nature microbiology

Article

https://doi.org/10.1038/s41564-022-01265-y

Evolution of CRISPR-associated endonucleases as inferred from resurrected proteins

In the format provided by the authors and unedited

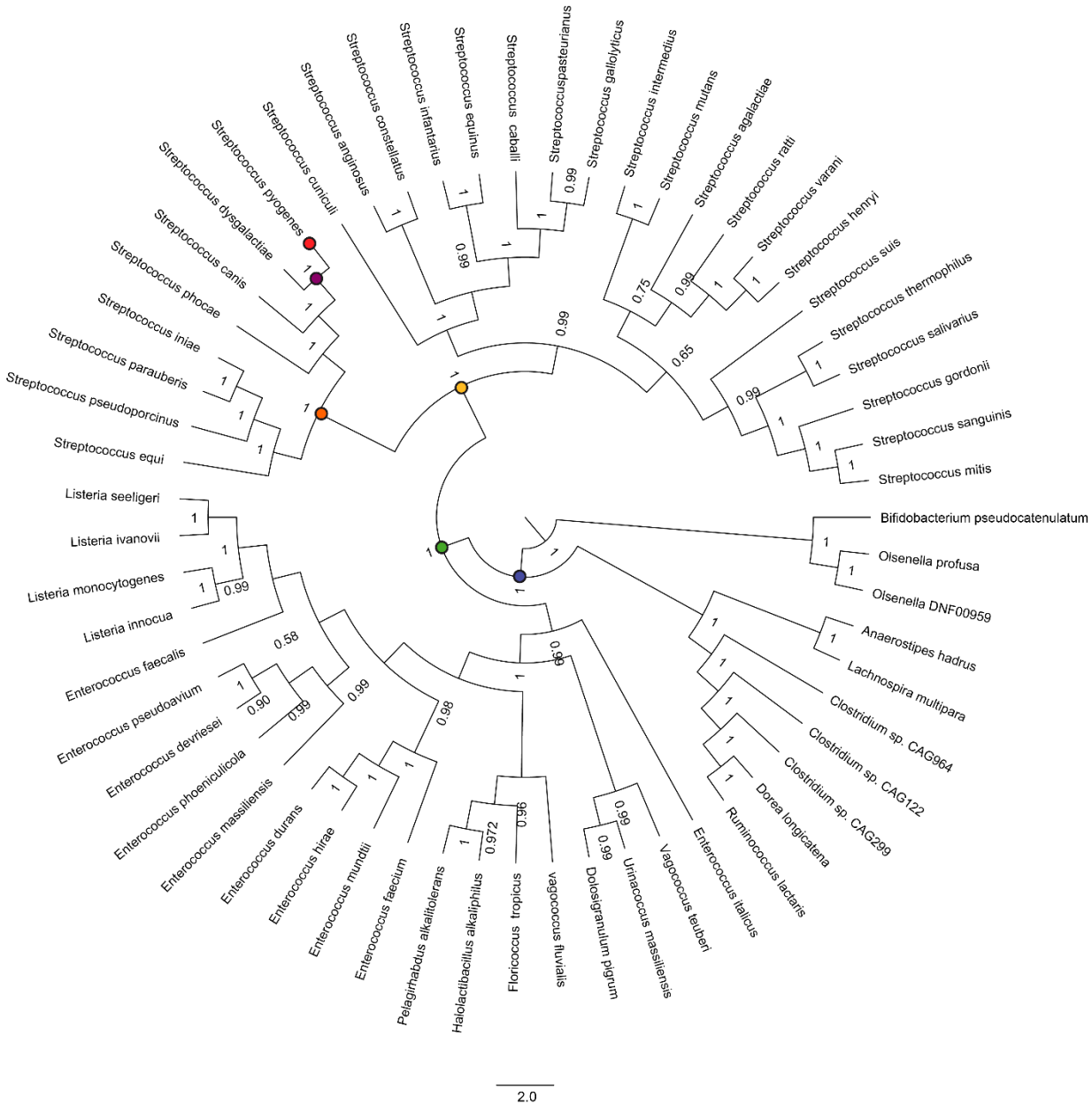
Supplemental Information for

Evolution of CRISPR-associated Endonucleases as Inferred from Resurrected Proteins

