Evolution of CRISPR-associated Endonucleases as Inferred from Resurrected Proteins

Borja Alonso-Lerma[1,16], Ylenia Jabalera[1,16], Sara Samperio[1], Matias Morin[2], Almudena Fernandez[3], Logan T. Hille[4,5], Rachel A. Silverstein [4,5], Ane Quesada-Ganuza[1], Antonio Reifs[1], Sergio Fernández-Peñalver[2], Yolanda Benitez[3,6] Lucia Soletto[2], Jose A Gavira[7], Adrian Diaz[8,9], Wim Vranken[8,9,10], Avencia Sanchez-Mejias[11], Marc Güell[11,12], Francisco JM Mojica[13], Benjamin P. Kleinstiver[4,14], Miguel A Moreno-Pelayo[2], Lluis Montoliu[3] & Raul Perez-Jimenez[1,15*]

[1]CIC nanoGUNE BRTA; San Sebastian, Spain.

[2]Servicio de Genética, Hospital Universitario Ramón y Cajal, IRYCIS and Centro de Investigaciones Biomédicas en Red de Enfermedades Raras (CIBERER); Madrid, Spain.

[3]Department of Molecular and Cellular Biology, National Centre for Biotechnology (CNB-CSIC) and Centre for Biomedical Network Research on Rare Diseases (CIBERER-ISCIII); Madrid, Spain.

[4]Center for Genomic Medicine and Department of Pathology, Massachusetts General Hospital; Boston, MA, 02114, USA.

[5]PhD Program in Biological and Biomedical Sciences, Harvard University; Boston, MA, 02115, USA.

[6]INGEMM, Hospital Universitario La Paz, CIBERER-ISCIII; Madrid, Spain.

[7]Laboratorio de Estudios Cristalográficos, IACT (CSIC-UGR); Armilla, Granada, Spain.

[8]Interuniversity Institute of Bioinformatics in Brussels, ULB-VUB; Brussels 1050, Belgium.

[9]Structural Biology Brussels, Vrije Universiteit Brussel; Brussels 1050, Belgium.

[10]Structural Biology Research Centre, VIB; Brussels 1050, Belgium.

[11]Integra Therapeutics S.L.; Barcelona, Spain.

[12]Department of Medicine and Life Sciences, Universitat Pompeu Fabra; Barcelona, Spain.

[13]Dpto. Fisiología, Genética y Microbiología and Instituto Multidisciplinar para el Estudio del Medio "Ramón Margalef", Universidad de Alicante; Alicante, Spain.

[14]Department of Pathology, Harvard Medical School; Boston, MA, 02115, USA.

[15]Ikerbasque Foundation for Science; Bilbao, Spain.

[16]These authors contributed equally to the work

*Corresponding author. Email: r.perezjimenez@nanogune.eu (R.P.-J.)

Summary

Clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas9 protein is an effector that plays a major role in a prokaryotic adaptive immune system, by which invading DNA can be targeted and cut for inactivation. The Cas9 endonuclease is directed to target sites by a guide RNA (gRNA) where Cas9 can recognize specific sequences (PAMs) in foreign DNA, which then serve as an anchoring point for cleavage of the adjacent RNA-matching DNA region. Although the CRISPR-Cas9 system has been widely studied and repurposed for diverse applications (notably, genome editing), its origin and evolution remain to be elucidated. Here, we investigate the evolution of Cas9 from resurrected ancient nucleases (anCas) in extinct firmicutes species as old as 2.6 By to the current day. Surprisingly, we demonstrate that these ancient forms were much more flexible in their PAM and gRNA scaffold requirements compared to modern day Cas9 enzymes. In addition, anCas portrays a gradual paleoenzymatic adaptation from nickase to double-strand break activity, suggesting a mechanism by which ancient CRISPR systems could propagate when harboring Cas enzymes with minimal PAMs. The oldest anCas also exhibit high levels of activity with ssDNA and ssRNA targets, resembling Cas nucleases in related system types. Finally, we illustrate editing activity of the anCas enzymes in human cells. The prediction and characterization of anCas proteins uncovers an evolutionary trajectory leading to functionally flexible ancient enzymes.

Introduction

The CRISPR-Cas molecular complexes of prokaryotes are very diverse in function [1] and composition [2], with over thirty subtypes currently categorized in six types and two classes (types I, III and IV in class 1, types II, V and VI in class 2). Most of them serve as defense systems that prevent infection by viruses and other transmissible genetic elements. To defend against foreign nucleic acids, CRISPR loci harbour repeat-intervening DNA fragments (spacers) acquired during the so-called adaptation stage from invading nucleic acid fragments that have attacked the cell lineage in past generations [3-6]. In many CRISPR-Cas systems, only DNA fragments next to specific short sequences called protospacer adjacent motifs (PAMs) [7] are incorporated as new spacers. Over time, adaptation has created a vast collection of invading DNA spacers that are continuously evolving [8]. Transcripts from the CRISPR array are processed into small RNA molecules (crRNA), each containing a fragment of a single spacer and part of the repeats. The following steps of the CRISPR mechanism may show substantial variations depending on the system type [9]. In the type II CRISPR-Cas system, associated with the signature Cas9 effector protein (referred to as the CRISPR-Cas9 system), the crRNA molecules bound to an accessory trans-activating crRNA (tracrRNA), serve as cognate guides (gRNAs) for the Cas9 nuclease. After binding to a compatible PAM in the invading dsDNA, Cas9 uses its RuvC and HNH nuclease domains [6] to catalyse a double-strand break (DSB) within PAM-adjacent sequences that match the spacer carried by the crRNA.

The accuracy of the targeting depends on the binding of Cas9 to the gRNA, and the Cas9-gRNA complex subsequently to the PAM and remainder of the target site [10, 11]. This recognition ability is also the basis for the utilization of CRISPR-Cas9 as a gene editor, often linking the crRNA and tracrRNA into a single-guide RNA (sgRNA) [6, 12, 13]. The incorporation of invading DNA in the

CRISPR locus is an adaptive process that is time-dependent, i.e., over time bacteria will acquire new DNA that will be transferred to new generations. The ability of the organisms to process new DNA for self-protection must have arisen due to evolutionary pressure, which might be responsible for the recognition and cleavage properties of Cas nucleases. Interestingly, the absence of the PAM in the spacer-proximal sequence of the type II CRISPR repeats protects the CRISPR locus from targeting. However, the acquisition of new spacers matching the host genome has the potential to induce autoimmunity [14]. In this vein, diverse mechanisms have been identified that prevent or minimize the deleterious consequences of CRISPR-driven DSB in prokaryotes [15]. However, how nucleases have achieved such abilities, as well as the origin of the system, is a matter of debate [16].

In this work, we resurrect ancestral forms of Cas9 to demonstrate how evolution has likely sculpted Cas endonucleases. Other types of ancestral proteins have demonstrated unique properties depicting their evolutionary history, as well as environmental and organismal characteristics [17, 18]. Importantly, numerous ancestral enzymes have shown abilities useful for biotechnological applications [19, 20], making Ancestral Sequence Reconstruction (ASR) technique a powerful enzyme design tool with advantages over other protein engineering techniques [21]. Here, we have resurrected five ancestors of Cas9 (anCas), from a 2600 million years ago (mya) old common ancestor of a set of modern firmicutes (FCA) tracing an evolutionary path to modern *Streptococcus pyogenes* Cas9 (SpCas9). Unlike SpCas9, the oldest anCas can be guided by diverse sgRNA scaffolds and exhibits remarkable PAM flexibility, indicating that Cas9 evolved from total PAMless to 5´-NGG-3´ specificity. The activity of the anCas proteins depicts a paleotrend for nuclease/nickase activity ratio and PAM recognition that confers special abilities for gene targeting and cleavage. Furthermore, we demonstrate robust genome editing activity with these anCas in human cells. Overall, the anCas not only shed light onto the evolution and appearance of the CRISPR-Cas9 system, but also show unprecedented abilities for PAM and sgRNA recognition, nickase and nuclease activity, functional promiscuity, immune response and editing of human cells that makes them a valuable addition to the growing list of Cas nucleases with novel properties [22].

Results

Reconstruction of ancestral nucleases: anCas.

To perform a reconstruction of ancestral nucleases, we searched the NCBI Protein database using the sequence of subtype II-A CRISPR-associated endonuclease Cas9 (formerly Csn1) from *S. pyogenes* as query (NCBI Reference Sequence WP_032464890.1 and UniProt Q99ZW2) with full annotation score. The NCBI database offers a BLAST search that allows an initial sequence homology search using the Blosum62 scoring matrix. We retrieved a total of fifty-nine Cas9 sequences from firmicutes and actinobacteria species (Supplementary Table 1). We constructed a sequence alignment that was first used to generate a protein phylogenetic tree using Bayesian inference and Markov Chain Monte Carlo (MCMC) (Supplementary Fig. 1), that was calibrated to obtain a chronogram (Fig. 1a,). When comparing the gene tree to the species tree extracted from the TimeTree of Life (TToL[23]), we observed one notable discrepancy, i.e., the position of the listeria group. Listeria species, from the order of Bacillales, branched with Enterococcus species which belongs to the order of Lactobacillales. According to the Time Tree of Life, Bacillales and Lactobaccillaes shared a common ancestor around 1800 Ma[23, 24]. It is possible that

gene transfer from Enterococci to Listeria occurred but given the position in the tree it seems to have happened late in evolution.

Several internal nodes were selected for laboratory resurrection of anCas tracing the evolutionary path from the set's most recent Firmicutes Common Ancestor (FCA), which supposedly lived in around 2600 MYA, to modern *S. pyogenes*. This path also contained anCas from a Bacilli Common Ancestor (BCA) that lived around 1000 MYA and a Streptococcus Common Ancestor (SCA) anCas from around 200 MYA. Two other ancestors of SpCas9 were reconstructed, Pyogenic Common Ancestor (PCA) and Pyogenes-Dysgalactiae Common Ancestor (PDCA), that, according to our tree calibration, lived around 137 and 37 Ma, respectively (Fig. 1a). Using maximum likelihood, the most probable ancestral sequences for these nodes can be reconstructed and synthetically resurrected (Supplementary Data 1). The posterior probability average of each resurrected form ranged from 0.74 to 0.98 (Extended Data Fig. 1).

We compared the reconstructed sequences to that of SpCas9, yielding amino acid sequence identities in the range 53-96% (Extended Data Fig. 2). The lowest identity corresponds to FCA anCas, which contains over 620 amino acid differences with respect to SpCas9. In the case of BCA and SCA anCas, the accumulated number of changes is 410 and 342 mutations, respectively. PCA and PDCA anCas are more similar with only 226 and 54 mutations occurring, respectively. Upon aligning all sequences, we observe that most of the important residues reported to have a role in Cas9 function such as D10, S15, R66, 70, 74, 78, 1333 and 1335 and tandem residues such as PWN in 475-477 are all conserved across the anCas sequences when compared to SpCas9 [10, 25, 26]. However, some notable differences exist (Extended Data Fig. 3). Residue 1103, equivalent to W1126 in SpCas9, which is reportedly associated to PAM recognition [26], is mutated to L in FCA anCas. Similarly, amino acids in positions 1084 and 1306 in FCA anCas (corresponding to K1107 and K1334 in SpCas9) are mutated to D and M, respectively. Precisely, the tandem RKR in positions 1333-1335 located in the PAM-interacting (PI) domain has been shown to play a pivotal role in PAM recognition [10]. Finally, residue S1109, also associated to PAM interaction [10], is mutated to T in FCA and BCA anCas enzymes. These alterations suggest possible differences in PAM recognition abilities in FCA anCas and possibly in BCA anCas.

We used the anCas sequences to predict their structures using the newly developed AlphaFold2 structure prediction platform [27, 28], which utilizes a neural network to infer structures with often near-experimental accuracy. Five structural models were predicted for each sequence and the model with the highest confidence 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 [29], is on average over 80 (Fig. 1e, Supplementary Table 2). We performed a structural alignment of each selected anCas model with the structure of SpCas9 bound to target DNA and gRNA (pdb:4oo8), (RMSD 2.52 Å) (Fig. 1b, d). This is expected given that most Cas9 structures available in databases are bound to target DNA through gRNA interactions. 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. Surprisingly, 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).

Activity of anCas nucleases: DSB and nickase activity.

The anCas genes were synthesized and cloned into pBAD/gIII expression vector, carrying an arabinose inducible promoter and a *gIII* encoding signal that directs the anCas to the periplasmic space. All anCas were expressed at high levels in *Escherichia coli* BL21 cells.

We began to assess anCas activity test by assuming a simplistic scenario in which anCas would recognize a sgRNA from *S. pyogenes* as well as its canonical 5'-NGG-3' PAM (Fig. 2a). We designed a sgRNA containing a 20 nt-long spacer region targeted towards a DNA fragment upstream of a TGG PAM, all placed in a 4007 bp supercoiled plasmid. *In vitro* cleavage assays were carried out by incubating anCas or SpCas9 with target DNA together with the sgRNA at different digestion times. Although with clear differences in cleavage efficiency, all enzymes tested produce both relaxed and linear products, indicative of nickase and DSB activity, respectively. As expected, SpCas9 shows nicked product at short times and linear products at longer times (Fig. 2b). However, in the case of anCas the behavior changes from the oldest FCA anCas to more recent enzymes (Fig. 2b). FCA anCas mostly shows nickase activity and only at times over 60 min, the DSB activity becomes prominent. Younger anCas nucleases show a progressive increasing DSB activity towards modern days (Fig. 2b). We quantified both the nicked and linear fraction for each anCas and SpCas9 and plotted them versus incubation time in three forms, total cleavage (Fig. 2c), nicked fraction (Fig. 2d) and linear fraction (Fig. 2e), demonstrating the progressive decrease of nicked fraction and increase of linear fraction. SpCas9 and younger anCas have the highest proportion of linear products, the oldest FCA anCas has the highest proportion of nicked fraction (Table 1). We plotted the fraction of linear and nicked products at 30 minutes reaction versus geological time demonstrating an evolutionary trend from nickase to DSB activity (Fig. 2f).

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) [30-33].

Activity of anCas nucleases: PAM and sgRNA recognition.

Compared to the more stringent gRNA scaffold and PAM preferences displayed by more evolutionary contemporary Cas enzymes, we were intrigued that the anCas enzymes showed nuclease activity *in vitro* when using the SpCas9 gRNA scaffold and against sites with NGG PAMs. We therefore investigated the ability of the anCas endonucleases to recognize different PAMs. To determine the preferred PAM of each anCas, we cloned a DNA library containing a target protospacer followed by seven random nucleotides (NNNNNNN) that corresponded to all possible PAMs (Fig. 3a). DNA libraries with random PAMs have been widely used in similar studies to characterize other Cas9 orthologs or mutated Cas9 [34-36]. An sgRNA was designed using the scaffold of *S. pyogenes*. An 844 bp dsDNA fragment containing both the target and randomized PAM was used as a substrate for anCas and SpCas9 RNPs complexed with an SpCas9 sgRNA. We performed *in vitro* digestion using the purified Cas protein, and the transcribed sgRNA. All five anCas produced two cleavage products, the smaller of which contains the PAM (Extended Data Fig. 6a). The small cleavage product was purified, sequenced by Next-generation Sequencing (NGS), and analysed to determine PAM preference of each anCas, enabling us to infer how the PAM preference might have changed over the course of evolution. The five anCas enzymes and SpCas9 exhibited a range of PAM preferences displayed in the form of PAM wheels (Krona plot), heatmaps (displaying the proportion of total sequencing reads attributable to each PAM), and weblogos (Fig. 3b, Extended Data Fig. 6b and 6c). Surprisingly, FCA anCas exhibited little preference for any PAM tested. For the other Cas proteins, we detected a preference for specific nucleotides in the target-proximal positions 2 and 3 (Extended Data Fig. 6b and Supplementary Fig. 2). Thus, in the case of BCA anCas, a slight preference for NGG was revealed, although additional PAMs were also detected (NNG). For the other, more recent anCas enzymes, the NGG bias was more prominent.

Plotting the percentage of reads with an NGG PAM versus the geological time of each anCas revealed a trend that reflects a trajectory towards NGG, suggesting that PAM fidelity is an evolutionary trait with gradual progression from PAMless to NGG preference in more recent *Streptococci* ancestors (Fig. 3c). This supports the hypothesis that as the number of spacers acquired by adaptive CRISPR immune systems increases over time, a need for a PAM preference arises to mitigate self-targeting of the host chromosome. More stringent PAMs can prevent self-cleavage of the CRISPR locus, especially, in a scenario in which the increment of DSB activity (deleterious in most prokaryotes) over nickase activity might further increase evolutionary pressure. Although more PAM-permissive Cas9 variants have been previously engineered [37-39], to the best of our knowledge, FCA anCas is the first Cas9-related endonuclease that naturally exhibits a more minimal PAM requirement.

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.

The above experiments suggest that at short times the oldest anCas act primarily as nickases with relaxed PAM preference. To investigate this further, we performed a high-throughput PAM determination assay (HT-PAMDA) [40] on each enzyme and compared to SpCas9 (Fig. 4a and Extended Data Fig. 7). As controls, we also performed HT-PAMDA on a wild-type SpCas9 and a recently engineered nearly PAMless SpCas9 variant named SpRY [39] (Fig. 4a). Additionally, we also performed HT-PAMDA on the two nickase variants of SpCas9 and SpRY (containing RuvC (D10A) or HNH (H840A) domain inactivating mutations) (Fig. 4b)**.** In accordance with the hypothesis that the oldest enzymes may contain an immature HNH domain, the oldest anCas show profiles that have less prominent PAM requirements acting as nickases (BCA and FCA anCas) during the timecourse of the HT-PAMDA assay (Fig. 4a). Furthermore, their activities were more consistent with the nickase versions of the more contemporary SpCas9 and SpRY enzymes (Fig. 4b). For SpCas9 and SpRY, both the RuvC or HNH nickases exhibited similar PAM preferences (Supplementary Fig. 3). Consistent with our other PAM determination assays, the younger anCas enzymes generally progress to a more constrained NGG PAM preference (Fig. 4a and Extended Data Fig. 7).

The promiscuity for PAM recognition exhibited by the oldest anCas made us wonder whether the anCas would also show flexibility towards gRNA recognition as well. The reconstruction of an ancestral gRNA would have been interesting; however, the variability in sequence of crRNA repeats and tracrRNA from different species makes this challenging, and building a reliable phylogeny is not even possible. In fact, recent work has shown a classification of tracrRNA at the level of secondary structure, forming clusters, suggesting that structure of tracrRNA is an evolutionary trait[41, 42], but sequence phylogenies have not been reported.

