JBC Papers in Press. Published on March 11, 2020 as Manuscript RA120.012933
The latest version is at https://www.jbc.org/cgi/doi/10.1074/jbc.RA120.012933

# CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects

Karthik Murugan [1, 2], Arun S. Seetharam [3], Andrew J. Severin [3] and Dipali G. Sashital [1, 2 *]

From the [1] Roy J. Carver Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, IA 50011, USA; [2] Molecular, Cellular, and Developmental Biology Interdepartmental Program, Iowa State University, Ames, IA 50011, USA; [3] Genome Informatics Facility, Office of Biotechnology, Iowa State University, Ames, IA 50011, USA

Running title: *Cas12a nickase activities*

*To whom correspondence should be addressed: Dipali G. Sashital: Roy J. Carver Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, IA 50011, USA; Email: sashital@iastate.edu; Tel: +1 (515)-294-5121; Fax: +1 (515)-294-7629

Keywords: CRISPR-Cas12a, specificity, cleavage activity, nickase, FnCas12a, LbCas12a, AsCas12a, bacterial immunity, mismatch tolerance, genome editing

## ABSTRACT

Cas12a (Cpf1) is an RNA-guided endonuclease in the bacterial type V-A CRISPR-Cas anti-phage immune system that can be repurposed for genome editing. Cas12a can bind and cut dsDNA targets with high specificity in vivo, making it an ideal candidate for expanding the arsenal of enzymes used in precise genome editing. However, this reported high specificity contradicts Cas12a's natural role as an immune effector against rapidly evolving phages. Here, we employed high-throughput in vitro cleavage assays to determine and compare the native cleavage specificities and activities of three different natural Cas12a orthologs (FnCas12a, LbCas12a, and AsCas12a). Surprisingly, we observed pervasive sequence-specific nicking of randomized target libraries, with strong nicking of DNA sequences containing up to four mismatches in the Cas12a-targeted DNA–RNA hybrid sequences. We also found that these nicking and cleavage activities depend on mismatch type and position and vary with Cas12a ortholog and CRISPR RNA (crRNA) sequence. Our analysis further revealed robust non-specific nicking of dsDNA when Cas12a is activated by binding to a target DNA. Together, our findings reveal that Cas12a has multiple nicking activities against dsDNA substrates and that these activities vary among different Cas12a orthologs.

Cas12a (formerly Cpf1) is an RNA-guided endonuclease that acts as the effector protein in type V-A CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated) immune systems (1). Within these systems, CRISPR arrays are DNA loci consisting of unique sequences called spacers that are flanked by repeat sequences. Spacer sequences are acquired from foreign nucleic acids during infections and serve as memories to defend against future infections (2). The CRISPR locus is transcribed into a long pre-CRISPR RNA (pre-crRNA), which is processed by Cas12a into small mature CRISPR RNA (crRNA) (3). Cas12a uses these crRNAs as guides to bind to complementary "protospacer" sequences within the invading DNA (1). Following DNA binding, Cas12a sequentially cleaves each strand of the DNA using a RuvC nuclease domain, creating a double-strand break (DSB) that eventually leads to neutralization of the infection (1, 4, 5). Because the crRNA can be changed to guide Cas12a cleavage at a sequence of interest, Cas12a is easy to repurpose for programmable genome editing and other biotechnological applications (1), similar to the widely used *Streptococcus pyogenes* (Sp) Cas9 (6). Several orthologs, including *Francisella novicida* (Fn) Cas12a, *Lachnospiraceae bacterium* (Lb) Cas12a and *Acidaminococcus sp.* (As) Cas12a, have been used for genome editing (1, 7).

The specificity of Cas endonucleases is an important consideration for both CRISPR-Cas

immunity and genome editing applications. While SpCas9 has been shown to tolerate mutations in the target sequence (8, 9), Cas12a orthologs are reported to be relatively specific (10–12). In searching for potential targets, both Cas9 and Cas12a identify a protospacer adjacent motif (PAM) (13, 14) to initiate R-loop formation with the crRNA-complementary strand of the dsDNA target (15, 16). However, they differ in terms of structure and target cleavage mechanisms that may modulate target cleavage specificity (17, 18). The first few nucleotides following the PAM sequence (i.e. the PAM-proximal sequence) where DNA-RNA hybrid nucleation begins is referred to as the "seed" region (19). Mismatches in this region have been reported to be more deleterious for binding and cleavage by Cas12a than by SpCas9 (8, 14, 19–21). This intrinsic low tolerance for mismatches in the target sequence is a desirable trait for high-fidelity genome editing but raises the question of how Cas12a can provide effective immunity to bacteria against invaders in its native role. Higher mismatch tolerance may limit the ability of phages to escape from CRISPR-Cas immunity via mutations and may provide broader defense against closely related phages (10, 22, 23).

In addition to its crRNA-guided target cleavage activity, recent reports showed that Cas12a can indiscriminately degrade single-stranded (ss) DNA in trans upon binding and activation by a crRNA-complementary dsDNA or ssDNA (24, 25). Structural studies of Cas12a have led to a model for this non-specific activity in which the RuvC catalytic pocket remains open due to conformational changes following 16 – 17 nucleotide crRNA-target strand (TS) hybridization and cleavage of at least the non-target strand (NTS) of the target dsDNA (5, 26). The release of PAM-distal products from the complex (21, 27) also increases the accessibility of the RuvC domain to non-specific substrates (5).

In this study, we developed a high-throughput *in vitro* method to determine the native specificity and cleavage activity of Cas12a orthologs. We show that Cas12a has crRNA-directed sequence-specific nicking activity against target sequences containing up to 4 mismatches with the guide RNA, where linearization does not always occur. We further show that when activated by target binding, Cas12a displays non-specific nicking activity in trans against dsDNA, similar to the previously observed activated ssDNA degradation activity (24, 25). Activated Cas12a also has weak dsDNA degradation activity for both target and non-specific DNA. Our results report several cleavage activities of Cas12a including cis (sequence-dependent or target-dependent) and activated trans (non-specific or target-independent) dsDNA nicking, and cis and trans dsDNA degradation resulting from extensive nicking activity.

## RESULTS

### Cleavage activity of Cas12a against a target library

Genome editing activity and specificity of Cas12a have previously been characterized *in vivo* (10–12). These studies show that Cas12a has low or no tolerance for mismatches in the target sequence in eukaryotic cells. However, the native cleavage specificity of Cas12a remains unclear, given that eukaryotic genomic structure may sequester potential off-targets (28–30) and that most off-target analyses account only for double-strand breaks (31, 32). *In vitro* analysis reveals the effect of single mismatches in the target sequence which slow the rate of R-loop formation and target-strand cleavage by Cas12a (27, 33, 34), but it remains unclear how multiple types and combination of mismatches in the target sequence affect the cleavage mechanism.

To directly observe the cleavage activity and specificity of Cas12a, we performed *in vitro* cleavage assays using a plasmid library followed by high-throughput sequencing analysis (Fig. 1A, Fig. S1). The libraries were designed to contain target sequences with different number and combinations of mismatches for a given guide sequence of the crRNA. We used a randomization frequency to generate a pLibrary pool with targets containing 2 or 3 mismatches maximally represented in the pool (see methods section – Library creation) (Fig. S2A) (35). The total number of possible sequences is $4^{20}$ ($1.1 \times 10^{12}$). However, due to the partial randomization, sequences with more than 9 mismatches compose less than 0.1% of the total pool. Therefore, we limited our analysis to target sequences with fewer than 10 mismatches. The 0 mismatch "perfect" target was spiked in as an internal control (Fig. S2A). We tested the cleavage activity of three Cas12a orthologs – FnCas12a, LbCas12a and AsCas12a (collectively referred to as

Cas12a hereafter). For each Cas12a ortholog, we used three different crRNA sequences and corresponding negatively supercoiled (nSC) plasmids (see methods section - Plasmid and nucleic acid preparation) containing the perfect target (pTarget) or target libraries (pLibrary) (Fig. S2B – D, Table S1). The three crRNA and library sequences were designed based on protospacer 4 sequence from *Streptococcus pyogenes* CRISPR locus (55% G/C), EMX1 gene target sequence (80% G/C) and CCR5 gene target sequence (20% G/C), henceforth referred to as pLibrary PS4, EMX1 and CCR5 respectively.

The use of negatively supercoiled dsDNA plasmid substrates enables detection of nicked and linear cleavage products resulting from Cas12a cleavage via product migration on agarose gel electrophoresis (36). Cas12a completely linearized all pTarget substrates within the time course tested, indicating complete cleavage of both strands of the dsDNA target (Fig. 1B, Fig. S2E). In contrast, only a fraction of pLibrary was linearized and some of the plasmid remained supercoiled, indicating the presence of sequences within the pool that could not be cleaved by Cas12a in the time span tested (Fig. 1B, C Fig. S2E, S3A, B). We also observed a nicked fraction for pLibrary that persisted through the longest time point in the assay (3 h) (Fig. 1B, C, Fig. S2E, S3A, B). To determine whether any cleavage might occur outside of the target region during pLibrary cleavage, we tested the cleavage activity of Cas12a against the empty plasmid backbone without and with the different crRNAs (Fig. S2F). Cas12a orthologs have been reported to have variable crRNA-independent nicking activity (37, 38). Similarly, we observed weak, variable amounts of plasmid nicking when incubated with Cas12a in the absence of crRNA (Fig 1B, Fig. S2E). To prevent this crRNA-independent nicking, we incubated Cas12a with excess crRNA in all cleavage assays (see methods section - *In vitro* cleavage assay). For most Cas12a-crRNA pairs, we observed minimal nicking of the empty plasmid backbone until the 3 h time point. However, more nicking occurred at the shorter time points for AsCas12a-CCR5 crRNA. Although some level of background non-specific nicking was observed for all crRNAs tested, further analysis indicates that pLibrary nicking is, in part, sequence dependent (see below).

Interestingly, we observed variable cleavage rates and patterns for the three pLibraries and Cas12a orthologs (Fig 1C, Fig. S3A, B). Both FnCas12a and LbCas12a cleaved most pLibraries with a relatively high efficiency, resulting in substantial depletion of supercoiled DNA to ~10% of the total DNA at the longest time point (3 h). One notable exception is pLibrary EMX1, for which FnCas12a showed relatively weak cleavage in comparison to LbCas12a (Fig. S3A). AsCas12a showed the least amount of overall cleavage, with a relatively large fraction of pLibrary remaining in the supercoiled pool for all three libraries, suggesting a large fraction of sequences remained uncleaved by this ortholog.

To determine which sequences were uncleaved and nicked, we extracted the plasmid DNA from gel bands for each of these fractions, PCR amplified the target region and performed high-throughput sequencing (HTS) followed by bioinformatic analysis (Fig. 1A, Fig. S1). Although we could not PCR amplify and analyze target sequences present in the linearized DNA pool due to the DSB generated in the target sequence by Cas12a upon cleavage, we assumed that sequences that were absent from both the supercoiled and nicked pools were linearized. PCR amplification of the target region from the plasmid for high-throughput sequencing facilitates analysis of target sequences, however, it also removes our ability to detect the relative abundance of sequences in the nicked versus supercoiled fraction. While we were unable to sequence the linearized pool, the analysis of target sequences in the nicked and supercoiled pool provides an understanding of the mismatch tolerance of Cas12a (see below).

To adjust for the abundance of sequences present in the supercoiled or nicked pool prior to PCR amplification, we normalized counts obtained from our high-throughput sequencing data using the fraction of DNA that was present in a given pool at a given time point (see methods section – HTS analysis). We normalized abundance at a given time point relative to the abundance in the original library, enabling analysis of the relative depletion of sequences from the supercoiled pool and enrichment of sequences within the nicked pools over time (Fig. 1D, E, Fig. S3C – F, S4). In addition, we plotted mismatch distribution curves for the normalized abundance of sequences present in the supercoiled and nicked pools to analyse the

proportion of sequences with a given number of mismatches in each pool at each time point (Fig. S5).

