

CasPEDIA Database: A Functional Classification System for Class 2 CRISPR-Cas Enzymes

Benjamin A. Adler^{1,2†}, Marena I. Trinidad^{1,3†}, Daniel Bellieny-Rabelo^{1,2}, Elaine Zhang^{1,3}, Hannah M. Karp^{1,4}, Petr Skopintsev^{1,2}, Brittney W. Thornton^{1,5}, Rachel F. Weissman^{1,5}, Peter H. Yoon^{1,5}, LinXing Chen^{1,6}, Tomas Hessler^{1,6-8}, Amy R. Eggers^{1,5}, David Colognori^{1,5}, Ron Boger^{1,2}, Erin E. Doherty^{1,2}, Connor A. Tsuchida^{1,9}, Ryan V. Tran⁴, Laura Hofman^{1,2,10}, Honglue Shi^{1,3}, Kevin M. Wasko^{1,5}, Zehan Zhou^{1,5}, Chenglong Xia^{1,2}, Muntathar J. Al-Shimary^{1,5}, Jaymin R. Patel¹, Vienna C.J.X. Thomas^{1,4}, Rithu Pattali^{1,5}, Matthew J. Kan^{1,11}, Anna Vardapetyan¹, Alana Yang^{1,5}, Arushi Lahiri⁵, Michaela F. Maxwell¹², Andrew G. Murdock¹, Glenn C. Ramit¹, Hope R. Henderson¹, Roland W. Calvert¹³, Rebecca S. Bamert¹³, Gavin J. Knott¹³, Audrone Lapinaite¹⁴⁻¹⁶, Patrick Pausch¹⁷, Joshua C. Cofsky¹⁸, Erik J. Sontheimer¹⁹⁻²¹, Blake Wiedenheft²², Peter C. Fineran²³⁻²⁶, Stan J.J. Brouns²⁷⁻²⁸, Dipali G. Sashital²⁹, Brian C. Thomas³⁰, Christopher T. Brown³⁰, Daniela S. A. Goltsman³⁰, Rodolphe Barrangou^{1,31}, Virginus Siksnyš³², Jillian F. Banfield^{1,6-8,33}, David F. Savage^{1,3,5}, Jennifer A. Doudna^{1-5,34-35*}

¹Innovative Genomics Institute, University of California, Berkeley, CA, 94720, USA.

²California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA, 94720, USA.

³Howard Hughes Medical Institute, University of California, Berkeley, CA, 94720, USA.

⁴Department of Chemistry, University of California, Berkeley, CA, 94720, USA.

⁵Department of Molecular and Cell Biology, University of California, Berkeley, CA, 94720, USA.

⁶Department of Earth and Planetary Sciences, University of California, Berkeley, CA, 94720, USA.

⁷Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, 94720, USA.

⁸EGSB Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA.

⁹University of California, Berkeley - University of California, San Francisco Graduate Program in Bioengineering, University of California, Berkeley, Berkeley, CA, 94720, USA.

¹⁰Graduate School of Life Sciences, Utrecht University, 3584 CS Utrecht, UT, the Netherlands.

¹¹Department of Pediatrics, Division of Allergy, Immunology, and Bone Marrow Transplantation, University of California, San Francisco, CA, 94158, USA.

¹²Department of Chemistry and Biochemistry, Hampton University, Hampton, VA, 23668, USA.

¹³Monash Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, VIC 3168, Australia.

¹⁴School of Molecular Sciences, Arizona State University, Tempe, AZ, 85281, USA.

¹⁵Arizona State University-Banner Neurodegenerative Disease Research Center at the Biodesign Institute, Arizona State University, Tempe, AZ, 85281, USA.

¹⁶Center for Molecular Design and Biomimetics, The Biodesign Institute, Arizona State University, Tempe, AZ, 85281, USA.

¹⁷LSC-EMBL Partnership Institute for Genome Editing Technologies, Life Sciences Center, Vilnius University, Vilnius, 10257, Lithuania.

¹⁸Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, 02115, USA.

¹⁹RNA Therapeutics Institute, University of Massachusetts Chan Medical School, Worcester, MA, 01655, USA.

²⁰Program in Molecular Medicine, University of Massachusetts Chan Medical School, Worcester, MA, 01655, USA.

²¹Li Weibo Institute for Rare Diseases Research, University of Massachusetts Chan Medical School, Worcester, MA, 01655, USA.

²²Department of Microbiology and Cell Biology, Montana State University, Bozeman, MT, 59717, USA.

²³Department of Microbiology and Immunology, University of Otago, Dunedin, 9016, New Zealand.

²⁴Genetics Otago, University of Otago, Dunedin, 9016, New Zealand.

²⁵Bioprotection Aotearoa, University of Otago, Dunedin, 9016, New Zealand.

²⁶Maurice Wilkins Centre for Molecular Biodiscovery, University of Otago, Dunedin, 9016, New Zealand.

²⁷Department of Bionanoscience, Delft University of Technology, 2629 HZ Delft, Netherlands.