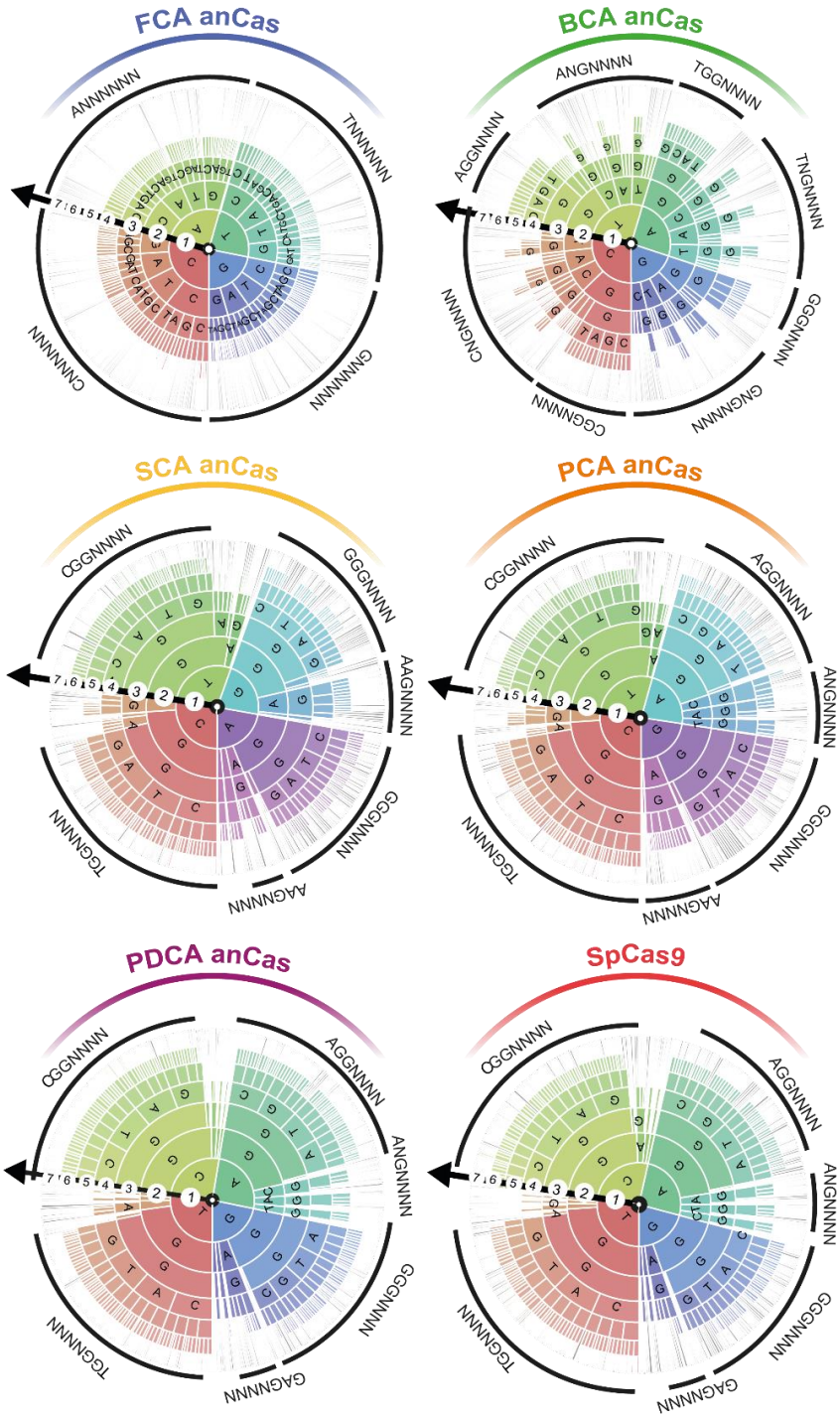
Table of Contents

Supplementary Notes:

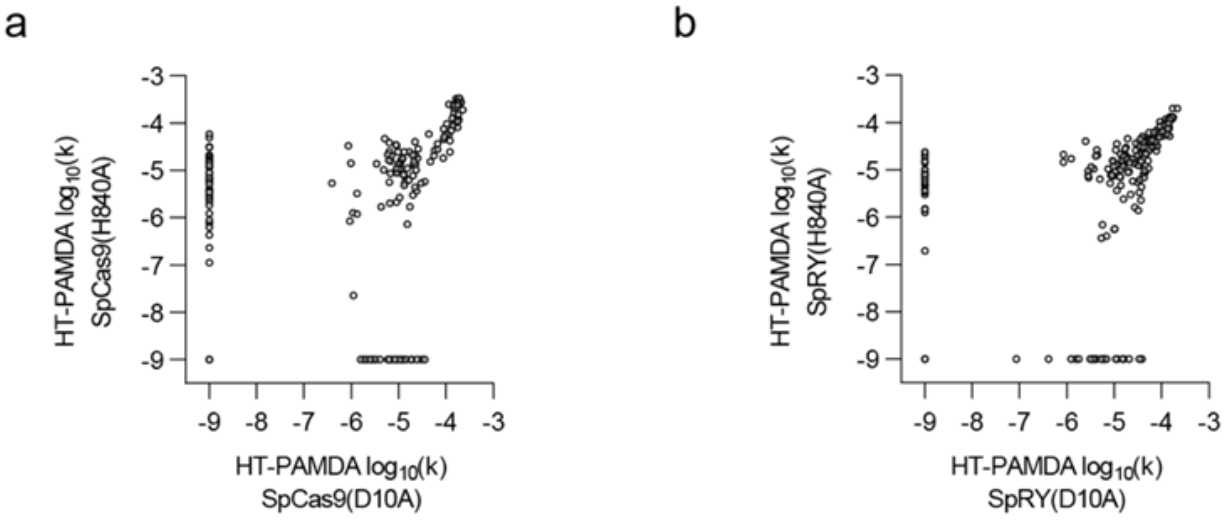
References



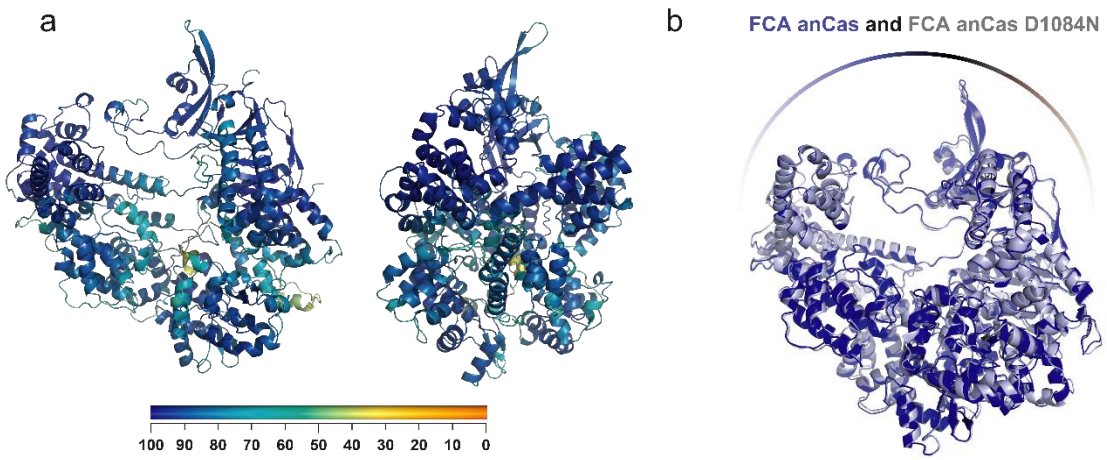
Supplementary Figure 1. Phylogeny tree of the fifty-nine species utilized in this study. Posterior probabilities for each node are indicated. Selected nodes for ancestral sequence reconstruction are marked with colored circles.



Supplementary Figure 2. PAM wheels (Krona plots) for all five anCas and SpCas9 including 7-nucleotides PAM analysis. These PAM wheels show seven concentric circles corresponding to the 7-nt PAM analysis with the arrow indicating the orientation. Preferred PAM are shown outside the wheel.

