kazlauskiene et al., 2017; Niewoehner et al., 2017).

To briefly summarize, CRISPR-Cas systems exhibit diverse locus and effector architectures that define their mechanisms of function and, subsequently, classification into the current CRISPR-Cas nomenclature.

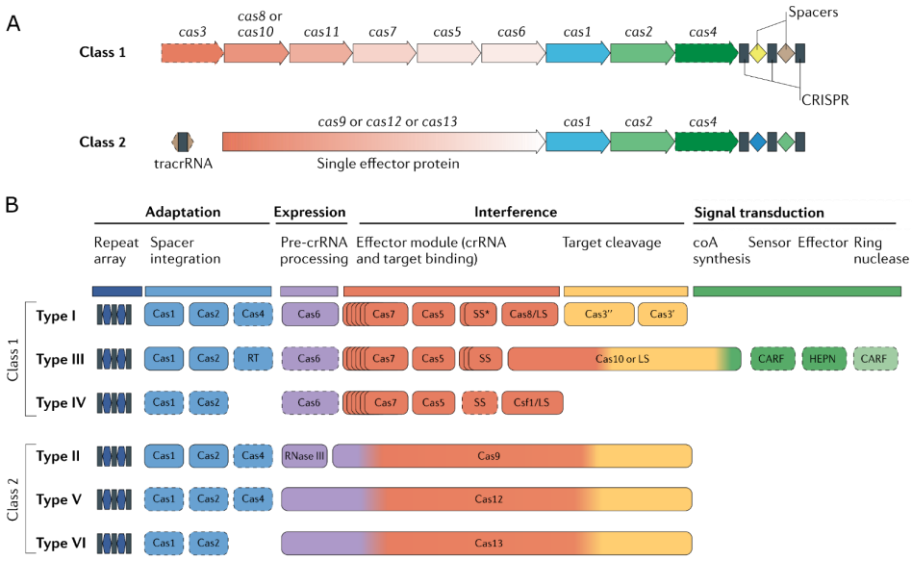


Figure 1.2. Classification and modular organization of CRISPR-Cas systems. (A) Schematic illustration of the organization of class 1 and class 2 CRISPR-Cas systems. This classification is based on the effector module complexity (genes marked in shades of red). In class 1 systems the effector is composed of multiple Cas proteins each mediating distinct processes such as pre-crRNA maturation, crRNA and target binding and cleavage, whereas class 2 systems encode a single multidomain effector protein facilitating the aforementioned activities. **(B)** Further subdivision of class 1 and class 2 CRISPR-Cas systems into 6 types based on the identity of their functional modules. Each module is involved in a particular stage of the CRISPR adaptive immunity process and the proteins responsible for the respective activites among the different types are color coded accordingly. Multidomain proteins Cas10, Cas9, Cas12 and Cas13 mediate multiple activites related to separate functional modules therefore they are marked in several colors. Dashed outlines indicate dispensable and/or missing, in some subtypes, components. SS –small subunit, LS – large subunit, RT – reverse transcriptase. Adapted from (Makarova et al., 2020)

1.3. Class 2 CRISPR-Cas systems

While class 1 CRISPR-Cas systems are significantly more abundant in nature (Makarova et al., 2020), the relative simplicity of class 2 effector modules being single proteins made them the most extensively studied CRISPR-Cas systems.

The class 2 effectors (Cas9, Cas12 and Cas13) fundamentally differ in their domain architecture and their interference mechanism (Figure 1.3). Type II Cas9 effectors contain two nuclease domains (HNH and RuvC) each responsible for the cleavage of a single target DNA strand (Chylinski et al., 2014). Cas9 recognizes targets using a dual RNA guide comprised of crRNA

and tracrRNA and introduces a blunt double-stranded break(Jinek et al., 2012). On the other hand, type V Cas12 effectors employ only a single RuvC nuclease domain which cleaves each DNA strand sequentially, leaving staggered DNA ends (Swarts & Jinek, 2019; Zetsche, Gootenberg, et al., 2015). Cas12a-mediated target cleavage requires only a single crRNA guide, but some effectors of other type V subtypes utilize a tracrRNA as well (B. Tong et al., 2021). Type VI CRISPR-Cas systems are distinct from other members of class 2 in the sense that their effector, Cas13, exclusively cleaves RNA (Abudayyeh et al., 2016). Cas13 enzymes contain two HEPN domains which mediate their RNase activity.

Both Cas9 and Cas12 enzymes require PAM sequences adjacent to the target for binding and cleavage, however Cas9 recognizes a PAM at the 3' end of the protospacer, while in the case of Cas12 the PAM is on the 5' end (Figure 1.3). Cas13 systems, on the other hand, seem to follow less restrictive rules regarding target recognition. Some of Type VI effectors require distinct sequences at the 3' end of the RNA target, which were named Protospacer Flanking Site (PFS), while others do not show any bias towards particular nucleotides (Abudayyeh et al., 2017; Smargon et al., 2017; Yan et al., 2018).

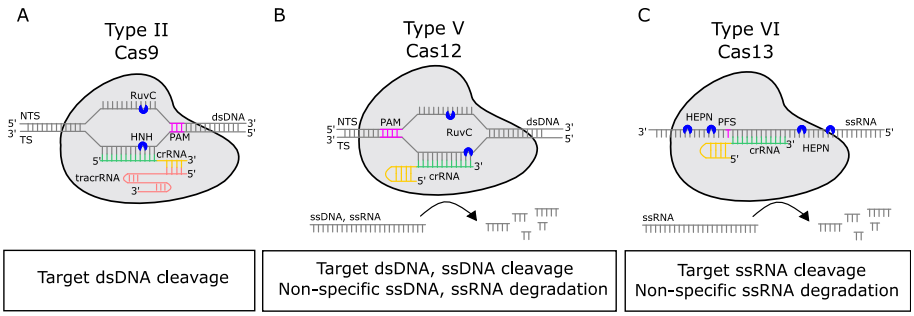


Figure 1.3. Schematic illustration of the class 2 CRISPR-Cas effector complexes. Type II nuclease effector Cas9 forms a complex with crRNA (yellow and green) and tracrRNA (red) to bind dsDNA at a site complementary to the crRNA spacer sequence (green), upstream of a PAM sequence (pink). DNA strands are then cleaved by RuvC and HNH nuclease domains (blue), leaving blunt DNA ends. Type V effector Cas12 forms a complex with crRNA to bind dsDNA at a site complementary to the crRNA spacer sequence, downstream of a PAM sequence. DNA strands are then cleaved by a single RuvC nuclease domain (blue), leaving staggered DNA ends. Subsequently, Cas12 carries out non-specific degradation of ssDNA and ssRNA molecules. Type VI effector Cas13 forms a complex with crRNA to bind ssRNA at a site complementary to the crRNA spacer sequence adjacent to a PFS sequence (pink). RNA is then cleaved by the dual HEPN nuclease domains (blue). Subsequently, Cas13 carries out non-specific degradation of ssRNA molecules. NTS – non-target strand, TS – target strand.

It was shown that in type VI systems self-interference is prevented by the inhibition of RNA cleavage when there is extended complementarity between the crRNA repeat sequence and the RNA substrate (Meeske & Marraffini, 2018).

Furthermore, unlike those of type II, some type V and all VI effectors process their own pre-crRNA into mature crRNA (East-Seletsky et al., 2016; Pausch et al., 2020; Yan et al., 2019; Zetsche, Gootenberg, et al., 2015).

Type V and VI CRISPR-Cas effectors exhibit non-specific cleavage activity of single stranded DNA and/or RNA upon binding the target substrate (Figure 1.3) (Bigelyte et al., 2021; J. S. Chen, Ma, et al., 2018; Fuchs et al., 2022; Yan et al., 2019).

1.4. Type II CRISPR-Cas systems

As stated previously, type II encompasses CRISPR-Cas systems containing the single nuclease effector Cas9. These systems are subdivided into 3 subtypes – II-A, II-B, and II-C, depending on the degree of homology between the Cas9 proteins, as well as the composition of the adaptation module and the organization of the CRISPR-Cas locus (Makarova et al., 2015). Subtypes II-A, -B, and -C comprise ~55%, ~3% and ~41% of type II systems identified in public sequence databases for, respectively (Shmakov et al., 2017).

Cas9 is large (~1000-1300 aa) multidomain DNA endonuclease. In its apo state, the structure of Cas9 can be separated into two lobes – the alpha-helical recognition (REC) lobe and the nuclease (NUC) lobe (Figure 1.4A). The NUC lobe contains an arginine-rich bridge helix (BH) domain, as well as the HNH and the split RuvC nuclease domains and the variable C-terminal domain (CTD) (Jinek et al., 2014). The CTD contains PAM-interacting sites required for PAM recognition and is sometimes referred to as the PAM-interacting (PI) domain (Nishimasu et al., 2014). The HNH nuclease domain is structurally similar to HNH endonucleases defined by a ββα-metal fold and a single divalent metal ion coordinated by three amino acid residues (Figure 1.4A) (Biertümpfel et al., 2007; Zuo & Liu, 2017). The RuvC nuclease domain exhibits an RNaseH-like fold where four catalytic residues coordinate two metal ions (Figure 1.4A) (Ariyoshi et al., 1994; Nishimasu et al., 2014).

Cas9 in its apo-state binds to DNA non-specifically and is easily detached in the presence of competitor RNA (Sternberg et al., 2014). It also undergoes significant conformational changes, particularly in the REC lobe, upon binding of guide RNA, therefore corroborating the experimental data that

Cas9 proteins are inactive nucleases in the absence of bound guide RNAs (Figure 1.4B-C) (F. Jiang et al., 2015; Jinek et al., 2012, 2014). A so-called seed region in the spacer is preordered in the Cas9-sgRNA complex and adopts a nearly A-form helical conformation, which is thermodynamically favorable for eventual guide:target duplex formation (Figure 1.4B) (F. Jiang et al., 2015). The Cas9-sgRNA complex scans DNA for a suitable PAM motif and upon its recognition the R-loop forms between the guide region of crRNA and the target DNA (F. Jiang et al., 2016; Sternberg et al., 2014). R-loop formation drives conformational rearrangements of the REC lobe and, subsequently, the HNH nuclease domain, bringing it near the scissile phosphate in the target DNA, which in turn activates the RuvC domain (Figure 1.4B-C) (Palermo et al., 2018; Raper et al., 2018). It is generally accepted that Cas9 generates a blunt double-strand DNA break, however some studies suggest that Cas9 possesses post-cleavage trimming activity mediated by its RuvC nuclease domain and thus can generate a heterogenous mixture of products (Suo, 2019; Zuo & Liu, 2016).

All type II CRISPR-Cas systems employ a tracrRNA for pre-crRNA processing and interference. The pre-crRNAs form a duplex with the partially complementary tracrRNA anti-repeat sequence and the dual RNAs are then processed by the nuclease activity of Cas9 and RNaseIII (Charpentier et al., 2015; Deltcheva et al., 2011). Most of the tracrRNAs of type II systems share conserved secondary structure motifs: lower and upper stem portions, which are often separated by a bulge, in the tracrRNA anti-repeat module which pairs with the crRNA repeat; a hairpin structure called the nexus in the immediate vicinity of the crRNA:tracrRNA duplex; and several hairpin structures at the 3’ end of the tracrRNA (Figure 1.5) (Briner et al., 2014). This conservation of tracrRNA enables the cross-functionality of closely related Cas9 systems (Briner et al., 2014; Fonfara et al., 2014). The dual RNAs of type II CRISPR-Cas have also been engineered to function as a single RNA molecule – single guide RNA (sgRNA), by fusing the crRNA and tracrRNA and also shortening the repeat:anti-repeat duplex, thus simplifying these systems for genome editing applications (Figure 1.5A) (Jacobi et al., 2017; Jinek et al., 2012).

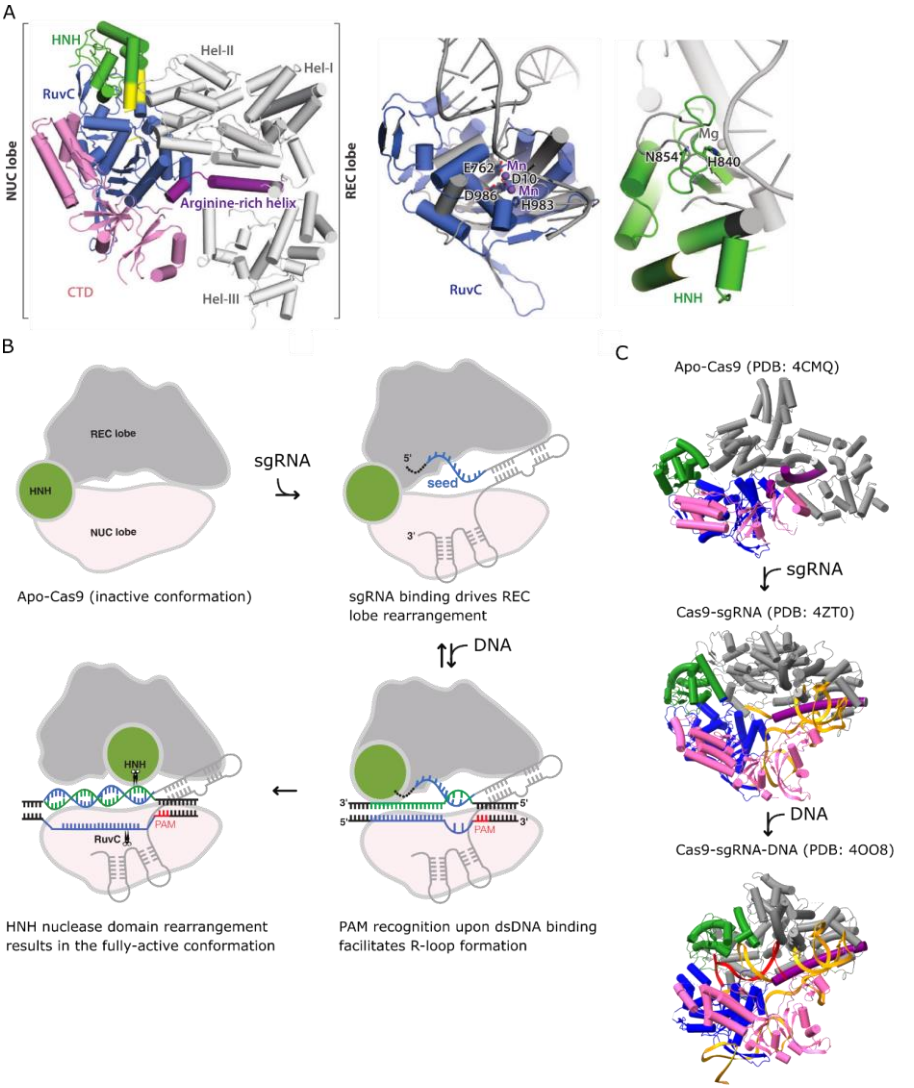


Figure 1.4. Cas9 structural and cleavage mechanism models. (A) Cartoons depicting the crystal structure of SpyCas9 (PDB ID: 4CMP) with individual domains color coded and close up views of the RuvC (blue) and the HNH (green) nuclease domains with the catalytic amino acids and divalent metal ions labelled. The RuvC domain is superimposed with the structure of the canonical *Thermus thermophilus* RuvC resolvase with a Holliday junction (gray) and the HNH domain is superimposed with the structure of the T4 Endo VII enzyme with a Holliday junction (gray) **(B)** Illustration of the mechanism of Cas9-mediated DNA cleavage. In its apo-form Cas9 adopts a cleavage-incompatible conformation. Binding of sgRNA initiates the rearrangement of the REC lobe and the helical preordering of the seed sequence in the guide RNA spacer (blue). The Cas9-sgRNA complex scans DNA for PAM sequences (red) and upon PAM recognition begins spacer:target DNA strand hybridization and R-loop formation. This drives further Cas9 domain rearrangement with the HNH nuclease domain (green) moving near the eventual cleavage site resulting in the

cleavage-compatible complex conformation. **(C)** The crystal structures of apo, binary and ternary Cas9 complexes showing the conformational rearrangements described in (B). The domains are color coded respectively to (A) and are as follows: REC lobe (gray), HNH (green), RuvC (blue), bridge helix (purple), CTD (pink), sgRNA (orange), target DNA (red). Adapted and modified from (F. Jiang et al., 2015; F. Jiang & Doudna, 2017).

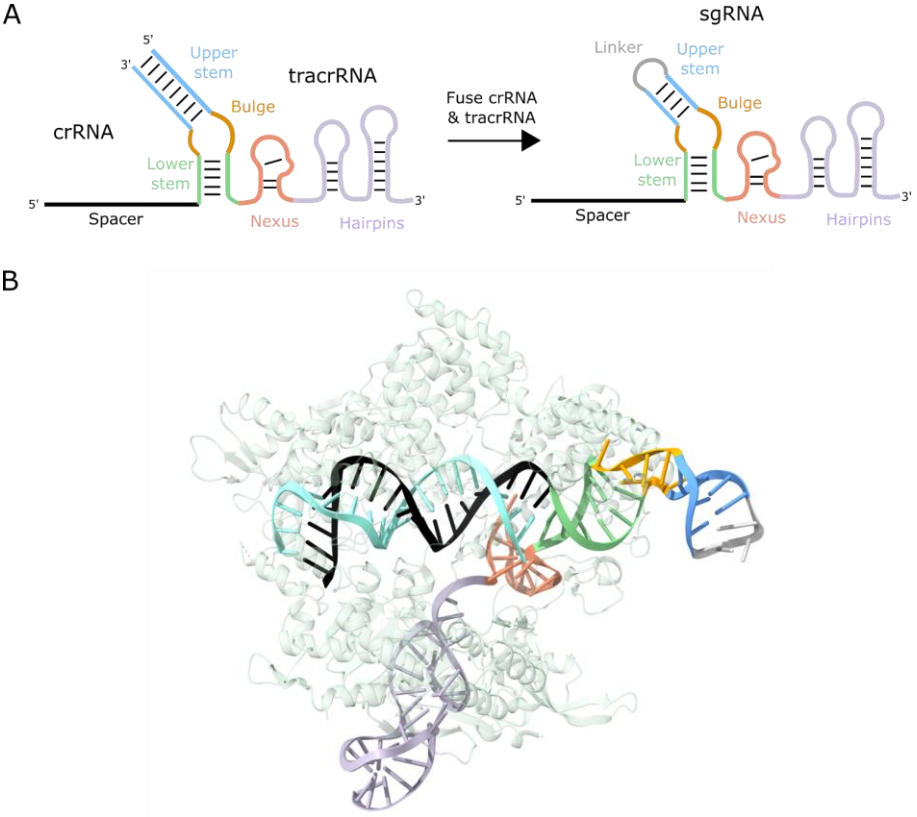


Figure 1.5. Structure of Cas9 guide RNA. (A) Illustration depicting the crRNA, tracrRNA and the chimeric sgRNA of *S. pyogenes* Cas9 and their characteristic modules: spacer sequence responsible for DNA targeting (black); lower stem (green), bulge (orange) and upper stem (blue) formed by the crRNA repeat:tracrRNA antirepeat duplex; nexus (red) and hairpin (purple) secondary structure motifs. The crRNA and tracrRNA can be fused into a sgRNA by shortening the crRNA:tracrRNA duplex and connecting the two molecules with a flexible linker (grey). Adapted from (Briner et al., 2014). **(B)** Crystal structure of *S. pyogenes* Cas9 in complex with sgRNA and target DNA strand (PDB ID: 4OO8). The sgRNA modules are color coded respectively to (A); GAAA tetraloop linking crRNA and tracrRNA colored in gray; DNA target strand colored in cyan.

Subtype II-A systems are defined by the presence of Csn2. It associates with the rest of Cas proteins from the CRISPR-Cas locus and *in vivo* deletions

of *csn2* genes inhibit integration of new spacers into the CRISPR array, suggesting its role in the adaptation process (Heler et al., 2015; Y. Wei et al., 2015; Wilkinson et al., 2019). Analysis of crystal structures and interactions between Cas1-Cas2 and Csn2 show that Csn2 is in close proximity to the prespacer binding surface of the Cas1-Cas2 complex, rather than the target DNA-binding site (Ka et al., 2016). This, together with data showing that *in vitro* Csn2 is dispensable in the integration of spacers when provided with proper substrates, suggests that Csn2 is involved with the early stages of prespacer recruitment and processing (Ka et al., 2018; Y. Xiao et al., 2017).

In subtype II-B, Csn2 is replaced by Cas4 (Makarova et al., 2020). In type I CRISPR-Cas systems the Cas4 nuclease is part of the adaptation module and is responsible for the recognition of prespacer PAM sequence, as well as the correct length and orientation of DNA fragments to be integrated into the CRISPR array (H. Lee et al., 2018; Shiimori et al., 2018). Also, Cas1 and Cas2 in subtype II-B loci phylogenetically are more closely related to those of type I systems, rather than those of subtype II-A, suggesting a similar evolutionary origin of type I and subtype II-B adaptation modules (Chylinski et al., 2014). Thus, it can be assumed that Cas4 in type II systems performs a similar role as in type I, even though concrete empirical data are lacking.

Subtype II-C systems are the simplest from the locus architecture standpoint (Shmakov et al., 2017). The absence of both Csn2 and Cas4 in II-C loci raises questions on the adaptation mechanism of these systems. It was shown that at least in one case adaptation occurred only in the presence of a phage-encoded Cas4-like protein (Hooton & Connerton, 2015). The mechanism might also function like in type I-E systems, where Cas1 and Cas2 is sufficient for adaptation, but prespacer processing is co-opted by cellular exonucleases (Yoganand et al., 2019). As of 2017, out of the ~4000 Cas9 orthologs in the NCBI database, ~1500 were type II-C (Shmakov et al., 2017). Type II-C systems appear to be overrepresented in pathogenic bacteria and the presence of these CRISPR-Cas loci can be associated increased pathogenicity (Chylinski et al., 2014; Louwen et al., 2013; Sampson et al., 2013). From a structural perspective, type II-C Cas9s stand out due their generally smaller size and more divergent PAM domain conferring the recognition of longer, more diverse PAM sequences (Fedorova, Vasileva, et al., 2020; Harrington et al., 2017; Hou et al., 2013; Karvelis et al., 2015; Yamada et al., 2017). While experimental validation is yet lacking, the tracrRNAs employed by type II-C Cas9s also are predicted to exhibit differences from their II-A and II-B counterparts, particularly in the form of additional pairing interactions and longer hairpin structures at the 3’ end (Mir et al., 2018).

1.5. Type V CRISPR-Cas systems

Type V CRISPR-Cas systems are the most diverse out of all Class 2 CRISPR-Cas systems and are divided into A-M subtypes encoding effectors of various size and interference mechanisms. The diversity stems from distinct evolutionary paths the various systems took diverging from their ancestral TnpB proteins, which resulted in Cas nucleases retaining a RuvC-like catalytic domain at the C-terminus, but developing high variability at the N-terminus. (Shmakov et al., 2017).While most of the type V effectors exhibit dsDNAse activity, some are only shown to cleave ssDNA or ssRNA, or display no nuclease activity whatsoever (Makarova et al., 2020; B. Tong et al., 2021). Key structural and functional features and differences pertaining to the respective subtypes are outlined in Table 1.1 and discussed in the following sections.

1.5.1. Structural and organizational diversity of type V CRISPR-Cas systems

While type V CRISPR-Cas systems encode effectors of various length ranging from 400 to 1500 aa, they typically exhibit similar domain architecture. However, significant differences are observed in their locus organization, identity of the guide RNA module and PAM recognition.

The first type V system to be identified and characterized contains the signature effector protein Cas12a (formerly called Cpf1) (Schunder et al., 2013; Zetsche, Gootenberg, et al., 2015). The locus in *Francisella novicida* encodes the Cas12a effector, as well as an adaptation module consisting of Cas4, Cas1 and Cas2, which are more closely related to orthologs from type I and III systems, rather than those of type II (Makarova et al., 2015; Zetsche, Gootenberg, et al., 2015). It was shown to act as a single crRNA-guided dsDNA endonuclease, recognizing a T-rich PAM sequence at the 5’ end of the target (Zetsche, Gootenberg, et al., 2015). As stated previously, the Cas12a effector itself contains a tri-split RuvC nuclease domain, but lacks the HNH domain inherent to type II Cas effectors. Nevertheless, it adopts a similar bilobed architecture, consisting of a recognition (REC) lobe, coordinating the crRNA – target DNA heteroduplex and a nuclease (NUC) lobe, containing the RuvC nuclease, bridge helix and PAM interacting domains (Dong et al., 2016; Swarts et al., 2017). Aside from the RuvC domain, Cas12a also harbors an additional catalytic moiety– a RNase motif in the WED domain responsible for the processing of pre-crRNA (Fonfara et al., 2016; Swarts et al., 2017).

Table 1.1 Features of the type V CRISPR-Cas effectors (adapted and modified from (B. Tong et al., 2021)

Subtype	Effector	Length, aa	PAM	tracrRNA	Pre-crRNA processing	*Cis*-cleavage	*Trans*-cleavage	References
V-A	Cas12a (Cpf1)	1200-1500	TTTV	-	+	dsDNA, ssDNA	ssDNA, ssRNA	(Zetsche, Gootenberg, et al., 2015)
V-B	Cas12b (C2c1)	~1300	TTN	tracrRNA	-	dsDNA, ssDNA	ssDNA	(H. Yang et al., 2016)
V-C	Cas12c (C2c3)	1200-1300	TG or TN	scoutRNA	+	-	-	(Harrington et al., 2020; Yan et al., 2019)
V-D	Cas12d (CasY)	~1200	TA or TG	scoutRNA	unconfirmed	dsDNA, ssDNA	ssDNA	(Harrington et al., 2020)
V-E	Cas12e (CasX)	~1000	TTCN	tracrRNA	-	dsDNA	ssDNA	(J.-J. Liu et al., 2019)
V-F	Cas12f (Cas14)	400-700	TTTR	tracrRNA	-	dsDNA, ssDNA	ssDNA	(Harrington et al., 2018; Karvelis et al., 2020)
V-G	Cas12g	~800	-	tracrRNA	-	ssRNA	ssDNA, ssRNA	(Yan et al., 2019)
V-H	Cas12h	~900	RTR	-	unconfirmed	dsDNA	ssDNA	(Yan et al., 2019)
V-I	Cas12i	~1100	TTN	-	+	dsDNA (nicking)	ssDNA	(Yan et al., 2019)
V-J	Cas12j (CasΦ)	~750	TBN	-	+	dsDNA, ssDNA	ssDNA	(Pausch et al., 2020)
V-K	Cas12k (V-U5)	~650	GTN	tracrRNA	-	-	-	(Strecker et al., 2019)
V-M	Cas12m (V-U1)	~600	TTN	-	+	-	-	(W. Y. Wu et al., 2022)

The next subtype of type V effectors to be identified and characterized was Cas12b (previously called C2c1) (Shmakov et al., 2015; H. Yang et al., 2016). While Cas12b employs crRNA and a tracrRNA for dsDNA cleavage like Cas9, its domain organization and architecture resembles that of Cas12a. Consequently, Cas12b also recognizes a 5‘ T-rich PAM sequence (H. Yang et al., 2016) The key difference between the V-A and V-B effectors is the lack of a clearly defined PI domain in Cas12b (L. Liu et al., 2017; H. Yang et al., 2016). The PAM recognition cleft in Cas12b is formed by helices and loops in the WED (sometimes referred to as OBD) and REC domains and is disordered in binary Cas12b-sgRNA complexes prior to substrate binding (D. Wu et al., 2017; H. Yang et al., 2016). Well defined amino acids contacts with the PAM duplex indicate a more stringent PAM requirement than that of Cas12a nucleases (D. Wu et al., 2017)

Subtype V-C & V-D effectors Cas12c (formerly C2c3) and Cas12d (formerly CasY) are also endonucleases guided by dual RNA molecules (Harrington et al., 2020; Shmakov et al., 2015; Yan et al., 2019). They are similar in size to Cas12a variants and, likewise, contain a single RuvC nuclease domain at the C-terminal region (Harrington et al., 2020; Kurihara et al., 2022). The PI domains of Cas12c effectors are significantly diverged from other type V nucleases and thus Cas12c orthologs recognize relatively short 5’-T(G/N)-3’ PAM sequences (Kurihara et al., 2022; Yan et al., 2019; B. Zhang et al., 2022). In contrast to the aforementioned subtypes Cas12c/d employ a novel type of noncoding RNA – a short-complementarity untranslated RNA (scoutRNA) (Harrington et al., 2020). Unlike tracrRNA, this transcript bears little complementarity to the repeat region of Cas12c/d crRNAs and there its predicted secondary structure differs from tracrRNA species (Harrington et al., 2020). This scoutRNA is also involved in a distinct dual-RNA pre-crRNA processing mechanism catalyzed by the Cas12c/d effectors without the need of cellular ribonucleases (Harrington et al., 2020). Like tracrRNA, scoutRNAs can be fused to crRNAs to yield functional sgRNAs (Huang et al., 2022). Structures of Cas12c-sgRNA complexes reveal that the sgRNA scaffolds are solvent exposed and form unique X-junctions not typical to type II dual RNA-guided effectors, but characteristic to type V effectors that employ tracrRNAs (Kurihara et al., 2022; B. Zhang et al., 2022).

Subtype V-E effector Cas12e (formerly CasX) is a dual RNA guided DNA endonuclease (Burstein et al., 2017). At the time of its discovery it was the smallest type V effector at a size of 980 amino acids. Distinct protein architecture of Cas12e facilitates unique mechanisms of substrate recognition where the DNA substrate is unwound by interactions with the N-terminal NTSB domain, and RNA-DNA hybrid duplex bending with the aid of the TSL

domain (J.-J. Liu et al., 2019). Structural data also showed that the guide RNA accounts for 26% of the mass in the Cas12e-sgRNA binary complex, a value that is significantly higher than observed for other Class II Cas effectors (~8% in Cas12a; ~20% in Cas12b; ~16% in Cas9) (J.-J. Liu et al., 2019). A trend of the guide RNA accounting for an increasing portion of the total RNP mass has since been observed with other compact Cas12 nucleases and the putative Cas nuclease evolutionary precursors IscB and TnpB (Bigelyte et al., 2021; Kato et al., 2022; Sasnauskas et al., 2023). Absence of guide RNA results in poorly resolved cryo-EM structures of Cas12e, suggesting that association with the RNA scaffold helps Cas12e adopt a defined conformation (J.-J. Liu et al., 2019).

Subtype V-F effectors Cas12f (previously called Cas14) represent the smallest Class 2 CRISPR-Cas nucleases to be discovered and characterized to date, being just 400 – 700 aa long (Harrington et al., 2018; Karvelis et al., 2020). Despite the small size, Cas12f proteins contain all the conventional domains of type V nucleases (R. Xiao, Li, et al., 2021). Uncharacteristically, Cas12f was shown to form a dimer, with two Cas12f molecules associating with one sgRNA molecule to assemble a functional RNP complex (Takeda et al., 2021; R. Xiao, Li, et al., 2021). Disrupting the dimerization interface abolished DNA cleavage activity (R. Xiao, Li, et al., 2021). The dimerization is asymmetrical, with a region of the RuvC nuclease domain of one protomer forming contacts with the sgRNA, while the equivalent region in the other molecule is solvent exposed (Takeda et al., 2021). Mutational analysis of the RuvC domains revealed that the cleavage of both of the substrate DNA strands is carried out by the RuvC domain of one of the protomers (Takeda et al., 2021).

Subtype V-G effector Cas12g exhibits some unique structural features pertaining to its distinct substrate recognition – since it specifically targets and cleaves ssRNA without any PAM or PFS recognition, no PI domain could be identified (Z. Li et al., 2021; Yan et al., 2019). Also, the spacer-derived guide RNA segment is not preordered in the Cas12g-sgRNA complex, as is common with other Cas12 enzymes, but adopts a somewhat disordered A-form structure more reminiscent of Cas13 systems (Z. Li et al., 2021).

Subtype V-H effector Cas12h is a ~900 aa long nuclease that uses a crRNA guide to target and cleave ss- and dsDNA in the presence of a 5'-RTR-3' PAM sequence (Yan et al., 2019).

Subtype V-I contains 1000-1100 aa long effectors Cas12i which were shown to target and cleave dsDNA substrates in the presence of a 5'-TTN-3' PAM, as well as catalyze the processing of their pre-crRNA (Yan et al., 2019). Structures of Cas12i reveal a lid located in the RuvC domain, which undergoes

a conformational change upon crRNA-DNA heteroduplex formation to facilitate substrate binding and cleavage (H. Zhang et al., 2020). The structural element equivalent to lid in Cas12i is conserved in Cas12a, Cas12b, Cas12e and Cas12g (Z. Li et al., 2021; J.-J. Liu et al., 2019; Stella et al., 2018; H. Yang et al., 2016). First, the lid engages the heteroduplex in a sequence-independent manner. Second, conformational changes of the lid open the RuvC catalytic center. Third, the lid contributes to substrate binding, thereby facilitating DNA cleavage (H. Zhang et al., 2020). The lid was demonstrated to be critical for RuvC activation in Cas12a, suggesting a conserved activation mechanism for Cas12 endonucleases (Stella et al., 2018).

Subtype V-J effector Cas12j (formerly CasΦ) nucleases are 700-800 aa long (Pausch et al., 2020). Despite their small size and sequency homology to other Cas12 enzymes limited to the ubiquitous RuvC domain, structural analysis revealed that Cas12j shares overall structural features with much larger type V counterparts (Pausch et al., 2021). Uncharacteristically to the trend observed with other compact (sub 1000 aa length) Cas12 effectors, Cas12j does not employ a large tracrRNA and requires only a crRNA for target recognition and cleavage (Pausch et al., 2020).

A unique locus organization can be observed in the subtype V-K system. It was initially discovered as containing a Cas12 effector (formerly referred to as C2c5 and V-U5), with a RuvC nuclease domain lacking the ubiquitous D-E-D catalytic triad, in association with a Tn7-like transposon, without additional *cas* genes (Faure et al., 2019; Peters et al., 2017). Additionally, the Cas12k effector is missing a Nuc domain and has an elongated lid motif, that does not change conformation upon target recognition (R. Xiao, Wang, et al., 2021).

Cas12k is not the sole type V effector without the strongly conserved D-E-D RuvC domain motif, as the compact subtype V-M nuclease Cas12m features variable residues equivalent to these metal-coordinating amino acids among its orthologs (W. Y. Wu et al., 2022).

1.5.2. Functional diversity of type V CRISPR-Cas systems

Diverse structural features of type V systems in turn confer varying enzymatic activities. Nevertheless, most type V CRISPR-Cas effectors are functional RNA-guided dsDNA and ssDNA nucleases exhibiting dsDNA and ssDNA cleavage activity. A feature ubiquitous to the single RuvC nuclease domain harboring type V effectors is the generation of staggered end dsDNA cleavage products (Harrington et al., 2020; Karvelis et al., 2020; J.-J. Liu et al., 2019; Zetsche, Gootenberg, et al., 2015). Moreover, Cas12 nucleases exhibit non-

specific collateral cleavage of ssDNA molecules upon binding their target dsDNA or ssDNA substrate (J. S. Chen, Ma, et al., 2018). However, among the active Cas12 variants there exists a discrepancy in the optimal conditions and efficiency of DNA cleavage. For example, initially, Cas12b was shown to mediate RNA-guided, crRNA/tracrRNA-dependent cleavage of dsDNA only at temperatures around 50 °C, precluding it from being used in mammalian cells (Shmakov et al., 2015). Other orthologs have since been discovered, which enable genome editing in human, animal and plant cells (Ming et al., 2020; Strecker et al., 2018; Teng et al., 2018). Currently characterized wild type Cas12f orthologs also exhibit peak dsDNA cleavage activity at elevated temperatures of around 45 °C, which likewise limits their efficiency in mammalian cell genome editing (Bigelyte et al., 2021; Karvelis et al., 2020; Z. Wu et al., 2021). However, protein engineering can be used to mitigate this issue, as shown with CasMINI, a Cas12f mutant (X. Xu et al., 2021).

Type V nucleases, including Cas12f, typically exhibit a different pattern of indels at their target sites in comparison to Cas9 – both Cas12a and Cas12f tend to induce more frequent and longer deletions, most likely stemming from the staggered end DNA cleavage products typical to type V systems. (Xin et al., 2022) The Cas12f nucleases are also characterized by higher specificity with few detectable off-target edits, which might arise from their generally lower editing efficiency than that of Cas9 or Cas12a (Xin et al., 2022).

The *trans* ssDNA cleavage activity is also susceptible to variation among type V effectors, particularly in the case of Cas12e, which exhibits significantly reduced collateral cleavage activity (J.-J. Liu et al., 2019).

Not all type V Cas nucleases facilitate nucleic acid cleavage and the exceptions are found among the V-C, V-G, V-K and V-M subtypes.

Typically, it is accepted that immunity provided by the single effector type V CRISPR-Cas systems relies on RNA-guided nuclease activity that targets phage or foreign molecules. However, no detectable RNA-guided cleavage activity on dsDNA, ssDNA or ssRNA targets by a subtype V-C effector orthologs Cas12c_4 and Cas12c2 could be observed (Harrington et al., 2020; Huang et al., 2022; Kurihara et al., 2022). Nevertheless, Cas12c_4 expression with gene-specific guides in *Escherichia coli* revealed silencing of target gene expression as well as restriction of phage growth, when targeting essential phage genes (Huang et al., 2022; Yan et al., 2019). These data support a model of DNase-free interference by Cas12c_4. Conversely, *cis*- and *trans*-DNA cleavage was demonstrated using an orthologous Cas12c1 effector (Z. Wang & Zhong, 2021; B. Zhang et al., 2022). What factors result in competent DNase activity in some Cas12c orthologs and the absence of which in others remains to be investigated.

Similarly, subtype V-M effector Cas12m exhibits no detectible cleavage of target nucleic acids, yet mediates interference through crRNA-guided binding of dsDNA targets, subsequently blocking either the replication or expression of the invading DNA (W. Y. Wu et al., 2022). Since almost half of currently known Cas12m systems are located on plasmids and contain plasmid-targeting spacers in their CRISPR arrays, it is hypothesized that they might be involved in inter-plasmid competition (Mohanraju et al., 2021; W. Y. Wu et al., 2022).

Subtype V-K effector Cas12k is not a functional DNA or RNA nuclease, due to a lack of the canonical RuvC active site motif (Shmakov et al., 2017). However, Cas12k loci containing CRISPR arrays are found in association with Tn7-like transposons (Peters et al., 2017). Biochemical and structural analysis have revealed that the CRISPR-Cas12k-associated Transposase, or CAST, mediates RNA-guided DNA transposition (Strecker et al., 2019; R. Xiao, Wang, et al., 2021). This suggests that the nuclease-deficient Cas protein had been co-opted by the Tn7-like transposon to facilitate site-directed transposon mobilization. Analogous associations have been discovered for type I and type IV CRISPR-Cas systems as well (Klompe et al., 2019; Rybarski et al., 2021).

As mentioned in the previous section, contrary to the other Cas12 systems, Cas12g specifically targets and cleaves ssRNA, which also enables non-specific hydrolysis of ssDNA and ssRNA *in trans* (Yan et al., 2019). Neither a PAM, nor a PFS is required for Cas12g mediated cleavage (Yan et al., 2019).

Subtype V-I effector Cas12i is a special case whereas it is a functional DNA nuclease, yet it predominantly cleaves the non-target strand on the dsDNA substrate, creating a nick (Yan et al., 2019).

Type V systems that feature single crRNA-guided effectors, namely V-A, V-I, V-J and V-M, use their respective Cas12 nucleases to process pre-crRNA transcribed from the CRISPR locus to mature crRNAs . Cas12h is also guided by a crRNA, however no data on its biogenesis has been put forward. The guide RNA processing of Cas12j is unique in the sense that it uses the RuvC domain instead of a separate catalytic center, as is the case with Cas12a (Pausch et al., 2020; Swarts et al., 2017).

1.5.3. dsDNA cleavage model of Cas12 enzymes

The main difference between dsDNA cleavage by Cas9 and Cas12 proteins is that both DNA strands need to be cleaved by a single active site in Cas12. Thus in order to produce a comprehensive model, the mechanism of strand exchange and the sequence of cleavage had to be deduced.

Firstly, like Cas9, Cas12a also adopts a dynamic cleavage-incompatible conformation prior to crRNA binding (Figure 1.6). However, the structural arrangement of apo-Cas12a might vary between different Cas12a orthologs with some assuming a particularly contrasting elongated apoprotein form with the REC lobe faced away from the NUC lobe(Dong et al., 2016; Jianwei et al., 2023; Min et al., 2018). After guide RNA binding the complex undergoes significant structural rearrangements resulting in the REC lobe rotating towards the NUC lobe, forming of a cleft in the interface between the REC and NUC lobes for crRNA- target DNA accommodation (Figure 1.6) (Jianwei et al., 2023; Swarts et al., 2017). Hybridization of the target dsDNA with the crRNA is initiated by the WED II-III, REC1 and PI domains. A conserved loop-lysine helix loop region in the PI domain is inserted into the PAM duplex at an 45° angle and promotes DNA unwinding (Paul & Montoya, 2020). Upon binding of the dsDNA substrate, the REC lobe moves slightly from the NUC lobe accommodating the substrate and unblocking the catalytic site in the RuvC domain (Figure 1.6) (Swarts et al., 2017). Molecular dynamics simulations indicate that after the binding of the substrate, the flexibility of the PI domain is notably reduced, whereas the flexibility of the REC2 and Nuc domains, which encompass the DNA binding cleft, increases (Saha et al., 2020). This most likely enables the conformational changes associated with DNA cleavage by Cas12a. Since the RuvC catalytic domain is highly rigid, the increased mobility of the adjacent Nuc and REC2 domains could contribute to the necessary exchange of the DNA strands for sequential cleavage by RuvC, release of the products and subsequent non-specific cleavage of ssDNA (Saha et al., 2020). It is accepted that the non-target DNA strand cleavage by Cas12a precedes the TS strand cleavage and can even occur independently of the TS (Stella et al., 2018; Swarts & Jinek, 2019). Recent studies further imply that the conformation change of REC2 and Nuc domains following the NTS cleavage result in the downstream dsDNA being clamped by the Nuc domain and the positioning of the TS near the RuvC active site (Figure 1.6) (Naqvi et al., 2022). Equivalent domains that could act in binding and positioning the TS have are also found in other type V effectors, namely Cas12b, Cas12e and Cas12f (J.-J. Liu et al., 2019; Takeda et al., 2021; H. Yang et al., 2016). However, the exact dynamics of dsDNA cleavage, especially as pertaining to the compact Cas12 effectors with large tracrRNAs, remain to be determined.

After *cis*-dsDNA cleavage, the PAM-distal dsDNA is released, while the PAM-proximal dsDNA remains bound to the Cas12a-crRNA complex (Figure 1.6). This maintains the complex in the catalytically competent conformation and with the RuvC active site solvent exposed and enables the non-specific

hydrolysis of ssDNA and, in some cases ssRNA, molecules (J. S. Chen, Ma, et al., 2018; Fuchs et al., 2022; Swarts & Jinek, 2019).

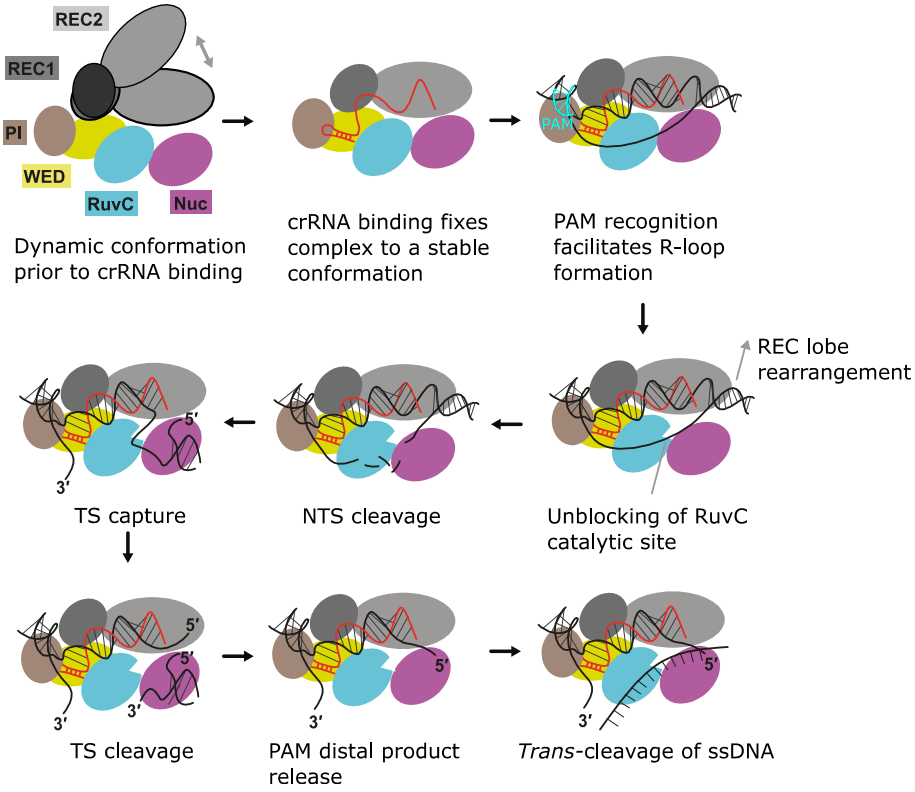


Figure 1.6. Schematic model of Cas12a-mediated *cis*- and *trans*- DNA cleavage. Cartoon illustration shows the color coded domain organization of Cas12a. The apo-protein exists in a dynamic conformation with a flexible REC lobe, which consists of the REC1 and REC2 domains (grey). Upon binding of the crRNA (red), the complex assumes a fixed, but closed conformation, with the Nuc domain (purple) in close proximity to the REC lobe. When the Cas12a-crRNA complex binds a target dsDNA substrate (black), PAM (cyan) recognition by the PI domain (brown) initiates dsDNA unwinding and R-loop formation. With the R-loop fully formed the REC lobe moves away from the Nuc domain, and the catalytic site in the RuvC domain (light blue) opens up for the cleavage of the DNA non-target strand (NTS). Subsequently, further motions of the REC2 and NUC domains enable the capture of the DNA target strand (TS) by the Nuc domain, positioning the TS near the RuvC catalytic site for cleavage. Afterwards, PAM-distal dsDNA products are released from the complex, which maintains a cleavage compatible conformation and proceeds to non-specifically degrade ssDNA *in trans*. Adapted and modified from (Naqvi et al., 2022).

1.6. CRISPR-Cas-based nucleic acid detection

While *cis* nucleic acid cleavage by Cas12 effectors is a single-turnover reaction, *trans* cleavage is multiple-turnover which enabled the use of Cas12 systems in rapid and sensitive nucleic acid detection applications (J. S. Chen, Ma, et al., 2018). The principle of these assays is the RNA-guided targeting of dsDNA or ssDNA by Cas12 nucleases and subsequent non-specific cleavage of short ssDNA molecules containing fluorophore and quencher moieties, triggering an easily detectable fluorescence signal (Figure 1.7) (J. S. Chen, Ma, et al., 2018). Combining the Cas12 activity with nucleic acid amplification enables detection of attomolar levels of target DNAs (J. S. Chen, Ma, et al., 2018). Coupling a reverse transcription step in turn permits the detection of RNAs as well (Broughton et al., 2020). The thermostability attributed to Cas12b systems was exploited to combine isothermal amplification of target nucleic acids with signal detection via Cas12 collateral nuclease activity in a single reaction (L. Li et al., 2019). These methods have been extensively applied in answer to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreak (Broughton et al., 2020; Ding et al., 2020; J. Joung et al., 2020).

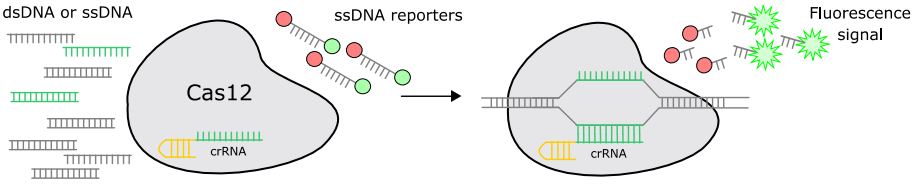


Figure 1.7. Principle of nucleic acid detection with Cas12 nucleases. Binding of dsDNA or ssDNA containing sequences complementary to the crRNA spacer (green) initiates non-specific collateral cleavage of ssDNA reporters. The reporters are labelled with fluorophore (light green) and quencher (red) moieties. The inhibition of fluorescence by the quencher is lost upon ssDNA cleavage and a fluorescence signal can be detected.

1.7. CRISPR-Cas-based genome engineering

The goal of a genome editing experiment is to introduce a desired change, or mutation, to a specific target genome sequence. Edits can be classified into several types, depending on the intended outcome: conversion of DNA base pairs (point mutations); deletion of DNA base pairs; insertion of DNA base pairs; or a combination of these changes. CRISPR-Cas nucleases or molecular

tools based on these nucleases (Figure 1.8) have been developed to mediate these types of alterations and are discussed in the following sections.

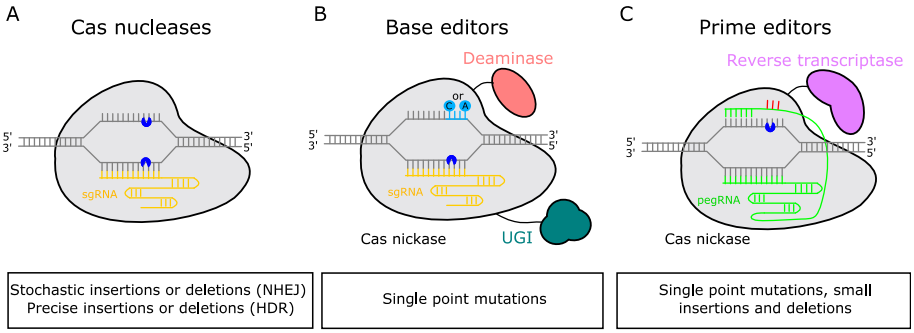


Figure 1.8. CRISPR-Cas-based genome editing molecular tools. **(A)** Cas nucleases are guided by a sgRNA molecule (yellow) to bind the DNA target and introduce a double strand break (blue) which is then repaired via non-homologous end joining (NHEJ) pathways to yield stochastic insertions or deletions, or repaired via homology directed repair (HDR) pathways to yield precise insertions or deletions, depending on the homology template provided. **(B)** Base editors contain a deaminase (light red) fused to a Cas nickase. The Cas nickase is guided by a sgRNA molecule (yellow) to bind the DNA target. The resulting R-loop displaces the non-target DNA strand which acts as a substrate for the deaminase moiety. Depending on the deaminase, either cytidines (C) or adenines (A) are then deaminated in the base editing window (light blue). The base pair mutation is achieved once the DNA is replicated using the deaminated strand as a template. This is stimulated by the nick on the sgRNA bound target DNA strand (blue), facilitating excision and repair of the nicked strand. An uracyl glycosylase inhibitor protein (UGI, dark green) fusion reduces the likelihood of uracil excision by DNA repair machinery. **(C)** Prime editors contain a reverse transcriptase domain (purple) fused to a Cas nickase. Instead of a sgRNA, the Cas nickase is guided by a prime-editing guide RNA (pegRNA, green). It consists of a canonical sgRNA module, which guides the complex to its target, and a reverse transcription template module, which harbors the desired edit (red). The 3‘ end of the pegRNA hybridizes to the DNA sequence adjacent to the nick site (blue) and the resulting heteroduplex acts as a substrate for the reverse transcriptase to extend the nicked DNA strand using the pegRNA as a template and introducing the intended edit. Prime editors can thus install single point mutations, as well as small insertions and deletions.

1.7.1. Genome editing with CRISPR-Cas nucleases

As already stated, the programmability of CRISPR-Cas systems to be directed to a desired DNA sequence and introduce a double-stranded break enabled their use in genome editing applications (Figure 1.8A). Class II single-effector protein systems offer an obvious advantage of reducing the complexity of expression and delivery in these kinds of experiments and thus are much more

widely used than their Class I counterparts. The toolbox of Cas9 and Cas12 nucleases that have been adapted for genome editing in mammalian cells is ever expanding and includes both wild type enzymes and engineered variants, in search of a tool that is efficient, specific and easy to introduce to the intended cell or tissue (Kleinstiver et al., 2019; Tsuchida et al., 2022; Walton et al., 2020; Xin et al., 2022).

While the Cas nuclease is responsible for generating a double-stranded break, the DNA repair mechanisms of the cell is what completes the sequence alteration. There are several DNA repair pathways and which is used in a particular event depends on the cell type, cell state and the nature of the double-stranded break (Chapman et al., 2012). The types of DNA repair can be divided into two branches: those that carry out re-ligation or end-joining of the broken DNA ends, often including additional nucleotide deletions or insertions at the break site (NHEJ); and those that use DNA templates for homology-directed repair (HDR) (Yeh et al., 2019). While end-joining processes are efficient in most mammalian cells, HDR is active only in dividing cells and require cell machinery that is expressed predominantly in the S and G2 cell-cycle phases (Heyer et al., 2010; Lieber, 2010). Since indel products resulting from end-joining repair usually generate frame-shifting mutations in coding sequences, CRISPR-Cas nucleases are most commonly used to selectively disrupt genes or regulatory elements (Korkmaz et al., 2016; Shalem et al., 2014; Zhu et al., 2016). However, to achieve precise genome edits, the DSB needs to be repaired by the HDR pathway in the presence of single- or double-stranded DNA donor templates, containing the intended DNA alteration (Jasin & Rothstein, 2013; Jinek et al., 2013). Methods to induce HDR in favor of NHEJ include the use of Cas nickases instead of wild type nucleases, suppression of proteins that mediate end-joining repair or expression of elements that facilitate HDR (Chu et al., 2015; Maruyama et al., 2015; Nambiar et al., 2019; Richardson et al., 2016).