²⁸Kavli Institute of Nanoscience, 2629 HZ Delft, Netherlands.

²⁹Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA, 50011, USA.

³⁰Metagenomi, Inc., Emeryville, CA, 94608, USA.

³¹Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC, 27606, USA.

³²Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, 10257, Lithuania.

³³The University of Melbourne, Parkville, VIC 3052, Australia.

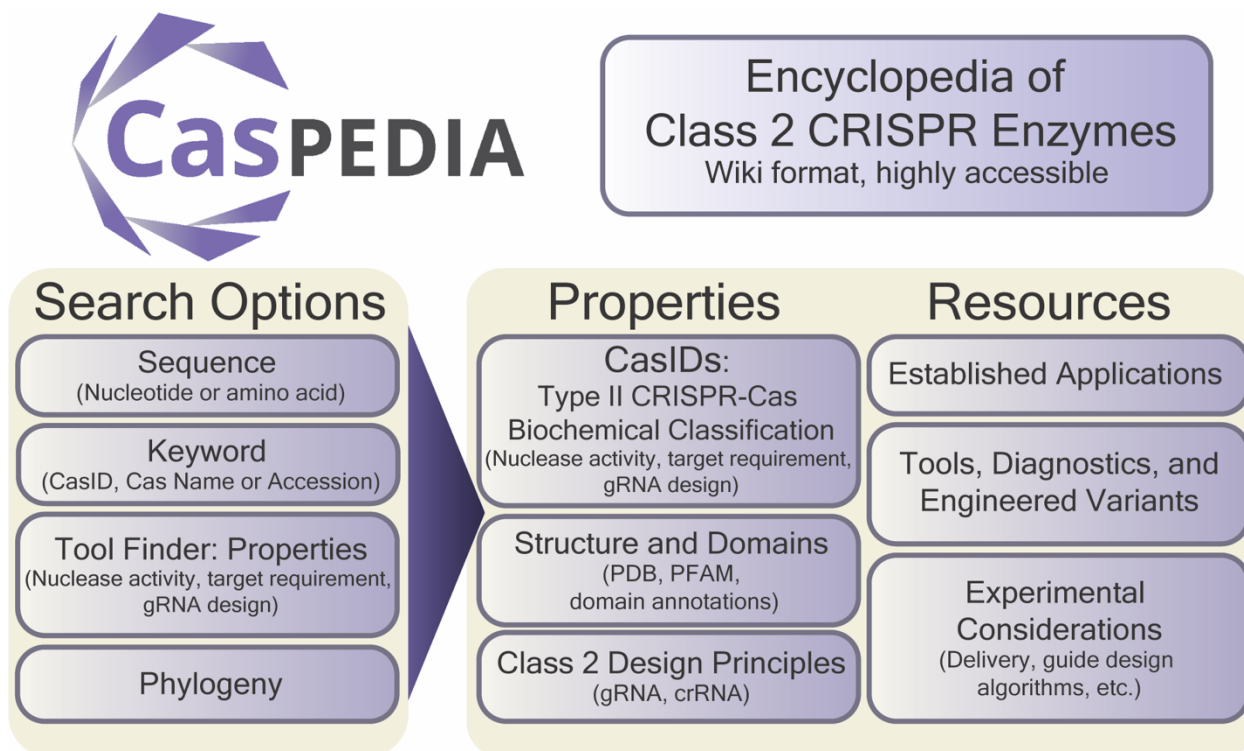
³⁴MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA.

³⁵Gladstone Institutes, University of California, San Francisco, CA, 94158, USA.

† Joint Authors

* To whom correspondence should be addressed. Email: doudna@berkeley.edu

GRAPHICAL ABSTRACT



ABSTRACT

CRISPR-Cas enzymes enable RNA-guided bacterial immunity and are widely used for biotechnological applications including genome editing. In particular, the Class 2 CRISPR-associated enzymes (Cas9, Cas12 and Cas13 families), have been deployed for numerous research, clinical and agricultural applications. However, the immense genetic and biochemical diversity of these proteins in the public domain poses a barrier for researchers seeking to leverage their activities. We present CasPEDIA (<http://caspedia.org>), the Cas Protein Effector Database of Information and Assessment, a curated encyclopedia that integrates enzymatic classification for hundreds of different Cas enzymes across 27 phylogenetic groups spanning the Cas9, Cas12 and Cas13 families, as well as evolutionarily related IscB and TnpB proteins. All enzymes in CasPEDIA were annotated with a standard workflow based on their primary nuclease activity, target requirements and guide-RNA design constraints. Our functional classification scheme, CasID, is described alongside current phylogenetic classification, allowing users to search related orthologs by enzymatic function and sequence similarity. CasPEDIA is a comprehensive data portal that summarizes and contextualizes enzymatic properties of widely used Cas enzymes, equipping users with valuable resources to foster biotechnological development. CasPEDIA complements phylogenetic Cas nomenclature and enables researchers to leverage the multi-faceted nucleic-acid targeting rules of diverse Class 2 Cas enzymes.