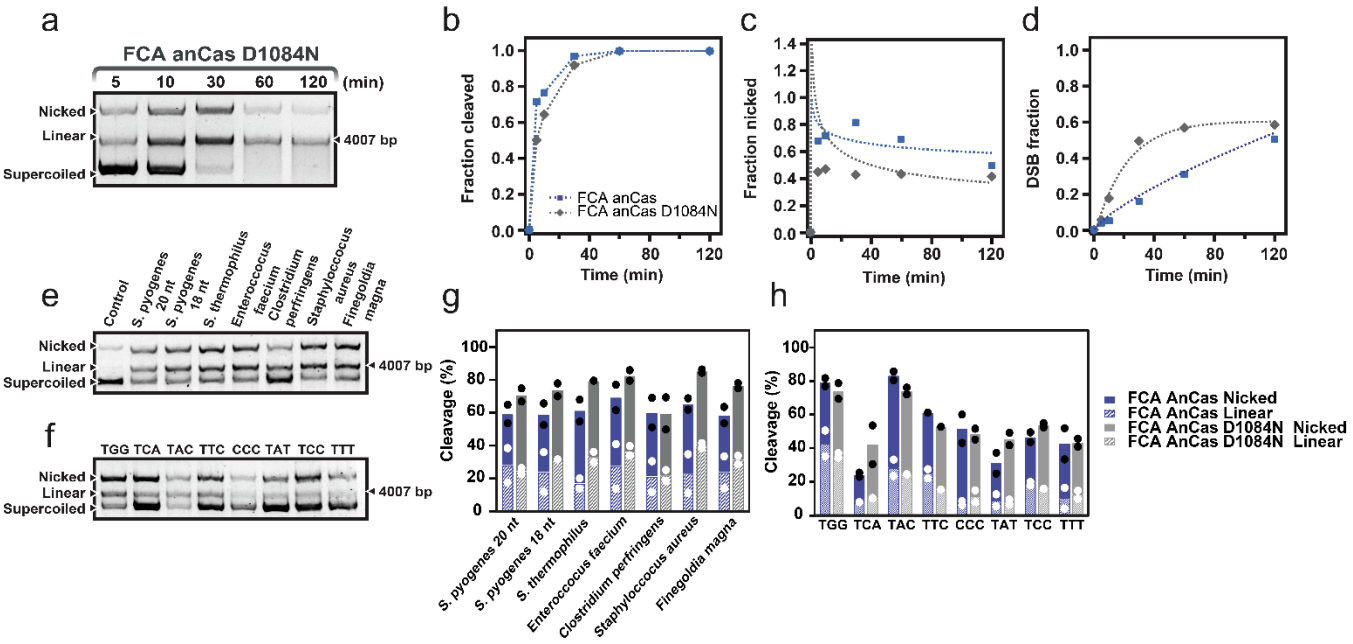


Supplementary Figure 3. HT-PAMDA rate constants from SpCas9 and SpRY nickases. **(a)** Correlation of HT-PAMDA $\log_{10}$ rate constants (*k*s) for nSpCas9(H840A) and **n**SpCas9(D10A) for each of the possible 256 NNNN PAMs, where the *k*s are the average of two HT-PAMDA experiments performed using libraries with distinct spacer sequences; $r(256) = 0.4466$, $p < 0.0001$ **(b)** Correlation of HT-PAMDA $\log_{10}$ rate constants (*k*s) for nSpRY(H840A) and nSpRY(D10A); $r(256) = 0.6541$, $p < 0.0001$. Reported correlations are Pearson correlation coefficients with statistical significance assessed with a two-tailed t-test.



RMSD values of different SpCas9 protein domains from FCA anCas D1084N		
Domains	**Over SpCas9**	**Over FCA anCas**
RuvCI	0.5520	0.1496
BH	0.5649	0.0686
REC1	0.8473	0.2796
REC2	1.6658	0.5796
REC	2.9402	0.4743
RUVC II	0.7830	0.5990
HNH	1.0741	0.6446
RUVC III	1.9090	0.9322
PI	1.7130	0.6321

Supplementary Figure 4. Characterization of FCA anCas D1084N mutant structure. **(a)** Structural prediction (front and side) of FCA anCas D1084N mutant by AlphaFold2. Structures are colored by pLDDT score according to the color bar. The estimated per-residue confidence score, pLDDT, for FCA anCas D1084N is 83.22. **(b)** Superposition of structural prediction of FCA (blue) and FCA D1084 (grey) using AlphaFold2. The structural alignment of FCA anCas and the mutant and the comparative structural analysis of the different domains alone by calculating RMSD values reveal no significant structural differences.



Kinetics parameters of the FCA anCas and FCA anCas D1084N determined from the single exponential decay curves.				
	FCA anCas		**FCA anCas D1084N**	
Kinetics parameters	Total	DSB	Total	DSB
K_{cleave} (min^{-1})	0.22 ± 0.04	0.007 ± 0.001	0.23 ± 0.03	0.03 ± 0.01
Max fraction	1	0.51	1	0.58
R^2	0.9790	0.9985	0.9909	0.9819

Supplementary Figure 5. Characterization of FCA anCas D1084N mutant activity. (a) *In vitro* cleavage assay for FCA anCas D1084N on a 4007 bp substrate at different reaction times showing nicked and linear fractions. **(b)** Quantification of total cleavage fraction at different reaction times and exponential fits (lines). Values reported as mean, where n = 2. **(c)** Quantification of fraction nicked at different times. Values reported as mean, where n = 2. **(d)** Quantification of DSB cleavage. Single-exponential fits were used to obtain k_{cleave} and maximum fraction cleaved (amplitude). Values reported as mean, where n = 2. As FCA anCas form, the mutant is able to produce nicked and linear products, showing a profile very similar to that exposed by FCA anCas. **(e)** *In vitro* cleavage assay on a supercoiled DNA substrate of FCA anCas D1084N using sgRNAs from different species. **(f)** *In vitro* cleavage assay using different PAM sequences. **(g)** Quantification of *in vitro* cleavage using different sgRNAs of FCA anCas and FCA anCas D1084N. **(h)** Quantification of in vitro cleavage using different PAM sequences of FCA anCas and FCA anCas D1084N. Bars represent the average value of two independent experiments indicated by the black dots. The mutant is able to nick and linearize the substrate DNA with all sgRNAs, proving its promiscuity for sgRNA. The results from PAM determination in vitro assay show the presence of both nicked and linear products. The percentage of cleavage of the mutant is similar to that exposed by FCA anCas, being mostly nicked product, as expected given the short incubation time.

