SUPPLEMENTARY INFORMATION

Characterization of Cas12a nucleases reveals distinct PAM profiles between closely-related orthologs

Thomas Jacobsen, Fani Ttofali, Chunyu Liao, Srinivas Manchalu,

Benjamin N. Gray, and Chase L. Beisel*

*Correspondence to chase.beisel@helmholtz-hiri.de (C.L.B.)

TABLE OF CONTENTS

Supplementary Figures 2

Supplementary Tables 14

Supplementary Data 16

Supplementary Documents 17

SUPPLEMENTARY FIGURES

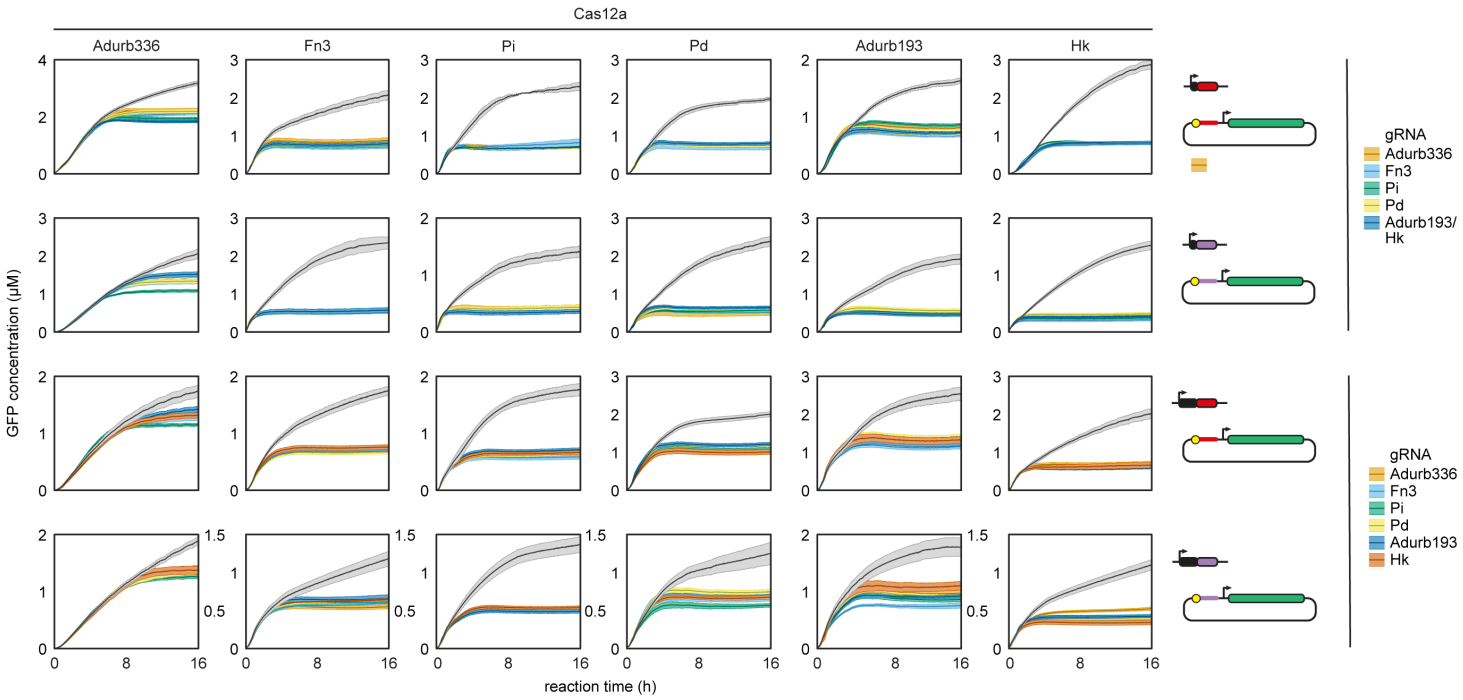


Figure S1. Time-series of GFP expression in TXTL reactions containing a Cas12a, targeting or non-targeting gRNA, and a GFP reporter containing a protospacer targeted by the gRNA. For this experiment, we tested both full-length (top two rows) and processed (bottoms two rows) gRNA that targeted two protospacers (red and purple) flanked by a TTTC PAM. See Figure 1E for more information. Note that the Adurb193Cas12a and HkCas12a share the same processed gRNA. Each solid line and shaded region represent the mean and standard deviation from three separate TXTL reactions. Gray lines reflect experiments with a non-targeting gRNA. End-point values are reported in Figure 1E.

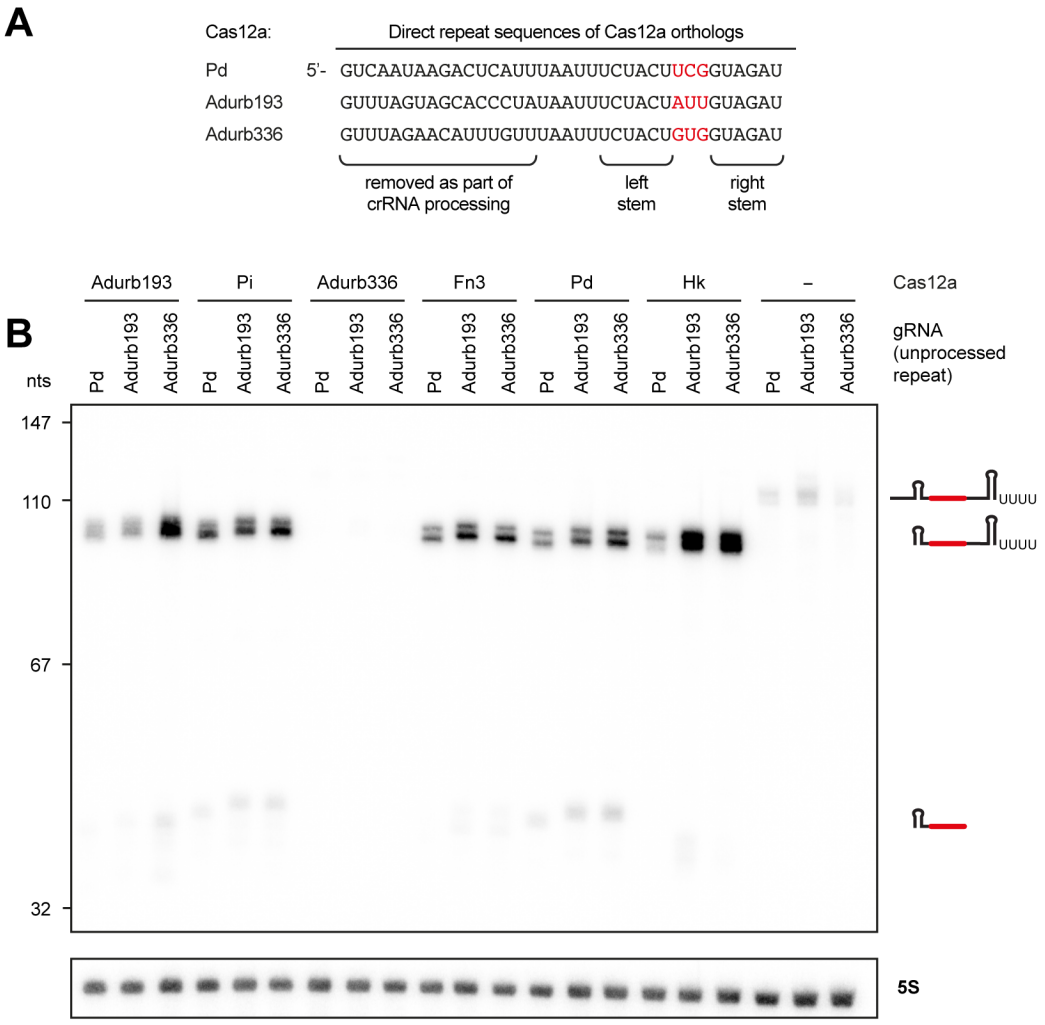


Figure S2. The Cas12a nucleases process different gRNAs with full-length repeats in TXTL. **(A)** Repeats used. The three repeats were selected because they exhibit the highest sequence diversity across the set of repeats in this work. **(B)** Northern blotting analysis of gRNA processing in TXTL. Each nuclease was co-expressed with one of the three gRNAs with unprocessed repeats in TXTL, and the extracted RNAs were subjected to Northern blotting analysis. The expected sizes are shown for the full-length gRNA, the gRNA processed at the 5' end, and a gRNA with the 3' terminator removed—presumably through non-specific ribonucleases. 5S RNA was probed as a loading control. Tick marks on the left designate the locations of an RNA ladder run with the samples. The image is representative of two independent experiments.