INTRODUCTION

CRISPR-Cas (clustered regularly interspaced short palindromic repeats, CRISPR-associated) systems provide adaptive immunity in bacteria and archaea through RNA-guided recognition and Cas-mediated destruction of foreign nucleic acids (1, 2). These immune systems are exceptionally diverse, occurring as 6 types and 27 subtypes in line with recent classification (3). Beginning with the discovery of RNA-guided endonuclease activity conferred by Cas9, insights into the enzymatic activities of CRISPR Cas enzymes have precipitated a veritable wave of biotechnological innovation (4–6).

In particular, the Class 2 Cas enzymes have been a driver of biotechnological development owing to their single-protein nature. Class 2 Cas enzymes can be separated into 3 families: Cas9, Cas12 and Cas13 from Type II, V and VI CRISPR systems respectively (3). Because these proteins all employ a processed CRISPR RNA (crRNA) to guide protein activity towards a sequence of interest, these proteins can all be easily “programmed” to target unique sequences of interest by simple design of a spacer (i.e. targeting) sequence. However, as researchers have explored the genetic diversity of these systems, it has become clear that (i) the RNA-guided biochemical activity, (ii) constraints on targeting context, and (iii) ways crRNAs are processed differ dramatically across - and within - families. While these differences reflect opportunities for biotechnological development, there does not yet exist a centralized resource for comparing biochemical activity to complement existing genetic classification efforts (3). It remains difficult for these enzymes to be functionally compared and contrasted across and within subfamilies.

Here, we present CasPEDIA, <http://caspedia.org>, providing users with summary information about the capabilities and limitations of Class 2 Cas technologies to facilitate tool selection and to highlight opportunities for future biotechnological development. We introduce CasID, a Cas enzyme classification scheme, to facilitate functional comparison between RNA-guided Class 2 Cas enzymes. The optimal selection of a CRISPR enzyme depends heavily on the intended application and CasPEDIA allows for efficient comparison between enzymes by both their biochemical properties and their previously established uses. As a flexible database, CasPEDIA can be updated to accommodate the emergence of novel CRISPR-Cas enzymes and their applications.

MAIN FEATURES of CasPEDIA

CasPEDIA introduces a systematic, enzymatic nomenclature for the functional classification of Class 2 Cas proteins, summarized in Table 1. This classification, termed CasID, is directly inspired by the ENZYME Classification (E.C.) system, but is tailored to the unique properties of these RNA-guided enzymes (7). Each enzyme in CasPEDIA receives a 3-decimal number reflecting its biochemical activities as RNA-guided enzymes. Briefly, CasPEDIA’s classification schema can be seen in Table 1 and is summarized here. Familiar to most CRISPR biotechnologists is “Nuclease Activity”, describing which nucleic acids are predominantly cut in *cis* (i.e. guide RNA-targeted) or in *trans* (i.e. non-guide RNA targeted). Additionally, we provide insight into “Targeting Context” or constraints on sequences that neighbor the targeted sequence (e.g. protospacer-

adjacent motif (PAM) or protospacer-flanking sequence (PFS)). Finally, we provide “gRNA Design and Multiplexability” to enable design of multiplexed guide RNAs (gRNAs) from a native CRISPR locus. For instance, as exhibited in Figure 1a, SpyCas9 can be summarized by CasID 1.1.1, meaning SpyCas9 is an RNA-guided enzyme with targeted double-stranded DNA (dsDNA) activity with blunt cleavage and no *trans*-activity, employs a 3' PAM positioning, and requires multiple synthetic gRNAs for multiplexable design. The enzymatic classification of Class 2 CRISPR proteins is intended to complement evolutionary classification efforts (3, 8). In tandem with phylogenetic classification, we hope that the consolidation of enzymatic and sequence information fosters the further development of CRISPR-based biotechnologies.

CasPEDIA is organized in wiki format, with dedicated web pages for an initial set of 31 nucleases. Shown in Figure 2, each wiki contains 7 sections: Quick Review, Summary, Applications, Experimental Considerations, Nucleotide Sequence, and Protein Structure. The Quick Review section, located at the top of the page, enables rapid access to essential information including: enzyme classification (a description of the CasID and phylogenetic classification), core properties (e.g. protospacer length, PAM/PFS, length of the nucleotide coding-sequence, etc.) and external resources (e.g. RefSeq identifiers for the gene and protein, UniProtKB ID, Conserved Domains Database IDs, etc.). Next, is a high-level Summary section, detailing the nuclease's origins, novel properties, and common uses. The Applications section then provides a literature review for the enzyme with sub-headers for Gene Editing examples in model organisms, Tools and Diagnostics utilizing the enzyme and Engineered Variants with expanded properties. This is followed by the Experimental Considerations section, a brief introduction to performing experiments with the Cas enzyme. It includes details on construct design, appropriate delivery modalities and a list of algorithms for gRNA design. Nucleotide Sequence is also discussed, complete with downloads and a genome browser, created with igv.js (9), demonstrating the nuclease's sequence and the architecture of its CRISPR array. The subsequent section covers Protein Structure, which includes a summary of the protein's domains from UniProt (10), Pfam annotations (11), and structures from the PDB (12), or predicted with AlphaFold2 (13), visualized using 3Dmol.js (14). Citations are provided for all wiki content and indexed at the bottom of the webpage. To assist users in locating relevant wiki entries, CasPEDIA includes extensive search features and navigational pages, discussed below.