Supplementary Table 1. List of Cas9 sequences utilized for Ancestral Reconstruction of ancient endonucleases.

Phylum	Class	Genus	Specie	Sequence
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus dysgalactiae*	WP_084916602.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus canis*	WP_003043819.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus gallolyticus*	WP_012962174.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus infantarius*	WP_014334983.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus phocae*	WP_054279288.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus equinus*	WP_157339387.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus pasteurianus*	WP_061100419
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus equi*	WP_037581760.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus mutans*	WP_002279859.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus iniae*	WP_003099269.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus agalactiae*	AFV72233.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus caballi*	WP_018363470.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus anginosus*	WP_003041502
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus varani*	WP_093650272
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus intermedius*	WP_082312238.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus ratti*	WP_003088697. 1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus sanguinis*	WP_002906454.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus mitis*	WP_084927115.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus gordonii*	WP_045635197.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus suis*	WP_044681799.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus salivarius*	WP_002891502.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus pseudoporcinus*	WP_007896501.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus henryi*	WP_074484960.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus thermophilus*	WP_065972475.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus parauberis*	WP_076751909.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus constellatus*	WP_006269658.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus pyogenes*	WP_032464890.1
Firmicutes	Bacilli	*Streptoctoccus*	*Streptococcus cuniculi*	WP_075103982.1
Firmicutes	Bacilli	*Enterococcus*	*Enterococcus italicus*	WP_007209003.1
Firmicutes	Bacilli	*Enterococcus*	*Enterococcus massiliensis*	WP_048604708.1
Firmicutes	Bacilli	*Enterococcus*	*Enterococcus phoeniculicola*	OJG69383.1
Firmicutes	Bacilli	*Enterococcus*	*Enterococcus faecalis*	QKR88578.1
Firmicutes	Bacilli	*Enterococcus*	*Enterococcus devriesei*	WP_071862060.1
Firmicutes	Bacilli	*Enterococcus*	*Enterococcus hirae*	WP_096708680.1
Firmicutes	Bacilli	*Enterococcus*	*Enterococcus faecium*	WP_087063969.1
Firmicutes	Bacilli	*Enterococcus*	*Enterococcus pseudoavium*	WP_067627992.1
Firmicutes	Bacilli	*Enterococcus*	*Enterococcus durans*	WP_081134165.1
Firmicutes	Bacilli	*Enterococcus*	*Enterococcus mundtii*	WP 023519017.1
Firmicutes	Bacilli	*Vagococcus*	*Vagococcus fluvialis*	WP_207053108.1
Firmicutes	Bacilli	*Vagococcus*	*Vagococcus teuberi*	WP 071456514.1
Firmicutes	Bacilli	*Listeria*	*Listeria monocytogenes*	WP_003739838.1
Firmicutes	Bacilli	*Listeria*	*Listeria ivanovii*	WP_038409211.1
Firmicutes	Bacilli	*Listeria*	*Listeria innocua*	WP_010991369.1
Firmicutes	Bacilli	*Listeria*	*Listeria seeligeri*	WP_075702521.1
Firmicutes	Bacilli	*Halolactibacillus*	*Halolactibacillus alkaliphilus*	WP_089800158.1
Firmicutes	Bacilli	*Pelagirhabdus*	*Pelagirhabdus alkalitolerans*	WP_090793453.1
Firmicutes	Bacilli	*Dolosigranulum*	*Dolosigranulum pigrum*	WP_004636532.1
Firmicutes	Bacilli	*Anaerostipes*	*Anaerostipes hadrus*	WP_173774801.1
Firmicutes	Bacilli	*Floricoccus*	*Floricoccus tropicus*	WP_070791099.1
Firmicutes	Bacilli	*Urinacoccus*	*Urinacoccus massiliensis*	WP_034440723
Firmicutes	Clostridia	*Clostridium*	*Clostridium sp CAG299*	CDD37961.1
Firmicutes	Clostridia	*Clostridium*	*Clostridium sp CAG964*	CDC80610.1
Firmicutes	Clostridia	*Clostridium*	*Clostridium_sp_CAG122*	CCZ42109.1
Firmicutes	Clostridia	*Lachnospira*	*Lachnospira multipara*	WP_027438114.1
Firmicutes	Clostridia	*Ruminococcus*	*Ruminococcus lactaris*	WP_005609677.1
Firmicutes	Clostridia	*Dorea*	*Dorea longicatena*	WP_055214841.1
Actinobacteria	Coriobacteriaceae	*Olsonella*	*Olsenella_DNF00959*	WP_062531800.1
Actinobacteria	Coriobacteriaceae	*Olsonella*	*Olsenella profusa*	WP_021725096.1
Actinobacteria	Bifidobacteriaceae	*Bifidobacterium*	*Bifidobacterium pseudocatenulatum*	WP_065439263.1

Supplementary Table 2. Statistics analysis of pLDDT values obtained from AlphaFold2 structure prediction of anCas and SpCas9.

	Mean	SD	Min	Max	Num Values
FCA anCas	82.24	12.60	35.81	98.33	1340
BCA anCas	85.87	11.28	33.75	98.32	1340
SCA anCas	85.97	11.72	33.15	98.31	1368
PCA anCas	86.88	10.75	31.90	98.42	1368
PDCA anCas	87.87	10.27	36.52	98.58	1368
SpCas9	88.33	9.92	38.10	98.61	1368

Supplementary Table 3. RMSD values of different SpCas9 protein domains obtained from anCas.