To overcome this limitation and still evaluate the potential gRNA scaffold promiscuity of anCas, we decided to test modern sgRNAs from different species. We selected a total of five sgRNAs from *Streptococcus thermophilus, Enterococcus faecium, Clostridium perfringens, Staphylococcus aureus* and *Finegoldia magna*, covering several Firmicutes classes. These sgRNAs were selected following previous studies on sgRNA classification and function, in which sgRNAs were divided into seven clusters [42]. These distinct sgRNAs were contrasted against *S. pyogenes* guides containing spacers of two sizes, 18 and 20 nucleotides long, referred to as 18 and 20 nt sgRNA, respectively. SpCas9 and the five anCas were incubated for 10 min at 37 °C with a target plasmid DNA and TGG PAM target sites. As expected, we observed that SpCas9 only linearized plasmid DNA when using its own sgRNA, although more efficiently when using the 20 nt spacer version, and sgRNAs from other species mostly resulted in nicked products leaving most supercoiled DNA substrate intact (Fig. 5a). On the contrary, FCA and BCA anCas were able to nick and linearize plasmid DNA with all sgRNAs, the *E. faecium* sgRNA showing a slightly increase in efficiency for FCA anCas. We also tested the other anCas,

observing that mostly FCA and BCA anCas had a marked promiscuity for sgRNA. All other anCas and SpCas9 seem to work best with 20 nt sgRNA from *S. pyogenes* (Fig. 5b).

Previous studies have indicated the contribution of the REC domain to sgRNA recognition specificity [43, 44]. As shown in Supplementary Table 3, this domain exhibits the highest RMSD differences, with a decreasing RMSD trend from the oldest to the newest anCas. These findings, along with the sgRNA promiscuity observed in oldest anCas, suggest that evolutionary pressure might have guided Cas nucleases to an enhanced guide specificity over time. In fact, this promiscuity has been already observed in type II-C Cas9, which has been suggested to be an ancient reminiscence of Cas9 nucleases [43]. In these nucleases, this promiscuity is also associated to PAM-independent ssDNA cleavage and weaker substrate DNA unwinding capabilities.

Activity of anCas nucleases on single-stranded targets: ssDNA and ssRNA.

As mentioned, the oldest anCas (FCA and BCA), show a remarkable nickase activity. This nickase activity may be related to ssDNA activity. ssDNA cutting activity has been suggested to be an ancestral trait present in smaller Cas9 such as subtype II-C Cas9 [43] which could also be reflected in the nickase activity of the ancestral forms from subtype II-A, such as anCas. Earlier forms of Cas9 with smaller catalytic domains might have been the origin of this ssDNA cutting activity that was still present in larger ancestral nucleases, which then gradually evolved towards DSB activity as part of a specialization process. To test this hypothesis, we examined the oldest anCas with an 85 nt ssDNA substrate containing target sequence complementary to the 20 nt spacer region of a SpCas9-sgRNA. As shown in Figure 5c, e, FCA and, especially BCA anCas, show higher levels of ssDNA cleavage than those of SpCas9. Exponential fits to the data show much faster rates for FCA and BCA anCas, reaching almost full cleavage for BCA anCas (Table 2). We also tested activity on a 60 nt ssRNA target showing comparable results (Fig. 5d, f). Exponential fits of the cleavage activity reveal maximum rate and amplitude for BCA anCas, again reaching full cleavage (Table 2). These results demonstrate that both ancient Cas, but BCA anCas in particular, behave as RNA-guided RNAses.

The activity of the oldest anCas on single-stranded substrates suggests that early Cas nucleases might have been active on those substrates, which seems to be an ancient trait, as mentioned before. These abilities may have additional important implications, given that the remarkable activity of anCas on ssDNA and ssRNA resembles that of Cas12a, Cas14 and Cas13a [30], respectively. This resemblance suggests a connection among the activities of all class 2 effector nucleases. In addition, these Cas nucleases share the ability to expose nonspecific/collateral ssDNA (Cas12a and Cas14) or ssRNA (Cas13a) cleavage after their activation with a complex formed by crRNA complementary to ssDNA/dsDNA/ssRNA that serves as an activator for the trans-cleavage activity [32]. This particular trait is also found in the case of the oldest anCas (FCA anCas), which shows trans-cleavage activity degrading circular ssDNA (M13 phage) in the presence of sgRNA and its complementary ssDNA activator (Extended Data Fig. 8), suggesting a common origin with Cas12a and perhaps Cas14.

We also investigated whether the oldest anCas endonucleases, which display more promiscuous features, may also have a different response towards an Anti-Cas9 antibody. We incubated BCA and FCA anCas with an Anti-Cas9 rabbit antibody and performed an ELISA test, which shows a diminished antibody binding (Fig. 5g). This would be expected given that host organisms

carrying these nucleases have been long extinct and therefore have not been in contact with any living organisms. We believe that antibodies against Cas9 may have a weaker response towards ancient Cas forms. This lower antibody response might be of interest for potential applications in *in vivo* editing, where the immune response towards SpCas9 and other modern endonucleases represents a current limitation [45].

All previous 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[46, 47]. 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[46].

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.

In vivo activity of anCas variants: determination of indels.

We then tested the genome editing activity of these ancestral nucleases in mammalian cells in culture to determine if these synthetic ancestral Cas can perform DNA cleavage and trigger editing in cells under similar conditions as those associated with the standard SpCas9. To do so, the endogenous Tyrosinase gene (*TYR)* and the Melanocyte-specific transporter protein gene (*OCA2)*, whose mutations are associated to albinism [48], were targeted in human HEK293T cells. The cells were co-transfected with plasmids containing the humanized versions of anCas or SpCas9, as well as the corresponding sgRNAs. We used standard sgRNA from *S. pyogenes* carrying a 20-nt spacer target. Seventy-two hours after co-transfection, cells were collected, and the genomic DNA extracted. The occurrence of indels (insertions/deletions) was first verified by the T7 endonuclease mismatch assay (Fig. 6a), where we could detect the expected DNA fragments. The size of the observed DNA fragments match those generated by DSB at the intended nucleotide, as well as the result of the subsequent cellular Non-Homologous End Joining (NHEJ) repairing pathway. T7-derived DNA fragments were obvious for the *OCA2* gene for all endonucleases except FCA anCas, whereas in the case of *TYR* locus the expected DNA

bands were clearly visible for SpCas9 and PDCA anCas, and much weaker for SCA and BCA anCas (Fig. 6a). No bands were detected with the oldest anCas FCA in both loci.

We then evaluated and quantified indels by Next-generation Sequencing (NGS) using advanced analysis with Mosaic Finder software [49], which allows high resolution reading of short DNA sequences after NHEJ repair. As shown in Figure 6b, anCas endonucleases perform robust gene editing in human genomic DNA, except for FCA anCas. This is expected given the unique features of FCA anCas, which presumably does not use the HNH domain for cleavage and seems to function more optimally on single-stranded substrates, akin to other types of Cas nucleases. In addition, FCA anCas has shown a prominent nickase activity, which makes it a potential candidate for double-nicking editing. For gene *TYR*, an evolutionary trend in indels proportion is observed (Fig. 6b), by which the indel proportion decreases over time. In the case of gene *OCA2* the proportion of indels reaches levels like those of SpCas9, e.g., PCA anCas, but even for oldest SCA and BCA anCas the levels are quite high, in the range of 40% (Fig. 6b). Comparable results were obtained analysing the NGS output files with the publicly available CrispRVariants R-software package [50]. Furthermore, the analysis of NGS reads also allows determination of preferred cleavage locations as well as the appearance of the most frequent genome-edited alleles, whose pattern is maintained across all anCas that cleave these DNA sequences (Fig. 6c, d and Extended Data Fig. 9). This suggests that the outcome of the genome editing experiments, irrespective of the Cas used (anCas or SpCas9), is strongly determined by the surrounding genomic DNA sequence, as has previously been reported for SpCas9 [51].

Finally, we also tested site-specific cleavage using a Traffic Light Reporter (TLR) based on RFP reconstitution (see methods). This method allows monitoring of DNA repair in HEK293T cells based on fluorescence activated cell sorting (FACS). Using conditions optimized for SpCas9, the results are in line with those determined by NGS (Extended Data Fig. 10), demonstrating their robustness.

Overall, we demonstrate that anCas nucleases can potentially be used to edit cells; however, it is important to mention that in our experiments the conditions tested are favorable to SpCas9. We expect that searching other conditions involving, for instance, different sgRNA scaffolds, targets with various sequence contexts, nuclear location signal (NLS) configurations on the Cas protein, timing, etc, could lead to improvements allowing for anCas mediated gene editing in conditions beyond the current SpCas9 capabilities.

Discussion

In the present study, we traced the evolutionary history of Cas9 endonucleases going back to the origin of Firmicutes species, a phylum where Cas9 is widely present. Given the age of the earliest forms studied here, we are likely facing some of the primordial forms of Class 2 effectors, perhaps after development from prior elements from Class 1 systems, as it has been suggested[16, 52].The ancestral variants, display features not observed in SpCas9 and other Cas enzymes studied to date. The evolutionary trends observed for the nuclease/nickase activity, NGG PAM recognition, gRNA recognition, and single-chain cleavage, highlight the adaptive character of the CRISPR-Cas systems in addition to offering features compatible with ancient functions. This adaptive process went likely from less restricted nickases to highly specific RNA-guided nucleases. Characteristics such as a PAM requirement or DNA unwinding are key

to secure proper recognition of foreign versus self-DNA, as well as R-loop formation for cleavage. Ancient endonucleases likely resulted from enzymes that worked with ssDNA or RNA [43], and thus did not require a potent dsDNA unwinding ability, perhaps only needing additional proteins or co-factors to exert DSB activity [43]. The observation that the oldest anCas work better than SpCas9 with single-chain substrates demonstrates that they have good cleavage abilities but are perhaps less efficient at DNA unwinding. One possible hypothesis is that ancient Cas nickases did not work much as a potent defence system and therefore did not have an extensive collection of spacers, thus not needing a strong PAM recognition nor specific gRNA. This less restricted nucleases progressively adapted its function as defense mechanisms necessitating both precise PAM and gRNA recognition. This is clearly shown in the functional trends observed toward newer reconstructed enzymes. These functional abilities seem to have been already established for early *Streptococci* species.

The cleavage abilities of FCA anCas resemble those of other Cas enzymes such as Cas12a, Cas14 or CasΦ, in which cleavage results only from the action of the RuvC domain [53]. These similarities suggest a common origin for all Cas enzymes, making them homologs [16]. This was already suggested for Class 2 CRISPR-Cas systems, which diverged to develop distinctive abilities after a specialization process [2]. Features such as efficient PAM recognition and cleavage efficiency of dsDNA involving the HNH domain were already present in ancient *Streptococci* forms such as SCA, PCA and PDCA anCas. This suggests that these abilities were an important evolutionary trait perhaps associated with the divergence of modern *Streptococcus* Cas proteins. Overall, we can say that the most prominent feature of the oldest endonucleases is their chemical promiscuity and flexibility for PAM, gRNA, and substrate. This promiscuity has been suggested as a feature in early enzymes[54, 55]. Nevertheless, the exact link between evolution and specificity is a matter of debate[56, 57], and perhaps is not a general process for all enzymes. Still, in the case of Cas enzymes it seems that specificity has been sculpted over millions of years of evolution, likely driven by a continuous speciation and specialization process[55, 58, 59], which is expected in a system that is adaptive, like CRISPR-Cas systems are.

Ancestral Sequence Reconstruction (ASR) has demonstrated its validity to test evolutionary theories and its capacity to redesign enzymes with exceptional abilities based on their origins [18, 19, 54, 60-63]. The excellent, abundant work on ancestral proteins and genes carried out in prior years has clearly placed this method as an exquisite biomolecular design technique with several advantages over existing techniques [20, 64, 65]. Amongst the properties reported on ancestral proteins, we can find increments in stability, functional promiscuity, pH tolerance, altered immunogenicity, structural alterations, and quite often, some or all these properties at once. Therefore, ASR should not only be seen as an exotic method for studying molecular evolution, but also as a method for exploring the sequence space of proteins by adding a time component that greatly expands the number of sequences of any protein[59, 66]. No other technique can currently yield a protein with hundreds of mutations with respect to existing ones, and still be fully folded and functional. These novel sequences, which no longer exist in nature, can thus be considered synthetic redesigned genes with known and unknown properties that, in most cases, can be explained under the light of evolution. Nevertheless, ancestral enzymes may also be the initial point for further improvement and modifications of proteins using rational design or directed evolution [64, 67]. Thus, ASR is a timely addition to the list of valid methods for protein design, providing perhaps the best possible approach for (re-)designing proteins and enzymes.

In the case of CRISPR-Cas, efforts have been made towards using evolutionary information to redesign Cas9 variants with the purpose of expanding their PAM range recognition, efficiency, and accuracy [37, 38]. Similarly, directed evolution approaches have been used to increase Cas9 specificity [68]. These new Cas9 proteins have expanded the list of Cas9 proteins that have been engineered for similar purposes [69, 70]. Still, the obtained variants are closely related orthologs whose sequences show only a few mutations, with most of them still showing limitations inherent to SpCas9. To circumvent CRISPR-Cas9 limitations, researchers have embarked on searches to identify new variants of Cas nucleases [22, 42]. This has resulted in an expanded catalogue of Cas nucleases that display a large diversity of molecular features[42]. These Cas nuclease sequences provide the basis for the present study and allowed us to build a phylogenetic chronogram. Thus, discovering new Cas nucleases will provide new opportunities to expand the search in primordial time for their ancestors, further increasing the list of orthologs. Our study contributes the five anCas tested in this work, but also the sequences of the untested anCas nodes in the tree, which will likely work with known and, perhaps, new features.

The anCas tested in this study show a variety of features that make them potentially useful in genome editing applications. Altogether, the anCas variants have shown PAMless activity, gRNA promiscuity, nickase activity, marked activity on single-chain substrates, high cleavage efficiency, trans-cleavage, and diminished immune response to an AntiCas9 antibody. All these features are desirable in emerging gene editing tools. In addition, the cleavage test in human cells demonstrates that most anCas can perform DSB followed by NHEJ in living cells, highlighting the validity of the ancestral reconstruction approach for the design of novel endonucleases, beyond the evolutionary implications. In addition, the prominent nickase activity of the oldest Cas such as FCA, makes it a potential candidate for double-nicking editing. This functional flexibility put anCas nucleases forward as highly versatile endonucleases for biotechnological applications. Thus, this work represents a proof of concept that opens new avenues towards expanding the catalog of genome editing tools. Exploring these editing abilities including off-target analysis, double-nicking cleavage, prime editing, complex DNA regions or other specific tests for biotechnology applications are clear directions for future work.

Finally, we believe anCas nucleases may be basic elements in which we can follow a modular Lego-like approach to fuse structural and molecular elements to redesign synthetic nucleases, or even combine them with previously developed Cas variants. Combining ASR and rational design would expand even further the sequence space, making these synthetic enzymes the basis of a collection of nucleases created *à la carte*. We also believe that any nuclease, as well as any enzyme, offers the potential to be reconstructed, offering countless possibilities not only for gene editing, but also for biotechnology in general.

Methods

Ancestral sequence reconstruction of Cas9

Fifty-nine Cas9 sequences were downloaded from the NCBI database (Supplementary Table 1). Sequences belong to two bacterial phyla: Firmicutes and Actinobacteria. NCBI database. We use the BLAST tool with custom parameters and criteria, i.e., maximum of 1000 hits, minimum identity 35% to ensure the selection of Cas9 type IIA sequences (individually inspected) and BLOSUM62 scoring matrix. E-values were virtually zero for all sequences. The sequences were

selected following similar proportions of sequences in each phylum as in the database. Alignment of the sequences was performed using MUSCLE software on the MEGA platform and manually edited to eliminate gaps, poorly aligned sites, and divergent regions. We inferred the best evolutionary model using MEGA, resulting in the Le and Gascuel (LG)[71] with gamma distribution model (8 categories), Yule model for speciation and length chain of 100 million generations, sampling every 1000 generations. Phylogeny was carried out using BEAST v1.10.4 package software (https://beast.community/) including the BEAGLE library for parallel processing and based on Bayesian inference using Markov chain Monte Carlo (MCMC). We set *Streptococci* species as monophyletic group. Outgroup sequences were selected from taxa falling outside of the Firmicutes phylum to place the root of the tree. Divergence times were estimated using the Reltime method[72] implemented in MEGA with discrete eight-categories Gamma distribution for evolutionary rates. We set calibration times using information from the TimeTree of Life (TToL 4th edition)[73, 74] in three major clades with 95% confidence intervals (CI). For FCA 2607 MYA, CI: 2448-2753 MYA; For BCA 924 MYA, CI: 568-1241 MYA; and for SCA 201 MYA, CI: 62-340 MYA. Calculations were run in a multicore server. From the generated trees we discarded the initial 25% as burn-in using the LogCombiner utility from BEAST. We verified the MCMC log file using TRACER and ensure all parameters showed effective sample size (ESS)>100. Posterior probabilities of all nodes were above 0.58 with most of them near 1(Supplementary Fig. 1). Figtree v1.4.3 was used to visualize and edit the phylogenetic tree. Finally, ancestral sequence reconstruction was performed by maximum likelihood using PAML v4.9 (http://abacus.gene.ucl.ac.uk/software/paml.html) with a gamma distribution of 8 categories for variable replacement rates across sites. Posterior probabilities were calculated for all amino acids and the residue with the highest posterior probability was chosen for each site. We selected from the tree to reconstruct different ancestors with different age named as anCas. These ancestors included an evolutionary line connecting an early ancestor of Firmicutes with modern *Streptococcus pyogenes.*

Structure prediction of anCas endonucleases using AlphaFold2

The AlphaFold2 models[27, 29] were calculated on the Flemish supercomputer centre (VSC) cluster at the University of Gent. The full AlphaFold2 pipeline was employed, with the main inference step executed five times, besides standard MSA search, template search, and constrained relaxation. pLDDT values were in turn mapped to the B-factor column of the corresponding coordinate model using a custom Python script. Visualization of the resulting models, with colouring according to pLDDT values (from 0 to 100), was performed with PyMol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). The spectrum bar was generated using the spectrumBar.py script (PyMOLWiki). Protein superposition was done with the SUPERPOSE software[75] of the CCP4 program suite[76].