1.7.2. Base editors

The difficulty of relying on HDR for precise editing prompted the development of alternative CRISPR-Cas based molecular tools. Base editors offer the ability of targeted point mutation introduction without the need of a DSB or donor DNA templates, thus eliminating the generation of random insertions and deletions resulting from the cellular repair of DSBs and bypassing the inefficient HDR process (Gaudelli et al., 2017; Komor et al., 2016). Base editors contain a catalytically impaired CRISPR-Cas nuclease (fully deactivated or nickase) fused to a DNA deaminase enzyme (Figure

1.8B) (Gaudelli et al., 2017; Komor et al., 2016). In currently developed base editors the deaminase enzyme is either a cytidine deaminase that converts cytosines to uracils, which are read by polymerases as thymines (cytosine base editor, CBE), or a deoxyadenosine deaminase which converts adenosines to inosines, which are read as guanines (adenine base editor, ABE) (Gaudelli et al., 2017; Komor et al., 2016; Lam et al., 2022). In both CBEs and ABEs, the Cas nuclease domain serves to localize the deaminase enzyme to a target sequence. Upon Cas binding, hybridization of the guide RNA spacer to the target DNA sequence causes the displacement of the non-target DNA strand to form a ssDNA R-loop, which is the substrate for the deaminase (Gaudelli et al., 2017; Komor et al., 2016). Deamination of nucleotides withing the editing window generates uridine or inosine, creating a mismatched DNA base pair in regards to the opposing non-deaminated strand. Stable base editing requires the replacement of the unedited strand, but the uracil and inosine intermediates are mutagenic and are rapidly excised by DNA repair enzymes uracil DNA N-glycosylase and N-methylpurine glycosylase, respectively (Asaeda et al., 2000; Savva et al., 1995). To improve base editing efficiency, revised CBE variants utilize fusions of uracil glycosylase inhibitor proteins (UGIs) (Komor et al., 2017; L. Wang et al., 2017). Furthermore, the dCas DNA-binding protein from early base editor variants has been replaced by a Cas nickase, to stimulate cellular repair of the non-deaminated DNA strand (Komor et al., 2016, 2017). Lastly, in the displaced ssDNA strand PAM-proximal nucleotides are occluded by the Cas effector domain and each deaminase has a distinct affinity for particular regions in the ssDNA, therefore the effective substrate for deamination is limited and defined as the base editing 'activity window', which in the canonical SpCas9-based CBE and ABEs spans the 4-8 protospacer positions (Gaudelli et al., 2017; Komor et al., 2016). This can be modulated by using different naturally occuring and engineered Cas variants, including SaCas9, Cas12a orthologs, as well as varying the deaminase domain (Y. B. Kim et al., 2017; Lam et al., 2022; X. Li et al., 2018; Z. Liu et al., 2019). Recently, dual deaminase base editing systems have been developed, enabling concurrent C-T and A-G conversions at the same target site (Y. Liang et al., 2022; X. Zhang et al., 2020).

1.7.3. Prime editors

Prime editing is a recent CRISPR-Cas-based genome editing technology that allows all 12 possible base conversions, as well as small insertions and deletions, without requiring DSBs or donor DNA templates (Anzalone et al., 2019). Prime editors (PEs) consist of a Cas nickase fused to a reverse

transcriptase paired with a prime-editing guide RNA (pegRNA), which both specifies the target site and encodes the desired edit (Figure 1.8C) (Anzalone et al., 2019). The pegRNA harbors a spacer sequence complementary to the target strand, the Cas nuclease binding scaffold (tracrRNA), a reverse transcriptase template encoding the desired edit and a primer binding site that can hybridize with the 3'end of the nicked non-target DNA strand (Anzalone et al., 2019). The associated reverse transcriptase can then extend the nicked ssDNA using the template encoded in the pegRNA (Anzalone et al., 2019). This results in a newly synthesized edited DNA strand, that exists as a 3' DNA flap in equillibrium with a 5' flap containing the original sequence. While the strand containing the unedited sequence is more thermodynamically favored to hybridize to the opposing strand, action by cellular structure-specific nucleases or 5' exonucleases result in the favored excision of the 5' flap and formation of a heteroduplex (Anzalone et al., 2019). Finally, the complete incorporation of the edit occurs after the opposing DNA strand is DNA repaired using the edited strand as a template, the process of which can be simulated by introducing another nick, this time using a separate sgRNA targeting the unedited non-target strand (Anzalone et al., 2019). Prime editing efficiency can be further enhanced by using pegRNAs protected from exonuclease activity by capping the 3' end with various structural motifs, as well as suppressing the mismatch repair machinery of the cell, so the edit-containing heteroduplex is not converted into its original sequence (P. J. Chen et al., 2021; Nelson et al., 2021).

Prime editing offers several advantages over other CRISPR-Cas based genome editing techniques. Firstly, prime editors are able to install point mutations further than Cas nuclease-mediated HDR approaches, which makes PAM availability less restrictive (Anzalone et al., 2019; X. Liang et al., 2017). Furthermore, Cas9-based off-target prime editing was found to be lower than the off-target indel generation by Cas9 nuclease, most likely due to the additional nucleic acid hybridization steps: nicked target strand - pegRNA primer binding site hybridization and 3' flap – target strand hybridization (Anzalone et al., 2019).

1.8. Limitations

While CRISPR-Cas systems have enabled various approaches of genetic manipulations, there still remain several challenges that need to be overcome, namely target space limitations (determined by PAM preferences of particular Cas nucleases), specificity of cleavage activity and potential off-target effects, and the delivery of CRISPR-Cas molecular tools to cells and tissues.

Some CRISPR-Cas applications, for example gene disruption through frame-shifting indels, permit a rather lenient target selection, since the DSB could be introduced at any region of the gene in question. However, in cases where precise editing of a specific location is required, particularly with base editing where the effective editing window needs to be fine-tuned, possessing flexibility in the PAM sequence is crucial. Expanding the PAM recognition space of Cas nucleases is carried out in two ways: searching for new Cas nucleases in nature or employing protein engineering to alter already characterized enzymes.

To date over 1100 genomes carrying CRISPR-Cas loci annotated as containing type II Cas nucleases and 150 containing type V nucleases are identified, with these values bound to increase with further sequencing efforts and more comprehensive classification and annotation of type V systems (Makarova et al., 2020). Exploring this diversity has yielded an assortment of Cas nucleases with varying PAM preferences (Amrani et al., 2018; Fedorova, Arseniev, et al., 2020; Jacobsen et al., 2019; Schmidt et al., 2019; Tsui et al., 2017; Zetsche et al., 2020). Alternatively, protein engineering is used to alter PAM recognition via rational or random mutagenesis of Cas nuclease PAM interacting domains (Collias et al., 2020; Kleinstiver et al., 2015, 2016). This has eventually resulted in the development of the SpRY Cas9 variant with the least stringent PAM preference of NR (R = A or G) to date (Walton et al., 2020). Advancements in modifying the PAM profile of Cas12 enzymes are limited in comparison to SpyCas9, with the work mostly reported on altering the PAM preferences rather than relaxing them (L. Gao et al., 2017; R. M. Liu et al., 2019; Tóth et al., 2020).

While the development of a PAM-less nuclease looks to be achievable in the near future, it would potentially introduce several challenges. Firstly, for Cas nuclease guide RNAs expressed from DNA constructs, self-targeting would either lead to the clearance of the gRNA-encoding plasmid or cell death in bacteria, and indel formation within the guide in eukaryotes. Furthermore, in theory a nuclease with no PAM preference would interrogate every sequence in the genome, possibly leading to extended timescales for the enzyme to find its target and subsequent reduction in overall editing efficiency and more susceptibility for off-targeting, which is already observed with variants recognizing relaxed PAMs (Walton et al., 2020).

Cas nucleases in some instances can tolerate mismatches between guide RNA and target sequences, leading to cleavage activity at undesired sites (Y. Fu et al., 2013; D. Kim, Kim, Hur, et al., 2016). For example, to minimize off-target activity with SpyCas9 the target sequence must differ by at least 3 nucleotides from any other sequence in the genome (D. Kim, Kim, Kim, et al.,

2016). In addition to unwanted local mutagenesis, off-target cleavage activity can generate chromosomal rearrangements between two DSB sites leading to further genomic instability (Kosicki et al., 2018; Rayner et al., 2019). Several strategies to mitigate the activity at unintended sites are being applied. Firstly, bioinformatic prediction of potential off-target sites aids during target selection and guide RNA design (Labun et al., 2019). Guide RNA engineering by truncating the spacer sequence or adding chemical modifications have also been shown to increase fidelity (Y. Fu et al., 2014; Ryan et al., 2018; Yin et al., 2018). Moreover, several Cas9 nuclease variants have been developed via protein engineering or directed evolution exhibiting increased accuracy (J. S. Chen, Dagdas, et al., 2018; Hu et al., 2018; Kleinstiver et al., 2016; J. K. Lee et al., 2018; Slaymaker et al., 2016). Natural Cas9 orthologs as well as Cas12 enzymes have also been shown to possess reduced off-target activity (Amrani et al., 2018; E. Kim et al., 2017; Ran et al., 2015; Xin et al., 2022). Lastly, editing approaches that do not introduce a double-strand break, like CRISPR-based transcriptional regulation, epigenetic modification, base and prime editing, are all less likely to cause significant genomic instability (Anzalone et al., 2019; Gilbert et al., 2013; Komor et al., 2016; X. S. Liu et al., 2016).

For any of the genome editing methods discussed here to be used therapeutically, the CRISPR-Cas molecular tools, which include the enzymes, guide RNA, any DNA repair templates, need to be introduced into the cells subject to genomic repair. The most common strategies of CRISPR-Cas system delivery use viral vectors, lipid nanoparticles or RNP electroporation. Viral delivery, particularly adeno-associated virus (AAV) vectors, offer the advantages of efficiency and tissue-selectivity (Flotte, 2004; Samulski & Muzyczka, 2014). Also, the AAV genome can be maintained as an episome for continuous expression of transgenes over long periods of time (Buchlis et al., 2012). However, AAV vectors have a cargo packaging limit of 4.7 kb (Z. Wu et al., 2010). Considering that the most widely used SpyCas9 nuclease gene is 4.2 kb long it can‘t be fit into a single AAV vector along with its guide RNA and necessary promoter and polyadenylation signals, requiring the use of separate vectors at the cost of reduced efficiency (Y. Yang et al., 2016; Zetsche, Volz, et al., 2015). This challenge is even more pronounced in the case of base editing and prime editing systems, which contain additional enzyme fusions and auxiliary elements (Y. Chen et al., 2020; Zhi et al., 2022). Alternatives to AAV include lentiviral and adenoviral vectors, both of which have higher capacities for transgene cargo (Stephens et al., 2019; Uchida et al., 2021). Recently, lipid nanoparticle delivery is gaining traction, enabling the delivery both of CRISPR-Cas mRNAs and RNPs (Kenjo et al., 2021; Qiu et al., 2021; Suzuki et al., 2021; T. Wei et al., 2020). Lastly, physical delivery

in the form of electroporation of cells with preformed Cas RNPs has been applied for *ex vivo* editing of stem cells and animal zygotes (Kagita et al., 2021; Qin & Wang, 2018; H. Xu et al., 2019; L. Xu et al., 2019). However, this approach is likely to have limited potential in *in vivo* editing applications, requiring the targeting of small organs or organisms at early stage of development (Taha et al., 2022).

Additionally, there is the challenge of immunogenicity against CRISPR-Cas proteins, particularly those from pathogenic bacteria (Ralph & Carapetis, 2012; S. Y. C. Tong et al., 2015). A significant portion of human population possess pre-existing antibodies against *S. pyogenes* and *Staphylococcus aureus* Cas9 enzymes (Charlesworth et al., 2019; Wagner et al., 2019). Also, viral-vectors used for delivery of CRISPR-Cas systems can likewise generate an immune response (Verdera et al., 2020). Such immunogenicity can potentially lead to failure of the intended editing or even damage to the cells (A. Li et al., 2020). Strategies to mitigate this effect include masking of immunogenic Cas9 epitopes, immunosuppression of targeted cells, as well as use of orthogonal Cas proteins (Ferdosi et al., 2019; Moreno et al., 2019; Rim et al., 2021).

Finally, CRISPR-Cas editing was shown to be influenced by p53-mediated response, potentially facilitating inadvertent selection for p53-supressed cells and posing the risk of tumorigenesis (Haapaniemi et al., 2018; Ihry et al., 2018).

In conclusion, the current limitations of CRISPR-Cas systems highlight the importance of the discovery and development of novel Cas enzyme orthologs as a means to expand the targeting space, ease the delivery and potentially alleviate the immune response generated by CRIPSR-Cas systems.

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Chemicals

All chemicals used in this study were obtained from Roth, Sigma-Aldrich, Thermo Fisher Scientific of the highest purity grade available.

2.1.2.Commercial proteins, dyes and kits

T4 DNA polymerase, Q5 DNA polymerase, Phusion High-Fidelity PCR Master Mix, T4 Ligase, Murine RNase Inhibitor, Exonuclease V, T5 Exonuclease, Proteinase K, DNaseI, EcoRI, HindIII enzymes used in this study were obtained from New England Biolabs.

DreamTaq DNA polymerase, RNase A, dNTP mix, Hi-Di Formamide, GeneScan LIZ 120 Size Standard were obtained from Thermo Fisher Scientific.

CircLigase™ single stranded (ss) DNA Ligase was obtained from Lucigen.

GelRed nucleic acid stain was obtained from Biotium.

6x Blue Gel loading dye was obtained from New England Biolabs.

HiScribe™ T7 Quick High Yield RNA Synthesis Kit, PURExpress bacterial IVT kit, NEBuilder HiFi DNA Assembly kit, Monarch RNA Cleanup Kit, Monarch PCR & DNA Cleanup purification kit, NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® were obtained from New England Biolabs.

2.1.3.Bacterial strains

E. coli strain DH5α (New England Biolabs) (*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) used for cloning procedures.

E. coli strain NiCo21(DE3) (New England Biolabs) (*can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*) used for Cas9 ortholog and Cas12l expression.

E. coli strain T7express lysY/Iq (New England Biolabs) (MiniF *lysY lacIq(CamR) / fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-*

73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr) 114::IS10) used for Cas9 ortholog expression.

E. coli strain NEB® Express Iq (New England Biolabs) (MiniF *lacIq (CamR) / fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10)* used for Cas9 ortholog expression.

E. coli strain Arctic express (DE3) (Agilent) (*F– ompT hsdS(rB – mB –) dcm+ Tetr gal λ(DE3) endA Hte [cpn10 cpn60 Gentr]*) used for Cas12l expression.

2.1.4.Buffers

In vitro cleavage reactions

Assembly buffer:

10 mM Tris-HCl pH 7.5 at 25°C, 100 mM NaCl, 1 mM DTT

Reaction buffer:

10 mM Tris-HCl pH 7.5 at 25°C, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT

CutSmart® buffer (New England Biolabs):

50 mM KOAc, 20 mM Tris-OAc, 10 mM $Mg(OAc)_2$, 100 μg/ml BSA pH 7.9 at 25°C

NEBuffer 1.1 (New England Biolabs):

10 mM Bis-Tris-Propane-HCl, 10 mM $MgCl_2$, 100 μg/ml BSA pH 7.9 at 25°C

NEBuffer 2.1 (New England Biolabs):

50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 100 μg/ml BSA pH 7.9 at 25°C

NEBuffer 3.1 (New England Biolabs):

100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 μg/ml BSA pH 7.9 at 25°C

Purification of Cas12l proteins

Resuspension and sonication buffer:

20 mM Tris-HCl pH 7.5 at 25°C, 300 mM NaCl, 40 mM imidazole, 1x Halt™ Protease Inhibitor (Thermo Fisher Scientific)

HisTrap column elution buffer:

20 mM Tris-HCl pH 7.5 at 25°C, 300 mM NaCl, 700 mM imidazole

HiTrap Heparin column loading buffer:

20 mM Tris-HCl pH 7.5 at 25°C, 500 mM NaCl

HiTrap Heparin column elution buffer:

20 mM Tris-HCl pH 7.5 at 25°C, 2 M NaCl
Protein storage buffer:
20 mM Tris-HCl, pH 7.5 at 25°C, 500 mM NaCl, 1 mM EDTA, 1 mM DTT, and 50% (v/v) glycerol.

2.1.5. Proteins and nucleic acids

Protein and guide RNA sequences are deposited online:

Cas9 ortholog amino acid sequences	https://benchling.com/urbaitis/f_/VJJfjl5L-cas9-ortholog-amino-acid-sequences/
Cas9 ortholog tracrRNA sequences	https://benchling.com/urbaitis/f_/PvyACMal-cas9-ortholog-tracrrna-sequences/
Cas12l ortholog amino acid sequences	https://benchling.com/urbaitis/f_/nlwtApCm-cas12l-ortholog-amino-acid-sequences/
Cas12l ortholog sgRNA sequences	https://benchling.com/urbaitis/f_/rf73Ikhf-cas12l-ortholog-sgrna-sequences/

Further information on DNA and RNA sequences are provided in appendices listed below:

- Appendix 1 – Cas12l DNA and RNA substrates
- Appendix 2 – Cas12l expression plasmids
- Appendix 3 – Cas12l target plasmids

2.2. Methods

2.2.1. Identification and phylogeny of Cas9 orthologs

Type II Cas9 endonucleases were identified by searching for the presence of an array of CRISPRs using PILER-CR 1.06 (Edgar, 2007). Following identification, the DNA sequences surrounding the CRISPR array (about 15 kb 5' and 3' of the CRISPR array) were examined for the presence of open-reading frames (ORFs) encoding proteins greater than 750 amino acids. Next, to identify cas genes encoding Cas9 orthologs, multiple sequence alignment of sequences from a diverse collection of Cas9 proteins was performed using

MUSCLE 3.8.31 (Edgar, 2004) and then used to build profile hidden Markov models (HMMs) for Cas9 sub-families using HMMER 3.2.1 (Eddy, 2011). The resulting HMMs were then utilized to search protein sequences translated from the cas ORFs for the presence of genes with homology to Cas9. Alternatively, Cas9 orthologs and the metagenomic sequence encoding them were obtained from publicly available datasets through the Joint Genome Institute's Integrated Microbial Genomes & Metagenomes resource (IMG/M): https://img.jgi.doe.gov/cgi-bin/m/main.cgi (I.-M. A. Chen et al., 2019). Only proteins containing the key HNH and RuvC nucleolytic domains and catalytic residues defining a type II Cas9 protein (Nishimasu et al., 2014) were selected. Through phylogenetic analyses (MEGA7 10.0.5 (Kumar et al., 2016)) Cas9 proteins were then parsed into distinct families and representative members of each group used to select orthologs for characterization. To place our collection in context with previously described Cas9 orthologs, a phylogenetic tree was built using type II-A, -B, and -C representatives (Makarova et al., 2020) and those we selected for characterization using MEGA7 employing Neighbor-Joining (Saitou & Nei, 1987) and Poisson correction (Zuckerkandl & Pauling, 1965) methods. Sequences of Cas9 proteins are listed in section 2.1.5.

2.2.2. Engineering Cas9 single guide RNA solutions

The trans-activating CRISPR RNA (tracrRNA) essential for CRISPR RNA (crRNA) maturation (Deltcheva et al., 2011) and Cas9 directed target site cleavage in type II systems (Jinek et al., 2012; Karvelis et al., 2013) was identified by searching for a region in the vicinity of the *cas9* gene, the anti-repeat, which may base-pair with the CRISPR repeat and was distinct from the CRISPR array(s). Once identified, the possible transcriptional directions of the putative tracrRNAs for each system were established by examining the secondary structures using UNAfold 3.9 (Markham & Zuker, 2008) and possible termination signals present in RNA versions corresponding to the sense and anti-sense transcription scenarios surrounding the anti-repeat. Based on the likely transcriptional direction of the tracrRNA and CRISPR array, single guide RNAs (sgRNAs), representing a fusion of the CRISPR RNA (crRNA) and tracrRNA (Jinek et al., 2012), were designed. For each ortholog, this was accomplished by linking 16 nt of the crRNA repeat to the complementary sequence of the tracrRNA anti-repeat by a 4 nt GAAA loop similar to that described previously for Spy Cas9 (Jinek et al., 2012). RNA sequences of Cas9 tracrRNAs are listed in section 2.1.5.

2.2.3. Computational analysis of Cas9 tracrRNAs

BLAST 2.7.3 (with parameters to optimize finding short sequences in highly-repetitive regions (-task blastn_short -dust no)) (Altschul et al., 1990) was used to identify sequences homologous to the 79 identified tracrRNAs. The resulting collection of identified sequences were grouped using CD-HIT 4.7 (L. Fu et al., 2012) at a 90% sequence similarity threshold. The resulting clusters were filtered to remove groups that did not contain at least one of the 79 reference tracrRNA sequences. Next, sequence homology and secondary structure models were constructed for each group using MAFFT 7.407 (Katoh & Standley, 2013) and RNAalifold 2.4.5 (Lorenz et al., 2011), respectively. Both models were then used to search for sequence/structural homology in the full set of reference and BLAST-identified sequences using the RNA structure search tools in the Infernal 1.1 software suite (Nawrocki & Eddy, 2013). Structural overlap between clusters was then generated by comparing the results of each covariance model (CM). To graph the relationship among tracrRNAs, vertices were first added for each representative CM (sequences with both shared secondary structure predictions and at least 90% sequence similarity). If two vertices shared a CM, they were connected with a line weighted by the percent similarity between shared vertices ($percent\ similarity = \frac{\#\ of\ shared\ sequences}{\min(\#\ found\ by\ model\ 1,\ \#\ found\ by\ model\ 2)}$).

2.2.4. Production of single guide RNAs

All single guide RNA (sgRNA) molecules used in this study were synthesized by in vitro transcription using HiScribe™ T7 Quick High Yield RNA Synthesis Kits (New England Biolabs), or transcribed directly in the in vitro translation (IVT) reaction. Templates for sgRNA transcription were generated by PCR amplifying synthesized fragments (IDT and Genscript) or by annealing a T7 primer oligo to a single stranded template oligonucleotide. Transcribed RNA products were treated with DNaseI (New England Biolabs) to remove DNA templates and cleaned up with Monarch RNA Cleanup Kit (50µg) (New England Biolabs) and eluted in nuclease-free water. RNA concentration and purity were measured by NanoDrop spectrophotometry and RNA integrity was visualized by agarose gel electrophoresis and staining with GelRed (Biotium).

2.2.5. PAM library cleavage using *in vitro* translated Cas9s

Cas9 was produced by IVT using PURExpress bacterial IVT kit (New England Biolabs), following the manufacturer's recommended protocol similar to that described previously (Karvelis et al., 2018). Plasmid DNA encoding *E. coli* codon optimized Cas9s were generated for use as templates for IVT reactions. Synthetic DNA fragments were synthesized by Twist Bioscience and assembled by NEBuilder HiFi DNA Assembly kit (New England Biolabs) into pET28a (EMD Millipore). Following IVT, 20 µl of supernatant containing soluble Cas9 protein was mixed with Murine RNase Inhibitor (40 U; New England Biolabs) and 2 µg of T7 in vitro transcribed sgRNA and incubated for 15 min. at room temperature. Alternatively, the sgRNA was transcribed directly in the IVT kit by supplying a DNA template containing a T7 promoter and sequence encoding the respective sgRNA. In this situation, 0.5 µg of plasmid encoding the cas9 gene and a 100-fold molar excess of sgRNA template was added to the IVT reaction mix. 10 µl (or series of ten-fold dilutions) of the resulting Cas9-sgRNA ribonucleoprotein (RNP) complex were then combined with 1 µg of the 7 bp randomized PAM library described previously (Karvelis et al., 2015) in a 100 µl reaction buffer (10 mM Tris-HCl pH 7.5 at 37°C, 100 mM NaCl, 10 mM MgCl2, 1 mM DTT) and incubated for 60 min. at 37°C.

2.2.6. Capture and sequencing of cleaved library fragments

Cleaved library fragments were captured by adapter ligation, enriched by PCR amplification, and deep sequenced as described in (Karvelis et al., 2015). Briefly, cleaved libraries were first subjected to DNA end-repair by incubation with 0.3 µl (1U) of T4 DNA polymerase (New England Biolabs) and 0.3 µl of 10 mM dNTP mix (Thermo Fisher Scientific) for 15 min. at 12°C and inactivated by heating (75°C for 20 min.). To efficiently capture free DNA ends, a 3'-dA overhang was added by incubating the reaction mixture with 0.3 µl (1.5 U) of DreamTaq polymerase (Thermo Fisher Scientific) for 30 min. at 72°C. The resulting DNA was then purified (Monarch PCR & DNA Cleanup purification column (New England Biolabs)) and ligated to adapters with a 3' dT overhang with 1 µl 400 U of T4 Ligase (New England Biolabs) in 25 µl of ligation buffer (50 mM Tris-HCl, pH 7.5 at 25°C, 10 mM MgCl2, 10 mM DTT, 1 mM ATP, 5% (w/v) PEG 4000). After 1 h at room temperature, 10 µl of the ligation reaction was used as template in a PCR reaction (Q5 DNA polymerase (New England Biolabs); 15 cycles; 100 µL final reaction volume) containing primers specific to the PAM-side of the library and the adapter.

DNA was next purified (Monarch PCR & DNA Cleanup purification column (New England Biolabs)) and the sequences and indexes required for Illumina deep sequencing were incorporated through two rounds of PCR (Phusion High-Fidelity PCR Master Mix in HF buffer (New England Biolabs); 10 cycles each round; 50 μL final reaction volume). The resulting products were then deep sequenced on a MiSeq Personal Sequencer (Illumina) with a 25% (v/v) spike of PhiX control v3 (Illumina).

2.2.7. Identification of PAM preferences

PAM sequences that supported double-stranded DNA target cleavage were determined as described in (Karvelis et al., 2015, 2018). Briefly, after sequencing, the location of cleavage within the library protospacer was first assessed by evaluating the position with the greatest number of adapter ligated reads using a custom script (provided at https://github.com/cortevaCRISPR/Cas12f-InformaticsTools). The PAMs associated with library fragments that supported cleavage were then extracted and used to evaluate the bias in the bp composition at each position within the randomized PAM library relative to that in the starting library by normalization ((Treatment Frequency)/((Control Frequency)/(Average Control Frequency))). Next, PAM preferences were quantified using position frequency matrices (PFMs) and displayed as a WebLogo. 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 IVT or *E. coli* cell lysate mixtures.

2.2.8. Cas9 expression and purification

Spy, *Streptococcus thermophilus* CRISPR3 (Sth3), and *S. thermophilus* CRISPR1 (Sth1) Cas9 proteins cloned into the pBAD-Chis vector were expressed in *E. coli* DH10B strain at 16°C for 20 h in the presence of 0.2 % (w/v) arabinose. Other orthologs were first *E. coli* codon optimized and cloned into the pET28 vector yielding constructs encoding fusion proteins comprising a C-terminal 6-His-tag. In some instances, sequences encoding nuclear localization sequences (SV40 origin) were incorporated onto the 5' and 3' ends of the cas9 gene. The expression of each ortholog was then tested in different *E. coli* strains (NiCo21(DE3), T7 Express lysY/Iq, NEB® Express Iq) under various growth conditions (media, temperature, induction) with the amount of protein produced being measured by SDS-PAGE analysis. Optimized conditions then were chosen for flask scale purification. Cells were

disrupted by sonication. The supernatant was loaded onto HiTrap DEAE Sepharose (GE Healthcare), followed by subsequent purification on Ni2+-charged HiTrap chelating HP column (GE Healthcare) and HiTrap Heparin HP (GE Healthcare) columns. Purified Cas9 proteins were stored at −20°C in 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 1 mM EDTA, 1 mM DTT, and 50% (v/v) glycerol.

2.2.9. Computational analysis of Cas9 PAM interacting domains

The Cas9 orthologs characterized here were aligned using MAFFT 7.407 (Katoh & Standley, 2013). Their PAM interacting (PI) regions corresponding to the C-terminal domain of *Streptococcus pyogenes* Cas9 (4ZT0_A:1090-1365) were extracted and used as queries for two iterations of PSI-BLAST 2.2.26 (Altschul et al., 1997) search against the NCBI NR protein collection, UniRef100 and Mgnify (Mitchell et al., 2019) databases. Hits were extracted, filtered to 80% identity using CD-HIT 4.6 (L. Fu et al., 2012) and clustered with CLANS 1.0 (Frickey & Lupas, 2004). CLANS is an implementation of the Fruchterman-Reingold force-directed layout algorithm, which treats protein sequences as point masses in a virtual multidimensional space, in which they attract or repel each other based on the strength of their pairwise similarities (CLANS P-values). CLANS P-values are calculated from BLAST E-values by dividing them by effective search space used. Resulting CLANS networks were visually inspected and clusters identified. For groups larger than 150 sequences, a phylogenetic analysis was performed recovering sequences which were filtered out during the previous step and removing identical ones. Next, multiple sequence alignments were performed for clusters 1 to 6 using MAFFT (options: "--ep 0.123 --maxiterate 20 --localpair") and regions with gaps removed with trimAL 1.2 (Capella-Gutiérrez et al., 2009) (option: "-gt 0.01"). Lengths of the resulting alignments varied from 359 to 652 residues in clusters 2 and 3, respectively. Phylogenetic trees were generated using IQtree 1.6.10 (Nguyen et al., 2015) with auto model selection and 1000 fast bootstrap (options: "-alrt 1000 -bb 1000").

2.2.10. Evaluation of protospacer cleavage patterns

To capture protospacer cleavage patterns with single molecule resolution, we developed a minicircle double stranded (ds) DNA substrate that allows both ends of target cleavage to be captured in a single Illumina sequence read. First, 124 nt oligonucleotides (IDT) were circularized using with CircLigase™

single stranded (ss) DNA Ligase (Lucigen) according to the manufacturers suggestion. Circularized ssDNA was next purified and concentrated using a Monarch® PCR & DNA Cleanup Kit (NEB). 20 pmol of the purified product was then incubated with 25 pmol of a complementary primer in 1X T4 DNA ligase buffer (NEB) supplemented with 40 µM dNTPs. To allow the primer to anneal, the reaction was then heated to 65°C for 30 seconds followed by a decrease in temperature to 25°C at a rate of 0.2°C/second. 6 units of T4 DNA polymerase and 400 units of T4 DNA ligase (NEB) were then added and the reaction was incubated at 12°C for 1 hour to allow second strand synthesis. Following purification with a Monarch® PCR & DNA Cleanup Kit and elution into 1X CutSmart® buffer (NEB) containing 1 mM ATP, 15 units of Exonuclease V (RecBCD; NEB) and T5 exonuclease (NEB) were added to the sample and incubated at 37°C for 45 min. 0.04 units of Proteinase K (NEB) was then added and the sample was incubated at 25°C for 15 min. prior to purification with a Monarch® PCR & DNA Cleanup Kit. After elution, the yield of circular double-stranded DNA was assessed using an Agilent 2100 Bioanalyzer.

For minicircle digestion, Cas9 RNPs were formed by incubating 1 pmol of sgRNA with 0.5 pmol of Cas9 protein in 1X NEBuffer™ 3.1 or 2.1 (NEB) at room temperature for 10 min. 0.1 pmol of circular dsDNA substrate was added, samples were incubated for 15 min at 37°C and then each 20 µl reaction was quenched by the addition of 5 µl of 0.16 M EDTA. Reactions were concentrated and purified with a Monarch® PCR & DNA Cleanup Kit and the entire 8 µl of eluted product was used as a substrate for Illumina sequencing library construction using a NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB) and the protocol provided with the kit. 15 cycles of PCR were used to add the Illumina priming sequences and index barcodes and then the concentration of each reaction was assessed on an Agilent 2100 Bioanalyzer. Libraries were pooled and sequenced on either an Illumina NovaSeq or NextSeq instrument with 2 X 150 paired-end sequencing runs. Cleavage sites were then mapped using custom scripts71 and visualized as heatmaps (representing proportion cleaved) using Microsoft Excel 16.36 and GraphPad Prism 8.

2.2.11. *In vitro* Cas9 cleavage assays for determining optimal buffer, temperature and spacer length

First, DNA substrates containing a canonical PAM for each ortholog were amplified from HEK293T genomic DNA by PCR using primers corresponding to WTAP and RUNX1. Forward primers were labelled with 5’-

FAM and 5'-ROX for WTAP and RUNX1, respectively. Reverse primers were unlabelled. 515 and 605 bp PCR products for WTAP and RUNX1, respectively, were then purified with a Monarch® PCR & DNA Cleanup Kit (5µg) (NEB T1030S) and DNA concentration and purity measured by NanoDrop™ spectrophotometry (ThermoFisher). Purified Cas9 protein was then diluted to 1µM in dilution buffer (300mM NaCl, 20mM Tris, pH7.5) and stored on ice. Next, sgRNAs were diluted to 2 µM in nuclease-free water. Cas9 and sgRNA were then combined in a 2:1 sgRNA:Cas9 molar ratio in reaction buffer at room temperature for 10 min. Substrate was added next at a Cas9:sgRNA:DNA ratio of 10:20:1 and incubated for 30 min. For buffer optimization and spacer length preference experiments, 1X NEBuffers 1.1, 2.1, 3.1, or CutSmart (NEB B7200S) were used as reaction buffers and incubations took place at 37ºC. For thermoactivity experiments, reactions were performed in NEBuffer 3.1. Here, RNPs were initially formed at room temperature and then transferred to a thermal cycler pre-heated or cooled to the various assay temperatures prior to DNA substrate addition. 10x DNA substrate (100nM) was separately equilibrated at the designated temperature prior to being added to the RNP containing reaction tubes. Reactions were quenched by adding SDS to 0.8% (v/v) and 80 mU Proteinase K (NEB P8107S). Cleavage products were diluted 4X in nuclease-free water and subjected to capillary electrophoresis (CE) to quantify the extent of cleavage72. The fraction of substrate cleaved at each temperature was then visualized as heatmaps, using Microsoft Excel 16.36 and GraphPad Prism 8.

2.2.12. Cas9 protein thermal stability

Purified Cas9 proteins were diluted in 300 mM NaCl, 20 mM Tris, pH7.5 to 5-10 µM at room temperature. 10µL of the diluted protein was loaded into NanoDSF Grade Standard Capillaries (NanoTemper) and melting temperatures were determined using a Prometheus NT4.8 NanoDSF instrument according to the manufacturer's instruction. Temperature was increased from 20ºC to 80ºC at the rate of 1ºC/min. Inflection points of melting curves are reported as the Tm.

2.2.13. Identification of CRISPR-Cas12l

First, arrays of CRISPRs were detected within microbial sequences using PILER-CR (Edgar, 2007) and MinCED (Bland et al., 2007) software programs. Next, known CRISPR-Cas systems were removed from the dataset by searching the proteins encoded in the vicinity (20 kb 5'and 20 kb 3' (where

possible)) of the CRISPR array for homology with known CRISPR associated (Cas) proteins utilizing a set of position specific scoring matrices (PSSMs) encompassing all known Cas protein families as described in Makarova, *et al*. 2015. To aid in the complete removal of known Class 2 CRISPR-Cas systems, multiple-sequence alignment of protein sequences from a collection of orthologs from each family of Class 2 CRISPR-Cas endonucleases (e.g. Cas9, Cpf1 (Cas12a), C2c1 (Cas12b), C2c2 (Cas13), C2c3 (Cas12c)) was performed using MUSCLE (Edgar, 2004). The alignments were examined, curated and used to build profile hidden Markov models (HMM) using HMMER (Eddy, 1998; Finn et al., 2011). The resulting HMM models were then utilized to further identify and remove known Class 2 CRISPR-Cas systems from the dataset. Next, using PSSM specific searches as described above, the CRISPR loci that remained were evaluated for the presence of genes encoding proteins implicated as being important for spacer insertion and adaptation, Cas1 and Cas2. CRISPR loci containing *cas1* and *cas2* genes were then selected and further examined to determine the proximity, order and directionality of the undefined genes encoded in the locus relative to the *cas1* and *cas2* genes and CRISPR array. Only those CRISPR loci forming an operon-like structure where a large (≥1500 bp open-reading frame) undefined gene was present close to and in the same transcriptional direction as the *cas1* and *cas2* genes were selected for further analysis. Next, the protein encoded by the undefined gene was analyzed for sequence and structural features indicative of a type V nuclease domain. First, depending on how much similarity existed between a candidate sequence and known proteins, various bioinformatics tools were employed to reveal its conserved functional features, from pairwise comparison, to family profile search, to structural threading, and to manually structural inspection. In general, homologous sequences for candidate proteins were first collected by a PSI-BLAST (Altschul et al., 1997) search against the National Center for Biotechnology Information (NCBI) non-redundant (NR) protein collection with cut-off e-value of 0.01. After redundancy reduction at ~90% identical level, groups of homologous sequences with various member inclusion thresholds (such as >60, 40, or 20% identity) were aligned to reveal conserved motifs by multiple-sequence alignment tools, MSAPRobs (Y. Liu et al., 2010) and ClustalW (Thompson et al., 1994). The most conserved homologous sequences underwent a sequence to family-profile search by HMMER (Eddy, 1998), against numerous domain databases including Pfam, Superfamily and SCOP (Murzin et al., 1995) and home-built structure-based profiles. Separately, the resulting candidate's homologous sequence alignment was also used to generate a candidate protein profile with addition of predicted secondary structures. The candidate profile was further used to

do a profile-profile search by HHSEARCH (Söding, 2005), against pdb70_hhm and Pfam_hhm profile databases. In the next step, all detected sequence-structure relationships and conserved RuvC-like motifs were threaded into a 3D structure template with MODELLER or manually mapped into the known structural reference on DiscoveryStudio (BIOVIA) and Pymol (Schrodinger). Finally, to verify and confirm the potential biological relevance, the catalytic or most conserved residues and key structural integrity were manually inspected and evaluated in light of the protein's ability to metabolize DNA. Following RuvC identification, the other proteins encoded within the locus (5 kb 5' and 5 kb 3' from the ends of the newly defined CRISPR-Cas system (where possible)) were next examined for homology to known proteins families using InterProScan software (EMBL-EBI, UK) and through comparison with the NCBI NR protein collection using the BLAST program (Altschul et al., 1990). Cas12l protein sequences are listed in section 2.1.5.

2.2.14. Engineering CRISPR-Cas12l systems to target a PAM library

Cas12l CRISPR systems were modified to target the 7 bp randomized PAM library described previously (Karvelis et al., 2018). The native CRISPR array was replaced with four repeat:spacer:repeat units, two of which encoded 37 nt spacers capable of base pairing with the anti-sense strand 5' of the randomized PAM sequence in the library, and two that encoded sequence capable of targeting the anti-sense strand 3' of the randomized PAM region. The resulting engineered CRISPR-Cas loci were synthesized (GenScript) and cloned into a low copy number *E. coli* plasmid (pET-Duet-1) modified to contain a single isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible T7 promoter (MilliporeSigma).

2.2.15. Detecting Cas12l dsDNA cleavage and PAM recognition

E. coli ArcticExpress (DE3) (Agilent Technologies) cells were transformed with the plasmid-borne engineered CRISPR-Cas12l systems. Cultures were grown in 20 ml of LB media containing 20 µg/ml gentamycin and 100 µg/ml carbenicillin. After the cultures reached OD_{600} of 0.6, expression was induced with 0.1 mM IPTG and the cultures were then incubated overnight at 16°C. Cells from aliquots of 10 ml were collected by centrifugation and resuspended in 1 ml of lysis buffer (20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 5% (v/v)

glycerol) supplemented with 1 mM PMSF and lysed by sonication. Cell debris was removed by centrifugation and 10 µl of the obtained supernatant were used in the plasmid cleavage reactions.
Lysate containing Cas12l RNP complexes were then used to cleave the 7 bp randomized PAM plasmid library. 10 µl of clarified *E. coli* lysate was mixed with 500 ng of PAM library in 50 µl of reaction buffer (1xCutSmart buffer (New England Biolabs): 50 mM KOAc, 20 mM Tris-OAc, 10 mM $Mg(OAc)_2$, 100 µg/ml BSA, pH 7.9) and incubated at 37°C for 1 hour.. The end-repair, capture of library cleavage products and sequencing was performed as described in the methodology for the PAM determination of Cas9 orthologs.

2.2.16. Expression and purification of Cas12l proteins

Cas12l proteins were expressed in *E. coli* NiCo21(DE3) strain (New England Biolabs) from pET-MBP14xHisSUMO-Cas12l plasmids. Cells were grown in LB media at 30°C. After the cultures reached an OD_{600} of 0.5, expression was induced with 0.4 mM IPTG and incubation continued at 16°C overnight. Cells were pelleted by centrifugation and resuspended in loading buffer (20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 40 mM imidazole) and subjected to disruption by sonication. Cell debris was removed by centrifugation and supernatant filtered before loading onto a HiTrap DEAE FF chromatography column (GE Healthcare). The flowthrough was then loaded onto a Ni^{2+}-charged HisTrap column (GE Healthcare) and eluted with a linear gradient of increasing imidazole concentration (from 40 mM to 700 mM). The fractions containing Cas12l proteins were pooled, and the MBP-14xHis-SUMO tags were cleaved by increasing the NaCl concentration to 500 mM, adding 1 mM DTT and 2% (v/v) glycerol and 100 nM SenP1 protease and incubating at 4°C overnight. To remove the cleaved tag and SenP1 protease the reaction mixture was loaded onto a HiTrap Heparin column (GE Healthcare) and eluted with a linear gradient of increasing NaCl concentration (from 0.5 M to 2 M). Fractions containing Cas12l proteins were pooled and dialyzed against 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 50% glycerol storage buffer and stored at -20°C. Expression plasmids are listed in Appendix 2.

2.2.17. Cas12l-sgRNA complex assembly for *in vitro* cleavage

Cas12l-sgRNA RNP complexes were assembled by mixing purified Cas12l protein with sgRNA at 1:2 molar ratio followed by incubation in a complex

assembly buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM DTT, 1 mM EDTA) at room temperature for 20 min.

2.2.18. Cas12l DNA substrate generation

Plasmid DNA substrates were generated either by cloning oligoduplexes assembled after annealing complementary oligonucleotides containing PAM and protospacer sequences into pUC19 plasmid over EcoRI (New England Biolabs) and HindIII (New England Biolabs) restriction sites or by cloning PCR products containing PAM and protospacer sequences via blunt end ligation over end repaired EcoRI and HindIII restriction sites.
Fluorescently labeled DNA substrates were generated by annealing partially complementary oligonucleotides containing PAM and protospacer sequences and PCR amplifying them with primers containing 5'-6-FAM (non-target strand) or 5'-6-ROX (target strand) (IDT) dyes.

2.2.19. Cas12l DNA substrate cleavage assays

DNA cleavage reactions were initiated by mixing DNA substrates (Appendix 1) with Cas12l RNP complexes. Plasmid DNA cleavage reactions were carried out at 37°C in 1xCutSmart buffer (New England Biolabs) with a 1:20 substrate:complex molar ratio, unless stated otherwise. Aliquots were removed at timed intervals (30 min if not indicated differently) and mixed with 6x Blue Gel Loading Dye (New England Biolabs). Reaction products were analyzed by agarose gel electrophoresis and GelRed (Biotium) staining. Fractions of cleaved substrate were calculated by densitometric analysis with ImageJ software according to the following equation: $= Fraction\ cleaved, \% = (\frac{I_{products}}{(I_{products}+I_{substrate})}) \times 100$, where $I_{products}$ is the intensity of cleavage product bands and $I_{substrate}$ is the intensity of the intact substrate band.
Reactions with fluorescently labelled DNA substrates were carried out at 37°C or 47°C in 1xCutSmart buffer with a 1:10 substrate:complex molar ratio or 1:5 substrate:complex molar ratio (suboptimal PAM substrate cleavage screens) (20 mM substrate and 200 mM or 100 mM Cas12l complex respectively, unless stated otherwise). Aliquots were removed at timed intervals and mixed with 9.5 µl Hi-Di Formamide (Thermo Fisher Scientific) supplemented with 0.5 µl GeneScan LIZ 120 Size Standard (Thermo Fisher Scientific). Reaction products were subjected to capillary electrophoresis using a 3500 Series Genetic Analyzer (Thermo Fisher Scientific) as per manufacturers' recommendations. Electrophoresis data were analyzed with

OSIRIS (NCBI) or Geneious Prime (Biomatters Ltd.) software. Fragment size was evaluated by comparing the retention time with Internal Lane Standard retention time and the fractions of cleaved substrate calculated by comparing the substrate and cleavage product fragment peak area (minimum RFU (relative fluorescence units) for peak detection – 50):

$$cleavage\% = \left(\frac{peak\ area\ (products)}{peak\ area\ (substrate) + peak\ area\ (products)}\right) \times 100$$

Cleavage graphs were plotted using GraphPad Prism Software. Where applicable, the resulting data were fit to a single exponential association curve (GraphPad Software), according to the following equation: $Fraction\ cleaved = A \times (1 - e^{(-k \times t)})$, where A is the amplitude of the curve; k is the pseudo-first-order rate constant and t is time.

2.2.20. M13 ssDNA cleavage assays

M13 ssDNA cleavage reactions were initiated by mixing M13 ssDNA (New England Biolabs) and ssDNA activator (oligonucleotide) or dsDNA activator (oligonucleotide duplex) (Appendix 1) with Cas12l RNP complexes at 37°C. Cleavage reactions were carried out in 1xCutSmart buffer. Final reaction mixture contained 5 nM M13 ssDNA, 100 nM ssDNA/dsDNA activator, 100 nM Cas12l-sgRNA complex (unless stated otherwise). Reactions were stopped by mixing with 6x Blue Gel Loading Dye (New England Biolabs) and products analyzed by agarose gel electrophoresis and GelRed staining.

2.2.21. Fluorophore quencher-labelled reporter assays

Asp2Cas12l-sgRNA complexes were preassembled by incubating 500 nM of Asp2Cas12l with 600 nM sgRNA in 1x 2.1 buffer (NEB) at room temperature for 15 min. The complexes were then diluted and ssDNA/dsDNA activators (Appendix 1) added to a final concentration of 105 nM Asp2Cas12l : 125 nM sgRNA : 26 nM ssDNA/dsDNA activator and incubated at 50°C for 30 min in 1x 2.1 buffer (NEB). The trans cleavage reactions were initiated by adding 50 pmol of ssDNA/ssRNA fluorophore quencher (FQ) (Appendix 1) reporter substrates with the final concentration being 100 nM Asp2Cas12l : 120 nM sgRNA : 25 nM ssDNA/dsDNA activator : 500 nM FQ substrate in a 100 µl reaction volume in a black 96-well plate. The reactions were immediately placed in a fluorescence plate reader and incubated at 37°C with fluorescence measurements taken every 30 seconds.

For trans-cleavage rate determination, raw fluorescence values were corrected by subtracting fluorescence values obtained from reactions with FQ reporter only. The relative fluorescence values were then converted to FQ substrate

fraction cleaved according to the following equation: $Fraction\ cleaved = \frac{F(t)}{F(cleaved)} \times 100$, where $F(t)$ is the fluorescence at a given time point; $F(cleaved)$ is the fluorescence of fully cleaved FQ reporter. The resulting data were fit to a single exponential association curve (GraphPad Software), according to the following equation: $Fraction\ cleaved = A \times (1 - e^{(-k \times t)})$, where A is the amplitude of the curve; k is the pseudo-first-order rate constant and t is time.

For Michaelis-Menten analysis, 400 nM Asp2Cas12l : 500 nM sgRNA : 4 nM ssDNA/dsDNA activator complexes were assembled by first incubating Asp2Cas12l & sgRNA in 1x 2.1 buffer (NEB) at room temperature for 15 min and then adding the ssDNA/dsDNA activators and incubating at 50°C for 30 min. The trans-cleavage reactions were initiated by diluting the effective complexes to final concentrations of 10 nM Asp2Cas12l : 12.5 nM sgRNA : 0.1 nM activator in a solution containing 1x 2.1 buffer and 0.001, 0.01, 0.1, 0.2, 0.5, 1, or 2 µM fluorophore quencher-labelled substrate in a 100 µl reaction volume in a black 96-well plate. The reactions were incubated in a fluorescence plate reader at 37°C for up to 60 minutes with fluorescence measurements taken at every 30 seconds (λ_{ex}=485 nm; λ_{em}=538 nm). Reactions containing no Asp2Cas12l : sgRNA : activator complexes were measured to obtain uncleaved reporter fluorescence versus concentration standard curves and reactions containing a 100-fold higher concentration of Asp2Cas12l : sgRNA : activator effective complexes were measured to obtain fully cleaved reporter fluorescence versus concentration standard curves.

Relative fluorescence values obtained were converted to cleaved fluorescent substrate concentrations according to the following equation: $c_{cl} = \frac{F(t) - F(c_0)}{S_{cl} - S_{ucl}}$, where c_{cl} is the cleaved substrate concentration; $F(t)$ is the relative fluorescence value at a given time point; $F(c_0)$ is the fluorescence value of uncleaved reporter at a given concentration; S_{cl} is the slope of a linear standard curve obtained by plotting relative fluorescence values of fully cleaved substrate versus substrate concentration and fitting to a linear regression; S_{ucl} is the slope of a linear standard curve obtained by plotting relative fluorescence values of uncleaved substrate versus substrate concentration and fitting to a linear regression. Fully cleaved substrate was generated by incubating with Asp2Cas12l until the RFU values reached a stable peak.

Initial velocity (v_0) was calculated by fitting the cleaved substrate concentration versus time traces to a linear regression and plotted against the initial substrate concentration. Michaelis-Menten constants were determined according to the following equation: $Y = \frac{V_{max} \times X}{K_M + X}$, where X is the substrate

concentration and Y is the enzyme velocity (GraphPad Prism Software). The turnover number (k_{cat}) was determined by the following equation: $k_{cat} = \frac{V_{max}}{E}$, where E is the effective complex concentration (0.1 nM).

2.2.22. Plasmid interference assay

Plasmid interference experiments were carried out in *E. coli* Arctic Express (DE3) (Agilent Technologies) strain. First, pETDuet plasmid bearing the *Asp2Cas12l* gene was engineered to carry spacers found in native Cas12l loci by cloning a synthesized DNA fragment (containing repeat-spacer1-repeat-spacer2-repeat-spacer3-repeat sequence) over restriction sites. Another pETDuet plasmid containing the *cas12l* gene was constructed by adding a sgRNA coding sequence bearing a RunXI spacer with T7 promoter, HDV ribozyme and terminator sequences, using Gibson assembly molecular cloning method (New England Biolabs) (Appendix 3). Next, pCDF-Duet plasmids were engineered to carry protospacer sequences, complementary to the native spacers and RunXI spacer found in the pETDuet plasmids, downstream of corresponding PAM 5'-CCC-3' by cloning oligoduplexes over restriction sites in the pCDF-Duet plasmid (Appendix 3).

E. coli Arctic Express (DE3) (Agilent Technologies) cells were first transformed with pETDuet plasmids containing the *Asp2Cas12l* gene, gRNA and spacer sequences. These cells were grown at 37°C to an OD600 of 0.5 and electroporated with 200 ng of target pCDF-Duet plasmids containing protospacer sequences or an empty pCDF-Duet control plasmid. The co-transformed cells were grown at 37°C for 16-20 h on plates containing 100 g/ml carbenicillin, 10 g/ml streptomycin, 10 g/ml gentamycin and 0.1 mM IPTG. Plasmid interference was evaluated by the amount of transformant colonies observed on the plate.

3. RESULTS AND DISCUSSION

This work is two-fold, with the first part reporting on a broad biochemical screen of a large set of diverse type II CRISPR-Cas nucleases, leading to the discovery of enzymes recognizing novel PAM sequences, exhibiting varied temperature dependencies and DNA cleavage patterns, as well as bioinformatical analysis on the phylogeny of their PAM interacting domains and how they correlate with their PAM preferences. The second part reports on the identification and a detailed *in vitro* examination of a new family of type V CRISPR-Cas nucleases and their cleavage activities.

3.1. Characterization of type II CRISPR-Cas9 orthologs

While Cas9 can be programmed to target any DNA site by altering the spacer sequence of its guide RNA, in practice target selection is restricted by the recognition of the PAM (Collias & Beisel, 2021). This is further exasperated by the need of careful target selection to minimize potential off-target binding due to tolerance for mismatches in the gRNA-target heteroduplex (D. Kim et al., 2019). It is also key in applications where the editing outcome is highly reliant on the proximity of the desired change to the binding site, particularly homology-directed repair, base editing or prime editing (Anzalone et al., 2020). Moreover, the various applications of Cas9s may be restricted by their biochemical and physical features, such as predominantly blunt-end DNA cleavage (Jinek et al., 2012), gRNA exchangeability (Fonfara et al., 2014), temperature dependence (Wiktor et al., 2016) and size (Lino et al., 2018). The diversity provided by Cas9 orthologs found in nature may offer unique nucleases, as well as provide insight for improvements of the Cas9 toolbox.

3.1.1. Selection of Cas9 orthologs

To balance the diversity and biochemical properties of the putative Cas9 orthologs, several distinct strategies of *cas9* gene selection were employed. 47 orthologs were chosen from most of the 10 major clades of a Cas9 evolutionary tree (Figure 3.1A). Clades giving rise to previously characterized proteins that were active in eukaryotic cells were mined at a rate of approximately 20% while all others were surveyed at a rate of approximately 10%. To enrich for proteins with potentially robust biochemical activity and thermostability, an additional 32 orthologs were selected based on their predicted physiochemical properties, classification as a type II-A subtype, which encompasses the closest orthologs of thoroughly characterized

CRISPR-Cas systems (Fonfara et al., 2014; Makarova et al., 2020), and affiliation with a thermophilic host organism. This approach yielded a set of Cas9 orthologs ranging from ~1,000 to ~1,600 residues with a bimodal distribution focused around sizes of ~1100 and ~1375 amino acids (Figure 3.1B).