Domains	FCA anCas	BCA anCas	SCA anCas	PCA anCas	PDCA anCas
RuvCI	0.5820	0.3722	0.3414	0.3415	0.3617
BH	0.5877	0.3153	0.3453	0.3240	0.3272
REC1	0.8105	0.5504	0.6405	0.5821	0.4768
REC2	1.6140	0.8739	0.3722	0.5694	0.5194
REC	2.9309	1.1052	1.0937	0.9087	1.1426
RUVC II	0.8643	0.7294	0.6052	0.7951	0.7577
HNH	1.1783	0.8971	0.8710	0.9024	0.7063
RUVC III	2.1138	1.0338	1.0573	1.1097	0.9155
PI	1.9637	1.1284	1.0394	0.9443	0.7475

Supplementary Table 4. Posterior probabilities of functionally important residues for FCA anCas nuclease. First and second most probable residues are shown. In red, the selected residue with a significant second most probable residue.

Position	First Most Probable Residue	Posterior Probability	Second Most Probable Residue	Posterior Probability
Conserved residues				
10	D	0.987	E	0.012
15	S	0.997	A/N/T*	0.001
66	R	0.994	K	0.006
70	R	0.994	K	0.006
74	R	0.994	K	0.006
78	R	0.994	K	0.006
475	P	1	-	-
476	W	1	-	-
477	N	0.996	D	0.002
838	H	1	-	-
1305	R	0.945	K	0.032
1307	R	0.729	K	0.22
Mutated residues				
1084	D	0.482	N	0.371
1086	T	0.904	S	0.066
1103	L	0.946	M	0.031
1306	M	0.984	L	0.005

**These residues show the same posterior probability values.*

Supplementary Table 5. PAM library cloned in pUC18 (Figure 3 and Supplementary Fig. 2). Target sequence is indicated in blue and random PAM in red.

PAM library sequence
tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgc cgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagag cagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccatt cgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggat gtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttgc atgcctgcaggtcgactctagagggatccagcaacaacggtcggccacaccttccattgtcgtggccacgctcggattacac ggcagaggtgcttgtgttccgacaggctagcatattgtcctaaggcgttaccccaaNNNNNNNggtaccgagctcga attcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaa gtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacct gtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctca ctgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatc aggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggc gtttttccataggctccgccccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggact ataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgc ctttctcccttcgggaagcgtggcgctttctcaaagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctg ggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagac acgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaa gtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagt tggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaa aggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatg agattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggt ctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcg tgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctc cagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtc tattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtgg tgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaa aaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactg cataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgc ggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattgga aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactga tcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggc gacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatat ttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccatta ttatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc

Supplementary Table 6. Primers for PAM determination assay (NGS analysis).

Name	Sequence
FW library	aataggcgtatcacgaggc
RV library	agcgagtcagtgagcgag

Supplementary Table 7. *In vitro* cleavage PAM sequences (Figure 3). Target sequence indicated in blue and PAM sequence in red.

PAM	Sequence
TAC	tgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaa ctgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaa ggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttg catgcctgcaggtcgactctagagggatccagcaacaacggtcggccacaccttccattgtcgtggccacgc tcggattacacggcagaggtgcttgtgttccgacaggctagcatattgtcctaaggcgttaccccaa**tac**gag gggtaccgagctcgaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccaca caacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtga
TGG	tgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaa ctgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaa ggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttg catgcctgcaggtcgactctagagggatccagcaacaacggtcggccacaccttccattgtcgtggccacgc tcggattacacggcagaggtgcttgtgttccgacaggctagcatattgtcctaaggcgttaccccaa**tgg**gag gggtaccgagctcgaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccaca caacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtga
TAT	tgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaa ctgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaa ggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttg catgcctgcaggtcgactctagagggatccagcaacaacggtcggccacaccttccattgtcgtggccacgc tcggattacacggcagaggtgcttgtgttccgacaggctagcatattgtcctaaggcgttaccccaa**tat**gag gggtaccgagctcgaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccaca caacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtga
TCC	tgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaa ctgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaa ggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttg catgcctgcaggtcgactctagagggatccagcaacaacggtcggccacaccttccattgtcgtggccacgc tcggattacacggcagaggtgcttgtgttccgacaggctagcatattgtcctaaggcgttaccccaa**tcc**gag gggtaccgagctcgaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccaca caacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtga
CCC	tgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaa ctgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaa ggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttg catgcctgcaggtcgactctagagggatccagcaacaacggtcggccacaccttccattgtcgtggccacgc tcggattacacggcagaggtgcttgtgttccgacaggctagcatattgtcctaaggcgttaccccaa**ccc**gag gggtaccgagctcgaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccaca caacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtga

PAM	Sequence
TTT	tgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaa ctgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaa ggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttg catgcctgcaggtcgactctagagggatccagcaacaacggtcggccacaccttccattgtcgtggccacgc tcggattacacggcagaggtgcttgtgttccgacaggctagcatattgtcctaaggcgttaccccaa**ttt**gagg ggtaccgagctcgaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacac aacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtga
TTC	tgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaa ctgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaa ggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttg catgcctgcaggtcgactctagagggatccagcaacaacggtcggccacaccttccattgtcgtggccacgc tcggattacacggcagaggtgcttgtgttccgacaggctagcatattgtcctaaggcgttaccccaa**ttc**gag gggtaccgagctcgaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccaca caacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtga
TCA	tgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaa ctgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaa ggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttg catgcctgcaggtcgactctagagggatccagcaacaacggtcggccacaccttccattgtcgtggccacgc tcggattacacggcagaggtgcttgtgttccgacaggctagcatattgtcctaaggcgttaccccaa**tca**gag gggtaccgagctcgaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccaca caacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtga

Supplementary Table 8. Plasmids used to generate Cas proteins for HT-PAMDA (Figure 4 and Extended Data Fig. 7).