Protein production and purification

anCas genes were synthesized, and codon optimized for *E. coli* cell expression by GenScript (New Jersey, U.S.). All anCas genes were cloned into pBAD/His expression vector (ThermoFisher Scientific) and transformed in *E. coli* BL21 (DE3) (Life Technologies) for protein expression. SpCas9 gene was purchased from addgene (Plasmid #62934). Cells were incubated in LB medium at 37 °C until OD600 reached 0.6, L-arabinose was added to 0.1% to anCas. IPTG was added to SpCas9 to 1 mM concentration for protein induction overnight at 20

ºC. Cells were pelleted by centrifugation at 4000 rpm. Pellets were resuspended in extraction buffer (20 mM Tris-HCl pH 7.9, 200 mM NaCl, 25 mM Imidazole, 0.5 mM TCEP). 100 mg/mL of lysozyme (Thermo Scientific) was added to the pellet and incubated for 15 min. Then pellet was sonicated for 3 cycles for 10 min at 30% amplitude. Cell debris was separated by ultracentrifugation at 33,000 x g for 1 h. For purification, the supernatants were mixed with His GraviTrap affinity column (GE Healthcare) and eluted in elution buffer (20 mM Tris-HCl pH 7.9, 200 mM NaCl, 300 mM Imidazole, 0.5 mM TCEP). Fractions were collected and loaded to a HiTrap Heparin HP column (GE Healthcare) and eluted using a linear gradient of 0.2-2 M NaCl. Proteins were further purified by size exclusion chromatography using a Superdex 200 HR column (GE Healthcare) and eluted in 20 mM Tris-HCl pH 7.9, 200 mM NaCl. For protein purification verification sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) was used with 8% gels. The protein concentration was calculated by measuring the absorbance at 280 nm in Nanodrop 2000C.

sgRNA synthesis

sgRNA with the complementary sequence to the target was synthesized and cloned into pUC18 vector. sgRNA sequence was amplified by PCR using Phusion® Hot Start Flex DNA Polymerase (NEB). PCR product was purified using mi-PCR Purification Kit (Metabion). sgRNA was synthesized using HiScribe T7 High Yield RNA Synthesis Kit (NEB). The PCR fragment includes the T7 promoter at 5´ end and the sequence from sgRNA of *S. pyogenes* at 3´ ends. The reaction was incubated overnight, and gRNA was purified following the protocol of the Monarch® RNA Purification Columns Kit. sgRNA integrity was analyzed by electrophoresis on 2% agarose gel in TBE buffer.

In vitro cleavage assay

In vitro cleavage assay was performed with purified anCas and SpCas9 endonucleases. In all assays, 30 nM enzyme was incubated for 15 min with 30 nM gRNA at 1:1 ratio in the cleavage buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/BSA, pH 7.9) at 37 ºC. Then, 3 nM target DNA (plasmid DNA containing TGG PAM) was added and incubated at different times depending on the experiment. Reaction was stopped by adding 6X loading dye (NEB) with EDTA and run on 2% agarose gel. Gels were dyed with SYBR gold (ThermoFisher Scientific) and imaged with ChemiDoc XRS + System (Bio-Rad). Cleavage was quantified by ImageJ.

PAM library construction

DNA library containing seven random nucleotides was designed and cloned into pUC18 plasmid by Genscript (Supplementary Table 5). This random library was transformed in XL1blue *E. coli* and amplified several times to achieve the maximal variability in the PAM sequences. A PCR fragment of 844 bp was amplified using the primers (Supplementary Table 6) from the DNA library containing the 7 random nucleotides.

PAM determination

PAM determination assay was performed by assembling in cleavage buffer: 3 nM of PCR fragment from the DNA library with 30 nM of each anCas and SpCas9 and 30 nM of gRNA

targeting the 20 nucleotides upstream the 7 random nucleotides. The reaction was incubated for 1 hour at 37 ºC and stopped by adding 6X loading dye (NEB) with EDTA and run on 2% agarose gel. Gels were dyed with SYBR gold (ThermoFisher Scientific) and imaged with ChemiDoc XRS + System (Bio-Rad). The small fragment of 278 bp was purified from the agarose gel with GeneJet Gel Extraction kit (ThermoFisher Scientific). The quantity and quality of the DNA were evaluated by Qubit dsDNA HS DNA kit (ThermoFisher Scientific) and agarose gels. Sequencing libraries were prepared following TruSeq Chip Sample Preparation Guide with the corresponding kit (Illumina Inc. Cat #IP- 202-1012 or IP-202-1024, (Set A and B). Input fragmented DNA (5-10 ng) was blunt-ended and phosphorylated. A single A nucleotide was added to the 3´ ends of the fragments in preparation for ligation to an adapter with a single T base overhang. The ligation products were purified and accurately size-selected by AMPure XP beads. Size-selected DNA was purified and PCR-amplified to enrich for fragments that have adapters on both ends. Libraries were visualized on an Agilent 2100 Bioanalyzer using an Agilent High Sensitivity DNA kit and quantified using Qubit dsDNA HS DNA kit (ThermoFisher Scientific). Libraries were sequenced in single end by NovaSeq 6000 Sequencer. Illumina reads were mapped to the reference library containing the random PAM sequence using Geneious Prime (2020 version). The reads aligned to the reference with zero mismatches and started 3 nucleotides right after the random PAM sequence were selected and extracted from the bulk. Each PAM sequence was quantified, and its frequency was calculated from the total PAM previously extracted. From the frequency of each PAM we generate PAM wheels following previously published methods[77].

For *in vitro* cleavage of PAM sequences, different DNA fragments carrying each PAM were cloned into Zero-Blunt TOPO plasmid (Supplementary Table 7). The cleavage assay was performed in cleavage buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/BSA, pH 7.9) at 37 ºC. 3 nM of anCas and SpCas9 were incubated for 15 min with 3 nM gRNA at 1:1 ratio in cleavage buffer and 3 nM DNA plasmid was added. After 10 min, reaction was stopped by adding 6X loading dye (NEB) with EDTA and run on 2% agarose gel. Similarly, gels were dyed with SYBR gold (ThermoFisher Scientific) and imaged with ChemiDoc XRS + System (Bio-Rad). Cleavage was quantified by ImageJ.

Human Cell Culture and Transfections for HT-PAMDA

Human HEK 293T cells (ATCC) were maintained in DMEM (Gibco) supplemented with 10% Heat-inactivated FBS (Gibco) and 1% penicillin/streptomycin (Gibco) in a 37 °C and 5% CO_2 incubator. Cells were confirmed to be free of mycoplasma by analyzing supernatant media monthly using MycoAlert™ PLUS (Lonza). For HT-PAMDA experiments, transfections were performed between 16 and 20 hours after seeding 150,000 cells per well in a 24-well tissue culture treated plate (Corning). Approximately 600-700 ng of Cas expression vector (Supplementary Table 8) were combined with 1.5 µL of TransIT-X2 (Mirus) into a total volume of 50 µL including Opti-MEM (Thermo Fisher). Reagent mixture was gently vortexed, incubated at room temperature for 15 minutes, and added to cells with gentle shaking of the plate.

HT-PAMDA Sample Preparation, Sequencing, and Data Analysis

The HT-PAMDA assay was performed essentially as previously described[39, 40]. Briefly, cell lysates containing Cas nucleases were harvested approximately 48 hours post-transfection by

carefully removing the culture media and resuspending cells in 100 µL of gentle lysis buffer (1X SIGMAFAST Protease Inhibitor Cocktail, EDTA-Free (Millipore Sigma), 20 mM Hepes pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 5% glycerol, 1 mM DTT, and 0.1% Triton X-100). The approximate quantity of Cas nuclease from each lysate was estimated based on EGFP fluorescence values and the lysates were normalized to approximately 150 nM Fluorescein (Sigma) based on a Fluorescein titration conducted in parallel and the resulting standard curve. Fluorescence was measured in a 384-well plate (volume of sample: 10 µL) on a DTX 880 Multimode Plate Reader (Beckman Coulter) with an excitation wavelength of 485 nm and emission wavelength of 535 nm. SpCas9 sgRNAs were *in vitro* transcribed at 37 °C for 16 hours from roughly 1 µg of HindIII linearized sgRNA T7-transcription plasmid template (Supplementary Table 9) using the T7 RiboMAX Express Large Scale RNA Production Kit (Promega). The DNA template was degraded by the addition of 1 µL RQ1 DNase at 37 °C for 15 minutes. sgRNAs were purified using a MinElute PCR Purification kit (Qiagen) and refolded by heating to 90 ºC for 5 minutes and then cooling to room temperature for 15 minutes.

To perform the *in vitro* cleavage assays, linearized plasmid libraries with 8 nucleotide randomized PAMs on the 3' end of the Cas target site were used as substrates (Supplementary Table 10)**.** To generate the Cas RNP, 4.375 µL of normalized Cas-containing lysates were complexed with 3.5 µL of 2.5 µM *in vitro* transcribed sgRNAs for 3-10 minutes at 37 °C. To perform the cleavage reaction, 1.75 µL of 25 nM randomized PAM library linearized plasmid was incubated with the RNP mixture at 37 °C with a final volume of 17.5 µL and final buffer composition of 10 mM Hepes pH 7.5, 150 mM NaCl, and 5 mM $MgCl_2$. Timepoints from each cleavage reaction were terminated at 1 min, 8 min, and 32 min by removing 5 µL aliquots into 5 µL of prepared stop buffer (50 mM EDTA and 2 mg/mL Proteinase K [NEB]) and incubating at room temperature for 10 minutes and 98 °C for 5 minutes. All Cas were assayed using two distinct PAM libraries comprising different spacer sequences.

Approximately 3 ng of the digested HT-PAMDA substrate from each Cas nuclease and timepoint were used as template for PCR reactions using Q5 polymerase (NEB) and primers that add unique i5 and i7 barcodes (Supplementary Table 11). The PCR products were pooled according to timepoint and purified twice via paramagnetic beads (AMPure XP; Beckman Coulter). The pooled clean PCR products were diluted to a concentration of approximately 0.125 ng/µL and used as the template in a subsequent PCR reaction using barcoded primer pairs (Supplementary Table 11). PCR products were not treated with Exonuclease I, differing from the HT-PAMDA protocol[40] to prevent unique barcode digestion from PCR ends. The resulting library was sequenced on a NextSeq sequencer using a 75-cycle NextSeq 500/550 High Output v2.5 kit (Illumina) at the Molecular Biology Core Facility at Dana-Farber Cancer Institute. Sequencing reads were analyzed via Python scripts (available on Zenodo: https://zenodo.org/record/3710516#.YlRtLdPMLGw). Rate constants were determined for each Cas enzymes on both substrate libraries, as previously described[39, 40].

In vitro cleavage assay for gRNA promiscuity

For *in vitro* cleavage for gRNA promiscuity, DNA plasmid carrying TGG PAM was used. The cleavage assay was performed in cleavage buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/BSA, pH 7.9) at 37 ºC. 3 nM of anCas and SpCas9 were incubated for 15 min

with 3 nM sgRNA of each species (Supplementary Table 12) at 1:1 ratio in cleavage buffer and 3 nM DNA plasmid was added. After 10 min, the reaction was stopped by adding 6X loading dye (NEB) with EDTA and run on 2% agarose gel. Similarly, gels were dyed with SYBR gold (ThermoFisher Scientific) and imaged with ChemiDoc XRS + System (Bio-Rad). Cleavage was quantified by ImageJ.

In vitro cleavage assay for ssDNA and ssRNA

In vitro cleavage assay was performed with purified FCA anCas, BCA anCas and SpCas9 endonucleases. In all assays, 30 nM enzyme was incubated for 15 min with 30 nM sgRNA (Spy-sgRNA 20 nt) at 1:1 ratio in the cleavage buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/BSA, pH 7.9) at 37 ºC. Then, 3 nM of target fragment (ssDNA or ssRNA, Supplementary Table 13) were added and incubated for different time intervals (0, 5, 10, 30 and 60 min). For ssDNA target, reaction was stopped by adding 6X loading dye (NEB) with urea. Samples were heated for 10 min at 80 ºC and were resolved by 2.5% denaturing urea agarose gel. In the case of trans-activity assay, 3 nM of M13 circular ssDNA (NEB) were added and incubated for 30 min at 37 ºC. For ssRNA target, reaction was stopped by adding 2X RNA gel-loading buffer (NEB) with urea. Samples were heated for 10 min at 95 ºC and were resolved by 15% denaturing urea polyacrylamide gel electrophoresis. In all cases, gels were dyed with SYBR gold (ThermoFisher Scientific) and imaged with ChemiDoc XRS + System (Bio-Rad). Cleavage was quantified by ImageJ and fitted with single-exponential decay curve.

ELISA test

Elisa test was performed by using a modified protocol described elsewhere[78]. Briefly, 1 µg/well of SpCas9, FCA anCas, BCA anCas and bovine serum albumin (BSA, Sigma Aldrich) were diluted in 1x bicarbonate buffer and coated onto 96-well plates (ThermoFisher Scientific) overnight at 4 ºC. Plates were washed with 1X wash buffer (TBST, ThermoFisher Scientific) and the blocking was performed with 1% BSA blocking solution for 1 hour at room temperature. Next, anti-Cas9 rabbit antibody (Rockland, 600-401-GK0) was diluted 1:25000 in 1% BSA blocking solution, added to the plates, and incubated for 2 hours at room temperature. Then, plates were washed and HRP-conjugated goat anti-Rabbit IgG (H+L) (Invitrogen), diluted 1:2000 in 1% BSA blocking solution, was added, and incubated for 1 hour at room temperature. Finally, 3,3′,5,5′-Tetramethylbenzidine ELISA substrate solution (ThermoFisher Scientific) was added and incubated for 10 min at room temperature. The reaction was stopped with 1 N sulfuric acid. The absorbance was measured at 450 nm by using a VICTOR X5 microplate reader (PerkinElmer).

Endogenous gene editing in human HEK 293T cells

Functional validation of ancestral Cas nucleases was carried out in human HEK293T cells, as described elsewhere[79]. Cells were grown in DMEM medium (Dulbecco' s Modified Eagle Medium, Gibco), supplemented with sterile-filtered 10% fetal bovine serum (FBS), 10 mM HEPES pH 7.4, 2 mM L-glutamine and penicillin (100 IU/ml)–streptomycin (100 µg/ml) and handled under aseptic conditions using a sterile hood. HEK293T cells were cultured in incubators at 37 ºC, 95% humidity and 5% CO_2. Humanized anCas were cloned into pcDNA3.1 plasmid expression vector (ThermoFisher). The sgRNA target sequences were designed with

Breaking-Cas web tool[80] and cloned into MLM3636 plasmid vector (Addgene #43860) through Golden Gate cloning method. HEK 293T cells were cultured in incubators at 37 ºC, 95% humidity and 5% CO_2. Humanized anCas (Supplementary Table 14) were cloned into pcDNA3.1 plasmid expression vector by GenScript (New Jersey, U.S.). The sgRNA target sequences were designed with Breaking-Cas web tool (*63*) and cloned into MLM3636 plasmid vector (Addgene #43860) through Golden Gate cloning method. SpCas9 from hCas9 plasmid (Addgene #41815) was used as a positive control. For the *in vivo* genome-editing tests, cells were plated in 24-well plates at a density of 4 x 10^5 cells/ml in a 0.5 ml volume of DMEM without antibiotics. To these cells, 1 µg of hCas/hanCas plasmid and 0.5 µg of the corresponding sgRNA plasmid were transfected with 2 µl of Lipofectamine 2000 (Life Technologies) diluted in 100 µl of Opti-MEM (Gibco) per well. 72 hours post-transfection genomic DNA was isolated with High Pure Template Preparation Kit (Roche). INDEL occurrence was assessed by T7 Endonuclease I assay by using a protocol described previously[81]. Briefly, DNA fragment surrounding the target DSB was amplified by PCR. The PCR product was melted and cooled as reported[79] to promote the formation of heteroduplexes. Finally, 10 µL of the PCR product were digested with 0.3 µL of T7 Endonuclease I (NEB) in a total reaction volume of 20 µL for 30 min at 37 ºC and run on 2% agarose gel.

Characterization of gene-edited allelic variants on TYR and OCA2 genes

DNA fragment covering the target region for the designed gRNAs at exon 1 of *TYR* (NM_000372.5; Gene ID: 7299) and exon 14 of *OCA2* (NM_000275.3; Gene ID: 4948), in HEK293T human cells was PCR-amplified using customized primers that included the adapters for NGS sequencing followed by the gene-specific sequence (Supplementary Table 15). Library preparation for all the samples in duplicate generated after cleavage and repair with humanized versions of SpCas9, PDCA, PCA, SCA, BCA and FCA anCas, was carried out as previously described[82]. Briefly, PCR products were indexed using the Nextera XT DNA library preparation kit (Illumina), purified, pooled, and sequenced on a MiSeq platform (Illumina, San Diego, CA, USA) using a 2x250 paired-end setting. The average depth of coverage was roughly 20.000x for each sample. Percentage of the different alleles after Cas9-mediated editing was determined using Mosaic Finder (MF)[49, 83]. MF's pipeline integrates NGS read mapping, normalization of read counts, mutation frequency calculation and genome-editing efficiency statistics at each position of the target region. MF takes, as input files, the fastq generated by pair-end sequencing and generates consensus sequences by joining the corresponding read pairs (forward and reverse). The repertoire of consensus sequences (allelic clusters) represents the allelic diversity generated by the Cas9-mediated editing. These clusters are then aligned against the sequence used as reference and are classified in allelic classes based on the different type of the identified mutation (mismatch, insertion/deletion). The frequency of each cluster is then calculated and plotted. For multiple sequence alignment visualization of the different alleles classified by MF we have used the Jalview program (https://www.jalview.org/). A similar analysis was carried out using other available software such as CrispRVariants R-software package[50] (data available upon request).

Traffic Light reporter (TLR) cell generation

HEK293T cells were infected with TLR lentivirus at a 0.2 multiplicity of f (MOI) to generate the TLR cell line as previously described [84]. After infection, cells were selected with puromycin

following standard protocols for 2 weeks before being used for subsequent assays. Cell transfection experiments were performed with Lipofectamine 3000 (Thermo Fisher Scientific) with 0.5 µg of nuclease and 0.5 µg of gRNA expression DNA vectors (Supplementary Table 16). Cells were analysed by flow cytometry using BD LSR Fortessa (BD Bioscience. Blue 488 nm laser with 530/30 filter and Yellow Green 561nm laser with 610/20 filter) 3-4 days after transfection. The relative NHEJ frequency was estimated by the number of RFP-positive cells. SpCas9 gRNA expression cassettes were cloned on TOPO vectors using manufacturing recommendations (Zero Blunt® TOPO®, Thermo Fisher Scientific).

Data Availability

We have made available sequencing data. Other data supporting the findings of this study are available from the corresponding authors upon reasonable request. Plasmids for gene editing in human cells are available through Addgene (Supplementary Table 14; see also https://www.addgene.org/Raul_Perez-Jimenez/). Plasmids for HT-PAMDA experiments are available through Addgene (Supplementary Tables 8-10; see also www.addgene.org/Benjamin_Kleinstiver/). Sequencing data for PAM determination and gene editing experiments will be made available through the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) under BioProject ID PRJNA832610. Sequencing data for HT-PAMDA experiments will be made available through the NCBI SRA under BioProject ID PRJNA832159 (Supplementary Data 2).