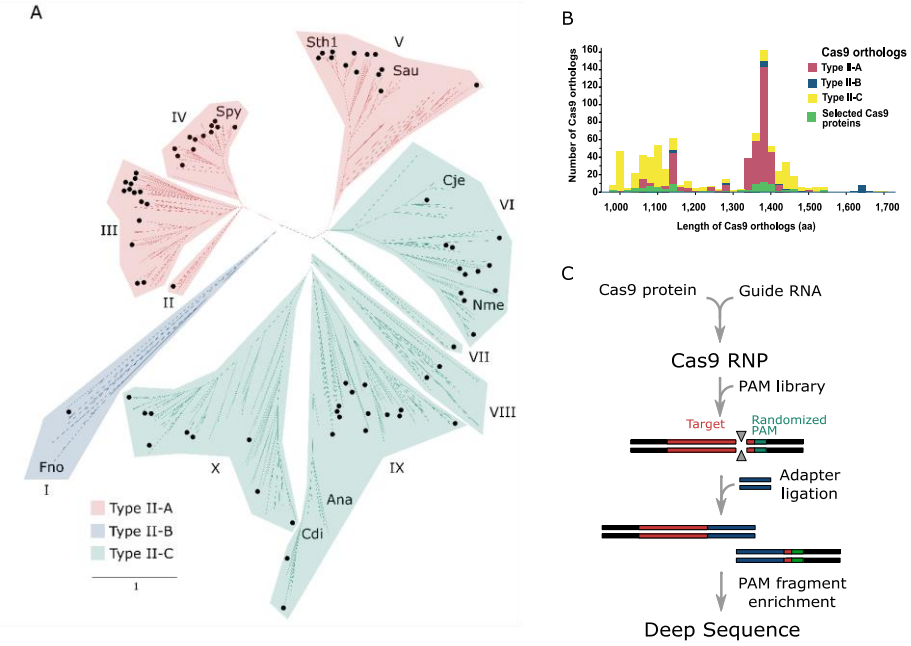


Figure 3.1. Cas9 diversity and PAM characterization approach. (A) Phylogenetic representation of the diversity provided by Cas9 orthologs. Type II-A, B, and C systems are color-coded, red, blue, and green, respectively. Distinct phylogenetic clades are numbered I–X. Those selected for the study are indicated with a black dot. Cas9s whose structure has been determined are also designated. **(B)** Cas9 protein size distribution according to subtype. Type II-A, B, and C systems are color-coded, red, blue, and yellow, respectively. The numbers of analysed Cas9s in each group are indicated in green. **(C)** Biochemical approach used to directly capture target cleavage and assess protospacer adjacent motif (PAM) recognition. Experiments were conducted using Cas9 protein produced by IVT.

3.1.2. Cas9 guide RNA determination and analysis

Prior to any biochemical testing of the selected Cas9 orthologs, the sequences of their tracrRNAs needed to be determined. This was done *in silico*, examining the Cas9 genomic loci for sequences complementary to the CRISPR array repeats and predicted conserved tracrRNA secondary structure motifs (Briner et al., 2014). The tracrRNAs were compared by computational analysis of covariant models (CMs) based on sequence and secondary

structure homology resulting in 7 distinct clusters (Figure 3.2). However, for some Cas9 orthologs, the tracrRNAs self-clustered or demonstrated weak similarity to other CMs and were not assigned to any particular group (Figure 3.2). In most of the cases, clusters were associated with a particular Cas9 phylogenetic clade (Figure 3.2).

For each ortholog, sgRNAs were designed by fusing the putative tracrRNAs to their respective crRNA repeats via a flexible linker. DNA templates for sgRNA transcription were generated and used in subsequent biochemical analysis.

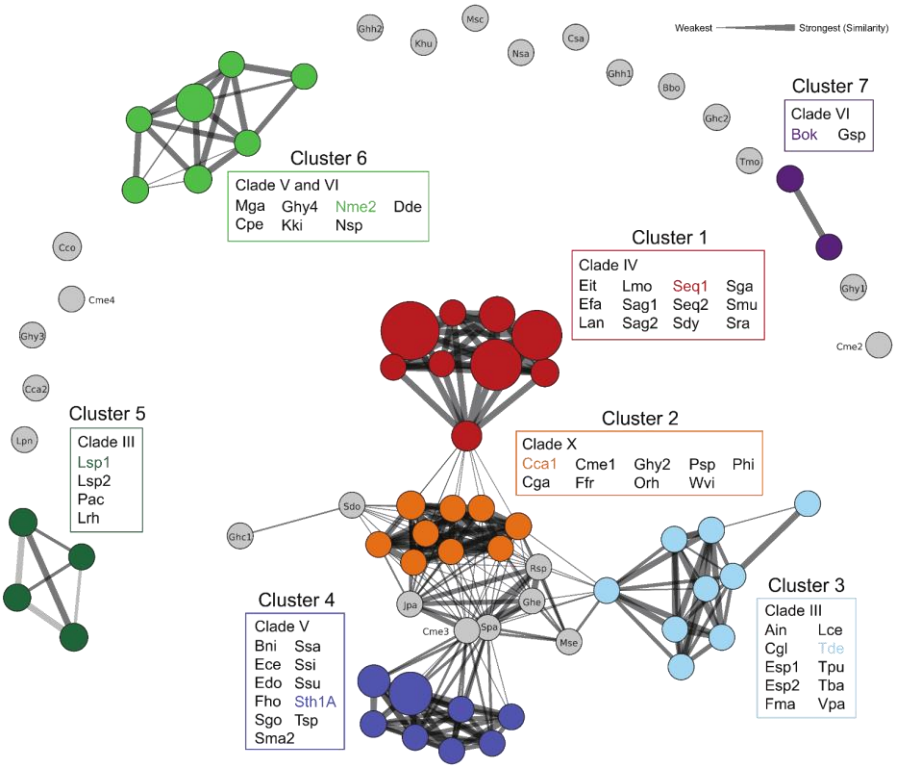


Figure 3.2. Cas9 tracrRNA sequence and secondary structure similarity. Circles are scaled based on the number of sequences belonging to each covariance model (CM) and colored according to the designated cluster. The width of the connecting lines indicates the percentage of similarity or relatedness among CMs. Representative tracrRNAs from each cluster are indicated with the associated color. CMs not assigned to a cluster are in gray.

3.1.3. Cas9 ortholog PAM requirements

To survey the target recognition properties of Cas9 orthologs in high throughput, we used a cell-free *in vitro* translation (IVT) method similar to

that described previously (Figure 3.1C) (Karvelis et al., 2018; Marshall et al., 2018). Crude IVT RNP mixtures were diluted (10^1 to 10^3 in 10-fold increments) to account for the difference in PAM recognition stringency dependent on the concentration of Cas9-guide RNA complex (Karvelis et al., 2015) and tested for their ability to facilitate cleavage when combined with a plasmid library containing a randomized PAM region adjacent to a Cas9 target site. The greatest dilution that still supported DNA cleavage activity was then used as a baseline for PAM recognition. Additionally, in cases where Cas9 orthologs exhibited PAM preferences at positions 6 or 7 and lacked PAM requirements in the first, second or third positions, the spacer targeting the PAM library was shifted to the 5' end by 1, 2 or 3 nts to examine potential PAM recognition beyond position 7, which is the length of randomized sequence in the PAM library. This permitted PAM identification to be extended to 8, 9, or 10 bp, respectively. Of the 20 orthologs that were surveyed for potentially longer PAM sequences, only 6, all belonging to phylogenetic clade VI (Figure 3.1A & 3.3), had PAM recognition that continued beyond the 7th position. Somewhat surprisingly, PAM preferences at the 8th position were always an A residue similar to the previously characterized Brevibacillus laterosporus (Blat) (Karvelis et al., 2015) and Geobacillus stearothermophilus (Geo) (Harrington et al., 2017) Cas9 proteins.

This approach resulted in the determination of PAM sequences for 79 Cas9 orthologs, exhibiting diverse PAM recognition motifs (Figure 3.3). This included nucleases with previously undescribed PAM requirements that varied in composition both in sequence and length. The PAM recognition of these Cas9 orthologs could be broadly sub-divided into A-, T-, and C-rich PAM recognition in addition to the G-rich PAM typical of the Spy Cas9 protein (Figure 3.3). Short PAMs (1-2 base pairs) were rare, with most orthologs exhibiting recognition at three or more positions. However, one case, Ghy4, recognized an exceptionally relaxed PAM sequence with only a single discrete G requirement at position 5, although with additional biases against particular bases at other positions. Orthologs recognizing PAMs composed of multiple residues of a single base pair (e.g. Efa, Nme2, Rsp, Ssi, and Ssu) were also much less common than Cas9s with composite PAM recognition containing at least two different base pairs. Furthermore, many proteins exhibited seemingly degenerate PAM recognition. Typically, this resulted in a strong requirement for at least one base pair in combination with positions that accepted more than one (typically two) base pairs (e.g. Lan, Mse, Nsa, and Sma2).

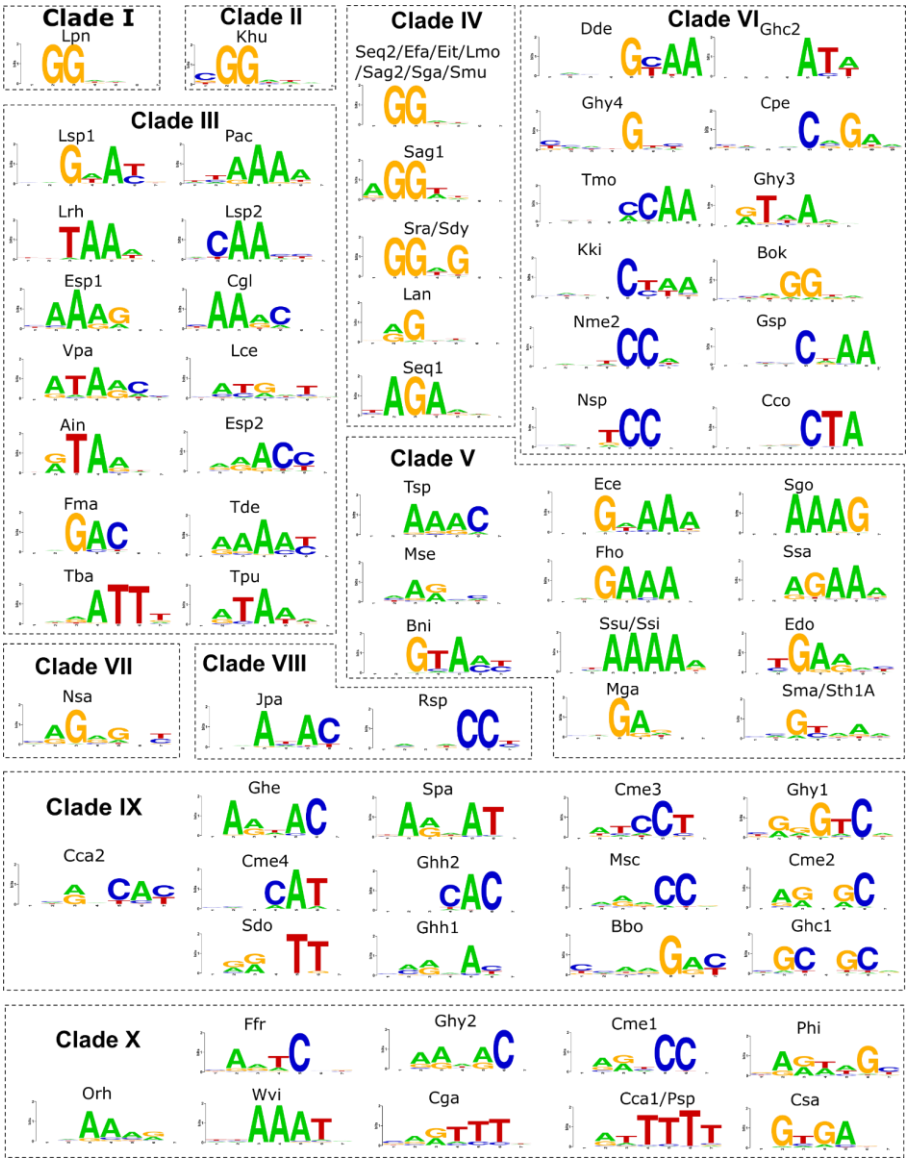


Figure 3.3. **PAM sequences recognized by Cas9 orthologs**. WebLogos of the PAM sequences recovered for each Cas9 ortholog, displaying the relative frequency of sequence recovery of a particular base in a given position 3‘ of the target. Cas9s were expressed by IVT and used to interrogate a randomized PAM plasmid library. Individual orthologs are separated by the phylogenetic clade they are attributed to according to Figure 3.1A.

3.1.4.Cas9 PAM interacting domain phylogeny

Next, the diversity of experimentally determined PAM preferences motivated us to evaluate the sequence relationship of Cas9 PAM interacting (PI) domains. The sequences of the PI regions from the characterized orthologs were extracted and used as queries for iterative searches against non-redundant collections of microbial proteins. Overall, 9,161 sequences having non-identical PI domains were discovered. Sequences of these domains were then clustered based on their pairwise similarity resulting in the identification of ten clusters (Figure 3.4). 93% of all sequences recovered were contained in clusters 1-4, while clusters 7-10 were considerably smaller and were comprised of 4 to 37 sequences. Broadly, PI domain sequence similarity could be correlated with the major phylogenetic branches of the Cas9 tree, with closely related clades grouped in the same cluster (Figure 3.1A and Figure 3.4).. Additionally, similar PI domains usually resulted in similar PAM recognition, however, in several instances PAM sequence diversity and length varied greatly even among members of the same group (Figure 3.3 and Figure 3.4)

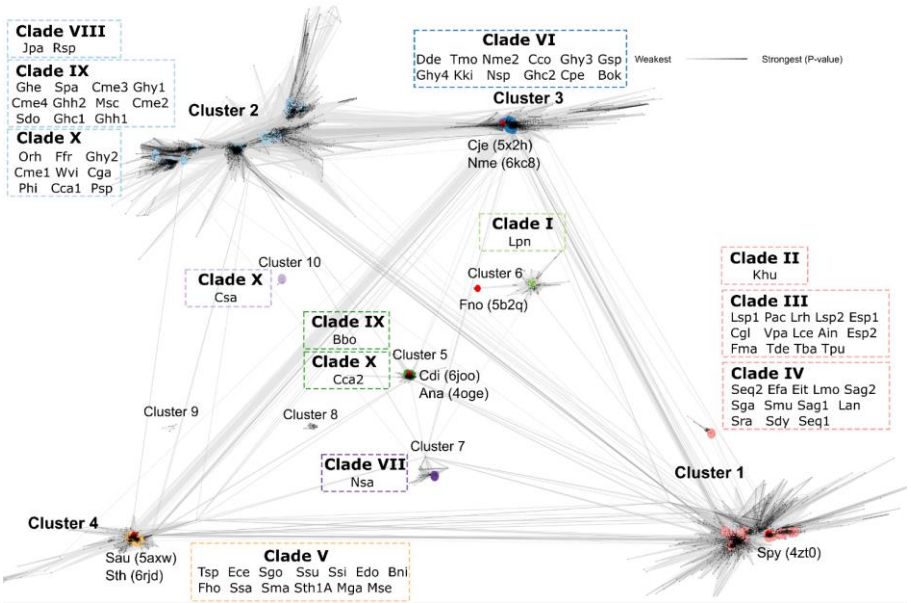


Figure 3.4. Cas9 PAM interacting (PI) domain similarity. Cas9 PI domains clustered by their pairwise sequence similarity. Sequences were clustered using CLANS (BLAST option). Lines connect sequences with P value $\leq 1e-11$. Line shading corresponds to P values according to the scale in the top-right corner (light and long lines connect distantly related sequences). The Cas9s which belong to the same clade are separated by dashed lines, colored according to the corresponding cluster. Sequences having known structures are marked red; their PDB code is shown in parentheses.

3.1.5. Cas9 biochemical activity

Fifty-two Cas9 orthologs from our collection were subjected for further biochemical characterization using purified components. The selection criteria for this subset of nucleases were based on short PAM recognition (≤3 bp) (where possible) while maintaining diversity in phylogenetic distribution and protein size.

Considering that it was previously reported that some Cas9 proteins require a guide RNA spacer longer than 20 nt (which is optimal for Spy Cas9) to function efficiently (Edraki et al., 2019; Harrington et al., 2017; E. Kim et al., 2017; Ran et al., 2015), we designed sgRNAs with two different spacer lengths, 20 and 24 nt, for each ortholog to initially gauge the influence of spacer length on Cas9 cleavage activity in vitro (Figure 3.5). While most orthologs performed similarly with both spacer lengths when evaluated across a panel of 5 different buffers, six orthologs, Cga, Cca1, Orh, Tmo, Nsa, and Ghh1 Cas9, required a spacer length of greater than 20 nt to effectively cut their DNA target (Figure 3.5, highlighted in red).

Next, we evaluated the thermal stability of thirty-eight orthologs showing efficient target cleavage using nano differential scanning fluorimetry (nanoDSF). 36 of 38 proteins showed a melting temperature of >37°C confirming stability under standard in vitro enzymatic reaction conditions. Interestingly, five orthologs had melting temperatures >50°C suggesting thermostability (Figure 3.6A). These included Cme2, Cme4, Ghy1, Esp1 and Nsa Cas9.

To corroborate nanoDSF predictions, DNA target cleavage was next measured in reactions at temperatures ranging from 10°C to 70°C. In all, Cas9 orthologs displayed a wide spectrum of temperature dependencies including both narrow and broad ranges of activity (Figure 3.6B). Consistent with thermal unfolding analysis, Cme2, Esp1, Nsa, Ain, Cme3, and Sth1A, were active at temperatures greater than 50°C with Nsa, isolated from the deep-sea hydrothermal vent chimney bacterium, Nitratifractor salsuginis, remaining active at temperatures greater than 60°C. Additionally, one ortholog, Ssa, retained 95% of its cleavage activity at 10°C. We also observed that 5 Cas9 orthologs (Cme2, Cme4, Nsp, Khu, and Fma) retained less than 25% activity at reaction temperatures of 25ºC or below.

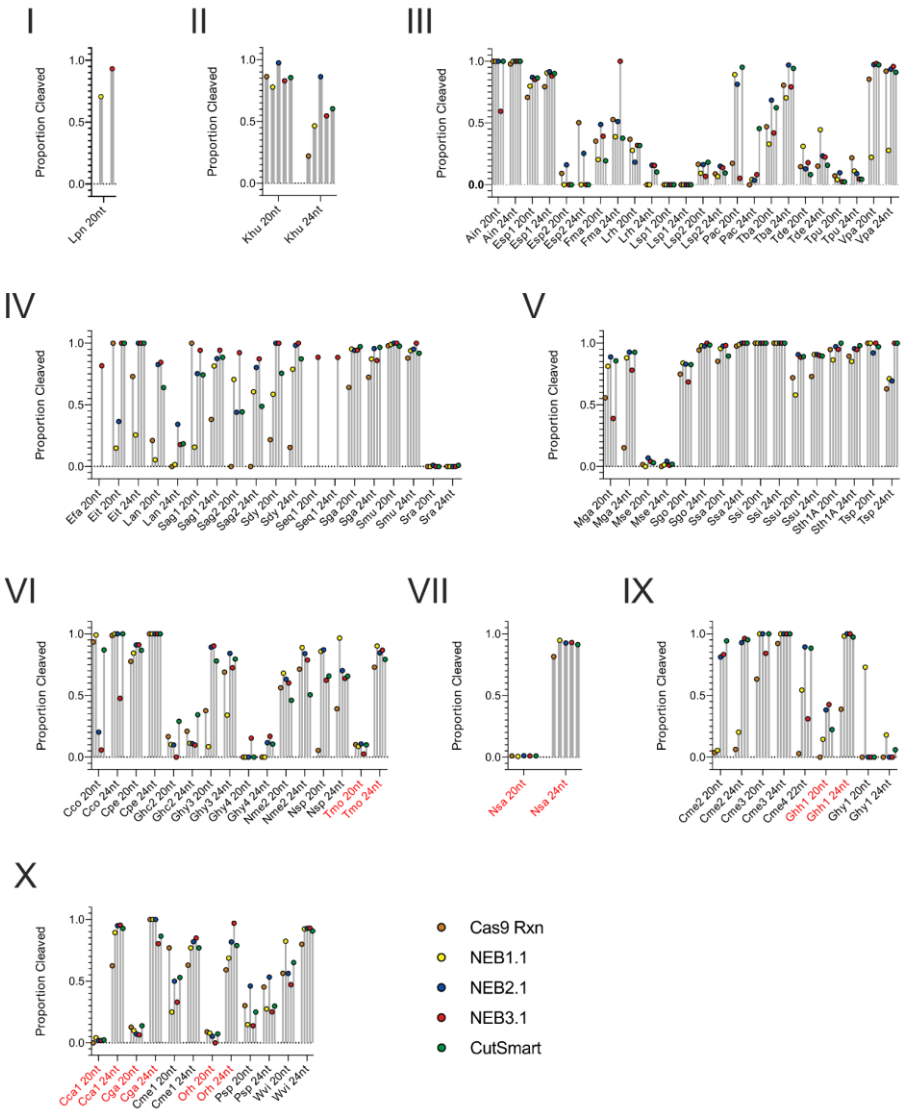


Figure 3.5. Spacer length preferences for purified Cas9 orthologs. The in vitro DNA cleavage activity of purified Cas9 proteins was screened using 20 and 24 nt spacers. Each point represents one cleavage determination in multiple buffers conditions as indicated by color. Brown, yellow, blue, red, and green circles represent reactions performed in Cas9 reaction buffer, NEBuffer 1.1, NEBuffer 2.1, NEBuffer 3.1, or CutSmart buffer, respectively. Clade membership is indicated by roman numerals. Orthologs exhibiting cleavage only with 24 nt spacers highlighted in red.

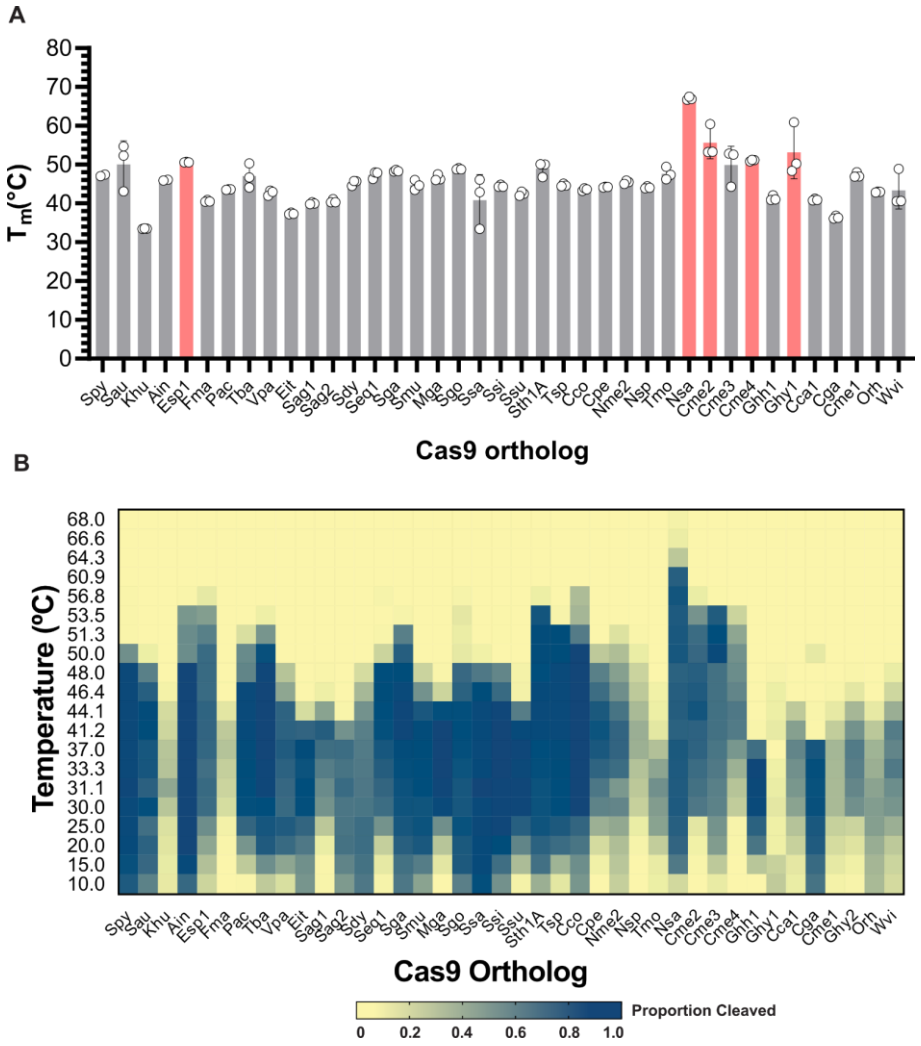


Figure 3.6. Cas9 thermal dependencies. **(A)** Melting temperatures of purified Cas9 proteins was determined by nanoDSF. Histograms of orthologs exhibiting melting temperatures above 50 °C marked in red. Open circles designate T_m values from individual experiments. Error bars indicate standard deviation of n = 3 independent measurements with the center of each denoting the mean of the determinations. **(B)** The cleavage activity of Cas9 orthologs was measured using in vitro DNA cleavage assays using fluorophore-labeled double-stranded DNA (dsDNA) substrates. Cleaved fragments were quantitated and are represented in a heatmap showing overall activity at temperatures ranging from 10 °C to 68 °C. The intensity of the blue color indicates the proportion of substrate cleaved.

To characterize the termini resulting from Cas9 DNA cleavage, we developed a method which allows both termini resulting from target cleavage to be captured simultaneously in a deep sequencing read. Some orthologs, as averaged across 5 different target sites, consistently generated overhanging

termini varying between 1 or more nts (e.g. Khu, Lpn, Nsa, and Esp1) (Figure 3.7). In these cases, only 5' staggered end cleavage products were recovered with the non-target strand tending to terminate at multiple positions suggesting variation in the positioning of or post-cleavage trimming by the RuvC domain while the target strand was cleaved predominantly between the 3rd and 4th positions of the protospacer (Figure 3.7).

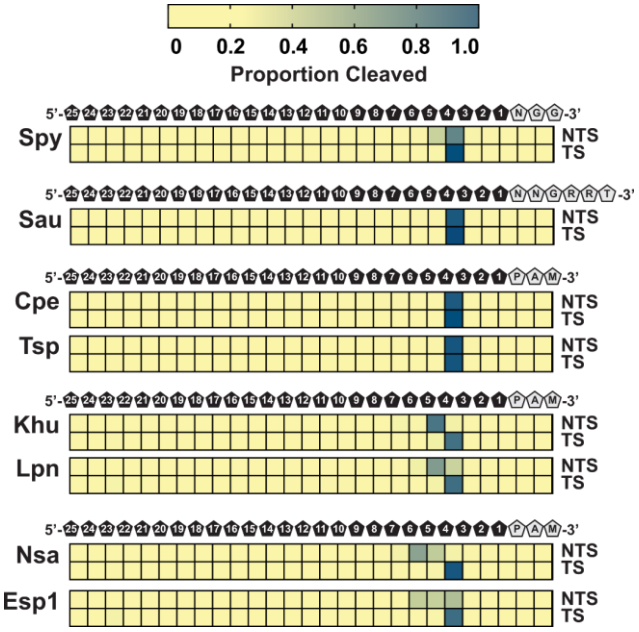


Figure 3.7. Target DNA cleavage patterns produced by Cas9 orthologs. Cleavage sites and resultant double-stranded DNA (dsDNA) ends are depicted as heatmaps that show the proportion of cleaved ends recovered by DNA sequencing at each position of a target DNA. The intensity of the blue color indicates the proportion of mapped cleavage ends. Examples of blunt, one base 5′-overhang staggered cleavage, and multiple base 5′-overhang cleavage show the proportion of cleaved ends as the averages at each position in five different dsDNA targets. The position of the DNA bases and PAM sequences is depicted above the heatmaps. NTS indicates a non-target strand; TS indicates the target strand.

3.2. Characterization of type V CRISPR-Cas nucleases

While the characterization screen of Cas9 nucleases yielded a set of enzymes exhibiting diverse biochemical features and their actual potential as bona fide genome editors is yet to be explored fully, they run into the challenge of being compared against the flagship *S. pyogenes* Cas9. In the meantime, the whole CRISPR-Cas field was seemingly shifting towards the significantly more diverse type V of CRISPR-Cas systems (Makarova et al., 2020). During the first half of this project, exciting studies describing the discovery and characterization of novel type V CRISPR-Cas systems encoding compact functional Cas nucleases were published (Yan 2019,Karvelis 2020,Pausch 2020). This encouraged us to probe the apparent natural diversity of the type V systems ourselves.

3.2.1. Identification of CRISPR-Cas12l systems

We set out by searching microbial sequence datasets for CRISPR associated nucleases that contain a single RuvC domain encoded in an operon-like organization with *cas1* and *cas2* genes. CRISPR locus gene architecture, structural inspection of the putative effector nuclease, and phylogenetic analysis were then used to define new systems. Using this methodology, a new family of type V CRISPR associated nuclease was discovered in species of *Armatimonadetes* whose sequence was captured in metagenomic studies aimed at identifying microbial communities involved in wastewater treatment (Kantor et al., 2017; Zhao et al., 2018). In each locus were genes that together encoded proteins required for adaptation (Cas1, Cas2, and Cas4) and a compact (~860 aa) effector nuclease adjacent to a CRISPR array (Figure 3.8A). The positioning and orientation of the putative effector and adaptation genes within the CRISPR locus (*cas nuclease*, *cas1*, *cas2*, *cas4* and CRISPR array) resembled that of type II-B CRISPR-Cas9 systems (Koonin & Makarova, 2019) (Figure 3.8A). Sequence examination of the putative nuclease confirmed the presence of a single tri-split RuvC-like domain located in the C-terminal half of the protein similar to other Cas12 proteins (Shmakov et al., 2017; Zetsche, Gootenberg, et al., 2015) (Figure 3.8A). In contrast, the sequence of the N-terminal half was highly divergent from other Cas12 nucleases. Despite this, it could be predicted to form an oligo binding domain (OBD) split by two helical regions containing a bridge-helix (BH)-like motif and a helix-turn-helix (HTH) DNA binding domain (Figure 3.8A). Phylogenetic analysis of the nuclease family confirmed their classification as a type V effector and showed that they form a new subgroup distinct from

previously described Cas12 proteins (Figure 3.8B). To simplify nomenclature and align with CRISPR nuclease naming conventions (Makarova et al., 2020), we proposed they be classified as Cas12l.

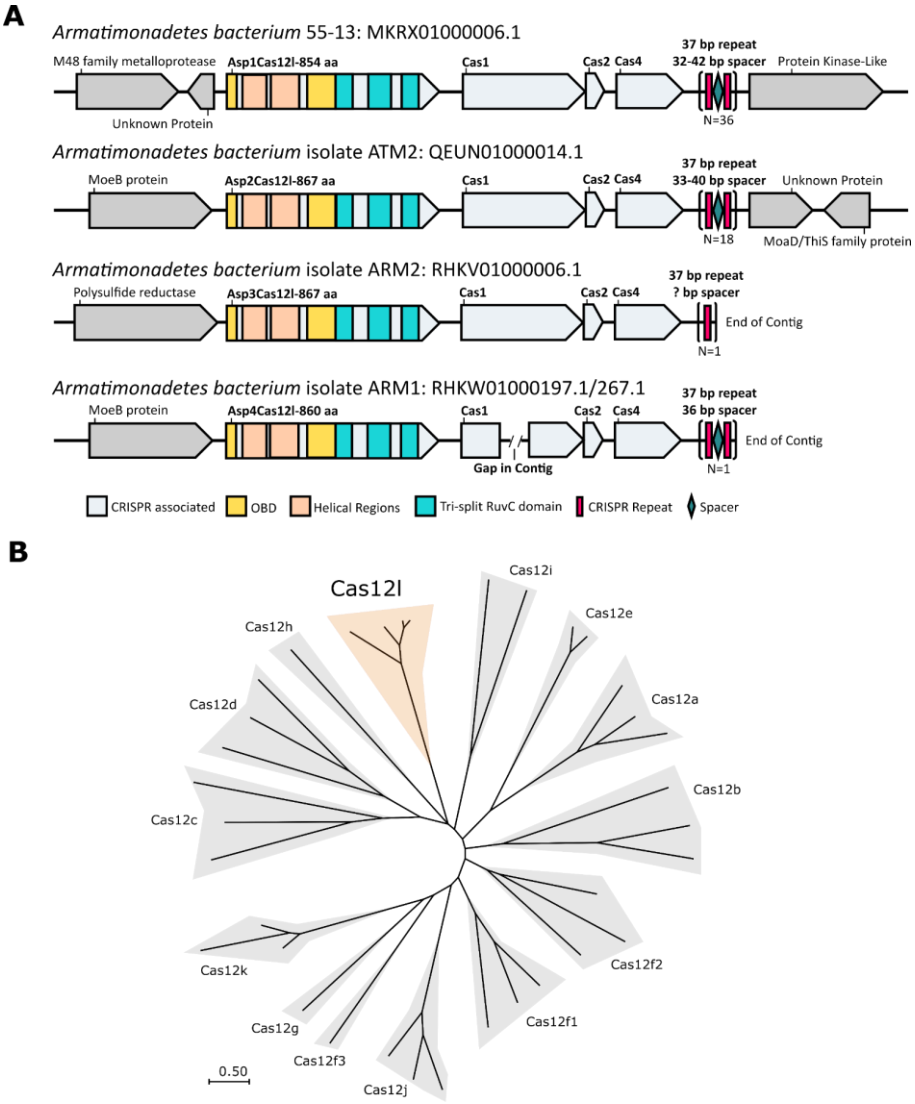


Figure 3.8. Cas12l is a new family of type V CRISPR associated nuclease. (A) A schematic of Cas12l locus architecture. CRISPR-associated genes are shown in light blue and each locus identified encodes one approximately 860 aa effector containing, an oligo-binding domain (OBD, yellow) a single tri-split RuvC nuclease domain and a helix-turn-helix motif (orange) followed by Cas1, Cas2, and Cas4. **(B)** Maximum likelihood phylogenetic tree illustrating the sequence relationship between Cas12l and other type V CRISPR-Cas proteins.

3.2.2. Cas12l PAM determination

To determine whether Cas12l systems were capable of dsDNA hydrolysis, lysate from *E. coli* expressing Cas12l proteins and putative guide RNAs were used to interrogate a 7N randomized PAM library similar to that described earlier (Karvelis et al., 2020) (Figure 3.9A). This was accomplished by first modifying Cas12l CRISPR arrays to encode spacer sequences capable of targeting either side of the PAM library, to account for both 5‘ and 3‘ possible PAM orientations (Figure 3.9A). Spacer altered Cas12l CRISPR systems were then synthesized and cloned into IPTG inducible expression plasmids and transformed into *E. coli*. After inducing expression, cells were disrupted, and the clarified lysate combined with the PAM library. Cleavage products were then captured by dsDNA adapter ligation, enriched for by PCR, and subjected to Illumina deep sequencing (Figure 3.9A). DNA cleavage was then detected by examining the frequency of adapter ligation at each position of the protospacer targets.

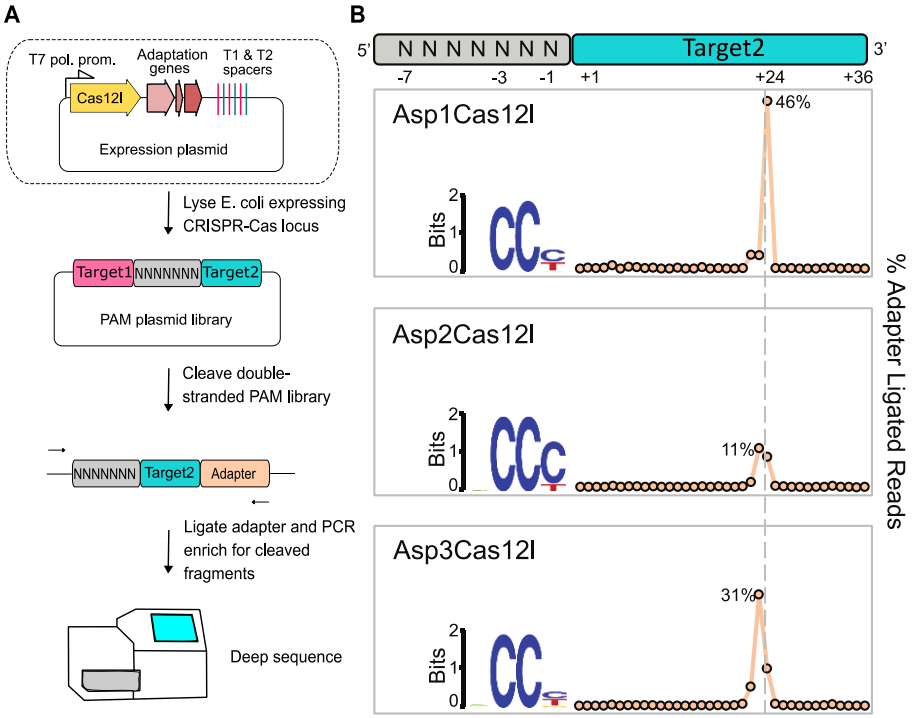


Figure 3.9. Cas12l nucleases cleave dsDNA in the presence of a 5‘ C-rich PAM. **(A)** Workflow used to detect dsDNA cleavage and associated PAM recognition of Cas12l CRISPR systems. **(B)** Weblogos of the PAM sequences that supported target recognition and cleavage as well as the frequency of adapter ligated reads are shown.

Sequence reads associated with spikes in the frequency of adapter ligation were then examined for biases in the PAM library and, if identified, used as further evidence of target cleavage (Figure 3.9B). Asp1, Asp2, and Asp3Cas12l all exhibited cleavage peaks after the 23rd and 24th positions of the T2 target region and corresponding library fragments showed a preference for a 5' C-rich motif (Figure 3.9B).

3.2.3. Cas12l guide RNA identity

The guide RNA responsible for the observed dsDNA cleavage by Cas12l effectors was next determined. The non-coding region between the nuclease and *cas1* genes was first explored *in silico* for the presence of a tracrRNA. Here, a 12-13 bp region with complementation to the CRISPR repeat, an anti-repeat, was identified for Asp1, Asp2, Asp3 and Asp4Cas12l (Figure 3.10A, orange). Sequence alignments and secondary structures predictions were then used to define regions of similarity. At the sequence level, Asp1, Asp2, Asp3 and Asp4Cas12l were observed to contain a 17 bp region of 100% similarity immediately 5' of the anti-repeat (Figure 3.10A, blue). Secondary structure analysis showed this region, when transcribed as an RNA, to form a stem-loop structure reminiscent of a nexus-like hairpin (Figure 3.10A and B) observed in the guide RNA of other Cas9 and Cas12 systems (Briner et al., 2014; Dooley et al., 2021; Faure et al., 2018) (Figure 3.10A and B). Analysis of other RNA structures revealed additional conservation among all four systems. This included Hairpin 1 and 3 located 5' of the nexus-like stem loop and 3' of the anti-repeat, respectively (Figure 3.10A and B, gray and red).

To confirm these sequences as bona fide tracrRNAs, sgRNAs were designed by linking the 3' end of the putative tracrRNAs with the 5' end of the CRISPR repeat with a tetraloop, 5'-GAAA-3'. Then using purified Asp2 and Asp3Cas12l nucleases, tested *in vitro* for the ability to guide dsDNA cleavage. Based on the respective putative tracrRNA (Figure 3.10A), four possible sgRNA design variants were generated for each effector protein. The sgRNAs differed in length at the 5' end of the tracrRNA, which was predicted to be unstructured, and by the presence or omission of the terminator-like hairpin (Figure 3.10B). Only the sgRNA variants bearing the 5' untruncated portion of the tracrRNA permitted dsDNA cleavage and inclusion of Hairpin 3 reduced cleavage efficiency (Figure 3.10C).

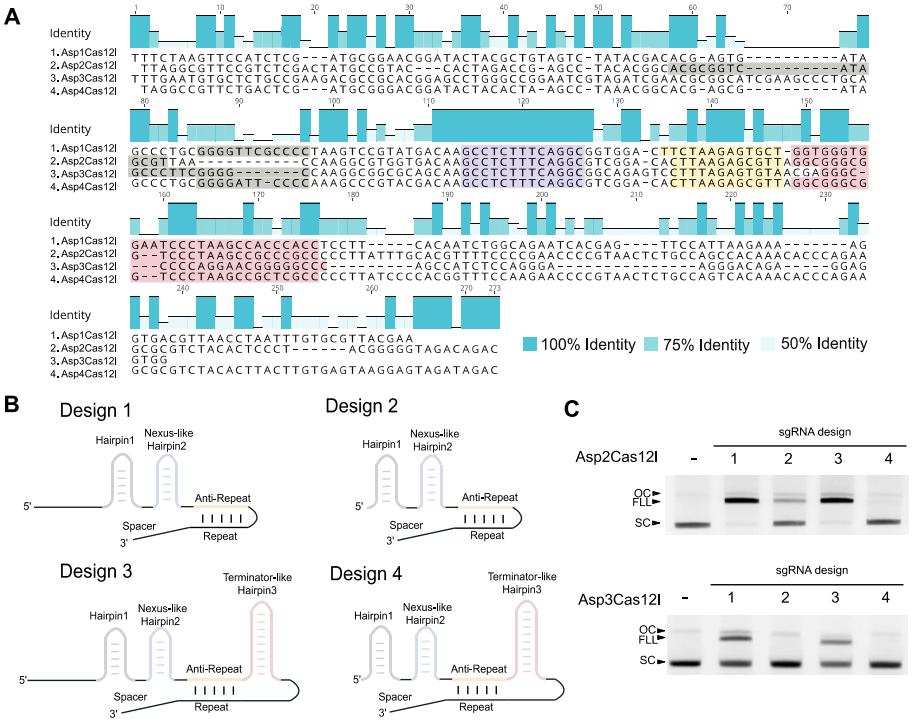


Figure 3.10. Identification and confirmation of Cas12l guide RNA(s). **(A)** Alignment of the sequences between the genes encoding the Cas12l effector and Cas1 exhibit features of a trans-activating CRISPR RNA (tracrRNA). Percent identity is shown in teal. These include a sequence encoding a 5' hairpin (Hairpin 1 – gray), a conserved sequence encoding a nexus-like stemloop (Hairpin 2 – blue), a region capable of base-pairing with the CRISPR repeat (Anti-repeat – orange) and a GC-rich sequence capable of forming a terminator-like hairpin (Hairpin 3 – red). **(B)** Four single guide RNA (sgRNA) designs were engineered for Asp2Cas12l and Asp3Cas12l based on the secondary structure prediction of tracrRNAs in (A). They differed by the presence or omission of a 5' region of the tracrRNA predicted to be unstructured and by the inclusion or exclusion of the terminator-like Hairpin 3. tracrRNA features are colored as described in (A). **(C)** Cleavage of supercoiled (SC) plasmid DNA substrates using purified Cas12l nuclease and respective sgRNA variants. Efficient full length linearization (FLL) of the substrate resulting from a complete double-strand break was only observed when using sgRNAs bearing the 5' most end of the tracrRNA (designs 1 and 3). Moreover, the terminator-like Hairpin 3 is not required for target cleavage (design 1). OC - open circular; FLL - full length linear; SC-supercoiled.

3.2.4. Biochemical characterization of Cas12l dsDNA cleavage activity

The positions of target strand (TS) and non-target strand (NTS) dsDNA cleavage were evaluated next. Purified Asp2 and Asp3Cas12l

ribonucleoproteins (RNPs) were used to digest the 7N plasmid library and cleavage products characterized by adapter ligation, PCR enrichment, and deep sequencing. The position of cleavage within the TS was as observed using *E. coli* lysate above, although most cleavage occurred immediately after the 24th nt downstream of the PAM, and NTS cleavage ensued just after positions 15-18 nts 3' of the PAM (Figure 3.11A).

The RuvC-like motif identified in Cas12l effectors was next confirmed to be responsible for the observed nuclease activity. This was accomplished by substituting alanine residues in place of the key catalytic D-E-D triad within the nuclease-fold for Asp2 and Asp3Cas12l effectors. Individually, each substitution abolished dsDNA cleavage (Figure 3.11B).

To confirm the findings of the PAM determination assay using E. coli lysates, Asp2 and Asp3Cas12l PAM recognition was interrogated further using purified protein and sgRNA components and substrates with fixed non-randomized PAM sequences. Here, the replacement of C at positions -2 and -3 completely abolished dsDNA target cleavage for both Asp2 and Asp3Cas12l effectors. At -1, Asp2Cas12l strongly preferred a C and tolerated a T (Figure 3.11C). In contrast, Asp3Cas12l was observed to be less stringent accepting C, T, and G and showed weak activity with an A at this position (Figure 3.11C).

Next, the effect of temperature on dsDNA target cleavage was evaluated. For these experiments, linear dsDNA fragments were used as substrates. As shown in Figure 3.12A, both Asp2 and Asp3Cas12l RNP complexes functioned optimally at elevated temperatures, around 50°C, with Asp2Cas12l demonstrating activity under a wider range of temperatures. After 30 min., it was able to cleave over 40% of its substrate at all temperatures except 27°C while Asp3Cas12l activity dropped-off precipitously at temperatures below 42°C. At temperatures above 52°C, activity diminished for both enzymes.

Following this, variation in Cas12l dsDNA target cleavage efficiency was assessed. For this, twelve targets were selected from two therapeutically relevant human genes, WTAP and RunXI, DNA regions PCR-amplified, and cleavage efficiency evaluated *in vitro* at 37°C using Asp2 and Asp3Cas12l nucleases. Interestingly, many of the targets were not efficiently cleaved (Figure 3.12B).

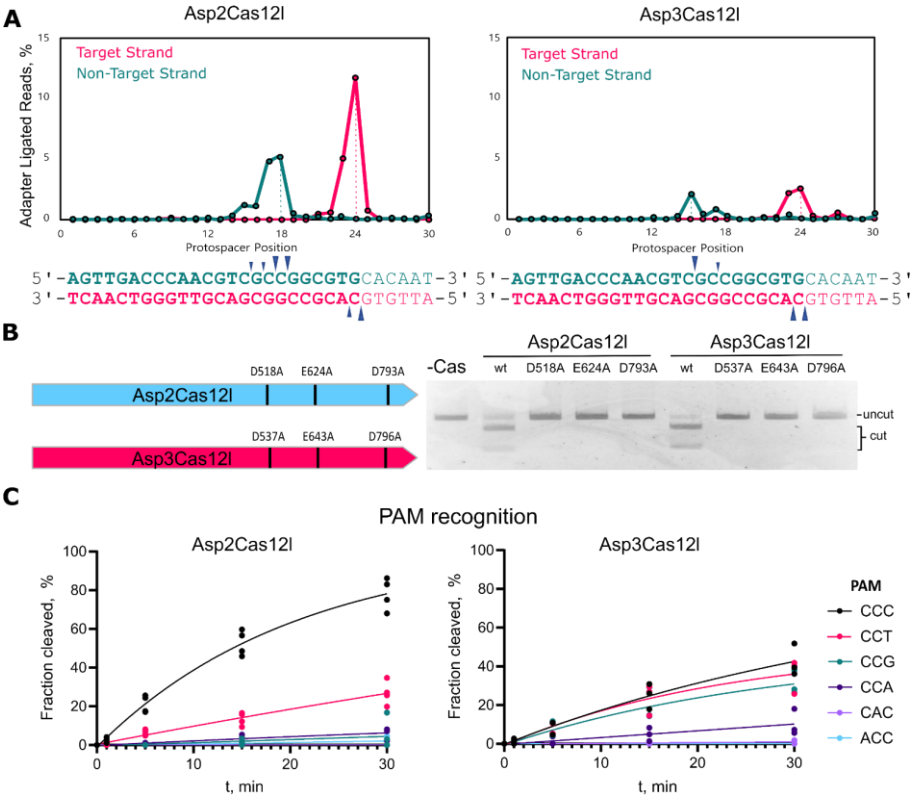


Figure 3.11. Biochemical characterization of Cas12l dsDNA cleavage. **(A)** Positions of Asp2Cas12l and Asp3Cas12l target and non-target strand cleavage. Frequency of adater ligated reads from PAM library cleavage experiments for both target and non-target strands. The target strand is cleaved 23-24 nt 3' of the PAM and the non-target strand is cleaved 15-18 nt downstream of PAM recognition. **(B)** Substitution of alanine residues disrupts linear dsDNA cleavage confirming key catalytic positions within the RuvC nuclease domain of Asp2Cas12l and Asp3Cas12l. wt – wildtype. **(C)** Cleavage of oligoduplex dsDNA substrates with purified RNP complexes confirm PAM recognition. The molar ratio of RNP to substrate was kept low (5:1) to increase reaction stringency. Asp2Cas12l predominantly recognizes a 5'-CCY-3' PAM and Asp3Cas12l a 5'-CCB-3' PAM. In **(C)**, individual data points are plotted, where n = 4 replicates from independent experiments (Asp2Cas12l) and n = 3 replicates from independent experiments (Asp3Cas12l). The data points were fitted to a single exponential association curve (solid lines).

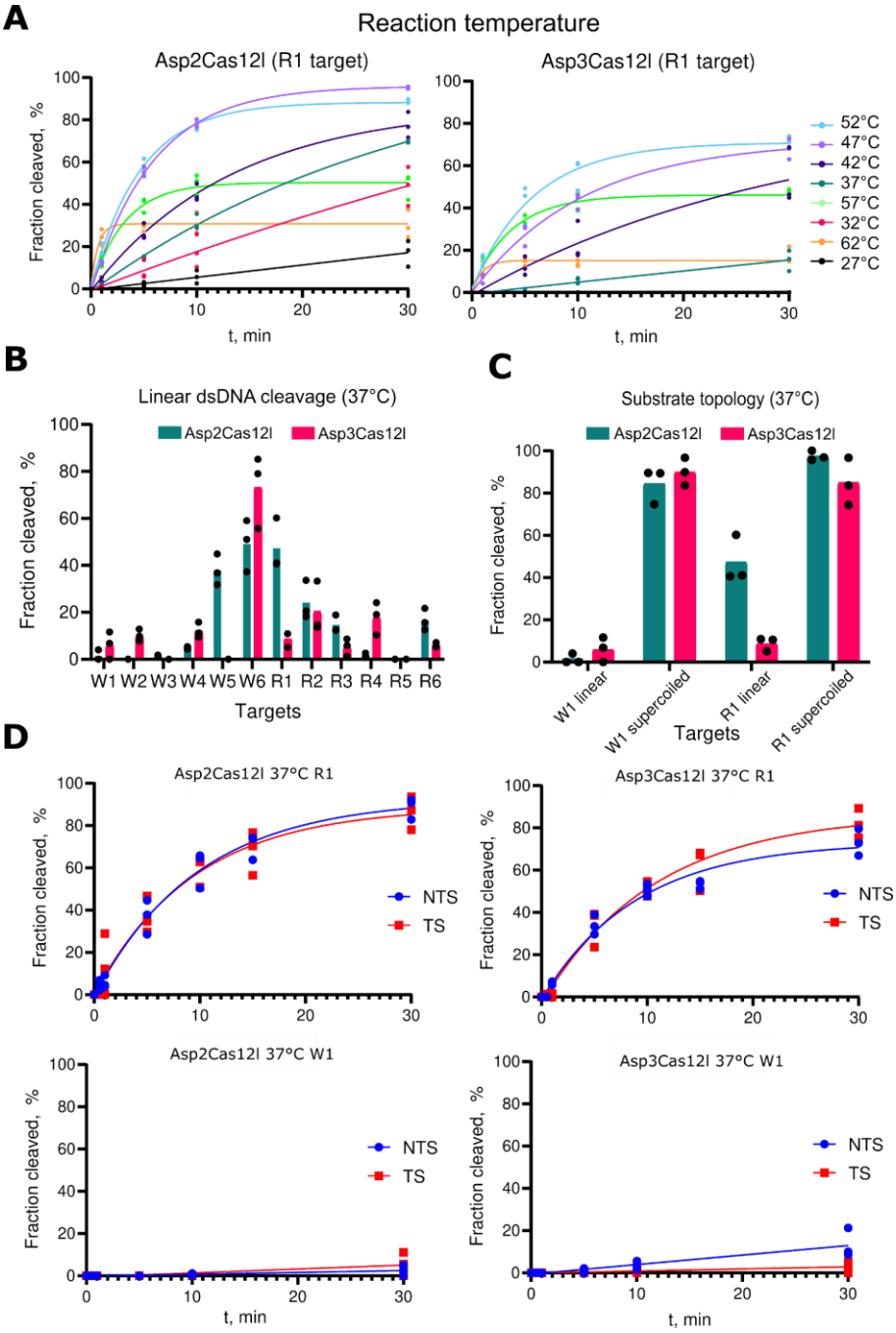


Figure 3.12. Cas12l dsDNA cleavage dependencies on temperature and substrate topology. (A) Effect of reaction temperature on Asp2Cas12l and Asp3Cas12l dsDNA hydrolysis. Optimal temperature for dsDNA cleavage by Asp2Cas12l and Asp3Cas12l RNP complexes is ~50°C. **(B)** Cas12l dsDNA hydrolysis efficiency varies depending on the target sequence. **(C)** Linear and supercoiled dsDNA substrates with different protospacer sequences (W1 & R1) were interrogated by

Asp2Cas12l and Asp3Cas12l RNPs. With a linear topology, only the R1 protospacer target was appreciably cleaved, however, both protospacers were cleaved with similar efficiencies by both proteins when presented in a supercoiled state. **(D)** Asp2Cas12l & Asp3Cas12l cleave both dsDNA strands at similar rates in vitro. Fluorescently labelled linear oligoduplex dsDNA substrates were used for hydrolysis experiments & cleavage rates of non-target (NTS) and target (TS) DNA strands are plotted. In (**B** and **C**), data are presented as mean with individual data points plotted, where n = 3 replicates from independent experiments. In (**A** and **D**), individual data points from n = 3 replicates from independent experiments are plotted and fitted to a single exponential association cuve (solid lines).