NAVIGATING CasPEDIA

CasPEDIA provides multiple search tools to connect users with pertinent CRISPR-Cas enzymes and wiki content (Figure 2). Scientists can use the search bar, located on the homepage, to search for Cas nucleases by name (e.g. AsCas12a, SpyCas9a), RefSeq protein ID, or function using CasID nomenclature. Each search returns a table containing matching protein entries and displays for each entry: Enzyme Name, CasID, Protein Accession (RefSeq protein ID, when available), Nuclease Activity, Targeting Requirement, gRNA Design and Multiplexability, and PAM. Similarly, the search bar can also be used to query the database for a protein sequence using DELTA-BLAST (15) with default parameters. This approach allows for remote homology detection with the support of NCBI's Conserved Domain Database (CDD) (16) for domain-

enhanced sequence searches across the CasPEDIA database. The resulting table is sortable by all fields, including E-value, to assist users in finding a nuclease of interest.

A separate page, entitled Advanced Tool Search, directs users through a series of drop-down menus (fields include: Cis-Activity Substrate, Trans-Activity Substrate, Targeting Requirements and gRNA Design and Multiplexability), which generates a table of all Class 2 systems within CasPedia that demonstrate or conservatively predicted to demonstrate the selected properties.

CasPEDIA also supports phylogenetic navigation, complementing evolutionary classifications from previous studies (3, 8). The Phylogeny page of the website provides summaries of Type II, Type V and Type VI systems which make up Class 2. We provide dedicated pages for each system type, containing subtype descriptions and an interactive tree whose leaves redirect to wiki entries.

While CasPEDIA wiki entries are organized by protein type (i.e. nuclease name and corresponding species) and CasID, users may also locate information for examples of engineered variants and gene-editing tools. Term searches for engineered variants are unsupported at this time, but variant details can be identified by searching the parental enzyme by name, and scanning the "Engineered Variants" section of the parental wiki entry. Furthermore, a designated page for fusion proteins is available (i.e. Tool Glossary), organizing the expanding list of base editors and prime editors by function, as well as proteins used for CRISPR interference (CRISPRi), CRISPR activation (CRISPRa) and other tools.

CasPEDIA DATA CURATION

CasPEDIA is a community project, curated from the literature by a panel of CRISPR researchers. Wiki content was managed through a series of forms, which were distributed amongst curators and editors for completion. To ensure accuracy and objectivity, citations from peer-reviewed publications and databases were required. Citations are provided at the base of each page. Structural and sequence information were taken from literature or databases like PDB, NCBI, UniProt and Pfam. The CasPEDIA Consortium and Scientific Communications Team at the Innovative Genomics Institute reviewed all entries prior to initial release.

Additionally, we visualized CasPEDIA's enzymatic classification efforts against the current genetic classification of Class 2 CRISPR systems (3, 8). Phylogenetic trees were constructed for each Class 2 Type from comprehensive datasets for Cas9, Cas12, and Cas13 proteins (3, 17–20). Trees were constructed with IQ-TREE from MUSCLE aligned sequences, and visualized in iTOL (21–23).

DATA AVAILABILITY

CasPEDIA is freely accessible at <http://caspedia.org>, and data is licensed under Creative Commons Attribution 4.0 International License (CC BY 4.0). The website is compatible with all devices, including tablets and mobile phones. A complete inventory of enzymes in CasPEDIA,

along with CasID numbers, can be downloaded on the Tool Finder page. Text content for the wikis is available upon request, with more information provided on the Contact page of the website. Illustrations from CasPEDIA are available for non-commercial use under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0). Please credit "Innovative Genomics Institute, University of California, Berkeley."

FUTURE DEVELOPMENTS

Currently, CasPEDIA only contains entries for the enzymatic activities of Cas effectors in Class 2 CRISPR-Cas systems, as there is limited distinction between the enzymatic activity of the protein and the mature CRISPR-Cas complex. The current CasPEDIA entries include representatives from all 27 phylogenetic subtypes encoded within the Cas9, Cas12, and Cas13 families. We also provide entries including related proteins IscB (HEARO) and TnpB, important variants used in biotechnological applications, and enzymatic subtypes (ex. Cas12c1 vs. Cas12c2). Class 2 CRISPR system derived enzymes represent only a fraction of the overall Cas protein diversity (3). Class 1 CRISPR-Cas systems and CRISPR adaptation, comprise the most abundant CRISPR systems and enzymes across bacterial and archaeal genomes (3). Owing to their multi-protein nature, Class 1 CRISPR-Cas interference complexes coordinate multiple enzymatic activities in target nucleic acid recognition and their adoption for biotechnology has thus been difficult (2, 24–26). Adaptations of CasID for these enzyme complexes would facilitate greater adoption and subsequent innovation by the biotechnology community and is a clear priority for future iterations. Additionally, new Class 2 CRISPR-Cas systems are emerging at a rapid pace. During the preparation of the CasPEDIA database alone, seven new systems were reported (19, 27–33). We anticipate that many new systems will emerge by the next update of CasPEDIA.