References

1. Mohanraju, P. et al. Alternative functions of CRISPR–Cas systems in the evolutionary arms race. *Nature Reviews Microbiology* (2022).
2. Makarova, K.S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* **18**, 67-83 (2020).
3. Garneau, J.E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71 (2010).
4. Barrangou, R. et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**, 1709-1712 (2007).
5. Karginov, F.V. & Hannon, G.J. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* **37**, 7-19 (2010).
6. Jinek, M. et al. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821 (2012).
7. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740 (2009).
8. McGinn, J. & Marraffini, L.A. Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nature Reviews Microbiology* **17**, 7-12 (2019).
9. Guzmán, N.M., Esquerra-Ruvira, B. & Mojica, F.J.M. Digging into the lesser-known aspects of CRISPR biology. *International Microbiology* **24**, 473-498 (2021).

10. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573 (2014).
11. Singh, D., Sternberg, S.H., Fei, J., Doudna, J.A. & Ha, T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nature Communications* **7**, 12778 (2016).
12. Mali, P., Esvelt, K.M. & Church, G.M. Cas9 as a versatile tool for engineering biology. *Nature Methods* **10**, 957-963 (2013).
13. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. & Doudna, J.A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67 (2014).
14. Wimmer, F. & Beisel, C.L. CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers. *Frontiers in Microbiology* **10** (2020).
15. Weissman, J.L., Stoltzfus, A., Westra, E.R. & Johnson, P.L.F. Avoidance of Self during CRISPR Immunization. *Trends Microbiol* **28**, 543-553 (2020).
16. Koonin, E.V. & Makarova, K.S. Origins and evolution of CRISPR-Cas systems. *Philos Trans R Soc Lond B Biol Sci* **374**, 20180087 (2019).
17. Manteca, A. et al. Mechanochemical evolution of the giant muscle protein titin as inferred from resurrected proteins. *Nat Struct Mol Biol* **24**, 652-657 (2017).
18. Perez-Jimenez, R. et al. Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat Struct Mol Biol* **18**, 592-596 (2011).
19. Barruetabeña, N. et al. Resurrection of efficient Precambrian endoglucanases for lignocellulosic biomass hydrolysis. *Communications Chemistry* **2**, 76 (2019).
20. Zakas, P.M. et al. Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nature Biotechnology* **35**, 35-37 (2017).
21. Spence, M.A., Kaczmarski, J.A., Saunders, J.W. & Jackson, C.J. Ancestral sequence reconstruction for protein engineers. *Current Opinion in Structural Biology* **69**, 131-141 (2021).
22. Burstein, D. et al. New CRISPR–Cas systems from uncultivated microbes. *Nature* **542**, 237-241 (2017).
23. Marin, J., Battistuzzi, F.U., Brown, A.C. & Hedges, S.B. The Timetree of Prokaryotes: New Insights into Their Evolution and Speciation. *Molecular Biology and Evolution* **34**, 437-446 (2016).
24. Sheridan, P.P., Freeman, K.H. & Brenchley, J.E. Estimated Minimal Divergence Times of the Major Bacterial and Archaeal Phyla. *Geomicrobiology Journal* **20**, 1-14 (2003).
25. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949 (2014).
26. Jinek, M. et al. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **343**, 1247997 (2014).
27. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
28. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596 (2021).
29. 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).
30. Feng, W. et al. CRISPR technology incorporating amplification strategies: molecular assays for nucleic acids, proteins, and small molecules. *Chem Sci* **12**, 4683-4698 (2021).
31. 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).
32. Chen, J.S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436-439 (2018).

33. Pausch, P. et al. CRISPR-Cas(Phi) from huge phages is a hypercompact genome editor. *Science* **369**, 333-337 (2020).
34. Karvelis, T. et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome biology* **16**, 253 (2015).
35. Chatterjee, P., Jakimo, N. & Jacobson, J.M. Minimal PAM specificity of a highly similar SpCas9 ortholog. *Science advances* **4**, eaau0766 (2018).
36. Kleinstiver, B.P. et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature biotechnology* **33**, 1293-1298 (2015).
37. Chatterjee, P. et al. An engineered ScCas9 with broad PAM range and high specificity and activity. *Nature Biotechnology* **38**, 1154-1158 (2020).
38. Miller, S.M. et al. Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nature Biotechnology* **38**, 471-481 (2020).
39. Walton, R.T., Christie, K.A., Whittaker, M.N. & Kleinstiver, B.P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290-296 (2020).
40. Walton, R.T., Hsu, J.Y., Joung, J.K. & Kleinstiver, B.P. Scalable characterization of the PAM requirements of CRISPR–Cas enzymes using HT-PAMDA. *Nature Protocols* **16**, 1511-1547 (2021).
41. Dooley, S.K., Baken, E.K., Moss, W.N., Howe, A. & Young, J.K. Identification and Evolution of Cas9 tracrRNAs. *Crispr j* **4**, 438-447 (2021).
42. Gasiunas, G. et al. A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nature Communications* **11**, 5512 (2020).
43. Ma, E., Harrington, L.B., O'Connell, M.R., Zhou, K. & Doudna, J.A. Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol Cell* **60**, 398-407 (2015).
44. Briner, A.E. et al. Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell* **56**, 333-339 (2014).
45. Charlesworth, C.T. et al. Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nature Medicine* **25**, 249-254 (2019).
46. 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).
47. Gaucher, E.A., Govindarajan, S. & Ganesh, O.K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704-707 (2008).
48. Fernández, A. et al. Genetics of non-syndromic and syndromic oculocutaneous albinism in human and mouse. *Pigment Cell & Melanoma Research* **34**, 786-799 (2021).
49. Cervera, S.T. et al. Therapeutic Potential of EWSR1–FLI1 Inactivation by CRISPR/Cas9 in Ewing Sarcoma. *Cancers* **13**, 3783 (2021).
50. Lindsay, H. et al. CrispRVariants charts the mutation spectrum of genome engineering experiments. *Nature Biotechnology* **34**, 701-702 (2016).
51. Leenay, R.T. et al. Large dataset enables prediction of repair after CRISPR–Cas9 editing in primary T cells. *Nature Biotechnology* **37**, 1034-1037 (2019).
52. Mohanraju, P. et al. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147 (2016).
53. Wang, J.A.-O., Pausch, P.A.-O. & Doudna, J.A.-O.X. Structural biology of CRISPR-Cas immunity and genome editing enzymes. LID - 10.1038/s41579-022-00739-4 [doi]. *Nature Review Microbiology* (2022).
54. Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A. & Sanchez-Ruiz, J.M. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. *Journal of the American Chemical Society* **135**, 2899-2902 (2013).

55. Risso, V.A., Gavira, J.A. & Sanchez-Ruiz, J.M. Thermostable and promiscuous Precambrian proteins. *Environ Microbiol* **16**, 1485-1489 (2014).
56. Wheeler, L.C., Lim, S.A., Marqusee, S. & Harms, M.J. The thermostability and specificity of ancient proteins. *Curr Opin Struct Biol* **38**, 37-43 (2016).
57. Siddiq, M.A., Hochberg, G.K. & Thornton, J.W. Evolution of protein specificity: insights from ancestral protein reconstruction. *Curr Opin Struct Biol* **47**, 113-122 (2017).
58. Jensen, R.A. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* **30**, 409-425 (1976).
59. Gumulya, Y. & Gillam, E.M. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro' approach to protein engineering. *Biochem J* **474**, 1-19 (2017).
60. Merkl, R. & Sterner, R. Ancestral protein reconstruction: techniques and applications. *Biological chemistry* **397**, 1-21 (2016).
61. Jones, B.J. et al. Larger active site in an ancestral hydroxynitrile lyase increases catalytically promiscuous esterase activity. *PLoS One* **15**, e0235341 (2020).
62. Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N. & Gouy, M. Parallel adaptations to high temperatures in the Archaean eon. *Nature* **456**, 942-945 (2008).
63. Blanquart, S. et al. Resurrection of Ancestral Malate Dehydrogenases Reveals the Evolutionary History of Halobacterial Proteins: Deciphering Gene Trajectories and Changes in Biochemical Properties. *Mol Biol Evol* **38**, 3754-3774 (2021).
64. Risso, V.A. et al. De novo active sites for resurrected Precambrian enzymes. *Nature Communications* **8**, 16113 (2017).
65. Santiago-Ortiz, J. et al. AAV ancestral reconstruction library enables selection of broadly infectious viral variants. *Gene Therapy* **22**, 934-946 (2015).
66. Harms, M.J. & Thornton, J.W. Analyzing protein structure and function using ancestral gene reconstruction. *Curr Opin Struct Biol* **20**, 360-366 (2010).
67. Alcalde, M. When directed evolution met ancestral enzyme resurrection. *Microb Biotechnol* **10**, 22-24 (2017).
68. Lee, J.K. et al. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nature Communications* **9**, 3048 (2018).
69. Slaymaker, I.M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88 (2016).
70. Kleinstiver, B.P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495 (2016).
71. Le, S.Q. & Gascuel, O. An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution* **25**, 1307-1320 (2008).
72. Tamura, K. et al. Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences* **109**, 19333-19338 (2012).
73. Kumar, S., Stecher, G., Suleski, M. & Hedges, S.B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution* **34**, 1812-1819 (2017).
74. Hedges, S.B. & Kumar, S. The timetree of life. (Oxford University Press, Oxford ; New York; 2009).
75. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268 (2004).
76. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D* **50**, 760-763 (1994).

77. Leenay, R.T. et al. Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Molecular cell* **62**, 137-147 (2016).
78. Wang, D. et al. Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses. *Hum Gene Ther* **26**, 432-442 (2015).
79. Harms, D.W. et al. Mouse Genome Editing Using the CRISPR/Cas System. *Curr Protoc Hum Genet* **83**, 15.17.11-27 (2014).
80. Oliveros, J.C. et al. Breaking-Cas-interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes. *Nucleic acids research* **44**, W267-271 (2016).
81. Seruggia, D., Fernández, A., Cantero, M., Pelczar, P. & Montoliu, L. Functional validation of mouse tyrosinase non-coding regulatory DNA elements by CRISPR–Cas9-mediated mutagenesis. *Nucleic acids research* **43**, 4855-4867 (2015).
82. Fernandez, A. et al. Simple Protocol for Generating and Genotyping Genome-Edited Mice With CRISPR-Cas9 Reagents. *Curr Protoc Mouse Biol* **10**, e69 (2020).
83. López-Márquez, A.A.-O. et al. CRISPR/Cas9-Mediated Allele-Specific Disruption of a Dominant COL6A1 Pathogenic Variant Improves Collagen VI Network in Patient Fibroblasts. *International Journal of Molecular Sciences* **23**, 4410 (2022).
84. Certo, M.T. et al. Tracking genome engineering outcome at individual DNA breakpoints. *Nature Methods* **8**, 671-676 (2011).

Acknowledgments

This work has been supported by grants PID2019-109087RB-I00 to R.P.-J and grant RTI2018-101223-B-I00 and and PID2021-127644OB-I00 to L.M. from Spanish Ministry of Science and Innovation. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964764 to R.P.-J. The content presented in this document represents the views of the authors, and the European Commission has no liability in respect to the content. We acknowledge financial support from Spanish Foundation for the promotion of research of Amyotrophic Lateral Sclerosis (FUNDELA). A.F acknowledges CIBERER intramural funds (ER19P5AC756/2021). F.J.M.M acknowledges research support by Conselleria d'Educació, Investigació, Cultura i Esport from Generalitat Valenciana, research projects PROMETEO/2017/129 and PROMETEO/2021/057. M.M acknowledges funding from the Spanish Center for Biomedical Network Research on Rare Diseases (CIBERER) grant ER19P5AC728/2021. The work has received funding from the Regional Government of Madrid (CAM) grant B2017/ BMD3721 to M.A.M-P and from Instituto de Salud Carlos III cofounded with the European Regional Development Fund (ERDF), "A way to make Europe") within the National Plans for Scientific and Technical Research and Innovation 2017–2020 and 2021–2024 (PI17/1659; PI20/0429 and IMP/00009) to M.A.M-P. B.P.K. was supported by an MGH ECOR Howard M. Goodman Award and NIH P01 HL142494. We thank H. Stutzman for assistance cloning plasmids, and Z. Herbert and M. Berkeley from the Molecular Biology Core Facilities at the Dana-Farber Cancer Institute for assistance with NextSeq sequencing.

Author contributions

R. P-J. conceived the project. R. P-J., B. A-L., B. P. K., A. S-M., M. G., F. J. M. M., M.A. M-P., L. M., designed research and planned experiments. R. P-J. performed the phylogenetic analysis and ancestral sequence reconstruction. B. A.-L., Y. J, S.S., cloned and expressed proteins and performed in vitro experiments. L.T.H. performed the HT-PAMDA experiments. L.T.H., R.A.S.

and B.P.K. analyzed HT-PAMDA data. B. A.-L., Y. J, S.S., M.M, A. F., Y. B., S.F-P, L.S., A. S-M., M.G., M.A. M-P., and L. M, performed functional validation of anCas in mammalian cells, sequencing experiments and bioinformatic analysis. A-L, Y. J, R. P-J., J.A.G., A. R., A. D., and W. V., designed analysed and represented structural data. All authors participated in discussions and provided ideas for the work. R.P-J., F. J. M. M., M.A. M-P., L. M wrote the original paper and all authors revised and edited the manuscript.

Competing interests

R. P-J., B. A-L. are co-inventors on patent application filed by CIC nanoGUNE and licenced to Integra Therapeutics S.L. relating to work in this article. A. S-M. and M.G. are co-founders of Integra Therapeutics S.L. B.P.K is an inventor on patents and/or patent applications filed by Mass General Brigham that describe genome engineering technologies. B.P.K. is a consultant for EcoR1 capital, and is an advisor to Acrigen Biosciences, Life Edit Therapeutics, and Prime Medicines.

Additional information

Extended data is available for this paper at…

Supplementary information The online version contains supplementary material available at..

Correspondence and requests for materials should be addressed to R. P.-J: r.perezjimenez@nanogune.eu

Tables

Table 1. Kinetics parameters of all anCas and SpCas9 determined from the single exponential decay curves in Figure 2c and 2e.

Kinetics parameters	**SpCas9**		**FCA anCas**		**BCA anCas**		**SCA anCas**		**PCA anCas**		**PDCA anCas**	
	Total	DSB	Total	DSB	Total	DSB	Total	DSB	Total	DSB	Total	DSB
k_{cleave} (min^{-1})	0.41 ± 0.03	0.28 ± 0.03	0.21 ± 0.03	0.008 ± 0.001	0.341 ± 0.009	0.021 ± 0.002	0.52 ± 0.02	0.16 ± 0.01	0.784 ± 0.003	0.56 ± 0.07	0.47 ± 0.02	0.28 ± 0.03
Max fraction	1	1	1	0.57	1	0.602	1	1	0.9994	0.92	1	1
R^2	0.9981	0.9915	0.9799	0.9983	0.9997	0.9968	0.9998	0.9901	0.9999	0.9983	0.9995	0.9926

Table 2. Kinetics parameters of FCA and BCA anCas and SpCas9 determined from the single exponential decay curves in Figure 4e and 4f.

Kinetics parameters	**SpCas9**		**FCA anCas**		**BCA anCas**	
	ssRNA	ssDNA	ssRNA	ssDNA	ssRNA	ssDNA
k_{cleave} (min^{-1})	0.036 ± 0.005	0.001 ± 0.003	0.003 ± 0.01	0.047 ± 0.005	0.03 ± 0.01	0.053 ± 0.009
Max fraction	0.43	0.37	0.69	0.60	0.97	0.97
R^2	0.9977	0.9989	0.9891	0.9978	0.9906	0.9951

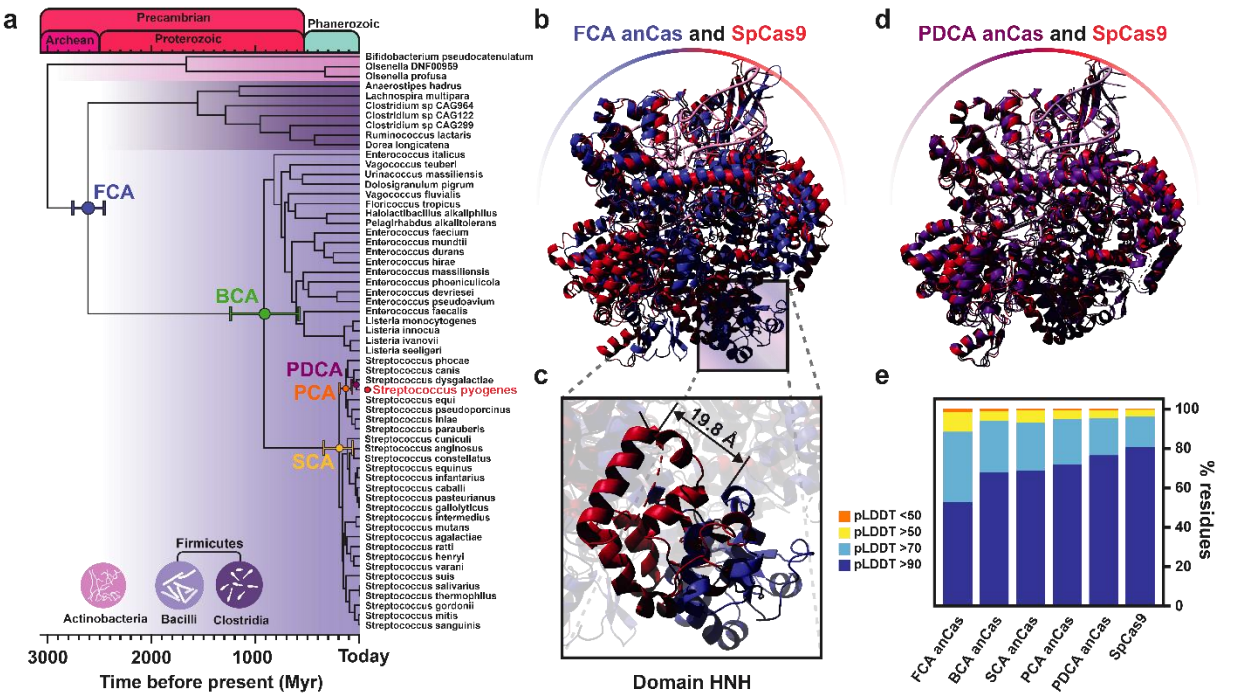


Figure 1. Phylogenetic and structural analysis of anCas endonucleases. (a) Phylogenetic chronogram of Cas9 endonucleases. Fifty-nine sequences were chosen from two phyla, Firmicutes and Actinobacteria, with two classes, Bacilli and Clostridia, belonging to Firmicutes. Identification codes of all sequence can be found in the Supplementary Information. Divergence times were estimated using Bayesian inference and information from the Time Tree of Life. Internal nodes from Firmicutes Common Ancestor (FCA), Bacilli Common Ancestor (BCA), Streptococci Common Ancestor (SCA), Pyogenic Common Ancestor (PCA) and Pyogenes-Dysgalactiae Common Ancestor (PDCA) were selected for testing. Node height error bars are indicated per each selected node. **(b)** Superposition of structural prediction of FCA anCas using AlphaFold2 (blue) with x-Ray structure of SpCas9 with guide RNA and target DNA (PDB:4oo8, red). **(c)** Isolated HNH domains of SpCas9 (red) and FCA anCas (blue) from whole structure coordinates. **(d)** Superposition of structural prediction of PDCA anCas with guide RNA and target DNA using AlphaFold2 (purple) with x-Ray structure of SpCas9 (PDB:4oo8, red). **(e)** pLDDT values for all anCas as estimated from AlphaFold2 prediction.