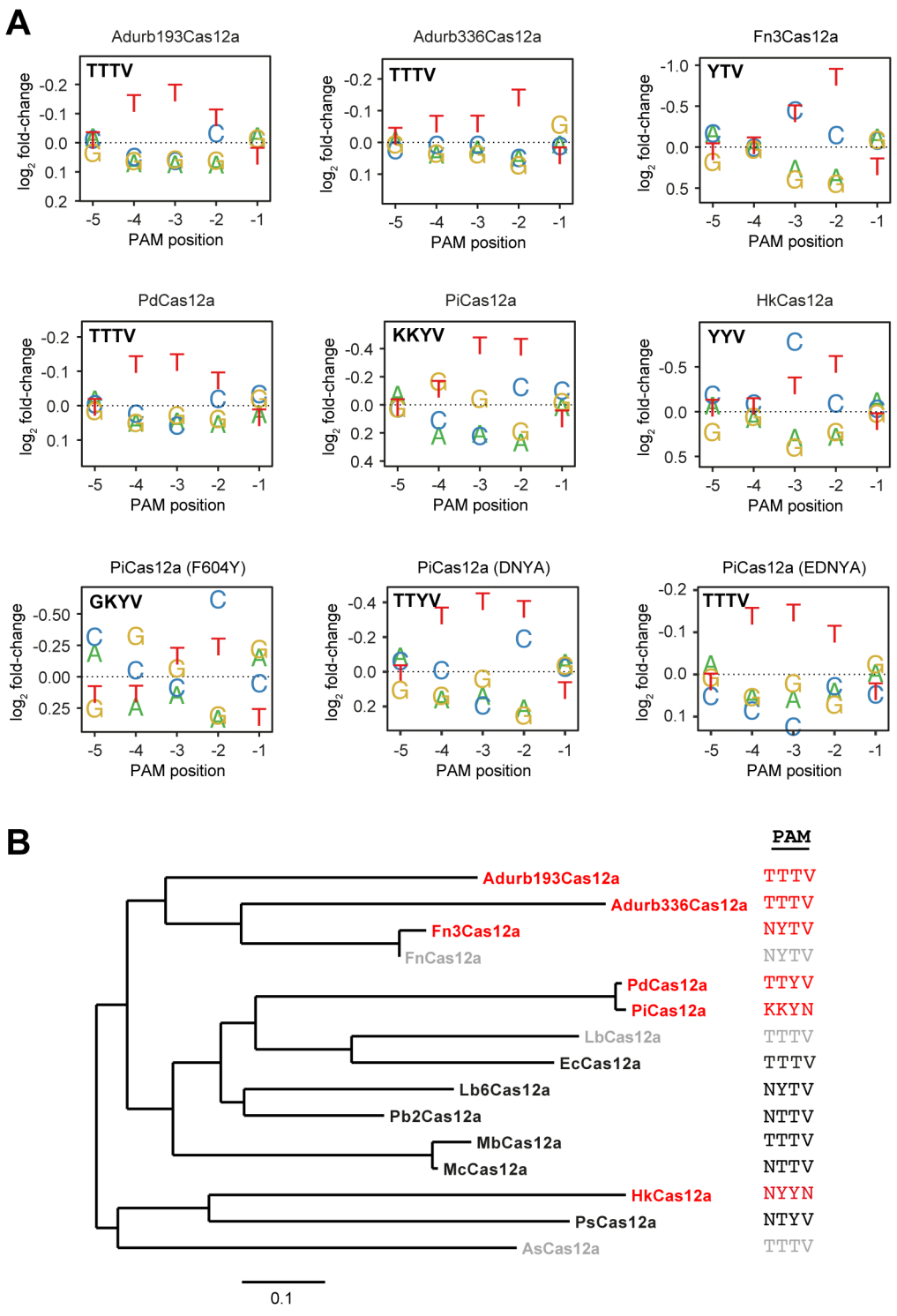


Figure S3. Comparison of elucidated PAMs for different Cas12a orthologs. **(A)** Fold-change plots for the Cas12a orthologs subjected to the TXTL-based PAM screen. Plots are rescaled from those in Figure 2B. See Figure 2B for more information. **(B)** PAM profiles do not trend with phylogeny. Phylogenetic tree and consensus PAMs of Cas12a nucleases. Red indicates nucleases

investigated in this study, black indicates nucleases previously investigated in our prior work (Marshall et al. *Mol Cell* 2018), and gray indicates the well-characterized and commonly used nucleases from *Francisella novicida* U112 (FnCas12a), *Lachnospiraceae* bacterium ND2006 (LbCas12a), and *Acidaminococcus* sp. BV3L6 (AsCas12a).

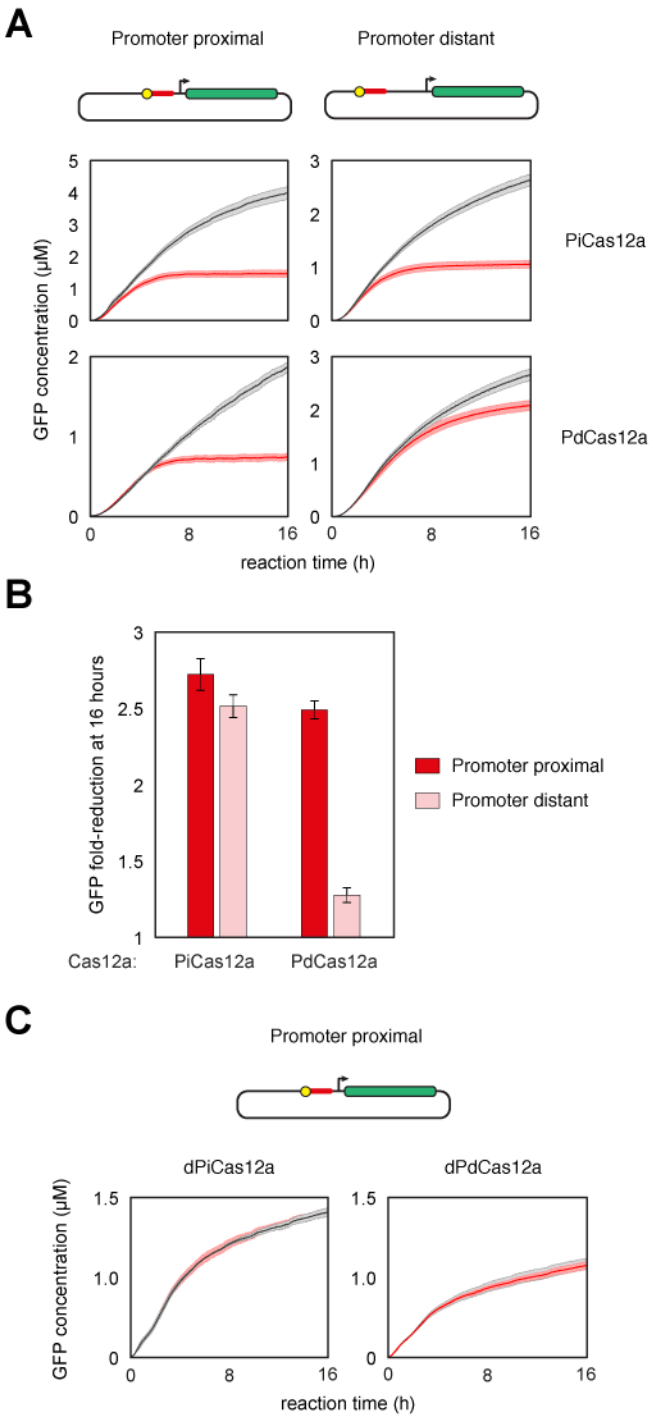


Figure S4. PdCas12a exhibits efficient transcriptional repression but poor DNA cleavage in TXTL. **(A)** Time-series of GFP expression in TXTL reactions containing PiCas12a or PdCas12a, a targeting (red) or non-targeting (gray) gRNA, and a GFP reporter containing the protospacer flanked by a 5' TTTG PAM. The protospacer was located immediately upstream (Promoter proximal) or 100 bps upstream (Promoter distant) of the -35 element of the promoter driving

deGFP expression. Each solid line and shaded region represent the mean and standard deviation from three separate TXTL reactions. **(B)** Fold-reduction of GFP expression from the time-series data in A at 16 hours. Error bars represent the standard deviation from three separate TXTL reactions. **(C)** Targeting the promoter-proximal target with dPiCas12a or dPdCas12a. Each solid line and shaded region represent the mean and standard deviation from three separate TXTL reactions.