Depletion of target sequences from the supercoiled pool indicates cleavage, and enrichment in the nicked pool indicates nicking by Cas12a. The decrease of target sequences present in the nicked pool over time implies linearization of these sequences. As expected, the perfect target (0 mismatch) was quickly depleted from the supercoiled pool (Fig. 1D, Fig. S3C, D). While the perfect target sequence was present in the nicked pool at earlier time points, over time it was depleted from the pool, indicating linearization (Fig. 1E, Fig. S3E, F). All Cas12a orthologs also cleaved target sequences with multiple mismatches (i.e. target sequences containing mismatches with the guide RNA) based on the depletion of sequences with 1 to 4 mismatches from the supercoiled pool over time (Fig. 1D, Fig. S3C, D). Single mismatch sequences were depleted from the supercoiled pool at similar rates to the perfect target sequence by LbCas12a for pLibrary EMX1 and CCR5 (Fig. S3C, D), but slightly more slowly for other orthologs and pLibraries (Fig. 1D, Fig. S3C, D). Single-mismatch sequences were also generally depleted more slowly from the nicked pool than the perfect target (Fig. 1E, Fig. S3E, F), indicating a decreased rate of linearization of single mismatch targets by Cas12a. For the A/T rich pLibrary CCR5, all three orthologs displayed the fastest rate of depletion of sequences with 2 to 4 mismatches (Fig. S3D). Target sequences with 2, and in some cases with 3, mismatches entered and were subsequently depleted from the nicked pool, suggesting that Cas12a can linearize these sequences with reduced kinetics for the second cleavage step (Fig. 1E, Fig. S3E, F).

Interestingly, target sequences with 5 or more mismatches were also depleted from the supercoiled pool by FnCas12a and LbCas12a for all three pLibraries (Fig. 1D, E, Fig. S3C, D, S4). Correspondingly, we observed an increase for these sequences in the nicked pool over time (Fig. 1E, Fig. S3E, F, S4). In contrast, AsCas12a showed minimal depletion of sequences with 5 to 10 mismatches from the supercoiled pool (Fig. S4); although these sequences were enriched in the nicked fraction, they remained a relatively small proportion of the overall library even at late time points (Fig. S4, S5). Importantly, unlike the target sequences with 1 to 4 mismatches, target sequences with 5 or more mismatches were generally cleaved at almost the same rate by all Cas12a orthologs (Fig. S4). These results suggest that sequences with large numbers of mismatches may have been nicked through a non-specific nicking activity that is more pronounced for LbCas12a and FnCas12a than for AsCas12a (see below).

**Sequence determinants of Cas12a cleavage activity**

We next looked at the sequences that were present in the supercoiled and nicked pools to determine the effects of mismatch position and type (Fig. S1). The heatmaps in Figure 2 and Figures S6, S7 show the relative abundance of target sequences containing 1 to 6 mismatches (MM) with all possible nucleotides at each position of the sequence in the supercoiled fraction, traced over time (see methods section – HTS analysis). Although target sequences with a single mismatch (1 MM) were quickly depleted from the supercoiled fraction (Fig. 2, Fig. S6, S7), some sequences with mismatches in the PAM-proximal "seed" region were depleted slowly, consistent with the expected importance of this region for R-loop formation and Cas12a binding affinity. We observed a short seed region of ~6 nucleotides for most pLibraries and Cas12a orthologs, when compared to the ~10 nucleotide length for SpCas9 (8, 14, 20). This is also in agreement with previously reported *in vivo* and *in vitro* specificity studies on Cas12a (10–12, 19). A single C substitution in the seed region is highly deleterious for cleavage by Cas12a, while G or A mismatches slowed the rate of cleavage or depletion from the supercoiled fractions to a lesser degree (Fig 2, Fig. S6, S7). In contrast, target sequences with a single T mismatch in the seed are generally tolerated for cleavage outside of the first PAM-proximal position for most pLibraries. Outside of the seed, most target sequences containing a single mismatch were rapidly depleted from the supercoiled pool, indicating that any type of single mismatch outside the seed can be similarly tolerated. However, in some cases single mismatches in the target sequence slowed the rate of cleavage regardless of mismatch position based on a slower decrease in abundance relative to other Cas12a orthologs, particularly for FnCas12a cleavage of pLibrary PS4 and both FnCas12a and

LbCas12a cleavage of pLibrary EMX1 (Fig. 2, Fig. S6).

A similar seed-dependent cleavage trend was observed for target sequences with 2, 3 and 4 mismatches (2 MM, 3 MM and 4 MM, respectively) (Fig 2, Fig. S6, S7). In general, C substitution is most deleterious for cleavage at all positions in the target sequence with 2 and 3 mismatches. A and G substitutions also slowed the rate of cleavage by Cas12a, particularly in pLibrary CCR5 (Fig. S7). For LbCas12a, most target sequences with 2 and 3 mismatches outside the seed region were eventually depleted in all three pLibraries. All three Cas12a orthologs tolerate up to 3 mismatches in the PAM-distal region for pLibrary CCR5 (Fig. S7), resulting in strong depletion of these sequences from the supercoiled fraction (Fig. S3D). For sequences with 5 and 6 mismatches, we observed a steady depletion of all sequences regardless of mismatch position or type for FnCas12a and LbCas12a (Fig 2, Fig. S6, S7). Together with the observation that sequences with 5 to 10 mismatches are depleted from the supercoiled fraction at the same rate (Fig. S4), these data suggest that Cas12a may cleave highly mismatched sequences present in pLibrary in a sequence-independent manner.

**Sequence determinants of Cas12a nickase activity**

To determine target sequences that are preferentially nicked by Cas12a, we performed analysis on the nicked pool similar to the supercoiled pool (Fig. S1). The heatmaps in Figure 3 and Figures S8, S9 show the relative abundance of target sequences containing 1 to 6 mismatches (MM) with all possible nucleotides at each position of the sequence in the nicked fraction, traced over time (see methods section – HTS analysis).

Cas12a linearized most single mismatches (1 MM) within the time frame tested (3 hours) (Fig. 3, Fig. S8, S9). While LbCas12a rapidly linearized target sequences with most single mismatches, FnCas12a and AsCas12a displayed slower kinetics for double-stranded cleavage for some single mismatch sequences in pLibrary PS4 (Fig. 3). Although they were a relatively small proportion of the overall library (Fig. S5), we observed enrichment of single mismatch sequences when the mismatch was located either in the seed or in the region where the non-target strand is cleaved (position 16 - 18 from the 5' end of the target on the non-target strand) (1, 33) (Fig. 3, Fig. S8, S9). Similarly, target sequences with 2 and 3 mismatches (2 MM and 3 MM, respectively) in the seed and PAM-distal region were enriched in the nicked fraction at later time points, especially for FnCas12a and AsCas12a (Fig. 3, Fig. S8, S9). For LbCas12a, these 2 and 3 mismatch targets in all pLibraries were eventually depleted, indicating that these sequences were linearized (Fig. 3, Fig. S8, S9). For AsCas12a cleavage of PS4 and EMX1 pLibraries, some target sequences with 2 and 3 mismatches in the PAM-distal region remained enriched throughout the time course (Fig. 3, Fig. S8), indicating that mismatches in the PAM-distal region block the second cleavage step for this Cas12a ortholog.

FnCas12a and LbCas12a can nick target sequences with 4 or more mismatches, observed as an increase in abundance of these target sequences for all pLibraries tested (Fig. 3, Fig. S8, S9). Similar to the supercoiled fraction, target sequences with 5 and 6 mismatches were uniformly enriched in the nicked fraction irrespective of the mismatch position or type for most pLibraries and Cas12a orthologs (Fig. 3, Fig. S8, S9). However, AsCas12a did not nick target sequences with 4 or more mismatches for pLibraries PS4 and EMX1 (Fig. 3, Fig. S8) but exhibited weak nicking of these target sequences in pLibrary CCR5 (Fig. S9).

**Effect of mismatch location and separation on Cas12a cleavage activity**

The heatmaps presented in Figures 2, 3 and Figures S6 – S9 show the overall effects of multiple mismatches in the target sequences. In these heatmaps, the effect of a single mismatch on Cas12a cleavage can be easily identified. We next generated heatmaps to show the effects of 2 mismatches on Cas12a cleavage as a function of mismatch location and distance between the mismatches (see methods section – HTS analysis). The total number of unique sequences with 2 mismatches in our pLibraries is 1,710, but the number of possible ways 2 mismatch sequences can occur is 190 (Fig. S10). Cas12a can cleave most targets with 2 mismatches, based on depletion of these sequences from the supercoiled pool over time (Fig. 4). While target sequences with 2 mismatches separated by a shorter distance were depleted slowly from the supercoiled pool,

sequences with one mismatch in the seed region were also enriched in the supercoiled pool irrespective of the distance from the other mismatch, highlighting the effect of a mismatch in the seed (Fig. 4, Fig. S10). Target sequences with 2 mismatches in G/C rich EMX1 pLibrary were depleted slowly by all three Cas12a orthologs in comparison to the other two pLibraries (Fig. 4B). Interestingly, FnCas12a can tolerate most double mismatches in A/T rich CCR5 pLibrary irrespective of the distance between the mismatches except when one mismatch is in the seed. In contrast, LbCas12a and AsCas12a are not tolerant of sequences with mismatches separated by a distance of 5 or fewer bases of pLibrary CCR5 (Fig. 4C).

Cas12a strongly nicks sequences with 2 mismatches and in some cases linearizes them, as observed by initial enrichment and subsequent depletion of sequences (Fig. 4). Among the three Cas12a orthologs tested, LbCas12a displays the most linearization while AsCas12a has the strongest sequence-specific nicking activity against target sequences with 2 mismatches. Interestingly, for the nicked pool, we observed less accumulation of 2 mismatch sequences containing mismatches in the seed region for AsCas12a than for FnCas12a and LbCas12a (Fig. 3, Fig. S8, S9). Consistently, we do not see representation of target sequences with a mismatch in the seed (i.e. when the mismatches are separated by a distance of 16 or more) in the nicked pool (Fig. 4, Fig. S10). It is possible that these sequences are linearized without accumulation of the nicked intermediate, as these sequences are depleted from the supercoiled pool (Fig. 4).

We observed an initial, overall enrichment of target sequences with 2 mismatches separated by a distance of 13 or more due to one of the mismatches being in the seed (Fig. 4, Fig. S10). To further study the bias and effect of a seed mismatch, we generated similar heatmaps for target sequences with 2 mismatches in the seed or outside the seed region. As expected, most double mismatches in the seed region are deleterious for Cas12a (Fig. S11). These sequences were slowly depleted from the supercoiled pool and appear in the nicked pool over time. This indicates that Cas12a mostly nicks target sequences with two mismatches in the seed.

Cas12a can tolerate most target sequences with 2 mismatches outside the seed seen as a depletion from the supercoiled pool over time for all pLibraries tested (Fig. S11D – F). We observed mismatched sequences were more rapidly depleted from the supercoiled pool from A/T rich pLibrary CCR5 by all three Cas12a orthologs compared to the other two pLibraries (Fig. S11D – F). Cas12a strongly nicks sequences when the two mismatches outside the seed are closer, seen as the enrichment of these sequences in the nicked pool. LbCas12a consistently linearizes sequences with two mismatches from all three pLibraries. Double mismatches separated by a distance of 10 or more are more rapidly linearized and depleted from the nicked pool by all three Cas12a orthologs. These observations also agree with sequence-specific nicking activity reported for Cas12a (39).

**Cas12a has non-specific activated nicking and dsDNA degradation activity**

Our HTS data suggests that Cas12a can nick, and in some cases linearize, sequences with several mismatches. To validate this observation, we selected sequences from pLibrary PS4 that were relatively enriched in the nicked fraction at the longest time point (3 hours). We cloned target sequences containing 2 to 8 mismatches and individually tested sequence-specific nicking activity of Cas12a against each plasmid. Consistent with our HTS results, we observed varying degrees of sequence-specific nicking and linearization of target sequences containing 2, 3 and 4 mismatches (MM) for different Cas12a orthologs (Fig. 5A). To compare among the Cas12a orthologs, we quantified the supercoiled, linearized and nicked fractions at the 3 hour time point (based on the time frame used for the pLibrary cleavage) for the perfect and mismatched target sequences (Fig. 5B, Fig. S12B, C). FnCas12a and AsCas12a showed higher sequence-specific nicking activity, while LbCas12a partially linearized most targets. All three Cas12a orthologs strongly nicked one of the 3 mismatch target sequences (3.2 MM) tested, with no or low partial linearization of the mismatched target. The distinct patterns of nicking observed for targets with different mismatches indicate that the nicking is target sequence-specific, and not a result of non-specific, background nicking of the plasmid backbone by Cas12a-crRNA (Fig. S2F). These data, together with our HTS data, suggests that LbCas12a has strong linearization activity against target sequences with multiple mismatches while

FnCas12a and AsCas12a can only nick these target sequences.