CasPEDIA is an actively evolving database, which will grow through community engagement and sustained content management. CRISPR scientists are encouraged to contact the CasPEDIA Consortium to suggest new wiki entries and features, as well as update current wikis with emergent discoveries. These efforts will maintain the relevancy of the database as a useful resource for future scientists. Prospective volunteers can follow detailed directions on the Contact page of the website to contribute.

AUTHOR CONTRIBUTIONS

Benjamin A. Adler: Conceptualization, Formal analysis, Methodology, Visualization, Writing - original draft. Marena I. Trinidad: Conceptualization, Methodology, Software, Visualization, Writing - original draft. Daniel Bellieny-Rabelo: Methodology, Software, Writing - original draft. Elaine Zhang: Project administration, Writing - original draft. Hannah M. Karp: Visualization, Writing - original draft. Petr Skopintsev: Formal analysis, Visualization. LinXing Chen: Conceptualization, Formal analysis, Visualization. Tomas Hessler: Formal analysis, Visualization. Jillian F. Banfield: Conceptualization, Supervision, Writing - original draft. David F. Savage: Conceptualization, Supervision, Writing - original draft. Jennifer A. Doudna: Conceptualization, Supervision, Writing - original draft. All authors: Data curation, Formal analysis, Resources, Writing - review & editing.

FUNDING

This work was supported by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils (m-CAFEs@lbl.gov), a Science Focus Area led by Lawrence Berkeley National Laboratory based upon work supported by the US Department of Energy, Office of Science, Office of Biological & Environmental Research [DE-AC02-05CH11231 to B.A.A.]; the Swiss National Science Foundation Mobility Fellowship [P500PB_214418 to P.S.]; the National Science Foundation Graduate Research Fellowship [B.W.T., R.F.W., P.H.Y., and M.J.A.-S.]; the National Institutes of Health [U01AI142817 to L.C., AI171110 to A.R.E. and X.C.]; the CIRM Training Program [EDUC4-12790 to E.E.D.]; the National Institutes of Health Ruth L. Kirschstein National Research Service Award F31 Pre-Doctoral Fellowship [F31HL156468-01 to C.A.T.]; the Siebel Scholarship from the Siebel Foundation [C.A.T.]; the Summer Undergraduate Research Fellowship from the Rose Hills Foundation [R.V.T.]; the U/SELECT Program at Utrecht University [L.H.]; the Jane Coffin Childs Fund for Medical Research Fellowship at HHMI [H.S.]; the Pediatric Scientist Development Program Fellowship [M.J.K.]; the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development [K12-HD000850 to M.J.K.]; the Monash Graduate Excellence Scholarship [R.W.C.]; the Snow Medical Fellowship and a National Health and Medical Research Council Investigator [EL1, 1175568 to G.J.K.]; the National Institutes of Health [DP2GM149550 to A.L.]; the Edson Initiative for Dementia Care and Solutions [A.L.]; the European Regional Development Fund with the Central Project Management Agency, Lithuania [01.2.2-CPVA-V-716-01-0001 to P.P.]; the Research Council of Lithuania (LMTLT) [S-MIP-22-10 to P.P.]; the European Molecular Biology Organization [5342-2023 to P.P.]; the Helen Hay Whitney Foundation Fellowship [J.C.C.]; the National Institutes of Health [R35GM134867 to B.W.]; the Montana State University Agricultural Experimental Station (USDA NIFA) [B.W.]; the Bioprotection Aotearoa and the Marsden Fund, Royal Society of New Zealand (Te Pūtea Rangahau a Marsden, Te Apārangi) [P.C.F.]; the European Research Council [101003229 to S.J.J.B.]; the Netherlands Organisation for Scientific Research [VI.C.192.027 to S.J.J.B.]; the National Institute of General Medical Sciences [GM140876 to D.G.S.]; and the Howard Hughes Medical Institute [D.F.S. and J.A.D.]. Funding for open access charge: Howard Hughes Medical Institute.