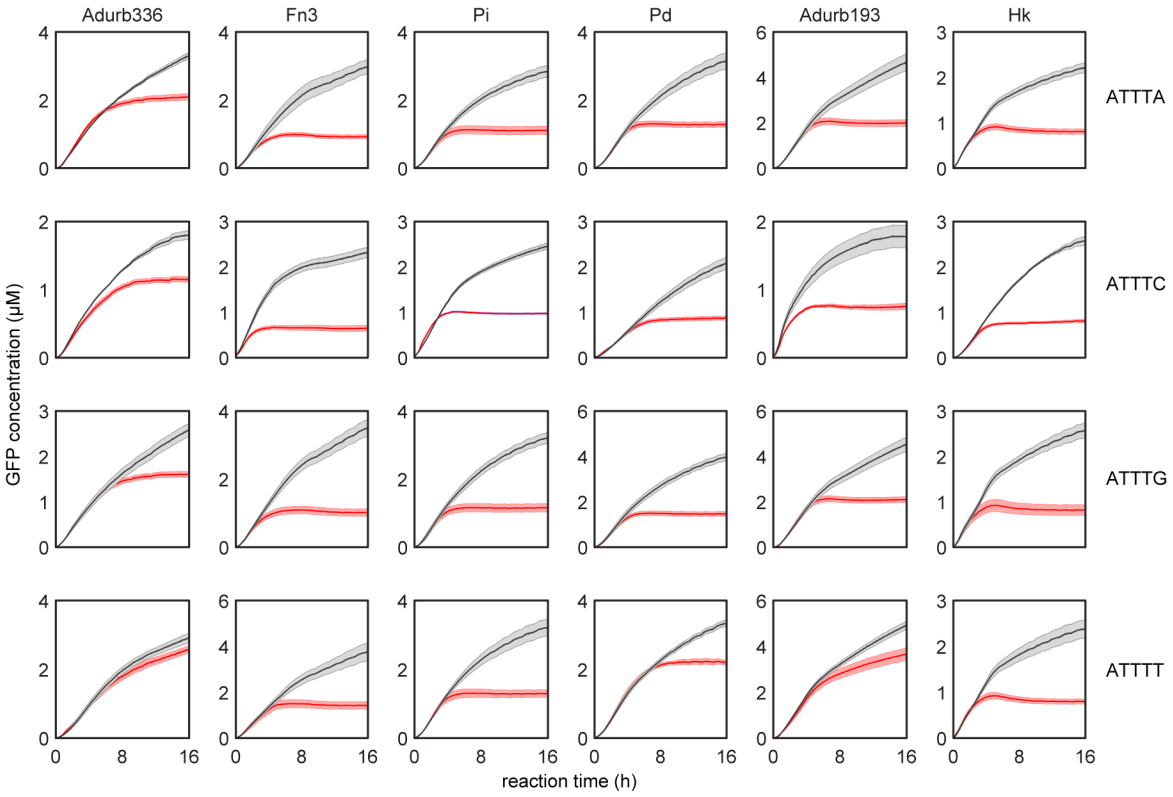


Figure S5. Time-series of GFP expression in TXTL reactions containing a Cas12a, targeting (red) or non-targeting (gray) gRNA, and a GFP reporter containing a protospacer targeted by the gRNA. Each protospacer was flanked on its 5' end by various non-canonical motifs (labeled on rows). Each solid line and shaded region represent the mean and standard deviation from three separate TXTL reactions. End-point values are reported in Figure 3.

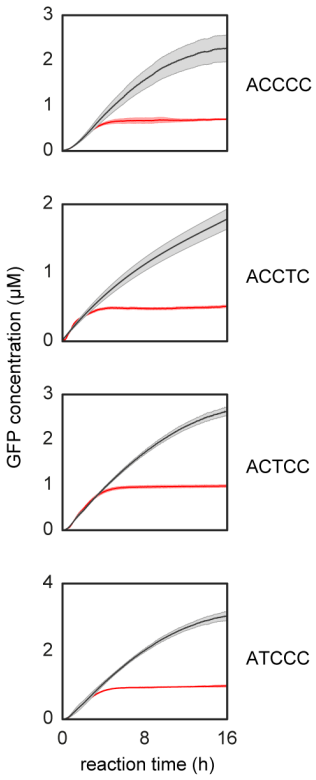


Figure S6. Time-series of GFP expression in TXTL reactions containing HkCas12a, targeting (red) or non-targeting (gray) gRNA, and a GFP reporter containing a protospacer targeted by the gRNA. Each protospacer was flanked on its 5' end by various non-canonical motifs (labeled on rows). Each solid line and shaded region represent the mean and standard deviation from three separate TXTL reactions. End-point values are reported in Figure 4A.

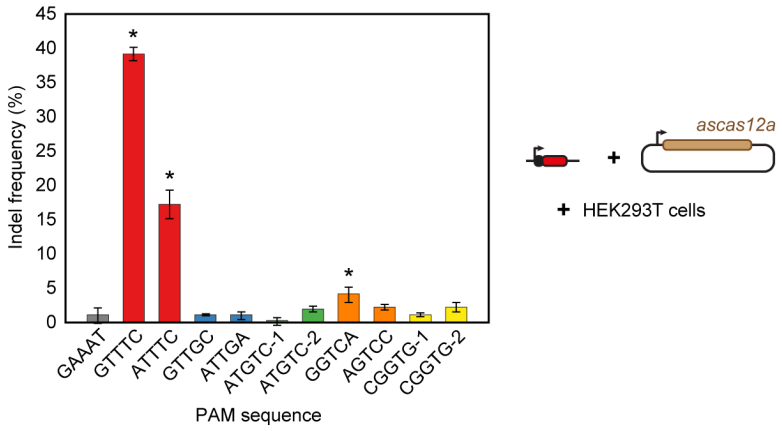


Figure S7. Indel formation in HEK293T cells using AsCas12a and different PAM-flanking sites. A plasmid encoding AsCas12a and linear DNA encoding a gRNA targeting a different location in the DNMT1 gene were transiently transfected, and indel formation was quantified by TIDE analysis. See Figure 7C for the corresponding target locations. The indel frequency data for each target sequence were compared to that of a target sequence flanked by a non-PAM (TAGCT). Error bars represent the standard deviation from three independent transfections. Asterisks (*) indicate indel frequencies that were significantly higher for a PAM than for the non-PAM control (GAAAT) ($p < 0.05$, $n = 3$). Colors correspond to those in Figure 7D.