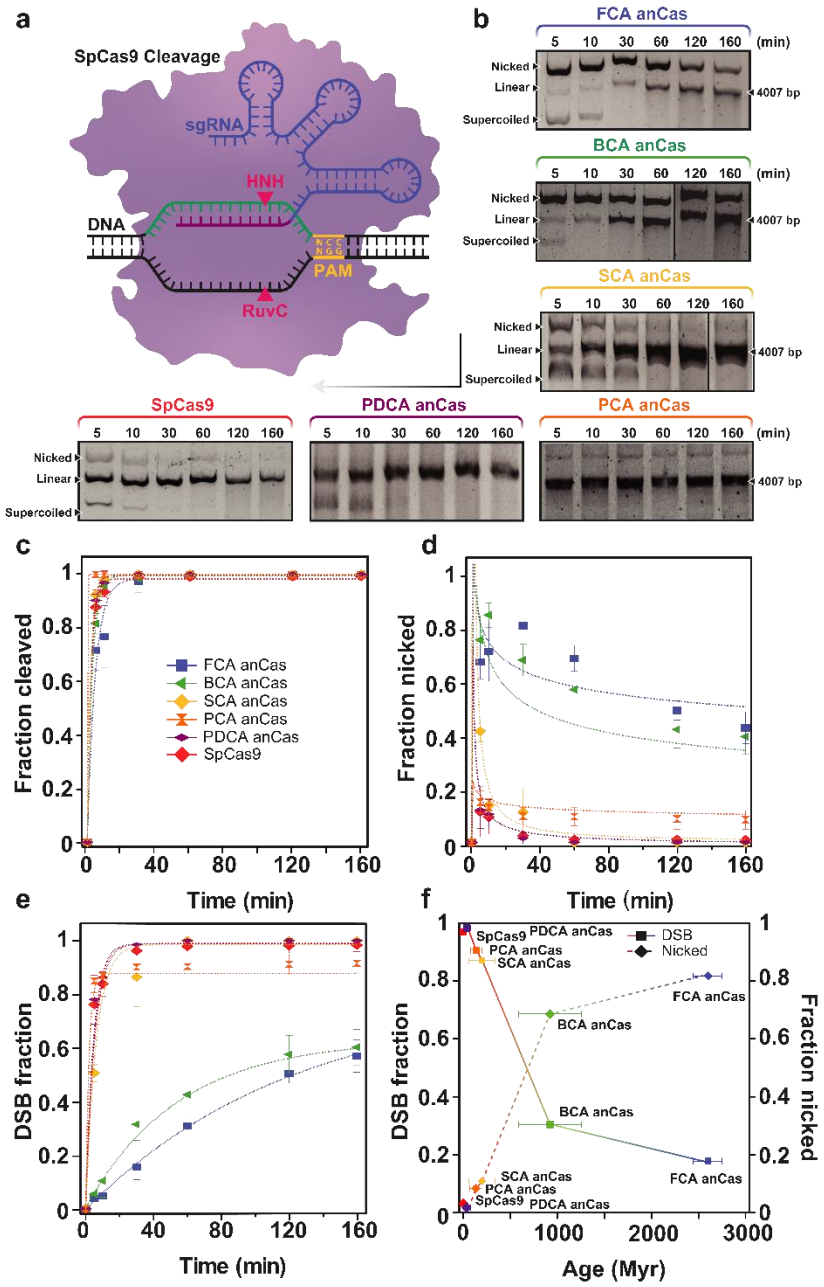


Figure 2. Activity of anCas endonucleases on a supercoiled DNA substrate. (a) Schematic representation of endonuclease activity on a supercoiled substrate. **(b)** *In vitro* cleavage assay for SpCas9 and all anCas on a 4007 bp substrate at different reaction times showing nicked and linear fractions. Black arrow top-down and right-left represents the chronological order. **(c)** Quantification of total cleavage at different reaction times and exponential fits (lines). **(d)** Quantification of nicked fraction for all anCas and SpCas9 at different times. **(e)** Quantification of DSB cleavage. Single-exponential fits were used to obtain k_{cleave} and maximum fraction cleaved (amplitude). Fitting parameters are summarized in Table 1. Values reported as mean ± SD, where n = 2. **(f)** DSB fraction (left axis) and nicked fraction (right axis) at 30 minutes reaction plotted against evolutionary time. Horizontal error bars represent the node height error per each anCas form.

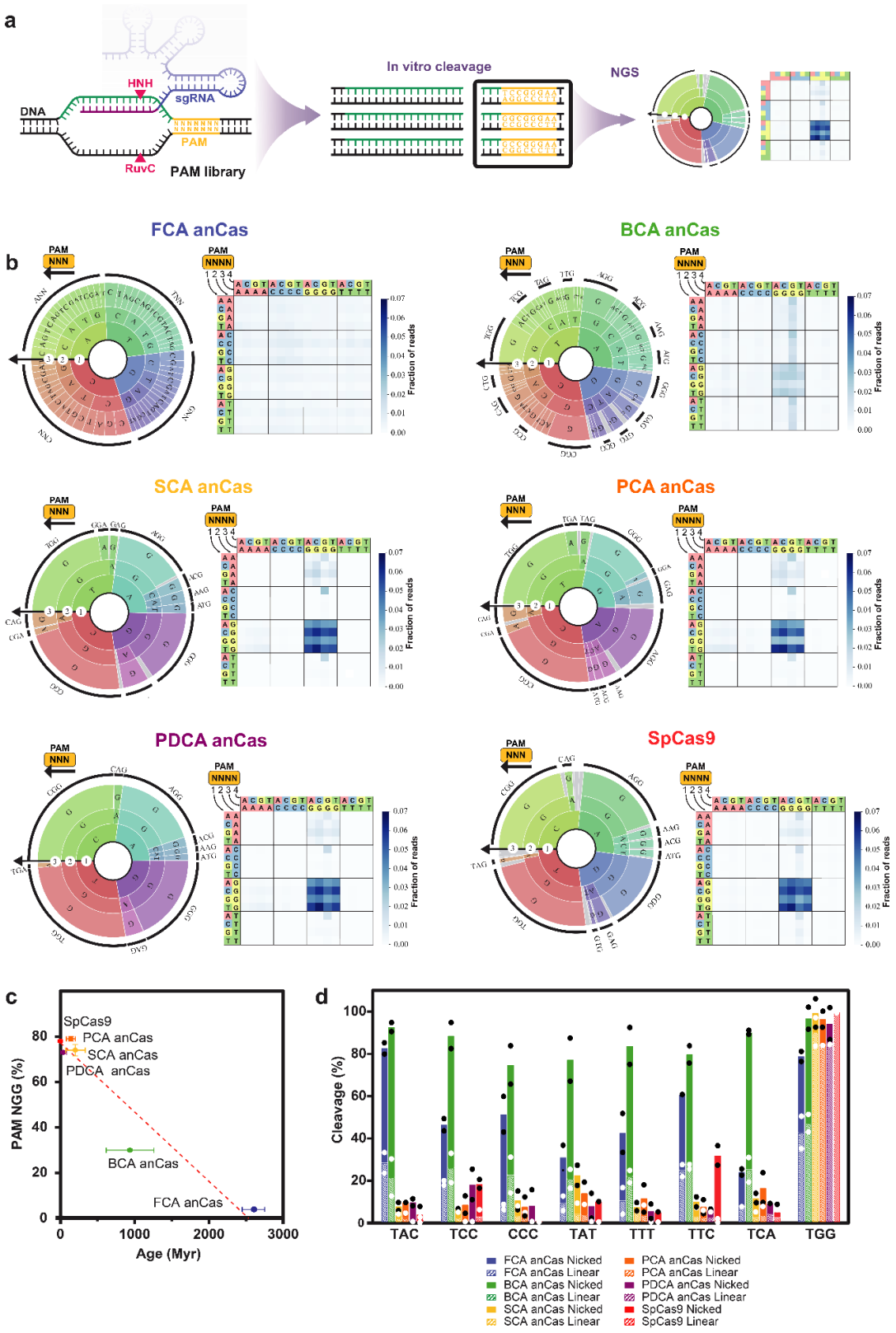


Figure 3. PAM determination of anCas. (a) Scheme of *in vitro* determination of PAM preference using a substrate library encoding a random 7-nt PAMs using Next-generation

Sequencing. Two fragments of 566 and 278 bp are generated after target cleavage. **(b)** PAM wheels (Krona plots) with 3nt PAM and heatmaps with 4nt PAM for all five anCas and SpCas9, used as control. **(c)** Percentage of reads containing an NGG PAMs 3-4 bp downstream from the cleavage position plotted against evolutionary time. Values reported as mean ± SD, where n = 2. Horizontal error bars represent the node height error per each anCas form **(d)** *In vitro* cleavage assay (DSB and nicked products) using a variety of PAMs represented by TNN and CCC as control. Incubation time was 10 min. Bars represent the average value of two independent experiments indicated by the black (nicked) and white (linear) dots.

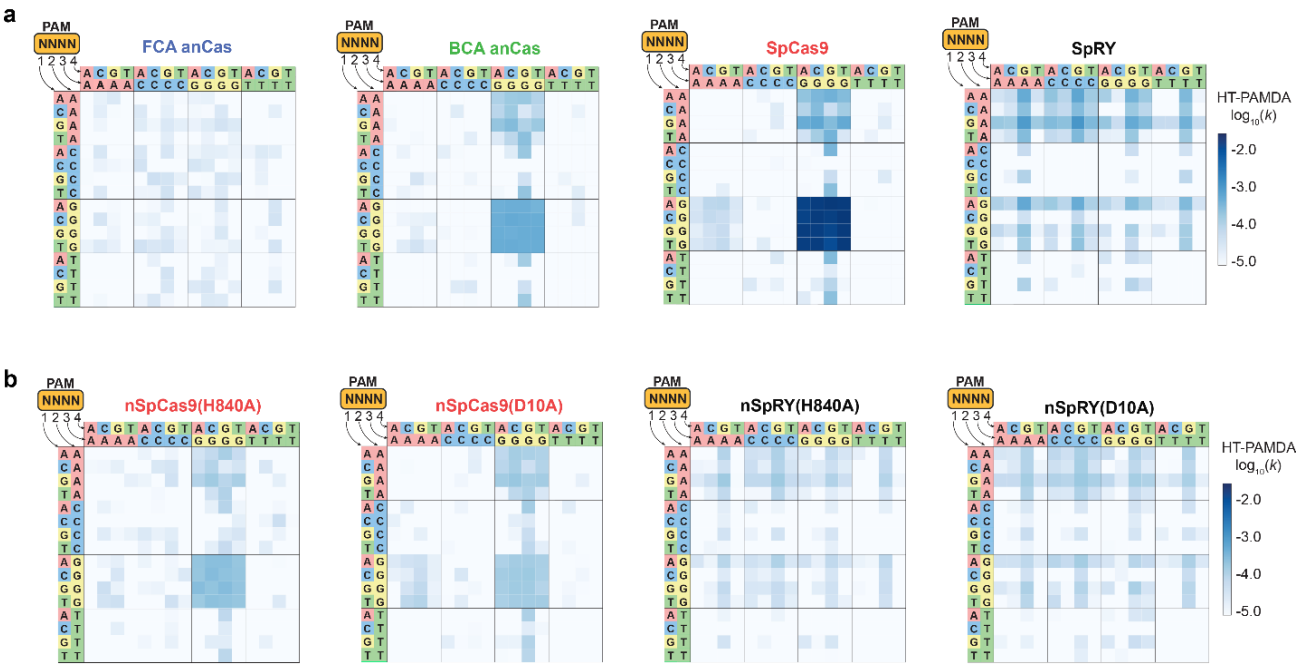


Figure 4. HT-PAMDA assay. **(a)** PAM profiles of anCas, SpCas9 and SpRY proteins as determined by HT-PAMDA. Rate constants corresponding to Cas cleavage activity are illustrated as $\log_{10}$ values and are the mean of cleavage reactions against two unique spacer sequences. **(b)** PAM profiles of contemporary Cas9 and variant SpRY proteins with inactivating mutations in the HNH (H840A) or RuvC (D10A) domain result in a single stranded DNA nickase enzyme with attenuated rate constant values relative to the double-stranded nuclease.

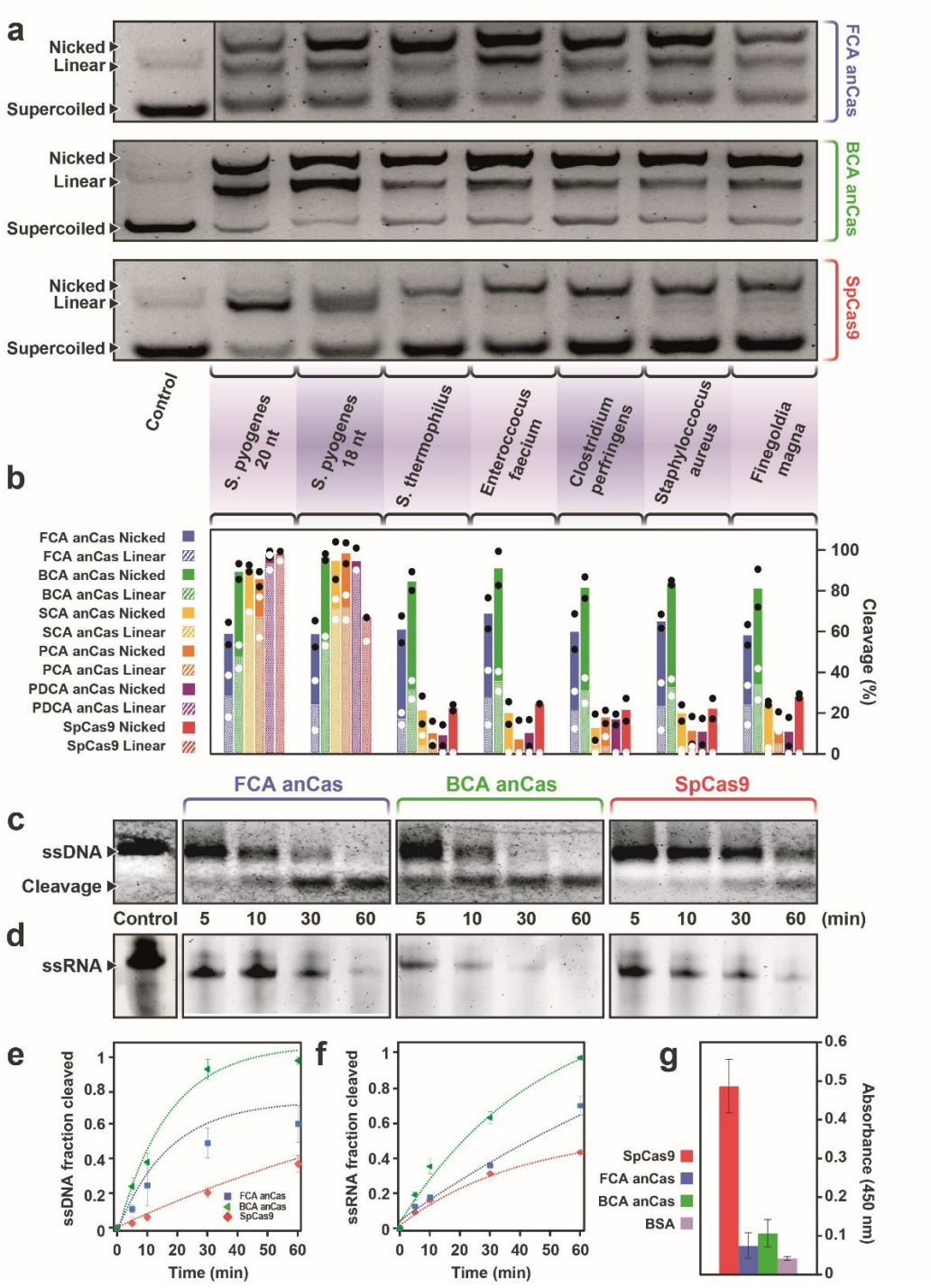


Figure 5. sgRNA test and nuclease activity of anCas on single-stranded substrates. (a) *In vitro* cleavage assay on a supercoiled DNA substrate of anCas and SpCas9 using sgRNAs from different species. FCA, BCA anCas and SpCas9 are shown. **(b)** Quantification of *in vitro* cleavage for all anCas and SpCas9 using the different sgRNAs. Bars represent the average value of two independent experiments indicated by the black (nicked) and white (linear) dots. **(c)** *In vitro* cleavage assay on an 85 nt ssDNA fragment at different incubation times for FCA, BCA anCas and SpCas9. **(d)** *In vitro* cleavage assay on a 60 nt ssRNA at different incubation times for

FCA, BCA anCas and SpCas9. In both (c) and (d), the control lane is the same for the three proteins. **(e)** Quantification of fraction cleavage of ssDNA at different times and exponential fits for determination of kinetics parameters. **(f)** Quantification of fraction cleavage of ssRNA at different times and exponential fits for determination of kinetics parameters. All kinetics parameters are summarized in Table 2. Values reported as mean ± SD, where n = 2. **(g)** Results from ELISA test of Anti-Cas9 rabbit antibody against SpCas9, FCA anCas, BCA anCas and BSA, used as control. Results are reported as average, and S.D. calculated from three independent experiments.