Plasmid ID	Addgene ID	Plasmid description
LTH1378	185484	pCMV-T7-NLS(SV40)-anCas(SCA)-BPNLS(SV40)-P2A-EGFP
LTH1430	185485	pCMV-T7-NLS(SV40)-anCas(PDCA)-BPNLS(SV40)-P2A-EGFP
LTH1380	185487	pCMV-T7-NLS(SV40)-anCas(FCA)-BPNLS(SV40)-P2A-EGFP
LTH1384	185488	pCMV-T7-NLS(SV40)-anCas(BCA)-BPNLS(SV40)-P2A-EGFP
LTH1441	185486	pCMV-T7-NLS(SV40)-anCas(PCA)-BPNLS(SV40)-P2A-EGFP
RTW3027	139987	pCMV-T7-SpCas9-BPNLS(SV40)-3xFLAG-P2A-EGFP
RTW4830	139989	pCMV-T7-SpRY-BPNLS(SV40)-3xFLAG-P2A-EGFP
HES24	185492	pCMV-T7-nSpCas9(D10A)-BPNLS(SV40)-3xFLAG-P2A-EGFP
HES636	185491	pCMV-T7-nSpRY(D10A)-BPNLS(SV40)-3xFLAG-P2A-EGFP
HES140	185490	pCMV-T7-nSpCas9(H840A)-BPNLS(SV40)-3xFLAG-P2A-EGFP
HES395	185489	pCMV-T7-nSpRY(H840A)-BPNLS(SV40)-3xFLAG-P2A-EGFP

Supplementary Table 9. Plasmids for sgRNA generation via *in vitro* transcription.

Plasmid ID	Description	Addgene ID	Spacer Sequence
RTW443	pT7-SpCas9_sgRNA-site1	160136	GGGCACGGGCAGCTTGCCGG
RTW448	pT7-SpCas9_sgRNA-site2	160137	GTCGCCCTCGAACTTCACCT

Supplementary Table 10. PAM Library Plasmids (HT-PAMDA substrates).

Plasmid ID	Addgene ID	Description
RTW554	160132	p11-3'_8xN_PAM-site1
RTW555	160133	p11-3'_8xN_PAM-site2

Supplementary Table 11. Primers for HT-PAMDA sample preparation.

Oligo ID	Description	Sequence
oRAS334	P5 sample barcode primer with GCAT barcode	ACACTCTTTCCCTACACGACGCTCTTCCGA TCTCCGCATGCATGCCTCGTGACCTGC
oRAS335	P5 sample barcode primer with GCCG barcode	ACACTCTTTCCCTACACGACGCTCTTCCGA TCTCCGCCGGCATGCCTCGTGACCTGC
oRAS336	P5 sample barcode primer with GCGC barcode	ACACTCTTTCCCTACACGACGCTCTTCCGA TCTCCGCGCGCATGCCTCGTGACCTGC
oRAS337	P5 sample barcode primer with GCTA barcode	ACACTCTTTCCCTACACGACGCTCTTCCGA TCTCCGCTAGCATGCCTCGTGACCTGC
oRAS338	P5 sample barcode primer with GGAA barcode	ACACTCTTTCCCTACACGACGCTCTTCCGA TCTCCGGAAGCATGCCTCGTGACCTGC

Oligo ID	Description	Sequence
oRAS340	P5 sample barcode primer with GGGG barcode	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGGGGGCATGCCTCGTGACCTGC
oRAS341	P5 sample barcode primer with GGTT barcode	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGGTTGCATGCCTCGTGACCTGC
oRAS342	P5 sample barcode primer with GTAC barcode	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGTACGCATGCCTCGTGACCTGC
oRAS343	P5 sample barcode primer with GTCA barcode	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGTCAGCATGCCTCGTGACCTGC
oRAS344	P5 sample barcode primer with GTGT barcode	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGTGTGCATGCCTCGTGACCTGC
oRAS345	P5 sample barcode primer with GTTG barcode	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGTTGGCATGCCTCGTGACCTGC
oRAS430	P7 sample barcode primer with GCAT barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGCATCGGTATTTCACACCGCATACGTAC
oRAS431	P7 sample barcode primer with GCCG barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGCCGCGGTATTTCACACCGCATACGTAC
oRAS432	P7 sample barcode primer with GCGC barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGCGCCGGTATTTCACACCGCATACGTAC
oRAS433	P7 sample barcode primer with GCTA barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGCTACGGTATTTCACACCGCATACGTAC
oRAS434	P7 sample barcode primer with GGAA barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGGAACGGTATTTCACACCGCATACGTAC
oRAS436	P7 sample barcode primer with GGGG barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGGGGCGGTATTTCACACCGCATACGTAC
oRAS437	P7 sample barcode primer with GGTT barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGGTTCGGTATTTCACACCGCATACGTAC
oRAS438	P7 sample barcode primer with GTAC barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGTACCGGTATTTCACACCGCATACGTAC
oRAS439	P7 sample barcode primer with GTCA barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGTCACGGTATTTCACACCGCATACGTAC
oRAS440	P7 sample barcode primer with GTGT barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGTGTCGGTATTTCACACCGCATACGTAC

Oligo ID	Description	Sequence
oRAS441	P7 sample barcode primer with GTTG barcode	CTGGAGTTCAGACGTGTGCTCTTCCGATCT TGGTTGCGGTATTTCACACCGCATACGTAC
oRW1117	P5 time point barcode primer Illumina 502 with ATAGAGAG barcode	AATGATACGGCGACCACCGAGATCTACAC CTCTCTATACACTCTTTCCCTACACGACGC TCTTCCGATCT
oRW1118	P5 time point barcode primer Illumina 503 with AGAGGATA barcode	AATGATACGGCGACCACCGAGATCTACAC TATCCTCTACACTCTTTCCCTACACGACGC TCTTCCGATCT
oRW1119	P5 time point barcode primer Illumina 504 with TCTACTCT barcode	AATGATACGGCGACCACCGAGATCTACAC AGAGTAGAACACTCTTTCCCTACACGACG CTCTTCCGATCT
oRW1105	P7 time point barcode primer Illumina 715 with ATCTCAGG barcode	CAAGCAGAAGACGGCATACGAGATCCTGA GATGTGACTGGAGTTCAGACGTGTGCTCTT CCGATCT
oRW1106	P7 time point barcode primer Illumina 716 with ACTCGCTA barcode	CAAGCAGAAGACGGCATACGAGATTAGCG AGTGTGACTGGAGTTCAGACGTGTGCTCTT CCGATCT
oRW1107	P7 time point barcode primer Illumina 718 with GGAGCTAC barcode	CAAGCAGAAGACGGCATACGAGATGTAGC TCCGTGACTGGAGTTCAGACGTGTGCTCTT CCGATCT