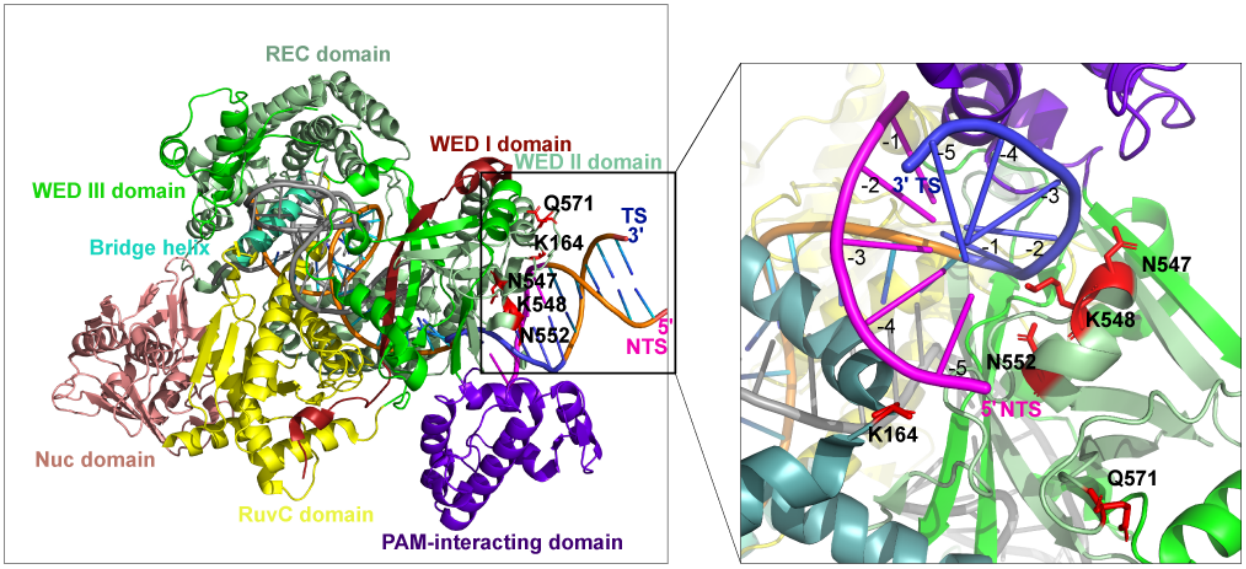


Figure S8. Structural view of the residues in the crystal structure of AsCas12a (PDB: 5B43) that are equivalent to the mutated residues in PiCas12a (K163, N599, S600, F604, T628). Left: the corresponding residues in AsCpf1 (K164, N547, K548, N552, Q571, respectively) are shown in red (left). Right: a close-up view of these residues in proximity with the PAM nucleotides. TS; DNA Targeting strand and NTS; non-targeting strand. PAM nucleotides on the TS and NTS are in blue and magenta, respectively. Numbering designate the position relative to the target sequence. K164 interacts with the two alpha helices in the REC domain and is adjacent to the phosphate backbone of the NTS at the -5 base. N547 and K548 interact with the -2 and -3 bases on the TS, respectively. N552 does not directly contact with the PAM, although a separate crystal of the RVR variant of AsCas12a (PDB: 5XH6) shows that R552 creates a new interaction with the -3 base of the NTS. Q571 interacts with two alpha helices in the WED 2 domain. See the original publication reporting the 5B43 crystal structure (Yamano et al. *Mol. Cell* 2017) for more information on the annotated domains. Structural images were generated using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

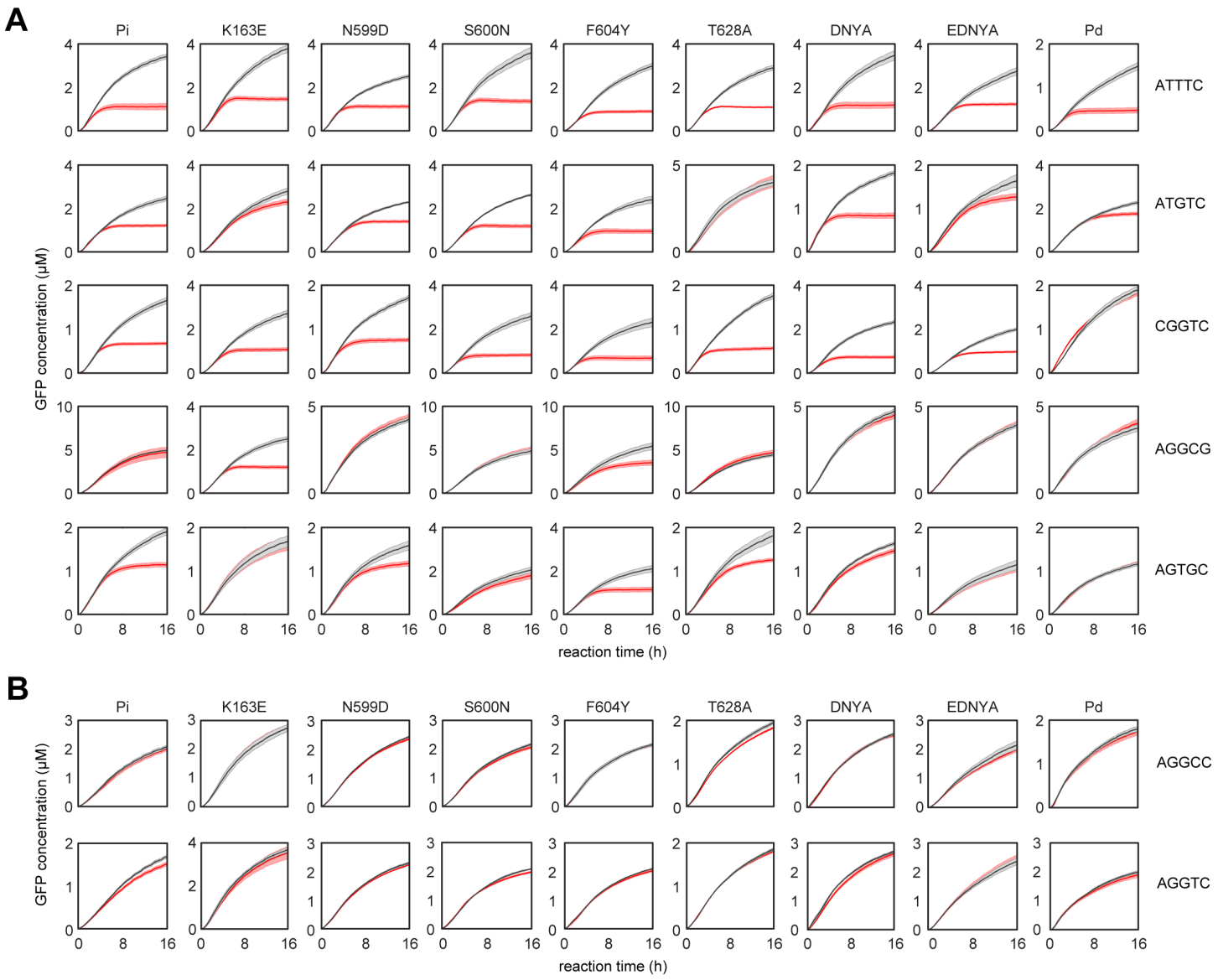


Figure S9. Time-series of GFP expression in TXTL reactions containing a PiCas12a variant (labeled on top), a targeting (red) or non-targeting (gray) gRNA, and a GFP reporter containing the protospacer flanked by various non-canonical motifs (labeled on rows). (**A**) PAM sequences with end-point values reported in Figure 5B. (**B**) PAM sequences with no observable loss of GFP expression across all tested Cas12a variants. Each solid line and shaded region represent the mean and standard deviation from three separate TXTL reactions.