Surprisingly, sequences with greater than 5 mismatches were not nicked by Cas12a even after 5 hours of incubation, although these sequences were enriched in the nicked pool of our HTS data (Fig. 1E, 5B, Fig. S3E, F, S4, S12). As noted above, the number of mismatches did not affect the rate of cleavage for target sequences with 5 or more mismatches (Fig. S4) and heatmaps for target sequences with 5 and 6 mismatches indicated enrichment of these sequences in the nicked pool was not sequence specific (Fig. 3, Fig. S8, S9). Although we observed similar levels of non-specific cleavage of the empty plasmid backbone by all three Cas12a orthologs (Fig. S2F), sequences with 5 or more mismatches were depleted from the supercoiled pool far faster for FnCas12a and LbCas12a than for AsCas12a (Fig. S5). Together, these observations suggest an additional non-specific cleavage activity occurred during pLibrary cleavage. This led us to hypothesize that FnCas12a and LbCas12a may have non-specific, target-activated nicking activity against dsDNA substrates with low or no homology to the crRNA. In the pLibrary cleavage assays, the mixed pool of sequences contains the perfect target sequence which may activate Cas12a for non-specific nicking activity (24, 25).

To test this, we used a short dsDNA oligonucleotide activator that was fully complementary to the crRNA to activate Cas12a. We formed a complex containing Cas12a, crRNA and the dsDNA activator and tested for non-specific cleavage activity against empty negatively supercoiled (nSC) plasmid that lacked a target sequence. Surprisingly, we observed robust non-specific, trans nicking and partial linearization of the empty negatively supercoiled plasmid by FnCas12a and LbCas12a (Fig. 6, top panels). This nicking activity is readily detected using negatively supercoiled dsDNA substrates where a single nick in the DNA is sufficient to cause a large shift in mobility of the DNA (36). The activated nicking activity is comparable to Cas12a ssDNase activity, with dsDNA nicking and ssDNA degradation observed in the same time frame (Fig. 6, bottom panels). Over time, FnCas12a and LbCas12a degraded the ssDNA, and similarly, further nicking events in the dsDNA eventually resulted in linearization. A discrete linear band initially appeared in the agarose gels, indicating nicking of both strands in close proximity. However, over time multiple cleavage events, either along the length of the plasmid or from the freed ends, lead to degradation.

The observation that trans-activated Cas12a could degrade linearized dsDNA suggests that degradation should also be observed for plasmids containing a target that can activate Cas12a in cis. We therefore tested cleavage of pTarget by Cas12a and performed the cleavage assay for longer time points (up to 24 h) to determine whether degradation eventually occurs. Negatively supercoiled pTarget was linearized quickly, likely due to specific cleavage within the target region. Following linearization, pTarget was slowly and processively degraded after 4 hours of incubation with LbCas12a or FnCas12a, indicating that these Cas12a orthologs can be activated for dsDNA degradation by a cis activator (Fig. S13A). In contrast, pTarget remained uncleaved by all three Cas12a orthologs programmed with a non-cognate crRNA and in the absence of an activator.

We next tested the types of trans activators that can trigger dsDNA nicking by Cas12a. Like the non-specific ssDNase activity (24, 25), Cas12a can be activated by crRNA-complementary ssDNA binding in a PAM-independent, RuvC-domain dependent manner for trans, non-specific dsDNA nicking and degradation (Fig. S13B – D, S14A). We also tested whether mismatched target sequences that were present in the pLibrary cleavage assays could act as activators for non-specific nicking activity. Interestingly, Cas12a was activated by some of these mismatched targets as well, especially those that were partially linearized by Cas12a (Fig. S14B). However, mismatched targets that were only nicked by Cas12a, like an activator based on mismatched pTarget 2.1 MM (Fig. 5B, Fig. S12B, C) were weak activators, indicating that double-strand cleavage of the mismatched target activator may be important for non-specific activated nicking of dsDNA.

Both FnCas12a and LbCas12a have strong activated, non-specific nicking activity at lower concentrations (20 nM). AsCas12a is not strongly activated as a nickase at this concentration, although it does display activated ssDNase activity (24) (Fig. 6, Fig. S13). The HTS data indicated that target sequences containing more than 4 mismatches were modestly enriched in the nicked

fraction by AsCas12a (Fig. 1E, 3, Fig. S4, S8, S9). Indeed, we observed a stronger non-specific activated nicking by AsCas12a at higher concentration (100 nM) at which the pLibrary cleavage assays were performed (Fig. S14C) (see Method section – *In vitro* cleavage assay). AsCas12a is also reported to have slower rates of PAM-distal product release after cleavage of a target (5, 27). The reduced non-specific activity may be a result of the cleaved target products hindering the RuvC domain from accessing dsDNA substrates. Some conformational changes or lack thereof may also not allow for non-specific activated nicking of dsDNA substrates by AsCas12a.

The non-specific nicking observed for plasmid substrates may be due to the partially single-stranded nature of negatively supercoiled DNA. To determine whether negative supercoiling is required for activated cleavage, we tested plasmid substrates in three different forms – negatively supercoiled (nSC), linear (li) and nicked (n). While activated Cas12a robustly nicked negatively supercoiled DNA, resulting in linearization and eventual degradation, nicked dsDNA was only weakly linearized and degraded (Fig. S15). This activity was also reproducible with commercially available enzymes and other crRNA-activator pairs, although the activation of the trans-nicking activity varied with the crRNA-activator pair (Fig. S15). These results suggest that nicking is dependent on the supercoiled state of the DNA, while degradation may require exposed termini.

## DISCUSSION

Cas12a has become a widely used tool for various biotechnological applications such as genome editing and diagnostic tools (7, 24, 40). Several reports show that Cas12a and engineered orthologs are highly specific for RNA-guided dsDNA cleavage activity (10–12). Despite these studies on Cas12a specificity, the cleavage activity and specificity outside of a eukaryotic setting remains unclear. The apparent high specificity of Cas12a in genome editing studies is paradoxical to its natural role as an immune system effector. Phages evolve rapidly and can escape from CRISPR-Cas immunity via mutations (22, 23). The high specificity of Cas12a may also limit targeting of closely related phages (41).

Here we show that Cas12a has additional sequence-specific and non-specific activated dsDNA nicking and degradation activities apart from previously described crRNA-mediated cis cleavage of dsDNA targets (1), activated trans cleavage of non-specific ssDNA substrates (24, 25) and RNA-independent dsDNA nicking and ssDNA degradation (37). Our results demonstrate that Cas12a can nick and, in some cases, create double-strand breaks in targets with up to 4 mismatches. Similarly, a recent study by Fu et al. (39) demonstrated that Cas12a and Cas9 have sequence-dependent nicking activity against targets with 1 or 2 mismatches. We also establish that Cas12a has non-specific dsDNA nicking activity upon binding to a crRNA-complementary DNA. While this manuscript was in preparation, a complementary study reported similar observations, demonstrating that these activities are reproducible *in vitro* (42).

Cas12a has a single active site, and it cleaves the dsDNA target in a sequential order with non-target strand (NTS) nicking followed by target-strand (TS) cleavage (5, 21). As a result, Cas12a cleaves the NTS faster than the TS (26, 33). In contrast, double-stranded cleavage by Cas9 is coordinated simultaneously by two different nuclease domains, although studies suggest that target sequences with mismatches result in differential rates of TS and NTS cleavage that may lead to a nicked intermediate (39, 43, 44). Similarly, the difference in rates of strand cleavage by Cas12a could potentially be amplified by the presence of mismatches in the target sequence, resulting in a nicked intermediate. Consistent with previous reports, we observed that single and multiple PAM-distal mismatches slow the linearization of the target sequence by Cas12a (33, 39). Target sequences with single mismatches reduce binding affinity and slow the rate of R-loop formation by Cas12a, which should result in a reduced rate of NTS cleavage. However, the observation of nicked intermediates suggests that specific types of mismatches may cause slower cleavage of the TS following NTS cleavage, resulting in sequence-specific nicking. In this case, TS cleavage may be the rate limiting step for double-stranded break formation. It is also possible that the target dissociates following NTS cleavage due to R-loop collapse, as mismatches in the PAM-distal region have previously been shown to accelerate the rate of target dissociation (33). Alternatively, the TS

may not be presented to the RuvC domain and/or the conformational change required for TS cleavage may not occur. Biochemical studies have revealed that TS recognition is a pre-requisite for NTS and ssDNA cleavage (24).

The release of the PAM-distal products upon cleavage exposes the active RuvC domain that can accept substrates for non-specific trans cleavage activities (5, 21, 27). Consistently, we observed that certain mismatched targets can only be nicked by Cas12a and these natural mismatched substrates are also ineffective activators of the non-specific nicking activity of Cas12a presumably due to failed cleavage and release of PAM-distal TS products. Targets with multiple mismatches also cause reduced binding affinity and defects in R-loop formation (21, 34), resulting in slower cleavage of both strands.

Cas9 cleavage specificity is not only dictated by position-specific mismatches but is also defined by the target sequence (20, 45). Here, by testing three different Cas12a orthologs against three pLibraries, we show the effects of target-crRNA sequence and different nucleotide substitutions across the target region in varying mismatch combinations on Cas12a cleavage activity. All three Cas12a orthologs can tolerate most types of PAM-distal mismatches in A/T rich target sequences in pLibrary CCR5 where these sequences are quickly depleted from the supercoiled pool. However, almost all target sequences with 1, 2 and 3 mismatches in G/C rich pLibrary EMX1 had slower cleavage rates, where we observed accumulation of nicked products. We also observe that LbCas12a tolerates the most mismatches and is likely to produce double-strand breaks in targets with mismatches while FnCas12a and AsCas12a may be more prone to nicking these target sequences.

A recent study demonstrated that topology plays a role in Cas12a-mediated target DNA cleavage (46). Similarly, we observed different rates of non-specific activated nicking and degradation of different forms of DNA. Nicking by activated Cas12a was easily detected for negatively supercoiled DNA. Cas12a is likely to mainly encounter negatively supercoiled DNA in both prokaryotic and eukaryotic cells, suggesting that the activated trans cleavage could enable dsDNA nicking *in vivo*. Negatively supercoiled DNA may undergo relaxation and breathing, exposing ssDNA. We also observe slow degradation of linear dsDNA by activated-Cas12a, where the linear ends of dsDNA may fray (47, 48). In both cases, the exposed ssDNA regions are likely degraded via the ssDNase activity of Cas12a (24, 25). With pre-nicked dsDNA, we observed slow and direct degradation by activated-Cas12a rather than an intermediate linear product. This may be attributed to decreased breathing and exposure of ssDNA sites. It remains unclear how negatively supercoiled plasmids are nicked multiple times leading to linearization; however, multiple events of the trans nicking activity may cause the DNA to eventually fall apart (42). This suggests that a cumulative effect of all the trans activities of Cas12a eventually leads to complete degradation of nucleic acid substrates (Fig. 7).

Cas12a has been successfully used for gene editing *in vivo* without any deleterious off-target effects (10–12, 49, 50). Our high-throughput pLibrary cleavage analysis indicates that Cas12a can bind and cleave sequences with multiple mismatches *in vitro*. Although we cannot establish with certainty how and what target sequences ended up in the nicked pool, our mismatch and position analysis along with the individual 2 to 4 mismatch target sequence cleavage assays strongly suggests that target sequences with 1 to 4 mismatches are nicked in a sequence-dependent manner. *In vivo*, nicked DNA may lead to genomic changes as DNA repair machinery can be recruited at these sites (51–53). However, chromatin structure may prevent nicks and nicks may be repaired by error-free DNA repair pathways (54). The cellular context also plays a role in Cas effector binding and cleavage. SpCas9 can bind to targets containing several mismatches depending on DNA breathing and supercoiling, both *in vitro* and *in vivo* (55, 56). Cas12a can stably bind to some targets with mismatches *in vitro* (27, 34), but *in vivo* studies suggest low or no off-target binding (57). This could reflect the inability of Cas12a to unwind and bind DNA in varying topological and cellular contexts which may result in overall lower off-target editing rates by Cas12a.