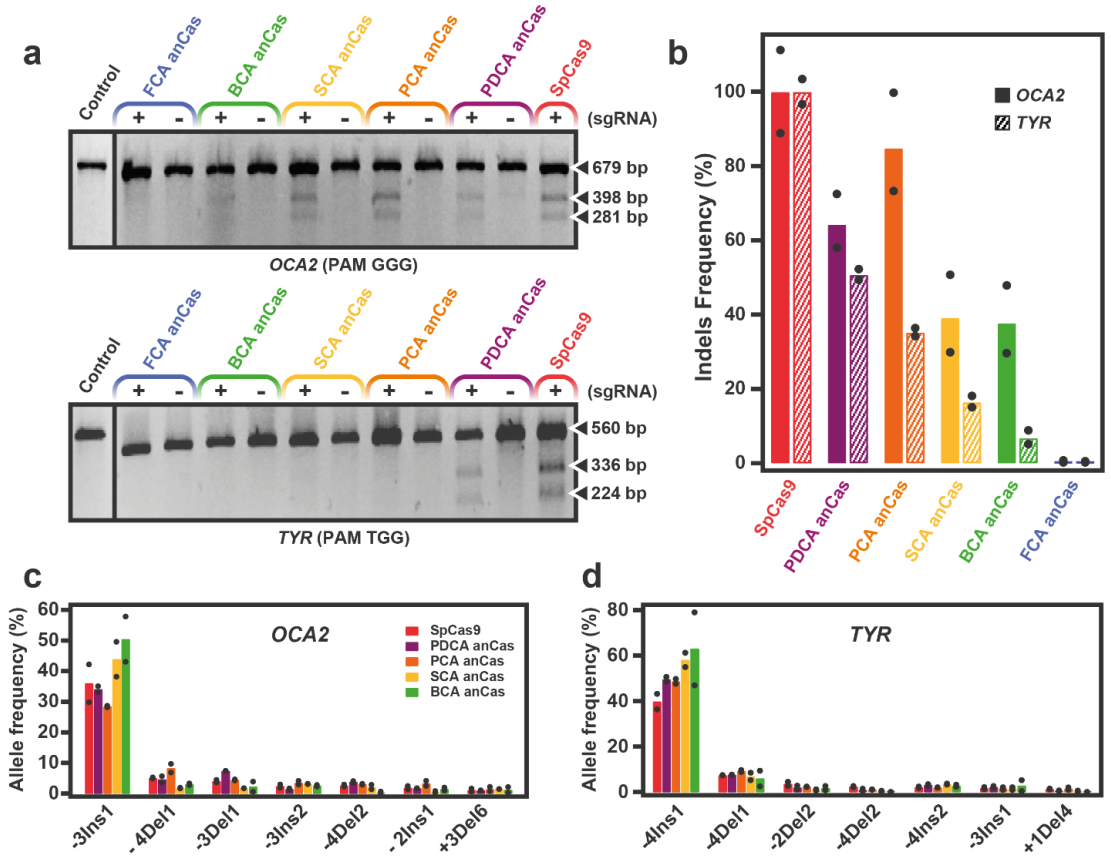
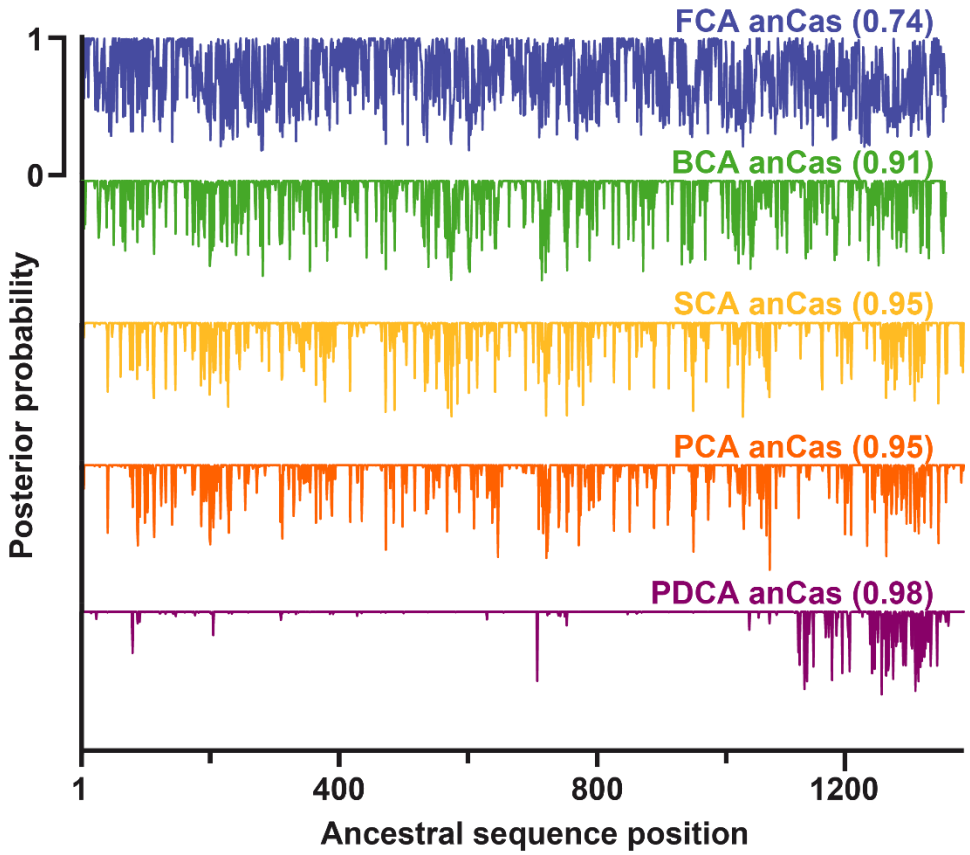
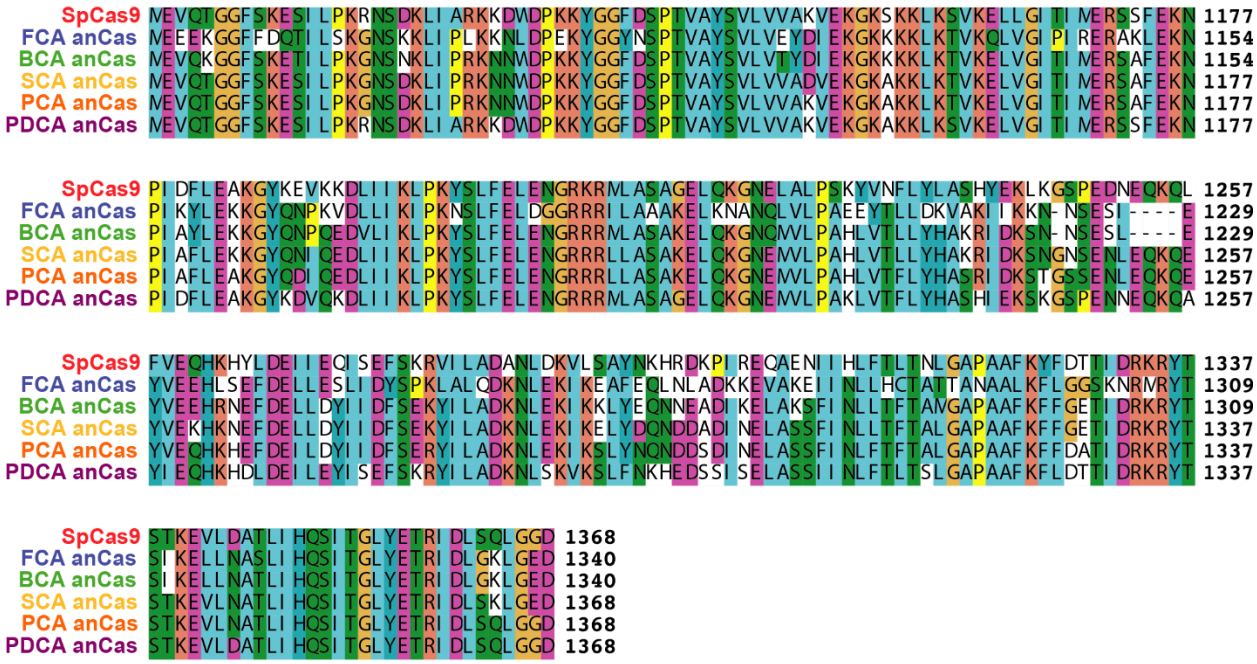


Figure 6. Activity of anCas endonucleases in HEK293T human cells. (a) T7 endonuclease mismatch assay for *OCA2* and *TYR* genes. Expected fragments for *OCA2* were 398 and 281 bp for a 679 bp amplified target, and 336 and 224 bp for *TYR* for a total 560 bp fragment. Experiments with (+) and without (-) sgRNA were run for each anCas. **(b)** Indels determination for *OCA2* and *TYR* by NGS of the experiments in (a) analyzed using Mosaic Finder. Indels frequency is normalized to SpCas9. **(c)** Allele frequency for *OCA2* and **(d)** *TYR*. Position of preferred cleavage was analyzed determining the preference, insertion (Ins) or deletion (Del) with position indicated upstream (-) or downstream (+) with respect to PAM. Length of the indel is indicated below the x axis (e.g. -3Ins1). The preferred allele for *OCA2* is one insertion at three nucleotides upstream the PAM (-3Ins1). In the case of *TYR*, we find one insertion at the fourth nucleotide upstream the PAM as the preferred allele (-4Ins1). Bars represent the average value of two independent experiments indicated by the black dots.

Extended data Figures



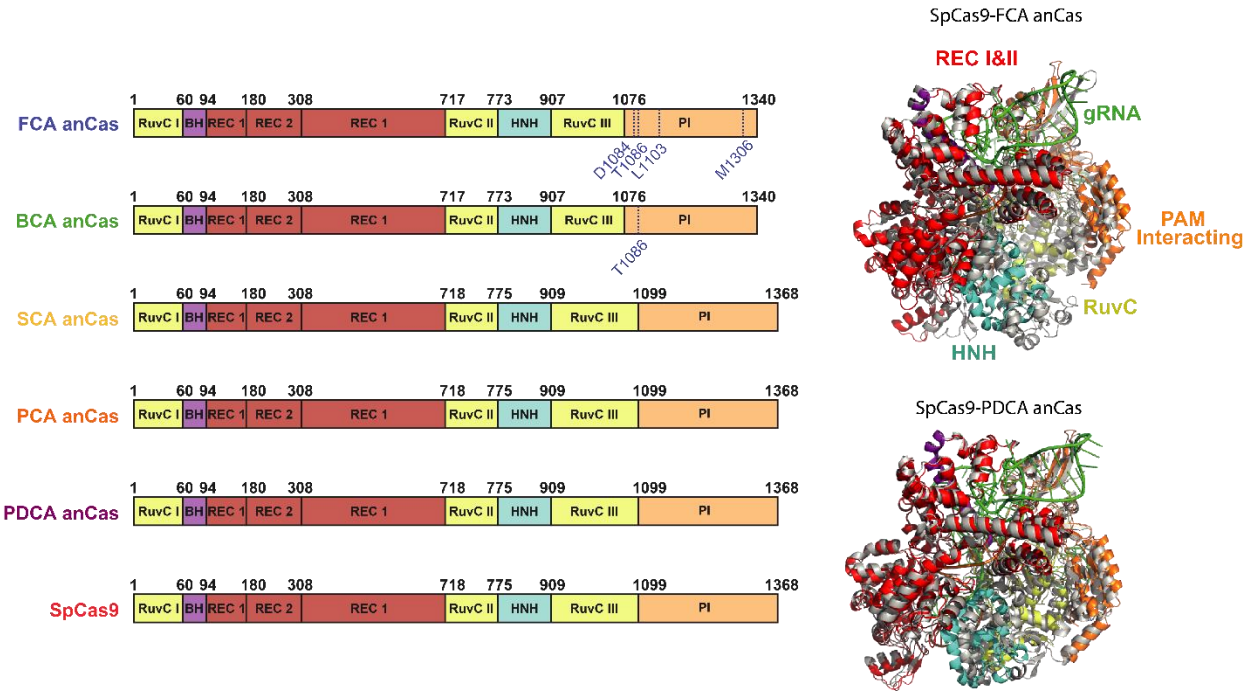
Extended Data Fig. 1. Posterior probability distribution for each inferred residue of all ancestral anCas endonucleases. The residue with the highest posterior probability is assigned at each position. The posterior probability average of each form is indicated in brackets. In all cases, posterior probability average is close to 1 except for FCA anCas which shows an average value of 0.74.



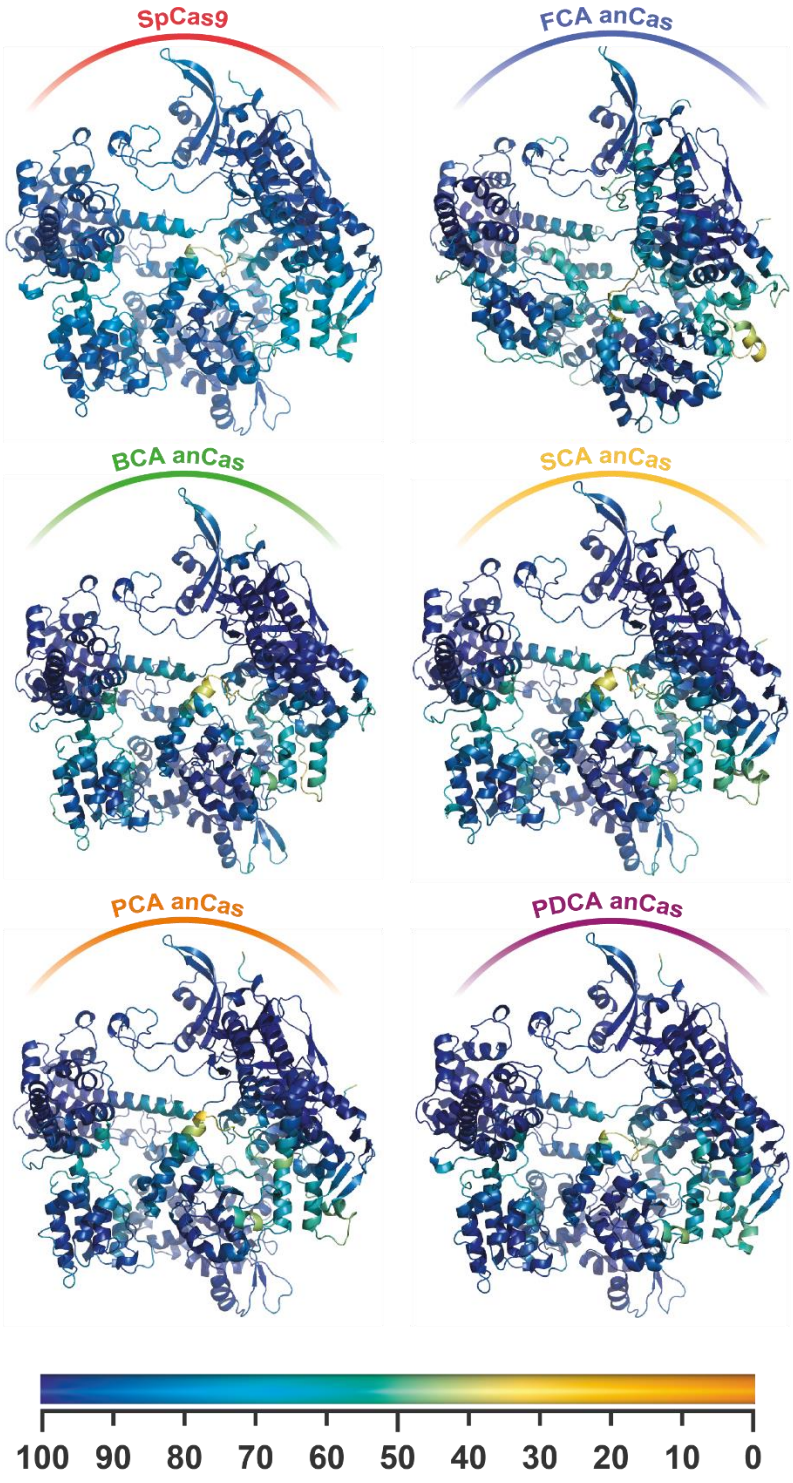
Amino acid sequence identity % of anCas with respect to the query sequence (SpCas9)					
	FCA anCas	BCA anCas	SCA anCas	PCA anCas	PDCA anCas
% Identity	54 %	70 %	75 %	83 %	96 %

Extended Data Fig. 2. Alignment of the amino acid sequences from PI domain of anCas and SpCas9. Percentage of identity of the different anCas sequences with respect to SpCas9.

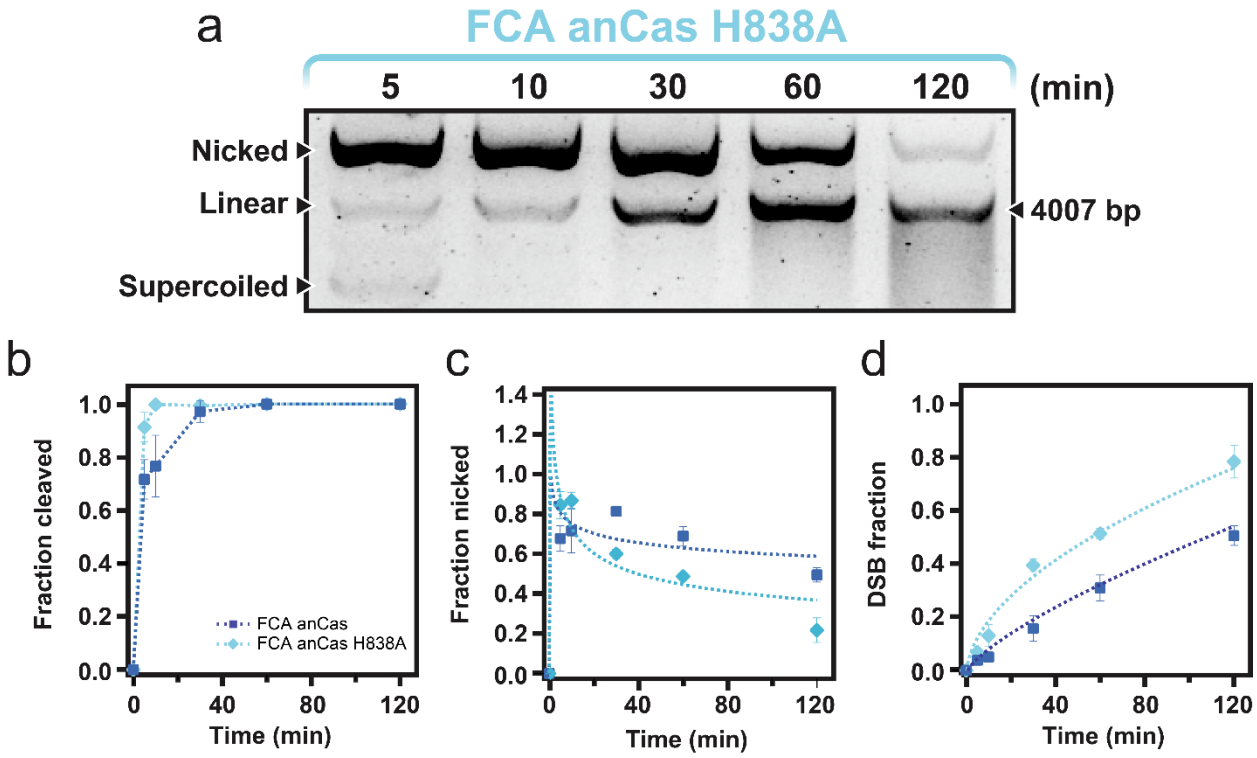
Main residues involved in SpCas9 activity and their correspondence with anCas sequences. Mutations are highlighted in blue.					
SpCas9	**FCA anCas**	**BCA anCas**	**SCA anCas**	**PCA anCas**	**PDCA anCas**
D10	D10	D10	D10	D10	D10
S15	S15	S15	S15	S15	S15
R66	R66	R66	R66	R66	R66
R70	R70	R70	R70	R70	R70
R74	R74	R74	R74	R74	R74
R78	R78	R78	R78	R78	R78
P475	P475	P475	P475	P475	P475
W476	W476	W476	W476	W476	W476
N477	N477	N477	N477	N477	N477
H840	H838	H838	H840	H840	H840
K1107	D1084	K1084	K1107	K1107	K1107
S1109	T1086	T1086	S1109	S1109	S1109
W1126	L1103	W1103	W1126	W1126	W1126
R1333	R1305	R1305	R1333	R1333	R1333
K1334	M1306	K1306	K1334	K1334	K1334
R1335	R1307	R1307	R1335	R1335	R1335



Extended Data Fig. 3. List of important mutations and domain organization of anCas compared to SpCas9. Mutations of the main residues involved in PAM recognition are marked in blue. Bottom figure depicts domain organization and structural alignment of SpCas9-FCA and SpCas9 PDCA anCas. Ancestral anCas are grey colored and SpCas9 colored by domains.