CONFLICT OF INTEREST

The authors and their respective institutions have filed for patent protection for some of the technology discussed in this manuscript. C.A.T. and the Regents of the University of California have patents pending or issued related to the use of CRISPR genome editing technologies. The Regents of the University of California have patents pending for CRISPR technologies on which G.J.K. and P.P. are inventors. E.J.S. is a co-founder and scientific advisory board member of Intellia Therapeutics and a scientific advisory board member of Tessera Therapeutics. B.W. is the founder of SurGene LLC and VIRIS Detection Systems Inc. and is an inventor on patent applications related to CRISPR-Cas systems and applications thereof. P.C.F. is an inventor on patent applications related to CRISPR-Cas systems and applications thereof. B.C.T., C.T.B., and D.S.A.G. are employees of and receive salary from Metagenomi, Inc., and might own equity in Metagenomi Technologies, LLC. R.B. is a co-founder of Intellia Therapeutics, Locus Biosciences, Ancilia Biosciences and TreeCo and a shareholder of Caribou Biosciences, Inari Ag, Tune Therapeutics and CRISPR QC. V.S. is a chairman and co-founder of CasZyme. J.F.B. is a co-founder of Metagenomi, Inc. D.F.S. is a co-founder and scientific advisory board member of Scribe Therapeutics. The Regents of the University of California have patents issued and pending for CRISPR technologies on which J.A.D. is an inventor. J.A.D. is a co-founder of Caribou Biosciences, Editas Medicine, Scribe Therapeutics, Intellia Therapeutics, and Mammoth Biosciences. J.A.D. is a scientific advisory board member of Vertex, Caribou Biosciences, Intellia Therapeutics, Scribe Therapeutics, Mammoth Biosciences, Algen Biotechnologies, Felix Biosciences, The Column Group and Inari. J.A.D. is Chief Science Advisor to Sixth Street, a Director at Johnson & Johnson, Altos and Tempus, and has research projects sponsored by Apple Tree Partners and Roche. All other authors declare no competing interests.

TABLES AND FIGURE LEGENDS

Table 1. Cas Enzyme Classification. See <http://caspedia.org> for diagrams.

Dimension	Value	Description
Primary Nuclease Activity	1	Targets dsDNA + no <i>trans</i>-activity. Cleavage products are predominantly blunt. However, additional trimming of DNA cleavage products may occur on a timescale much slower than that of the initial cuts. RNA-guided RuvC domains are also capable of targeted, PAM-independent ssDNA cleavage.
	2	Targets dsDNA + no <i>trans</i>-activity. Staggered cleavage products. RNA-guided RuvC domains are also capable of targeted, PAM-independent ssDNA cleavage.
	3	Targets dsDNA (or ssDNA) + <i>trans</i>-ssDNase activity. Staggered cleavage products. RNA-guided RuvC domains are also capable of PAM-independent ssDNA cleavage.
	4	Targets dsDNA (primarily nicking) + <i>trans</i>-ssDNase activity. Cleavage products are predominantly nicked on a single strand on short timescales, but these enzymes retain the capacity to create double-strand breaks at a slow rate.
	5	Targets dsDNA (binding only)
	6	Targets RNA + <i>trans</i>-RNase activity
	7	Targets RNA + <i>trans</i>-RNase + <i>trans</i>-ssDNase activity
	8	Targets RNA + <i>trans</i>-RNase + <i>trans</i>-ssDNase + <i>trans</i>-dsDNase activity
	-	Unknown
Target Requirement	1	3' Protospacer-adjacent motif (PAM). This is a required sequence encoded in the non-targeted strand. 3' positioning also means the 3' CRISPR repeat is used.
	2	3' Protospacer-flanking sequence (PFS). This is a prohibited sequence encoded in the targeted strand also referred to as an anti-tag. 3' positioning also means the 3' CRISPR repeat is used.
	3	3' No constraints. 3' positioning means the 3' CRISPR repeat is used.
	4	5' Protospacer-adjacent motif (PAM). This is a required sequence encoded in the non-targeted strand. 5' positioning also means the 5' CRISPR repeat is used.
	5	5' Protospacer-flanking sequence (PFS). This is a prohibited sequence encoded in the targeted strand also referred to as an anti-tag. 5' positioning also means the 5' CRISPR repeat is used.
	6	5' No constraints. 5' positioning means the 5' CRISPR repeat is used.
	-	Unknown
	1	crRNA + tracrRNA + non-CRISPR-associated endogenous factors in the native host required for CRISPR processing

Guide RNA (gRNA) design + Multiplexing	2	crRNA + tracrRNA required for CRISPR processing
	3	crRNA required for CRISPR processing
	4	ωRNA
	-	Unknown