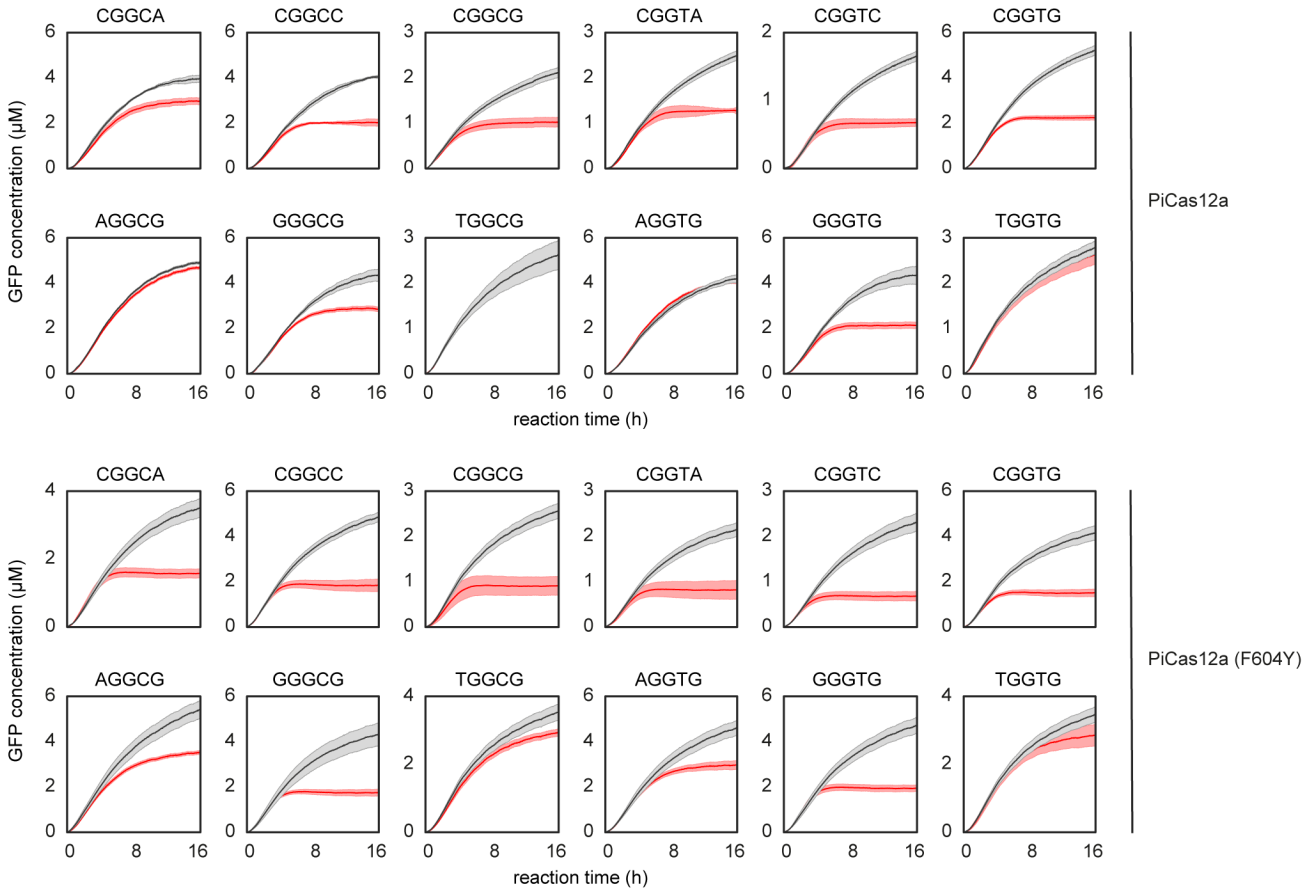


Figure S10. Time-series of GFP expression in TXTL reactions containing PiCas12a or the F604Y variant, a targeting (red) or non-targeting (gray) gRNA, and a GFP reporter containing a protospacer targeted by the gRNA. Each protospacer was flanked on its 5' end by various non-canonical motifs (labeled on top of each plot). Each solid line and shaded region represent the mean and standard deviation from three separate TXTL reactions. End-point values are reported in Figure 7B.

SUPPLEMENTARY TABLES

Table S1. DNA constructs used in this work. See the associated excel file Table S1.xlsx.

Table S2. Target sequences and flanking PAMs used in this work.

	5' PAM	Target sequence
PAMs and target sequence(s) used in DNA cleavage assay	ATTTA	GTCGAGTGCAAAACCTTTCGCGGT
	ATTTC	
	ATTTG	
	ATTTT	
	ACCCC	
	ACCTC	
	ACTCC	
	ATCCC	
	ATGTC	
	CGGTC	
	AGGCG	
	AGTGC	
	AGGCC	
	AGGTC	
	CGGCA	
	CGGCC	
	CGGCG	
	CGGTA	
	CGGTG	
	GGGCG	
	TGGCG	
	AGGTG	
	GGGTG	
	TGGTG	
PAMs and target sequence(s) used in orthogonality testing	ATTTC	GTCGAGTGCAAAACCTTTCGCGGT
	TTTTC	ACTGGCGTTGTTCCCATCCTGGTC
PAMs and target sequence(s) used in *DNMT1* indel formation assay	CGGTC	CTGATGGTCCATGTCTGTTACTC
	CGGCG	CACTCATACAGTGGTAGATTTGA
	CGGTG	AGGGCCAGCCCAGCAGCCAACCT
	AGGTG	TGTTTCTGGCACCAGGAATCCCC
	GGGTG	TCAAGTGGCGTGACACCGGGCGT
	CGGCA	ACACAACAGCTTCATGTCAGCCA
	CGGCC	GCAGGCACCTGCCTCAGCTGCTC
	AGGCG	TACAGTGGTAGATTTGATATAAT
	GGGCG	TCACTCCTGCTCGGTGAATTTGG
	TGGCG	GAATGCACAAAGTACTGCACAAT
	TGGTG	ACGTGTCAAGTGCTTAGAGCAGG
	AGTGC	CATGTTAAAAACACAACATCAGT
	GTTTG	TGTTACTCGCCTGTCAAGTGGCG
	ATTTC	AGCCAAGGCCACAAACACCATGT
	CGTTG	GGGATTCCTGGTGCCAGAAACAG
	TTTTC	CCTTAGCACTCTGCCACTTATTG
	GGTCA	GCTGTTAACATCAGTACGTTAAT
	GTTTC	TTAGCAGCTTCCTCCTCCTTTAT

Table S3. List of PAM sequences that were unable to be assessed as part of the TXTL-based PAM screen with Fn3Cas12a. Sequences in black were absent in the targeted PAM library, while sequences in blue were absent from the targeted and non-targeted libraries. The number of reads was generally low in both libraries. All PAM sequences were present for all other Cas12a variants subjected to the screen.

AAAAC	CGAAC
AAATG	CGACC
AACCC	CTAAT
AACGC	CTATA
AACTA	CTCAC
AATCG	CTCCT
AATTA	CTCGG
AATTG	CTCTA
AATTT	CTTTC
ACACG	GAACG
ACATT	GAGCA
ACCGC	GAGCT
ACCGG	GATCA
ACCTA	GATTA
ACCTG	GCCCA
ACTTA	GCCGT
AGACC	GCCTG
AGATT	GCTCA
AGCGC	GCTTG
AGCGT	GTCCT
AGGGC	GTTCC
AGTTA	GTTTC
ATATA	TAACT
ATCCC	TAATA
ATCTA	TACCA
ATGCG	TACCG
CACCA	TACTA
CACTA	TCCCA
CAGCG	TCCCT
CAGCT	TCGCC
CATGC	TCTTA
CATTC	TGCAC
CCAGT	TGCCC
CCCCT	TGCTA
CCCGA	TTTCG
CCGCG	TTTTC

SUPPLEMENTARY DATA

Data S1. Set of Krona plots for all TXTL-based PAM screens. See the associated zip file Data S1.zip.