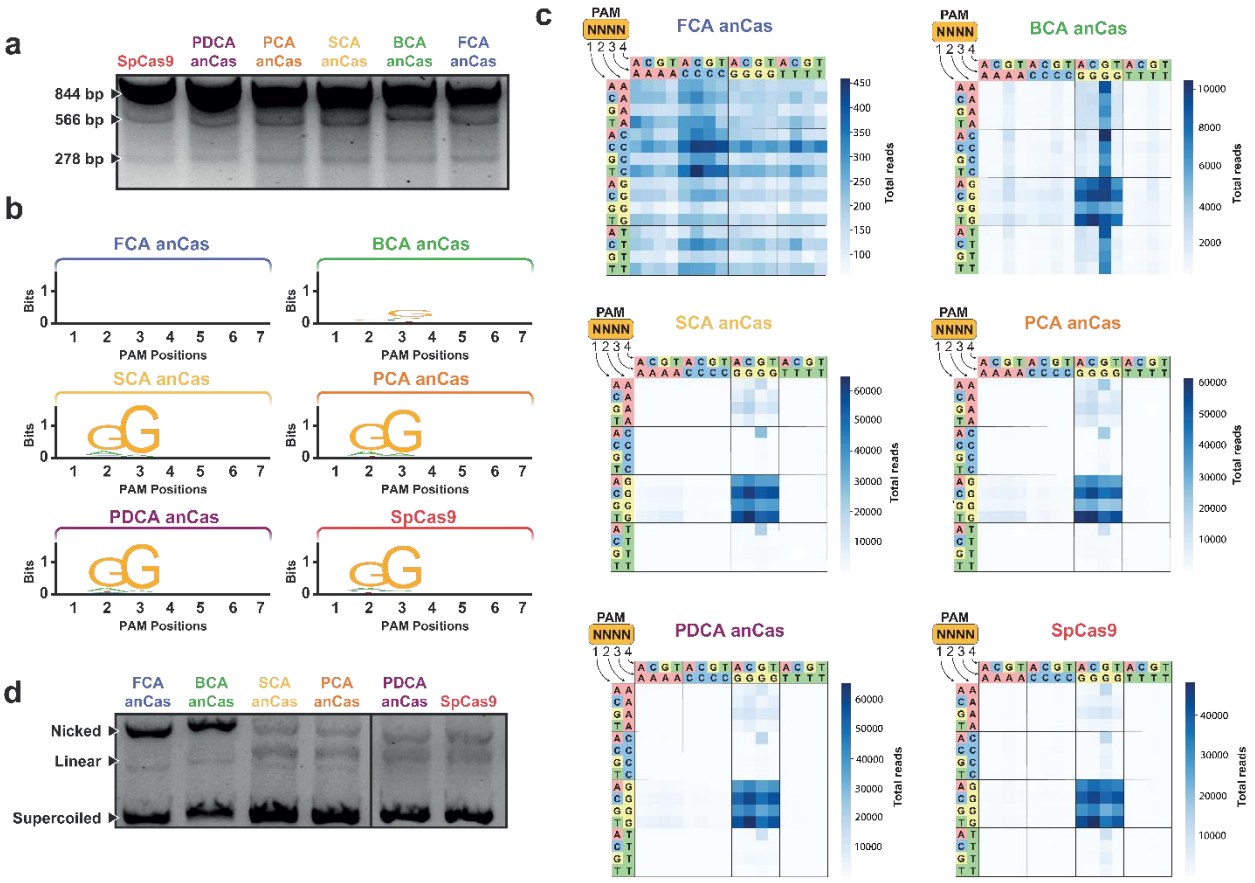
Extended Data Fig. 4. Structural predictions of anCas and SpCas9 by AlphaFold2. Structures are colored by pLDDT score according to the color bar.



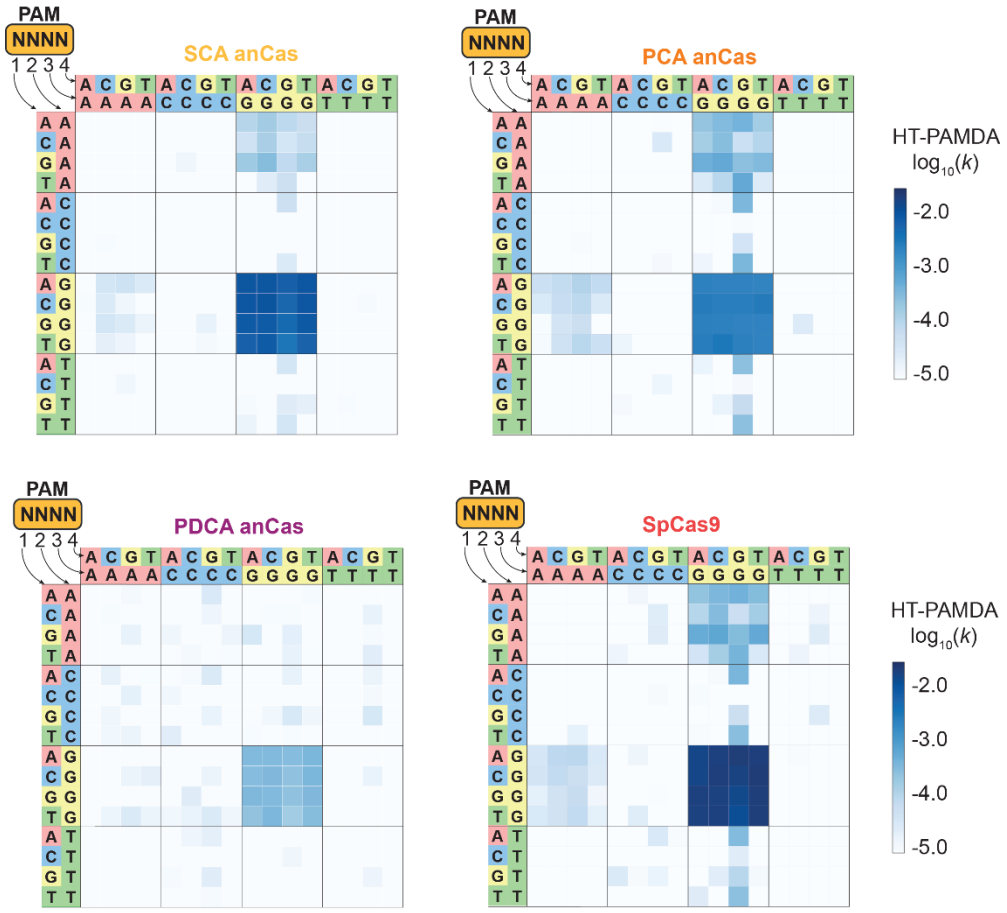
Kinetics parameters of the FCA anCas and FCA anCas H838A determined from the single exponential decay curves.

	FCA anCas		FCA anCas H838A	
Kinetics parameters	Total	DSB	Total	DSB
K_{cleave} (min^{-1})	0.22 ± 0.04	0.007 ± 0.001	0.33 ± 0.01	0.015 ± 0.004
Max fraction	1	0.51	1	0.78
R^2	0.9790	0.9985	0.9999	0.9899

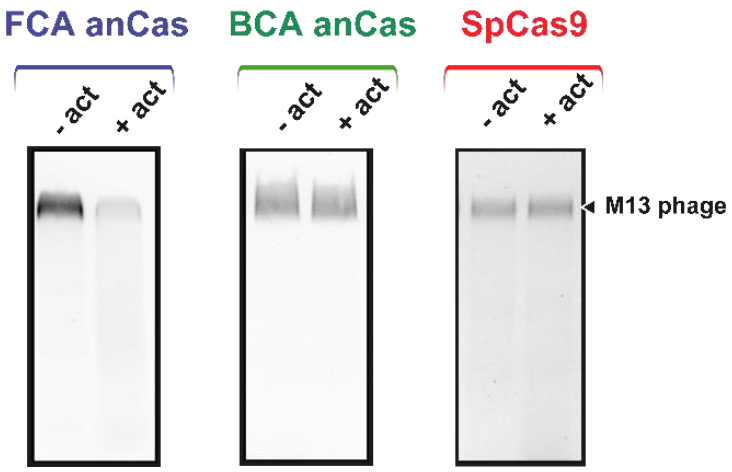
Extended Data Fig. 5. Activity of FCA anCas H838A endonuclease on a supercoiled DNA substrate (a) *In vitro* cleavage assay for anCas FCA H838A on a 4007 bp substrate at different reaction times showing nicked and linear fractions. **(b)** Quantification of total cleavage fraction at different reaction times and exponential fits (lines). **(c)** Quantification of fraction nicked at different times. **(d)** Quantification of DSB cleavage. Single-exponential fits were used to obtain k_{cleave} and maximum fraction cleaved (amplitude). Values reported as mean ± SD, where n = 2.



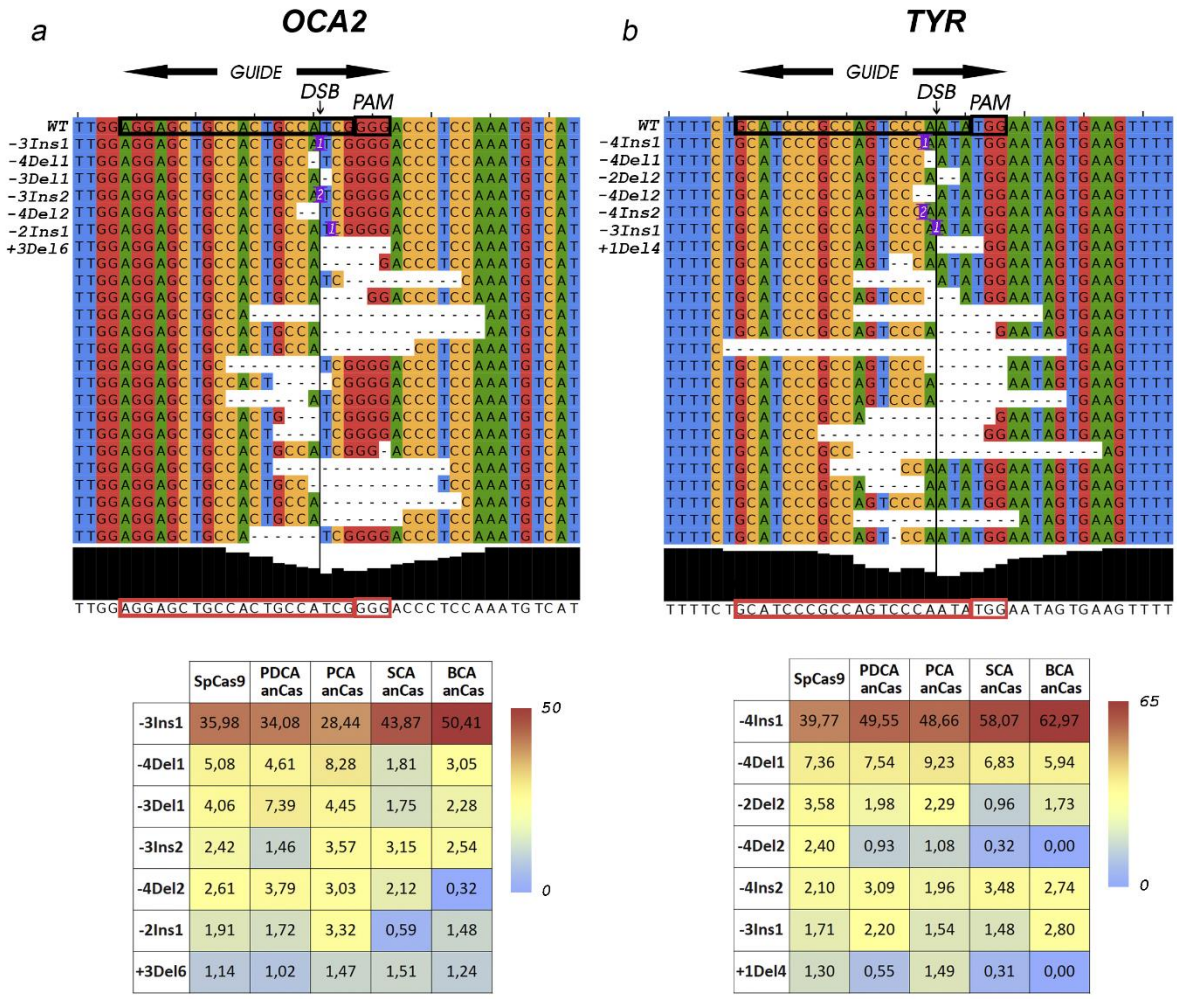
Extended Data Fig. 6. PAM determination of anCas enzymes. (a) Example of *in vitro* cleavage assay to obtain 278 bp fragment for NGS analysis. **(b)** Weblogo of the different PAM recognized by anCas and SpCas9. **(c)** Heatmaps illustrating the total reads for each of the possible 256 NNNN PAMs, analyzed from NGS of the of 278 bp cleaved DNA fragments from panel a. **(d)** *In vitro* cleavage assay using the PAM sequence TCC.



Extended Data Fig. 7. HT-PAMDA-determined PAM profiles of Cas enzymes. PAM profiles of anCas enzymes and SpCas9 as determined by HT-PAMDA. Rate constants corresponding to Cas cleavage activity on each of the 256 NNNN PAMs are illustrated as mean $\log_{10}$ values of cleavage reactions against two unique spacer sequences. For comparison, the SpCas9 is re-plotted from **Fig. 4**.



Extended Data Fig. 8. Trans-activity of FCA anCas, BCA anCas and SpCas on M13 phage ssDNA. Nonspecific M13 ssDNA cleavage with sgRNA and complementary (or not) 85nt ssDNA activator with no sequence homology to M13 circular ssDNA. FCA anCas can cleave the ssDNA substrate in the presence of the activator, whereas BCA anCAs and SpCas9 do not cleave the same substrate.



	SpCas9	PDCA anCas	PCA anCas	SCA anCas	BCA anCas
-3Ins1	35,98	34,08	28,44	43,87	50,41
-4Del1	5,08	4,61	8,28	1,81	3,05
-3Del1	4,06	7,39	4,45	1,75	2,28
-3Ins2	2,42	1,46	3,57	3,15	2,54
-4Del2	2,61	3,79	3,03	2,12	0,32
-2Ins1	1,91	1,72	3,32	0,59	1,48
+3Del6	1,14	1,02	1,47	1,51	1,24

50

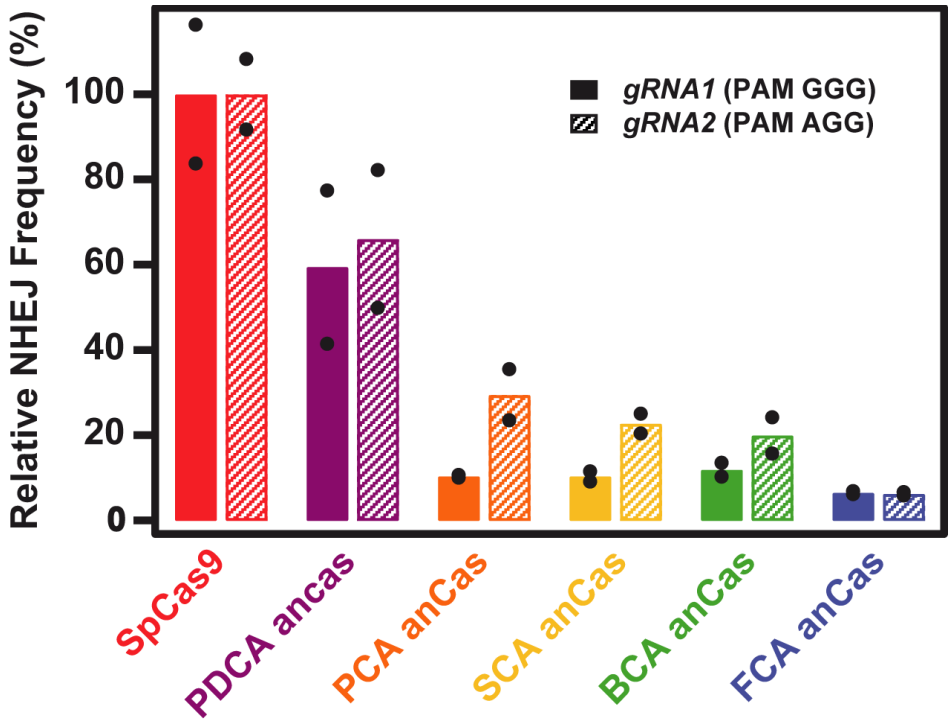
0

	SpCas9	PDCA anCas	PCA anCas	SCA anCas	BCA anCas
-4Ins1	39,77	49,55	48,66	58,07	62,97
-4Del1	7,36	7,54	9,23	6,83	5,94
-2Del2	3,58	1,98	2,29	0,96	1,73
-4Del2	2,40	0,93	1,08	0,32	0,00
-4Ins2	2,10	3,09	1,96	3,48	2,74
-3Ins1	1,71	2,20	1,54	1,48	2,80
+1Del4	1,30	0,55	1,49	0,31	0,00

65

0

Extended Data Fig. 9. Analysis of the *in vivo* activity of anCas variants. Alignments generated by Jalview program of the wild-type and the most frequent edited alleles (indels) detected by Mosaic Finder in **(a)** *OCA2* and **(b)** *TYR* genes after NHEJ cell repair in HEK 293T cells. Heatmaps are shown underneath the alignments highlighting the frequencies of the top-7 most frequent alleles generated after cleavage and repair with SpCas9, PDCA, PCA, SCA and BCA anCas, once normalized with respect to the total number of indels for each Cas. The guide, the PAM and the DSB theoretical site are marked in the figure. For the mutation nomenclature of each allele we consider the first nucleotide of the PAM as +1. Numbers within the allele sequences represent the length of insertions or deletion in the exact location indicated by the first figure. Example: *-4Ins1*, insertion of 1 nucleotide four bases upstream the PAM.



Extended Data Fig. 10. Traffic Light Reporter cleavage assay targeting gene *TLR*. The relative NHEJ frequency is estimated by the number of RFP-positive cells and is normalized to SpCas9. Bars represent the average value of two independent experiments indicated by the black dots.

