Methods are described in the order they appear in the Results section

Standard methods are cited

Details necessary for replication

EXPERIMENTAL PROCEDURES

Generation of Heterologous Plasmids

To generate the FnCpf1 locus for heterologous expression, genomic DNA from Francisella novicida (generous gift from Wayne Conlan) was PCR amplified using Herculase II polymerase (Agilent Technologies) and cloned into pACYC-184 using Gibson cloning (New England Biolabs). Cells harboring plasmids were made competent using the Z-competent kit (Zymo). Sequences of all bacterial expression plasmids can be found in Table S1 .

Bacterial RNA Sequencing

RNA was isolated from stationary-phase bacteria by first resuspending F. novicida (generous gift from David Weiss) or E. coli in TRIzol and then homogenizing the bacteria with zirconia/silica beads (BioSpec Products) in a BeadBeater (BioSpec Products) for three 1-min cycles. Total RNA was purified from homogenized samples with the Direct-Zol RNA miniprep protocol (Zymo), DNase treated with TURBO DNase (Life Technologies), and 3 ^{0} dephosphorylated with T4 Polynucleotide Kinase (New England Biolabs). rRNA was removed with the bacterial Ribo-Zero rRNA removal kit (Illumina). RNA libraries were prepared from rRNA-depleted RNA using NEBNext Small RNA Library Prep Set for Illumina (New England Biolabs) and size selected using the Pippin Prep (Sage Science).

For heterologous E. coli expression of the FnCpf1 locus, RNA-sequencing libraries were prepared from rRNA-depleted RNA using a derivative of the previously described CRISPR RNA-sequencing method (Heidrich et al., 2015). In brief, transcripts were poly-A tailed with E. coli Poly(A) Polymerase (New England Biolabs), ligated with 5 ^{0} RNA adapters using T4 RNA Ligase 1 (ssRNA Ligase) High Concentration (New England Biolabs), and reverse transcribed with AffinityScript Multiple Temperature Reverse Transcriptase (Agilent Technologies). cDNA was PCR amplified with barcoded primers using Herculase II polymerase (Agilent Technologies).

RNA-Sequencing Analysis

The prepared cDNA libraries were sequenced on a MiSeq (Illumina). Reads from each sample were identified on the basis of their associated barcode and aligned to the appropriate RefSeq reference genome using BWA (Li and Durbin, 2009). Paired-end alignments were used to extract entire transcript sequences using Picard tools (http://broadinstitute.github.io/picard), and these sequences were analyzed using Geneious 8.1.5 (Biomatters).

In Vivo FnCpf1 PAM Screen

Randomized PAM plasmid libraries were constructed using synthesized oligonucleotides (IDT) consisting of eight or seven randomized nucleotides either upstream or downstream, respectively, of the FnCpf1 spacer 1. The randomized ssDNA oligos (Table S1) were made double stranded by annealing to a short primer and using the large Klenow fragment (New England Biolabs) for second-strand synthesis. The dsDNA product was assembled into a linearized pUC19 using Gibson cloning (New England Biolabs). Competent Stbl3 E. coli (Invitrogen) were transformed with the cloned products, and >10 ^{7} cells were collected and pooled. Plasmid DNA was harvested using a Maxi-prep kit (QIAGEN). We transformed 30 ng of the pooled library into E. coli cells carrying the FnCpf1 locus or pACYC184 control. After transformation, cells were plated on ampicillin. After 16 hr of growth, >4E6 cells were harvested and plasmid DNA was extracted using a Maxi-prep kit (QIAGEN). The target PAM region was amplified and sequenced using a MiSeq (Illumina) with a single-end 150 cycle kit.

Computational PAM Discovery Pipeline

PAM regions were extracted, counted, and normalized to total reads for each sample. For a given PAM, enrichment was measured as the log ratio compared to pACYC184 control, with a 0.01 psuedocount adjustment. PAMs above a 3.5 enrichment threshold were collected and used to generate sequence logos (Crooks et al., 2004).

PAM Validation

Sequences corresponding to both PAMs and non-PAMs were cloned into digested pUC19 and ligated with T4 ligase (Enzymatics). Competent E. coli with either the FnCpf1 locus plasmid or pACYC184 control plasmid were transformed with 20 ng of PAM plasmid and plated on LB agar plates supplemented with ampicillin and chloramphenicol. Colonies were counted after 18 hr.