While off-target sites can be predicted and avoided by careful design of crRNAs, the robust non-specific activated nicking activity we observed *in vitro* may lead to off-target editing as nicked DNA can result in indels as discussed above. Although there are a limited number of target sites

in the genome where Cas12a can be activated for non-specific nicking, it is unclear how long Cas12a remains bound to a target. Prolonged binding could result in nicking of proximal DNA sequences. The unpredictable nature of the non-specific activated nicking makes it difficult to detect the outcomes of nicking. In addition, the commonly used methods to verify off-target editing do not detect nicks (31, 32), meaning that detection of potential off-target effects due to non-specific nicking would require whole genome sequencing or specialized detection methods (58, 59). Use of Cas12a orthologs, such as AsCas12a, that display reduced non-specific nickase activity may reduce these unpredictable effects during genome editing experiments. Notably, our *in vitro* specificity analysis also suggests that AsCas12a is less prone to creating double-strand breaks at sites with 2 or 3 mismatches, suggesting that this ortholog may be less prone to off-target, sequence-specific nicking and cleavage at highly homologous sites.

The activities reported in our study add to the growing number of target sequence specific and sequence non-specific Cas12a cleavage activities (Fig. 7) and may provide a possible explanation of how Cas12a compensates for its highly specific targeted cleavage activities as an immune effector. The sequence-specific nicking, non-specific activated nicking and degradation activities, along with previously described dsDNA cleavage (1) and trans ssDNase activity (24, 25), could allow Cas12a to mount a strong defence against different types of invading phages. In the event of phage evolution via mutations, Cas12a may tolerate some mutations and still nick or fully cleave phage DNA via the sequence-specific nicking activity. Cas12a could also be activated by the evolved/mutated target region of the phage DNA, enabling non-specific activated nicking and degradation activities.

While our *in vitro* results indicate that Cas12a has robust nicking activities, it remains uncertain whether this activity occurs *in vivo*. During phage infection, multiple copies of phage DNA are injected into the bacterial cell, each carrying targets that can activate Cas12a. The effective activated-Cas12a concentration in the bacterial cell is therefore relative to the level of infection occurring at a given time in the cell. It is also interesting to note that recent studies reported several Cas12 proteins with strong nickase activities, even against perfectly matching target DNA (60, 61). Similarly, the recent discovery of the Can1 nickase that is activated by signalling molecules produced by type III Cas effectors upon target recognition suggests that targeted nicking or target-activated nicking may be a common mechanism among CRISPR-Cas systems to slow phage replication (62). The non-specific nature of the nicking and degradation activities may also be harmful to the host bacteria. In type III and type VI CRISPR-Cas systems, Cas nucleases can be activated for non-specific cleavage of RNA (4). Perhaps, the non-specific nicking activity is a means to prevent phage proliferation by initiating programmed cell death (PCD), abortive infection or dormancy in order to save the bacterial population (63, 64) . Further studies are required to investigate the cost-benefit relation of such non-specific activities of Cas12a to the bacteria.

## EXPERIMENTAL PROCEDURES

### Cas12a cloning

The gene sequences for *Francisella novicida* (Fn) and *Acidaminococcus sp.* (As) Cas12a were synthesized as *Escherichia coli* codon-optimized gBlocks (purchased from Integrated DNA Technologies, IDT) and inserted into pSV272 using Gibson assembly (New England Biolabs) as per the manufacturer's protocol, to generate a protein expression construct encoding Cas12a fused with N-terminal 6X-His sequence, a maltose binding protein (MBP) and a Tobacco Etch Virus (TEV) protease cleavage site. The AsCas12a clone inadvertently carried two mutations P874G and G1292V, which we confirmed did not affect the activity of this protein (data not shown). Experiments shown in Fig. 6 and Fig. S14C were performed with WT AsCas12a. *Lachnospiraceae bacterium* (Lb) Cas12a was expressed using expression plasmid pMAL-his-LbCpf1-EC. pMAL-his-LbCpf1-EC was a gift from Jin-Soo Kim (Addgene plasmid # 79008 ; http://n2t.net/addgene:79008;RRID:Addgene_79008) (10). Catalytically inactive (pre-crRNA processing and DNase dead) Cas12a mutants were generated via site-directed mutagenesis (SDM) or Round-The-Horn (RTH) PCR and verified by Sanger sequencing (Eurofins Genomics, Kentucky, USA) (see Table S1 for SDM primers).

### Cas12a expression and purification

All Cas12a proteins were expressed in Escherichia coli BL21 (DE3) cells. 2X TY broth supplemented with corresponding antibiotics was inoculated with overnight cultures of cells in 1:100 ratio. Cultures were grown to an optical density (600 nm) of 0.5 − 0.6 at 37 °C and protein expression was induced by the addition of IPTG to a final concentration of 0.2 mM. The incubation was continued at 18 °C overnight.

FnCas12a was purified by the following protocol adapted from a previous method (65) to initially test purification of the protein. Cells were resuspended in Lysis Buffer II (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10mM imidazole, and 10% glycerol) supplemented with PMSF. Cells were lysed by sonication or a homogenizer and the lysate was centrifuged to remove insoluble material. The clarified lysate was applied to a HisPur™ Ni-NTA Resin (ThermoFisher Scientific) column. The column was washed with 10 column volumes of Lysis Buffer and bound protein was eluted in Elution Buffer II (Lysis Buffer II + 250 mM imidazole final concentration). The elution was concentrated and run on a HiLoad 16/600 Superdex 200 gel filtration column (GE Healthcare) pre-equilibrated with SEC Buffer A (20 mM Tris-HCl, pH 8.0, and 500 mM NaCl). Fractions containing 6X His-MBP tagged Cas12a were collected and treated with TEV protease in a 1:100 (w/w) ratio, overnight at 4 °C. Samples were reapplied to HisPur™ Ni-NTA Resin (ThermoFisher Scientific) to remove the His-tagged TEV, free 6X His-MBP, and any remaining tagged protein. The flow-through was collected, concentrated and further purified by using a HiLoad 16/600 S200 gel filtration column in SEC Buffer B (20 mM Tris-HCl, pH 8.0, 200 mM KCl, and 1mM EDTA). Peak fractions were combined, concentrated, and flash frozen in liquid nitrogen and stored at −80°C until further use.

After further optimization of the purification protocol, LbCas12a, AsCas12a and all Cas12a mutants were purified using a modified protocol adapted from a previous method (66). Cells were harvested by centrifugation and the cell pellet was resuspended in Lysis Buffer I (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole), supplemented with protease inhibitors PMSF, cOmplete™ Protease Inhibitor Cocktail Tablet or Halt Protease Inhibitor Cocktail. Cells were lysed by sonication or a homogenizer and the lysate was centrifuged to remove insoluble material. The clarified lysate was applied to a HisPur™ Ni-NTA Resin (ThermoFisher Scientific) column. The column was washed with 10 column volumes of Wash Buffer (Lysis Buffer + 15 mM imidazole final concentration) and bound protein was eluted in Elution Buffer I (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 250 mM imidazole). Fractions containing Cas12a were pooled and TEV protease was added in a 1:100 (w/w) ratio. The sample was dialyzed in Dialysis Buffer (10 mM HEPES-KOH pH 7.5, 200 mM KCl, 1 mM DTT) at 4°C overnight. For further purification the protein was diluted 1:1 with 20 mM HEPES-KOH (pH 7.5) and loaded on a HiTrap Heparin HP (GE Healthcare) column. The column was washed with Buffer A (20 mM HEPES-KOH pH 7.5, 100 mM KCl) and eluted with Buffer B (20 mM HEPES-KOH pH 7.5, 2 M KCl) by applying a gradient from 0% to 50% over a total volume of 60 ml. Peak fractions were analyzed by SDS-PAGE and fractions containing Cas12a were combined, and DTT was added to a final concentration of 1 mM. The protein was fractionated on a HiLoad 16/600 Superdex 200 gel filtration column (GE Healthcare), eluting with SEC buffer (20 mM HEPES-KOH pH 7.5, 500 mM KCl, 1 mM DTT). Peak fractions were combined, concentrated, and flash frozen in liquid nitrogen and stored at −80°C until further use.

Commercially available Cas12a were purchased from New England Biolabs (NEB) – LbCas12a and Integrated DNA Technologies (IDT) – AsCas12a.

**Library creation**

To generate a pool of sequences containing mismatches (i.e. target sequences containing mismatches with the guide RNA), the library was partially randomized (35). The following probability distribution function was used to determine the randomization/doping frequency,

$$P\,(n, L, f) \;=\; \frac{L!}{n!\,(L \;-\; n)!}(f^{n})(1 \;-\; f)(L - n)$$

where, P is the fraction of the population, L is the sequence length, n is the number of mutations/template and f is the probability of mutation/position (doping level or frequency). The number of different mutation combinations (MMc) for a given number of mutations, n and sequence

length, L, regardless of the doping level/frequency is determined by,

$$MM_c = 3^n \frac{\mathrm{L!}}{\mathrm{n!\,(L\ -\ n)!}}$$

For example, the total number of unique target sequences with a single mismatch is 60, with 2 mismatches is 1,710, and with 3 mismatches is 30,780, etc. The three library sequences tested were a modified protospacer 4 sequence from *Streptococcus pyogenes* CRISPR locus (55% GC), EMX1 gene target sequence (80% GC) and CCR5 gene target sequence (20% GC) (see Table S1 for target sequence). A randomization/doping frequency (f) of 15% was selected to optimize the library to contain a maximum of sequences with 2 or 3 mismatches.

**Plasmid and nucleic acid preparation**

DNA oligonucleotides were synthesized by Integrated DNA Technologies (IDT) or Thermo Scientific. All RNAs and single-stranded library oligonucleotides with 15% doping frequency in the target region were ordered from IDT. Sequences of DNA oligonucleotides and RNA used are listed in Table S1.

The oligonucleotides (for the libraries, target and mismatched targets) were diluted to 0.2 µM in 1X NEBuffer 2. pUC19 vector was amplified via PCR to insert homology arms, followed by DpnI digestion and PCR cleanup (Promega Wizard SV Gel and PCR Clean-Up System). 30 ng of PCR amplified pUC19, 5 µL of oligonucleotide (0.2 µM) and ddH2O to bring the volume to 10 µL were mixed with 10 µL 2X NEBuilder HiFi DNA Assembly Master mix (New England Biolabs) and incubated at 50 °C for 1 hour. 2 µL of the assembled product was transformed into NEB Stable competent cells as per the manufacturer's protocol. After the recovery step, all of cells in the outgrowth media were used to inoculate a 50 mL LB supplemented with ampicillin and incubated overnight at 37 °C. Cells were cooled on ice before harvesting for the plasmid library (pLibrary) extraction using QIAGEN Plasmid Midi Kit. All the initial steps (lysis to neutralization) for plasmid extractions (pTarget, pLibrary and empty plasmid) were performed on ice with minimum mechanical stress to ensure that the plasmid retained its supercoiled state following purification. Plasmid were stored as aliquots that were used for up to 10 freeze-thaw cycles. Three different pLibrary assembly reactions and preparations were used for the three replicates of the high-throughput *in vitro* cleavage assays.

Target plasmids and empty pUC19 were linearized by restriction enzyme digestion using BsaI-HF and nicked using a nicking enzyme Nt.BspQI (New England Biolabs). All restriction digestion and Gibson Assembly reactions were carried out as per the manufacturer's protocols. All sequences were verified by Sanger sequencing (Eurofins Genomics, Kentucky, USA). The topology of the extracted and restriction digested plasmids was verified on an agarose gel before using in cleavage assays.

***In vitro* cleavage assay**

The protocol was adapted from previously described methods (67). Briefly, Cas12a:crRNA complex was formed by incubating Cas12a and crRNA (1:1.5 ratio) in 1X CutSmart buffer (50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 100 µg/ml BSA, pH 7.9) and 5 mM DTT at 37 °C for 10 min. For activator-mediated cleavage assays, Cas12a, crRNA, and dsDNA oligonucleotide activator (1:1.5:1.5) were incubated at 37 °C for 10 min. Cleavage reactions were initiated by mixing the Cas12a complex with pTarget, pLibrary or empty plasmid (150 ng) or M13mp18 ssDNA (250 ng) (NEB) and incubating at 37 °C. 10 µL aliquots were drawn from the reaction at each time point and quenched with phenol-chloroform. The aqueous layer was extracted and separated on a 1% agarose gel via electrophoresis and stained with SYBR safe or RED safe stain for dsDNA and SYBR gold for ssDNA for visualization. Excess crRNA was used in cleavage assays to prevent any crRNA-independent cleavage activity (37, 38). For library and mismatched target plasmid cleavage assays, 100 nM Cas12a and 150 nM crRNA was used. For activator-mediated cleavage assays, 20 nM Cas12a, 30 nM crRNA and 30 nM activator DNA oligonucleotide were used. Concentrations of pLibrary, pTarget and empty pUC19 used were at 150 ng/10 µL (8.6 nM) reaction, and M13mp18 ssDNA was at 250 ng/10 µL (10.6 nM).