Supplemental Information for

Evolution of CRISPR-associated Endonucleases as Inferred from Resurrected Proteins

Borja Alonso-Lerma[1,16], Ylenia Jabalera[1,16], Sara Samperio[1], Matias Morin[2], Almudena Fernandez[3], Logan T. Hille[4,5], Rachel A. Silverstein [4,5], Ane Quesada-Ganuza[1], Antonio Reifs[1], Sergio Fernández-Peñalver[2], Yolanda Benitez[3,6] Lucia Soletto[2], Jose A Gavira[7], Adrian Diaz[8,9], Wim Vranken[8,9,10], Avencia Sanchez-Mejias[11], Marc Güell[11,12], Francisco JM Mojica[13], Benjamin P. Kleinstiver[4,14], Miguel A Moreno-Pelayo[2], Lluis Montoliu[3] & Raul Perez-Jimenez[1,15*]

[1]CIC nanoGUNE BRTA; San Sebastian, Spain.

[2]Servicio de Genética, Hospital Universitario Ramón y Cajal, IRYCIS and Centro de Investigaciones Biomédicas en Red de Enfermedades Raras (CIBERER); Madrid, Spain.

[3]Department of Molecular and Cellular Biology, National Centre for Biotechnology (CNB-CSIC) and Centre for Biomedical Network Research on Rare Diseases (CIBERER-ISCIII); Madrid, Spain.

[4]Center for Genomic Medicine and Department of Pathology, Massachusetts General Hospital; Boston, MA, 02114, USA.

[5]PhD Program in Biological and Biomedical Sciences, Harvard University; Boston, MA, 02115, USA.

[6]INGEMM, Hospital Universitario La Paz, CIBERER-ISCIII; Madrid, Spain.

[7]Laboratorio de Estudios Cristalográficos, IACT (CSIC-UGR); Armilla, Granada, Spain.

[8]Interuniversity Institute of Bioinformatics in Brussels, ULB-VUB; Brussels 1050, Belgium.

[9]Structural Biology Brussels, Vrije Universiteit Brussel; Brussels 1050, Belgium.

[10]Structural Biology Research Centre, VIB; Brussels 1050, Belgium.

[11]Integra Therapeutics S.L.; Barcelona, Spain.

[12]Department of Medicine and Life Sciences, Universitat Pompeu Fabra; Barcelona, Spain.

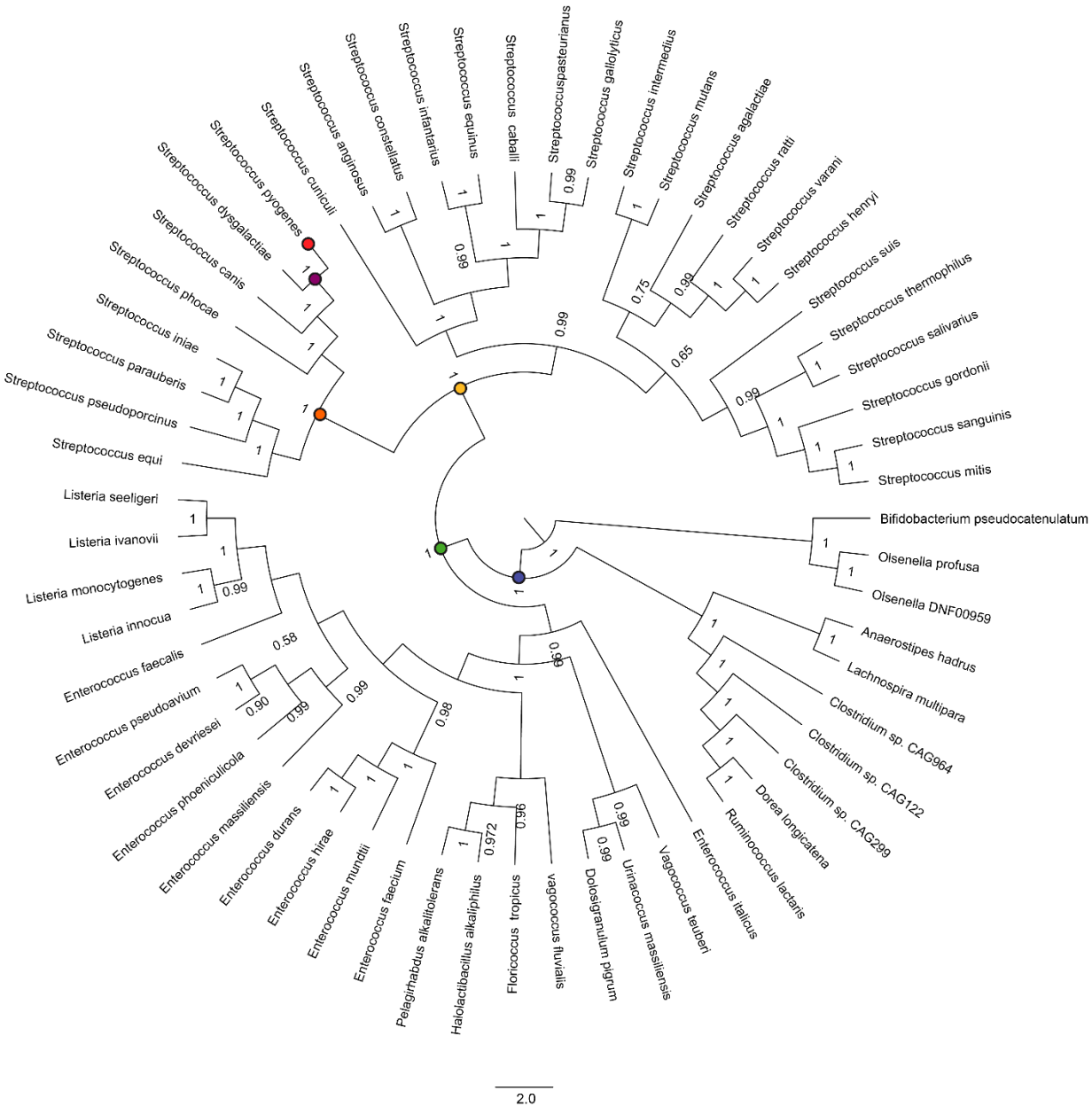
[13]Dpto. Fisiología, Genética y Microbiología and Instituto Multidisciplinar para el Estudio del Medio "Ramón Margalef", Universidad de Alicante; Alicante, Spain.

[14]Department of Pathology, Harvard Medical School; Boston, MA, 02115, USA.

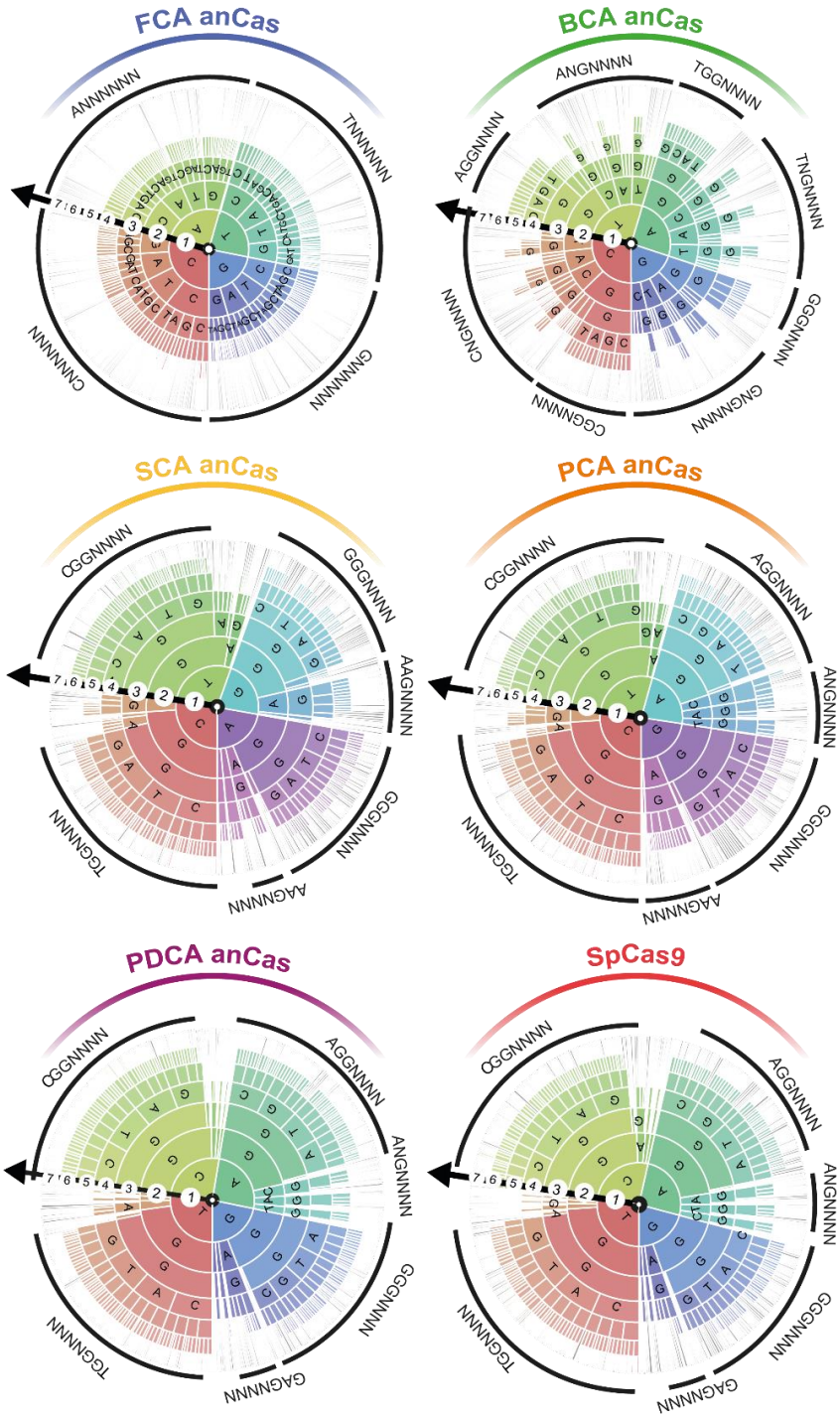
[15]Ikerbasque Foundation for Science; Bilbao, Spain.

[16]These authors contributed equally to the work

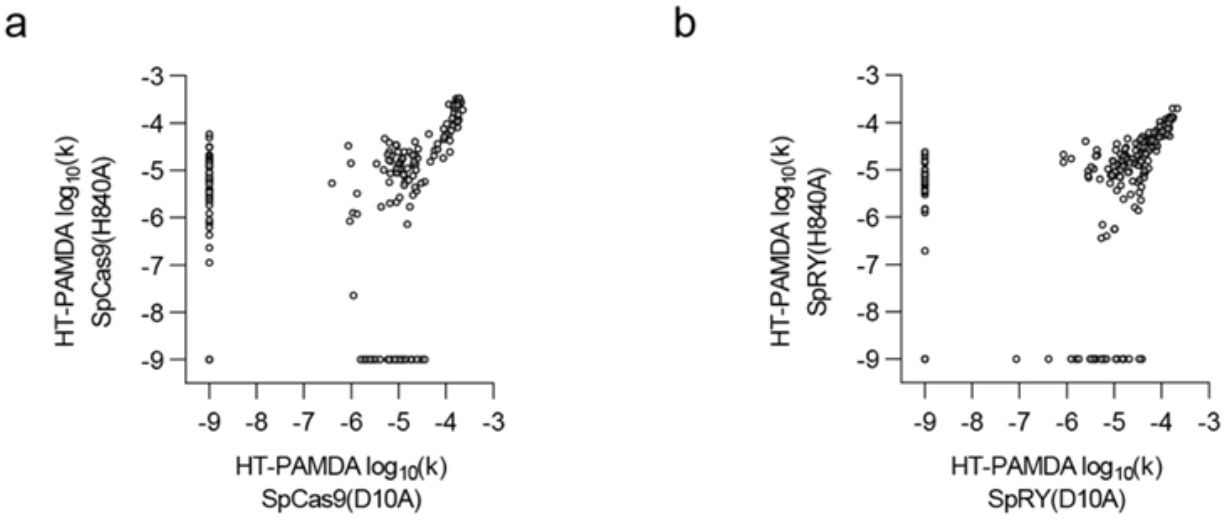
*Corresponding author. Email: r.perezjimenez@nanogune.eu (R.P.-J.)



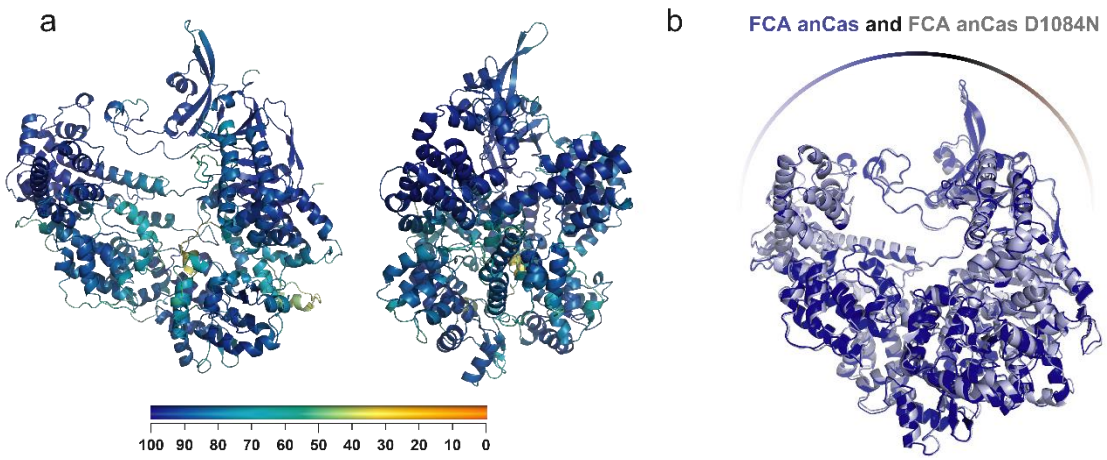
Supplementary figure 1. Phylogeny tree of the fifty-nine species utilized in this study. Posterior probabilities per each node are indicated. Selected nodes for ancestral sequence reconstruction are marked with colored circles.



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.

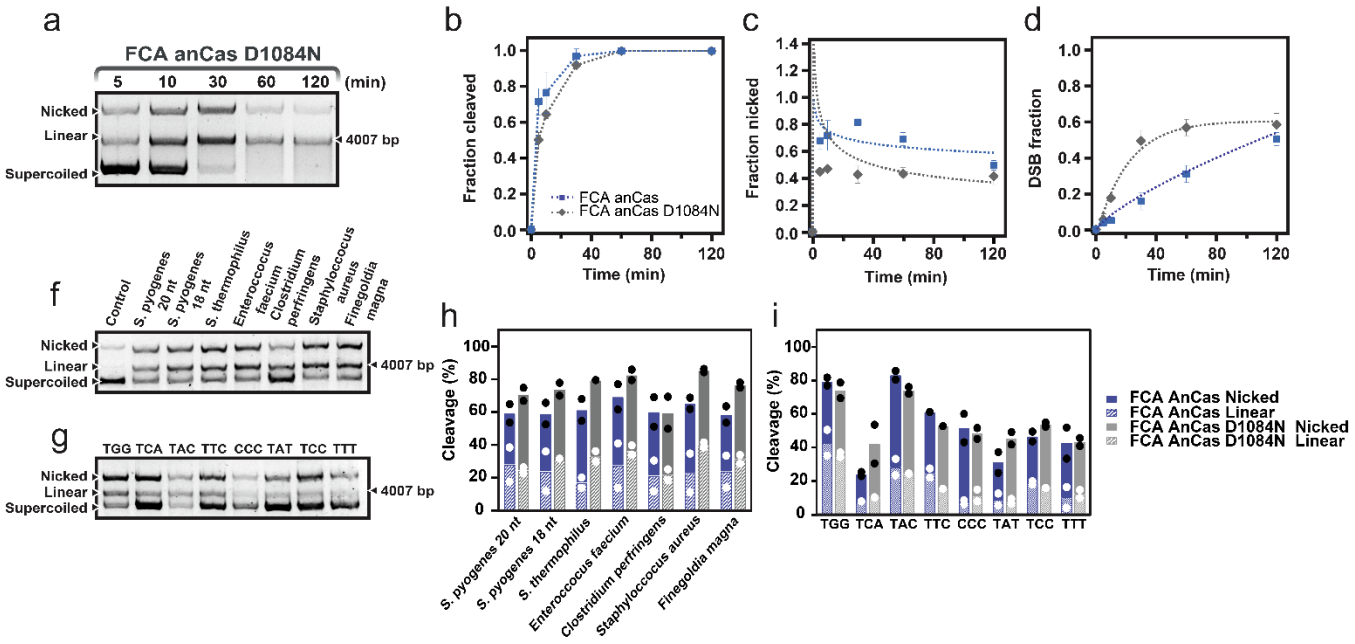


Supplementary figure 3. HT-PAMDA rate constants from SpCas9 and SpRY nickases. **(a)** Correlation of HT-PAMDA $\log_{10}$ rate constants (*k*s) for nSpCas9(H840A) and **n**SpCas9(D10A) for each of the possible 256 NNNN PAMs, where the *k*s are the average of two HT-PAMDA experiments performed using 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). **(c)** Quantification of fraction nicked at different times. **(d)** Quantification of DSB cleavage. Single-exponential fits were used to obtain k_{cleave} and maximum fraction cleaved (amplitude). Values reported as mean ± SD, 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. **(f)** *In vitro* cleavage assay on a supercoiled DNA substrate of FCA anCas D1084N using sgRNAs from different species. **(g)** *In vitro* cleavage assay using different PAM sequences. **(h)** Quantification of *in vitro* cleavage using different sgRNAs of FCA anCas and FCA anCas D1084N. **(i)** 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.

Table S1. 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

Table S2. 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

Table S3. 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

Table S4. 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.*

Table S5. 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 gtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggact ataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgc ctttctcccttcgggaagcgtggcgctttctcaaagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctg ggctgtgtgcacgaacccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagac acgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaa gtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagt tggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaa aggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatg agattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggt ctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcg tgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctc cagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtc tattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtgg tgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaa aaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactg cataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgc ggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattgga aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactga tcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggc gacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatat ttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccatta ttatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc

Table S6. Primers for PAM determination assay (NGS analysis).

Name	Sequence
FW library	aataggcgtatcacgaggc
RV library	agcgagtcagtgagcgag

Table S7. *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

Table S8. 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

Table S9. 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

Table S10. 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

Table S11. 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

Table S12. 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

Table S13. Sequences of ssDNA and ssRNA (Figure 5). Target sequences shown in blue.

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

Table S14. 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)

Table S15. 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'

Table S16. 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