Supplementary Table 12. sgRNA template for sgRNA promiscuity assay (Figure 5). Target sequences shown in blue.

sgRNA	Sequence
sgRNA 20 nt *Streptococcus pyogenes* (WP_032464890.1)	taatacgactcactatagtcctaaggcgttaccccaagttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcttt
sgRNA 18 nt *Streptococcus pyogenes* (WP_032464890.1)	taatacgactcactatagctaaggcgttaccccaagttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttt
sgRNA *Enterococcus faecium* (WP_119364770.1)	taatacgactcactatagtcctaaggcgttaccccaagttttagagctatgctgagaaatcaatatagcaagttaaaataaggctttgtccgtcatcagctttttaaagcagcgctgttctcggcgctttttt
sgRNA *Streptococcus thermophilus* (LMG 18311) (WP_011225725.1)	taatacgactcactatagtcctaaggcgttaccccaagttttgtactctcagaaatgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgtttt
sgRNA *Clostridium perfringens* (WP_003473526.1)	taatacgactcactatagtcctaaggcgttaccccaagttatagttcctagtgaaaactagttactataacaaggcattaagccgtaaagtatcccctatgttcatttgaacctagggtatcttttcattt
sgRNA *Staphylococcus aureus* (AYD60528.1)	taatacgactcactatagtcctaaggcgttaccccaagttttagtactctggaaacagaatctactaaaacaaggcaaaatgccgtgtttatctcgtcaacttgttggcgagatt
sgRNA *Finegoldia magna* (WP_012290141.1)	taatacgactcactatagtcctaaggcgttaccccaagtttgagaatgatgtaatgaaaattacatcatgagttcaaataaaagtttactcaaatcgcccgaaagagcccacattggtggactaaacaaatcttcggatttgttttttt

Supplementary Table 13. Sequences of ssDNA and ssRNA (Figure 5). Target sequences shown in blue.

Sample	Sequence
ssDNA	ggatcctaatacgactcactataggctgtccgatcgtataa**caggattccgcaatggggtt**accgcttaagcattaggggagctc
ssRNA	gcuguccgaucguauaa**caggauuccgcaaugggguu**accgcuuaagcauuaggggagcuc

Supplementary Table 14. Plasmids used to generate Cas proteins for endogenous gene editing in human HEK 293T cells (Figure 6 and Extended Data Fig 9, 10).

Plasmid ID	Addgene ID	Plasmid description
pcDNA3.1-anCas(FCA)	185701	pCMV-T7-NLS(SV40)-anCas(FCA)-6xHis-NLS(SV40)
pcDNA3.1-anCas(BCA)	185702	pCMV-T7-NLS(SV40)-anCas(BCA)-6xHis-NLS(SV40)
pcDNA3.1-anCas(SCA)	185703	pCMV-T7-NLS(SV40)-anCas(SCA)-6xHis-NLS(SV40)
pcDNA3.1-anCas(PCA)	185704	pCMV-T7-NLS(SV40)-anCas(PCA)-6xHis-NLS(SV40)
pcDNA3.1-anCas(PDCA)	185705	pCMV-T7-NLS(SV40)-anCas(PDCA)-6xHis-NLS(SV40)
pcDNA3.1-(Spyo_Cas9)	185706	pCMV-T7-NLS(SV40)-(Spyo_Cas9)-6xHis-NLS(SV40)

Supplementary Table 15. Customized Primers for *in vivo* NGS analysis assay (Figure 6 and Extended Data Fig. 9). Gene-specific sequences are indicated in blue.

	sgRNA (5'-3') (PAM)	Primer sgRNA (5'-3')	Primer sequence (5'-3')
***TYR* gene**	gcatcccgccagtcccaata (TGG)	FW: acaccgcatcccgccagtcccaatag	FW: ttcgatttgagtgccccaga
		RV: aaaactattgggactggcgggatgcg	RV: ccttgatgggggctgcaat
***OCA2* gene**	aggagctgccactgccatcg (GGG)	FW: acaccaggagctgccactgccatcgg	FW: aactctcggagtgagctgtg
		RV: aaaaccgatggcagtggcagctcctg	RV: tctccagtgagagggaacagg

		Sequences
***TYR* gene**	TYREX1CRISP-F	5'-tcgtcggcagcgtcagatgtgtataagagacagtcaatggatgcactgcttgg-3'
	TYREX1CRISP-R	5'- gtctcgtgggctcggagatgtgtataagagacagtcaatggatgcactgcttgg-3'
***OCA2* gene**	OCA2EX14CRISP-F	5'-tcgtcggcagcgtcagatgtgtataagagacagcgcctcccttatacgagcaa-3'
	OCA2EX14CRISP-R	5'- gtctcgtgggctcggagatgtgtataagagacagttactgtgaagaggtggcgt-3'

		Sequences
***TYR* gene**	TYREX1CRISP-F	5'-tcgtcggcagcgtcagatgtgtataagagacagtcaatggatgcactgcttgg-3'
	TYREX1CRISP-R	5'- gtctcgtgggctcggagatgtgtataagagacagtcaatggatgcactgcttgg-3'
***OCA2* gene**	OCA2EX14CRISP-F	5'-tcgtcggcagcgtcagatgtgtataagagacagcgcctcccttatacgagcaa-3'
	OCA2EX14CRISP-R	5'- gtctcgtgggctcggagatgtgtataagagacagttactgtgaagaggtggcgt-3'

Supplementary Table 16. Traffic Light reporter assay (Extended Data Fig. 10).

	Sequence
gRNA 1 (PAM GGG)	tacgcaaataagagctcacc
gRNA 2 (PAM AGG)	aggtgagctcttatttgcgt
SpCas9 U6 driven expression cassette	ttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaa caccgaggtgagctcttatttgcgtgttttagagctagaaatagcaagttaaaataagg ctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttttt

Supplementary Note 1. Structural analysis of anCas nucleases.