SUPPLEMENTARY DOCUMENTS

Document S1. Alignment of protein sequence for the tested Cas12a variants.

```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Hk              -----MFEKLSNIVSISKTIRFKLI-PVGKTLENIEK------LGKLEKDFERSDFYPIL
As              ---MTQFEGFTNLYQVSKTLRFELI-PQGKTLKHIQE------QGFIEEDKARNDHYKEL
Pd              ---MENYQEFTNLFQLNKTLRFELK-PIGKTCELLEEGKIFASGSFLEKDKVRADNVSYV
Pi              MKVMENYQEFTNLFQLNKTLRFELK-PIGKTCELLEEGKIFASGSFLEKDKVRADNVSYV
Adurb193        --MSSVFSDFTNQYQLTKTLRFELK-PVGKTLEYIKE------NGLLSDDKLRAENYTKI
Adurb336        ---------MTNKYSLSKTLRFELKKPIGKTQEFIEK------NGLIEEDEQRNKDFLCG
Fn3             ---MSIYQEFVNKYSLSKTLRFELI-PQGKTLENIKA------RGLILDDEKRAKDYKKA
                         : *  .:.**:**:*  * *** : ::        . : .*  * .

Hk              KNISDDYYRQYIKEKLSDLNLDWQKLYDAHELLDSSKKES-QKNLEMIQAQYRKVLFNIL
As              KPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYF
Pd              KKEIDKKHKIFIEETLSSFSISNDLLKQYFDCYNELKAFK-KDCKSDEEEVKKTALRNKC
Pi              KKEIDKKHKIFIEETLSSFSISNDLLKQYFDCYNELKAFK-KDCKSDEEEVKKTALRNKC
Adurb193        KDIIDDFHKDFIDKSLKQADIRWDDLKNGIENYRKDNSKKNKDELEKLQKRYRNEIVECF
Adurb336        KELLEGYYSDLIERRLGKIEIDSNLLEEYFKKYKEFNKLK-QKGKEVDAKELKNS-KDSL
Fn3             KQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSD-DDNLQKDFKSAKDTIKKQI
                *   :  :    .  *    :  : *    .   . :  .  .         .    .

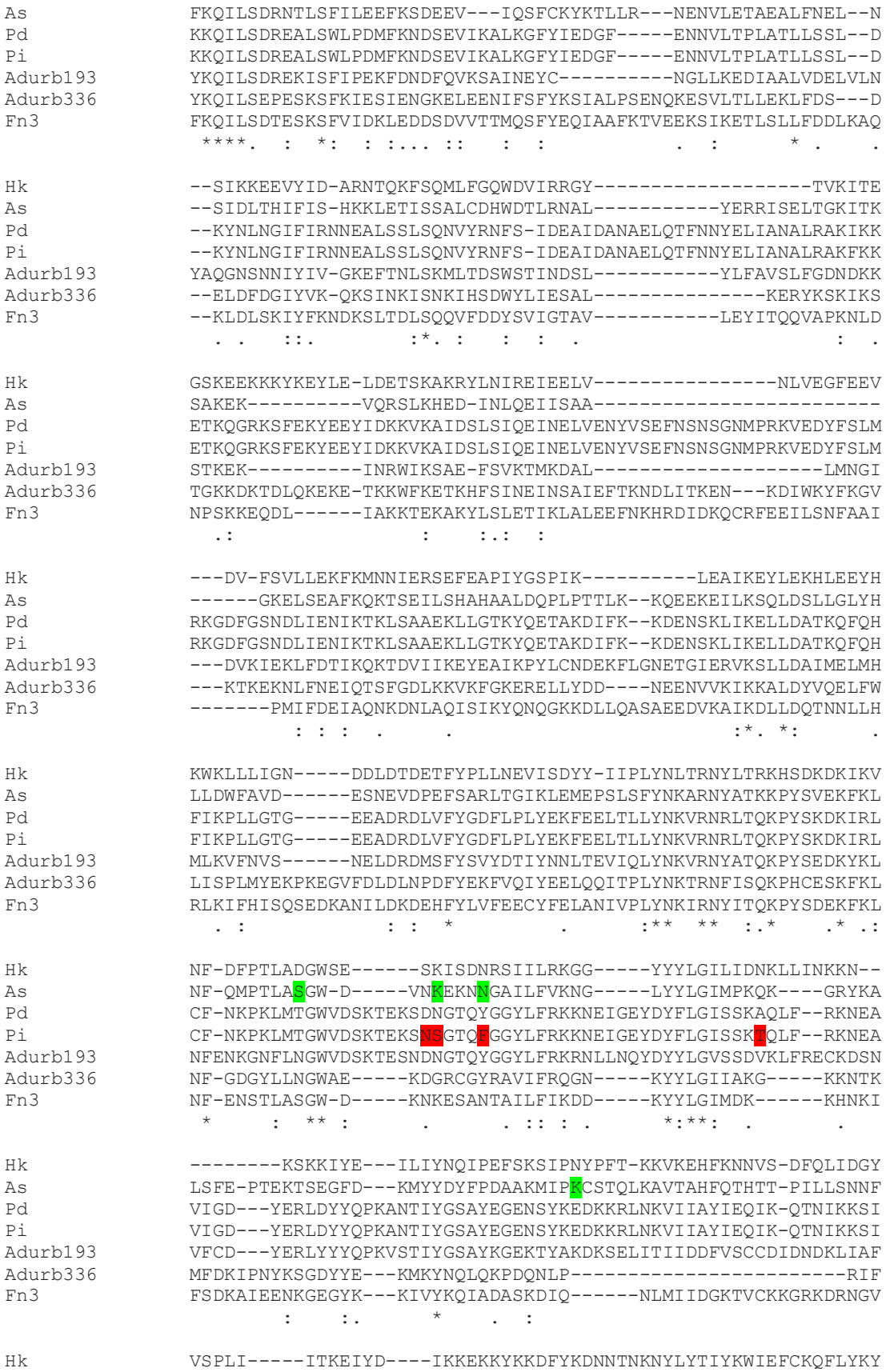
Hk              S---GELDKSGEKNSKDLIKN------------NK-----ALYGKLFKKQFILEVLPDFV
As              I---GRTDNLTDAINKRHAEI------------YK--------GLFKAELFNGKVLKQLG
Pd              T---SIQRAMREAISQAFLKS------------PQ------------KKLLAIKNL----
Pi              T---SIQRAMREAISQAFLKS------------PQ------------KKLLAIKNL----
Adurb193        K-------KDKEIFS-----------------------------YLFKKELITEILPEYI
Adurb336        K---ASQRKLREELHKSFFDNDALFKDEEPFKEKRDNEDVIIAAVKFHRKKLLEKRNEGG
Fn3             SKYINDSEKFKNLFNQNLIDA------------KKGQESDLILWLKQSKDNGIELFKANS
                          :                                        :

Hk              NNNDSYSEEDLEGLNLYSKF-----TTRLKNFWETRKNVFTDKDIVTAIPFRAVNENFGF
As              TVTTTEHENAL--LRSFDKF-----TTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPK
Pd              -IENVFKADEN--VQHFSEF-----TSYFSGFETNRENFYSDEEKSTSIAYRLVHDNLPI
Pi              -IENVFKADEN--VQHFSKF-----TSYFSGFETNRENFYSDEEKSTSIAYRLVHDNLPI
Adurb193        KNKETFEAERD-ILSSFSRF-----ATYFTSFQETRKNIYSKNAISTAISYRLVHQNFPK
Adurb336        VDNNLEIEEDK--INCLYRFVNDHCTTYFTKFFNNRKNVFSDKAISTSIFYRTIHENLHF
Fn3             DITDIDEALEI--IKSFKGW-----TTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPK
                             :     :     :: :. *  .*:*.:: :   *:* .* : :*:

Hk              YYDNIKIFNKNIEYLENK--IPNLENELKEADILDD----------NRSVKDYFTPNGFN
As              FKENCHIFTRLITAVP---SLREHFENVKKAIGIFV----------STSIEEVFSFPFYN
Pd              FIKNIYIFEKLKEQFDAK-TLSEIFENYKLYVA-------------GSSLDEVFSLEYFN
Pi              FIKNIYIFEKLKEQFDAK-TLSEIFENYKLYVA-------------GSSLDEVFSLEYFN
Adurb193        FISNIEVFNKISEVCADI--ILDTERELAEILD-------------GRRLRDVFSIEYYN
Adurb336        FVKNIENFEKLSEIIDEK-ELLEIEKELQKELG-------------MFNLKEVFNINYFN
Fn3             FLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFN
                : .*   :              :   :                       : : *    :*

Hk              YVITQDGIDVYQAIRGG-FTKENG-EKVQGINEILNLTQQQLRR-KPETKNVKLGVLTKL
As              QLLTQTQIDLYNQLLGG-ISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPH-RFIPL
Pd              NTLTQKGIDNYNAVIGK-IVKEDK-QEIQGLNEHINLYNQKHKD-----RRLPF--FISL
Pi              NTLTQKGIDNYNAVIGK-IVKEDK-QEIQGLNEHINLYNQKHKD-----RRLPF--FISL
Adurb193        NVLCQSGIDDFNRCIGG-YSKDDG-VKIRGLNEFVNLFNQKNNG-------LMLSKMTPL
Adurb336        NCLNQKGIDKYNQIISG-VKKKD--IQKEGLNNLINKLSQKKDKDQEERKNIRKLRMIQL
Fn3             NYLNQSGITKFNTIIGGKFVNGEN-TKRKGINEYINLYSQQIND-----KTLKKYKMSVL
                  : *  *  ::   .    .     :  *:*: :*   *:          :    :  *

Hk              RKQILEYSESTSFLIDQIEDDNDLVDRINKFNVSF-FESTEVSPSLFEQIERLYNAL--K
```