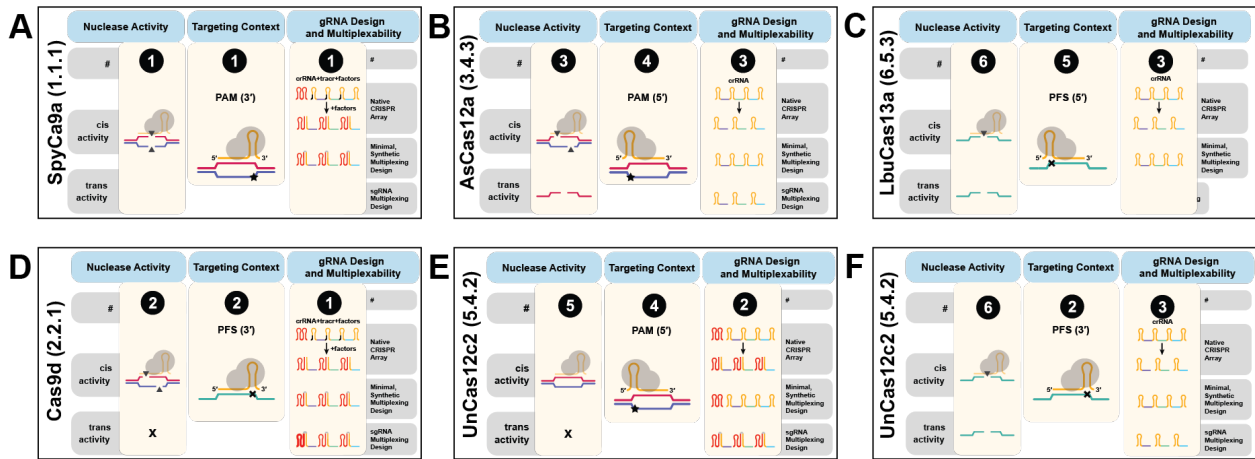


Figure 1. CasID enzymatic labels for biotechnologically important Cas enzymes. Examples are shown for (a) SpyCas9a (1.1.1), (b) Cas9d (2.2.1), (c) AsCas12a (3.4.3), (d) UnCas12c2 (5.4.2), (e) LbuCas13a (6.5.3), (f) PspCas13b (6.2.3). For a complete description and list of CasID values and their definition, please refer to <http://caspedia.org/>.

CasPEDIA Innovative Genomics Institute

SpyCas9a

A Cas ID

Nuclease Activity

1

cis activity

trans activity **X**

Targeting Context

1

PAM (3')

gRNA Design and Multiplexability

1

crRNA+tracr+ factors

+factors

Native CRISPR Array

Minimal, Synthetic Multiplexing Design

sgRNA Multiplexing Design

B RESOURCES

Species *Streptococcus pyogenes* M1 GAS strain SF370

Genome Assembly LS483338.1

Gene ID 901176

Protein ID WP_010922251.1

UniProtKB Q99ZW2

CDD COG3513

Nucleotide Sequence [FASTA](#)

Amino Acid Sequence [FASTA](#)

crRNA (Ex) [FASTA](#)

sgRNA Sequence (Ex) [FASTA](#)

tracrRNA Sequence (Ex) [FASTA](#)

Direct Repeat (Native) [FASTA](#)

C CLASSIFICATION

Cas ID 1.1.1

Nuclease Activity Target dsDNA + no trans-activity. Blunt cut.

Targeting Requirement 3' PAM

gRNA and Multiplexability crRNA + tracr + endogenous

Class II subtype II-A

D PROPERTIES

Protospacer Length 20

PAM NGG

Protein Weight (kDa) 158

RNP Weight (kDa) 190

CDS Length (nt) 4107

Number Amino Acids 1368

tracrRNA Length (nt) 75

crRNA Length (nt) 56

sgRNA Length (nt) 96

E DESCRIPTION

(Below) Summary, Applications, Experimental Considerations, Locus Browser, Protein Structure

...

F

G

H

I

J

?

K

Figure 2. Overview of CasPEDIA Entry for SpyCas9a (1.1.1) from the database. (a) CasID diagram and functional description. (b) Resources for accessing native sequences and gRNA design for the Cas enzyme. (c) Functional and phylogenetic classification of SpyCas9 (CasID 1.1.1). (d) Biological properties of this Cas enzyme, including protein, gene, and gRNA properties. (e) Overview of the Cas enzyme including a summary of the enzyme, applications, experimental considerations, protein structure, and gene browser (below the visualized portion). (f) Link to homepage containing CasID Definitions and search bar, accommodating queries for Cas enzymes by CasID, protein name, or protein family. (g) Icon for Tool Finder, where users can search CasPEDIA for enzymes with specific properties. (h) Redirects to Cas Phylogeny page for browsing the website by protein family. (i) Tool Glossary of common CRISPR-Cas systems. (j) Contact Page. (k) FAQ and general information.