**Library preparation for HTS**

The library plasmid cleavage products were run on an agarose gel as described above to separate the cleaved (linear and nicked) and uncleaved (supercoiled) products. The bands from the nicked and supercoiled fractions from various time points were excised and gel purified using QIAquick Gel Extraction Kit (Qiagen). PCR was used to add Nextera Adapters (NEA), followed by another round to add unique indices/barcodes for each sample. Samples were cleaned using QIAquick PCR Purification Kit (Qiagen). Although the plasmid DNA conformation plays a role in PCR quantification (68), the nicked and supercoiled DNA yielded PCR products when using primers that flanked the randomized target region in the pLibraries (see Table S1 for NEA primers). PCR product size and quality were verified using DNA 1000 kit and Agilent 2100 Bioanalyzer. Samples were sent for MiSeq or NextSeq for paired-end reads of 75 or 150 cycles to Iowa State DNA Facility or Admera Health, LLC (New Jersey, USA). 15% PhiX was spiked in.

**HTS analysis**

HTS data were obtained as compressed fastq files and were processed with custom bash scripts (see associated GitHub repository https://github.com/sashital-lab/Cas12a_nickase). A simple workflow of the analysis is described in figure S1. Briefly, the files were renamed based on the sample information (pLibrary name, replicate and Cas12a ortholog), stored in separate folders identified by the library, and the target sequences were extracted. Bash scripts were used for obtaining the counts of the extracted target sequences, determining the number of mismatches, calculating the fractions in each replicate, as well as preparing summary tables for total counts of each mismatched target sequence. Once all the processing was done on the command-line, they were imported into Microsoft Excel for plotting and summarizing.

After imaging the agarose gels with the pLibrary cleavage products (Fig. 1B, S2E), the plasmid bands were quantified with ImageQuant TL (GE Heathcare) or Image J (https://imagej.nih.gov/ij/). Intensities of the band (I) in the uncleaved (supercoiled - SC) and cleaved fractions (nicked - N and linearized - L) were measured. Fraction (FR) cleaved and uncleaved were calculated as follows.

$$Fraction\ cleaved\ (FR_C) = \frac{I_N + I_L}{I_N + I_L + I_{SC}}$$

$$Fraction\ supercoiled\ (uncleaved)(FR_{SC}) = \frac{I_{SC}}{I_N + I_L + I_{SC}}$$

$$Fraction\ nicked\ (FR_N) = \frac{I_N}{I_N + I_L + I_{SC}}$$

$$Fraction\ linearized\ (FR_L) = \frac{I_L}{I_N + I_L + I_{SC}}$$

The $FR_C$, $FR_{SC}$, $FR_N$ or $F_{RL}$ were determined for each of the time points 't' where $I_{SC}$, $I_N$ and $I_L$ were the intensities of the supercoiled (SC), nicked (N) and linearized (L) bands at that time point.

The fraction of target sequences containing 'n' mismatches (MM) ($F_{n\text{-}MM}$) in the pool was calculated.

$$F_{n-MM} = \frac{total\ counts\ of\ sequences\ with\ n\ mismatches}{total\ counts\ of\ all\ sequences\ in\ the\ pool}$$

$F_{n\text{-}MM}$ was normalized to the fraction (FR) of DNA present in the supercoiled or nicked pool at a given time point 't' to generate an estimated abundance (EA) of a given set of sequences at a given timepoint.

$$EA_{n-MM} = (F_{n-MM}\ of\ S\ at\ t) * (FR\ at\ t)$$

These values were plotted against number of mismatches to generate mismatch distribution curves (Fig. S5). The relative abundance (enrichment and/or depletion) (RA) of a sequence 'S' containing 'n' mismatches at each time point 't' compared to the pLibraries in the negative control, (i.e. pLibrary incubated in reaction conditions for the longest time point, labelled as (-) in all the gels) was calculated and plotted versus time to generate normalized curves shown in Fig. 1D – E, S3C - F, S4.

$$RA_S = \frac{EA_{n-MM}\ of\ S\ at\ t}{EA_{n-MM}\ in\ the\ pLibrary}$$

For the heatmaps in Fig. 2, 3, S6 - 9, the estimated abundance (EA) of sequences containing a particular nucleotide (N = A, G, C, T) at a particular position (P = 1 to 20) for target sequences

containing 'n' mismatches was calculated as above. Relative abundance (RA) was calculated by normalizing EA against the fraction of DNA in the original library to eliminate variability in aberrant nicking that may have occurred for individual pLibraries in the negative control.

$$RA_{S-NP} = \frac{EA_{n-MM}\ of\ S\ with\ N\ at\ P\ at\ t}{F_{n-MM}\ in\ the\ pLibrary}$$

Each RA value was normalized to the maximum RA value present in either the supercoiled or nicked pool to scale the relative abundance from 0 to 1 for the heatmaps. For the analysis of target sequences with two mismatches, the sequences with 2 mismatches were extracted. The distance between the two mismatches and the total counts for sequences separated by that distance were determined. The counts were normalized to the number of possible ways the two mismatches can occur (Fig. S10), and the normalized RA was calculated as described above for heatmaps.

**Acknowledgement:** We thank all the former and current members of the Sashital Lab for helpful discussions and suggestions on various aspects of the project. We thank Michael Baker from the DNA Facility for assistance with HTS data collection, and the Protein Facility for providing access to ImageQuant TL, both a branch of the Iowa State University Office of Biotechnology. We also thank Heather S. Lewin and Megan N. O'Donnell from the University Library for helping with the data deposition to DataShare.

**Conflict of interest:** The authors declare that they have no conflicts of interest with the contents of this article.

**Author contributions:** All experiments and HTS sample preparation were performed by K.M. The HTS data extraction was performed by A.S.S and A.J.S. HTS results were analyzed and interpreted by K.M and D.G.S. The manuscript was written by K.M and D.G.S, with input from A.S.S and A.J.S. All authors read the final draft of the manuscript. Funding for this project was secured by D.G.S.

**References:**


1. Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., and Zhang, F. (2015) Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*. **163**, 759–771
2. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007) CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*. **315**, 1709–1712
3. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*. **532**, 517–521
4. Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., and Charpentier, E. (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell*. **172**, 1239–1259
5. Swarts, D. C., and Jinek, M. (2019) Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Molecular Cell*. **73**, 589-600.e4
6. Knott, G. J., and Doudna, J. A. (2018) CRISPR-Cas guides the future of genetic engineering. *Science*. **361**, 866–869
7. Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S., Wu, W. Y., Scott, D. A., Severinov, K., van der Oost, J., and Zhang, F. (2017) Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nature Biotechnology*. **35**, 31–34
8. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., and Zhang, F. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*. **31**, 827–832
9. Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., and Liu, D. R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*. **31**, 839–843
10. Kim, D., Kim, J., Hur, J. K., Been, K. W., Yoon, S., and Kim, J.-S. (2016) Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nature Biotechnology*. **34**, 863–868
11. Kleinstiver, B. P., Tsai, S. Q., Prew, M. S., Nguyen, N. T., Welch, M. M., Lopez, J. M., McCaw, Z. R., Aryee, M. J., and Joung, J. K. (2016) Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nature Biotechnology*. **34**, 869–874
12. Kim, H. K., Song, M., Lee, J., Menon, A. V., Jung, S., Kang, Y.-M., Choi, J. W., Woo, E., Koh, H. C., Nam, J.-W., and Kim, H. (2017) *In vivo* high-throughput profiling of CRISPR–Cpf1 activity. *Nature Methods*. **14**, 153–159
13. Yamano, T., Zetsche, B., Ishitani, R., Zhang, F., Nishimasu, H., and Nureki, O. (2017) Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1. *Molecular Cell*. **67**, 633-645.e3
14. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*. **507**, 62–67
15. Stella, S., Alcón, P., and Montoya, G. (2017) Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature*. 10.1038/nature22398
16. Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E., and Doudna, J. A. (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*. **351**, 867–871
17. Swarts, D. C., and Jinek, M. (2018) Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. *Wiley Interdisciplinary Reviews: RNA*. **9**, e1481
18. Murugan, K., Babu, K., Sundaresan, R., Rajan, R., and Sashital, D. G. (2017) The Revolution Continues: Newly Discovered Systems Expand the CRISPR-Cas Toolkit. *Molecular Cell*. **68**, 15–25

19. Swarts, D. C., Oost, J. van der, and Jinek, M. (2017) Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Molecular Cell*. **66**, 221-233.e4
20. Liu, X., Homma, A., Sayadi, J., Yang, S., Ohashi, J., and Takumi, T. (2016) Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Scientific Reports*. **6**, 19675
21. Jeon, Y., Choi, Y. H., Jang, Y., Yu, J., Goo, J., Lee, G., Jeong, Y. K., Lee, S. H., Kim, I.-S., Kim, J.-S., Jeong, C., Lee, S., and Bae, S. (2018) Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. *Nature Communications*. 10.1038/s41467-018-05245-x
22. Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., and Moineau, S. (2008) Phage Response to CRISPR-Encoded Resistance in Streptococcus thermophilus. *Journal of Bacteriology*. **190**, 1390–1400
23. Tao, P., Wu, X., and Rao, V. (2018) Unexpected evolutionary benefit to phages imparted by bacterial CRISPR-Cas9. *Science Advances*. **4**, eaar4134
24. Chen, J. S., Ma, E., Harrington, L. B., Costa, M. D., Tian, X., Palefsky, J. M., and Doudna, J. A. (2018) CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*. **360**, 436–439
25. Li, S.-Y., Cheng, Q.-X., Liu, J.-K., Nie, X.-Q., Zhao, G.-P., and Wang, J. (2018) CRISPR-Cas12a has both cis - and trans -cleavage activities on single-stranded DNA. *Cell Research*. **28**, 491
26. Stella, S., Mesa, P., Thomsen, J., Paul, B., Alcón, P., Jensen, S. B., Saligram, B., Moses, M. E., Hatzakis, N. S., and Montoya, G. (2018) Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell*. **175**, 1856-1871.e21
27. Singh, D., Mallon, J., Poddar, A., Wang, Y., Tippana, R., Yang, O., Bailey, S., and Ha, T. (2018) Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *PNAS*. **115**, 5444–5449
28. Hinz, J. M., Laughery, M. F., and Wyrick, J. J. (2015) Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. *Biochemistry*. **54**, 7063–7066
29. Isaac, R. S., Jiang, F., Doudna, J. A., Lim, W. A., Narlikar, G. J., and Almeida, R. (2016) Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife*. **5**, e13450
30. Yarrington, R. M., Verma, S., Schwartz, S., Trautman, J. K., and Carroll, D. (2018) Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proceedings of the National Academy of Sciences*. **115**, 9351–9358
31. Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A. J., Le, L. P., Aryee, M. J., and Joung, J. K. (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology*. **33**, 187–197
32. Tsai, S. Q., Nguyen, N. T., Malagon-Lopez, J., Topkar, V. V., Aryee, M. J., and Joung, J. K. (2017) CIRCLE-seq: a highly sensitive *in vitro* screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nature Methods*. **14**, 607–614
33. Strohkendl, I., Saifuddin, F. A., Rybarski, J. R., Finkelstein, I. J., and Russell, R. (2018) Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Molecular Cell*. **71**, 816-824.e3
34. Zhang, L., Sun, R., Yang, M., Peng, S., Cheng, Y., and Chen, C. (2019) Conformational Dynamics and Cleavage Sites of Cas12a Are Modulated by Complementarity between crRNA and DNA. *iScience*. **19**, 492–503
35. Pollard, J. (1998) Pool Design, Complexity, and Purification. *https://molbio.mgh.harvard.edu/szostakweb/protocols/rnapool/rnapool.html*. [online] https://molbio.mgh.harvard.edu/szostakweb/protocols/rnapool/rnapool.html (Accessed April 17, 2019)
36. Oppenheim, A. (1981) Separation of closed circular DNA from linear DNA by electrophoresis in two dimensions in agarose gels. *Nucleic Acids Res*. **9**, 6805–6812
37. Sundaresan, R., Parameshwaran, H. P., Yogesha, S. D., Keilbarth, M. W., and Rajan, R. (2017) RNA-Independent DNA Cleavage Activities of Cas9 and Cas12a. *Cell Reports*. **21**, 3728–3739