Since most of the previous biochemical experimentation used plasmid DNA targets, we reasoned that Cas12l nucleases may prefer supercoiled substrates. To test this, the efficiency of dsDNA target cleavage at the R1 and W1 sites were assayed in both linear and supercoiled states. Asp2 and Asp3Cas12l rapidly cleaved both sites when presented in a supercoiled form (Figure 3.12C). In contrast, only the R1 target was partially cleaved by Asp2Cas12l when linear substrates were used (Figure 3.12C).

Since dsDNA nicking products due to incomplete target cleavage cannot be resolved with linear dsDNA substrate, we aimed to check for the possibility of Cas12l enzymes predominately targeting a particular dsDNA strand. Fluorescently labelled oligoduplexes containing R1 and W1 targets were interrogated with Asp2 and Asp3Cas12l RNP complexes to assay rates of NTS and TS cleavage. As shown in Figure 3.12D, Asp2 and Asp3Cas12l RNPs cleaved NTS and TS at similar rates.

3.2.5. Cas12l collateral nucleic acid cleavage

A feature shared by most type V Cas effectors is the non-specific collateral degradation of ssDNA after DNA target recognition (J. S. Chen, Ma, et al., 2018; Yan et al., 2019). To determine if Cas12l nucleases have this attribute, Asp2 and Asp3Cas12l RNPs were incubated in the presence or absence of a ssDNA or dsDNA target substrate and bacteriophage M13 ssDNA. As shown in Figure 3.13A, both target substrates triggered the rapid loss of the M13 ssDNA resulting in its near complete degradation after 30 min. Also, when both RNPs were incubated with a ss or dsDNA molecule with no sequence complementary to the guide RNA, a non-specific (NS) activator, M13 ssDNA remained intact confirming that the ssDNase-like activity is only stimulated by the presence of a target DNA (Figure 3.13A).

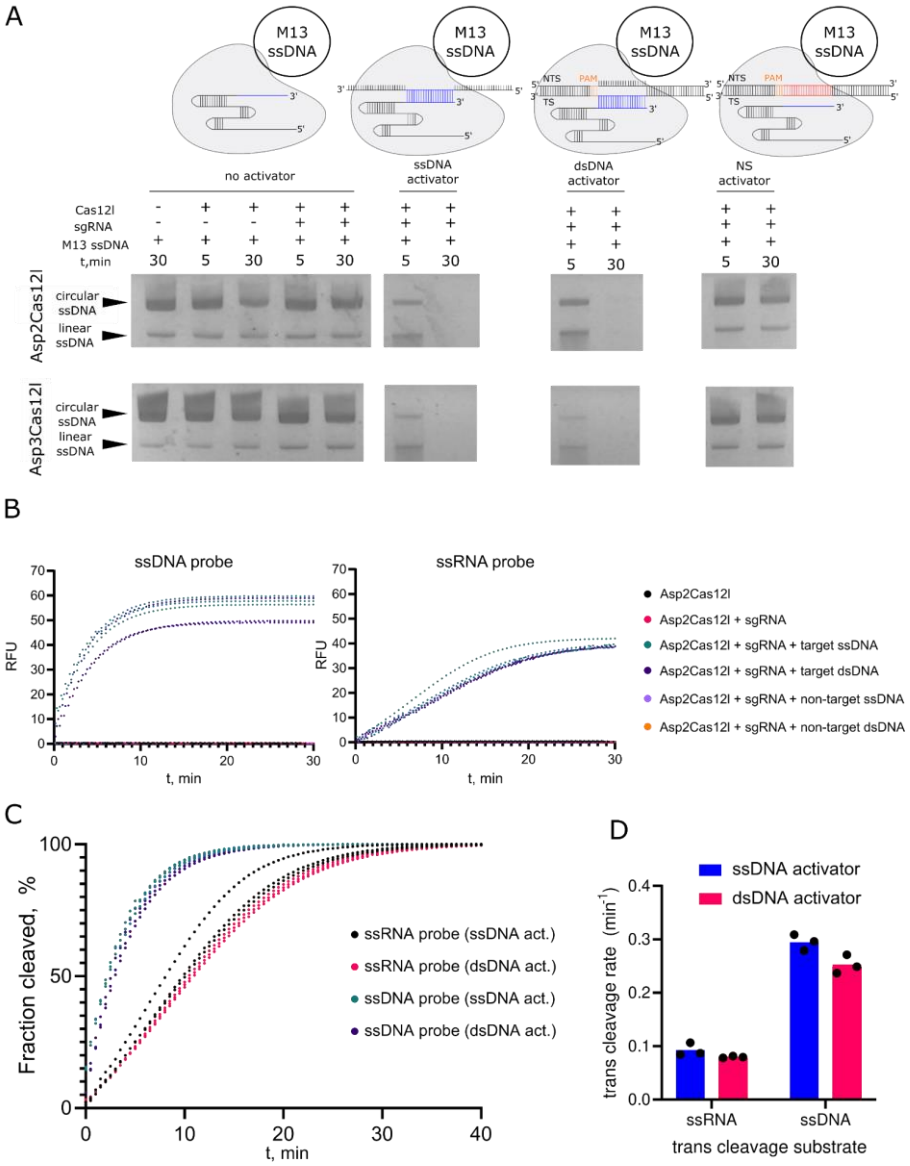
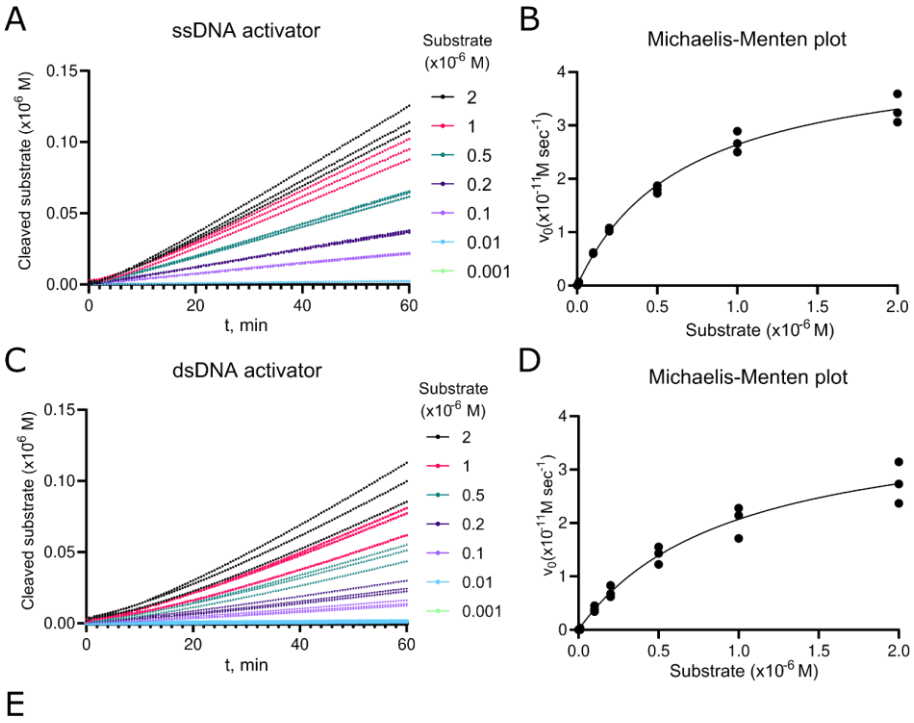


Figure 3.13. Cas12l collateral nuclease activity. (A) Asp2 and Asp3Cas12l RNP complexes degrade M13 ssDNA in the presence of single-stranded (ss) DNA or dsDNA (including PAM) activators with target sequences complementary to the gRNA spacer. NS activator – non specific activator (ssDNA oligonucleotide or dsDNA duplex) with no sequence complementarity to the spacer of the gRNA. **(B)** Asp2Cas12l RNP complexes activated with ss or dsDNA degrade quenched fluorescent ssDNA or ssRNA probes. Background-subtracted traces from fluorescent reporter assays with favored ssDNA (5‘-CCCCCCCC-3‘) or ssRNA (5‘-CCCCCCCC-3‘) probes. RFU – relative fluorescence units. **(C)** Background-subtracted traces of fluorescent reporter cleavage using ssDNA or ssRNA probes. **(D)** Rates of trans-degradation with ssRNA or ssDNA reporters activated with either

ssDNA or dsDNA targets. Collateral ssDNAse activity is about 3-fold higher than the rate of ssRNA degradation. Fluorescence intensities were normalized against the fluorescence of a reaction containing only the probe to account for imperfect quenching or degradation of reporters.

ssDNA sequences separated by a quencher moiety and fluorophore were next used to measure rates of *trans*-degradation for Asp2Cas12l as described earlier using Cas12a (J. S. Chen, Ma, et al., 2018). Collateral nuclease activity, using ssDNA and ssRNA probes, was confirmed to be triggered only in the presence of a ssDNA or dsDNA target and the rate of ssDNA and ssRNA degradation calculated (Figure 3.13B-D). Altogether, it was found that ssRNA was non-specifically cleaved about 3 times slower than ssDNA similar to that observed earlier with Cas12a (Fuchs et al., 2022).

The kinetics of ssDNA collateral cleavage by Asp2Cas12l were next examined. For this, the rate of ssDNA collateral degradation was measured using a fixed concentration of Asp2Cas12l-sgRNA RNP complex, 0.1 nM, activated with either a ssDNA or dsDNA target at different ssDNA reporter concentrations, 0.001 x 10^{-6} to 2 x 10^{-6} M, at 37°C. Collateral nuclease activity was measured by fluorescence continuously for 60 min. Raw fluorescence values were converted to cleaved substrate concentrations using standard curves based on data from experiments assembled without RNP complex and ones allowed to proceed to completion. Michaelis-Menten plots were then fitted to the data and rate of Asp2Cas12l collateral nuclease activity under substrate saturating (k_{cat}) and limiting conditions (k_{cat}/K_M) calculated as well as the reporter concentration that provided half maximal velocity (K_M). Altogether, it was shown to non-specifically degrade ssDNA at a rate of 0.44 and 0.41 molecules per second with a k_{cat}/K_m catalytic efficiency of ˜6.5×10^5 or 4.2×10^5 s^{-1} M^{-1} when using a ssDNA or dsDNA activator, respectively (Figure 3.14).



E

Activator	k_{cat} (sec^{-1})	K_M (M)	k_{cat}/K_M (sec^{-1} M^{-1})
ssDNA	0.44 ± 0.03	0.67 ± 0.11 × 10^{-6}	6.54 ± 2.72 × 10^5
dsDNA	0.40 ± 0.06	0.95 ± 0.29 × 10^{-6}	4.21 ± 2.06 × 10^5

Figure 3.14. Michaelis-Menten analysis of Asp2Cas12l collateral ssDNA cleavage activity. (A) Background-subtracted traces and corresponding linear trendlines of cleaved substrate concentration versus time for a ssDNA activator, using 0.1 nM effective Asp2Cas12l-sgRNA-activator complex and increasing ssDNA reporter concentration. **(B)** Michaelis-Menten fits for the ssDNA activator. **(C)** Background-subtracted traces and corresponding linear trendlines of cleaved substrate concentration versus time for a dsDNA activator, using 0.1 nM effective Asp2Cas12l-sgRNA-activator complex and increasing ssDNA reporter concentration. **(D)** Michaelis-Menten fits for the dsDNA activator. **(E)** Calculated kinetic constant values. Data are presented as mean ± s.d., where n = 3 replicates from independent experiments.

3.2.6. Asp2Cas12l activity in *E. coli*

Since Asp2Cas12l was more active at 37°C than Asp3Cas12l (Figure 3.12A), it was next tested for its ability to interfere with DNA plasmid transformation in *E. coli*. Asp2Cas12l effector and guide RNA were expressed from a plasmid containing an ampicillin resistance marker (AmpR) in two configurations. The first contained the Asp2Cas12l CRISPR locus while the second encoded only the *Asp2Cas12l* nuclease gene and sgRNA (Figure 3.15).

In both cases, guide RNAs encoded in the plasmid (either CRISPR array or sgRNA) were engineered to target a second plasmid containing a streptomycin resistance (aadA) gene (Figure 3.15). *E. coli* cells were then co-transformed with the respective AmpR and aadA plasmids and plated on media containing ampicillin and streptomycin. Only in the case of a control plasmid, lacking a target sequence, were any colonies observed (Figure 3.15).

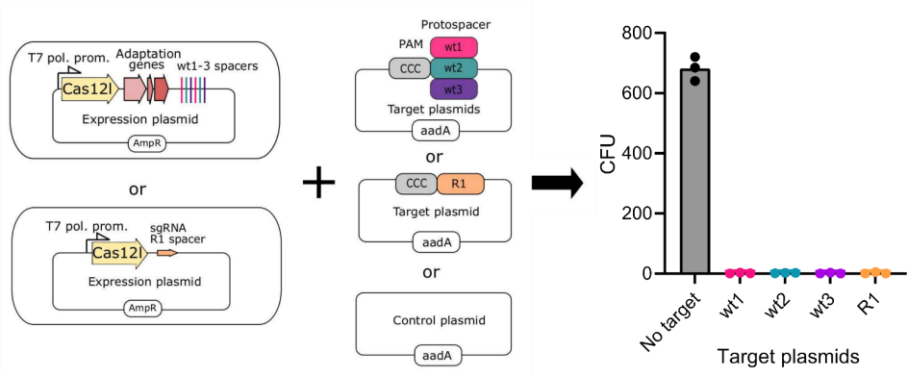


Figure 3.15. Asp2Cas12l functions in a heterologous cell to protect against invading dsDNA. *E. coli* were co-transformed with plasmids encoding an inducible Asp2Cas12l gene and guide RNA (expression plasmid) and a second plasmid containing protospacer target adjacent to a suitable PAM for Asp2Cas12l (5'-CCC-3') (target plasmid) or a control plasmid with no PAM or protospacer target sequence. The transformants were plated on medium containing T7 expression inductor (IPTG) and appropriate antibiotics (carbenicillin & streptomycin). Colonies were only recovered when the control plasmid lacking a target sequence was used. AmpR – ampicillin (carbenicillin) resistance gene, aadA – streptomycin resistance gene.

3.3. Final remarks

Here we showed that the natural diversity of CRISPR-Cas systems can be harnessed to identify and develop novel RNA-guided site specific nucleases. This was achieved by applying broad screens as in the case of the large set of type II Cas9 orthologs, resulting in the identification of functional Cas9 nucleases exhibiting diverse PAM recognition sequences of varying compositions. Proteins recognizing shorter PAMs could help increase the targeting space susceptible to editing by CRISPR-Cas systems, while orthologs with longer PAM recognition may afford higher specificity (C. M. Lee et al., 2016; Müller et al., 2016). Analysis of the relationship between the primary protein structure of the PAM-interacting domains and the empirically determined PAM recognition of the respective Cas9 orthologs may assist in developing more thorough models for *in silico* PAM preference prediction for novel putative Cas9 nucleases, as well as facilitate the generation of chimeric

Cas9 mutants by exchanging their PI domains (Ciciani et al., 2022; Ma et al., 2019). Further biochemical characterization of the different Cas9 nuclease dsDNA cleavage activity *in vitro* revealed a range of temperature dependencies, resulting in nucleases that could potentially be used in psychrophiles (Yusof et al., 2021) or thermophiles (Le & Sun, 2022), as well as enzymes that have a well-defined range of tolerated temperatures, suggesting of the possibility of thermal control of activity (Zhuo et al., 2021). Additionaly, preferences for longer tracts of gRNA spacer – DNA substrate complementarity as well as generation of varying DNA termini post-cleavage, which could facilitate alternative DNA repair outcomes in mammalian cell editing (Y.-W. Fu et al., 2021), were observed. In all, this provides a selection of orthologous Cas9 nucleases with properties suitable for applications where features attributed to other currently employed Cas9 enzymes may be disadvanteguous.

Alternatively, we identified a new family of type V Cas12 nuclease and applied a more in-depth strategy towards its characterization. Phylogenetic analysis revealed the system to be distinct from other Cas12 subtypes designating it as a separate subtype Cas12l. It exhibited recognition of a C-rich PAM sequence, providing a counterbalance to the T-rich PAM recognition characteristic to most Cas12 effectors. Next, we discerned that a tracrRNA is required for dsDNA cleavage, which could be fused with the crRNA into a sgRNA. Further biochemical characterization *in vitro* revealed the optimal cleavage temperature of around 50°C, a preference for supercoiled plasmid DNA substrates and collateral non-specific cleavage of ssDNA and ssRNA. The elevated temperature required for ideal activity is a feature that may be advantageous for editing in cell types permissible to heat shock treatments, as well as in nucleic acid detection strategies employing isothermal amplification methods (J. Joung et al., 2020; Nandy et al., 2019; Q. Wang et al., 2020). The increased cleavage efficiency of negatively supercoiled DNA has been observed with other Cas enzymes as well (Aelst et al., 2019; Westra et al., 2012) and is most likely explained as a way of promoting DNA strand-opening and R-loop formation (López-García, 1999). Finally, one member of the Cas12l family facilitated interference against invading plasmid DNA in a heterologous *E. coli* host. In all, this expands upon the increasing landscape of the type V CRISPR-Cas systems and posits the Cas12l family for further development as potential compact genome editing tools.

One of the Cas9 orthologs described in this work, Mga, was utilized in a biosensor setup which allowed the discrimination of a single nucleotide polymorphism (SNP) in unamplified genomic DNA (Balderston et al., 2021). The PAM sequence recognized by Mga Cas9 overlapped with the SNP associated with sickle-cell disease and was disrupted in mutant allelles, compromising binding and cleavage at higher sensitivity than in the case of the SNP being located within a Cas9 protospacer (Balderston et al., 2021). This highlights the advantage of flexible target selection permitted by a diverse collection of Cas9 enzymes.

A growing amount of other Cas9 nuclease orthologs are being applied in base editing systems resulting in expanded targeting space and altered editing windows (Kweon et al., 2023; M. Li et al., 2023; Trasanidou et al., 2023).

New orthologous Cas9 nucleases continue to be discovered and characterized (Cui et al., 2022; S. Gao et al., 2023; J. Wei et al., 2022). Furthermore, the increasing amount of Cas nuclease sequences provide the basis for building more thorough algorithms for the discovery of ever more phylogenetically distant Cas species, exemplified by the examination of ancient Cas nucleases (Alonso-Lerma et al., 2023) or the discovery of a new subtype of Cas9 effectors (Goltsman et al., 2022).

Recently, a separate group reported on their own characterization of the Cas12l system (Sun et al., 2023). Overall, their results corroborate those determined in our study. However, they were able to solve the 3D structure of one member of the family, Asp3Cas12l. This highlighted the divergent nature of the Cas12l family, exhibiting novel structural motifs, not characteristic to other Cas12 effectors. Furthermore, the group showed that Cas12l nucleases mediate editing in mammalian cells, in turn confirming the potential of these systems to be used in genome editing applications.

On a more fundamental level, profound advancements have been made in the characterization of the likely evolutionary precursors to Cas9 and Cas12 nucleases (Shmakov et al., 2017). Cas9 is thought to be evolved from IscB proteins of the IS200/IS605 transposable element family, members of which were examined by (Altae-Tran et al., 2021). It was determined that the IscB proteins are compact (400 aa) functional dsDNA nucleases guided by a large RNA species called the ω-RNA, which are encoded in a multitude of configurations across different IscB loci and are the likely ancestors for CRISPR tracrRNAs (Altae-Tran et al., 2021). Furthermore, cryo-EM structures of IscB-ω-RNA complexes display a similar bi-lobed protein architecture to Cas9, yet lacking the REC lobe of Cas9 (Kato et al., 2022;

Schuler et al., 2022). With a structured portion of the ω-RNA observed in an analogous position to the REC lobe it suggests that parts of the ω-RNA were replaced by protein domains throughout the evolution of Cas9. Also, the IscB complexes require a PAM-like motif for target recognition, which is referred to as a target-adjacent motif (TAM) and located at the 3‘ end of the target (Altae-Tran et al., 2021). Simultaneously, the in-depth characterization of the TnpB protein from the same IS200/IS605 transposable element family was reported in (Karvelis et al., 2021). It was shown that the TnpB protein is a compact (400 aa) functional DNA endonuclease guided by a RNA species derived from the 3‘ end of the TnpB gene and palindromic terminal end of the transposable element, with a variable sequence at the RNA 3‘ end conferring target recognition (Karvelis et al., 2021). Like IscB, TnpB requires a TAM sequence adjacent to the target, however it is located at the 5‘ end, akin to Cas12 systems. Cryo-EM structures of TnpB further verified the relationship between TnpB and Cas12 effectors, with significant similarities in the complex architectures, particularly compared to the compact Cas12f nucleases (Sasnauskas et al., 2023). Lastly, both IscB and TnpB were shown to cleave targets in human cells (Altae-Tran et al., 2021; Karvelis et al., 2021). Taking all this together, these findings not only provide essential information on the evolution of CRISPR-Cas systems, but also on what constitutes a minimal functional genome editor, due to their comparatively small size.

CONCLUSIONS

1. 79 type II Cas9 orthologs were shown to cleave dsDNA, facilitated by diverse PAM and tracRNA requirements.
2. Cas9 orthologs exhibited varied temperature dependencies, spacer length preferences and dsDNA cleavage patterns *in vitro*.
3. New type V CRISPR-Cas nucleases, Cas12l, were shown to cleave dsDNA, facilitated by dual guide RNAs and a C-rich PAM recognition.
4. Cas12l enzymes exhibited optimal dsDNA cleavage activity *in vitro* at temperatures around 50°C.
5. Cas12l exhibited collateral ssRNA and ssDNA cleavage activity upon target recognition.
6. Asp2Cas12l mediated protection against invading plasmid DNA in a heterologous *E. coli* host.

LIST OF PUBLICATIONS

The thesis is based on the following original publications:

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.L., Mabuchi, M., Fuchs, R.T., Schildkraut, E., Weigele, P., Jack, W., Robb, G.B., Venclovas, C., Siksnys, V. A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nature Communications.* 2020 November; 11(1): 5512 doi: 10.1038/s41467-020-19344-1

2. **Urbaitis, T.***, Gasiunas, G.*, Young, J.K.*, Hou, Z., Paulraj, S., Godliauskaite, E., Juskeviciene, M.M., Stitilyte, M., Jasnauskaite, M., Mabuchi, M., Robb, G.B., Siksnys, V. A new family of CRISPR-type V nucleases with C-rich PAM recognition. *EMBO reports*. 2022 December; 23(12): e55481 doi: 10.15252/embr.202255481

*- these authors contributed equally.

CONFERENCE PRESENTATIONS

Oral presentations

1. **Urbaitis, T.**, Gasiunas G., Young, J.K., Hou, Z., Paulraj, S., Godliauskaite, E., Grusyte, M., Stitilyte, M., Jasnauskaite, M., Schidlkraut, E., Robb, G.B., Siksnys, V. *A new family of CRISPR type V nuclease with C-rich PAM recognition.* Keystone Symposia, Precision Genome Engineering. Keystone, USA, 2022 04 27 – 05 01.

Poster presentations

1. **Urbaitis, T.**, Gasiunas, G., Young, J.K., Jasnauskaite, M., Grusyte, M., Paulraj, S., Curcuru, J.L., Mabuchi, M., Fuchs, R.T., Schildkraut, E., Robb, G.B., Siksnys, V.. *Harnessing the diversity of Cas9 orthologs for genome editing*. International Conference of Life Sciences The Coins 2020. Vilnius, Lithuania, 2020 02 27.
2. **Urbaitis, T.**, Gasiunas, G., Young, J.K., Jasnauskaite, M., Grusyte, M., Paulraj, S., Curcuru, J.L., Mabuchi, M., Fuchs, R.T., Schildkraut, E., Robb, G.B., Siksnys, V. *Harnessing the diversity of Cas9 orthologs for genome editing*. Keystone eSymposia Precision Engineering of the Genome, Epigenome and Transcriptome. 2021 03 08
3. **Urbaitis, T.**, Gasiunas, G., Young, J.K., Jasnauskaite, M., Grusyte, M., Paulraj, S., Curcuru, J.L., Mabuchi, M., Fuchs, R.T., Schildkraut, E., Robb, G.B., Siksnys, V. *Harnessing the diversity of Cas9 orthologs for genome editing*. 64th International Conference for Students of Physics and Natural Sciences Open Readings 2021. 2021 03 16-19.
4. **Urbaitis, T.**, Gasiunas, G., Young, J.K., Hou, Z., Paulraj, S., Godliauskaite, E., Grusyte, M., Alarcon, C., Chilcoat, N.D., Schildkraut, E., Robb, G.B., Siksnys, V. *Introducing Cas-beta: a new family of compact class 2 CRISPR associated nucleases.* Cold Spring Harbor Laboratory Genome Engineering: CRISPR Frontiers. 2021 08 18-20

REFERENCES

1. Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., Verdine, V., Cox, D. B. T., Kellner, M. J., Regev, A., Lander, E. S., Voytas, D. F., Ting, A. Y., & Zhang, F. (2017). RNA targeting with CRISPR–Cas13. *Nature*, *550*(7675), 280–284. https://doi.org/10.1038/nature24049
2. Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B. T., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E. S., Koonin, E. V., & Zhang, F. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, *353*(6299), aaf5573. https://doi.org/10.1126/science.aaf5573
3. Aelst, K. van, Martínez-Santiago, C. J., Cross, S. J., & Szczelkun, M. D. (2019). The Effect of DNA Topology on Observed Rates of R-Loop Formation and DNA Strand Cleavage by CRISPR Cas12a. *Genes*, *10*(2), 169. https://doi.org/10.3390/genes10020169
4. Alonso-Lerma, B., Jabalera, Y., Samperio, S., Morin, M., Fernandez, A., Hille, L. T., Silverstein, R. A., Quesada-Ganuza, A., Reifs, A., Fernández-Peñalver, S., Benitez, Y., Soletto, L., Gavira, J. A., Diaz, A., Vranken, W., Sanchez-Mejias, A., Güell, M., Mojica, F. J. M., Kleinstiver, B. P., … Perez-Jimenez, R. (2023). Evolution of CRISPR-associated endonucleases as inferred from resurrected proteins. *Nature Microbiology*, *8*(1), 77–90. https://doi.org/10.1038/s41564-022-01265-y
5. Altae-Tran, H., Kannan, S., Demircioglu, F. E., Oshiro, R., Nety, S. P., McKay, L. J., Dlakić, M., Inskeep, W. P., Makarova, K. S., Macrae, R. K., Koonin, E. V., & Zhang, F. (2021). The widespread IS200/605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science*, *374*(6563), 57–65. https://doi.org/10.1126/science.abj6856
6. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/s0022-2836(05)80360-2
7. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389
8. Amrani, N., Gao, X. D., Liu, P., Edraki, A., Mir, A., Ibraheim, R., Gupta, A., Sasaki, K. E., Wu, T., Donohoue, P. D., Settle, A. H., Lied, A. M., McGovern, K., Fuller, C. K., Cameron, P., Fazzio, T. G., Zhu, L. J., Wolfe, S. A., & Sontheimer, E. J. (2018). NmeCas9 is an intrinsically high-fidelity genome-editing platform. *Genome Biology*, *19*(1), 214. https://doi.org/10.1186/s13059-018-1591-1
9. Anzalone, A. V., Koblan, L. W., & Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology*, *38*(7), 824–844. https://doi.org/10.1038/s41587-020-0561-9

10. Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4
11. Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., & Morikawa, K. (1994). Atomic structure of the RuvC resolvase: A holliday junction-specific endonuclease from E. coli. *Cell*, *78*(6), 1063–1072. https://doi.org/10.1016/0092-8674(94)90280-1
12. Asaeda, A., Ide, H., Asagoshi, K., Matsuyama, S., Tano, K., Murakami, A., Takamori, Y., & Kubo, K. (2000). Substrate Specificity of Human Methylpurine DNA N-Glycosylase †. *Biochemistry*, *39*(8), 1959–1965. https://doi.org/10.1021/bi9917075
13. Aziz, R. K., Dwivedi, B., Akhter, S., Breitbart, M., & Edwards, R. A. (2015). Multidimensional metrics for estimating phage abundance, distribution, gene density, and sequence coverage in metagenomes. *Frontiers in Microbiology*, *6*, 381. https://doi.org/10.3389/fmicb.2015.00381
14. Balderston, S., Taulbee, J. J., Celaya, E., Fung, K., Jiao, A., Smith, K., Hajian, R., Gasiunas, G., Kutanovas, S., Kim, D., Parkinson, J., Dickerson, K., Ripoll, J.-J., Peytavi, R., Lu, H.-W., Barron, F., Goldsmith, B. R., Collins, P. G., Conboy, I. M., … Aran, K. (2021). Discrimination of single-point mutations in unamplified genomic DNA via Cas9 immobilized on a graphene field-effect transistor. *Nature Biomedical Engineering*, *5*(7), 713–725. https://doi.org/10.1038/s41551-021-00706-z
15. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, *315*(5819), 1709–1712. https://doi.org/10.1126/science.1138140
16. Biertümpfel, C., Yang, W., & Suck, D. (2007). Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature*, *449*(7162), 616–620. https://doi.org/10.1038/nature06152
17. Bigelyte, G., Young, J. K., Karvelis, T., Budre, K., Zedaveinyte, R., Djukanovic, V., Ginkel, E. V., Paulraj, S., Gasior, S., Jones, S., Feigenbutz, L., Clair, G. St., Barone, P., Bohn, J., Acharya, A., Zastrow-Hayes, G., Henkel-Heinecke, S., Silanskas, A., Seidel, R., & Siksnys, V. (2021). Miniature type V-F CRISPR-Cas nucleases enable targeted DNA modification in cells. *Nature Communications*, *12*(1), 6191. https://doi.org/10.1038/s41467-021-26469-4
18. Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., & Hugenholtz, P. (2007). CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*, *8*(1), 209–209. https://doi.org/10.1186/1471-2105-8-209
19. Briner, A. E., Donohoue, P. D., Gomaa, A. A., Selle, K., Slorach, E. M., Nye, C. H., Haurwitz, R. E., Beisel, C. L., May, A. P., & Barrangou, R. (2014). Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality. *Molecular Cell*, *56*(2), 333–339. https://doi.org/10.1016/j.molcel.2014.09.019
20. Broughton, J. P., Deng, X., Yu, G., Fasching, C. L., Servellita, V., Singh, J., Miao, X., Streithorst, J. A., Granados, A., Sotomayor-Gonzalez, A., Zorn,

K., Gopez, A., Hsu, E., Gu, W., Miller, S., Pan, C.-Y., Guevara, H., Wadford, D. A., Chen, J. S., & Chiu, C. Y. (2020). CRISPR–Cas12-based detection of SARS-CoV-2. *Nature Biotechnology*, *38*(7), 870–874. https://doi.org/10.1038/s41587-020-0513-4

21. Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & Oost, J. van der. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689
22. Buchlis, G., Podsakoff, G. M., Radu, A., Hawk, S. M., Flake, A. W., Mingozzi, F., & High, K. A. (2012). Factor IX expression in skeletal muscle of a severe hemophilia B patient 10 years after AAV-mediated gene transfer. *Blood*, *119*(13), 3038–3041. https://doi.org/10.1182/blood-2011-09-382317
23. Burstein, D., Harrington, L. B., Strutt, S. C., Probst, A. J., Anantharaman, K., Thomas, B. C., Doudna, J. A., & Banfield, J. F. (2017). New CRISPR-Cas systems from uncultivated microbes. *Nature*, *542*(7640), 237–241. https://doi.org/10.1038/nature21059
24. Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348
25. Chapman, J. R., Taylor, M. R. G., & Boulton, S. J. (2012). Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell*, *47*(4), 497–510. https://doi.org/10.1016/j.molcel.2012.07.029
26. Charlesworth, C. T., Deshpande, P. S., Dever, D. P., Camarena, J., Lemgart, V. T., Cromer, M. K., Vakulskas, C. A., Collingwood, M. A., Zhang, L., Bode, N. M., Behlke, M. A., Dejene, B., Cieniewicz, B., Romano, R., Lesch, B. J., Gomez-Ospina, N., Mantri, S., Pavel-Dinu, M., Weinberg, K. I., & Porteus, M. H. (2019). Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nature Medicine*, *25*(2), 249–254. https://doi.org/10.1038/s41591-018-0326-x
27. Charpentier, E., Richter, H., Oost, J. van der, & White, M. F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiology Reviews*, *39*(3), 428–441. https://doi.org/10.1093/femsre/fuv023
28. Chen, I.-M. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J. R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S. P., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N., & Kyrpides, N. C. (2019). IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research*, *47*(Database issue), D666–D677. https://doi.org/10.1093/nar/gky901
29. Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., Sternberg, S. H., Joung, K. J., Yildiz, A., & Doudna, J. A. (2018). Enhanced Proofreading Governs CRISPR-Cas9 Targeting Accuracy. *Biophysical Journal*, *114*(3), 194a. https://doi.org/10.1016/j.bpj.2017.11.1082
30. Chen, J. S., Ma, E., Harrington, L. B., Costa, M. D., Tian, X., Palefsky, J. M., & Doudna, J. A. (2018). CRISPR-Cas12a target binding unleashes

indiscriminate single-stranded DNase activity. *Science*, *360*(6387), 436–439. https://doi.org/10.1126/science.aar6245
31. Chen, P. J., Hussmann, J. A., Yan, J., Knipping, F., Ravisankar, P., Chen, P.-F., Chen, C., Nelson, J. W., Newby, G. A., Sahin, M., Osborn, M. J., Weissman, J. S., Adamson, B., & Liu, D. R. (2021). Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell*, *184*(22), 5635-5652.e29. https://doi.org/10.1016/j.cell.2021.09.018
32. Chen, Y., Zhi, S., Liu, W., Wen, J., Hu, S., Cao, T., Sun, H., Li, Y., Huang, L., Liu, Y., Liang, P., & Huang, J. (2020). Development of Highly Efficient Dual-AAV Split Adenosine Base Editor for In Vivo Gene Therapy. *Small Methods*, *4*(9), 2000309. https://doi.org/10.1002/smtd.202000309
33. Choulika, A., Perrin, A., Dujon, B., & Nicolas, J. F. (1995). Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *15*(4), 1968–1973. https://doi.org/10.1128/mcb.15.4.1968
34. Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K. L., Davidson, A. R., Fischer, E. R., Lander, G. C., & Wiedenheft, B. (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell*, *169*(1), 47-57.e11. https://doi.org/10.1016/j.cell.2017.03.012
35. Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., & Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, *33*(5), 543–548. https://doi.org/10.1038/nbt.3198
36. Chylinski, K., Makarova, K. S., Charpentier, E., & Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research*, *42*(10), 6091–6105. https://doi.org/10.1093/nar/gku241
37. Ciciani, M., Demozzi, M., Pedrazzoli, E., Visentin, E., Pezzè, L., Signorini, L., Blanco-Míguez, A., Zolfo, M., Asnicar, F., Casini, A., Segata, N., & Cereseto, A. (2022). *Tailored identification of Cas9 proteins with specific PAM requirements using a computational prediction pipeline*. https://doi.org/10.21203/rs.3.rs-1652795/v1
38. Collias, D., & Beisel, C. L. (2021). CRISPR technologies and the search for the PAM-free nuclease. *Nature Communications*, *12*(1), 555. https://doi.org/10.1038/s41467-020-20633-y
39. Collias, D., Leenay, R. T., Slotkowski, R. A., Zuo, Z., Collins, S. P., McGirr, B. A., Liu, J., & Beisel, C. L. (2020). A positive, growth-based PAM screen identifies noncanonical motifs recognized by the S. pyogenes Cas9. *Science Advances*, *6*(29), eabb4054. https://doi.org/10.1126/sciadv.abb4054
40. Condò, I. (2022). Rare Monogenic Diseases: Molecular Pathophysiology and Novel Therapies. *International Journal of Molecular Sciences*, *23*(12), 6525. https://doi.org/10.3390/ijms23126525
41. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143
42. Cui, Z., Tian, R., Huang, Z., Jin, Z., Li, L., Liu, J., Huang, Z., Xie, H., Liu, D., Mo, H., Zhou, R., Lang, B., Meng, B., Weng, H., & Hu, Z. (2022). FrCas9 is a CRISPR/Cas9 system with high editing efficiency and fidelity.

Nature Communications, *13*(1), 1425. https://doi.org/10.1038/s41467-022-29089-8

43. Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602–607. https://doi.org/10.1038/nature09886
44. Ding, X., Yin, K., Li, Z., Lalla, R. V., Ballesteros, E., Sfeir, M. M., & Liu, C. (2020). Ultrasensitive and visual detection of SARS-CoV-2 using all-in-one dual CRISPR-Cas12a assay. *Nature Communications*, *11*(1), 4711. https://doi.org/10.1038/s41467-020-18575-6
45. Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., Ma, C., Wang, S., Wu, D., Ma, Y., Fan, S., Wang, J., Gao, N., & Huang, Z. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature*, *532*(7600), 522–526. https://doi.org/10.1038/nature17944
46. Dooley, S. K., Baken, E. K., Moss, W. N., Howe, A., & Young, J. K. (2021). Identification and Evolution of Cas9 tracrRNAs. *The CRISPR Journal*, *4*(3), 438–447. https://doi.org/10.1089/crispr.2020.0093
47. East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H. D., Tjian, R., & Doudna, J. A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*, *538*(7624), 270–273. https://doi.org/10.1038/nature19802
48. Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England)*, *14*(9), 755–763. https://doi.org/10.1093/bioinformatics/14.9.755
49. Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Computational Biology*, *7*(10), e1002195. https://doi.org/10.1371/journal.pcbi.1002195
50. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340
51. Edgar, R. C. (2007). PILER-CR: Fast and accurate identification of CRISPR repeats. *BMC Bioinformatics*, *8*(1), 18. https://doi.org/10.1186/1471-2105-8-18
52. Edraki, A., Mir, A., Ibraheim, R., Gainetdinov, I., Yoon, Y., Song, C.-Q., Cao, Y., Gallant, J., Xue, W., Rivera-Pérez, J. A., & Sontheimer, E. J. (2019). A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. *Molecular Cell*, *73*(4), 714-726.e4. https://doi.org/10.1016/j.molcel.2018.12.003
53. Endo, A., Masafumi, M., Kaya, H., & Toki, S. (2016). Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida. *Scientific Reports*, *6*(1), 38169. https://doi.org/10.1038/srep38169
54. Faure, G., Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Crawley, A. B., Barrangou, R., & Koonin, E. V. (2018). Comparative genomics and evolution of trans-activating RNAs in Class 2 CRISPR-Cas systems. *RNA Biology*, *16*(4), 435–448. https://doi.org/10.1080/15476286.2018.1493331
55. Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., Makarova, K. S., & Koonin, E. V. (2019). CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nature Reviews Microbiology*, *17*(8), 513–525. https://doi.org/10.1038/s41579-019-0204-7

56. Fedorova, I., Arseniev, A., Selkova, P., Pobegalov, G., Goryanin, I., Vasileva, A., Musharova, O., Abramova, M., Kazalov, M., Zyubko, T., Artamonova, T., Artamonova, D., Shmakov, S., Khodorkovskii, M., & Severinov, K. (2020). DNA targeting by Clostridium cellulolyticum CRISPR–Cas9 Type II-C system. *Nucleic Acids Research*, *48*(4), 2026–2034. https://doi.org/10.1093/nar/gkz1225
57. Fedorova, I., Vasileva, A., Selkova, P., Abramova, M., Arseniev, A., Pobegalov, G., Kazalov, M., Musharova, O., Goryanin, I., Artamonova, D., Zyubko, T., Shmakov, S., Artamonova, T., Khodorkovskii, M., & Severinov, K. (2020). PpCas9 from Pasteurella pneumotropica — a compact Type II-C Cas9 ortholog active in human cells. *Nucleic Acids Research*, *48*(21), gkaa998-. https://doi.org/10.1093/nar/gkaa998
58. Ferdosi, S. R., Ewaisha, R., Moghadam, F., Krishna, S., Park, J. G., Ebrahimkhani, M. R., Kiani, S., & Anderson, K. S. (2019). Multifunctional CRISPR-Cas9 with engineered immunosilenced human T cell epitopes. *Nature Communications*, *10*(1), 1842. https://doi.org/10.1038/s41467-019-09693-x
59. Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, *39*(suppl), W29–W37. https://doi.org/10.1093/nar/gkr367
60. Flotte, T. R. (2004). Gene Therapy Progress and Prospects: Recombinant adeno-associated virus (rAAV) vectors. *Gene Therapy*, *11*(10), 805–810. https://doi.org/10.1038/sj.gt.3302233
61. Fonfara, I., Rhun, A. L., Chylinski, K., Makarova, K. S., Lécrivain, A.-L., Bzdrenga, J., Koonin, E. V., & Charpentier, E. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Research*, *42*(4), 2577–2590. https://doi.org/10.1093/nar/gkt1074
62. Fonfara, I., Richter, H., Bratovič, M., Rhun, A. L., & Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, *532*(7600), 517–521. https://doi.org/10.1038/nature17945
63. Frickey, T., & Lupas, A. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*, *20*(18), 3702–3704. https://doi.org/10.1093/bioinformatics/bth444
64. Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565
65. Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, *31*(9), 822–826. https://doi.org/10.1038/nbt.2623
66. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology*, *32*(3), 279–284. https://doi.org/10.1038/nbt.2808
67. Fu, Y.-W., Dai, X.-Y., Wang, W.-T., Yang, Z.-X., Zhao, J.-J., Zhang, J.-P., Wen, W., Zhang, F., Oberg, K. C., Zhang, L., Cheng, T., & Zhang, X.-B. (2021). Dynamics and competition of CRISPR–Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. *Nucleic Acids Research*, *49*(2), 969–985. https://doi.org/10.1093/nar/gkaa1251

68. Fuchs, R. T., Curcuru, J. L., Mabuchi, M., Noireterre, A., Weigele, P. R., Sun, Z., & Robb, G. B. (2022). Characterization of Cme and Yme thermostable Cas12a orthologs. *Communications Biology*, *5*(1), 325. https://doi.org/10.1038/s42003-022-03275-2
69. Gao, L., Cox, D. B. T., Yan, W. X., Manteiga, J. C., Schneider, M. W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., & Zhang, F. (2017). Engineered Cpf1 variants with altered PAM specificities. *Nature Biotechnology*, *35*(8), 789–792. https://doi.org/10.1038/nbt.3900
70. Gao, S., Wang, Y., Qi, T., Wei, J., Hu, Z., Liu, J., Sun, S., Liu, H., & Wang, Y. (2023). Genome editing with natural and engineered CjCas9 orthologs. *Molecular Therapy*, *31*(4), 1177–1187. https://doi.org/10.1016/j.ymthe.2023.01.029
71. Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, *109*(39), E2579–E2586. https://doi.org/10.1073/pnas.1208507109
72. Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464–471. https://doi.org/10.1038/nature24644
73. Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M., & Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, *159*(3), 647–661. https://doi.org/10.1016/j.cell.2014.09.029
74. Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, *154*(2), 442–451. https://doi.org/10.1016/j.cell.2013.06.044
75. Goltsman, D. S. A., Alexander, L. M., Lin, J.-L., Ocampo, R. F., Freeman, B., Lamothe, R. C., Rivas, A. P., Temoche-Diaz, M. M., Chadha, S., Nordenfelt, N., Janson, O. P., Barr, I., Devoto, A. E., Cost, G. J., Butterfield, C. N., Thomas, B. C., & Brown, C. T. (2022). Compact Cas9d and HEARO enzymes for genome editing discovered from uncultivated microbes. *Nature Communications*, *13*(1), 7602. https://doi.org/10.1038/s41467-022-35257-7
76. Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., & Zhang, F. (2018). Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science*, *360*(6387), 439–444. https://doi.org/10.1126/science.aaq0179
77. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine*, *24*(7), 927–930. https://doi.org/10.1038/s41591-018-0049-z
78. Harrington, L. B., Burstein, D., Chen, J. S., Paez-Espino, D., Ma, E., Witte, I. P., Cofsky, J. C., Kyrpides, N. C., Banfield, J. F., & Doudna, J. A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*, *362*(6416), 839–842. https://doi.org/10.1126/science.aav4294
79. Harrington, L. B., Ma, E., Chen, J. S., Witte, I. P., Gertz, D., Paez-Espino, D., Al-Shayeb, B., Kyrpides, N. C., Burstein, D., Banfield, J. F., & Doudna,

J. A. (2020). A scoutRNA Is Required for Some Type V CRISPR-Cas Systems. *Molecular Cell*, *79*(3), 416-424.e5. https://doi.org/10.1016/j.molcel.2020.06.022

80. Harrington, L. B., Paez-Espino, D., Staahl, B. T., Chen, J. S., Ma, E., Kyrpides, N. C., & Doudna, J. A. (2017). A thermostable Cas9 with increased lifetime in human plasma. *Nature Communications*, *8*(1), 1424. https://doi.org/10.1038/s41467-017-01408-4
81. He, L., James, M. St. J., Radovcic, M., Ivancic-Bace, I., & Bolt, E. L. (2020). Cas3 Protein—A Review of a Multi-Tasking Machine. *Genes*, *11*(2), 208. https://doi.org/10.3390/genes11020208
82. Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D., & Marraffini, L. A. (2015). Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature*, *519*(7542), 199–202. https://doi.org/10.1038/nature14245
83. Heyer, W.-D., Ehmsen, K. T., & Liu, J. (2010). Regulation of Homologous Recombination in Eukaryotes. *Annual Review of Genetics*, *44*(1), 113–139. https://doi.org/10.1146/annurev-genet-051710-150955
84. Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., & Gersbach, C. A. (2015). Epigenome editing by a CRISPR/Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology*, *33*(5), 510–517. https://doi.org/10.1038/nbt.3199
85. Hooton, S. P. T., & Connerton, I. F. (2015). Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. *Frontiers in Microbiology*, *5*, 744. https://doi.org/10.3389/fmicb.2014.00744
86. Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L.-F., Sontheimer, E. J., & Thomson, J. A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proceedings of the National Academy of Sciences*, *110*(39), 15644–15649. https://doi.org/10.1073/pnas.1313587110
87. Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., & Liu, D. R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, *556*(7699), 57–63. https://doi.org/10.1038/nature26155
88. Huang, C. J., Adler, B. A., & Doudna, J. A. (2022). A naturally DNase-free CRISPR-Cas12c enzyme silences gene expression. *Molecular Cell*, *82*(11), 2148-2160.e4. https://doi.org/10.1016/j.molcel.2022.04.020
89. Ihry, R. J., Worringer, K. A., Salick, M. R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., Randhawa, R., Kulkarni, T., Yang, Z., McAllister, G., Russ, C., Reece-Hoyes, J., Forrester, W., Hoffman, G. R., Dolmetsch, R., & Kaykas, A. (2018). p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine*, *24*(7), 939–946. https://doi.org/10.1038/s41591-018-0050-6
90. Jackson, R. N., & Wiedenheft, B. (2015). A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses. *Molecular Cell*, *58*(5), 722–728. https://doi.org/10.1016/j.molcel.2015.05.023

91. Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., & Brouns, S. J. J. (2017). CRISPR-Cas: Adapting to change. *Science*, *356*(6333). https://doi.org/10.1126/science.aal5056
92. Jacobi, A. M., Rettig, G. R., Turk, R., Collingwood, M. A., Zeiner, S. A., Quadros, R. M., Harms, D. W., Bonthuis, P. J., Gregg, C., Ohtsuka, M., Gurumurthy, C. B., & Behlke, M. A. (2017). Simplified CRISPR tools for efficient genome editing and streamlined protocols for their delivery into mammalian cells and mouse zygotes. *Methods (San Diego, Calif.)*, *121–122*, 16–28. https://doi.org/10.1016/j.ymeth.2017.03.021
93. Jacobsen, T., Liao, C., & Beisel, C. L. (2019). The Acidaminococcus sp. Cas12a nuclease recognizes GTTV and GCTV as non-canonical PAMs. *FEMS Microbiology Letters*, *366*(8). https://doi.org/10.1093/femsle/fnz085
94. Jasin, M., & Rothstein, R. (2013). Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harbor Perspectives in Biology*, *5*(11), a012740. https://doi.org/10.1101/cshperspect.a012740
95. Jiang, F., & Doudna, J. A. (2017). CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics*, *46*(1), 505–529. https://doi.org/10.1146/annurev-biophys-062215-010822
96. Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E., & Doudna, J. A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, *351*(6275), 867–871. https://doi.org/10.1126/science.aad8282
97. Jiang, F., Zhou, K., Ma, L., Gressel, S., & Doudna, J. A. (2015). A Cas9–guide RNA complex preorganized for target DNA recognition. *Science*, *348*(6242), 1477–1481. https://doi.org/10.1126/science.aab1452
98. Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, *31*(3), 233–239. https://doi.org/10.1038/nbt.2508
99. Jianwei, L., Jobichen, C., Machida, S., Meng, S., Read, R. J., Hongying, C., Jian, S., Yuan, Y. A., & Sivaraman, J. (2023). Structures of apo Cas12a and its complex with crRNA and DNA reveal the dynamics of ternary complex formation and target DNA cleavage. *PLOS Biology*, *21*(3), e3002023. https://doi.org/10.1371/journal.pbio.3002023
100. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829
101. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *ELife*, *2*, e00471. https://doi.org/10.7554/elife.00471
102. Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., & Doudna, J. A. (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science*, *343*(6176), 1247997. https://doi.org/10.1126/science.1247997
103. Jones, D., Unoson, C., Leroy, P., Curic, V., & Elf, J. (2017). Kinetics of dCas9 Target Search in Escherichia Coli. *Biophysical Journal*, *112*(3), 314a. https://doi.org/10.1016/j.bpj.2016.11.1700

104. Joung, J. K., & Sander, J. D. (2012). TALENs: a widely applicable technology for targeted genome editing. *Nature Reviews. Molecular Cell Biology*, *14*(1), 49–55. https://doi.org/10.1038/nrm3486
105. Joung, J., Ladha, A., Saito, M., Kim, N.-G., Woolley, A. E., Segel, M., Barretto, R. P. J., Ranu, A., Macrae, R. K., Faure, G., Ioannidi, E. I., Krajeski, R. N., Bruneau, R., Huang, M.-L. W., Yu, X. G., Li, J. Z., Walker, B. D., Hung, D. T., Greninger, A. L., … Zhang, F. (2020). Detection of SARS-CoV-2 with SHERLOCK One-Pot Testing. *The New England Journal of Medicine*, *383*(15), NEJMc2026172. https://doi.org/10.1056/nejmc2026172
106. Ka, D., Jang, D. M., Han, B. W., & Bae, E. (2018). Molecular organization of the type II-A CRISPR adaptation module and its interaction with Cas9 via Csn2. *Nucleic Acids Research*, *46*(18), 9805–9815. https://doi.org/10.1093/nar/gky702
107. Ka, D., Lee, H., Jung, Y.-D., Kim, K., Seok, C., Suh, N., & Bae, E. (2016). Crystal Structure of Streptococcus pyogenes Cas1 and Its Interaction with Csn2 in the Type II CRISPR-Cas System. *Structure*, *24*(1), 70–79. https://doi.org/10.1016/j.str.2015.10.019
108. Kagita, A., Lung, M. S. Y., Xu, H., Kita, Y., Sasakawa, N., Iguchi, T., Ono, M., Wang, X. H., Gee, P., & Hotta, A. (2021). Efficient ssODN-Mediated Targeting by Avoiding Cellular Inhibitory RNAs through Precomplexed CRISPR-Cas9/sgRNA Ribonucleoprotein. *Stem Cell Reports*, *16*(4), 985–996. https://doi.org/10.1016/j.stemcr.2021.02.013
109. Kantor, R. S., Huddy, R. J., Iyer, R., Thomas, B. C., Brown, C. T., Anantharaman, K., Tringe, S., Hettich, R. L., Harrison, S. T. L., & Banfield, J. F. (2017). Genome-Resolved Meta-Omics Ties Microbial Dynamics to Process Performance in Biotechnology for Thiocyanate Degradation. *Environmental Science & Technology*, *51*(5), 2944–2953. https://doi.org/10.1021/acs.est.6b04477
110. Karvelis, T., Bigelyte, G., Young, J. K., Hou, Z., Zedaveinyte, R., Budre, K., Paulraj, S., Djukanovic, V., Gasior, S., Silanskas, A., Venclovas, Č., & Siksnys, V. (2020). PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Research*, *48*(9), 5016–5023. https://doi.org/10.1093/nar/gkaa208
111. Karvelis, T., Druteika, G., Bigelyte, G., Budre, K., Zedaveinyte, R., Silanskas, A., Kazlauskas, D., Venclovas, Č., & Siksnys, V. (2021). Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature*, *599*(7886), 692–696. https://doi.org/10.1038/s41586-021-04058-1
112. Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P., & Siksnys, V. (2013). crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. *RNA Biology*, *10*(5), 841–851. https://doi.org/10.4161/rna.24203
113. Karvelis, T., Gasiunas, G., Young, J., Bigelyte, G., Silanskas, A., Cigan, M., & Siksnys, V. (2015). Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biology*, *16*(1), 253. https://doi.org/10.1186/s13059-015-0818-7
114. Karvelis, T., Young, J. K., & Siksnys, V. (2018). A pipeline for characterization of novel Cas9 orthologs. *Methods in Enzymology*, *616*, 219–240. https://doi.org/10.1016/bs.mie.2018.10.021