Sequences of the different anCAs were used to predict their structure using AlphaFold2. The model with the highest confidence provided by AlphaFold2 was chosen for the structural analysis comparison (Extended Data Fig. 4). For all anCas, the estimated per-residue confidence score (Local Difference Distance Test), pLDDT [1], is on average over 80 (Fig. 1e, Supplementary Table 2). We performed a structural alignment of each anCas with the structure of SpCas9 bound to target DNA and gRNA (pdb:4oo8), (RMSD 2.52 Å) (Fig. 1b, d). Upon close examination of the structural alignment of FCA anCas and SpCas9, we observe a drastic structural difference in the position of the HNH domains, with a clear displacement of around 20 Å (red and blue domains in Fig. 1c). Considering that this domain is involved in DNA cleavage, an alteration in function would be expected. The rest of the anCas align well with SpCas9, with an increasing pLDDT as they approach present time (Fig. 1e, Supplementary Table 2). We also performed a comparative structural analysis of the different domains alone by calculating RMSD values (Supplementary Table 3). As expected, the highest RMSD values correspond to FCA anCas; however, direct domain comparison demonstrates considerable differences of over 2 Å in domains REC and RuvC III, further suggesting that structural constraints might alter the function of FCA anCas. When comparing the HNH domain alone between SpCas9 and FCA anCas, the RMSD is approximately 1 Å, demonstrating that the displacement observed in the complete structure (Fig. 1c) must be due to structural alterations to a global level in the FCA anCas structure. Finally, it is worth noting that RMSD values for PI domains seem to follow a decreasing trend, from the oldest to the newest anCas (Supplementary Table 3).

Supplementary Note 2. Effect of mutation H838A in FCA anCas activity.

The structural differences observed in FCA anCas protein related to HNH domain displacement, together with the evolutionary trend from nickase to DSB activity, suggests the oldest anCas, FCA, could display an ancestral HNH domain with a reduced or suppressed activity. To examine this, we tested the *in vitro* activity of the H838A FCA anCas mutant. This mutation is equivalent to H840A with respect to SpCas9, which is known to render SpCas9 as a nickase. The mutant H838A in FCA anCas was able to produce nicked and, surprisingly, linear products, showing a profile practically identical to that obtained with FCA anCas (Extended Data Fig. 5). These results suggest that FCA anCas may contain an immature HNH domain that does not participate in any cleavage, leaving the RuvC domain as the only responsible for both the nickase and DSB activity observed in FCA anCas. This behavior has been previously shown in some type V effector nucleases that lack the HNH domain, such as Cpf1 (Cas12a), Cas14 (Cas12f) or CasΦ (Cas12j) [2-5].

Supplementary Note 3. *In vitro* determination of PAMless activity.

To further probe the PAMless ability of FCA anCas, we designed an *in vitro* PAM determination assay to test cleavage of a target DNA adjacent to a total of seven selected PAM of 3-nt length (TAC, TCC, TAT, TTT, TTC, TAC and TGG) within the general TNN PAM. A CCC PAM was also included in the set to verify possibilities other than an initial T nucleotide. The anCas effectors together with the sgRNA were incubated with each of the target DNAs for 10 min and cleavage products were verified by agarose gel (Extended Data Fig. 6d). We observed both nicked and linear products, demonstrating the cleavage activity with all TNN PAM. In the case of SpCas9, only TGG PAM demonstrated double-stranded break of the supercoiled DNA substrate. We quantified the percentage of nicked and linear products for each PAM, and observed that for the oldest anCas, FCA and BCA, the percentage of cleavage was similar for all PAM tested, with mostly nicked products, as expected given the short incubation time (Fig. 3d). In the case of the more contemporary anCas (SCA, PCA and PDCA) and SpCas9, the cleavage fraction reached high levels for TGG PAM, corroborating the NGG PAM preference. In the case of CCC control, the cleavage profile was similar to those obtained from non-NGG PAM.

Supplementary Note 4. Testing the robustness of ancestral sequence reconstruction.

Experiments have been carried out with the ancestral proteins in which the most probable residue is chosen for each site. However, uncertainty in the reconstruction may arise, for instance, when the second most probable residue for a given site shows a significant probability. This is especially important for key functional residues. In the case of FCA anCas, for most key functional residues the most probable residue has a high probability compared to the second option (Supplementary Table 4). However, we noted one exception, which is residue 1048 for which the first and second option show probability values that are not far apart, i.e., 0.482 versus 0.371. This constitutes an ambiguity. It is accepted that ASR aims at determining functional and structural characteristics of ancestral proteins and genes, i.e. phenotypes, not the precise sequence of a protein or gene[6, 7]. With this consideration in mind, the robustness of the reconstruction must also be evaluated by solid functional and structural features. This is evaluated by testing second most probable residues[6].

To test the accuracy of the reconstruction of the ambiguous site of FCA anCas, we created a mutant form of the protein with the second most probable residue at the site 1084 (D1084N mutation). The protein has been obtained in the laboratory and a set of experiments have been performed to compare it with anCas FCA. These experiments are reported in Supplementary Figure 4 and 5, and compare structural prediction, cleavage function, and sgRNA and PAM recognition abilities. The characterization of the D1084N mutant validates the accuracy of the reconstruction showing phenotypes and characteristics practically identical to those exposed by the first most probable FCA anCas form.

References

1. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**, D439-D444 (2021).
2. Feng, W. et al. CRISPR technology incorporating amplification strategies: molecular assays for nucleic acids, proteins, and small molecules. *Chem Sci* **12**, 4683-4698 (2021).
3. Dai, Y. et al. Exploring the Trans-Cleavage Activity of CRISPR-Cas12a (cpf1) for the Development of a Universal Electrochemical Biosensor. *Angew Chem Int Ed Engl* **58**, 17399-17405 (2019).
4. Chen, J.S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436-439 (2018).
5. Pausch, P. et al. CRISPR-Cas(Phi) from huge phages is a hypercompact genome editor. *Science* **369**, 333-337 (2020).
6. Eick, G.N., Bridgham, J.T., Anderson, D.P., Harms, M.J. & Thornton, J.W. Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. *Molecular Biology and Evolution* **34**, 247-261 (2016).
7. Gaucher, E.A., Govindarajan, S. & Ganesh, O.K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704-707 (2008).