REFERENCES

1. Barrangou,R., Fremaux,C., Deveau,H., Richards,M., Boyaval,P., Moineau,S., Romero,D.A. and Horvath,P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.
2. 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. and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
3. 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., *et al.* (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.*, **18**, 67–83.
4. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
5. Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, E2579–86.
6. Wang,J.Y. and Doudna,J.A. (2023) CRISPR technology: A decade of genome editing is only the beginning. *Science*, **379**, eadd8643.
7. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.
8. Koonin,E.V., Gootenberg,J.S. and Abudayyeh,O.O. (2023) Discovery of Diverse CRISPR-Cas Systems and Expansion of the Genome Engineering Toolbox. *Biochemistry*, 10.1021/acs.biochem.3c00159.
9. Robinson,J.T., Thorvaldsdottir,H., Turner,D. and Mesirov,J.P. (2023) igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*, **39**.
10. UniProt Consortium (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.*, **51**, D523–D531.
11. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L.L., Tosatto,S.C.E., Paladin,L., Raj,S., Richardson,L.J., *et al.* (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
12. wwPDB consortium (2019) Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.*, **47**, D520–D528.
13. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Židek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
14. Rego,N. and Koes,D. (2015) 3Dmol.js: molecular visualization with WebGL. *Bioinformatics*, **31**, 1322–1324.
15. Boratyn,G.M., Schäffer,A.A., Agarwala,R., Altschul,S.F., Lipman,D.J. and Madden,T.L. (2012) Domain enhanced lookup time accelerated BLAST. *Biol. Direct*, **7**, 12.
16. Lu,S., Wang,J., Chitsaz,F., Derbyshire,M.K., Geer,R.C., Gonzales,N.R., Gwadz,M., Hurwitz,D.I., Marchler,G.H., Song,J.S., *et al.* (2020) CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.*, **48**, D265–D268.
17. Gasiunas,G., Young,J.K., Karvelis,T., Kazlauskas,D., Urbaitis,T., Jasnauskaite,M.,

- Grusyte, M.M., Paulraj, S., Wang, P.-H., Hou, Z., *et al.* (2020) A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nat. Commun.*, **11**, 5512.
18. 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. and Doudna, J.A. (2020) CRISPR-Cas Φ from huge phages is a hypercompact genome editor. *Science*, **369**, 333–337.
 19. Al-Shayeb, B., Skopintsev, P., Soczek, K.M., Stahl, E.C., Li, Z., Groover, E., Smock, D., Eggers, A.R., Pausch, P., Cress, B.F., *et al.* (2022) Diverse virus-encoded CRISPR-Cas systems include streamlined genome editors. *Cell*, **185**, 4574–4586.e16.
 20. Adler, B.A., Hessler, T., Cress, B.F., Lahiri, A., Mutalik, V.K., Barrangou, R., Banfield, J. and Doudna, J.A. (2022) Broad-spectrum CRISPR-Cas13a enables efficient phage genome editing. *Nat Microbiol*, **7**, 1967–1979.
 21. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
 22. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*, **32**, 268–274.
 23. Letunic, I. and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.*, **49**, W293–W296.
 24. Goldberg, G.W., Jiang, W., Bikard, D. and Marraffini, L.A. (2014) Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature*, **514**, 633–637.
 25. Kazlauskienė, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. and Siksnys, V. (2017) A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, **357**, 605–609.
 26. Niewoehner, O., Garcia-Doval, C., Rostøl, J.T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L.A. and Jinek, M. (2017) Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature*, **548**, 543–548.
 27. Aliaga Goltsman, D.S., Alexander, L.M., Lin, J.-L., Fregoso Ocampo, R., Freeman, B., Lamothe, R.C., Perez Rivas, A., Temoche-Diaz, M.M., Chadha, S., Nordenfelt, N., *et al.* (2022) Compact Cas9d and HEARO enzymes for genome editing discovered from uncultivated microbes. *Nat. Commun.*, **13**, 7602.
 28. Urbaitis, T., Gasiunas, G., Young, J.K., Hou, Z., Paulraj, S., Godliauskaite, E., Juskeviciene, M.M., Stitilyte, M., Jasnauskaite, M., Mabuchi, M., *et al.* (2022) A new family of CRISPR-type V nucleases with C-rich PAM recognition. *EMBO Rep.*, **23**, e55481.
 29. Sun, A., Li, C.-P., Chen, Z., Zhang, S., Li, D.-Y., Yang, Y., Li, L.-Q., Zhao, Y., Wang, K., Li, Z., *et al.* (2023) The compact Cas π (Cas12I) ‘bracelet’ provides a unique structural platform for DNA manipulation. *Cell Res.*, **33**, 229–244.
 30. 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., *et al.* (2022) The miniature CRISPR-Cas12m effector binds DNA to block transcription. *Mol. Cell*, **82**, 4487–4502.e7.
 31. Chen, W., Ma, J., Wu, Z., Wang, Z., Zhang, H., Fu, W., Pan, D., Shi, J. and Ji, Q. (2023) Cas12n nucleases, early evolutionary intermediates of type V CRISPR, comprise a distinct family of miniature genome editors. *Mol. Cell*, 10.1016/j.molcel.2023.06.014.
 32. Bravo, J.P.K., Hallmark, T., Naegle, B., Beisel, C.L., Jackson, R.N. and Taylor, D.W. (2023) RNA targeting unleashes indiscriminate nuclease activity of CRISPR–Cas12a2. *Nature*,

613, 582–587.

33. Dmytrenko, O., Neumann, G.C., Hallmark, T., Keiser, D.J., Crowley, V.M., Vialletto, E., Mougiakos, I., Wandera, K.G., Domgaard, H., Weber, J., *et al.* (2023) Cas12a2 elicits abortive infection through RNA-triggered destruction of dsDNA. *Nature*, **613**, 588–594.