As FKQILSDRNTLSFILEEFKSDEEV---IQSFCKYKTLLR---NENVLETAEALFNEL--N
Pd KKQILSDREALSWLPDMFKNDSEVIKALKGFYIEDGF-----ENNVLTPLATLLSSL--D
Pi KKQILSDREALSWLPDMFKNDSEVIKALKGFYIEDGF-----ENNVLTPLATLLSSL--D
Adurb193 YKQILSDREKISFIPEKFDNDFQVKSAINEYC----------NGLLKEDIAALVDELVLN
Adurb336 YKQILSEPESKSFKIESIENGKELEENIFSFYKSIALPSENQKESVLTLLEKLFDS---D
Fn3 FKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
Hk --SIKKEEVYID-ARNTQKFSQMLFGQWDVIRRGY-------------------TVKITE
As --SIDLTHIFIS-HKKLETISSALCDHWDTLRNAL-----------YERRISELTGKITK
Pd --KYNLNGIFIRNNEALSSLSQNVYRNFS-IDEAIDANAELQTFNNYELIANALRAKIKK
Pi --KYNLNGIFIRNNEALSSLSQNVYRNFS-IDEAIDANAELQTFNNYELIANALRAKFKK
Adurb193 YAQGNSNNIYIV-GKEFTNLSKMLTDSWSTINDSL-----------YLFAVSLFGDNDKK
Adurb336 --ELDFDGIYVK-QKSINKISNKIHSDWYLIESAL---------------KERYKSKIKS
Fn3 --KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAV-----------LEYITQQVAPKNLD
Hk GSKEEKKKYKEYLE-LDETSKAKRYLNIREIEELV----------------NLVEGFEEV
As SAKEK----------VQRSLKHED-INLQEIISAA--------------------------
Pd ETKQGRKSFEKYEEYIDKKVKAIDSLSIQEINELVENYVSEFNSNSGNMPRKVEDYFSLM
Pi ETKQGRKSFEKYEEYIDKKVKAIDSLSIQEINELVENYVSEFNSNSGNMPRKVEDYFSLM
Adurb193 STKEK----------INRWIKSAE-FSVKTMKDAL--------------------LMNGI
Adurb336 TGKKDKTDLQKEKE-TKKWFKETKHFSINEINSAIEFTKNDLITKEN---KDIWKYFKGV
Fn3 NPSKKEQDL------IAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILSNFAAI
Hk ---DV-FSVLLEKFKMNNIERSEFEAPIYGSPIK----------LEAIKEYLEKHLEEYH
As ------GKELSEAFKQKTSEILSHAHAALDQPLPTTLK--KQEEKEILKSQLDSLLGLYH
Pd RKGDFGSNDLIENIKTKLSAAEKLLGTKYQETAKDIFK--KDENSKLIKELLDATKQFQH
Pi RKGDFGSNDLIENIKTKLSAAEKLLGTKYQETAKDIFK--KDENSKLIKELLDATKQFQH
Adurb193 ---DVKIEKLFDTIKQKTDVIIKEYEAIKPYLCNDEKFLGNETGIERVKSLLDAIMELMH
Adurb336 ---KTKEKNLFNEIQTSFGDLKKVKFGKERELLYDD----NEENVVKIKKALDYVQELFW
Fn3 -------PMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEEDVKAIKDLLDQTNNLLH
Hk KWKLLLIGN-----DDLDTDETFYPLLNEVISDYY-IIPLYNLTRNYLTRKHSDKDKIKV
As LLDWFAVD------ESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKL
Pd FIKPLLGTG-----EEADRDLVFYGDFLPLYEKFEELTLLYNKVRNRLTQKPYSKDKIRL
Pi FIKPLLGTG-----EEADRDLVFYGDFLPLYEKFEELTLLYNKVRNRLTQKPYSKDKIRL
Adurb193 MLKVFNVS------NELDRDMSFYSVYDTIYNNLTEVIQLYNKVRNYATQKPYSEDKYKL
Adurb336 LISPLMYEKPKEGVFDLDLNPDFYEKFVQIYEELQQITPLYNKTRNFISQKPHCESKFKL
Fn3 RLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKL
Hk NF-DFPTLADGWSE------SKISDNRSIILRKGG-----YYYLGILIDNKLLINKKN--
As NF-QMPTLASGW-D-----VNKEKNNGAILFVKNG-----LYYLGIMPKQK----GRYKA
Pd CF-NKPKLMTGWVDSKTEKSDNGTQYGGYLFRKKNEIGEYDYFLGISSKAQLF--RKNEA
Pi CF-NKPKLMTGWVDSKTEKSNSGTQFGGYLFRKKNEIGEYDYFLGISSKTQLF--RKNEA
Adurb193 NFENKGNFLNGWVDSKTESNDNGTQYGGYLFRKRNLLNQYDYYLGVSSDVKLFRECKDSN
Adurb336 NF-GDGYLLNGWAE-----KDGRCGYRAVIFRQGN-----KYYLGIIAKG-----KKNTK
Fn3 NF-ENSTLASGW-D-----KNKESANTAILFIKDD-----KYYLGIMDK------KHNKI
Hk --------KSKKIYE---ILIYNQIPEFSKSIPNYPFT-KKVKEHFKNNVS-DFQLIDGY
As LSFE-PTEKTSEGFD---KMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTT-PILLSNNF
Pd VIGD---YERLDYYQPKANTIYGSAYEGENSYKEDKKRLNKVIIAYIEQIK-QTNIKKSI
Pi VIGD---YERLDYYQPKANTIYGSAYEGENSYKEDKKRLNKVIIAYIEQIK-QTNIKKSI
Adurb193 VFCD---YERLYYYQPKVSTIYGSAYKGEKTYAKDKSELITIIDDFVSCCDIDNDKLIAF
Adurb336 MFDKIPNYKSGDYYE---KMKYNQLQKPDQNLP-----------------------RIF
Fn3 FSDKAIEENKGEGYK---KIVYKQIADASKDIQ------NLMIIDGKTVCKKGRKDRNGV
Hk VSPLI-----ITKEIYD----IKKEKKYKKDFYKDNNTNKNYLYTIYKWIEFCKQFLYKY

```
As              IEPLE-----ITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKY
Pd              IESISKYPNISDDDKVTPSSLLEKIKKVSIDSYNGILSFKSFQSVNKEVIDNLLKTISPL
Pi              IESISKYPNISDDDKVTPSSLLEKIKKVSIDSYNGILSFKSFQSVNKEVIGNLLKTISCL
Adurb193        KEKVE----------LTPNGYINWLKKEYPSIFEKLICNQEFKNKNNEIIGNLKKALKDL
Adurb336        FAP-------SNEKYYNPSEEIKSIKN--------KGSFRTNIEDCHKMIEFYKKSISKK
Fn3             NRQLL---SLKRKHLPENIYRIKETKSYLKN------EARFSRKDLYDFIDYYKDRL---
                                    :  *.           .      . *    . :

Hk              KGPNKESYKEMY-DFS-TLKDTSLYVNLNDFYADVNSCAYR--VLFNKIDENTIDNAVED
As              ------TKTTSI-DLS-SLRPSSQYKDLGEYYAELNPLLYH--ISFQRIAEKEIMDAVET
Pd              K-----NKAEFL-DLI-----NKDYQIFTEVQAVIDEICKQKTFIYFPISNVELEKEMGD
Pi              K-----NKEEFH-DLI-----KKDYQIFTEVQAVIDEICKQKTFIYFPVSNVELEKEMGD
Adurb193        I-----PKMPYVQQYI-----EKQYTIVNEIISDIEEISHNKILTYIKINRREMENALSR
Adurb336        ------SEWDIY-KFK--FKETNDYVSIDEFYKDFSNQAYK--IEFQKISVDDIHELVDS
Fn3             ---------DYY-DFEFELKPSNEYSDFNDFTNHIGSQGYK--LTFENISQDYINSLVNE
                           .        . *  . :    .     .  . :  :    : . :

Hk              G--KLLLFQIYNKDFS---------PESKGKKNLHTLYWLSMFSEENLRTRKLKLNGQAE
As              G--KLYLFQIYNKDFA---------KGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAE
Pd              KDKPLCLFQISNKDLSFAKTFSANLRKKRGAENLHTMLFKALMEGNQ---DNLDL-GSGA
Pi              KDKPLCLFQISNKDLSFAKTFSANLRKKRGAENLHTMLFKALMSGNQ---DNLDL-GSGA
Adurb193        GTKRLYLFKITNKDLSFAESFEKGYRKEIKNKNLHTLYFEQLMSGGN---SVFDL-GTGS
Adurb336        G--KLYLFTIYNKDFSEKKFKK---DFSKRKPNLHTMYWKELFSDENLKNVVYQLNGGAE
Fn3             G--KLYLFQIYSKDFS---------AYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAE
                    * ** * .**::               ****: :  ::.  :      .* * .

Hk              IFYRKK-LEKKPIIHKEGSILLNKIDKEGN-TIPENIYHECYRYLNKKIGREDLSDEAIA
As              LFYRPKSRM-KRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSH-DLSDEARA
Pd              IFYRAKSLDGNKPTHPANEAIKCRNVANKD-K------VSLF------------------
Pi              IFYRAKSLDGNKPTHPANEAIKCRNVANKD-K------VSLF------------------
Adurb193        IFYRPASIE-SPFVHKKGEILIDRKDKNGN-PIPMDIYDQLFNHFNY--GS-TLSEEAKK
Adurb336        IFFRKKSPSNKKIIHIKNKAVPNEKPINGR-K------ESTF------------------
Fn3             LFYRKQSIP-KKITHPAKETIANKNKDNPK-K------ESVF------------------