115. Kato, K., Okazaki, S., Kannan, S., Altae-Tran, H., Demircioglu, F. E., Isayama, Y., Ishikawa, J., Fukuda, M., Macrae, R. K., Nishizawa, T., Makarova, K. S., Koonin, E. V., Zhang, F., & Nishimasu, H. (2022). Structure of the IscB–ωRNA ribonucleoprotein complex, the likely ancestor of CRISPR-Cas9. *Nature Communications*, *13*(1), 6719. https://doi.org/10.1038/s41467-022-34378-3
116. Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010
117. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, *357*(6351), 605–609. https://doi.org/10.1126/science.aao0100
118. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č., & Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Molecular Cell*, *62*(2), 295–306. https://doi.org/10.1016/j.molcel.2016.03.024
119. Kenjo, E., Hozumi, H., Makita, Y., Iwabuchi, K. A., Fujimoto, N., Matsumoto, S., Kimura, M., Amano, Y., Ifuku, M., Naoe, Y., Inukai, N., & Hotta, A. (2021). Low immunogenicity of LNP allows repeated administrations of CRISPR-Cas9 mRNA into skeletal muscle in mice. *Nature Communications*, *12*(1), 7101. https://doi.org/10.1038/s41467-021-26714-w
120. Kieper, S. N., Almendros, C., Behler, J., McKenzie, R. E., Nobrega, F. L., Haagsma, A. C., Vink, J. N. A., Hess, W. R., & Brouns, S. J. J. (2018). Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Reports*, *22*(13), 3377–3384. https://doi.org/10.1016/j.celrep.2018.02.103
121. Kim, D., Kim, J., Hur, J. K., Been, K. W., Yoon, S., & Kim, J.-S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nature Biotechnology*, *34*(8), 863–868. https://doi.org/10.1038/nbt.3609
122. Kim, D., Kim, S., Kim, S., Park, J., & Kim, J.-S. (2016). Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Research*, *26*(3), 406–415. https://doi.org/10.1101/gr.199588.115
123. Kim, D., Luk, K., Wolfe, S. A., & Kim, J.-S. (2019). Evaluating and Enhancing Target Specificity of Gene-Editing Nucleases and Deaminases. *Annual Review of Biochemistry*, *88*(1), 1–30. https://doi.org/10.1146/annurev-biochem-013118-111730
124. Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H.-Y., Song, D. W., Lee, K. J., Jung, M. H., Kim, S., Kim, J. H., Kim, J. H., & Kim, J.-S. (2017). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nature Communications*, *8*(1), 14500. https://doi.org/10.1038/ncomms14500
125. Kim, H., Kim, S.-T., Ryu, J., Kang, B.-C., Kim, J.-S., & Kim, S.-G. (2016). CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nature Communications*, *8*(1), 14406. https://doi.org/10.1038/ncomms14406

126. Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., & Liu, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology*, *35*(4), 371–376. https://doi.org/10.1038/nbt.3803
127. Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, *529*(7587), 490–495. https://doi.org/10.1038/nature16526
128. Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, *523*(7561), 481–485. https://doi.org/10.1038/nature14592
129. Kleinstiver, B. P., Sousa, A. A., Walton, R. T., Tak, Y. E., Hsu, J. Y., Clement, K., Welch, M. M., Horng, J. E., Malagon-Lopez, J., Scarfò, I., Maus, M. V., Pinello, L., Aryee, M. J., & Joung, J. K. (2019). Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nature Biotechnology*, *37*(3), 276–282. https://doi.org/10.1038/s41587-018-0011-0
130. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*, *571*(7764), 219–225. https://doi.org/10.1038/s41586-019-1323-z
131. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420–424. https://doi.org/10.1038/nature17946
132. Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., Kim, Y. B., Badran, A. H., & Liu, D. R. (2017). Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Science Advances*, *3*(8), eaao4774. https://doi.org/10.1126/sciadv.aao4774
133. Koonin, E. V., & Dolja, V. V. (2013). A virocentric perspective on the evolution of life. *Current Opinion in Virology*, *3*(5), 546–557. https://doi.org/10.1016/j.coviro.2013.06.008
134. Koonin, E. V., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1772), 20180087. https://doi.org/10.1098/rstb.2018.0087
135. Koonin, E. V., Makarova, K. S., & Wolf, Y. I. (2016). Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annual Review of Microbiology*, *71*(1), 1–29. https://doi.org/10.1146/annurev-micro-090816-093830
136. Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., & Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 192–198. https://doi.org/10.1038/nbt.3450
137. Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*, *36*(8), 765–771. https://doi.org/10.1038/nbt.4192

138. Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054
139. Kurihara, N., Nakagawa, R., Hirano, H., Okazaki, S., Tomita, A., Kobayashi, K., Kusakizako, T., Nishizawa, T., Yamashita, K., Scott, D. A., Nishimasu, H., & Nureki, O. (2022). Structure of the type V-C CRISPR-Cas effector enzyme. *Molecular Cell*, *82*(10), 1865-1877.e4. https://doi.org/10.1016/j.molcel.2022.03.006
140. Kweon, J., Jang, A.-H., Kwon, E., Kim, U., Shin, H. R., See, J., Jang, G., Lee, C., Koo, T., Kim, S., & Kim, Y. (2023). Targeted dual base editing with Campylobacter jejuni Cas9 by single AAV-mediated delivery. *Experimental & Molecular Medicine*, *55*(2), 377–384. https://doi.org/10.1038/s12276-023-00938-w
141. Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research*, *47*(W1), W171–W174. https://doi.org/10.1093/nar/gkz365
142. Lam, D. K., Feliciano, P. R., Arif, A., Bohnuud, T., Fernandez, T. P., Gehrke, J. M., Grayson, P., Lee, K. D., Ortega, M. A., Sawyer, C., Schwaegerle, N. D., Peraro, L., Young, L., Lee, S.-J., Ciaramella, G., & Gaudelli, N. M. (2022). Improved cytosine base editors generated from TadA variants. *Nature Biotechnology*, 1–12. https://doi.org/10.1038/s41587-022-01611-9
143. Le, Y., & Sun, J. (2022). CRISPR/Cas genome editing systems in thermophiles: Current status, associated challenges, and future perspectives. *Advances in Applied Microbiology*, *118*, 1–30. https://doi.org/10.1016/bs.aambs.2022.02.001
144. Lee, C. M., Cradick, T. J., & Bao, G. (2016). The Neisseria meningitidis CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells. *Molecular Therapy*, *24*(3), 645–654. https://doi.org/10.1038/mt.2016.8
145. Lee, H., Zhou, Y., Taylor, D. W., & Sashital, D. G. (2018). Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. *Molecular Cell*, *70*(1), 48-59.e5. https://doi.org/10.1016/j.molcel.2018.03.003
146. Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y., Lee, K., Jung, I., Kim, D., Kim, S., & Kim, J.-S. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. *Nature Communications*, *9*(1), 3048. https://doi.org/10.1038/s41467-018-05477-x
147. Li, A., Tanner, M. R., Lee, C. M., Hurley, A. E., Giorgi, M. D., Jarrett, K. E., Davis, T. H., Doerfler, A. M., Bao, G., Beeton, C., & Lagor, W. R. (2020). AAV-CRISPR Gene Editing Is Negated by Pre-existing Immunity to Cas9. *Molecular Therapy*, *28*(6), 1432–1441. https://doi.org/10.1016/j.ymthe.2020.04.017
148. Li, L., Li, S., Wu, N., Wu, J., Wang, G., Zhao, G., & Wang, J. (2019). HOLMESv2: A CRISPR-Cas12b-Assisted Platform for Nucleic Acid Detection and DNA Methylation Quantitation. *ACS Synthetic Biology*, *8*(10), 2228–2237. https://doi.org/10.1021/acssynbio.9b00209
149. Li, M., Zhao, Y., Xue, X., Zhong, J., Lin, J., Zhou, J., Yu, W., Chen, J., & Qiao, Y. (2023). Cas9-orthologue-mediated cytosine and adenine base

editors recognizing NNAAAA PAM sequences. *Biotechnology Journal*, *18*(5), e2200533. https://doi.org/10.1002/biot.202200533
150. Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., Lu, Z., Zhang, Y., Wu, J., Huang, X., Yang, L., & Chen, J. (2018). Base editing with a Cpf1–cytidine deaminase fusion. *Nature Biotechnology*, *36*(4), 324–327. https://doi.org/10.1038/nbt.4102
151. Li, Z., Zhang, H., Xiao, R., Han, R., & Chang, L. (2021). Cryo-EM structure of the RNA-guided ribonuclease Cas12g. *Nature Chemical Biology*, *17*(4), 387–393. https://doi.org/10.1038/s41589-020-00721-2
152. Liang, X., Potter, J., Kumar, S., Ravinder, N., & Chesnut, J. D. (2017). Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *Journal of Biotechnology*, *241*, 136–146. https://doi.org/10.1016/j.jbiotec.2016.11.011
153. Liang, Y., Xie, J., Zhang, Q., Wang, X., Gou, S., Lin, L., Chen, T., Ge, W., Zhuang, Z., Lian, M., Chen, F., Li, N., Ouyang, Z., Lai, C., Liu, X., Li, L., Ye, Y., Wu, H., Wang, K., & Lai, L. (2022). AGBE: a dual deaminase-mediated base editor by fusing CGBE with ABE for creating a saturated mutant population with multiple editing patterns. *Nucleic Acids Research*, *50*(9), 5384–5399. https://doi.org/10.1093/nar/gkac353
154. Lieber, M. 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. https://doi.org/10.1146/annurev.biochem.052308.093131
155. Lino, C. A., Harper, J. C., Carney, J. P., & Timlin, J. A. (2018). Delivering CRISPR: a review of the challenges and approaches. *Drug Delivery*, *25*(1), 1234–1257. https://doi.org/10.1080/10717544.2018.1474964
156. Liu, J.-J., Orlova, N., Oakes, B. L., Ma, E., Spinner, H. B., Baney, K. L. M., Chuck, J., Tan, D., Knott, G. J., Harrington, L. B., Al-Shayeb, B., Wagner, A., Brötzmann, J., Staahl, B. T., Talyor, K. L., Desmarais, J., Nogales, E., & Doudna, J. A. (2019). CRISPR-CasX is an RNA-dominated enzyme active for human genome editing. *Nature*, *566*(7743), 218–223. https://doi.org/10.1038/s41586-019-0908-x
157. Liu, L., Chen, P., Wang, M., Li, X., Wang, J., Yin, M., & Wang, Y. (2017). C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism. *Molecular Cell*, *65*(2), 310–322. https://doi.org/10.1016/j.molcel.2016.11.040
158. Liu, R. M., Liang, L. L., Freed, E., Chang, H., Oh, E., Liu, Z. Y., Garst, A., Eckert, C. A., & Gill, R. T. (2019). Synthetic chimeric nucleases function for efficient genome editing. *Nature Communications*, *10*(1), 5524. https://doi.org/10.1038/s41467-019-13500-y
159. Liu, T. Y., & Doudna, J. A. (2020). Chemistry of Class 1 CRISPR-Cas effectors: Binding, editing, and regulation. *Journal of Biological Chemistry*, *295*(42), 14473–14487. https://doi.org/10.1074/jbc.rev120.007034
160. Liu, X. S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R. A., & Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. *Cell*, *167*(1), 233-247.e17. https://doi.org/10.1016/j.cell.2016.08.056
161. Liu, Y., Schmidt, B., & Maskell, D. L. (2010). MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function

posterior probabilities. *Bioinformatics*, *26*(16), 1958–1964. https://doi.org/10.1093/bioinformatics/btq338

162. Liu, Z., Shan, H., Chen, S., Chen, M., Zhang, Q., Lai, L., & Li, Z. (2019). Improved base editor for efficient editing in GC contexts in rabbits with an optimized AID-Cas9 fusion. *The FASEB Journal*, *33*(8), 9210–9219. https://doi.org/10.1096/fj.201900476rr
163. López-García, P. (1999). DNA Supercoiling and Temperature Adaptation: A Clue to Early Diversification of Life? *Journal of Molecular Evolution*, *49*(4), 439–452. https://doi.org/10.1007/pl00006567
164. Lorenz, R., Bernhart, S. H., Siederdissen, C. H. Z., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology : AMB*, *6*(1), 26. https://doi.org/10.1186/1748-7188-6-26
165. Louwen, R., Horst-Kreft, D., Boer, A. G. de, Graaf, L. van der, Knegt, G. de, Hamersma, M., Heikema, A. P., Timms, A. R., Jacobs, B. C., Wagenaar, J. A., Endtz, H. P., Oost, J. van der, Wells, J. M., Nieuwenhuis, E. E. S., Vliet, A. H. M. van, Willemsen, P. T. J., Baarlen, P. van, & Belkum, A. van. (2013). A novel link between Campylobacter jejuni bacteriophage defence, virulence and Guillain–Barré syndrome. *European Journal of Clinical Microbiology & Infectious Diseases*, *32*(2), 207–226. https://doi.org/10.1007/s10096-012-1733-4
166. Ma, D., Xu, Z., Zhang, Z., Chen, X., Zeng, X., Zhang, Y., Deng, T., Ren, M., Sun, Z., Jiang, R., & Xie, Z. (2019). Engineer chimeric Cas9 to expand PAM recognition based on evolutionary information. *Nature Communications*, *10*(1), 560. https://doi.org/10.1038/s41467-019-08395-8
167. Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., Oost, J. van der, … Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), 722–736. https://doi.org/10.1038/nrmicro3569
168. Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, *18*(2), 67–83. https://doi.org/10.1038/s41579-019-0299-x
169. Markham, N. R., & Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization. *Methods in Molecular Biology (Clifton, N.J.)*, *453*, 3–31. https://doi.org/10.1007/978-1-60327-429-6_1
170. Marshall, R., Maxwell, C. S., Collins, S. P., Jacobsen, T., Luo, M. L., Begemann, M. B., Gray, B. N., January, E., Singer, A., He, Y., Beisel, C. L., & Noireaux, V. (2018). Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free Transcription-Translation System. *Molecular Cell*, *69*(1), 146-157.e3. https://doi.org/10.1016/j.molcel.2017.12.007
171. Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., & Ploegh, H. L. (2015). Increasing the efficiency of precise genome

editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature Biotechnology*, *33*(5), 538–542. https://doi.org/10.1038/nbt.3190
172. Meeske, A. J., & Marraffini, L. A. (2018). RNA Guide Complementarity Prevents Self-Targeting in Type VI CRISPR Systems. *Molecular Cell*, *71*(5), 791-801.e3. https://doi.org/10.1016/j.molcel.2018.07.013
173. Min, K., Yoon, H., Jo, I., Ha, N.-C., Jin, K. S., Kim, J.-S., & Lee, H. H. (2018). Structural insights into the apo-structure of Cpf1 protein from Francisella novicida. *Biochemical and Biophysical Research Communications*, *498*(4), 775–781. https://doi.org/10.1016/j.bbrc.2018.03.057
174. Ming, M., Ren, Q., Pan, C., He, Y., Zhang, Y., Liu, S., Zhong, Z., Wang, J., Malzahn, A. A., Wu, J., Zheng, X., Zhang, Y., & Qi, Y. (2020). CRISPR–Cas12b enables efficient plant genome engineering. *Nature Plants*, *6*(3), 202–208. https://doi.org/10.1038/s41477-020-0614-6
175. Mir, A., Edraki, A., Lee, J., & Sontheimer, E. J. (2018). Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chemical Biology*, *13*(2), 357–365. https://doi.org/10.1021/acschembio.7b00855
176. Mitchell, A. L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M. R., Kale, V., Potter, S. C., Richardson, L. J., Sakharova, E., Scheremetjew, M., Korobeynikov, A., Shlemov, A., Kunyavskaya, O., Lapidus, A., & Finn, R. D. (2019). MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Research*, *48*(D1), D570–D578. https://doi.org/10.1093/nar/gkz1035
177. Mohanraju, P., Saha, C., Baarlen, P. van, Louwen, R., Staals, R. H. J., & Oost, J. van der. (2021). Alternative functions of CRISPR-Cas systems in the evolutionary arms race. *Nature Reviews. Microbiology*, *20*(6), 351–364. https://doi.org/10.1038/s41579-021-00663-z
178. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), 733–740. https://doi.org/10.1099/mic.0.023960-0
179. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3
180. Moreno, A. M., Palmer, N., Alemán, F., Chen, G., Pla, A., Jiang, N., Chew, W. L., Law, M., & Mali, P. (2019). immune-orthogonal orthologues of AAV capsids and of Cas9 circumvent the immune response to the administration of gene therapy. *Nature Biomedical Engineering*, *3*(10), 806–816. https://doi.org/10.1038/s41551-019-0431-2
181. Müller, M., Lee, C. M., Gasiunas, G., Davis, T. H., Cradick, T. J., Siksnys, V., Bao, G., Cathomen, T., & Mussolino, C. (2016). Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Molecular Therapy*, *24*(3), 636–644. https://doi.org/10.1038/mt.2015.218
182. Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology*, *247*(4), 536–540. https://doi.org/10.1016/s0022-2836(05)80134-2

183. Nambiar, T. S., Billon, P., Diedenhofen, G., Hayward, S. B., Taglialatela, A., Cai, K., Huang, J.-W., Leuzzi, G., Cuella-Martin, R., Palacios, A., Gupta, A., Egli, D., & Ciccia, A. (2019). Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nature Communications*, *10*(1), 3395. https://doi.org/10.1038/s41467-019-11105-z
184. Nandy, S., Pathak, B., Zhao, S., & Srivastava, V. (2019). Heat-shock-inducible CRISPR/Cas9 system generates heritable mutations in rice. *Plant Direct*, *3*(5), e00145. https://doi.org/10.1002/pld3.145
185. Naqvi, M. M., Lee, L., Montaguth, O. E. T., Diffin, F. M., & Szczelkun, M. D. (2022). CRISPR–Cas12a-mediated DNA clamping triggers target-strand cleavage. *Nature Chemical Biology*, *18*(9), 1014–1022. https://doi.org/10.1038/s41589-022-01082-8
186. Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*(22), 2933–2935. https://doi.org/10.1093/bioinformatics/btt509
187. Nelson, J. W., Randolph, P. B., Shen, S. P., Everette, K. A., Chen, P. J., Anzalone, A. V., An, M., Newby, G. A., Chen, J. C., Hsu, A., & Liu, D. R. (2021). Engineered pegRNAs improve prime editing efficiency. *Nature Biotechnology*, *40*(3), 402–410. https://doi.org/10.1038/s41587-021-01039-7
188. Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von, & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300
189. Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A., & Jinek, M. (2017). Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature*, *548*(7669), 543–548. https://doi.org/10.1038/nature23467
190. Niewoehner, O., Jinek, M., & Doudna, J. A. (2014). Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Research*, *42*(2), 1341–1353. https://doi.org/10.1093/nar/gkt922
191. Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., & Nureki, O. (2014). Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*, *156*(5), 935–949. https://doi.org/10.1016/j.cell.2014.02.001
192. Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., & Doudna, J. A. (2015). Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature*, *527*(7579), 535–538. https://doi.org/10.1038/nature15760
193. Özcan, A., Krajeski, R., Ioannidi, E., Lee, B., Gardner, A., Makarova, K. S., Koonin, E. V., Abudayyeh, O. O., & Gootenberg, J. S. (2021). Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. *Nature*, *597*(7878), 720–725. https://doi.org/10.1038/s41586-021-03886-5
194. Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., & Randau, L. (2019). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nature Microbiology*, *4*(1), 89–96. https://doi.org/10.1038/s41564-018-0274-8

195. Palermo, G., Chen, J. S., Ricci, C. G., Rivalta, I., Jinek, M., Batista, V. S., Doudna, J. A., & McCammon, J. A. (2018). Key role of the REC lobe during CRISPR–Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Quarterly Reviews of Biophysics*, *51*, e9. https://doi.org/10.1017/s0033583518000070
196. Paques, F., & Duchateau, P. (2007). Meganucleases and DNA Double-Strand Break-Induced Recombination: Perspectives for Gene Therapy. *Current Gene Therapy*, *7*(1), 49–66. https://doi.org/10.2174/156652307779940216
197. Paul, B., & Montoya, G. (2020). CRISPR-Cas12a: Functional overview and applications. *Biomedical Journal*, *43*(1), 8–17. https://doi.org/10.1016/j.bj.2019.10.005
198. Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., Tsuchida, C. A., Li, Z., Cress, B. F., Knott, G. J., Jacobsen, S. E., Banfield, J. F., & Doudna, J. A. (2020). CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science*, *369*(6501), 333–337. https://doi.org/10.1126/science.abb1400
199. Pausch, P., Soczek, K. M., Herbst, D. A., Tsuchida, C. A., Al-Shayeb, B., Banfield, J. F., Nogales, E., & Doudna, J. A. (2021). DNA interference states of the hypercompact CRISPR–CasΦ effector. *Nature Structural & Molecular Biology*, *28*(8), 652–661. https://doi.org/10.1038/s41594-021-00632-3
200. Peters, J. E., Makarova, K. S., Shmakov, S., & Koonin, E. V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proceedings of the National Academy of Sciences*, *114*(35), E7358–E7366. https://doi.org/10.1073/pnas.1709035114
201. Porteus, M. (2014). Genome Editing: A New Approach to Human Therapeutics. *Annual Review of Pharmacology and Toxicology*, *56*(1), 1–28. https://doi.org/10.1146/annurev-pharmtox-010814-124454
202. Qin, W., & Wang, H. (2018). Delivery of CRISPR-Cas9 into Mouse Zygotes by Electroporation. *Methods in Molecular Biology (Clifton, N.J.)*, *1874*, 179–190. https://doi.org/10.1007/978-1-4939-8831-0_10
203. Qiu, M., Glass, Z., Chen, J., Haas, M., Jin, X., Zhao, X., Rui, X., Ye, Z., Li, Y., Zhang, F., & Xu, Q. (2021). Lipid nanoparticle-mediated codelivery of Cas9 mRNA and single-guide RNA achieves liver-specific in vivo genome editing of Angptl3. *Proceedings of the National Academy of Sciences*, *118*(10), e2020401118. https://doi.org/10.1073/pnas.2020401118
204. Ralph, A. P., & Carapetis, J. R. (2012). Group a streptococcal diseases and their global burden. *Current Topics in Microbiology and Immunology*, *368*, 1–27. https://doi.org/10.1007/82_2012_280
205. Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., Koonin, E. V., Sharp, P. A., & Zhang, F. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, *520*(7546), 186–191. https://doi.org/10.1038/nature14299
206. Raper, A. T., Stephenson, A. A., & Suo, Z. (2018). Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *Journal of the American Chemical Society*, *140*(8), 2971–2984. https://doi.org/10.1021/jacs.7b13047
207. Rayner, E., Durin, M.-A., Thomas, R., Moralli, D., O'Cathail, S. M., Tomlinson, I., Green, C. M., & Lewis, A. (2019). CRISPR-Cas9 Causes Chromosomal Instability and Rearrangements in Cancer Cell Lines,

Detectable by Cytogenetic Methods. *The CRISPR Journal*, *2*(6), 406–416. https://doi.org/10.1089/crispr.2019.0006

208. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L., & Corn, J. E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature Biotechnology*, *34*(3), 339–344. https://doi.org/10.1038/nbt.3481
209. Richter, C., Dy, R. L., McKenzie, R. E., Watson, B. N. J., Taylor, C., Chang, J. T., McNeil, M. B., Staals, R. H. J., & Fineran, P. C. (2014). Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Research*, *42*(13), 8516–8526. https://doi.org/10.1093/nar/gku527
210. Richter, F., Fonfara, I., Bouazza, B., Schumacher, C. H., Bratovič, M., Charpentier, E., & Möglich, A. (2016). Engineering of temperature- and light-switchable Cas9 variants. *Nucleic Acids Research*, *44*(20), 10003–10014. https://doi.org/10.1093/nar/gkw930
211. Rim, J. H., Gopalappa, R., & Gee, H. Y. (2021). CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *New England Journal of Medicine*, *385*(18), 1721–1723. https://doi.org/10.1056/nejmc2114592
212. Ryan, D. E., Taussig, D., Steinfeld, I., Phadnis, S. M., Lunstad, B. D., Singh, M., Vuong, X., Okochi, K. D., McCaffrey, R., Olesiak, M., Roy, S., Yung, C. W., Curry, B., Sampson, J. R., Bruhn, L., & Dellinger, D. J. (2018). Improving CRISPR–Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Research*, *46*(2), 792–803. https://doi.org/10.1093/nar/gkx1199
213. Rybarski, J. R., Hu, K., Hill, A. M., Wilke, C. O., & Finkelstein, I. J. (2021). Metagenomic discovery of CRISPR-associated transposons. *Proceedings of the National Academy of Sciences*, *118*(49), e2112279118. https://doi.org/10.1073/pnas.2112279118
214. Saha, A., Arantes, P. R., Hsu, R. V., Narkhede, Y. B., Jinek, M., & Palermo, G. (2020). Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a. *Journal of Chemical Information and Modeling*, *60*(12), 6427–6437. https://doi.org/10.1021/acs.jcim.0c00929
215. Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454
216. Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y.-L., & Weiss, D. S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, *497*(7448), 254–257. https://doi.org/10.1038/nature12048
217. Samulski, R. J., & Muzyczka, N. (2014). AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annual Review of Virology*, *1*(1), 427–451. https://doi.org/10.1146/annurev-virology-031413-085355
218. Sasnauskas, G., Tamulaitiene, G., Druteika, G., Carabias, A., Silanskas, A., Kazlauskas, D., Venclovas, Č., Montoya, G., Karvelis, T., & Siksnys, V. (2023). TnpB structure reveals minimal functional core of Cas12 nuclease family. *Nature*, *616*(7956), 384–389. https://doi.org/10.1038/s41586-023-05826-x

219. Savva, R., McAuley-Hecht, K., Brown, T., & Pearl, L. (1995). The structural basis of specific base-excision repair by uracil–DNA glycosylase. *Nature*, *373*(6514), 487–493. https://doi.org/10.1038/373487a0
220. Schmidt, S. T., Yu, F. B., Blainey, P. C., May, A. P., & Quake, S. R. (2019). Nucleic acid cleavage with a hyperthermophilic Cas9 from an uncultured Ignavibacterium. *Proceedings of the National Academy of Sciences*, *116*(46), 23100–23105. https://doi.org/10.1073/pnas.1904273116
221. Schuler, G., Hu, C., & Ke, A. (2022). Structural basis for RNA-guided DNA cleavage by IscB-ωRNA and mechanistic comparison with Cas9. *Science*, *376*(6600), 1476–1481. https://doi.org/10.1126/science.abq7220
222. Schunder, E., Rydzewski, K., Grunow, R., & Heuner, K. (2013). First indication for a functional CRISPR/Cas system in Francisella tularensis. *International Journal of Medical Microbiology*, *303*(2), 51–60. https://doi.org/10.1016/j.ijmm.2012.11.004
223. Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., Oost, J. van der, Brouns, S. J. J., & Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences*, *108*(25), 10098–10103. https://doi.org/10.1073/pnas.1104144108
224. Shah, S. A., Erdmann, S., Mojica, F. J. M., & Garrett, R. A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biology*, *10*(5), 891–899. https://doi.org/10.4161/rna.23764
225. Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G., & Zhang, F. (2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science*, *343*(6166), 84–87. https://doi.org/10.1126/science.1247005
226. Shiimori, M., Garrett, S. C., Graveley, B. R., & Terns, M. P. (2018). Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci. *Molecular Cell*, *70*(5), 814-824.e6. https://doi.org/10.1016/j.molcel.2018.05.002
227. Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., Zhang, F., & Koonin, E. V. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular Cell*, *60*(3), 385–397. https://doi.org/10.1016/j.molcel.2015.10.008
228. Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O. O., Gootenberg, J. S., Makarova, K. S., Wolf, Y. I., Severinov, K., Zhang, F., & Koonin, E. V. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. *Nature Reviews Microbiology*, *15*(3), 169–182. https://doi.org/10.1038/nrmicro.2016.184
229. Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science*, *351*(6268), 84–88. https://doi.org/10.1126/science.aad5227
230. Smargon, A. A., Cox, D. B. T., Pyzocha, N. K., Zheng, K., Slaymaker, I. M., Gootenberg, J. S., Abudayyeh, O. A., Essletzbichler, P., Shmakov, S., Makarova, K. S., Koonin, E. V., & Zhang, F. (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. https://doi.org/10.1016/j.molcel.2016.12.023

231. Söding, J. (2005). Protein homology detection by HMM–HMM comparison. *Bioinformatics*, *21*(7), 951–960. https://doi.org/10.1093/bioinformatics/bti125
232. Stella, S., Mesa, P., Thomsen, J., Paul, B., Alcón, P., Jensen, S. B., Saligram, B., Moses, M. E., Hatzakis, N. S., & Montoya, G. (2018). Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell*, *175*(7), 1856-1871.e21. https://doi.org/10.1016/j.cell.2018.10.045
233. Stephens, C. J., Lauron, E. J., Kashentseva, E., Lu, Z. H., Yokoyama, W. M., & Curiel, D. T. (2019). Long-term correction of hemophilia B using adenoviral delivery of CRISPR/Cas9. *Journal of Controlled Release*, *298*, 128–141. https://doi.org/10.1016/j.jconrel.2019.02.009
234. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), 62–67. https://doi.org/10.1038/nature13011
235. Strecker, J., Jones, S., Koopal, B., Schmid-Burgk, J., Zetsche, B., Gao, L., Makarova, K. S., Koonin, E. V., & Zhang, F. (2018). Engineering of CRISPR-Cas12b for human genome editing. *Nature Communications*, *10*(1), 212. https://doi.org/10.1038/s41467-018-08224-4
236. Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., & Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, *365*(6448), 48–53. https://doi.org/10.1126/science.aax9181
237. Sun, A., Li, C.-P., Chen, Z., Zhang, S., Li, D.-Y., Yang, Y., Li, L.-Q., Zhao, Y., Wang, K., Li, Z., Liu, J., Liu, S., Wang, J., & Liu, J.-J. G. (2023). The compact Casπ (Cas12l) 'bracelet' provides a unique structural platform for DNA manipulation. *Cell Research*, *33*(3), 229–244. https://doi.org/10.1038/s41422-022-00771-2
238. Suo, Z. (2019). Bidirectional Degradation of DNA Cleavage Products Catalyzed by CRISPR/Cas9. *The FASEB Journal*, *33*(S1), 633.30-633.30. https://doi.org/10.1096/fasebj.2019.33.1_supplement.633.30
239. Suzuki, Y., Onuma, H., Sato, R., Sato, Y., Hashiba, A., Maeki, M., Tokeshi, M., Kayesh, M. E. H., Kohara, M., Tsukiyama-Kohara, K., & Harashima, H. (2021). Lipid nanoparticles loaded with ribonucleoprotein–oligonucleotide complexes synthesized using a microfluidic device exhibit robust genome editing and hepatitis B virus inhibition. *Journal of Controlled Release*, *330*, 61–71. https://doi.org/10.1016/j.jconrel.2020.12.013
240. Swarts, D. C., & Jinek, M. (2019). Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Molecular Cell*, *73*(3), 589-600.e4. https://doi.org/10.1016/j.molcel.2018.11.021
241. Swarts, D. C., Oost, J. van der, & Jinek, M. (2017). Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Molecular Cell*, *66*(2), 221-233.e4. https://doi.org/10.1016/j.molcel.2017.03.016
242. Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., & Seidel, R. (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. https://doi.org/10.1073/pnas.1402597111

243. Taha, E. A., Lee, J., & Hotta, A. (2022). Delivery of CRISPR-Cas tools for in vivo genome editing therapy: Trends and challenges. *Journal of Controlled Release*, *342*, 345–361. https://doi.org/10.1016/j.jconrel.2022.01.013
244. Takeda, S. N., Nakagawa, R., Okazaki, S., Hirano, H., Kobayashi, K., Kusakizako, T., Nishizawa, T., Yamashita, K., Nishimasu, H., & Nureki, O. (2021). Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Molecular Cell*, *81*(3), 558-570.e3. https://doi.org/10.1016/j.molcel.2020.11.035
245. Tamulaitis, G., Venclovas, Č., & Siksnys, V. (2017). Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends in Microbiology*, *25*(1), 49–61. https://doi.org/10.1016/j.tim.2016.09.012
246. Teng, F., Cui, T., Feng, G., Guo, L., Xu, K., Gao, Q., Li, T., Li, J., Zhou, Q., & Li, W. (2018). Repurposing CRISPR-Cas12b for mammalian genome engineering. *Cell Discovery*, *4*(1), 63. https://doi.org/10.1038/s41421-018-0069-3
247. Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673
248. Thyme, S. B., Boissel, S. J. S., Quadri, S. A., Nolan, T., Baker, D. A., Park, R. U., Kusak, L., Ashworth, J., & Baker, D. (2014). Reprogramming homing endonuclease specificity through computational design and directed evolution. *Nucleic Acids Research*, *42*(4), 2564–2576. https://doi.org/10.1093/nar/gkt1212
249. Tong, B., Dong, H., Cui, Y., Jiang, P., Jin, Z., & Zhang, D. (2021). The Versatile Type V CRISPR Effectors and Their Application Prospects. *Frontiers in Cell and Developmental Biology*, *8*, 622103. https://doi.org/10.3389/fcell.2020.622103
250. Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L., & Fowler, V. G. (2015). Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clinical Microbiology Reviews*, *28*(3), 603–661. https://doi.org/10.1128/cmr.00134-14
251. Toro, N., Martínez-Abarca, F., & González-Delgado, A. (2017). The Reverse Transcriptases Associated with CRISPR-Cas Systems. *Scientific Reports*, *7*(1), 7089. https://doi.org/10.1038/s41598-017-07828-y
252. Toro, N., Mestre, M. R., Martínez-Abarca, F., & González-Delgado, A. (2019). Recruitment of Reverse Transcriptase-Cas1 Fusion Proteins by Type VI-A CRISPR-Cas Systems. *Frontiers in Microbiology*, *10*, 2160. https://doi.org/10.3389/fmicb.2019.02160
253. Tóth, E., Varga, É., Kulcsár, P. I., Kocsis-Jutka, V., Krausz, S. L., Nyeste, A., Welker, Z., Huszár, K., Ligeti, Z., Tálas, A., & Welker, E. (2020). Improved LbCas12a variants with altered PAM specificities further broaden the genome targeting range of Cas12a nucleases. *Nucleic Acids Research*, *48*(7), 3722–3733. https://doi.org/10.1093/nar/gkaa110
254. Trasanidou, D., Barendse, P., Bouzetos, E., Haan, L. de, Bouwmeester, H., Staals, R. H. J., Mougiakos, I., & Oost, J. van der. (2023). Efficient Genome

and Base Editing in Human Cells Using ThermoCas9. *The CRISPR Journal*. https://doi.org/10.1089/crispr.2023.0005

255. Tsuchida, C. A., Zhang, S., Doost, M. S., Zhao, Y., Wang, J., O'Brien, E., Fang, H., Li, C.-P., Li, D., Hai, Z.-Y., Chuck, J., Brötzmann, J., Vartoumian, A., Burstein, D., Chen, X.-W., Nogales, E., Doudna, J. A., & Liu, J.-J. G. (2022). Chimeric CRISPR-CasX enzymes and guide RNAs for improved genome editing activity. *Molecular Cell*, *82*(6), 1199-1209.e6. https://doi.org/10.1016/j.molcel.2022.02.002
256. Tsui, T. K. M., Hand, T. H., Duboy, E. C., & Li, H. (2017). The Impact of DNA Topology and Guide Length on Target Selection by a Cytosine-Specific Cas9. *ACS Synthetic Biology*, *6*(6), 1103–1113. https://doi.org/10.1021/acssynbio.7b00050
257. Uchida, N., Drysdale, C. M., Nassehi, T., Gamer, J., Yapundich, M., DiNicola, J., Shibata, Y., Hinds, M., Gudmundsdottir, B., Haro-Mora, J. J., Demirci, S., & Tisdale, J. F. (2021). Cas9 protein delivery non-integrating lentiviral vectors for gene correction in sickle cell disease. *Molecular Therapy - Methods & Clinical Development*, *21*, 121–132. https://doi.org/10.1016/j.omtm.2021.02.022
258. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., & Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, *11*(9), 636–646. https://doi.org/10.1038/nrg2842
259. Venclovas, Č. (2016). Structure of Csm2 elucidates the relationship between small subunits of CRISPR-Cas effector complexes. *FEBS Letters*, *590*(10), 1521–1529. https://doi.org/10.1002/1873-3468.12179
260. Verdera, H. C., Kuranda, K., & Mingozzi, F. (2020). AAV Vector Immunogenicity in Humans: A Long Journey to Successful Gene Transfer. *Molecular Therapy*, *28*(3), 723–746. https://doi.org/10.1016/j.ymthe.2019.12.010
261. Vorontsova, D., Datsenko, K. A., Medvedeva, S., Bondy-Denomy, J., Savitskaya, E. E., Pougach, K., Logacheva, M., Wiedenheft, B., Davidson, A. R., Severinov, K., & Semenova, E. (2015). Foreign DNA acquisition by the I-F CRISPR–Cas system requires all components of the interference machinery. *Nucleic Acids Research*, *43*(22), 10848–10860. https://doi.org/10.1093/nar/gkv1261
262. Wagner, D. L., Amini, L., Wendering, D. J., Burkhardt, L.-M., Akyüz, L., Reinke, P., Volk, H.-D., & Schmueck-Henneresse, M. (2019). High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population. *Nature Medicine*, *25*(2), 242–248. https://doi.org/10.1038/s41591-018-0204-6
263. Walton, R. T., Christie, K. A., Whittaker, M. N., & Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science (New York, N.Y.)*, *368*(6488), 290–296. https://doi.org/10.1126/science.aba8853
264. Wang, L., Xue, W., Yan, L., Li, X., Wei, J., Chen, M., Wu, J., Yang, B., Yang, L., & Chen, J. (2017). Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Research*, *27*(10), 1289–1292. https://doi.org/10.1038/cr.2017.111
265. Wang, Q., Alariqi, M., Wang, F., Li, B., Ding, X., Rui, H., Li, Y., Xu, Z., Qin, L., Sun, L., Li, J., Zou, J., Lindsey, K., Zhang, X., & Jin, S. (2020). The application of a heat-inducible CRISPR/Cas12b (C2c1) genome editing

system in tetraploid cotton (G. hirsutum) plants. *Plant Biotechnology Journal*, *18*(12), 2436–2443. https://doi.org/10.1111/pbi.13417
266. Wang, Z., & Zhong, C. (2021). Cas12c-DETECTOR: A specific and sensitive Cas12c-based DNA detection platform. *International Journal of Biological Macromolecules*, *193*(Pt A), 441–449. https://doi.org/10.1016/j.ijbiomac.2021.10.167
267. Wei, J., Hou, L., Liu, J., Wang, Z., Gao, S., Qi, T., Gao, S., Sun, S., & Wang, Y. (2022). Closely related type II-C Cas9 orthologs recognize diverse PAMs. *ELife*, *11*, e77825. https://doi.org/10.7554/elife.77825
268. Wei, T., Cheng, Q., Min, Y.-L., Olson, E. N., & Siegwart, D. J. (2020). Systemic nanoparticle delivery of CRISPR-Cas9 ribonucleoproteins for effective tissue specific genome editing. *Nature Communications*, *11*(1), 3232. https://doi.org/10.1038/s41467-020-17029-3
269. Wei, Y., Terns, R. M., & Terns, M. P. (2015). Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. *Genes & Development*, *29*(4), 356–361. https://doi.org/10.1101/gad.257550.114
270. Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., & Brouns, S. J. J. (2013). Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLoS Genetics*, *9*(9), e1003742. https://doi.org/10.1371/journal.pgen.1003742
271. Westra, E. R., van Erp, P. B. G., Künne, T., Wong, S. P., Staals, R. H. J., Seegers, C. L. C., Bollen, S., Jore, M. M., Semenova, E., Severinov, K., de Vos, W. M., Dame, R. T., de Vries, R., Brouns, S. J. J., & van der Oost, J. (2012). CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Molecular Cell*, *46*(5), 595–605. https://doi.org/10.1016/j.molcel.2012.03.018
272. Wiktor, J., Lesterlin, C., Sherratt, D. J., & Dekker, C. (2016). CRISPR-mediated control of the bacterial initiation of replication. *Nucleic Acids Research*, *44*(8), 3801–3810. https://doi.org/10.1093/nar/gkw214
273. Wilbie, D., Walther, J., & Mastrobattista, E. (2019). Delivery Aspects of CRISPR/Cas for in Vivo Genome Editing. *Accounts of Chemical Research*, *52*(6), 1555–1564. https://doi.org/10.1021/acs.accounts.9b00106
274. Wilkinson, M., Drabavicius, G., Silanskas, A., Gasiunas, G., Siksnys, V., & Wigley, D. B. (2019). Structure of the DNA-Bound Spacer Capture Complex of a Type II CRISPR-Cas System. *Molecular Cell*, *75*(1), 90-101.e5. https://doi.org/10.1016/j.molcel.2019.04.020
275. Wu, D., Guan, X., Zhu, Y., Ren, K., & Huang, Z. (2017). Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA. *Cell Research*, *27*(5), 705–708. https://doi.org/10.1038/cr.2017.46
276. Wu, W. Y., Mohanraju, P., Liao, C., Adiego-Pérez, B., Creutzburg, S. C. A., Makarova, K. S., Keessen, K., Lindeboom, T. A., Khan, T. S., Prinsen, S., Joosten, R., Yan, W. X., Migur, A., Laffeber, C., Scott, D. A., Lebbink, J. H. G., Koonin, E. V., Beisel, C. L., & Oost, J. van der. (2022). The miniature CRISPR-Cas12m effector binds DNA to block transcription. *Molecular Cell*, *82*(23), 4487-4502.e7. https://doi.org/10.1016/j.molcel.2022.11.003
277. Wu, Z., Yang, H., & Colosi, P. (2010). Effect of Genome Size on AAV Vector Packaging. *Molecular Therapy*, *18*(1), 80–86. https://doi.org/10.1038/mt.2009.255

278. Wu, Z., Zhang, Y., Yu, H., Pan, D., Wang, Y., Wang, Y., Li, F., Liu, C., Nan, H., Chen, W., & Ji, Q. (2021). Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nature Chemical Biology*, *17*(11), 1132–1138. https://doi.org/10.1038/s41589-021-00868-6
279. Xiao, R., Li, Z., Wang, S., Han, R., & Chang, L. (2021). Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR–Cas12f nuclease. *Nucleic Acids Research*, *49*(7), 4120–4128. https://doi.org/10.1093/nar/gkab179
280. Xiao, R., Wang, S., Han, R., Li, Z., Gabel, C., Mukherjee, I. A., & Chang, L. (2021). Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition. *Molecular Cell*, *81*(21), 4457-4466.e5. https://doi.org/10.1016/j.molcel.2021.07.043
281. Xiao, Y., Ng, S., Nam, K. H., & Ke, A. (2017). How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature*, *550*(7674), 137–141. https://doi.org/10.1038/nature24020
282. Xin, C., Yin, J., Yuan, S., Ou, L., Liu, M., Zhang, W., & Hu, J. (2022). Comprehensive assessment of miniature CRISPR-Cas12f nucleases for gene disruption. *Nature Communications*, *13*(1), 5623. https://doi.org/10.1038/s41467-022-33346-1
283. Xu, H., Wang, B., Ono, M., Kagita, A., Fujii, K., Sasakawa, N., Ueda, T., Gee, P., Nishikawa, M., Nomura, M., Kitaoka, F., Takahashi, T., Okita, K., Yoshida, Y., Kaneko, S., & Hotta, A. (2019). Targeted Disruption of HLA Genes via CRISPR-Cas9 Generates iPSCs with Enhanced Immune Compatibility. *Cell Stem Cell*, *24*(4), 566-578.e7. https://doi.org/10.1016/j.stem.2019.02.005
284. Xu, L., Wang, J., Liu, Y., Xie, L., Su, B., Mou, D., Wang, L., Liu, T., Wang, X., Zhang, B., Zhao, L., Hu, L., Ning, H., Zhang, Y., Deng, K., Liu, L., Lu, X., Zhang, T., Xu, J., … Chen, H. (2019). CRISPR-Edited Stem Cells in a Patient with HIV and Acute Lymphocytic Leukemia. *New England Journal of Medicine*, *381*(13), 1240–1247. https://doi.org/10.1056/nejmoa1817426
285. Xu, X., Chemparathy, A., Zeng, L., Kempton, H. R., Shang, S., Nakamura, M., & Qi, L. S. (2021). Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Molecular Cell*, *81*(20), 4333-4345.e4. https://doi.org/10.1016/j.molcel.2021.08.008
286. Yamada, M., Watanabe, Y., Gootenberg, J. S., Hirano, H., Ran, F. A., Nakane, T., Ishitani, R., Zhang, F., Nishimasu, H., & Nureki, O. (2017). Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Molecular Cell*, *65*(6), 1109-1121.e3. https://doi.org/10.1016/j.molcel.2017.02.007
287. Yan, W. X., Chong, S., Zhang, H., Makarova, K. S., Koonin, E. V., Cheng, D. R., & Scott, D. A. (2018). Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. *Molecular Cell*, *70*(2), 327-339.e5. https://doi.org/10.1016/j.molcel.2018.02.028
288. Yan, W. X., Hunnewell, P., Alfonse, L. E., Carte, J. M., Keston-Smith, E., Sothiselvam, S., Garrity, A. J., Chong, S., Makarova, K. S., Koonin, E. V., Cheng, D. R., & Scott, D. A. (2019). Functionally diverse type V CRISPR-Cas systems. *Science*, *363*(6422), 88–91. https://doi.org/10.1126/science.aav7271

289. Yang, H., Gao, P., Rajashankar, K. R., & Patel, D. J. (2016). PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell*, *167*(7), 1814-1828.e12. https://doi.org/10.1016/j.cell.2016.11.053
290. Yang, Y., Wang, L., Bell, P., McMenamin, D., He, Z., White, J., Yu, H., Xu, C., Morizono, H., Musunuru, K., Batshaw, M. L., & Wilson, J. M. (2016). A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nature Biotechnology*, *34*(3), 334–338. https://doi.org/10.1038/nbt.3469
291. Yeh, C. D., Richardson, C. D., & Corn, J. E. (2019). Advances in genome editing through control of DNA repair pathways. *Nature Cell Biology*, *21*(12), 1468–1478. https://doi.org/10.1038/s41556-019-0425-z
292. Yin, H., Song, C.-Q., Suresh, S., Kwan, S.-Y., Wu, Q., Walsh, S., Ding, J., Bogorad, R. L., Zhu, L. J., Wolfe, S. A., Koteliansky, V., Xue, W., Langer, R., & Anderson, D. G. (2018). Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nature Chemical Biology*, *14*(3), 311–316. https://doi.org/10.1038/nchembio.2559
293. Yoganand, K. N., Muralidharan, M., Nimkar, S., & Anand, B. (2019). Fidelity of prespacer capture and processing is governed by the PAM-mediated interactions of Cas1-2 adaptation complex in CRISPR-Cas type I-E system. *Journal of Biological Chemistry*, *294*(52), 20039–20053. https://doi.org/10.1074/jbc.ra119.009438
294. Yusof, N. A., Hashim, N. H. F., & Bharudin, I. (2021). Cold Adaptation Strategies and the Potential of Psychrophilic Enzymes from the Antarctic Yeast, Glaciozyma antarctica PI12. *Journal of Fungi*, *7*(7), 528. https://doi.org/10.3390/jof7070528
295. Zetsche, B., Abudayyeh, O. O., Gootenberg, J. S., Scott, D. A., & Zhang, F. (2020). A Survey of Genome Editing Activity for 16 Cas12a Orthologs. *The Keio Journal of Medicine*, *69*(3), 59–65. https://doi.org/10.2302/kjm.2019-0009-oa
296. Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., & Zhang, F. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, *163*(3), 759–771. https://doi.org/10.1016/j.cell.2015.09.038
297. Zetsche, B., Volz, S. E., & Zhang, F. (2015). A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature Biotechnology*, *33*(2), 139–142. https://doi.org/10.1038/nbt.3149
298. Zhang, B., Lin, J., Perčulija, V., Li, Y., Lu, Q., Chen, J., & Ouyang, S. (2022). Structural insights into target DNA recognition and cleavage by the CRISPR-Cas12c1 system. *Nucleic Acids Research*, *50*(20), 11820–11833. https://doi.org/10.1093/nar/gkac987
299. Zhang, H., Li, Z., Xiao, R., & Chang, L. (2020). Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nature Structural & Molecular Biology*, *27*(11), 1069–1076. https://doi.org/10.1038/s41594-020-0499-0
300. Zhang, X., Zhu, B., Chen, L., Xie, L., Yu, W., Wang, Y., Li, L., Yin, S., Yang, L., Hu, H., Han, H., Li, Y., Wang, L., Chen, G., Ma, X., Geng, H., Huang, W., Pang, X., Yang, Z., … Li, D. (2020). Dual base editor catalyzes

both cytosine and adenine base conversions in human cells. *Nature Biotechnology*, *38*(7), 856–860. https://doi.org/10.1038/s41587-020-0527-y

301. Zhao, Y., Liu, S., Jiang, B., Feng, Y., Zhu, T., Tao, H., Tang, X., & Liu, S. (2018). Genome-Centered Metagenomics Analysis Reveals the Symbiotic Organisms Possessing Ability to Cross-Feed with Anammox Bacteria in Anammox Consortia. *Environmental Science & Technology*, *52*(19), 11285–11296. https://doi.org/10.1021/acs.est.8b02599
302. Zhi, S., Chen, Y., Wu, G., Wen, J., Wu, J., Liu, Q., Li, Y., Kang, R., Hu, S., Wang, J., Liang, P., & Huang, J. (2022). Dual-AAV delivering split prime editor system for in vivo genome editing. *Molecular Therapy*, *30*(1), 283–294. https://doi.org/10.1016/j.ymthe.2021.07.011
303. Zhu, S., Li, W., Liu, J., Chen, C.-H., Liao, Q., Xu, P., Xu, H., Xiao, T., Cao, Z., Peng, J., Yuan, P., Brown, M., Liu, X. S., & Wei, W. (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nature Biotechnology*, *34*(12), 1279–1286. https://doi.org/10.1038/nbt.3715
304. Zhuo, C., Zhang, J., Lee, J.-H., Jiao, J., Cheng, D., Liu, L., Kim, H.-W., Tao, Y., & Li, M. (2021). Spatiotemporal control of CRISPR/Cas9 gene editing. *Signal Transduction and Targeted Therapy*, *6*(1), 238. https://doi.org/10.1038/s41392-021-00645-w
305. Zuckerkandl, E., & Pauling, L. (1965). Evolving Genes and Proteins. *PART III: EVOLUTION OF PROTEINS I*, 97–166. https://doi.org/10.1016/b978-1-4832-2734-4.50017-6
306. Zuo, Z., & Liu, J. (2016). Cas9-catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations. *Scientific Reports*, *5*(1), 37584. https://doi.org/10.1038/srep37584
307. Zuo, Z., & Liu, J. (2017). Structure and Dynamics of Cas9 HNH Domain Catalytic State. *Scientific Reports*, *7*(1), 17271. https://doi.org/10.1038/s41598-017-17578-6

APPENDICES

Appendix 1. Cas12l DNA and RNA substrates.

Name	Details	Figures	Sequence
CCC_W1	Fluorescently (5' 6-FAM (NTS) ; 5' 6-ROX (TS)) labelled dsDNA duplex W1 target, CCC PAM	3.11; 3.12D	ACCAGCAGGACTACAGCTTCCCCACAGTTCGATTACCTTTCCCACTCAGCCT TTTTGTGGGTGTACGTTTTGG
CCC_R1	Fluorescently (5' 6-FAM (NTS) ; 5' 6-ROX (TS)) labelled dsDNA duplex R1 target, CCC PAM	3.11; 3.12D	ACCAGCAGGACTACAGCTTCCCCGCCTTCAGAAGAGGGTGCATTTTCAGCCT TTTTGTGGGTGTACGTTTTGG
CCT_R1	Fluorescently (5' 6-FAM (NTS) ; 5' 6-ROX (TS)) labelled dsDNA duplex R1 target, CCT PAM	3.11	ACCAGCAGGACTACAGCTTCCCTGCCTTCAGAAGAGGGTGCATTTTCAGCCT TTTTGTGGGTGTACGTTTTGG
CCA_R1	Fluorescently (5' 6-FAM (NTS) ; 5' 6-ROX (TS)) labelled dsDNA duplex R1 target, CCA PAM	3.11	ACCAGCAGGACTACAGCTTCCCAGCCTTCAGAAGAGGGTGCATTTTCAGCCT TTTTGTGGGTGTACGTTTTGG
CCG_R1	Fluorescently (5' 6-FAM (NTS) ; 5' 6-ROX (TS)) labelled dsDNA duplex R1 target, CCG PAM	3.11	ACCAGCAGGACTACAGCTTCCCGGCCTTCAGAAGAGGGTGCATTTTCAGCCT TTTTGTGGGTGTACGTTTTGG

CAC_R1	Fluorescently (5' 6-FAM (NTS) ; 5' 6-ROX (TS)) labelled dsDNA duplex R1 target, CAC PAM	3.11	ACCAGCAGGACTACAGCTTCCACGCCTTCAGAAGAGGGTGCATTTTCAGCCT TTTTGTGGGTGTACGTTTTGG
ACC_R1	Fluorescently (5' 6-FAM (NTS) ; 5' 6-ROX (TS)) labelled dsDNA duplex R1 target, ACC PAM	3.11	ACCAGCAGGACTACAGCTTCACCGCCTTCAGAAGAGGGTGCATTTTCAGCCT TTTTGTGGGTGTACGTTTTGG
CCC_R1	dsDNA activator for Asp2Cas12l trans cleavage, R1 target, CCC PAM	3.13; 3.14	ACCAGCAGGACTACAGCTTCCCCGCCTTCAGAAGAGGGTGCATTTTCAGCCT TTTTGTGGGTGTACGTTTTGG
CCC_W1	dsDNA activator for Asp2Cas12l trans cleavage, W1 target CCC PAM	3.13A-B	ACCAGCAGGACTACAGCTTCCCCACAGTTCGATTACCTTTCCCACTCAGCCT TTTTGTGGGTGTACGTTTTGG
R1_TS	ssDNA activator for Asp2Cas12l trans cleavage, R1 target	3.13; 3.14	CTGAAAATGCACCCTCTTCTGAAGGCGGGGAAGCTGTAGTCCTGCTGGT
W1_TS	ssDNA activator for Asp2Cas12l trans cleavage, W1 target	3.13A-B	GTGAGTGGGAAAGGTAATCGAACTGTGGGGAAGCTGTAGTCCTGCTGGT
CZ-2901	ssDNA reporter 5' 6-FAM; 3' Iowa Black® FQ	3.13; 3.14	CCCCCCCC
CZ-3295	ssRNA FQ reporter 5'-6-FAM; 3'-IABkFQ	3.13B-C	CCCCCCCC
R1 900bp	Linear dsDNA substrate R1 target	3.12A	https://benchling.com/s/seq-fqxyiTkGPCpRXnm8D47V?m=slm-mERYE4SvSjKv4AJ5Czj5

RunXI_2kb	Linear dsDNA substrate R1-R6 targets	3.12B-C	https://benchling.com/s/seq-zBmRqeyLVCgbka3H2ACa?m=slm-GZorptjH213nyolLmGDI
WTAP_2kb	Linear dsDNA substrate W1-W6 targets	3.12B-C	https://benchling.com/s/seq-rFd2T6lCPLlCiPeiLUQR?m=slm-HJvKruM7VxK3QWa5XqfS

Appendix 2. Cas12l expression plasmids.