38. Li, B., Yan, J., Zhang, Y., Li, W., Zeng, C., Zhao, W., Hou, X., Zhang, C., and Dong, Y. (2020) CRISPR-Cas12a Possesses Unconventional DNase Activity that Can Be Inactivated by Synthetic Oligonucleotides. *Molecular Therapy - Nucleic Acids*. **19**, 1043–1052
39. Fu, B. X. H., Smith, J. D., Fuchs, R. T., Mabuchi, M., Curcuru, J., Robb, G. B., and Fire, A. Z. (2019) Target-dependent nickase activities of the CRISPR–Cas nucleases Cpf1 and Cas9. *Nature Microbiology*. 10.1038/s41564-019-0382-0
40. Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., and Zhang, F. (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science*. **360**, 439–444
41. Andersson, A. F., and Banfield, J. F. (2008) Virus Population Dynamics and Acquired Virus Resistance in Natural Microbial Communities. *Science*. **320**, 1047–1050
42. Fuchs, R. T., Curcuru, J., Mabuchi, M., Yourik, P., and Robb, G. B. (2019) Cas12a trans-cleavage can be modulated in vitro and is active on ssDNA, dsDNA, and RNA. *bioRxiv*. 10.1101/600890
43. Sternberg, S. H., LaFrance, B., Kaplan, M., and Doudna, J. A. (2015) Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature*. **527**, 110–113
44. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. **337**, 816–821
45. Huston, N. C., Tycko, J., Tillotson, E. L., Wilson, C. J., Myer, V. E., Jayaram, H., and Steinberg, B. E. (2019) Identification of Guide-Intrinsic Determinants of Cas9 Specificity. *The CRISPR Journal*. **2**, 172–185
46. van Aelst, K., Martínez-Santiago, C. J., Cross, S. J., and Szczelkun, M. D. (2019) The Effect of DNA Topology on Observed Rates of R-Loop Formation and DNA Strand Cleavage by CRISPR Cas12a. *Genes*. **10**, 169
47. Andreatta, D., Sen, S., Pérez Lustres, J. L., Kovalenko, S. A., Ernsting, N. P., Murphy, C. J., Coleman, R. S., and Berg, M. A. (2006) Ultrafast Dynamics in DNA: "Fraying" at the End of the Helix. *J. Am. Chem. Soc.* **128**, 6885–6892
48. Fei, J., and Ha, T. (2013) Watching DNA breath one molecule at a time. *Proc Natl Acad Sci U S A*. **110**, 17173–17174
49. Tang, X., Liu, G., Zhou, J., Ren, Q., You, Q., Tian, L., Xin, X., Zhong, Z., Liu, B., Zheng, X., Zhang, D., Malzahn, A., Gong, Z., Qi, Y., Zhang, T., and Zhang, Y. (2018) A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biology*. 10.1186/s13059-018-1458-5
50. Moon, S. B., Lee, J. M., Kang, J. G., Lee, N.-E., Ha, D.-I., Kim, D. Y., Kim, S. H., Yoo, K., Kim, D., Ko, J.-H., and Kim, Y.-S. (2018) Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA with a uridinylate-rich 3′-overhang. *Nature Communications*. **9**, 3651
51. Kuzminov, A. (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks. *PNAS*. **98**, 8241–8246
52. Vriend, L. E. M., Prakash, R., Chen, C.-C., Vanoli, F., Cavallo, F., Zhang, Y., Jasin, M., and Krawczyk, P. M. (2016) Distinct genetic control of homologous recombination repair of Cas9-induced double-strand breaks, nicks and paired nicks. *Nucleic Acids Res*. **44**, 5204–5217
53. Vriend, L. E. M., and Krawczyk, P. M. (2017) Nick-initiated homologous recombination: Protecting the genome, one strand at a time. *DNA Repair*. **50**, 1–13
54. Fukui, K. (2010) DNA Mismatch Repair in Eukaryotes and Bacteria. *Journal of Nucleic Acids*. 10.4061/2010/260512
55. Newton, M. D., Taylor, B. J., Driessen, R. P. C., Roos, L., Cvetesic, N., Allyjaun, S., Lenhard, B., Cuomo, M. E., and Rueda, D. S. (2019) DNA stretching induces Cas9 off-target activity. *Nature Structural & Molecular Biology*. **26**, 185
56. Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., Cheng, A. W., Trevino, A. E., Konermann, S., Chen, S., Jaenisch, R., Zhang, F., and Sharp, P. A. (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology*. **32**, 670–676

57. Zhang, X., Wang, J., Cheng, Q., Zheng, X., Zhao, G., and Wang, J. (2017) Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discovery*. **3**, 17018
58. Cao, B., Wu, X., Zhou, J., Wu, H., DeMott, M. S., Gu, C., Wang, L., You, D., and Dedon, P. C. (2019) *Nick-seq for single-nucleotide resolution genomic maps of DNA modifications and damage*, Molecular Biology, 10.1101/845768
59. Elacqua, J. J., Ranu, N., Dilorio, S. E., and Blainey, P. C. (2019) *NickSeq for genome-wide strand-specific identification of DNA single-strand break sites with single nucleotide resolution*, Genomics, 10.1101/867937
60. Yan, W. X., Hunnewell, P., Alfonse, L. E., Carte, J. M., Keston-Smith, E., Sothiselvam, S., Garrity, A. J., Chong, S., Makarova, K. S., Koonin, E. V., Cheng, D. R., and Scott, D. A. (2019) Functionally diverse type V CRISPR-Cas systems. *Science*. **363**, 88–91
61. Strecker, J., Jones, S., Koopal, B., Schmid-Burgk, J., Zetsche, B., Gao, L., Makarova, K. S., Koonin, E. V., and Zhang, F. (2019) Engineering of CRISPR-Cas12b for human genome editing. *Nature Communications*. **10**, 212
62. McMahon, S. A., Zhu, W., Graham, S., Rambo, R., White, M. F., and Gloster, T. M. (2019) Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *bioRxiv*. 10.1101/784280
63. Meeske, A. J., Nakandakari-Higa, S., and Marraffini, L. A. (2019) Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature*. 10.1038/s41586-019-1257-5
64. Koonin, E. V., and Zhang, F. (2017) Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. *BioEssays*. **39**, e201600186
65. Chen, H., Choi, J., and Bailey, S. (2014) Cut Site Selection by the Two Nuclease Domains of the Cas9 RNA-guided Endonuclease. *J Biol Chem*. **289**, 13284–13294
66. Mohanraju, P., Oost, J., Jinek, M., and Swarts, D. (2018) Heterologous Expression and Purification of the CRISPR-Cas12a/Cpf1 Protein. *BIO-PROTOCOL*. 10.21769/BioProtoc.2842
67. Anders, C., and Jinek, M. (2014) In Vitro Enzymology of Cas9. in *Methods in Enzymology*, pp. 1–20, Elsevier, **546**, 1–20
68. Lin, C.-H., Chen, Y.-C., and Pan, T.-M. (2011) Quantification Bias Caused by Plasmid DNA Conformation in Quantitative Real-Time PCR Assay. *PLOS ONE*. **6**, e29101

**FOOTNOTES**

DATA AVAILABILITY: HTS data and processed data files from this study have been deposited in the Iowa State University Library's DataShare, and can be found at https://doi.org/10.25380/iastate.8178938. HTS data were processed with custom bash scripts which can be found at the GitHub repository https://github.com/sashital-lab/Cas12a_nickase.
All other information and data are available from the authors upon request.

FUNDING: This work was supported by funds from startup funds to D.G.S from Iowa State University College of Liberal Arts and Sciences and the Roy J. Carver Charitable Trust, the National Science Foundation [grant number 1652661] to D.G.S, and National Institute of Food and Agriculture [grant number IOW05480] to D.G.S.

The abbreviations used are: CRISPR, clustered regularly interspaced short palindromic repeat; Cas, CRISPR associated; Sp, *Streptococcus pyogenes*; Fn, *Francisella novicida*; Lb, *Lachnospiraceae bacterium*; As, *Acidaminococcus sp*., crRNA, CRISPR RNA; PAM, protospacer adjacent motif; MM, mismatch; HTS, high-throughput sequencing; nSC, negatively supercoiled, n, nicked; li, linear

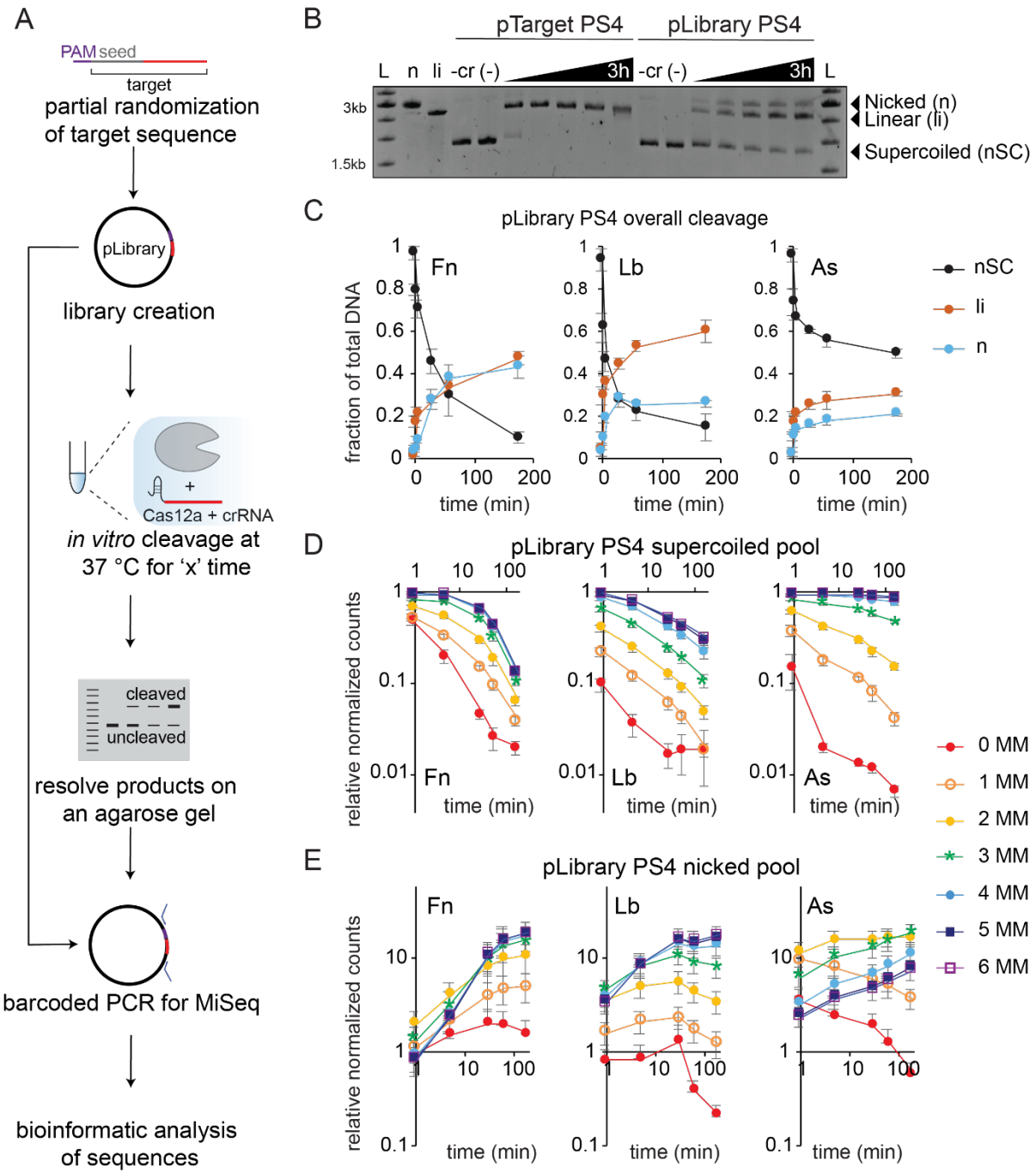


**Figure 1**. High-throughput in vitro analysis of Cas12a mismatch tolerance.
(A) Outline and workflow of the high-throughput in vitro cleavage assay.
(B) Representative agarose gel showing time course cleavage of negatively supercoiled (nSC) plasmid containing a fully matched target (pTarget, left) and plasmid library (pLibrary, right) pLibrary PS4 by LbCas12a, resulting in linear (li) and/or nicked (n) products. Time points at which the samples were collected are 1 min, 5 min, 30 min, 1 h, and 3 h.
All controls were performed under the same conditions as the longest time point for the experimental samples. Controls: -cr = pTarget or pLibrary incubated with Cas12a only at 37 °C for the longest time point

in the assay (3 h); (-) = pTarget or pLibrary alone incubated at 37 °C for the longest time point in the assay (3 h); n = Nt.BspQI nicked pUC19; li = BsaI-HF linearized pUC19
(C) Overall cleavage of the pLibrary by Cas12a indicating the decrease in supercoiled (nSC) pool and appearance of nicked (n) and linear (li) pools over time. The 0 time point is quantification of the (-) control shown in (B) and Fig. S2E. Error bars are SD, n = 2 for LbCas12a, n = 3 for FnCas12a and AsCas12a.
(D, E) Normalized counts relative to the negative control (-) for targets sequences with different number of mismatches (MM) plotted against time in a log scale for pLibrary PS4 in the (D) supercoiled and (E) nicked pool for different Cas12a orthologs. Fn = FnCas12a, Lb = LbCas12a, As = AsCas12a. Error bars are propagation of SEM, n = 2 for LbCas12a, n = 3 for FnCas12a and AsCas12a.