                :*:*      .   *   . :     :           . :

Hk              LFNKDVLKYKEARFDIIKDRRYSESQFFFHVPITFNWDIKTNK-N-VNQIVQGMIKD-GE
As              L--LPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSK-FNQRVNAYLKE-HP
Pd              ------------TYDIYKNRRYMENKFLFHLSIVQNYKAANDSAQ-LNSSATEYIRK-AD
Pi              ------------TYDIYKNRRYMENKFLFHLSIVQNYKAAKDSAQ-LNSSATEYIRK-AD
Adurb193        Y--KDIVVVKNAKHDIIKDKRFTQDKFLFHLSITINYKATGSG-Y-INAKVLNAISN-NP
Adurb336        ------------EYDLIKDKRYFEDKYLFHCPITLNFKSEKNNVESFNNYINEFIVKNND
Fn3             ------------EYDLIKDKRFTEDKFFFHCPITINFKSSGAN-K-FNDEINLLLKEKAN
                            .:: *:.*: ..:::** .*. *:.       .*      : .

Hk              IKHIIGIDRGERHLLYYSVIDLEGNIVEQGSLNTLEQNRFDNSTVKVDYQNKLRTREEDR
As              ETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQ---------QFDYQKKLDNREKER
Pd              DLHIIGIDRGERNLLYYSVIDMKGNIVEQDSLNIIR----NN-DLETDYHDLLDKREKER
Pi              DLHIIGIDRGERNLLYYSVIDMKGNIVEQDSLNIIR----NN-DLETDYHDLLDKREKER
Adurb193        NVNIIGIDRGERNLLYVSLINQRGDIITQKSFNIAG---------GFDYAAKLTQYEKNR
Adurb336        EINILGLDRGERNLLYYYLLNPNGEKIKSGSFNIFK----DEFGREFDYHKMLSEREDER
Fn3             DVHILSIDRGERHLAYYTLVDGKGNIIKQDNFNIIG----ND-RMKTNYHDKLAAIEKDR
                   *:.:*****:* *  :::  *. : . .:*             :*   *   *.:*

Hk              DRARKNWTNINKIKELKDGYLSHVVHKLSRLIIKYEAIVIMENLNQGFKRGRFKV-ERQV
As              VAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAV
Pd              KANRQNWEAVEGIKDLKKGYLSQAVHQIAQLMLKYNAIIALEDLGQMFVTRGQKI-EKAV
Pi              KANRQNWEAVEGIKDLKKGYLSQAVHQIAQLMLKYNAIIALEDLGEMFVTRGQKI-EKAV
Adurb193        DIARKDWKQIGNIKELKEGYLSVVIHEIIQMMIKNNAIIVMEDLNFGFKRGRFKF-ERQV
Adurb336        DKDQKEWKNIKNIKDLKEGYLSRIIHEISKIVIDNNAIVVLEDLNPDFKRNRFKF-ENQV
Fn3             DSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKV-EKQV
                   .: *  :  **::*.****  :*::  :::  :*:: :*:*.  *      . *. *

Hk              YQKFELALMNKLSALSFKEKYDERKNLEPSGILNPIQACYPVDAYQEL--QGQNGIVFYL
```

```
As              YQQFEKMLIDKLNCLVLK----DYPAEKVGGVLNPYQLTDQFTSFAKM--GTQSGFLFYV
Pd              YQQFEKSLVDKLSYLVDK----KRPYNELGGILKAYQLASSITKNNS---DKQNGFLFYV
Pi              YQQFEKSLVDKLSYLVDK----KRPYNELGGILKAYQLASSITKSNS---DKQNGFLFYV
Adurb193        YQKFEKMLIDKLNYLADK----KKNPADEGGILRGYQLTEAFTSFQKL--GKQSGFIFYV
Adurb336        YQKFEEKLIKKLNYLVFK----NRKSTDSGGLFRAYQLTSKFVSFKNLREQKQSGILYYV
Fn3             YQKLEKMLIEKLNYLVFK----DNEFDKTGGVLRAYQLTAPFETFKKM--GKQTGIIYYV
                **::*  *:.**. *  *    .    . .*::.  *    .    .     *.*:::*:

Hk              PAAYTSVIDPVTGFTNLFRLKSINSSKYEE-FIKKFKNIYFDNEEEDFKFIFNYKD---F
As              PAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSF
Pd              PAWNTSKIDPVTGFTDLLRPKAMTIKEAQD-FFGAFDNISYN-DKGYFEFETNYDK---F
Pi              PAWNTSKIDPVTGFTDLLRPKAMTIKEAQD-FFGAFDNISYN-DKGYFEFETNYDK---F
Adurb193        PAGYTSKIDPTTGFVSLINLKWTSIENAQK-IISAMDFIRYNSSEDYFEFGIDYDK---L
Adurb336        NANYTSKIDPKTGFVNLLYPNYKNIEKSKL-FFDKFDSIKYNKSENMFEFEFDYSN---F
Fn3             PAGFTSKICPVTGFVNQLYPKYESVSKSQE-FFSKFDKICYNLDKGYFEFSFDYKN---F
                 *  ** * * ***.. :  :  .  :    ::  :. : :: .   * :  . .    :

Hk              AKANLVILNNIKSKDWKISTR---------------GERISYNSK--KKEYFY-----VQ
As              QRG----LPGFMPA-WDIVFEKNETQFDAKGTPFIAGKRIVPVIE---NHRFTGRYRDLY
Pd              KIR----MKSAQTR-WTICTF---------------GNRIKRKKD---KNYWN--YEEVE
Pi              KIR----MKGAQTR-WTICTF---------------GNRIKHKKD---KNYWN--YEEVE
Adurb193        RTA----QTDFRKK-WTVCTF---------------GDRYTHTQN---KATSNHTTEKVN
Adurb336        TKN----LKLKKNN-WTVFTN---------------GERIIQVKT--KNNIYE--PKKIN
Fn3             G------DKAAKGK-WTIASF---------------GSRLINFRNSDKNHNWD--TREVY
                              * :                  *.*         :         :

Hk              PTEFLINKLKELNIDYENIDIIPLIDNLEEKAKRKILKALFDTFKYSVQLRNYDF--END
As              PANELIALLEEKGIVFRDGS--NILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGED
Pd              LTEEFKKLFKDSNIDYENC---NLKEEIQNKDNRKFFDDLIKLLQLTLQMRNSDDK-GND
Pi              LTEEFKKLFDKYDIDYRDG---NLKEEILKIDNRKFFDALIKLLQLTLQLRNSDDK-GND
Adurb193        LTAKLKAVLDKYSIDYDKGE--DIRDILCKSNSKELLETVLYVLKLAMQMRSTSPEDDVD
Adurb336        LTNEMKNLFESEGISFNDGK--NLKEEIINSNSKKLHTNLTNLLKYTLQLRNSNNKTGED
Fn3             PTKELEKLLKDYSIEYGHGE--CIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELD
                 :  :   :.. .* :       :   :    .      :   :.  :*:*. .     *

Hk              YIISPTADDNGNYYNSNEIDIDKTNLPNNGDANGAFNIARKGLLLKDRIVNSNE--SKVD
As              YINSPVRDLNGVCFDSRFQN---PEWPMDADANGAYHIALKGQLLLNHLKESKD--LKLQ
Pd              YIISPVANAEGQFFDSRNGD---KKLPLDADANGAYNIARKGLWNIRQIKQTKN-DKKLN
Pi              YIVSPIANADGKFFNSNDGC---KELPLDADANGAYNIARKGLWVVRQIKDKKD---KIS
Adurb193        MIISPVKNDDGKFFVSGN-D---EKLPIDADANGAFHIALKGLMYIKRIKNGKI--ECFD
Adurb336        YILSCVKDDEKNFFNSNNAK---DSEPENADANGAYHIGLKGIMLIKKMKKQSDKNQKID
Fn3             YLISPVADVNGNFFDSRQAP---KNMPQDADANGAYHIGLKGLMLLDRIKNNQE-GKKLN
                 : *   : :   : *        . * :.*****::*. **     .: . .     ..

Hk              LKIKNEDWINFIIS---------
As              NGISNQDWLAYIQELRN------
Pd              LSISSTEWLDFVREKPYLK----
Pi              -KLSNQEWLKFAQEKPYLK----
Adurb193        KGMATYEWLKFMQNREYKS----
Adurb336        LRISNEEYFNEMCSKEQSKKKES
Fn3             LVIKNEEYFEFVQNRNN------
                  : . :::    .
```

Green highlights indicate AsCas12a residues mutated in the RR, RVR, and enAsCas12a variants (Gao L et al, *Nature Biotechnology* 2017; Kleinstiver BP et al, *Nature Biotechnology* 2019).
Red highlights indicate PiCas12a residues mutated in this work.