Name	Details	Figures	Sequence
petduet-cas12lbeta-native-crispr_locus	Plasmid interference in E. coli Asp2Cas12l locus T7 expression plasmid; wt1-3 spacers	3.15	https://benchling.com/s/seq-EW3oVUDnNkX52s4tIZ1h?m=slm-mWhmeakGBxdXiErIj8mp
petduet_cas12lbeta-sgrna-R1	Plasmid interference in E. coli Asp2Cas12l & sgRNA T7 expression plasmid; R1 spacer	3.15	https://benchling.com/s/seq-BECj5nOotCwO6FHSeXDK?m=slm-lFIISUCj4rKfEBQIyBJO
pet-mbp14hissumo-cas12lbeta	Asp2Cas12l with solubility and affinity chromatography tags; T7 expression		https://benchling.com/s/seq-wjRAqt3Nr7gPX2LYwelw?m=slm-HVxuqJbB5Nt2MeETcnQ4
pet-mbp14hissumo-cas12lbeta-d518a-ruvc-mut	Asp2Cas12l (D518A) with solubility and affinity chromatography tags; T7 expression		https://benchling.com/s/seq-Akuh6etzc8QF0VBBUGyq?m=slm-4XdBk5veB5T6ZxR778XQ
pet-mbp14hissumo-cas12lbeta-e624a-ruvc-mut	Asp2Cas12l (E624A) with solubility and affinity chromatography tags; T7 expression		https://benchling.com/s/seq-XEDvYQHVIcdEama0M6Wp?m=slm-evkbGJ6JGHGaZ1PloKwc
pet-mbp14hissumo-cas12lbeta-d793a-ruvc-mut	Asp2Cas12l (D793A) with solubility and affinity chromatography tags; T7 expression		https://benchling.com/s/seq-4ScG6Ij8oGSgvTVQJIZ2?m=slm-DYVswCCaQBwdtwGyyOLM

pet-mbp14hissumo-cas12lgamma	Asp3Cas12l with solubility and affinity chromatography tags; T7 expression	https://benchling.com/s/seq-Uqz2SymbHU8KBdaoySLH?m=slm-BvsCsfICCZefdunqx58g
pet-mbp14hissumo-cas12lgamma-d537a-ruvc-mut	Asp3Cas12l (D537A) with solubility and affinity chromatography tags; T7 expression	https://benchling.com/s/seq-CpGyjerGkHPPZ84rUDOL?m=slm-mKZbAPbhDMhis1CFnEdF
pet-mbp14hissumo-cas12lgamma-e643a-ruvc-mut	Asp3Cas12l (E643A) with solubility and affinity chromatography tags; T7 expression	https://benchling.com/s/seq-dN0w0SvlCCXVP1gfCVxP?m=slm-TWaycdNRVzeaBMti7eXG
pet-mbp14hissumo-cas12lgamma-d796a-ruvc-mut	Asp3Cas12l (D793A) with solubility and affinity chromatography tags; T7 expression	https://benchling.com/s/seq-gAOroycH4HO7tVo8tBL1?m=slm-Rg0lT5r4OEzusYPrapnd

Appendix 3. Cas12l target plasmids.

Name	Details	Figures	Sequence
pTZ57_7NPAM library	7N randomized PAM plasmid library	3.1; 3.3; 3.9; 3.11A	https://benchling.com/s/seq-hkJm7gQGpvaT1fmpzwdk?m=slm-DcAJ0C6oG11lldlXwNko
pUC19_R1	Plasmid substrate; CCC PAM R1 protospacer	3.12C	https://benchling.com/s/seq-mzOAzAKa0quurwwGOI24?m=slm-XRThjYYge9E4yUujL9aZ
pUC19_W1	Plasmid substrate; CCC PAM W1 protospacer	3.12C	https://benchling.com/s/seq-g9zYovcmAiVAdHdpKaF9?m=slm-NXh0nDHbKTBujqcARkET
pUC19_T2	Plasmid substrate; CCC PAM T2 protospacer	3.10C	https://benchling.com/s/seq-vIpeTSPO0Bcv9KDQwPT4?m=slm-bJP09JqcfuQSwhftrtES
pCDFDuet_CCC-wt1	Plasmid interference in E. coli target plasmid; CCC PAM wt1 protospacer	3.15	https://benchling.com/s/seq-Mw72CYu24koXLdkTQouF?m=slm-kFSIcaZvdhVBTrOLDNwb
pCDFDuet_CCC-wt2	Plasmid interference in E. coli target plasmid; CCC PAM wt2 protospacer	3.15	https://benchling.com/s/seq-cwypYzD8K2EyylNICcRV?m=slm-G6vUUmhspwfYHv0W76Hr
pCDFDuet_CCC-wt3	Plasmid interference in E. coli target plasmid; CCC PAM wt3 protospacer	3.15	https://benchling.com/s/seq-tqKyMU3EDWaKntPpW91a?m=slm-LlT6lOTycRt0Wbo6Mu2E
pCDFDuet_CCC-R1	Plasmid interference in E. coli target plasmid; CCC PAM R1 protospacer	3.15	https://benchling.com/s/seq-YSADP3TBNqcIFKWvWARI?m=slm-cx4CskMdxSs9bQew9hUO

SANTRAUKA

SANTRUMPOS

Asp	*Armatimonadetes* rūšis
AAV	adeno-asocijuotas virusas
BH	angl. *bridge helix*, tiltinė spiralė
bp	bazių pora
Cas	angl. *CRISPR-associated*, asocijuotas su CRISPR
CRISPR	angl. *clustered regularly interspaced short palindromic repeats*
crRNR	CRISPR RNR
CTD	C-galinis domenas
DNase	deoksiribonukleazė
dg	dvigrandinis
DSB	dvigrandinis trūkis
DTT	ditiotreitolis
FLL	angl. *full length linear*, pilno ilgio linijinis fragmentas
gRNR	angl. *guide RNA*, vedančioji RNR
HDR	angl. *homology-directed recombination*, homologija paremta DNR rekombinacija
His	histidinas
HMM	angl. hidden Markov models
IPTG	isopropil-β-D-1-tiogalaktopiranozidas
IVT	in vitro transliacija
LB	Luria-Bertani terpė
MBP	maltozę surišantis baltymas
NHEJ	angl. *non-homologous end-joining*, nehomologinis DNR galų sujungimas
nt	nukleotidas
NTG	ne taikinio grandinė
NUC	nukleazės skiltis
OC	angl. *open circle*, atviro žiedo
PAGE	poliakrilamido gelio elektroforezė
PAM	angl. *protospacer adjacent motif*
PCR	polimerazės grandininė reakcija
PFM	angl. *position frequency matrix*, padėties dažnio matrica
PFS	angl. *protospacer flanking site*
PI	angl. *PAM-interacting*, su PAM sąveikaujantis
PMSF	fenilmetilsulfonil fluoridas

pre-crRNR	angl. *precursor crRNA*, pirmtakinė crRNR
REC	angl. *recognition lobe*, atpažinimo skiltis
RNAse	ribonukleazė
RNP	ribonukleobaltymas
SC	superspiralizuota
SDS	natrio dodecil sulfatas
sgRNR	angl. *single-guide RNA*, pavienė gidinė RNR
vg	viengrandinė
SUMO	angl. *small ubiquitin-like modifier*
tracrRNR	angl. *trans-activating CRISPR RNA,* trans-aktyvuojanti CRISPR RNR
Tris	tris(hidroksimetil)aminometanas
TG	taikinio grandinė

ĮVADAS

Prokariotinės CRISPR (angl. *clustered regularly interspaced palindromic repeats*) – Cas (angl. *CRISPR associated*) sistemos bakterijose ir archėjose yra prisitaikančios gynybos nuo pašalinių nukleorūgščių, virusų ir mobilių genomo elementų pavidalu, forma (Barrangou et al., 2007). Ši apsauga suteikiama surenkant invazinės DNR fragmentus ir integruojant juos į CRISPR regioną, nuo kurio transkribuojamos CRISPR RNR molekulės, kurios sudaro kompleksus su Cas baltymais ir nutaiko juos hidrolizuoti komplementarias nukleorūgštis per pakartotines infekcijas (Barrangou et al., 2007; Mojica et al., 2005). Pakeičiant vedančiosios crRNR seką, Cas nukleazės teoriškai gali būti nukreipiamos į bet kokią nukleorūgščių seką – ši savybė lėmė šių fermentų pritaikymą genomų redagavimui (Cong et al., 2013; Gasiunas et al., 2012; Jinek et al., 2012). Tačiau, realus Cas baltymams pasiekiamas taikinių spektras yra ribojamas šalia taikinio aptinkamos ir atpažinimui reikalingos trumpos nukleorūgščių sekos, vadinamos PAM (angl. *protospacer adjacent motif*), skirtingos kiekvienam Cas baltymui (Mojica et al., 2009; Shah et al., 2013). Taipogi, didelis dažniausiai genomų redagavimui naudojamų Cas baltymų dydis apsunkina šių sistemų pristatymą į tikslines ląsteles, ypač naudojant įprastą adeno-asocijuotų virusų transporto metodą (Wilbie et al., 2019). Turint omenyje, jog CRISPR-Cas sistemos yra plačiai paplitusios prokariotuose ir stebimos ~85% archėjų bei ~40% bakterijų genomuose, ši natūrali įvairovė gali būti pasitelkta ieškant naujų CRISPR-Cas nukleazių potencialiai pasižyminčių plataus spektro struktūrinėmis bei biocheminėmis savybėmis (Makarova et al., 2020).

Pagrindinis šios disertacijos objektas – naujos II bei V tipų CRISPR-Cas nukleazės. **Tikslas** buvo atrasti bei charakterizuoti naujus fermentus, pasižyminčius biocheminėmis ir struktūrinėmis savybėmis, palankiomis genų redagavimo reikmėms, kaip naujų PAM sekų atpažinimas, reikšmingas DNR hidrolizės aktyvumas bei mažesnis dydis. Tikslui pasiekti buvo išsikelti šie **uždaviniai:**

1. Identifikuoti rinkinį II tipo CRISPR-Cas nukleazių, patikrinti jų gebėjimą hidrolizuoti dgDNR bei nustatyti jų vedančiųjų RNR ir PAM sekų reikalavimus.
2. Atkurti ir charakterizuoti II tipo CRISPR-Cas nukleazių kompleksus *in vitro*.
3. Identifikuoti V tipo CRISPR-Cas nukleazes, patikrinti jų gebėjimą hidrolizuoti dgDNR bei nustatyti jų vedančiųjų RNR ir PAM sekų reikalavimus.

4. Atkurti ir charakterizuoti V tipo CRISPR-Cas nukleazių kompleksus *in vitro.*

Mokslinis naujumas ir praktinė vertė:

Šis mokslinis darbas apima naujų II bei V tipo CRISPR-Cas nukleazių identifikavimą bei charakterizavimą.

Darbas buvo pradėtas biochemiškai įvertinant didžiausią ligšiol imtį spėjamų Cas9 nukleazių, ko pasekoje identifikavome funkcionalius Cas9 baltymus, atpažįstančius įvairių konfiguracijų G, C, A, ir T nukleotidais praturtintas PAM sekas, iš esmės praplečiant Cas9 baltymais pasiekiamų taikinių galimybes. Vienu atveju, šiame tyrime identifikuotas Mga Cas9 ortologas buvo pritaikytas vieno nukleotido polimorfizmo (angl. single nucleotide polymorphism (SNP)) detekcijai dėka nukleazės atpažįstamos PAM sekos persidengimo su aktualiu SNP. Tai lėmė jautresnę diskriminaciją, negu kuomet tikslinis SNP buvo lokalizuotas Cas9 taikinio srityje, pabrėžiant privalumą, suteikiamą plataus spektro PAM sekų atpažinimo. Nors didžioji dalis aptiktų ortologų atpažino PAM sekas ilgesnes nei 2 bp, kaip būdinga kanoninei *Streptococcus pyogenes* Cas9 nukleazei, tai galėtų būti naudinga genomų redagavimo aplikacijoms, turint omenyje, jog Cas9 ortologai su ilgesnėmis atpažįstamomis PAM sekomis gali sąlygoti didesnį aktyvumo specifiškumą (C. M. Lee et al., 2016; Müller et al., 2016). Tolimesnis tikslinių baltymų DNR hidrolizės aktyvumo biocheminis įvertinimas parodė daugiau skirtumų tarp ortologų. Tarp jų buvo įvairus temperatūros poveikis kirpimo efektyvumui. Iš visos imties išsiskyrė Cme2 Cas9, kuris buvo aktyvus tik 30 °C - 55 °C temperatūrų ruože, sudarantis galimybę riboti DNR modifikavimo aktyvumą pasitelkiant temperatūrų kontrolę (F. Richter et al., 2016; Zhuo et al., 2021). Taipogi, Nsa ir Ssa Cas9 baltymų DNR hidrolizės aktyvumas skirtingose temperatūrose rodo jų potencialą būti panaudotiems atitinkamai termo- ar psichrofilinių organizmų kontekste. Toliau mes išskyrėmė Cas9 ortologus, pasižyminčius potencialiai naudingomis savybėmis lyginant su *S. pyogenes* Cas9. Tarp jų buvo baltymai po DNR hidrolizės paliekantys skirtingo ilgio DNR grandinių galus, kas gali lemti distinktyvius dvigrandžio trūkio reparacijos kelius (Y.-W. Fu et al., 2021), bei baltymai rodantys poreikį ilgesnei gRNR ir DNR taikinio homologijai. Galop, parodyta, jog kai kurie mažesni, apie 1100 aa ilgio, Cas9 ortologai, kaip Nsa ir Tsp Cas9, išlaiko aukštą DNR hidrolizės aktyvumą ir todėl galėtų būti pranašesni, kuomet didelių Cas9 baltymų transportas yra sudėtingas (E. Kim et al., 2017).

Sekančioje tyrimo dalyje mes identifikavome naują V tipo CRISPR-Cas nukleazių šeimą. Naujieji Cas12l fermentai yra ~860 aa ilgio, t.y. ženkliai trumpesni už plačiai naudojamus Cas9 ir Cas12a baltymus ir pralenkiami tik V-F ir V-J potipių sistemų (Bigelyte et al., 2021; Pausch et al., 2020). Cas12l baltymai sąlygoja DNR hidrolizę 5‘ C-praturtintos PAM sekos aplinkoje. Nors aptinkamos CRISPR-Cas9 nukleazės, tarp jų keletas identifikuotų ir šiame tyrime, atpažįstančios C-praturtintas PAM sekas, ši Cas12l tendencija būtų reikšminga atsvara tarp V tipo Cas12 baltymų paplitusiam T-praturtintų PAM sekų atpažinimui. Taipogi, parodyta, jog Cas12l baltymai nespecifiškai degraduoja vgDNR ir vgRNR po taikinio atpažinimo – savybė, kuri buvo adaptuota nukleorūgščių detekcijos sistemose naudojant kitus Cas12 baltymus (Gootenberg et al., 2018; J. Joung et al., 2020). Kadangi optimalus Cas12l hidrolizės aktyvumas stebimas temperatūrose aplink 50 °C, šios nukleazės galėtų pasitarnauti izotermininiuose metoduose, apjungiant nukleorūgščių amplifikaciją ir detekciją (J. Joung et al., 2020). Galiausiai, Asp2Cas12l lėmė *Escherichia coli* transformacijos plazmidine DNR ribojimą. Tai leidžia manyti, jog šis naujo tipo efektorius galėtų būti pritaikytas redagavimui ir kituose ląstelių tipuose.

Pagrindiniai ginamieji disertacijos teiginiai:

1. 79 filogenetiškai skirtingi Cas9 ortologai pasižymi dgDNR hidrolizės aktyvumu, lemiamu įvairių PAM sekų bei vedančiųjų RNR reikalavimais.
2. Cas9 ortologai pasižymi įvairiomis biocheminėmis savybėmis, kaip temperatūros bei vedančiosios RNR skirtuko ilgio įtaka aktyvumui bei po hidrolizės sudaromų DNR grandinių galų ilgis.
3. Nauja V tipo CRISPR-Cas nukleazių šeima, Cas12l, vykdo nuo C-praturtintos PAM sekos ir dvigubų vedančiųjų RNR priklausomą dgDNR hidrolizę.
4. Cas12l lemia nespecifinę vgDNR ir vgRNR degradaciją po taikinio atpažinimo.
5. Asp2Cas12l raiška *Escherichia coli* riboja transformaciją plazmidine DNR.

Cas9 ortologų identifikavimas

II tipo Cas9 endonukleazės identifikuotos ieškant CRISPR regionų naudojant PILER-CR 1.06 (Edgar, 2007a). Aptikus CRISPR, aplinkinės DNR sekos (15 kb į abi puses nuo CRISPR regiono) tikrinti ieškant atviro skaitymo rėmelių (ORF), koduojančių baltymus ilgesnius nei 750 aminorūgčių. Toliau, norint identifikuoti *cas* genus, koduojančius Cas9 ortologus, sudaryti sekų palyginiai iš plačios Cas9 baltymų kolekcijos naudojant MUSCLE 3.8.31 (Edgar, 2004a) ir panaudoti kuriant užslėptų Markov modelių (HMM) profilius Cas9 subšeimoms pasitelkiant HMMER 3.2.1 (Edgar, 2004a). Gautieji HMM profiliai naudoti baltymų sekų, transliuojamų nuo spėjamų *cas* ORF ir homologiškų Cas9, paieškai. Kitu būdu, Cas9 ortologai ir metagenominės juos koduojančios sekos gauti iš viešai prieinamų duomenų bazių https://img.jgi.doe.gov/cgi-bin/m/main.cgi (I.-M. A. Chen et al., 2019). Tik baltymai turintys Cas9 baltymams charakteringus HNH ir RuvC nukleazinius domenus ir katalitines aminorūgštis buvo atrenkami. Filogenetinės analizės (MEGA7 10.0.5 (Kumar et al., 2016)) būdu Cas9 baltymai išskirstyti į skirtingas šeimas ir reprezentatyvūs atstovai iš kiekvienos grupės panaudoti charakterizavimo atrankai. Iš atsirinktų baltymų ir II-A, II-B bei II-C potipių reprezentatyvių atstovų sudarytas filogenetinis medis panaudojant MEGA7 ir Neighbor-Joining (Saitou & Nei, 1987) ir Poisson correction (Zuckerkandl & Pauling, 1965) metodus.

Spėjamų Cas9 vedančiųjų RNR dizainas

Trans-aktyvuojančios CRISPR RNR (tracrRNR) būtinos CRISPR RNR (crRNR) brendimui (Deltcheva et al., 2011) ir Cas9 medijuojamui taikinio srities kirpimui buvo aptikos ieškant regionų *cas9* geno aplinkoje, kurie pasižymėtų komplementarumu CRISPR pasikartojimams. Aptikus šias sritis, potencialios spėjamų tracrRNR transkripcijos sritys buvo nustatytos tikrinant spėjamas sudaromas antrines RNR struktūras naudojant UNAfold 3.9 (Markham & Zuker, 2008) bei potencialių transkripcijos terminacijos signalų buvimą. Pagal tikėtinas tracrRNR ir CRISPR RNR transkripcijos kryptis buvo suprojektuotos pavienės gidinės RNR (sgRNR) sujungiant crRNR ir tracrRNR sekas (Jinek et al., 2012). 16 nukleotidų iš crRNR pasikartojimo buvo prilieta prie komplementaraus tracrRNR anti-pasikartojimo per 4 nt ilgio GAAA jungtuką (Jinek et al., 2012).

Skaitmeninė Cas9 tracrRNR analizė

Sekų, homologiškų 79 identifikuotoms tracrRNR, paieškai panaudota BLAST 2.7.3 (naudojant parametrus optimizuotus trumpų sekų pasikartojančiuose regionuose paieškai (-task blastn_short -dust no)) (Altschul et al., 1990a). Gauta identifikuotų sekų kolekcija grupuota naudojant CD-HIT 4.7 (L. Fu et al., 2012) su 90% sekų panašumo atskirties tašku. Iš gautų klasterių pašalintos grupės, neturinčios bent vienos iš 79 tracrRNR sekos. Toliau, sekų homologijos ir antrinės RNR struktūros modeliai sukonstruoti naudojant MAFFT 7.407 (Katoh & Standley, 2013) ir RNAalifold 2.4.5 (Lorenz et al., 2011). Abu modeliai naudoti homologų sekų ar struktūriniame lygmeniuose paieškai tarp visų BLAST dėka identifikuotų variantų naudojant RNR struktūrų paieškos įrankius Infernal 1.1 programinėje įrangoje (Nawrocki & Eddy, 2013). Struktūriniai persidengimai tarp klasterių sugeneruoti lyginant kiekvieno kovariacijos modelio (CM) rezultatus. Norint grafiškai pateikti tracRNR gimiringumą, kiekvienam CM priskirti taškai. Jei keletas taškų buvo priskirti vienam CM, jie buvo apjungtos tiesėmis, kurių storis priklauso nuo konkrečių taškų sekų panašumo pagal šią formulę ($panašumas = \frac{\#\ priskirtų\ sekų}{\min(\#sekų\ aptiktų\ 1\ modeliu,\ \#sekų\ aptiktų\ 2\ modeliu)}$)

sgRNR sintezė

Visos sgRNR molekulės susintetintos in vitro transkripcijos būdu naudojant HiScribe™ T7 Quick High Yield RNR Synthesis Kits (New England Biolabs) rinkinius, arba transkribuotos kartu su baltymų genais in vitro transliacijos (IVT) reakcijoje. Matricos sgRNR transkripcijai gautos PGR amplifikacijos nuo sintetinių DNR fragmentų (IDT arba Genscript) būdu arba prilydant T7 RNR polimerazės promotoriaus pradmenį prie viengrandinio matricinio oligonukleotido. Transkribuoti RNR produktai buvo paveikti DNase I (New England Biolabs) matricinės DNR pašalinimui, išvalyti Monarch RNR Cleanup Kit (50 µg) (New England Biolabs) rinkiniais ir eliuoti vandenyje. RNR koncentracija ir grynumas įvertintas NanoDrop spektrofotometru bei agarozinio gelio elektroforeze, vizuolizuojant RNR naudojant GelRed (Biotium) dažą.

PAM bibliotekos hidrolizė *in vitro* transliuotais Cas9

Cas9 baltymai susintetinti IVT reakcijose naudojant PURExpress bacterial IVT kit (New England Biolabs) rinkinius, pagal gamintojo nustatytas rekomendacijas ir panašiai, kaip aprašyta anksčiau (Karvelis et al., 2018a). DNR plazmidės, koduojančios E. coli kodonais optimizuotus Cas9 baltymus, naudotos kaip matricos IVT reakcijoms. Sintetiniai DNR fragmentai buvo susintetinti Twist Bioscience ir surinkti NEBuilder HiFi DNA Assembly kit (New England Biolabs) rinkiniais į pET28a (EMD Millipore) plazmides. Po IVT reakcijų, 20 µl supernatanto sumaišyta su RNRzių slopikliu Murine RNRse Inhibitor (40 U; New England Biolabs) ir 2 µg transkribuotos sgRNR ir inkubuoti 15 min kambario temperatūroje RNP kompleksų surinkimui. Kitu būdu, sgRNR buvo transkribuota iškart IVT reakcijoje į reakciją dedant matricinės DNR fragmentus, koduojančius sgRNR seką. Tokiu atveju, 0.5 µg cas9 geną koduojančios plazmidės ir 100-kartinis molinis sgRNR matricos perviršis buvo dedami į IVT reakcijos mišinį. 10 µl (arba serijiniai 10-kartiniai skiedimai) gautų Cas9-sgRNR kompleksų buvo apjungti su 1 µg 7 bp atsitiktinių PAM sekų biblioteka, aprašyta anksčiau (Karvelis 2015), 100 µl reakcijoje (reakcijos buferis 10 mM Tris-HCl pH 7.5 at 37°C, 100 mM NaCl, 10 mM MgCl2, 1 mM DTT) ir inkubuoti 60 min 37 °C temperatūroje.

PAM bibliotekos hidrolizės produktų paruošimas ir sekvenavimas

Kirpti bibliotekos fragmentai buvo surinkti prie jų priligavus oligonukleotidinį jungtuką, pagausinti PGR amplifikacija ir sekvenuoti, kaip aprašyta (Karvelis et al., 2015). Trumpai, hidrolizuotoms bibliotekoms pirmiausia buvo atliktas DNR galų taisymas inkubuojant su 0,3 µl (1U) T4 DNR polimerazės (New England Biolabs) ir 0,3 µl 10 mM dNTP mišinio (Thermo Fisher Scientific) 15 min. 12 °C temperatūroje ir inaktyvuojant kaitinant (75 °C 20 min.). Siekiant veiksmingai surinkti laisvus DNR galus, į reakcijos mišinį buvo pridėta 3'-dA iškyša, inkubuojant reakcijos mišinį su 0,3 µl (1,5 U) "DreamTaq" polimerazės (Thermo Fisher Scientific) 30 min. 72 °C temperatūroje. Tada gauta DNR buvo išgryninta (Monarch PCR & DNA Cleanup rinkiniu (New England Biolabs)) ir suliguota su jungtukais, turinčiais 3' dT iškyšą, naudojant 1 µl 400 U T4 ligazės (New England Biolabs) 25 µl ligavimo buferio (50 mM Tris-HCl, pH 7,5, 10 mM MgCl2, 10 mM DTT, 1 mM ATP, 5 % (w/v) PEG 4000). Po 1 val. kambario temperatūroje 10 µl ligavimo reakcijos buvo panaudota kaip matricinė DNR PGR reakcijoje (Q5 DNR polimerazė (New England Biolabs); 15 ciklų; 100 µl galutinis reakcijos tūris) su pradmenimis, komplementariais bibliotekos PAM pusei ir adapteriui. Po to DNR buvo išgryninta (Monarch PCR & DNA Cleanup rinkiniu (New England Biolabs)), o Illumina sekoskaitai reikalingos

sekos ir indeksai buvo pridėti dviem PGR raundais (Phusion High-Fidelity PCR Master Mix HF buferyje (New England Biolabs); 10 ciklų kiekviename raunde; 50 μL galutinis reakcijos tūris). Su gautais produktais buvo atlikta sekoskaita MiSeq Personal Sequencer (Illumina) su 25 % (v/v) PhiX control v3 (Illumina).

Cas baltymų atpažintų PAM sekų nustatymas

PAM sekos, palaikančios dvigubos grandinės DNR taikinio hidrolizę, buvo nustatytos, kaip aprašyta (Karvelis et al., 2015, 2018). Trumpai, po sekoskaitos pirmiausia buvo įvertinta hidrolizės vieta bibliotekos protoskirtuke, (algoritmas pateiktas https://github.com/cortevaCRISPR/Cas12f-InformaticsTools) įvertinant vietą, kurioje aptikta daugiausia jungtuko prisiligavimo įvykių. Tada buvo išskirtos PAM sekos, susietos su bibliotekos fragmentais, kurios palaikė skilimą, ir panaudotos bazių sudėties nuokrypiui kiekvienoje pozicijoje atsitiktinės atrankos PAM bibliotekoje, palyginti su pradinės bibliotekos sudėtimi, įvertinti normalizuojant pagal formulę: normalizuotas dažnis = (kirpimo dažnis)/((kontrolinis dažnis)/(vidutinis kontrolinis dažnis)). Toliau PAM pirmenybės buvo kiekybiškai įvertintos pasitelkiant padėties dažnio matricas (PFM) ir pavaizduotos kaip WebLogo. Siekiant sumažinti foninio triukšmo, atsirandančio dėl kitų IVT arba E. coli ląstelių lizato mišinių komponentų nespecifinio skilimo, poveikį, analizuota tik 10 % dažniausiai pasitaikančių PAM.

Cas9 raiška ir gryninimas

Streptococcus pyogenes, *Streptococcus thermophilus* CRISPR3 (Sth3) ir S. thermophilus CRISPR1 (Sth1) Cas9 baltymai, klonuoti į pBAD-Chis vektorius, buvo ekspresuojami E. coli DH10B kamiene 16 °C temperatūroje 20 val., esant 0,2 % (w/v) arabinozės. Kitų ortologų genai pirmiausia buvo optimizuoti E. coli kodonams ir klonuoti į pET28 vektorių, priliejant C-galinį 6xHis inkarą. Kai kuriais atvejais į Cas9 geno 5' ir 3' galus buvo įterptos sekos, koduojančios branduolio lokalizacijos sekas (SV40 kilmės). Tada kiekvieno ortologo raiška buvo išbandyta skirtinguose E. coli kamienuose (NiCo21(DE3), T7 Express lysY/Iq, NEB® Express Iq) įvairiomis augimo sąlygomis (terpė, temperatūra, indukcija), o susidariusių baltymų kiekis išmatuotas NDS-PAAGE analize. Tada buvo pasirinktos optimalios sąlygos gryninimui kolbos mastu. Ląstelės buvo suardytos ultragarsu. SupeRNRtantas buvo užneštas ant HiTrap DEAE Sepharose kolonėlės (GE Healthcare), po to gryninta Ni2+ įkrauta HiTrap cheluojančia HP kolonėle (GE Healthcare) ir

HiTrap Heparin HP ("GE Healthcare") kolonėlėmis. Išgryninti Cas9 baltymai saugoti -20 °C temperatūroje 20 mM Tris-HCl, pH 7,5, 500 mM KCl, 1 mM EDTA, 1 mM DTT ir 50 % (v/v) glicerolio buferyje.

Cas9 su PAM sąveikaujančių domenų skaitmeninė analizė

Iš čia aprašytų Cas9 ortologų sekų buvo sudarytas sekų palyginys naudojant MAFFT 7.407 (Katoh & Standley, 2013). Jų su PAM sąveikaujančios (PI) sritys, atitinkančios *S. pyogenes* Cas9 (4ZT0_A:1090-1365) C galinį domeną, buvo išskirtos ir panaudotos kaip užklausos dviem PSI-BLAST 2.2.26 (Altschul et al., 1997) paieškos iteracijoms NCBI NR baltymų kolekcijoje, UniRef100 ir Mgnify (Mitchell et al., 2019) duomenų bazėse. Sutapimai buvo išskirti, filtruoti iki 80 % tapatumo naudojant CD-HIT 4.6 (L. Fu et al., 2012) ir klasterizuoti naudojant CLANS 1.0 (Frickey & Lupas, 2004). CLANS yra paremntas Fruchtermano-Reingoldo jėgų nukreipto išdėstymo algoritmu, kuris baltymų sekas traktuoja kaip taškines mases virtualioje daugiamatėje erdvėje, kurioje jos traukia arba atstumia viena kitą pagal jų porinio panašumo stiprumą (CLANS P vertės). CLANS P vertės apskaičiuojamos iš BLAST E-verčių, jas dalijant iš naudojamos efektyvios paieškos erdvės. Gauti CLANS tinklai buvo vizualiai apžiūrėti ir nustatyti klasteriai. Didesnėms nei 150 sekų grupėms buvo atlikta filogenetinė analizė, atkuriant ankstesniame etape išfiltruotas sekas ir pašalinant identiškas sekas. Toliau 1-6 klasterių daugybiniai sekų palyginiai sudaryti naudojant MAFFT (parinktys: --ep 0.123 --maxiterate 20 --localpair), o regionai su spragomis pašalinti naudojant trimAL 1.2 (Capella-Gutiérrez et al., 2009) (parinktis: -gt 0.01). Gautų palyginių ilgis svyravo nuo 359 iki 652 aminorūgščių atitinkamai 2 ir 3 klasteriuose. Filogenetiniai medžiai sukurti naudojant IQtree 1.6.10 (Nguyen et al., 2015) su automatiniu modelio parinkimu ir 1000 greitųjų bootstrapų (parinktys: -alrt 1000 -bb 1000).

Protoskirtuko kirpimo vietų nustatymas

Norėdami užfiksuoti protoskirtuko kirpimo modelius vienos molekulės skiriamąja geba, sukūrėme minižiedinį dvigubos grandinės (ds) DNR substratą, kuris leidžia užfiksuoti abu tikslinio kirpimo galus vienu Illumina sekoskaitos rodmeniu. Pirmiausia 124 nt oligonukleotidai (IDT) buvo paversti į žiedinę DNR naudojant CircLigase™ vienos grandinės (ss) DNR ligazę (Lucigen) pagal gamintojo rekomendacijas. Žiedinė ssDNR buvo išgryninta ir koncentruota naudojant Monarch® PCR & DNA Cleanup Kit (New England Biolabs). Tada 20 pmol išgryninto produkto buvo inkubuota su 25 pmol komplementaraus pradmens 1X T4 DNR ligazės buferiniame tirpale (New England Biolabs), papildytame 40 μM dNTP. Kad pradmuo galėtų

susilieti, reakcija 30 sekundžių buvo kaitinama iki 65 °C, po to temperatūra sumažinama iki 25 °C 0,2 °C per sekundę greičiu. Tada buvo pridėta 6 vienetai T4 DNR polimerazės ir 400 vienetų T4 DNR ligazės (New England Biolabs) ir reakcija inkubuota 12 °C temperatūroje 1 valandą, kad vyktų antrosios grandinės sintezė. Po gryninimo su Monarch® PCR & DNA Cleanup Kit ir eliucijos į 1X CutSmart buferį (New England Biolabs), kuriame yra 1 mM ATP, į mėginį pridėta 15 vienetų Egzonukleazės V (RecBCD; New England Biolabs) ir T5 egzonukleazės (New England Biolabs) ir 45 min. inkubuota 37 °C temperatūroje. Po išskyrimo, naudojant Agilent 2100 Bioanalyzer, buvo įvertinta žiedinės dvigubos grandinės DNR išeiga.

Minižiedo hidrolizei Cas9 RNP kompleksai buvo suformuoti inkubuojant 1 pmol sgRNR su 0,5 pmol Cas9 baltymo 1X NEBuffer 3.1 arba 2.1 (New England Biolabs) kambario temperatūroje 10 min. Pridėta 0,1 pmol žiedinio dgDNR substrato, mėginiai inkubuoti 15 min. 37 °C temperatūroje, tada kiekviena 20 µl reakcija sustabdyta įpilant 5 µl 0,16 M EDTA. Reakcijos buvo sukoncentruotos ir išgrynintos naudojant Monarch® PCR & DNA Cleanup Kit, o visas 8 µl eliuoto produkto buvo panaudotas kaip substratas Illumina sekoskaitos bibliotekai kurti naudojant NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs) pagal su rinkiniu pateiktą protokolą. 15 PGR ciklų buvo naudojama pridedant Illumina pradmenis ir barkoduotus indeksus, tada kiekvienos reakcijos koncentracija buvo įvertinta Agilent 2100 Bioanalyzer. Bibliotekos buvo apjungtos ir sekvenuotos Illumina NovaSeq arba NextSeq prietaisu, atliekant 2 X 150 poruotų-galų sekoskaitos ciklų. Po to, naudojant pasirinktinius algoritmus , buvo įvertinamos kirpimo vietos ir vizualizuojamos kaip *heatmap* diagramos (rodančios skaldymo dalį) naudojant Microsoft Excel 16.36 ir GraphPad Prism 8.

Cas9 *in vitro* DNR hidrolizės tyrimai optimalaus buferio, temperatūros ir skirtuko ilgio nustatymui

Pirmiausia iš HEK293T genominės DNR, naudojant pradmenis, atitinkančius WTAP ir RUNX1, PGR būdu buvo amplifikuoti DNR substratai, kuriuose yra kiekvieno ortologo kanoninis PAM. Priekiniai pradmenys buvo paženklinti 5'-FAM ir 5'-ROX atitinkamai WTAP ir RUNX1. Atvirkštiniai pradmenys buvo nežymėti. Po to 515 ir 605 bp PGR produktai, atitinkamai WTAP ir RUNX1, buvo išgryninti naudojant Monarch® PCR & DNA Cleanup Kit (New England Biolabs), o DNR koncentracija ir grynumas išmatuoti NanoDrop™ spektrofotometru (ThermoFisher). Išgrynintas Cas9 baltymas buvo praskiestas iki 1 µM skiedimo buferyje (300 mM NaCl, 20 mM Tris, pH7,5) ir laikomas ant ledo. Tada sgRNR buvo praskiestos iki 2 µM vandenyje. Cas9

ir sgRNR buvo apjungti 2:1 sgRNR:Cas9 moliniu santykiu reakcijos buferiniame tirpale 10 min. kambario temperatūroje. Po to buvo pridėtas substratas, kurio santykis Cas9:sgRNR:DNR buvo 10:20:1, ir inkubuota 30 min. Buferio optimizavimo ir skirtuko ilgio pasirinkimo eksperimentams kaip reakcijos buferiai buvo naudojami 1X NEB buferiai 1.1, 2.1, 3.1 arba CutSmart (New England Biolabs), o inkubacijos vyko 37 ºC temperatūroje. Atliekant termoaktyvumo eksperimentus, reakcijos buvo atliekamos NEBuffer 3.1. Iš pradžių RNP kompleksai buvo formuojami kambario temperatūroje, o po to perkeliami į termociklerį, iš anksto pašildytą arba atvėsintą iki įvairių tyrimo temperatūrų, prieš pridedant DNR substrato. 10x DNR substratas (100 nM) prieš dedant į RNP turinčius reakcijos mėgintuvėlius buvo atskirai laikomas nurodytoje temperatūroje. Reakcijos buvo stabdytos pridedant 0,8 % (v/v) NDS ir 80 mU proteinazės K (New England Biolabs). Hidrolizės produktai buvo 4 kartus praskiesti ir atlikta kapiliarinė elektroforezė (CE), siekiant kiekybiškai nustatyti hidrolizės efektyvumą. Tada, naudojant Microsoft Excel 16.36 ir GraphPad Prism 8, kiekvienoje temperatūroje suskaidyto substrato dalis buvo vizualizuota kaip *heatmap* diagramos.

Cas9 baltymų termostabilumo tyrimai

Išgryninti Cas9 baltymai buvo praskiesti 300 mM NaCl, 20 mM Tris, pH7,5 buferyje iki 5-10 µM kambario temperatūroje. 10 µl praskiesto baltymo buvo užnešta į NanoDSF Grade Standard Capillaries (NanoTemper) ir lydymosi temperatūra nustatyta naudojant Prometheus NT4.8 NanoDSF prietaisą pagal gamintojo instrukciją. Temperatūra buvo didinama nuo 20ºC iki 80ºC 1ºC/min greičiu. Lydymosi kreivių lūžio taškai pateikiami kaip T_m.

CRISPR-Cas12l identifikavimas

Pirmiausia CRISPR regionai mikrobų genomuose buvo aptikti naudojant PILER-CR (Edgar, 2007) ir MinCED (Bland et al., 2007) programas. Tada iš duomenų rinkinio buvo pašalintos žinomos CRISPR-Cas sistemos, ieškant netoli (20 kb 5' ir 20 kb 3' (jei įmanoma)) CRISPR regiono koduojamų baltymų homologiškų žinomiems Cas baltymams, naudojant pozicijai būdingų vertinimo matricų (PSSM) rinkinį, apimantį visas žinomas Cas baltymų šeimas, kaip aprašyta Makarova, *et al*. 2015. Siekiant padėti visiškai pašalinti žinomas 2 klasės CRISPR-Cas sistemas, naudojant MUSCLE (Edgar, 2004) buvo sudarytas daugialypis baltymų sekų iš kiekvienos 2 klasės CRISPR-Cas endonukleazių šeimos (pvz., Cas9, Cpf1 (Cas12a), C2c1 (Cas12b), C2c2 (Cas13), C2c3 (Cas12c)) ortologų rinkinio palyginys. Palyginiai buvo išnagrinėti, kuruoti ir naudoti profiliniams paslėptiesiems

Markovo modeliams (HMM) kurti naudojant HMMER (Eddy, 1998; Finn et al., 2011). Tada gauti HMM modeliai buvo panaudoti toliau nustatant ir šalinant žinomas 2 klasės CRISPR-Cas sistemas iš duomenų rinkinio. Toliau, naudojant pirmiau aprašytą specifinę PSSM paiešką, likusieji CRISPR lokusai buvo įvertinti, ar juose yra genų, koduojančių baltymus, kurie, kaip įtariama, yra svarbūs skirtuko integracijoje ir adaptacijoje, Cas1 ir Cas2. Tada buvo atrinkti CRISPR lokusai, kuriuose yra *cas1* ir *cas2* genų, ir toliau tiriami siekiant nustatyti lokuse koduojamų neapibrėžtų genų artumą, eiliškumą ir kryptingumą *cas1* ir *cas2* genų ir CRISPR regiono atžvilgiu. Tolesnei analizei buvo atrinkti tik tie CRISPR lokusai, sudarantys į operoną panašią struktūrą, kuriuose didelis (≥1500 bp atviro skaitymo rėmelis) neapibrėžtasis genas buvo arti *cas1* ir *cas2* genų ir ta pačia transkripcijos kryptimi. Toliau neapibrėžto geno koduojamas baltymas buvo analizuojamas ieškant sekos ir struktūrinių savybių, būdingų V tipo nukleazių domenams. Pirmiausia, priklausomai nuo to, kiek panašumo buvo tarp kandidato sekos ir žinomų baltymų, jo konservatyvioms funkcinėms savybėms atskleisti buvo pasitelktos įvairios bioinformatinės priemonės, pradedant poriniu palyginimu, šeimos profilio paieška ir baigiant rankiniu struktūriniu patikrinimu. Bendrai, sekos, homologiškos kandidatiniams baltymams, pirmiausia buvo renkamos atliekant PSI-BLAST (Altschul et al., 1997) paiešką Nacionalinio biotechnologijų informacijos centro (NCBI) neperteklinėje (NR) baltymų kolekcijoje, kurios ribinė e. vertė yra 0,01. Sumažinus perteklių ~90 % tapatumo lygmeniu, homologinių sekų grupės su įvairiomis narių įtraukimo ribomis (pvz., >60, 40 arba 20 % tapatumo) buvo sulygiuotos siekiant atskleisti konservatyvius motyvus naudojant daugialypio sekų lyginimo įrankius MSAPRobs (Y. Liu et al., 2010) ir ClustalW (Thompson et al., 1994). Labiausiai konservatyvios homologiškos sekos buvo ieškomos pagal HMMER (Eddy, 1998) įvairiose duomenų bazėse, įskaitant Pfam, Superfamily ir SCOP (Murzin et al., 1995). Atskirai, gautas kandidato homologinės sekos palyginys taip pat buvo naudojamas kandidato baltymo profiliui sukurti, pridedant prognozuojamas antrines struktūras. Toliau kandidato profilis buvo naudojamas profilio ir profilio paieškai atlikti naudojant HHSEARCH (Söding, 2005), lyginant su pdb70_hhm ir Pfam_hhm profilių duomenų bazėmis. Kitame etape visi aptikti sekos ir struktūros ryšiai ir konservatyvūs RuvC panašūs motyvai buvo suvesti į 3D struktūros šabloną naudojant MODELLER arba rankiniu būdu atvaizduoti į žinomą struktūrinį etaloną DiscoveryStudio (BIOVIA) ir Pymol (Schrodinger). Galiausiai, siekiant patikrinti ir patvirtinti galimą biologinę svarbą, rankiniu būdu buvo patikrintos ir įvertintos katalitinės arba labiausiai konservatyvios aminorūgštys ir pagrindiniai struktūriniai motyvai, atsižvelgiant į baltymo

gebėjimą metabolizuoti DNR. Nustačius RuvC, kiti lokuse koduojami baltymai (5 kb 5' ir 5 kb 3' nuo naujai apibrėžtos CRISPR-Cas sistemos galų (kur įmanoma)) toliau buvo tiriami dėl homologijos su žinomomis baltymų šeimomis naudojant InterProScan programinę įrangą (EMBL-EBI, Jungtinė Karalystė) ir lyginant su NCBI NR baltymų kolekcija naudojant BLAST programą (Altschul et al., 1990).

CRISPR-Cas12l sistemų nutaikymas į PAM biblioteką

Cas12l CRISPR sistemos buvo modifikuotos taip, kad būtų nukreiptos į anksčiau aprašytą 7 atsitiktinių bp PAM biblioteką (Karvelis et al., 2018). Natyvus CRISPR regionas buvo pakeistas keturiais pasikartojimo:skirtuko:pasikartojimo vienetais, iš kurių du kodavo 37 nt skirtukus, komplementarius antiprasminei DNR grandinei 5’ nuo atsitiktinės PAM sekos plazmidinėje bibliotekoje, o du kodavo seką, komplementarią taikiniui 3’ už PAM. Sukurti CRISPR-Cas lokusai buvo susintetinti (GenScript) ir klonuoti į mažakopijinę *E. coli* plazmidę (pET-Duet-1), modifikuotą taip, kad joje būtų vienas izopropil β-D-1-tiogalaktopiranozido (IPTG) indukuojamas T7 promotorius (MilliporeSigma).

Cas12l DNR hidrolizės aktyvumo ir PAM atpažinimo nustatymas

ArcticExpress (DE3) (Agilent Technologies) *E. coli* buvo transformuotos plazmidėmis, turinčiomis modifikuotas CRISPR-Cas12l sistemas. Kultūros buvo auginamos 20 ml LB terpės, kurioje yra 20 µg/ml gentamicino ir 100 µg/ml karbenicilino. Kultūroms pasiekus 0,6 OD_{600}, raiška buvo indukuota 0,1 mM IPTG ir kultūros inkubuotos per naktį 16 °C temperatūroje. Ląstelės iš 10 ml terpės buvo surinktos centrifuguojant ir resuspenduotos 1 ml lizės buferinio tirpalo (20 mM Tris-HCl, pH 7,5, 500 mM NaCl, 5 % (v/v) glicerolio), papildyto 1 mM PMSF, ir lizuotos ardant ultragarsu. Ląstelių liekanos pašalintos centrifuguojant, o 10 µl gauto supernatanto panaudota plazmidžių hidrolizės reakcijoms.

Lizatas, kuriame yra Cas12l RNP kompleksų, buvo naudojamas 7 bp atsitiktinių PAM plazmidžių bibliotekai hidrolizuoti. 10 µl *E. coli* lizato buvo sumaišyta su 500 ng PAM bibliotekos 50 µl reakcijos buferio (1xCutSmart buferis (New England Biolabs): 50 mM KOAc, 20 mM Tris-OAc, 10 mM Mg(OAc)2, 100 µg/ml BSA, pH 7,9) ir 1 valandą inkubuojama 37 °C temperatūroje. Kirpimo produktų DNR galų taisymas, surinkimas ir sekoskaita atlikti kaip aprašyta Cas9 ortologų PAM nustatymo metodikoje.

Cas12l baltymų raiška ir gryninimas

Cas12l baltymai buvo išreikšti *E. coli* NiCo21(DE3) kamiene (New England Biolabs) iš pET-MBP14xHisSUMO-Cas12l plazmidžių. Ląstelės augintos LB terpėje 30 °C temperatūroje. Kultūroms pasiekus 0,5 OD_{600}, raiška buvo indukuota 0,4 mM IPTG ir inkubuota 16 °C temperatūroje per naktį. Ląstelės centrifuguotos, resuspenduotos užnešimo buferyje (20 mM Tris-HCl, pH 7,5, 300 mM NaCl, 40 mM imidazolo) ir suardytos ultragarsu. Ląstelių liekanos pašalintos centrifuguojant, o supernatantas filtruotas prieš užnešant ant HiTrap DEAE FF chromatografijos kolonėlės (GE Healthcare). Po to neprikibusi frakcija buvo užnešta ant Ni^{2+} įkrautos HisTrap kolonėlės (GE Healthcare) ir eliuotas linijiniu gradientu didinant imidazolo koncentraciją (nuo 40 mM iki 700 mM). Frakcijos, kuriose buvo Cas12l baltymų, buvo apjungtos, o MBP-14xHis-SUMO inkarai buvo nukirpti padidinus NaCl koncentraciją iki 500 mM, pridėjus 1 mM DTT, 2 % (v/v) glicerolio ir 100 nM SenP1 proteazės bei inkubuojant 4 °C temperatūroje per naktį. Kad būtų pašalinta nukirptas aminorūgščių inkaras ir SenP1 proteazė, reakcijos mišinys buvo užneštas ant HiTrap Heparin kolonėlės (GE Healthcare) ir eliuotas linijiniu gradientu didinant NaCl koncentraciją (nuo 0,5 M iki 2 M). Frakcijos, kuriose buvo Cas12l baltymų, buvo apjungtos ir dializuotos prieš 20 mM Tris-HCl, pH 7,5, 500 mM NaCl, 1 mM DTT, 0,1 mM EDTA, 0,1 mM EDTA, 50 % glicerolio saugojimo buferį ir patalpintos -20 °C temperatūroje.

Cas12l-sgRNR kompleksų surinkimas *in vitro* hidrolizės reakcijoms

Cas12l ir sgRNR RNP kompleksai buvo surinkti sumaišius išgrynintą Cas12l baltymą su sgRNR 1:2 moliniu santykiu, po to 20 min. inkubuojant komplekso sudarymo buferyje (10 mM Tris-HCl, pH 7,5, 100 mM NaCl, 1 mM DTT, 1 mM EDTA) kambario temperatūroje.

Cas12l DNR substratų gamyba

Plazmidinės DNR substratai buvo sukurti liguojant oligodupleksus, surinktus sulydant oligonukleotidus, turinčius PAM ir protoskirtuko sekas, į pUC19 plazmidę per EcoRI (New England Biolabs) ir HindIII (New England Biolabs) restrikcijos endonukleazių sritis arba liguojant PGR produktus, turinčius PAM ir protoskirtuko sekas, per EcoRI ir HindIII restrikcijos endonukleazių sritis užbukintais DNR galais.

Fluorescentiškai pažymėti DNR substratai buvo sukurti sulydant iš dalies komplementarius oligonukleotidus, turinčius PAM ir protoskirtuko sekas, ir PGR dauginant juos su pradmenimis, turinčiais 5'-6-FAM (ne-taikinio grandinė) arba 5'-6-ROX (taikinio grandinė) (IDT) žymes.

Cas12l DNR substratų hidrolizės eksperimentai

DNR hidrolizės reakcijos buvo pradėtos sumaišius DNR substratus su Cas12l RNP kompleksais. Plazmidinės DNR hidrolizės reakcijos buvo atliekamos 37 °C temperatūroje 1xCutSmart buferiniame tirpale (New England Biolabs), esant substrato ir komplekso moliniam santykiui 1:20, jei nenurodyta kitaip. Reakcijos mėginiai buvo paimami nustatytais laiko intervalais (30 min, jei nenurodyta kitaip) ir sumaišomi su 6x Blue Gel Loading Dye (New England Biolabs). Reakcijos produktai buvo analizuojami agarozės gelio elektroforezės metodu ir vizualizuojant GelRed dažu (Biotium). Suskaidyto substrato frakcijos buvo apskaičiuotos atliekant densitometrinę analizę su ImageJ programine įranga pagal šią lygtį: $Hidrolizuota\ substrato\ dalis, \% = (\frac{I_{produktų}}{(I_{produktų}+I_{substrato})}) \times 100,$ kur $I_{produktų}$ yra hidrolizės produktų juostų intensyvumas, o $I_{substrato}$ - nepažeisto substrato juostos intensyvumas.