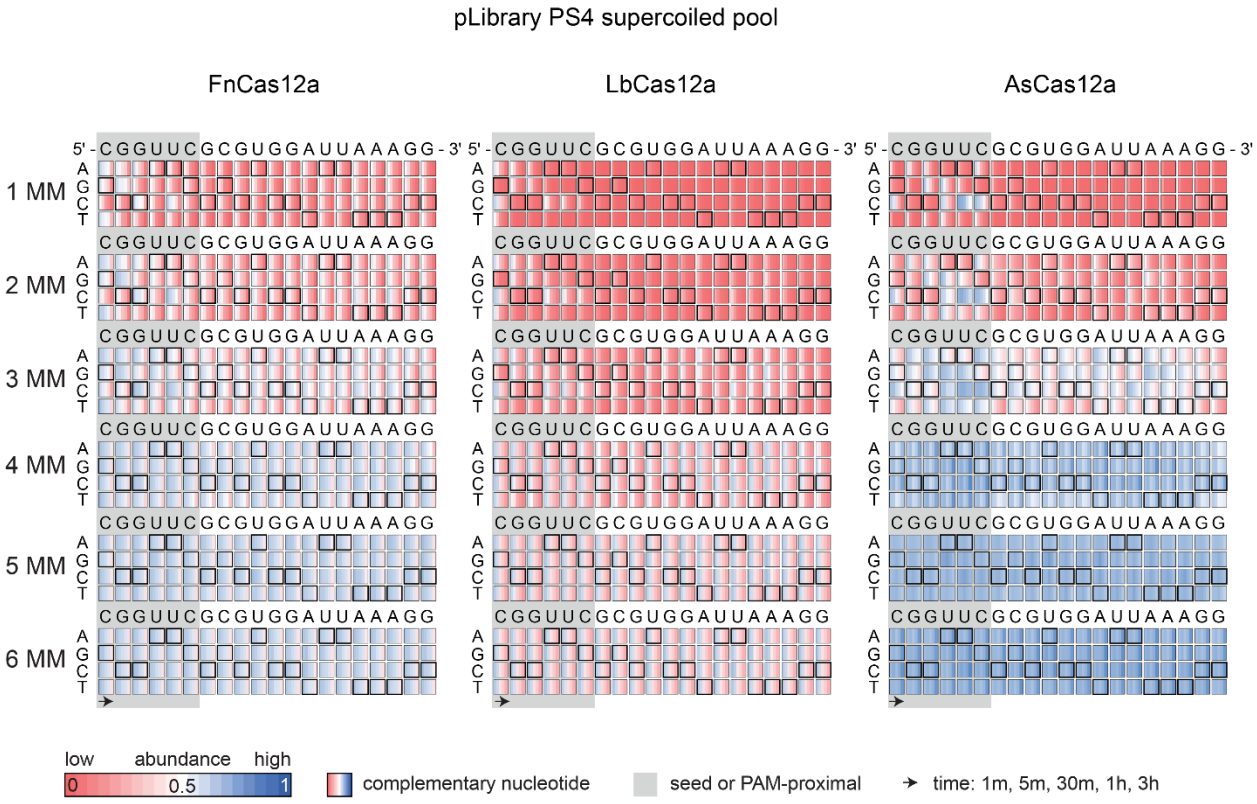


**Figure 2**. Sequence determinants of Cas12a cleavage activity for pLibrary PS4.
Heatmaps showing the relative abundance of different mismatched sequences over time for the supercoiled pool in pLibrary PS4 upon cleavage by Cas12a orthologs. The crRNA sequence is indicated on the top. The nucleotides on the left side of the heatmaps indicate the potential base pair or mismatches formed. The crRNA complementary nucleotides are highlighted by bold black boxes in the heatmap which result in Watson–Crick base pairs. The PAM-proximal "seed" sequence is highlighted by the grey box. Each box represents normalized proportion of sequences containing each nucleotide at a given position across time. Values plotted represent average of two replicates for LbCas12a and three replicates for FnCas12a and AsCas12a. MM = mismatch.

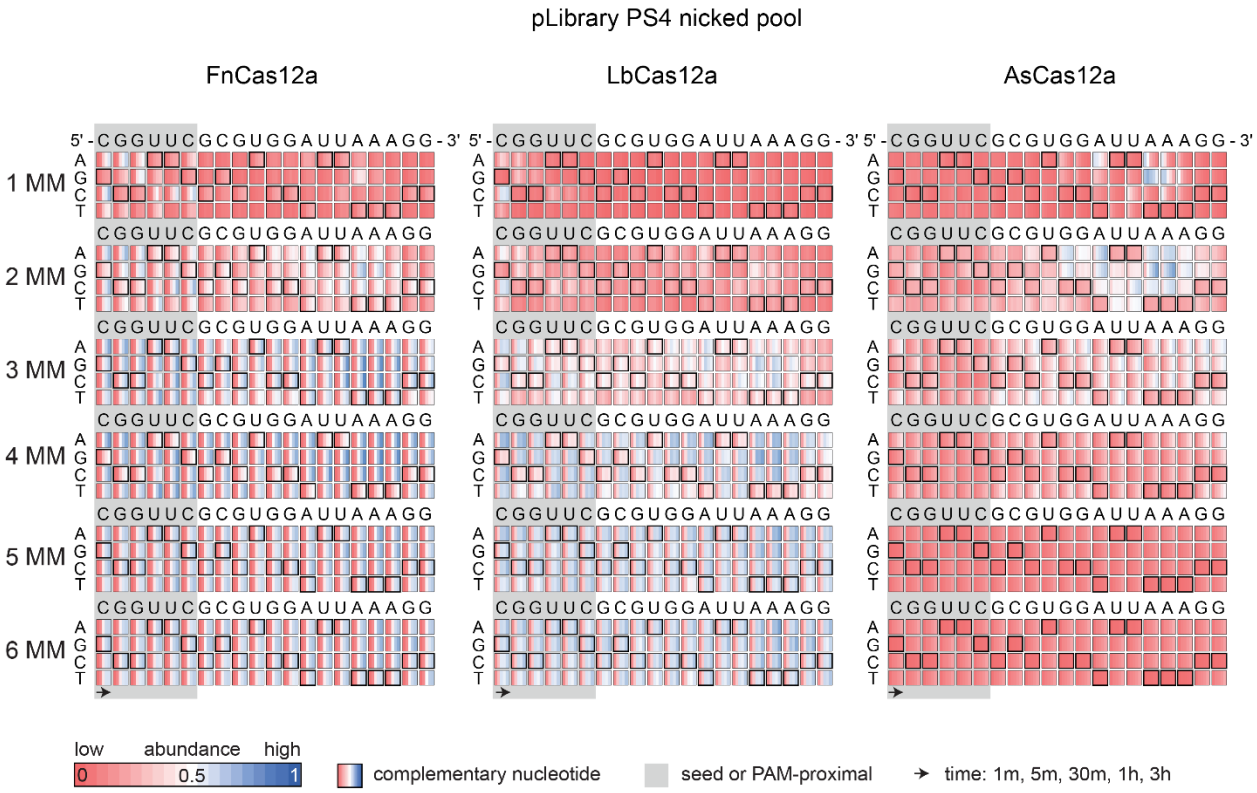


**Figure 3**. Cas12a has pervasive nicking activity against mismatched sequences in pLibrary PS4. Heatmaps showing the relative abundance of different mismatched sequences over time for the nicked pool in pLibrary PS4 upon cleavage by Cas12a orthologs. The crRNA sequence is indicated on the top. The nucleotides on the left side of the heatmaps indicate the potential base pair or mismatches formed. The crRNA complementary nucleotides are highlighted by bold black boxes in the heatmap which result in Watson–Crick base pairs. The PAM-proximal "seed" sequence is highlighted by the grey box. Each box represents the normalized proportion of sequences containing each nucleotide at a given position across time. Values plotted represent average of two replicates for LbCas12a and three replicates for FnCas12a and AsCas12a. MM = mismatch.

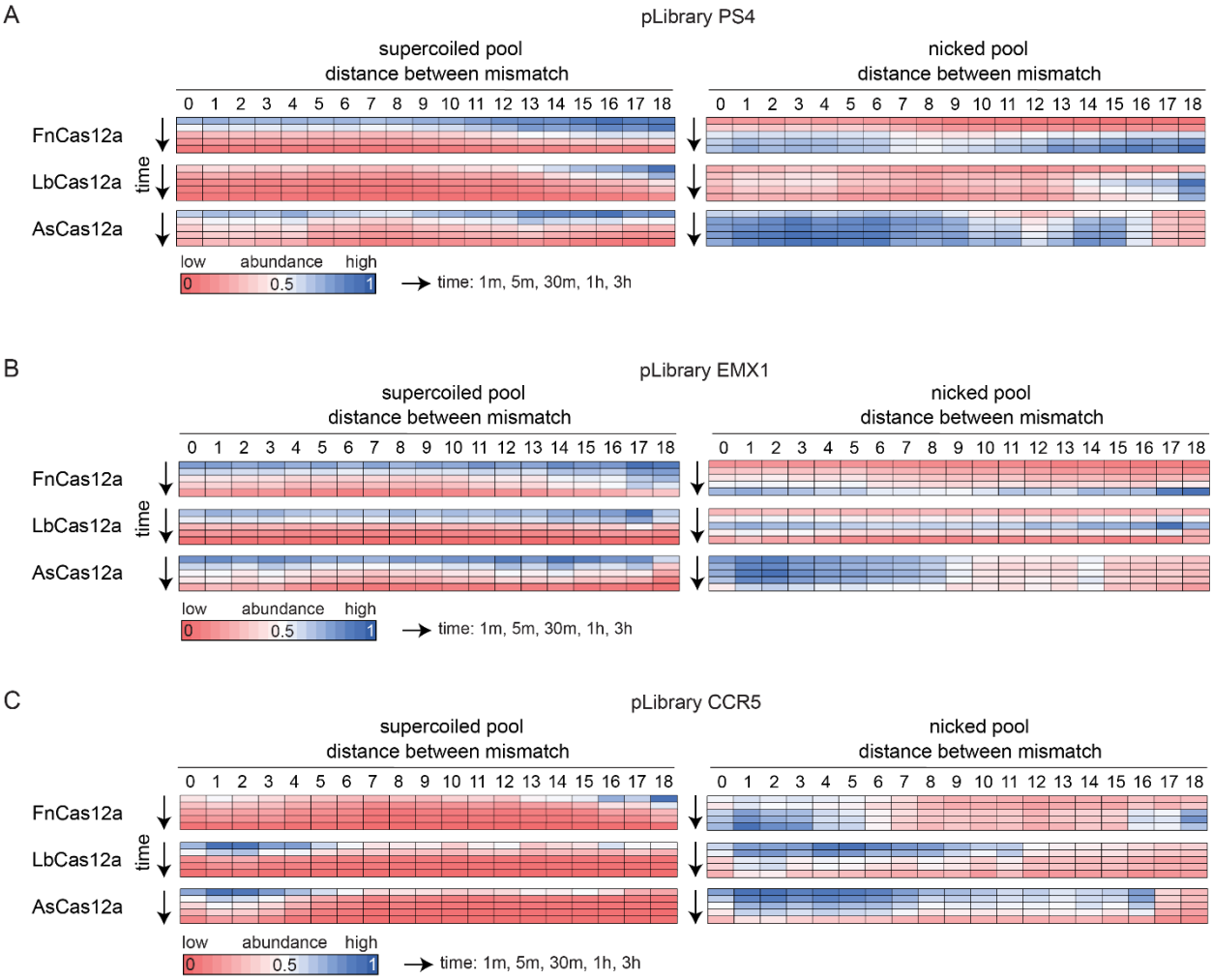


**Figure 4**. Effect of double mismatches in the target sequence on Cas12a cleavage activity.
Heatmaps showing the relative abundance of target sequences with two mismatches over time for the supercoiled (left) and nicked (right) pools in pLibrary (A) PS4, (B) EMX1 and (C) CCR5 upon cleavage by Cas12a as a function of distance between the two mismatches and time. Time points indicated on the left by the arrow are 1 min, 5 min, 30 min, 1 h, and 3 h. Values plotted represent average of two (for LbCas12a, pLibrary PS4) and three (all other Cas12a and libraries) replicates.