Reakcijos su fluorescentiškai pažymėtais DNR substratais buvo atliekamos 37 °C arba 47 °C temperatūroje 1xCutSmart buferyje, kurioje substrato ir komplekso molinis santykis yra 1:10 arba 1:5 (suboptimalių PAM substratų hidrolizės eksperimentai) (20 mM substrato ir atitinkamai 200 mM arba 100 mM Cas12l komplekso, jei nenurodyta kitaip). Reakcijų mėginiai buvo paimti nustatytais laiko intervalais ir sumaišyti su 9,5 µl Hi-Di formamido (Thermo Fisher Scientific), papildyto 0,5 µl GeneScan LIZ 120 Size Standard (Thermo Fisher Scientific). Reakcijos produktai buvo tiriami kapiliarine elektroforeze naudojant 3500 Series Genetic Analyzer (Thermo Fisher Scientific) pagal gamintojo rekomendacijas. Elektroforezės duomenys buvo analizuojami naudojant OSIRIS (NCBI) arba Geneious Prime (Biomatters Ltd.) programinę įrangą. Fragmentų dydis įvertintas palyginus sulaikymo laiką su vidinio juostos standarto sulaikymo laiku, o suskaidyto substrato frakcijos apskaičiuotos palyginus substrato ir skilimo produkto fragmento smailės plotą (mažiausias RFU (santykiniai fluorescencijos vienetai) smailės aptikimui - 50): $Hidrolizės\ efektyvumas, \% = (\frac{smailės\ plotas\ (produktų)}{smailės\ plotas\ (substrato)+smailės\ plotas\ (produktų)}) \times 100.$ Hidrolizės efektyvumo grafikai buvo nubraižyti naudojant GraphPad Prism programinę įrangą. Atitinkamais atvejais, gauti duomenys atvaizduoti grafiškai pritaikius eksponentinės asociacijos kreivės modelį (GraphPad Prism programinė įranga), pagal šią lygtį: $Hidrolizuota\ dalis = A \times (1 - e^{(-k \times t)},$ kur A - kreivės amplitudė, k - pseudopirmosios eilės greičio konstanta, o t - laikas.

M13 vgDNR hidrolizės tyrimai

M13 vgDNR hidrolizės reakcijos pradėtos 37 °C temperatūroje sumaišius M13 vgDNR (New England Biolabs) ir vgDNR aktyvatorių (oligonukleotidas) arba dgDNR aktyvatorių (oligonukleotidų dupleksas) su Cas12l RNP kompleksais. Hidrolizės reakcijos buvo atliekamos 1xCutSmart buferyje. Galutiniame reakcijos mišinyje buvo 5 nM M13 ssDNA, 100 nM vgDNR/dgDNR aktyvatoriaus, 100 nM Cas12l-sgRNA komplekso (jei nenurodyta kitaip). Reakcijos sustabdytos sumaišant su 6x Blue Gel Loading Dye (New England Biolabs), o produktai analizuoti agarozės gelio elektroforezės metodu ir vizualizuojant GelRed dažu (Biotium).

Fluoroforais ir slopikliais žymėtų reporterių matavimai

Asp2Cas12l ir sgRNR kompleksai buvo iš anksto surinkti inkubuojant 500 nM Asp2Cas12l su 600 nM sgRNR 1x2.1 buferyje (New England Biolabs) 15 min. kambario temperatūroje. Po to kompleksai buvo praskiesti ir pridėta vgDNR/dgDNR aktyvatorių iki galutinių 105 nM Asp2Cas12l : 125 nM sgRNR : 26 nM vgDNR/dgDNR koncentracijų ir 30 min. inkubuoti 50 °C temperatūroje 1x 2.1 buferiniame tirpale (New England Biolabs). Trans skilimo reakcijos buvo pradėtos pridedant 50 pmol vgDNR/vgRNR fluoroforais ir slopikliais žymėto (FQ) reporterinio substrato, galutinė reakcijos komponentų koncentracija buvo 100 nM Asp2Cas12l : 120 nM sgRNR : 25 nM vgDNR/dgDNR aktyvatoriaus : 500 nM FQ substrato 100 µl reakcijos tūryje juodoje 96 šulinėlių plokštelėje. Reakcijos iš karto buvo dedamos į fluorescentinį plokštelių skaitytuvą ir inkubuojamos 37 °C temperatūroje, fluorescenciją matuojant kas 30 sekundžių.

Nustatant *trans*-nukleazinio aktyvumo greitį, neapdorotos fluorescencijos vertės buvo pakoreguotos atimant fluorescencijos vertes, gautas iš reakcijų tik su FQ reporteriu. Tada santykinės fluorescencijos vertės buvo perskaičiuotos į hidrolizuoto FQ substrato dalį pagal šią lygtį: $Hidrolizuota\ dalis = \frac{F(t)}{F(suskaidyta)} \times 100$, kur F(t) - fluorescencija tam tikru laiko momentu; F(suskaidyta) - visiškai suskaidyto FQ reporterio fluorescencija. Gauti duomenys buvo priderinti prie eksponentinės asociacijos kreivės modelio (GraphPad Software) pagal šią lygtį: $Hidrolizuota\ dalis = A \times (1 - e^{(-k \times t)}$, kur A - kreivės amplitudė, k - pseudopirmosios eilės greičio konstanta, o t - laikas.

Michaelis-Menten analizei 400 nM Asp2Cas12l : 500 nM sgRNR : 4 nM vgDNR/dgDNR aktyvatorių kompleksai buvo sudaryti pirmiausia 15 min. inkubuojant Asp2Cas12l ir sgRNR 1x 2.1 buferiniame tirpale (New England Biolabs) kambario temperatūroje, tada pridedant vgDNR/dgDNR aktyvatorių

ir 30 min. inkubuojant 50 °C temperatūroje. *Trans*-nukleazinio aktyvumo reakcijos buvo pradėtos praskiedus surinktus kompleksus iki galutinės 10 nM Asp2Cas12l : 12,5 nM sgRNR : 0,1 nM aktyvatoriaus koncentracijos tirpale, kuriame yra 1x2.1 buferis ir 0,001, 0,01, 0,1, 0,2, 0,5, 1 arba 2 µM FQ substrato 100 µl reakcijos tūryje juodoje 96 šulinėlių plokštelėje. Reakcijos buvo inkubuojamos fluorescentiniame plokštelių skaitytuve 37 °C temperatūroje iki 60 minučių, fluorescenciją matuojant kas 30 sekundžių (λ_{ex} = 485 nm; λ_{em} = 538 nm). Reakcijos, kuriose nebuvo Asp2Cas12l : sgRNR : aktyvatoriaus kompleksų, buvo matuojamos, norint gauti nesuskaidytų reporterių fluorescencijos priklausomybės nuo koncentracijos standartines kreives, o reakcijos, kuriose buvo 100 kartų didesnė Asp2Cas12l : sgRNR : aktyvatoriaus efektyvių kompleksų koncentracija, buvo matuojamos, kad būtų gautos visiškai suskaidytų reporterių fluorescencijos priklausomybės nuo koncentracijos standartinės kreivės.

Gautos santykinės fluorescencijos vertės buvo perskaičiuotos į suskaidyto fluorescencinio substrato koncentraciją pagal šią lygtį: $c_{cl} = \frac{F(t) - F(c_0)}{S_{cl} - S_{ucl}}$, kur c_{cl}- suskaidyto substrato koncentracija; $F(t)$ - santykinė fluorescencijos vertė tam tikru laiko momentu; $F(c_0)$ - nesuskaidyto reporterio fluorescencijos vertė tam tikroje koncentracijoje; S_{cl} - tiesinės standartinės kreivės, gautos nubrėžus visiškai suskaidyto substrato santykinės fluorescencijos verčių priklausomybę nuo substrato koncentracijos ir pritaikius tiesinę regresiją, nuolydis; S_{ucl}- tiesinės standartinės kreivės, gautos nubrėžus nesuskaidyto substrato santykinės fluorescencijos verčių priklausomybę nuo substrato koncentracijos ir pritaikius tiesinę regresiją, nuolydis. Visiškai suskaidytas substratas buvo gaunamas inkubuojant su Asp2Cas12l, kol RFU vertės pasiekė stabilią viršūnę.

Pradinis greitis (v_0) apskaičiuotas pritaikius suskaidyto substrato koncentracijos priklausomybės nuo laiko kreives pagal tiesinę regresiją ir pateikta grafiškai pagal pradinę substrato koncentraciją. Michaelis-Menten konstantos buvo nustatytos pagal šią lygtį: $Y = \frac{V_{max} \times X}{K_M + X}$, kur X - substrato koncentracija, o Y - fermento greitis (GraphPad Prism Software). Katalizinė konstanta, arba fermento apsisukimų skaičius (k_{cat}) buvo nustatytas pagal šią lygtį: $k_{cat} = \frac{V_{max}}{E}$, kur E - efektyvi komplekso koncentracija (0,1 nM). V_{max} šiomis sąlygomis prilygintas pradiniam greičiui v_0.

Transformacijos ribojimo tyrimas

E. coli transformacijos plazmidėmis ribojimo eksperimentai buvo atliekami su Arctic Express (DE3) (Agilent Technologies) kamieno *E. coli*. Pirmiausia

pETDuet plazmidė su *Asp2Cas12l* genu buvo modifikuota taip, kad joje būtų įterpti natyviuose Cas12l lokusuose esantys skirtukai, klonuojant susintetintą DNR fragmentą (kuriame yra pasikartojimas-skirtukas1- pasikartojimas-skirtukas2-pasikartojimas-skirtukas3-pasikartojimas seka) per restrikcijos vietas. Kita pETDuet plazmidė, kurioje yra *Cas12l* genas, buvo sukurta pridėjus sgRNR koduojančią seką, turinčią RunXI skirtuką su T7 promotoriaus, HDV ribozimo ir terminatoriaus sekomis, naudojant *Gibson assembly* molekulinio klonavimo metodą (New England Biolabs). Po to pCDF-Duet plazmidės buvo sukurtos taip, kad į jas būtų įterptos protoskirtukų sekos, komplementarios pETDuet plazmidėse esantiems natyviems skirtukams bei RunXI skirtukui, už Cas12l būdingų 5'-CCC-3' PAM sekų, klonuojant oligodupleksus per pCDF-Duet plazmidėje esančias restrikcijos vietas.

E. coli Arctic Express (DE3) (Agilent Technologies) ląstelės pirmiausia buvo transformuotos pETDuet plazmidėmis, kuriose yra *Asp2Cas12l* genas, gRNR ir skirtukų sekos. Šios ląstelės buvo auginamos 37 °C temperatūroje iki 0,5 OD_{600} ir elektroporuojamos 200 ng taikininių pCDF-Duet plazmidžių su protoskirtukų sekomis arba tuščia kontroline pCDF-Duet plazmide. Ko-transformuotos ląstelės buvo auginamos 37 °C temperatūroje 16-20 val. plokštelėse, kuriose buvo 100 g/ml karbenicilino, 10 g/ml streptomicino, 10 g/ml gentamicino ir 0,1 mM IPTG. Transformacijos ribojimas buvo įvertintas pagal plokštelėje stebėtų transformantų kolonijų kiekį.

REZULTATAI IR JŲ APTARIMAS

Šis darbas susideda iš dviejų dalių: pirmoje dalyje aprašomas plataus masto biocheminis įvairių II tipo CRISPR-Cas nukleazių biocheminį tyrimas, kurio metu atrasti fermentai, atpažįstantys naujas PAM sekas, pasižymintys įvairiomis temperatūrinėmis priklausomybėmis ir DNR hidrolizės modeliais, taip pat atlikta bioinformatinė jų su PAM sąveikaujančių domenų filogenetinė analizė ir jų sąsajos su PAM atpažinimu. Antroje dalyje pateikiama informacija apie naujos V tipo CRISPR-Cas nukleazių šeimos identifikavimą ir nukleazinio aktyvumo biocheminį charakterizavimą *in vitro.*

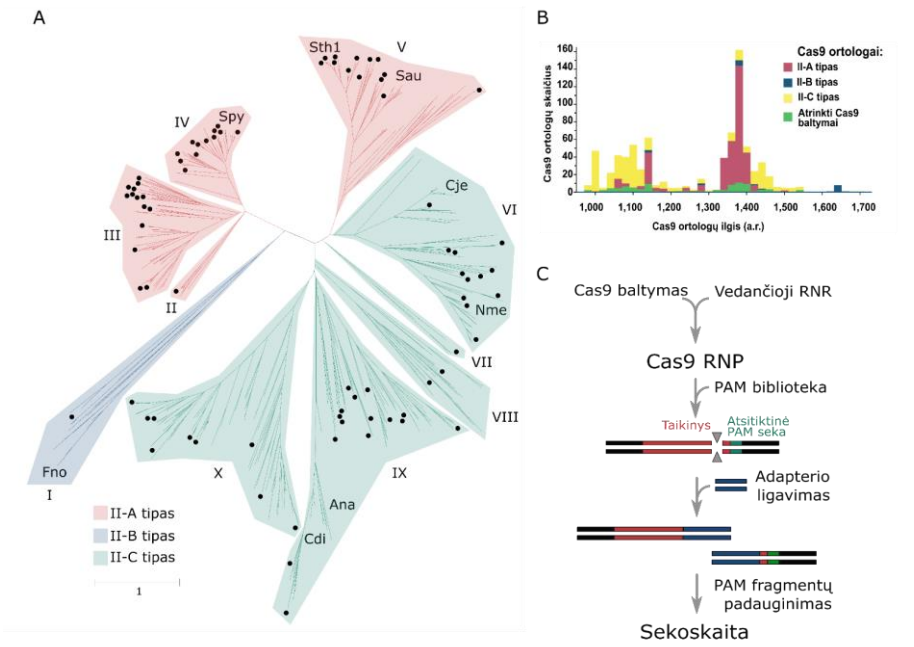
II tipo CRISPR-Cas9 ortologų charakterizavimas

Nors Cas9 nukleazes galima užprogramuoti taip, kad jos būtų nukreiptos į bet kurią DNR vietą, keičiant jų vedančiosios RNR skirtuko seką, praktikoje taikinio pasirinkimą riboja PAM atpažinimas (Collias & Beisel, 2021). Tai dar labiau apsunkina būtinybė kruopščiai parinkti taikinį, kad būtų sumažintas galimas netikslinis prisijungimas dėl tolerancijos gRNR ir taikinio heteroduplekso sekų daliniam nekomplementarumui (D. Kim et al., 2019). Tai taip pat labai svarbu tuomet, kai redagavimo rezultatas labai priklauso nuo pageidaujamos mutacijos pozicijos surišimo vietos atžvilgiu, ypač homologija-paremtos DNR reparacijos, bazių redagavimo ar pradinio redagavimo (angl. *prime editing*) atvejais (Anzalone et al., 2020). Be to, įvairias Cas9 taikymo sritis gali riboti jų biocheminės ir fizikinės savybės, pavyzdžiui, DNR hidrolizė paliekant bukus DNR galus (Jinek et al., 2012), gRNR pakeičiamumas tarp skirtingų ortologų (Fonfara et al., 2014), priklausomybė nuo temperatūros (Wiktor et al., 2016) ir baltymo dydis (Lino et al., 2018). Gamtoje aptinkamų Cas9 ortologų įvairovė gali pasiūlyti unikalių nukleazių, taip pat suteikti įžvalgų, kaip patobulinti Cas9 įrankių rinkinį.

Cas9 ortologų atranka

Siekiant subalansuoti numanomų Cas9 ortologų įvairovę ir biochemines savybes, taikytos kelios skirtingos *cas9* genų atrankos strategijos. Buvo atrinkti 47 ortologai iš 10 pagrindinių Cas9 evoliucinio medžio grupių (1A pav.). Iš grupių, iš kurių kilo anksčiau viešai charakterizuoti baltymai, aktyvūs eukariotinėse ląstelėse, buvo atrinkta maždaug 20 %, o visų kitų - maždaug 10 % genų. Siekiant praturtinti imtį baltymais, pasižyminčiais potencialiai stipriu biocheminiu aktyvumu ir termostabilumu, buvo atrinkti papildomi 32

ortologai, atsižvelgiant į jų prognozuojamas fizikines ir chemines savybes, priskyrimą II-A tipo potipiui, kuris apima artimiausius nuodugniai apibūdintų CRISPR-Cas sistemų ortologus (Fonfara et al., 2014; Makarova et al., 2020), ir priklausymą termofiliniam organizmui šeimininkui. Taikant šį metodą buvo gautas Cas9 ortologųnuo ~1000 iki ~1600 aminorūgščių dydžio rinkinys su bimodaliniu pasiskirstymu, sutelktu apie ~1100 ir ~1375 aminorūgštis (1B pav.).



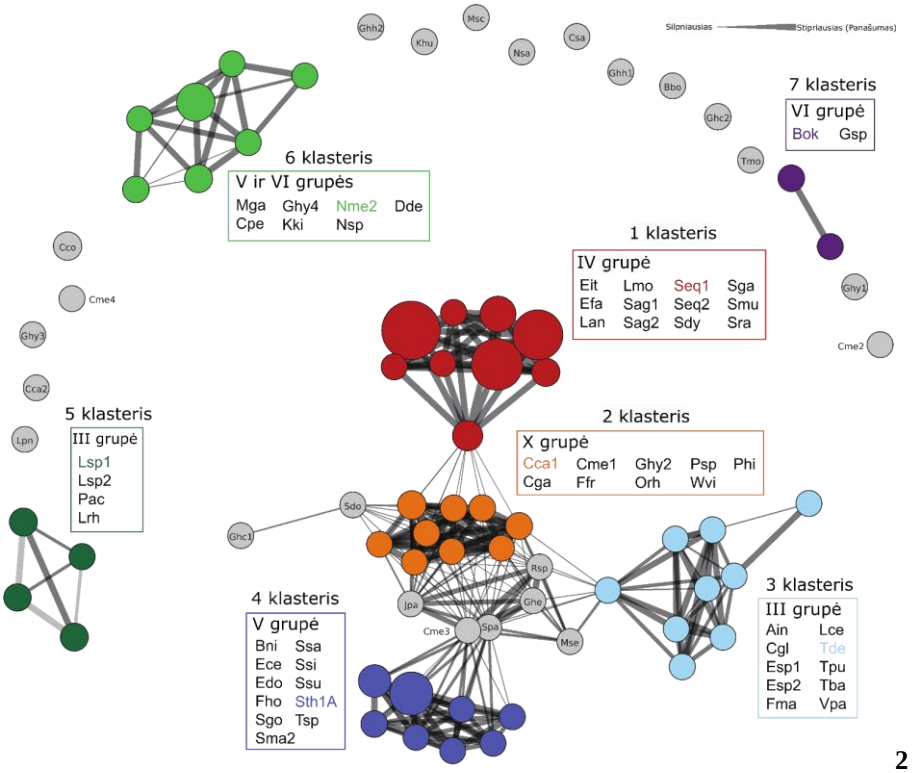
1 pav. Cas9 įvairovė ir PAM charakterizavimo metodas. (A) Cas9 ortologų įvairovės filogenetinis vaizdavimas. II-A, B ir C tipo sistemos pažymėtos spalvomis, atitinkamai raudona, mėlyna ir žalia. Skirtingos filogenetinės grupės sunumeruotos I-X numeriais. Tyrimui atrinktieji baltymai pažymėti juodu tašku. Taip pat pažymėti Cas9, kurių struktūra buvo nustatyta anksčiau. **(B)** Cas9 baltymų dydžio pasiskirstymas pagal potipį. II-A, B ir C tipo sistemos žymimos spalvomis, atitinkamai raudona, mėlyna ir geltona. Analizuotų Cas9s skaičius kiekvienoje grupėje pažymėtas žaliai. **(C)** Biocheminis metodas, naudotas tiesiogiai fiksuoti taikinio hidrolizę ir įvertinti protoskirtukui artimo motyvo (angl. *protospacer adjacent motif*) (PAM) atpažinimą. Eksperimentai atlikti naudojant Cas9 baltymus, pagamintus IVT būdu.

Cas9 vedančiųjų RNR nustatymas ir analizė

Prieš atliekant atrinktų Cas9 ortologų biocheminius tyrimus, reikėjo nustatyti jų tracrRNR sekas. Tai atlikta *in silico*, tiriant Cas9 genomo lokusus, ieškant sekų, komplementarių CRISPR regionų pasikartojimams ir pasižyminčių spėjamais konservatyviais antrinės struktūros motyvais (Briner et al., 2014).

TracrRNR buvo lyginamos atliekant kompiuterinę kovariacijos modelių (CM) analizę pagal sekų ir antrinės struktūros homologiją, kurios rezultatas - 7 skirtingi klasteriai (2 pav.). Tačiau kai kurių Cas9 ortologų atveju tracrRNR savaime susibūrė į klasterius arba pasižymėjo silpnu panašumu į kitus CM ir nebuvo priskirtos jokiai konkrečiai grupei (2 pav.). Daugeliu atvejų klasteriai buvo susiję su tam tikra Cas9 filogenetiniu grupe (1 ir 2 pav.).

Kiekvienam ortologui sgRNR buvo sukurtos sujungiant numanomas tracrRNR su atitinkamais crRNR pasikartojimais per lankstų jungtuką. Buvo susintetintos matricinės DNR sgRNR transkripcijai, kurios vėliau buvo naudojamos biocheminei analizei.



2

2 pav. Cas9 tracrRNR sekų ir antrinių struktūrų panašumas. Apskritimų mastelis priklauso pagal sekų, priklausančių kiekvienam kovariacijos modeliui (CM), skaičių ir apskritimai nuspalvinti pagal nurodytą klasterį. Jungiamųjų linijų plotis rodo CM panašumo arba gimininguмo procentinę dalį. Kiekvieno klasterio reprezentatyvios tracrRNR pažymėtos atitinkama spalva. Klasteriams nepriskirti CM yra pilkos spalvos.

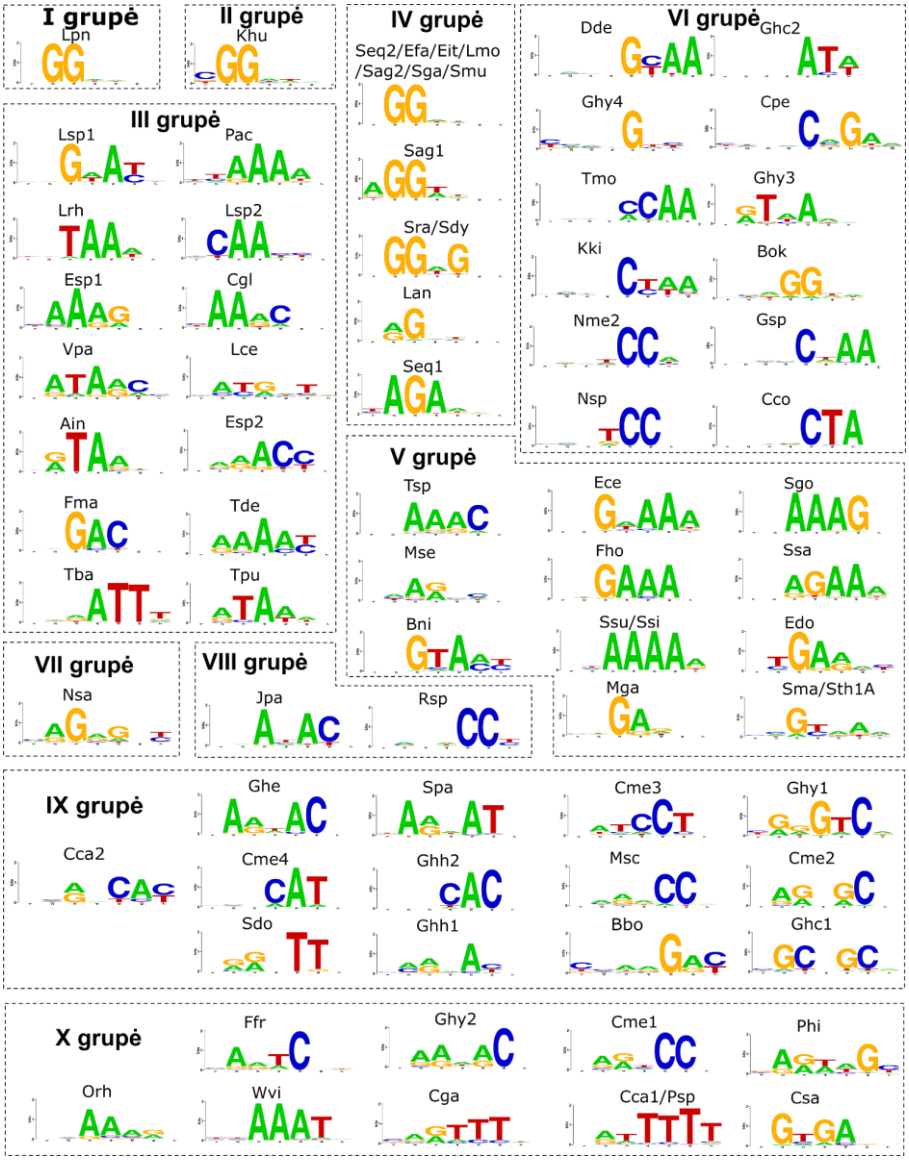
Cas9 ortologų PAM atpažinimas

Siekdami ištirti Cas9 ortologų taikinio atpažinimo savybes dideliu našumu, naudojome in vitro transliacijos (IVT) metodą be ląstelių, panašų į anksčiau aprašytą (Karvelis et al., 2018; Marshall et al., 2018) (1C pav.). Neapdoroti

IVT RNP mišiniai buvo praskiesti (10^1-10^3 kartų), kad būtų atsižvelgta į PAM atpažinimo griežtumo skirtumus, priklausančius nuo Cas9-vedančiosios RNR komplekso koncentracijos (Karvelis et al., 2015), ir patikrintas jų gebėjimas lemti hidrolizę, kai jie buvo apjungti su plazmidžių biblioteka, kurioje yra atsitiktinių nukleotidų PAM sritis, besiribojanti su Cas9 taikinio vieta.

Tuomet, kaip PAM atpažinimo atskaitos taškas buvo naudojamas didžiausias praskiedimas, kuris vis dar palaikė DNR hidrolizės aktyvumą. Be to, tais atvejais, kai Cas9 ortologai pirmenybę teikė PAM 6 arba 7 padėčiai ir neturėjo PAM reikalavimų pirmoje, antroje arba trečioje padėtyje, į PAM biblioteką nukreiptas skirtukas buvo perkeltas į 5' pusę 1, 2 arba 3 nukleotidais, kad būtų galima ištirti galimą PAM atpažinimą už 7 pozicijos, t. y. už PAM bibliotekos atsitiktinės sekos ilgio. Tai leido išplėsti PAM atpažinimą atitinkamai iki 8, 9 arba 10 bazių porų. Iš 20 ortologų, kurie buvo tiriami dėl galimai ilgesnių PAM sekų, tik 6 iš jų, visi priklausantys VI filogenetinei grupei (1A ir 3 pav.), atpažino PAM sekas, kurios tęsėsi toliau nei 7 pozicijoje. Šiek tiek netikėta, kad PAM pirmenybė 8-ojoje padėtyje visada buvo A nukleotidas, panašiai kaip anksčiau apibūdintuose Brevibacillus laterosporus (Blat) (Karvelis et al., 2015) ir Geobacillus stearothermophilus (Geo) (Harrington et al., 2017) Cas9 baltymuose.

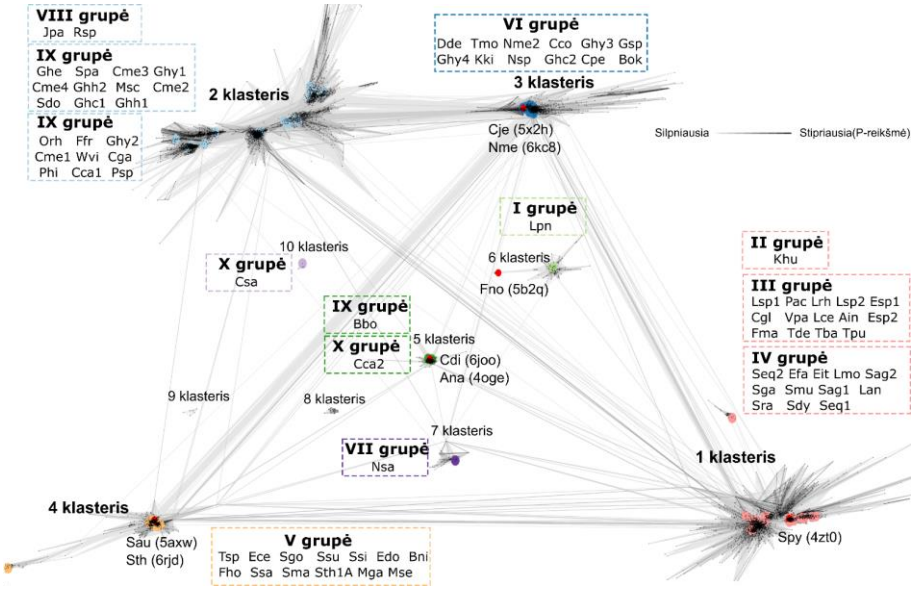
Taikant šį metodą nustatytos 79 Cas9 ortologų PAM sekos, pasižyminčios įvairiais PAM atpažinimo motyvais (3 pav.). Tai apėmė nukleazes su anksčiau neaprašytais PAM reikalavimais, kurių sudėtis skyrėsi ir pagal seką, ir pagal ilgį. Šių Cas9 ortologų PAM atpažinimo būdus galima iš esmės suskirstyti į A, T ir C turtingus PAM atpažinimo būdus, šalia Spy Cas9 baltymui būdingo G turtingo PAM (3 pav.). Trumpi PAM (1-2 bazių poros) buvo reti, o dauguma ortologų atpažino tris ar daugiau pozicijų. Tačiau vienas atvejis, Ghy4, atpažino išskirtinai laisvą PAM seką su tik vienu atskiru G reikalavimu 5 padėtyje, nors ir su papildomais nukrypimais nuo tam tikrų bazių kitose padėtyse. Ortologai, atpažįstantys PAM, sudarytas iš kelių pasikartojančių nukleotidų (pvz., Efa, Nme2, Rsp, Ssi ir Ssu), taip pat buvo daug retesni nei Cas9, atpažįstantys sudėtines PAM, turinčias bent du skirtingus nukleotidus. Be to, daugelis baltymų pasižymėjo iš pažiūros degeneruotu PAM atpažinimu. Paprastai tai lėmė stiprų reikalavimą, kad bent viena bazinė pora būtų derinama su pozicijomis, kurios priimdavo daugiau nei vieną (paprastai du) nukleotidus (pvz., Lan, Mse, Nsa ir Sma2).



3 pav. Cas9 ortologų atpažįstamos PAM sekos. Kiekvieno Cas9 ortologo atpažintų PAM sekų WebLogo diagramos, nurodančios aptiktą santykinį atitinkamo nukleotido dažnį konkrečioje pozicijoje 3’ už taikinio. Cas9 buvo išreikšti IVT būdu ir naudojami atsitiktinių nukleotidų PAM plazmidžių bibliotekai tirti. Ortologai atskirti pagal filogenetines grupes, kuriam jie priskiriami pagal 1A pav.

Cas9 su PAM sąveikaujančių domenų filogenetinė analizė

Toliau dėka eksperimentiškai nustatytų PAM pirmenybių įvairovės įvertinome Cas9 su PAM sąveikaujančių (PI) domenų sekų ryšį. PI sričių sekos iš apibūdintų ortologų buvo išskirtos ir panaudotos kaip užklausos iteracinei paieškai neperteklinuose mikrobų baltymų rinkiniuose. Iš viso buvo aptikta 9 161 seka, turinti netapačius PI domenus. Tada šių domenų sekos buvo suskirstytos į klasterius pagal jų porinį panašumą ir buvo nustatyta dešimt klasterių (4 pav.). 93 % visų rastų sekų pateko į 1-4 klasterius, o 7-10 klasteriai buvo gerokai mažesni ir juos sudarė nuo 4 iki 37 sekų. Apskritai PI domenų sekų panašumą galima susieti su pagrindinėmis filogenetinio Cas9 medžio šakomis, o glaudžiai susiję klasteriai sugrupuoti tame pačiame klasteryje (1A pav. ir 4 pav.). Be to, panašūs PI domenai paprastai lemdavo panašų PAM atpažinimą, tačiau keliais atvejais PAM sekų įvairovė ir ilgis labai skyrėsi net tarp tos pačios grupės narių (3 pav. ir 4 pav.).



4 pav. Cas9 su PAM sąveikaujančių (PI) domenų panašumas. Cas9 PI domenai, sugrupuoti pagal porinį sekų panašumą. Sekos klasterizuotos naudojant CLANS (BLAST parinktis). Linijos jungia sekas, kurių P reikšmės ≤ 1e - 11. Linijų ryškumas atitinka P-reikšmes pagal viršutiniame dešiniajame kampe esančią skalę (šviesios ir ilgos linijos jungia toli susijusias sekas). Tam pačiam klasteriui priklausančios Cas9 sekos atskirtos brūkšninėmis linijomis, nuspalvintomis pagal atitinkamą klasterį. Sekos, kurių struktūra žinoma, pažymėtos raudonai; skliausteliuose nurodytas jų PDB kodas.

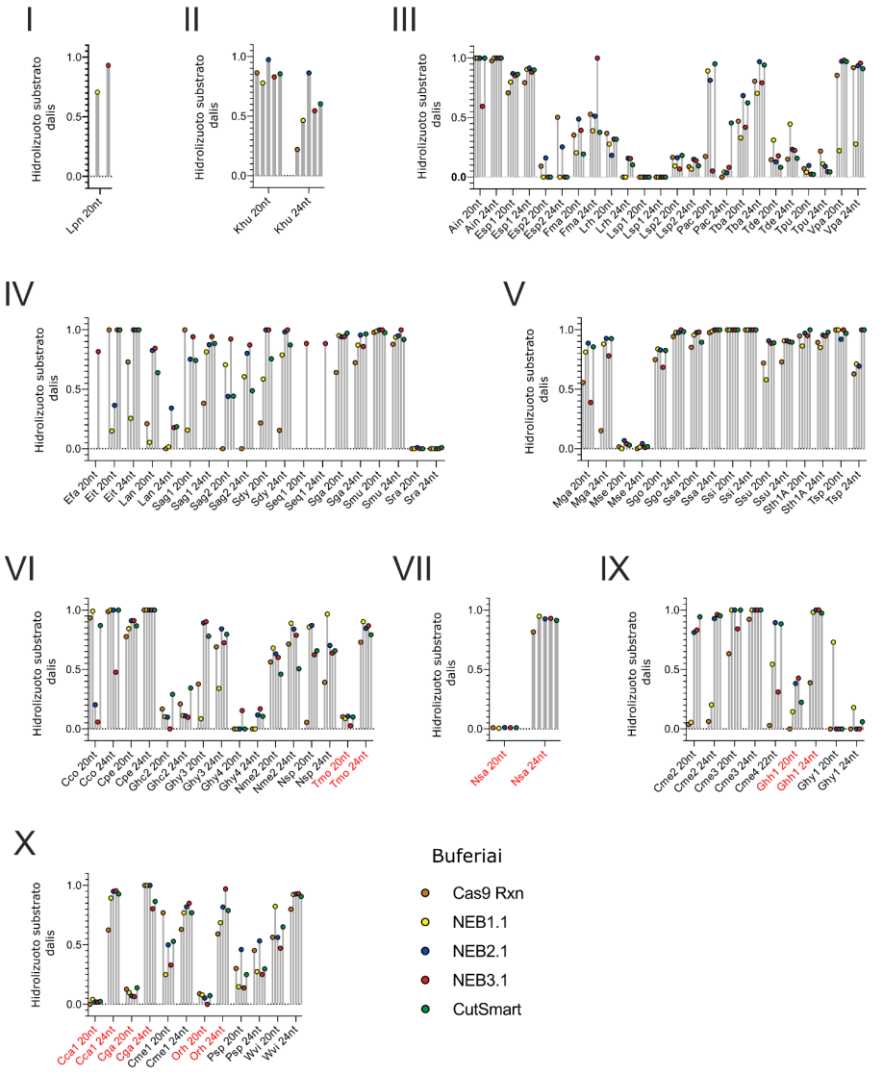
Biocheminė Cas9 aktyvumų analizė

Penkiasdešimt du Cas9 ortologai iš mūsų kolekcijos buvo toliau biochemiškai charakterizuojami naudojant išgrynintus komponentus (baltymus ir sgRNR). Šios nukleazių imties atrankos kriterijai buvo grindžiami trumpu PAM atpažįstamumu (≤3 bp) (jei įmanoma), išlaikant filogenetinio pasiskirstymo ir baltymų dydžio įvairovę.

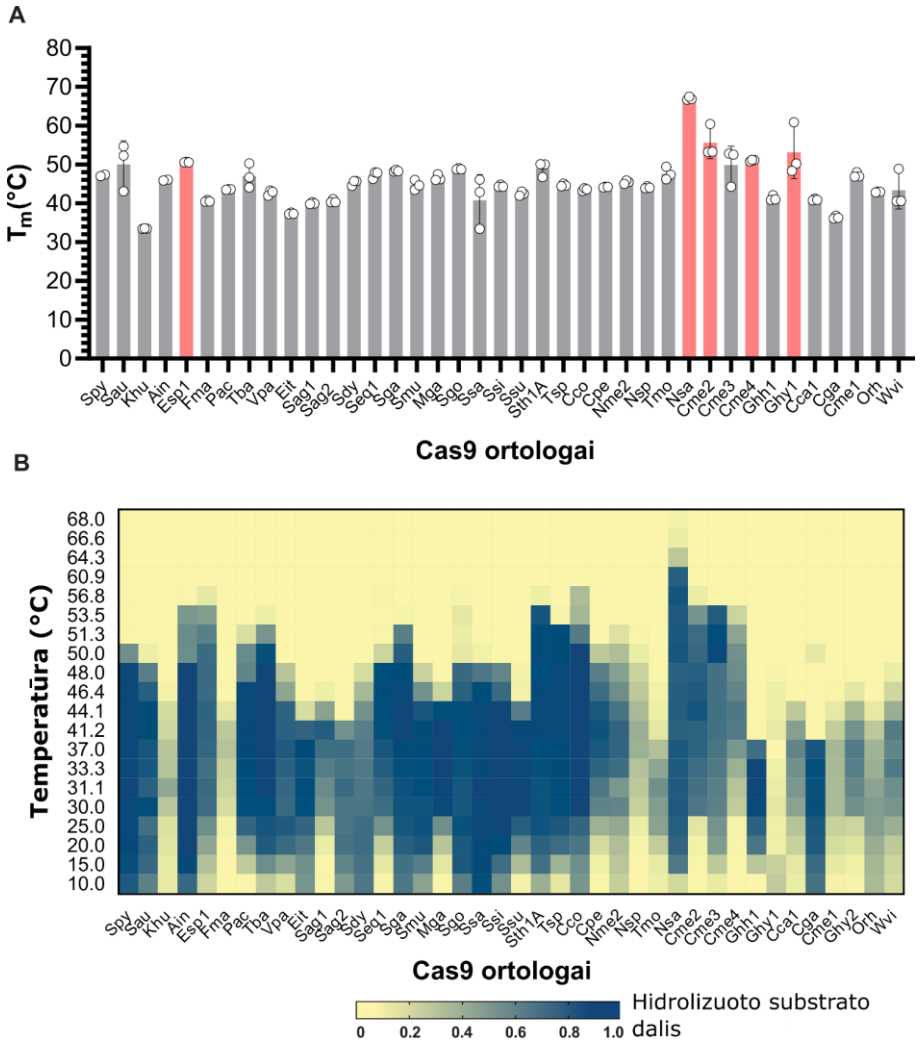
Atsižvelgiant į tai, kad anksčiau buvo pranešta, jog kai kuriems Cas9 baltymams, kad jie veiktų efektyviai, reikalingas ilgesnis nei 20 nt (kuris yra optimalus Spy Cas9) vedančiosios RNR skirtukas (Edraki et al., 2019; Harrington et al., 2017; E. Kim et al., 2017; Ran et al., 2015), kiekvienam ortologui sukūrėme sgRNR su dviem skirtingais skirtukų ilgiais, 20 ir 24 nt, kad iš pradžių įvertintume skirtuko ilgio įtaką Cas9 hidrolizės aktyvumui *in vitro* (5 pav.). Dauguma ortologų veikė panašiai su abiejų ilgių skirtukais, tikrinant nukleazinį aktyvumą 5 skirtinguose buferiuose, tačiau šešiems ortologams - Cga, Cca1, Orh, Tmo, Nsa ir Ghh1 Cas9 - norint efektyviai nukirpti DNR taikinį, reikėjo ilgesnių nei 20 nt skirtukų (5 pav., pažymėta raudonai).

Toliau įvertinome trisdešimt aštuonių ortologų, pasižyminčių efektyviu taikinio kirpimu, terminį stabilumą, naudodami nano-diferencinę skenuojančią fluorimetriją (nanoDSF). 36 iš 38 baltymų lydymosi temperatūra buvo >37 °C, o tai patvirtina stabilumą standartinėmis *in vitro* fermentinių reakcijų sąlygomis. Įdomu tai, kad penkių ortologinių baltymų lydymosi temperatūra buvo >50 °C, o tai rodo termostabilumą (6A pav.). Tai Cme2, Cme4, Ghy1, Esp1 ir Nsa Cas9.

Siekiant patvirtinti nanoDSF prognozes, DNR taikinio hidrolizė buvo matuojama reakcijose, kurių temperatūra svyravo nuo 10 °C iki 70 °C. Visi Cas9 ortologai pasižymėjo plačiu temperatūrinių priklausomybių spektru, įskaitant siaurus ir plačius aktyvumo diapazonus (6B pav.). Kaip stebėta ir nanoDSF analize, Cme2, Esp1, Nsa, Ain, Cme3 ir Sth1A buvo aktyvios esant aukštesnei nei 50 °C temperatūrai, o Nsa, išskirtas iš giliavandenės hidroterminės ventiliacijos kamino bakterijos *Nitratifractor salsuginis*, išliko aktyvus esant aukštesnei nei 60 °C temperatūrai. Be to, vienas ortologas, Ssa, išlaikė 95 % savo hidrolizės aktyvumo 10 °C temperatūroje. Taip pat pastebėjome, kad 5 Cas9 ortologai (Cme2, Cme4, Nsp, Khu ir Fma) išlaikė mažiau nei 25 % aktyvumą, kai reakcijos temperatūra buvo 25 °C arba žemesnė.



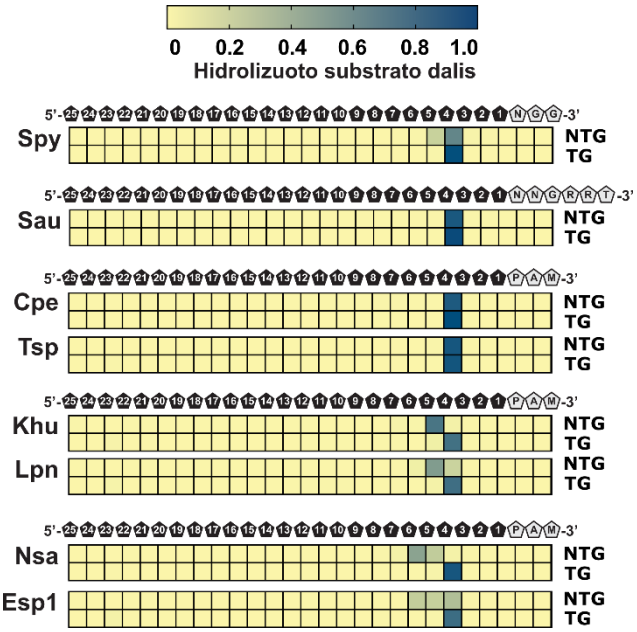
5 pav. Išgrynintų Cas9 ortologų optimalių skirtukų ilgis. Išgrynintų Cas9 baltymų *in vitro* DNR hidrolizės aktyvumas buvo tikrinamas naudojant 20 ir 24 nt sgRNR skirtukus. Kiekviename taške pavaizduotas vienas hidrolizės tikrinimas keliomis buferių sąlygomis, kaip nurodyta atitinkama spalva. Rudi, geltoni, mėlyni, raudoni ir žali apskritimai žymi reakcijas, atliktas atitinkamai su Cas9 reakcijos buferiu (Cas9 Rxn), NEBuffer 1.1, NEBuffer 2.1, NEBuffer 2.1, NEBuffer 3.1 arba CutSmart buferiu. Priklausomybė filogenetinei grupei žymima romėniškais skaitmenimis. Raudonai pažymėti ortologai, kuriems reikalingi 24 nt skirtukai.



6 pav. Cas9 temperatūrinė priklausomybė. **(A)** Išgrynintų Cas9 baltymų lydymosi temperatūra nustatyta naudojant nanoDSF. Raudonai pažymėtos ortologų, kurių lydymosi temperatūra viršija 50 °C, histogramos. Atviri apskritimai žymi atskirų eksperimentų T_m vertes. Paklaidų juostos rodo n = 3 nepriklausomų matavimų standartinį nuokrypį, o kiekvienos iš jų centras žymi nustatymų vidurkį. **(B)** Cas9 ortologų DNR hidrolizės aktyvumas buvo matuojamas atliekant DNR hidrolizės tyrimus *in vitro*, naudojant fluoroforais žymėtus dvigrandinės DNR (dgDNR) substratus. Suskaidyti fragmentai buvo įvertinti kiekybiškai ir pavaizduoti *heatmap* diagramoje, kurioje parodytas bendras aktyvumas esant nuo 10 °C iki 68 °C temperatūrai. Mėlynos spalvos intensyvumas rodo hidrolizuoto substrato dalį.

Siekdami ištirti Cas9 DNR hidrolizės metu susiformuojančius DNR galus, sukūrėme metodą, leidžiantį vienu metu užfiksuoti abi taikinio skilimo metu susidariusias galūnes. Kai kurie ortologai, išvedus vidurkį iš 5 skirtingų

taikinio vietų, nuolat generavo 1 ar daugiau nukleotidų išsikišusius galus (pvz., Khu, Lpn, Nsa ir Esp1) (7 pav.). Šiais atvejais buvo atgauti tik į 5' išsikišę skilimo produktai, kuomet ne-taikinio grandinė dažniausiai baigdavosi keliose padėtyse, o tai rodo RuvC domeno medijuojamos hidrolizės pozicijos arba po hidrolizės atliekamo RuvC domeno medijuojamo "trumpinimo" variaciją. Taikinio grandinė buvo kerpama beveik išimintinai tarp 3 ir 4 protoskirtuko padėčių (7 pav.).



7 pav. Cas9 ortologų taikomi dgDNR hidrolizės modeliai. Kirpimo vietos ir susiformavę dgDNR galai atvaizduoti *heatmap* diagramomis, rodantys, kokia dalis nukirptų galų aptikta kiekvienoje taikinio DNR vietoje atlikus DNR sekoskaitą. Mėlynos spalvos intensyvumas rodo, kokia dalis atvaizduotų nukirptų galų buvo aptikta. Bukų, vienos bazės 5′-iškyšų ir kelių bazių 5′ iškyšų produktų pavyzdžiai rodo, kokia dalis hidrolizės produktų DNR galų yra vidutinė kiekvienoje padėtyje penkiuose skirtinguose dgDNR taikiniuose. DNR bazių ir PAM sekų padėtis pavaizduota virš *heatmap* diagramų. NTG - ne-taikinio grandinė; TG - taikinio grandinė.

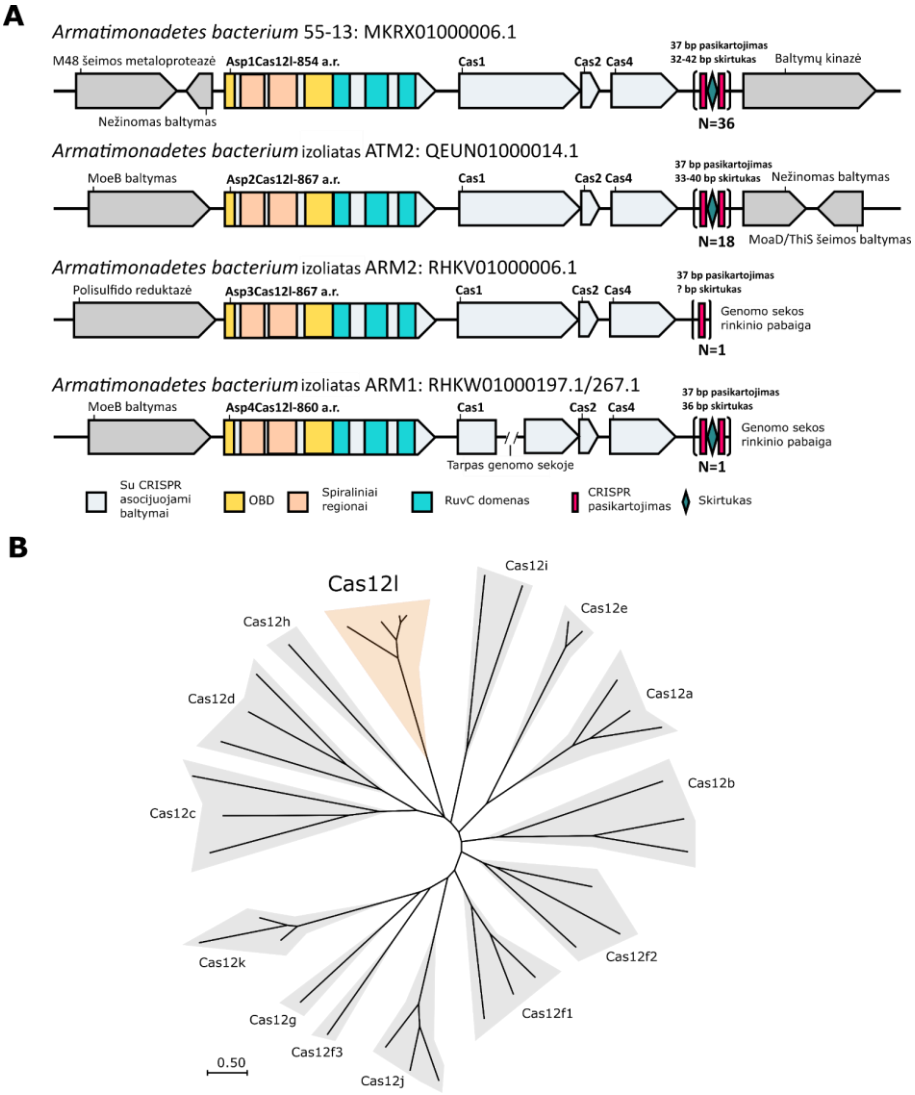
V tipo CRISPR-Cas nukleazių charakterizavimas

Nors atlikus Cas9 nukleazių charakterizavimo tyrimą buvo gautas fermentų, pasižyminčių įvairiomis biocheminėmis savybėmis, rinkinys ir jų, kaip tikrų genomo redaktorių, potencialas dar nėra visapusiškai ištirtas, jie susiduria su iššūkiu, kai yra lyginami su kanoniniu *Streptococcus pyogenes* Cas9. Tuo tarpu visa CRISPR-Cas tyrimų sfera, regis, orientavosi į gerokai įvairesnes V tipo CRISPR-Cas sistemas (Makarova et al., 2020). Pirmosios šio projekto pusės darbų metu buvo paskelbti įdomūs atradimai, aprašantys naujų V tipo

CRISPR-Cas sistemų, koduojančių kompaktiškas funkcines Cas nukleazes, atradimą ir apibūdinimą (Karvelis et al., 2020; Pausch et al., 2020; Yan et al., 2019). Tai paskatino mus pačius ištirti akivaizdžią natūralią V tipo sistemų įvairovę.

CRISPR-Cas12l sistemų identifikavimas

Pradėjome ieškoti su CRISPR susijusių nukleazių, turinčių vieną RuvC domeną, užkoduotą operono tipo organizacijoje su *cas1* ir *cas2* genais, mikrobų sekų duomenų rinkiniuose. Tuomet naujoms sistemoms apibrėžti buvo naudojama CRISPR lokusų genų architektūra, struktūrinė numanomos efektorinės nukleazės patikra ir filogenetinė analizė. Taikant šią metodiką, su V tipo CRISPR susijusi nauja nukleazių šeima atrasta *Armatimonadetes* rūšyje, kurių seka buvo užfiksuota metagenominiuose tyrimuose, skirtuose nuotekų valyme dalyvaujančių mikrobų bendrijoms nustatyti (Kantor et al., 2017; Zhao et al., 2018). Kiekviename lokuse buvo genai, kartu koduojantys adaptacijai reikalingus baltymus (Cas1, Cas2 ir Cas4) ir kompaktišką (~860 aa) efektorinę nukleazę, besiribojančią su CRISPR regionu (8A pav.). Spėjamų efektoriaus ir adaptacijos genų išdėstymas ir orientacija CRISPR lokuse (*cas nukleazė*, *cas1*, *cas2*, *cas4* ir CRISPR regionas) priminė II-B tipo CRISPR-Cas9 sistemų (Koonin & Makarova, 2019) (8A pav.) padėtį ir orientaciją. Spėjamos nukleazės sekos analizė patvirtino, kad baltymo C-galinėje pusėje yra vienas į RuvC panašus trigubai perskeltas domenas, panašiai į kitus Cas12 baltymus (Shmakov et al., 2017; Zetsche, Gootenberg, et al., 2015) (8A pav.). Priešingai, N-galinės pusės seka labai skyrėsi nuo kitų Cas12 nukleazių. Nepaisant to, galima numatyti, kad ji sudaro oligonukleotidų surišimo domeną (angl. *oligo binding domain*, OBD), padalytą dviejų spiralinių sričių, kuriose yra į tilto spiralę (angl. *bridge helix,* BH) panašus motyvas ir spiralės-posūkio-spiralės (angl. *helix-turn-helix,* HTH) DNR surišimo domenas (8A pav.). Nukleazių šeimos filogenetinė analizė patvirtino jų priskyrimą V tipo efektoriams ir parodė, kad jos sudaro naują potipį, kuris skiriasi nuo anksčiau aprašytų Cas12 baltymų (8B pav.). Siekdami supaprastinti nomenklatūrą ir suderinti ją su CRISPR nukleazių pavadinimų suteikimo tvarka (Makarova et al., 2020), pasiūlėme jas klasifikuoti kaip Cas12l.