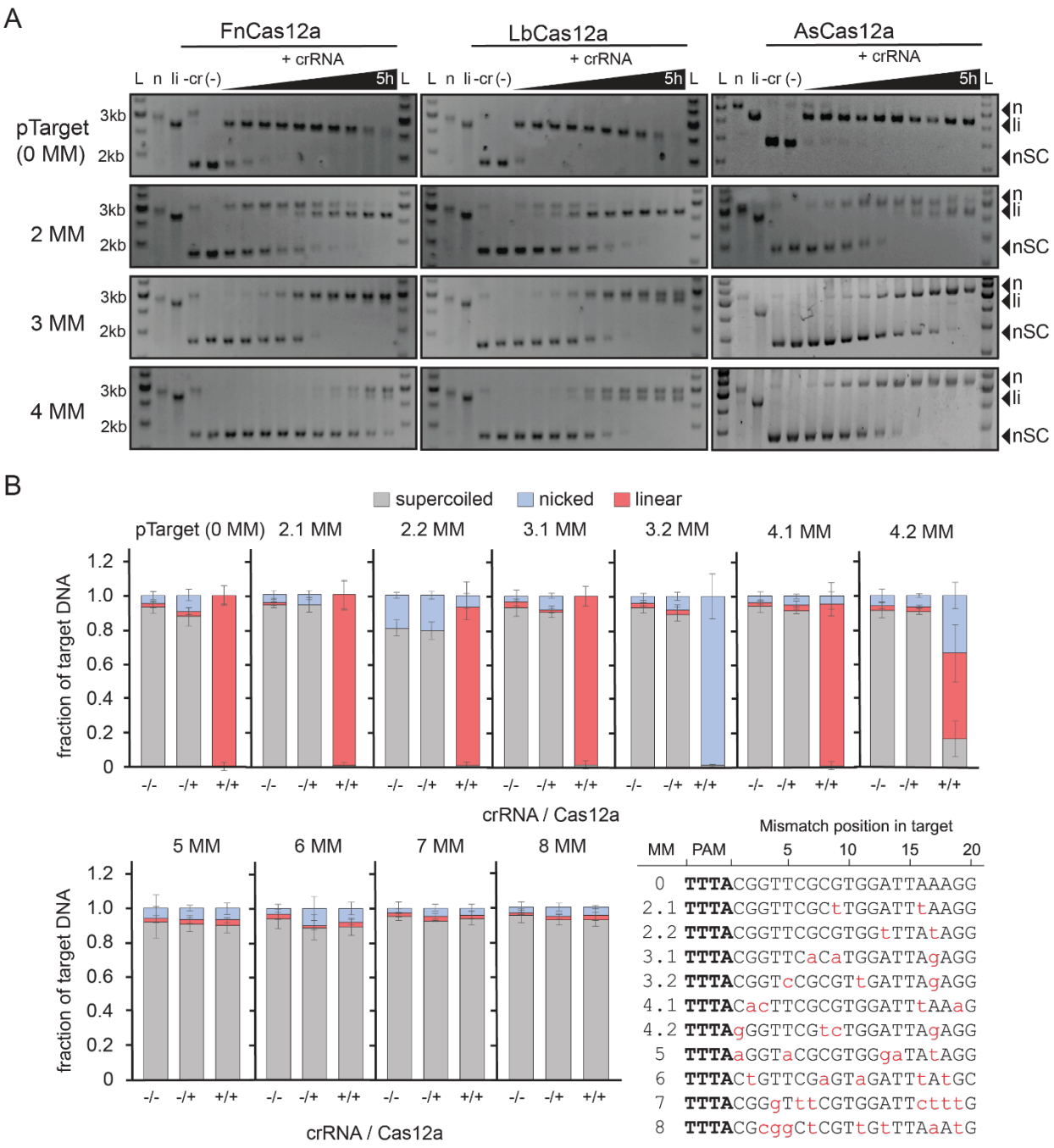


**Figure 5**. Cas12a orthologs have distinct nicking patterns against mismatched targets.
(A) Representative agarose gels showing cleavage of a negatively supercoiled (nSC) plasmid containing the perfect target (pTarget) or mismatched (MM) target over a time course by Cas12a orthologs, FnCas12a (left), LbCas12a (center) and AsCas12a (right), resulting in linear (li) and/or nicked (n) products. Time points at which the samples were collected are 15 sec, 30 sec, 1 min, 2 min, 5 min, 15 min, 30 min, 1 h, 3 h, and 5 h.
All controls were performed under the same conditions as the longest time point for the experimental samples. Controls: -cr = target plasmid incubated with Cas12a only at 37 °C for the longest time point in the assay (5 h); (-) = target plasmid alone incubated at 37 °C for the longest time point in the assay (5 h); n = Nt.BspQI nicked pUC19; li = BsaI-HF linearized pUC19

(B) Quantification of supercoiled, linear and nicked fractions from cleavage of perfect or fully crRNA-complementary and mismatched (MM) target plasmid by LbCas12a after 3 hours. The different target sequences tested are listed where the PAM is in bold and mismatches are in lowercase and red.
-/- indicates a cleavage reaction with the target plasmid without Cas12a and crRNA. -/+ indicates a cleavage reaction with the target plasmid and Cas12a only, and +/+ indicates a cleavage reaction with the target plasmid, Cas12a and cognate crRNA. Average of the intensity fraction values are plotted with SD as error bars, n = 3 replicates.

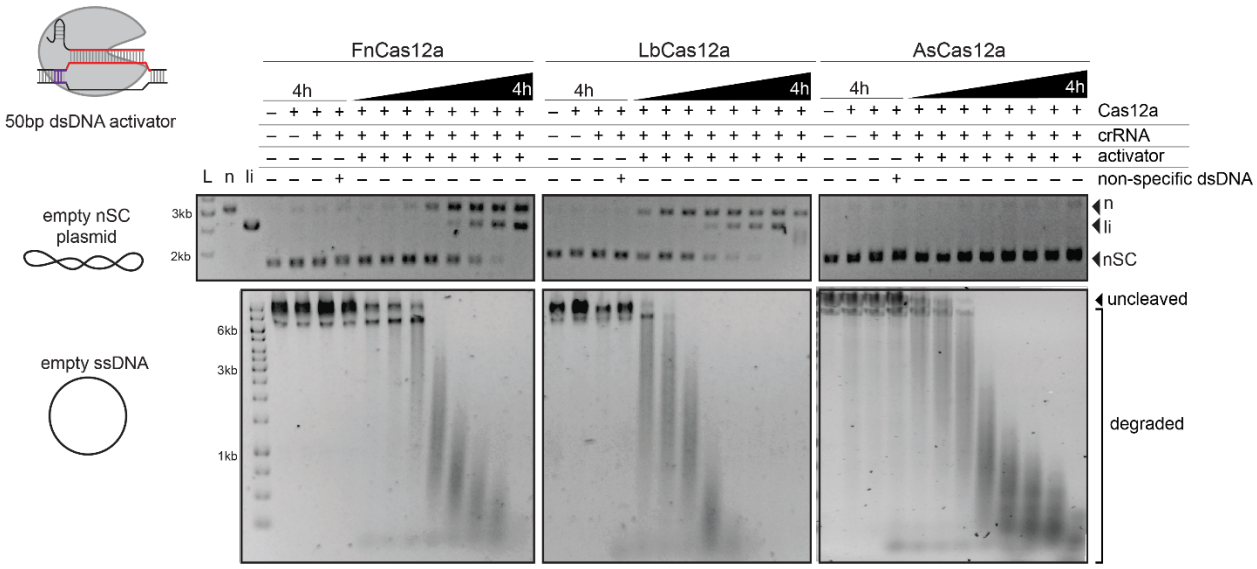


**Figure 6**. FnCas12a and LbCas12a can be activated for non-specific, trans dsDNA nicking
Representative agarose gels showing non-specific nicking and linearization of negatively supercoiled (nSC) dsDNA plasmid, and degradation of linearized dsDNA plasmid (top) and degradation of ssDNA (bottom) over time by Cas12a. Cas12a (20 nM) and crRNA (30 nM) were complexed with a crRNA-complementary 50 bp dsDNA activator (30 nM) (perfect target from pLibrary PS4) and incubated with empty negatively supercoiled (nSC) dsDNA (top) and ssDNA (bottom). Time points at which the samples were collected are 15 sec, 30 sec, 1 min, 5 min, 15 min, 30 min, 1 h and 4 h. All controls were performed under the same conditions as the longest time point for the experimental samples. Control reactions were incubated at 37 °C for the longest time point in the assay (4 h); n = Nt.BspQI nicked pUC19; li = BsaI-HF linearized pUC19

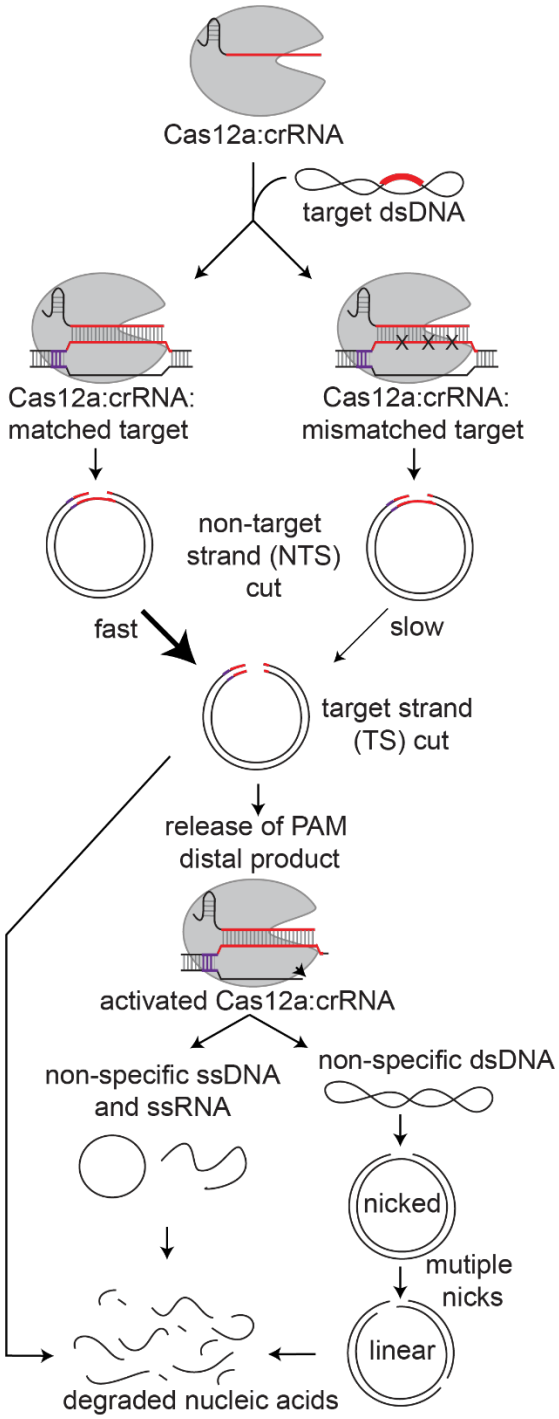


**Figure 7**. Cas12a has cis and activated, trans nuclease activities.
Cas12a-crRNA complex can bind complementary target dsDNA and cleave it. Some mismatched targets can be nicked rapidly by Cas12a but undergo slow linearization. After this cleavage event, Cas12a releases the PAM distal cleavage product and remains bound to at least the PAM proximal cleaved target-strand, thus remaining in an activated confirmation. The active RuvC domain can then accept other nucleic acid substrates. Target-activated Cas12a can further nick, linearize and degrade non-specific dsDNA, ssDNA and ssRNA substrates.