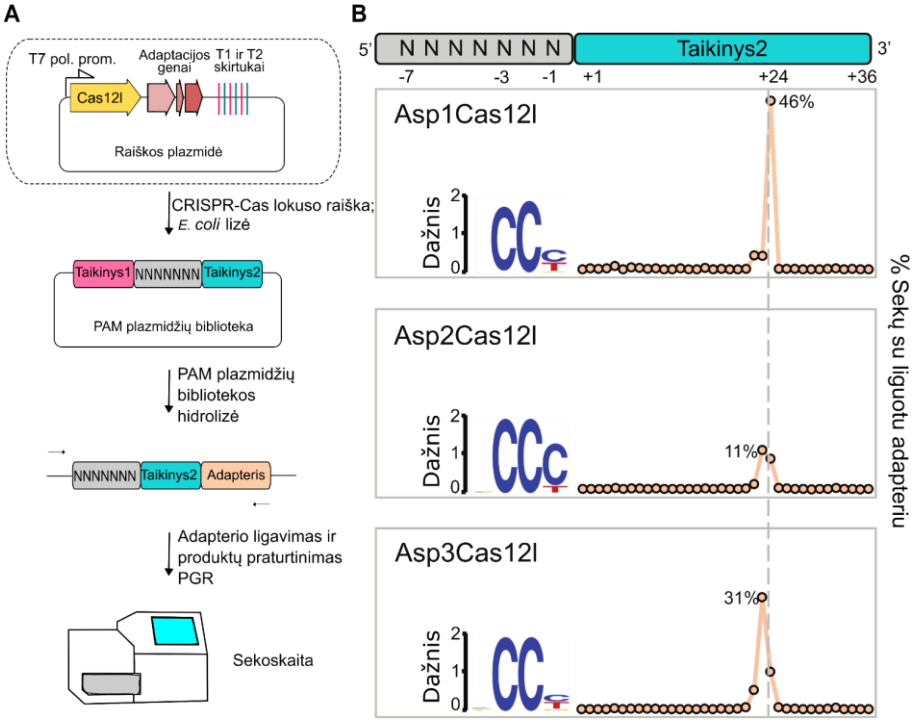


8 pav. Cas12l yra nauja V tipo su CRISPR susijusių nukleazių šeima. (A) Cas12l lokuso architektūros schema. Su CRISPR susiję genai pavaizduoti šviesiai mėlyna spalva, o kiekvienas identifikuotas lokusas koduoja vieną maždaug 860 aa efektorių, turintį, oligonukleotidus surišantį domeną (OBD, geltona spalva), vieną trigubai perskeltą RuvC nukleazės domeną (žydra spalva) ir spiralės-posūkio-spiralės motyvą (rožinė spalva), po kurio seka Cas1, Cas2 ir Cas4 adaptacijos baltymai. (B) Didžiausio tikėtinumo filogenetinis medis, iliustruojantis Cas12l ir kitų V tipo CRISPR-Cas baltymų sekų ryšį.

Cas12l PAM nustatymas

Siekdami nustatyti, ar Cas12l sistemos gali hidrolizuoti dgDNR, Cas12l baltymus ir numanomas vendančiąsias RNR *E. coli*, kurių lizatus naudojome 7N atsitiktinių nukleotidų PAM bibliotekos hidrolizei, panašiai kaip aprašyta

anksčiau (Karvelis et al., 2020) (9A pav.). Tai buvo atlikta pirmiausia modifikuojant Cas12l CRISPR regionus, įterpiant skirtukų sekas, galinčias nukreipti į abi bibliotekos PAM regiono puses, kad būtų atsižvelgta į 5' ir 3' galimas PAM orientacijas (9A pav.). Tada sistemos buvo susintetintos ir klonuotos į IPTG indukuojamas išraiškos plazmides ir transformuotos į *E. coli*. Indukavus raišką, ląstelės buvo suardytos, o nuskaidrintas lizatas apjungtas su PAM biblioteka. Tada hidrolizės produktai buvo užfiksuoti dgDNR adapterio ligavimu, praturtinti PGR ir apdoroti Illumina sekoskaitos metodu (9A pav.). Tuomet DNR hidrolizė buvo įvertinta pagal su adapteriu liguotų sekų dažnį kiekvienoje protoskirtuko taikinio padėtyje.

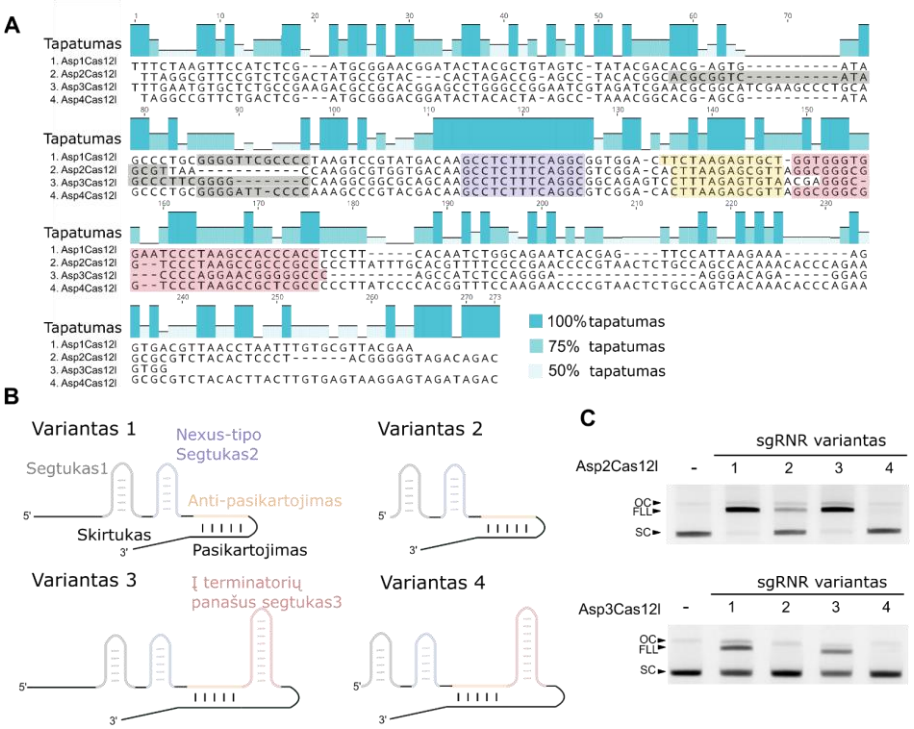


9 pav. Cas12l nukleazės hidrolizuoja dgDNR, esant 5' C-praturtintai PAM sekai. **(A)** Darbo eiga, naudota siekiant nustatyti dsDNA hidrolizę ir Cas12l CRISPR sistemų PAM atpažinimą. **(B)** WebLogo diagramos PAM sekų, kurios palaikė taikinio atpažinimą ir hidrolizę bei aptiktų sekų su liguotu adapteriu dažnis ir pozicija taikinio atžvilgiu.

Sekos, asocijuojamos su adapterio ligavimosi dažnio šuoliais, buvo patikrintos dėl tapatumo PAM bibliotekos taikinio srityje, ir, jei buvo nustatyta, panaudotos kaip taikinio hidrolizės įrodymai (9B pav.). Asp1, Asp2 ir Asp3Cas12l po 23-iosios ir 24-iosios T2 taikinio srities pozicijų buvo stebimi hidrolizės signalai, o atitinkami bibliotekos fragmentai pirmenybę teikė 5' C-praturtintam PAM motyvui (9B pav.).

Cas12l vedančiosios RNR identifikavimas

Toliau buvo nustatyta vedančioji RNR, atsakinga už stebėtą Cas12l efektorių atliekamą dgDNR hidrolizę. Pirmiausia buvo ištirta nekoduojanti sritis tarp nukleazės ir *cas1* genų, ieškant potencialios tracrRNR. Čia buvo nustatytas 12-13 bp regionas, kuriame yra seka komplementari CRISPR pasikartojimui - antipasikartojimas - Asp1, Asp2, Asp3 ir Asp4Cas12l baltymams (10A pav., oranžinė spalva). Tada sekų palyginiai ir antrinių struktūrų prognozės buvo naudojami panašumo sritims nustatyti. Pastebėta, kad Asp1, Asp2, Asp3 ir Asp4Cas12l sekos lygmeniu iš karto 5' nuo antipasikartojimo yra 17 bp sritis, kuri yra 100% identiška tarp sistemų (10A pav., mėlyna spalva). Antrinės struktūros analizė parodė, kad ši sritis, transkribuota kaip RNR, gali sudaryti stiebo kilpos struktūrą, primenančią *nexus* tipo segtuką (10A ir B pav.), pastebėtą kitų Cas9 ir Cas12 sistemų (Briner et al., 2014; Dooley et al., 2021; Faure et al., 2018) tracrRNR. Kitų RNR struktūrų analizė atskleidė papildomą visų keturių sistemų konservatyvumą. Tarp jų buvo



10 pav. Cas12l vedančiosios (-iųjų) RNR identifikavimas ir patvirtinimas. (A) Sekų, tarp genų, koduojančių Cas12l efektorių ir Cas1, palyginys rodo transaktyvuojančios CRISPR RNR (tracrRNR) požymius. Procentinis tapatumas parodytas žydros spalvos atspalviais. Tarp šių požymių aptinkamos sekos: koduojanti 5' segtuką (segtukas1 - pilka spalva), konservatyvi seka, koduojanti į *nexus* panašią stiebo kilpą (nexus-tipo segtukas2 - mėlyna spalva), sritis, galinti sudaryti bazių poras su CRISPR pasikartojimu (Anti-pasikartojimas - oranžinė spalva), ir GC turtinga

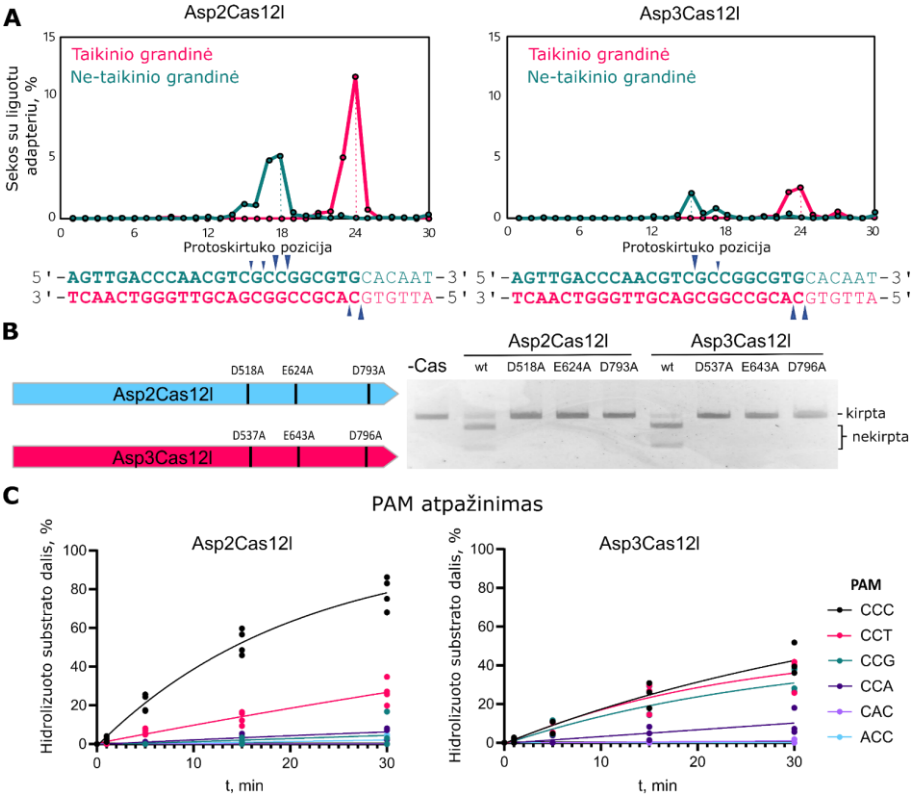
seka, galinti suformuoti į terminatorių panašų segtuką (segtukas3 - raudona spalva). **(B)** Keturi sgRNR dizaino variantai buvo sukurti Asp2Cas12l ir Asp3Cas12l, remiantis antrinės struktūros prognoze, pateikta A dalyje. Jos tarpusavyje skyrėsi spėjamo nestruktūrizuoto 5' tracrRNR regiono ir į terminatorių panašaus segtuko3 buvimu. TracrRNR elementai nuspalvinti analogiškai, kaip pažymėta **(A)**. **(C)** Superspiralizuotos (SC) plazmidinės DNR substratų hidrolizė naudojant išgrynintas Cas12l nukleazes ir atitinkamus sgRNA variantus. Efektyvus viso ilgio substrato linijinis suskaidymas (FLL) dėl visiško dvigubos grandinės trūkio buvo pastebėtas tik naudojant sgRNR, turinčias 5' nesutrumpintą tracrRNR galą (1 ir 3 variantai). Be to, į terminatorių panašus segtukas3 nėra būtinas taikiniui hidrolizuoti (1 variantas). OC - atvira žiedinė DNR; FLL - pilno ilgio linijinė DNR; SC – superspiralizuota DNR.

1 ir 3 segtukai, esantys atitinkamai 5' nuo į *nexus* panašios kamieno kilpos ir 3' nuo antipasikartojimo (10A ir B pav., pilka ir raudona spalvos).

Norint patvirtinti šias sekas kaip tikras tracrRNR, buvo sukurtos sgRNR, susiejant 3' spėjamų tracrRNR galą su 5' CRISPR pasikartojimo galu su lanksčiu juntuku, 5'-GAAA-3'. Tada, naudojant išgrynintas Asp2 ir Asp3Cas12l nukleazes, *in vitro* patikrintas gebėjimas medijuoti dgDNR hidrolizę. Remiantis atitinkama numanoma tracrRNR seka (10A pav.), kiekvienam efektoriniam baltymui buvo sukurti keturi galimi sgRNR dizaino variantai. sgRNR variantai skyrėsi savo ilgiu tracrRNR 5' gale, kuris, kaip pagal antrinės struktūros spėjimas, buvo nestruktūrizuotas, ir į terminatorių panašaus segtuko buvimu (10B pav.). Galop tik sgRNR variantai, turintys 5' nenutrumpintą tracrRNR dalį, leido skelti dsDNA, o segtukas 3 sumažino hidrolizės efektyvumą (10C pav.).

Biocheminis Cas12 dgDNR hidrolizės aktyvumo charakterizavimas

Toliau buvo įvertintos taikinio grandinės (TG) ir ne-taikinio grandinės (NTG) dgDNR hidrolizės vietos. Išgrynintų Asp2 ir Asp3Cas12l baltymų ir sgRNR kompleksai (RNP) buvo naudojami 7N PAM plazmidžių bibliotekai hidrolizuoti, o produktai tikrinti naudojant adapterio ligavimą, produktų PGR praturtinimą ir sekoskaitą. Kirpimo vieta TG buvo tokia pati, kaip stebėta naudojant anksčiau minėtą *E. coli* lizatais paremtą metodą, nors dauguma skilimų vyko iš karto po 24 nt į 3’ pusę nuo PAM, o NTG kirpimas vyko 15-18 nt 3' nuo PAM (11A pav.).



11 pav. Cas12l dgDNR hidrolizės biocheminis charakterizavimas. (A) Asp2Cas12l ir Asp3Cas12l taikinio ir ne-taikinio grandinių kirpimo vietos. Sekų su liguotu adapteriu dažnis iš PAM bibliotekos hidrolizės eksperimentų susietas su protoskirtuko taikinio ir ne-taikinio grandinėms. Taikinio grandinė kerpama 23-24 nt 3' nuo PAM, o ne-taikinio grandinė kerpama 15-18 nt nuo PAM. **(B)** Alanino aminorūgščių mutacijos sutrikdo dgDNR hidrolizę, patvirtinant pagrindines katalitines aminorūgštis Asp2Cas12l ir Asp3Cas12l RuvC nukleazės domene. wt - laukinio tipo. **(C)** Oligodupleksinių dgDNR substratų hidrolizė su išgrynintais RNP kompleksais patvirtina PAM atpažinimą. RNP ir substrato molinis santykis buvo nedidelis (5:1), kad reakcija būtų griežtesnė. Asp2Cas12l pagrinde atpažįsta 5'-CCY-3' PAM, o Asp3Cas12l - 5'-CCB-3' PAM. Taškai žymi duomenis iš n = 4 (Asp2Cas12l) ir n = 3 (Asp3Cas12l) nepriklausomų eksperimentų. Duomenys atvaizduoti grafiškai pritaikius eksponentinės asociacijos kreivės modelį (vientisos linijos).

Tuomet buvo patvirtinta, kad į RuvC panašus motyvas, nustatytas Cas12l efektoriuose, yra atsakingas už stebėtą nukleazinį aktyvumą. Tai buvo atlikta įvedus alanino aminorūgštis vietoje pagrindinės katalitinės D-E-D triados nukleazės branduolio Asp2 ir Asp3Cas12l efektoriuose. Kiekvienas pakeitimas atskirai panaikino dgDNR hidrolizę (11B pav.).

Siekiant patvirtinti PAM nustatymo tyrimo rezultatus naudojant *E. coli* lizatus, Asp2 ir Asp3Cas12l PAM atpažinimas buvo dar kartą tikrintas

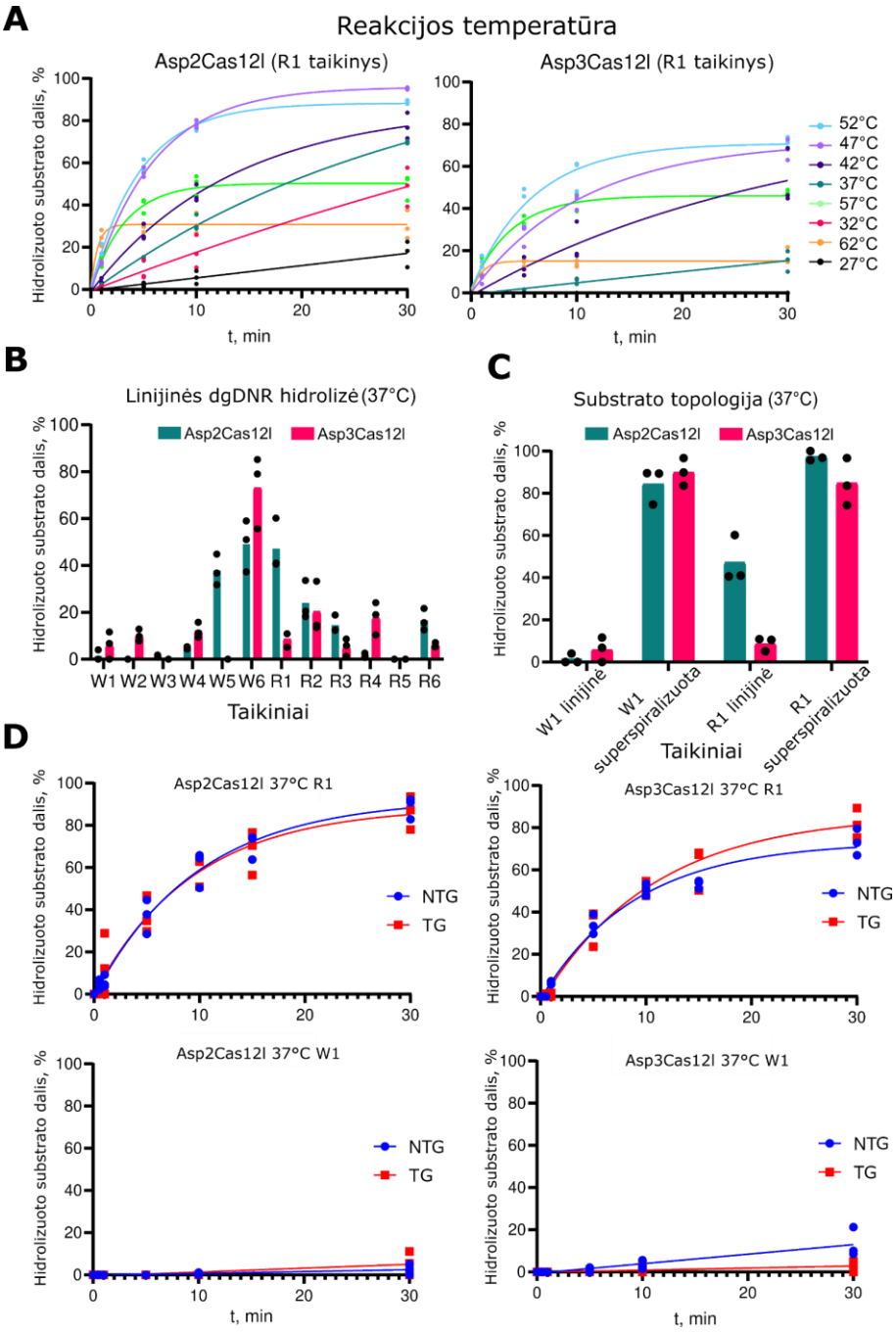
naudojant išgrynintus baltymų ir sgRNR preparatus bei substratus su fiksuotomis neatsitiktinėmis PAM sekomis. Šiuo atveju C pakeitimas -2 ir -3 pozicijose protoskirtuko atžvilgiu visiškai panaikino dsDNA taikinio hidrolizę tiek Asp2, tiek Asp3Cas12l efektoriams. Asp2Cas12l -1 padėtyje pirmenybę teikė C ir toleravo T (11C pav.). Asp3Cas12l šioje pozicijoje ne taip griežtai priima C, T ir G, o su A aktyvumas buvo silpnas (11C pav.).

Toliau buvo įvertintas temperatūros poveikis dgDNR taikinio hidrolizei. Šiems eksperimentams kaip substratai buvo naudojami linijiniai dgDNR fragmentai. Kaip parodyta 12A paveiksle, tiek Asp2, tiek Asp3Cas12l RNP kompleksai optimaliai veikė esant aukštesnei temperatūrai, maždaug 50 °C, o Asp2Cas12l aktyvumas pasireiškė esant platesniam temperatūrų diapazonui. Po 30 min. Asp2Cas12l hidrolizavo daugiau kaip 40 % substrato esant visoms temperatūroms, išskyrus 27 °C, o Asp3Cas12l aktyvumas smarkiai sumažėjo esant žemesnei nei 42 °C temperatūrai. Aukštesnioji temperatūros riba buvo virš 52 °C, kuomet abiejų fermentų aktyvumas mažėjo.

Po to buvo įvertintas Cas12l dgDNR hidrolizės aktyvumo variacija tarp skirtingų taikinių. Tam buvo atrinkta dvylika taikinių iš dviejų terapeutiškai aktualių žmogaus genų - WTAP ir RunXI, PGR būdu amplifikuoti DNR fragmentai ir 37 °C temperatūroje įvertintas hidrolizės aktyvumas naudojant Asp2 ir Asp3Cas12l nukleazes *in vitro*. Stebėtina tai, kad dauguma taikinių nebuvo veiksmingai hidrolizuoti (12B pav.).

Kadangi daugumoje ankstesnių biocheminių eksperimentų buvo naudojami plazmidinės DNR substratai, manėme, kad Cas12l nukleazėms gali būti priimtinesni superspiralizuoti substratai. Norint tai patikrinti, dgDNR hidrolizės efektyvumas R1 ir W1 taikiniuose buvo įvertintas tiek linijinėje, tiek superspiralizuotoje formose. Asp2 ir Asp3Cas12l efektyviai kirpo abu taikinius, kai jie buvo superspiralizuotoje DNR (12C pav.). Priešingai, kai buvo naudojami linijiniai substratai, Asp2Cas12l iš dalies hidrolizavo tik R1 taikinį (12C pav.).

Kadangi dgDNR nikavimo (tik vienos grandinės hidrolizės) produktų dėl nepilnai kirpto taikinio negalima diskriminuoti naudojant linijinį dsDNA substratą, siekėme patikrinti, ar Cas12l fermentai išimintinai taikosi į tam tikrą DNR grandinę. Fluoroforais pažymėti oligodupleksai, kuriuose buvo R1 ir W1 taikiniai, buvo tikrinti su Asp2 ir Asp3Cas12l RNP kompleksais, siekiant nustatyti atitinkamai NTG ir TG hidrolizės efektyvumą. Kaip parodyta 12D pav., Asp2 ir Asp3Cas12l RNP kirpo NTG ir TG panašiu greičiu.



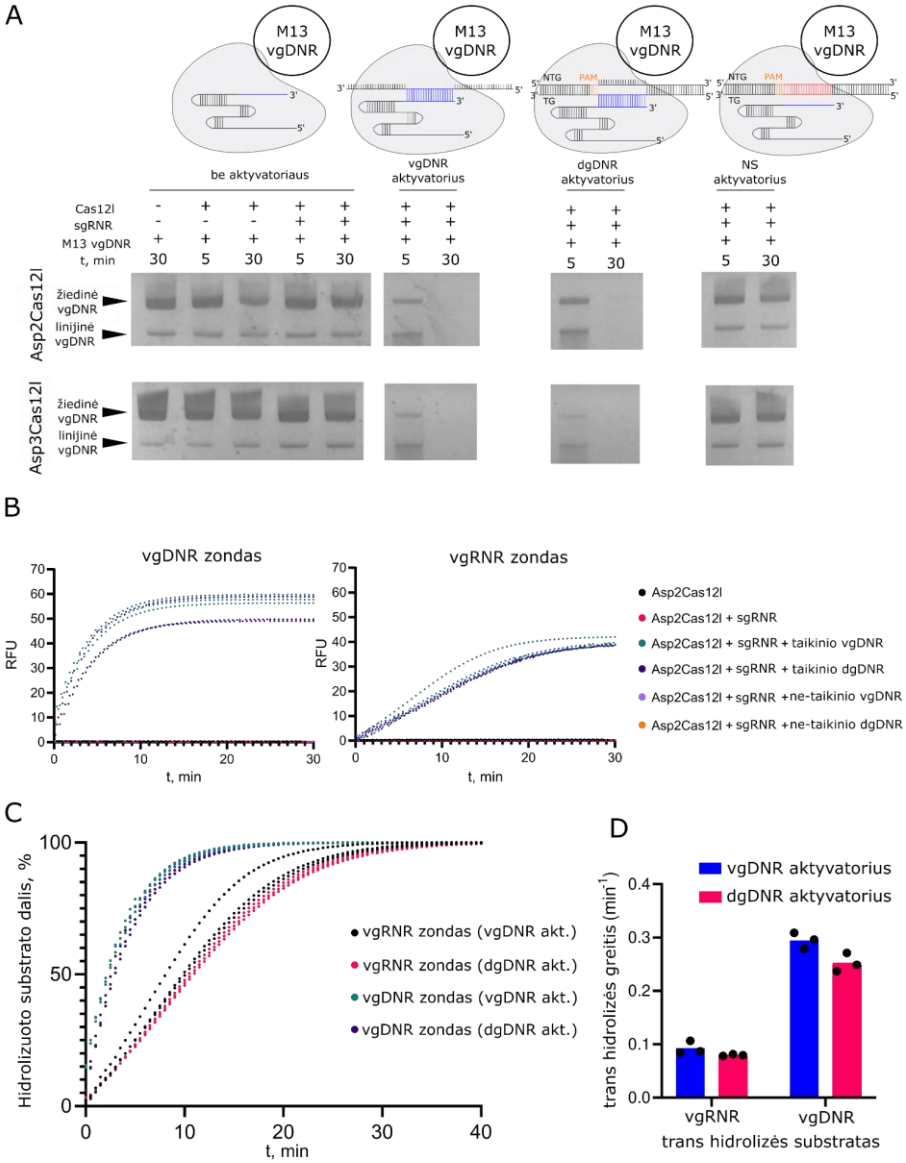
12 pav. Cas12l dsDNA skilimo priklausomybė nuo temperatūros ir substrato topologijos. **(A)** Reakcijos temperatūros poveikis Asp2Cas12l ir Asp3Cas12l dgDNR hidrolizei. Optimali Asp2Cas12l ir Asp3Cas12l RNP kompleksų dgDNR hidrolizės temperatūra yra ~50 °C. **(B)** Cas12l dgDNR hidrolizės efektyvumas priklauso nuo taikinio sekos. **(C)** Asp2Cas12l ir Asp3Cas12l RNP kompleksais paveikti linijiniai ir superspiralizuoti dgDNR substratai su skirtingomis protoskirtukų sekomis (W1 ir R1).

Linijinės topologijos taikinys pastebimai kirptas tik R1 protospkirtuko atveju, tačiau abu baltymai panašiai efektyviai kirpo abu taikinius, kai jie buvo superspiralizuoti. **(D)** Asp2Cas12l ir Asp3Cas12l panašiu greičiu kerpa abi dgDNR grandines *in vitro*. Hidrolizės eksperimentams buvo naudojami fluoroforais žymėti linijiniai oligodupleksiniai dgDNR substratai. Ne-taikinio (NTG, mėlyna spalva) ir taikinio (TG, raudona spalva) DNR grandinės skaldymo efektyvumas pavaizduotas grafiškai. (B ir C) duomenys pateikti kaip n = 3 pakartojimų iš nepriklausomų eksperimentų vidurkiai. Visais atvejais taškai žymi atskirų n = 3 pakartojimų iš nepriklausomų eksperimentų duomenis. (A ir D) Duomenys atvaizduoti grafiškai pritaikius eksponentinės asociacijos kreivės modelį (vientisos linijos).

Cas12l nespecifinis nukleorūgščių degradavimas

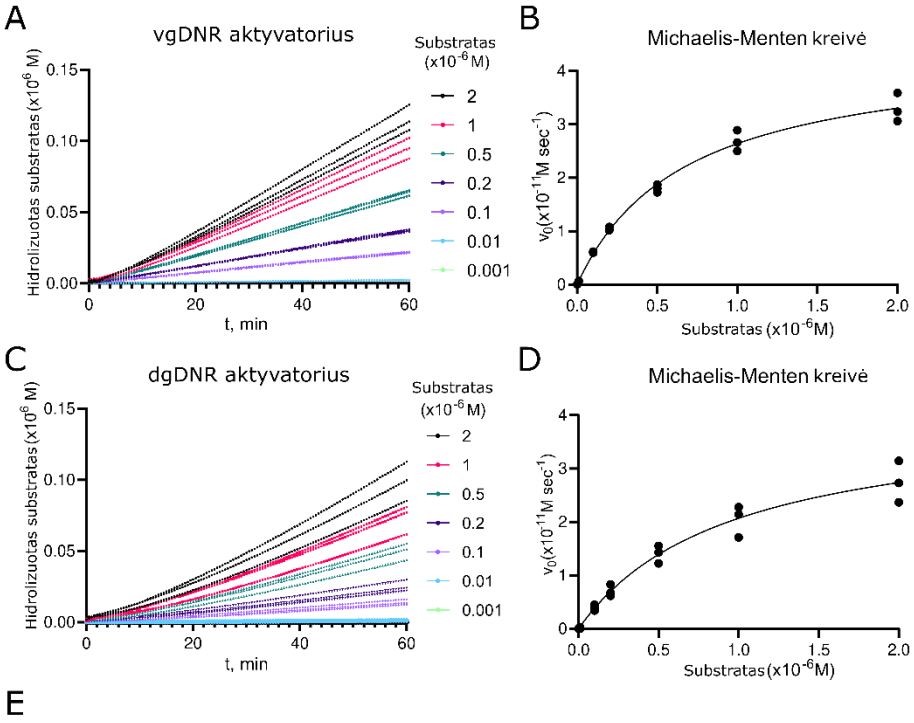
Daugumai V tipo Cas efektorių būdingas bendras bruožas yra nespecifinis *trans* vgDNR degradavimas atpažinus DNR taikinį (J. S. Chen, Ma, et al., 2018; Yan et al., 2019). Siekiant nustatyti, ar Cas12l nukleazės turi šią savybę, Asp2 ir Asp3Cas12l RNP kompleksai buvo inkubuojami su M13 bakteriofago vgDNR esant vgDNR ar dgDNR taikinio substratams, arba be jų. Kaip parodyta 13A paveiksle, abu taikinio substratai lėmė greitą M13 vgDNR degradaciją. Be to, kai abu Cas12l RNP kompleksai buvo inkubuoti su vg arba dgDNR molekule, neturinčia sekos, komplementarios vedančiajai RNR, t. y. nespecifiniu (NS) aktyvatoriumi, M13 vgDNR išliko nepažeista (13A pav.). Tai patvirtina, kad į vgDNazinį panašų Cas12l aktyvumą lemia tik taikinio DNR substrato, komplementaraus Cas12l komplekso sgRNR skirtukui, buvimas.

Toliau buvo tirta nespecifinės vgDNR hidrolizės kinetika, Asp2Cas12l atveju. Tam vgDNR *trans*-hidrolizės greitis buvo matuojamas naudojant fiksuotos koncentracijos Asp2Cas12l-sgRNA RNP kompleksą (0,1 nM), aktyvuotą vgDNR arba dgDNR taikiniais, esant skirtingoms vgDNR reporterio koncentracijoms (nuo 0,001 x 10^{-6} iki 2 x 10^{-6} M) 37 °C temperatūroje. *Trans*-nukleazinis aktyvumas buvo įvertintas nepertraukiamai matuojant fluorescenciją 60 min. Neapdorotos fluorescencijos vertės buvo perskaičiuotos į suskaidyto substrato koncentraciją naudojant standartines kreives, pagrįstas eksperimentų, atliktų be RNP komplekso, ir eksperimentų, kurių metu substratas buvo pilnai hidrolizuotas, duomenimis. Tada duomenys buvo pritaikyti Michaelis-Menten modeliui, atvaizduoti grafiškai ir apskaičiuotas Asp2Cas12l nespecifinės vgDNR hidrolizės aktyvumo greitis esant substrato prisotinimui (k_{cat}) ir ribinėms sąlygoms (k_{cat}/K_M), taip pat ir reporterio koncentracija, užtikrinanti pusę maksimalaus reakcijos greičio (K_M). Iš viso nustatyta, kad fermentas nespecifiškai degraduoja vgDNR 0,44 ir 0,41 molekulių per sekundę greičiu, o k_{cat}/K_m katalitinis efektyvumas yra ~6,5 × 10^5 arba 4,2 × 10^5 s^{-1} M^{-1}, kai naudojamas atitinkamai vgDNR arba dgDNR aktyvatorius (14 pav.).



13 pav. Cas12l nespecifinis nukleazinis aktyvumas. (A) Asp2 ir Asp3Cas12l RNP kompleksai hidrolizuoja M13 vgDNR, esant viengrandinės (vg) DNR arba dvigrandinės (dg) DNA aktyvatoriams su taikinio sekomis, komplementariomis sgRNR skirtukui. NS aktyvatorius - nespecifinis aktyvatorius (vgDNR oligonukleotidas arba dgDNR dupleksas), kurio seka nėra komplementari sgRNR skirtukui. **(B)** Asp2Cas12l RNP kompleksai, aktyvuoti vg arba dgDNR, hidrolizuoja fluoroforu ir slopikliu žymėtus vgDNR (5'-CCCCCCCCCC-3') arba vgRNR (5'-CCCCCCCC-3') zondus. Atvaizduotos fluorescencijos kreivės, pašalinus foninę fluorescenciją. RFU - santykiniai fluorescencijos vienetai. **(C)** Fluorescuojančių vgDNR ir vgRNR zondų hidrolizės efektyvumo kreivės. (D) Nespecifinės *trans*-hidrolizės greičiai, kai vgRNR arba vgDNR reporteriai aktyvuojami vgDNR arba

dgDNR taikiniais. Šalutinis vgDNR hidrolizės aktyvumas yra maždaug 3 kartus didesnis už vgRNR hidrolizės greitį. Fluorescencijos intensyvumai normalizuoti pagal reakcijų, kuriose buvo tik zondas, fluorescenciją, kad būtų atsižvelgta į netobulą reporterių slopinimą arba jų degradaciją. B-D pavaizduoti individualūs taškai atspindi n = 3 nepriklausomų eskperimentų rezultatus. D histogramų aukštis atspindi n = 3 pakartojimų vidurkį.



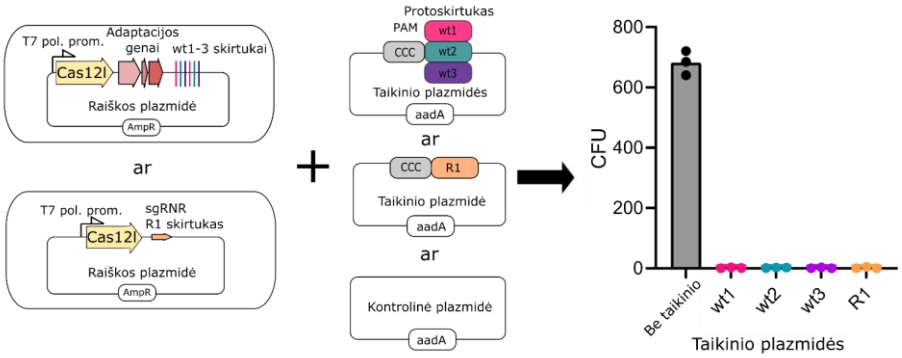
E

Aktyvatorius	k_{cat} (sec^{-1})	K_M (M)	k_{cat}/K_M (sec^{-1} M^{-1})
vgDNR	0.44 ± 0.03	0.67 ± 0.11 × 10^{-6}	6.54 ± 2.72 × 10^5
dgDNR	0.40 ± 0.06	0.95 ± 0.29 × 10^{-6}	4.21 ± 2.06 × 10^5

14 pav. Asp2Cas12l nespecifinio vgDNR hidrolizės aktyvumo Michaelis-Menten analizė. (A) Reporterinio substrato hidrolizės priklausomybės nuo koncentracijos ir laiko kreivės, naudojant vgDNR aktyvatorių. Duomenys pritaikyti linijinės regresijos modeliui, naudojant 0,1 nM efektyvios koncentracijos Asp2Cas12l-sgRNA-aktyvatoriaus kompleksą ir didėjančią vgDNR reporterio koncentraciją. **(B)** Michaelis-Menten modelio pritaikymas aktyvacijai vgDNR. **(C)** Reporterinio substrato hidrolizės priklausomybės nuo koncentracijos ir laiko kreivės, naudojant dgDNR aktyvatorių. Duomenys pritaikyti linijinės regresijos modeliui, naudojant 0,1 nM efektyvios koncentracijos Asp2Cas12l-sgRNA-aktyvatoriaus kompleksą ir didėjančią vgDNR reporterio koncentraciją. **(D)** Michaelis-Menten modelio pritaikymas aktyvacijai vgDNR **(E)** Apskaičiuotos kinetinių konstantų vertės. Duomenys pateikti kaip vidurkis ± standartinis nuokrypis, kur n = 3 pakartojimai iš nepriklausomų eksperimentų. Detali metodologija pateikta Metodų skyriuje.

Asp2Cas12l aktyvumas *E. coli*

Kadangi Asp2Cas12l buvo aktyvesnis 37 °C temperatūroje nei Asp3Cas12l (12A pav.), toliau buvo tikrintas šio fermento gebėjimas riboti *E. coli* transformaciją DNR plazmidėmis. Asp2Cas12l efektorius ir vedančioji RNR buvo išreikšti iš plazmidės, turinčios atsparumo ampicilinui geną (AmpR), dviem konfigūracijomis. Pirmojoje buvo Asp2Cas12l CRISPR lokusas, o antrojoje - tik Asp2Cas12l nukleazės genas ir sgRNR (15 pav.). Abiem atvejais plazmidėje koduotos vedančiosios RNR (CRISPR regione arba sgRNR) buvo suprojektuotos taip, kad būtų nutaikytos į antrą plazmidę, kurioje yra atsparumo streptomicinui (aadA) genas (15 pav.). Tada *E. coli* ląstelės buvo ko-transformuotos su atitinkamomis AmpR ir aadA plazmidėmis ir išsėtos ant terpės, kurioje yra ampicilino ir streptomicino. Tik kontrolinės plazmidės, neturinčios taikinio sekos, atveju buvo stebėtos *E. coli* kolonijos (15 pav.).



15 pav. Asp2Cas12l veikia heterologinėje ląstelėje, apsaugodama nuo invazinės dgDNR. *E. coli* buvo ko-transformuota plazmide, koduojančia indukuojamą Asp2Cas12l geną ir vedančiąją RNR (ekspresijos plazmidės), ir antrąja plazmide, kurioje yra protoskirtuko taikinys šalia Asp2Cas12l tinkamo PAM (5'-CCC-3') (taikinio plazmidė), arba kontroline plazmide be PAM ar protoskirtuko taikinio sekos. Transformantai buvo išsėti ant terpės, kurioje yra T7 ekspresijos induktoriaus (IPTG) ir atitinkamų antibiotikų (karbenicilino ir streptomicino). Kolonijos aptiktos tik tada, kai buvo naudojama kontrolinė plazmidė, neturinti taikinio sekos. AmpR - atsparumo ampicilinui (karbenicilinui) genas, aadA - atsparumo streptomicinui genas. Taškai atspindi duomenis iš n = 3 nepriklausomų eksperimentų, histogramos aukštis rodo duomenų vidurkį.

Apibendrinimas

Čia parodėme, kad natūralią CRISPR-Cas sistemų įvairovę galima pasitelkti naujų RNR-vedamų specifinių nukleazių identifikavimui ir vystymui. Tai pavyko pasiekti taikant plataus spektro atrankinius tyrimus, kaip didelio II tipo Cas9 ortologų rinkinio atveju, kuomet buvo identifikuotos veiklios Cas9

nukleazės, pasižyminčios įvairios sudėties PAM atpažinimo sekomis. Trumpesnes PAM atpažįstantys baltymai galėtų padėti padidinti CRISPR-Cas sistemų redagavimui prieinamą taikinių erdvę, o ilgesnes PAM atpažįstantys ortologai gali užtikrinti didesnį specifiškumą (C. M. Lee et al., 2016; Müller et al., 2016). PAM sąveikaujančių domenų pirminės baltymų struktūros ir empiriškai nustatytų atitinkamų Cas9 ortologų PAM atpažinimo ryšio analizė gali padėti sukurti išsamesnius *in silico* PAM šališkumo prognozavimo modelius naujoms spėjamoms Cas9 nukleazėms, taip pat palengvinti chimerinių Cas9 mutantų generavimą, keičiant jų PI domenus (Ciciani et al., 2022; Ma et al., 2019). Tolesnis biocheminis ortologiškų Cas9 nukleazių dgDNR hidrolizės aktyvumo *in vitro* charakterizavimas atskleidė įvairias temperatūrines priklausomybes, taip išskiriant nukleazes, kurios potencialiai galėtų būti naudojamos psichrofiluose (Yusof et al., 2021) arba termofiluose (Le & Sun, 2022), taip pat fermentus, turinčius gerai apibrėžtą toleruojamų temperatūrų diapazoną, o tai rodo terminę aktyvumo valdymo galimybę (Zhuo et al., 2021). Be to, aptiktos pirmenybės ilgesniems gRNR skirtuko - DNR substrato komplementarumo traktams, taip pat skirtingų DNR galūnių susidarymas po kirpimo, kas galėtų lemti alternatyvius DNR reparacijos rezultatus žinduolių ląstelių redagavime (Y.-W. Fu et al., 2021). Bendrai, tai leidžia atrinkti ortologiškas Cas9 nukleazes, pasižyminčias potencialiai palankiomis savybėmis kitų šiuo metu taikomų Cas9 baltymų atžvilgiu.

Taipogi, identifikavome naują V tipo Cas12 nukleazių šeimą ir taikėme išsamesnę jos apibūdinimo strategiją. Filogenetinė analizė parodė, kad ši sistema skiriasi nuo kitų Cas12 potipių ir gali būti įvardinta, kaip atskiras Cas12l potipis. Šie baltymai atpažįsta C-praturtintą PAM seką, taip sudarydami atsvarą T-turtingoms PAM sekoms, būdingoms daugumai Cas12 efektorių. Toliau nustatėme, kad dgDNR hidrolizei reikalinga tracrRNR, kuri gali būti sujungta su crRNR į sgRNR. Tolesnis biocheminis apibūdinimas *in vitro* atskleidė, kad optimali Cas12l dgDNR hidrolizės temperatūra yra apie 50 °C, superspiralizuotos plazmidinės DNR substratami hidrolizuojami efektyviau nei linijiniai, bei parodytas nespecifinis vgDNR ir vgRNR degradavimas. Idealiam aktyvumui reikalinga padidinta temperatūra yra savybė, kuri gali būti naudinga redagavimui ląstelių tipuose, kuriems priimtinas veikimas terminiu šoku, taip pat nukleorūgščių aptikimo strategijose, kuriose taikomi izoterminio amplifikavimo metodai (J. Joung et al., 2020; Nandy et al., 2019; Q. Wang et al., 2020). Didesnis neigiamai superspiralizuotos DNR hidrolizės efektyvumas stebėtas ir naudojant kitus Cas fermentus (Aelst et al., 2019; Westra et al., 2012), ir, tikėtina, paaiškinamas kaip būdas, skatinantis DNR grandinės išvyniojimą ir R kilpos susidarymą (López-García, 1999). Galiausiai, vienas Cas12l šeimos narys

lėmė *E. coli* transformacijos plazmidine DNR ribojimą. Apibendrinant, tai išplečia vis didėjantį V tipo CRISPR-Cas sistemų kraštovaizdį ir rodo, kad Cas12l šeima gali būti toliau plėtojama kaip potenciali kompaktiška genomų redagavimo priemonė.

Dabartinė pažanga šioje srityje

Vienas iš šiame darbe aprašytų Cas9 ortologų, Mga, buvo panaudotas biojutiklio sistemoje, kuri leido atskirti vieno nukleotido polimorfizmą (SNP) neamplifikuotoje genominėje DNR (Balderston et al., 2021). Mga Cas9 atpažįstama PAM seka sutapo su SNP, asocijuojamu su pjautuvine anemija, kuris mutavusiuose aleliuose buvo pažeistas, todėl buvo sutrikdytas prisirišimas ir hidrolizė didesniu jautrumu nei tuo atveju, kai SNP yra Cas9 protoskirtuko viduje (Balderston et al., 2021). Tai išryškina lanksčios taikinio atrankos, kurią leidžia atlikti įvairi Cas9 fermentų kolekcija, pranašumą.

Vis daugiau kitų Cas9 nukleazės ortologų taikoma bazių redagavimo sistemose, tokiu būdu išplečiant potencialią taikinių erdvę ir keičiant redagavimo langus (Kweon et al., 2023; M. Li et al., 2023; Trasanidou et al., 2023).

Toliau atrandamos ir apibūdinamos naujos ortologiškos Cas9 nukleazės (Cui et al., 2022; S. Gao et al., 2023; J. Wei et al., 2022). Be to, didėjantis Cas nukleazių sekų kiekis suteikia pagrindą kurti kruopštesnius algoritmus, leidžiančius atrasti vis labiau filogenetiškai nutolusias Cas rūšis, kaip pavyzdį galima pateikti senovinių Cas nukleazių tyrimą (Alonso-Lerma et al., 2023) arba naujo Cas9 efektorių potipio atradimą (Goltsman et al., 2022).

Neseniai atskira grupė pranešė apie savo atliktą Cas12l sistemos charakterizavimą (Sun et al., 2023). Bendrai, jų rezultatai patvirtina mūsų tyrime nustatytus rezultatus. Tačiau jiems pavyko išspręsti vieno šeimos nario, Asp3Cas12l, 3D struktūrą. Tai išryškino Cas12l šeimos, pasižyminčios naujais struktūriniais motyvais, nebūdingais kitiems Cas12 efektoriams, skirtingą pobūdį. Be to, grupė įrodė, Cas12l aktyvumą žinduolių ląstelėse, o tai patvirtina šių sistemų potencialą naudoti genomo redagavimo reikmėms.

Labiau fundamentaliame lygmenyje, padaryta didelė pažanga charakterizuojant tikėtinus evoliucinius Cas9 ir Cas12 nukleazių pirmtakus (Shmakov et al., 2017). Manoma, kad Cas9 išsivystė iš IS200/IS605 transpozicinių elementų šeimos IscB baltymų, kurių narius tyrė (Altae-Tran et al., 2021). Nustatyta, kad IscB baltymai yra kompaktiškos (400 aa) funkcionalios dgDNR nukleazės, vedamos didelių RNR molekulių, vadinamų ω-RNR, kurios koduojamos daugybe konfigūracijų skirtinguose IscB lokusuose ir yra tikėtini CRISPR tracrRNR pirmtakai (Altae-Tran et al., 2021). Be to, IscB-ω-RNA kompleksų krio-EM struktūrose matoma panaši į

Cas9 dvisluoksnė baltymų architektūra, tačiau nėra Cas9 REC skilties (Kato et al., 2022; Schuler et al., 2022). Pastebėjus, kad struktūrinė ω-RNR dalis yra analogiškoje padėtyje kaip ir REC skiltis, galima daryti prielaidą, kad Cas9 evoliucijos metu ω-RNR dalys buvo pakeistos baltymų domenais. Be to, IscB kompleksams atpažinti reikalingas į PAM panašus motyvas, kuris vadinamas taikinio gretutiniu motyvu (angl. target-adjacent motif, TAM) ir yra taikinio 3' gale (Altae-Tran et al., 2021). Tuo pat metu buvo pateiktas išsamus tos pačios IS200/IS605 transpozicinių elementų šeimos TnpB baltymo charakterizavimas (Karvelis et al., 2021). Parodyta, kad TnpB baltymas yra kompaktiška (400 aa) funkcionali DNR endonukleazė, vedama RNR molekulės, gautos iš TnpB geno 3' galo ir transpozicinio elemento palindrominio galo, o RNR 3' gale esanti kintama seka lemia taikinio atpažinimą (Karvelis et al., 2021). Kaip ir IscB, TnpB reikia TAM sekos, esančios šalia taikinio, tačiau ji yra 5' gale, panašiai kaip Cas12 sistemose. TnpB krio-EM struktūros dar kartą patvirtino TnpB ir Cas12 efektorių giminystę - kompleksų architektūros labai panašios, ypač lyginant su kompaktiškomis Cas12f nukleazėmis (Sasnauskas et al., 2023). Galiausiai buvo įrodyta, kad tiek IscB, tiek TnpB hidrolizuoja taikinius žmogaus ląstelėse (Altae-Tran et al., 2021; Karvelis et al., 2021). Visa tai apibendrinus, šie rezultatai ne tik suteikia esminės informacijos apie CRISPR-Cas sistemų evoliuciją, bet ir apie tai, kas gali sudaryti minimalų funkcionalų genomo redagavimo įrankį, dėl palyginti mažo jų dydžio.

IŠVADOS

1. 79 II tipo Cas9 ortologams parodytas dgDNR hidrolizės aktyvumas, nulemtas įvairių PAM sekų bei tracrRNR reikalavimų.
2. Cas9 ortologai pasižymi įvairiomis temperatūrinėmis priklausomybėmis, gRNR skirtuko ilgio polinkiais bei dgDNR hidrolizės modeliais *in vitro.*
3. Naujoms V tipo CRISPR-Cas nukleazėms, Cas12l, nusatytas dgDNR hidrolizės aktyvumas, nulemtas dvigubos gidinės RNR ir C-praturtintos PAM sekos atpažinimo.
4. Cas12l fermentai pasižymi optimaliu dgDNR hidrolizės aktyvumu *in* vitro apie 50°C temperatūroje.
5. Cas12l pasižymi pašaliniu nespecifiniu vgRNR ir vgDNR hidrolizės aktyvumu po taikinio atpažinimo.
6. Asp2Cas12l ir gRNR raiška lėmė *E. coli* transformacijos plazmidine DNR ribojimą.

ACKNOWLEDGEMENTS

This doctoral thesis would not have seen the light of day without the significant contributions of several parties.

First of all, I would like to express my gratitude to my supervisor dr. Giedrius Gasiūnas, who provided me the opportunity to work in the CRISPR-Cas field as well as carry out the doctoral studies in the Caszyme company. His guidance and extraordinary patience has been invaluable for my growth as a scientist and overall sanity during the doctoral project.

Furthermore, the doctoral thesis and associated publications would not have been made without the effort of our collaborators at Corteva Agriscience, particularly Joshua K. Young and his team, as well as New England Biolabs and G. Brett Robb and his team.

I would also like to thank the doctoral school in Vilnius University Life Sciences Center for all the assistance provided regarding the organization of the studies, doctoral exams and related processes. I am also grateful to everyone from the LSC community that lent their time to attend and evaluate my doctoral exams.

My thanks go to the rest of the Caszyme team both for all the work done on this study and associated projects, and for fostering a friendly and supportive environment, to be a part of which is an absolute pleasure.

Lastly, to my friends and family – thank you for your support and encouragements during this journey, especially when things looked bleak.

CURRICULUM VITAE

Tomas Urbaitis

LLC “Caszyme”
7C Saulėtekio Ave, LT-10257 Vilnius, Lithuania
Phone no. +370 622 81505
E-mail: tomas.urbaitis@caszyme.com

EDUCATION

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

SCIENTIFIC WORK EXPERIENCE

LLC “Caszyme”
2022-present Research Scientist
2017-2022 Senior Technician
Vilnius University, Institute of Biotechnology, Department of Protein-DNA interactions
2015-2017 Lab Assistant

PUBLISHED SCIENTIFIC RESEARCH

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.L., Mabuchi, M., Fuchs, R.T., Schidlkraut, E., Weigele, P., Jack, W., Robb, G.B., Venclovas, Č., Siksnys, V. A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nature Communications.* 2020 November; 11(1): 5512 doi: 10.1038/s41467-020-19344-1

Urbaitis, T., Gasiunas, G., Young, J.K.*, Hou, Z., Paulraj, S., Godliauskaite, E., Juskeviciene, M.M., Stitilyte, M., Jasnauskaite, M., Mabuchi, M., Robb, G.B., Siksnys, V. A new family of CRISPR-type V nucleases with C-rich PAM recognition. *EMBO reports*. 2022 December; 23(12): e55481 doi: 10.15252/embr.202255481

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.