Synthesis of crRNAs and sgRNAs

All crRNAs and sgRNAs used in biochemical reactions were synthesized using the HiScribe T7 High Yield RNA Synthesis Kit (NEB). ssDNA oligos (Table S2) corresponding to the reverse complement of the target RNA sequence were synthesized from IDT and annealed to a short T7 priming sequence. T7 transcription was performed for 4 hr, and then RNA was purified using the MEGAclear Transcription Clean-Up Kit (Ambion).

Purification of Cpf1 Protein

FnCpf1 protein was cloned into a bacterial expression vector (6-His-MBP-TEV-Cpf1, a pET based vector generously provided by Doug Daniels). Two liters of Terrific Broth growth media with 100 mg/ml ampicillin were inoculated with 10 ml overnight culture Rosetta (DE3) pLyseS (EMD Millipore) cells containing the Cpf1 expression construct. Growth media plus inoculant was grown at 37 °C until the cell density reached 0.2 OD600, then the temperature was decreased to 21 °C. Growth was continued until OD600 reached 0.6 when a final concentration of 500 mM IPTG was added to induce MBP-Cpf1 expression. The culture was induced for 14–18 hr before harvesting cells and freezing at −80°C until purification.

Cell paste was resuspended in 200 ml of Lysis Buffer (50 mM HEPES [pH 7], 2M NaCl, 5 mM $MgCl_2$, 20 mM imidazole) supplemented with protease inhibitors (Roche cOmplete, EDTA-free) and lysozyme (Sigma). Once homogenized, cells were lysed by sonication (Branson Sonifier 450) and then centrifuged at 10,000 3 g for 1 hr to clear the lysate. The lysate was filtered through 0.22 micron filters (Millipore, Stericup) and applied to a nickel column (HisTrap FF, 5 ml), washed, and then eluted with a gradient ofimidazole. Fractions containing protein of the expected size were pooled, TEV protease (Sigma) was added, and the sample was dialyzed overnight into TEV buffer (500 mM NaCl, 50 mM HEPES [pH 7], 5 mM MgCl, 2 mM DTT). After dialysis, TEV cleavage was confirmed by SDS-PAGE, and the sample was concentrated to 500 ml prior to loading on a gel filtration column (HiLoad 16/600 Superdex 200) via FPLC (AKTA Pure). Fractions from gel filtration were analyzed by SDS-PAGE; fractions containing Cpf1 were pooled and concentrated to 200 ml and either used directly for biochemical assays or frozen at −80°C for storage. Gel filtration standards were run on the same column equilibrated in 2M NaCl, HEPES (pH 7.0) to calculate the approximate size of FnCpf1.

Generation of Cpf1 Protein Lysate

Cpf1 proteins codon optimized for human expression were synthesized with a C-terminal nuclear localization tag and cloned into the pcDNA3.1 expression plasmid by Genscript (Table S1). 2,000 ng of Cpf1 expression plasmids were transfected into 6-well plates of HEK293FT cells at 90% confluency using Lipofectamine 2000 reagent (Life Technologies). 48 hr later, cells were harvested by washing once with DPBS (Life Technologies) and scraping in lysis buffer (20 mM HEPES [pH 7.5], 100 mM KCl, 5mM MgCl $_2$, 1 mM DTT, 5% glycerol, 0.1% Triton X-100, 1X cOmplete Protease Inhibitor Cocktail Tablets [Roche]). Lysate was sonicated for 10 min in a Biorupter sonicator (Diagenode) and then centrifuged. Supernatant was frozen for subsequent use in in vitro cleavage assays.

In Vitro Cleavage Assay

Cleavage in vitro was performed either with purified protein (25 nM) or mammalian lysate with protein at 37 °C in cleavage buffer (NEBuffer 3, 5 mM DTT) for 20 min. The cleavage reaction used 500 ng of synthesized crRNA or sgRNA and 200 ng of target DNA. Target DNA involved either protospacers cloned into pUC19 or PCR amplicons of gene regions from genomic DNA isolated from HEK293 cells. Reactions were cleaned up using PCR purification columns (QIAGEN) and were run on 2% agarose E-gels (Life Technologies). For native and denaturing gels to analyze cleavage by nuclease mutants, cleaned-up reactions were run on TBE 6% polyacrylamide or TBE-Urea 6% polyacrylamide gels (Life Technologies).

In Vitro Cpf1-Family Protein PAM Screen

In vitro cleavage reactions with Cpf1-family proteins were run on 2% agarose E-gels (Life Technologies). Bands corresponding to un-cleaved target were gel extracted using QIAquick Gel Extraction Kit (QIAGEN), and the target PAM region was amplified and sequenced using a MiSeq (Illumina) with a single-end 150 cycle kit. Sequencing results were entered into the PAM discovery pipeline.

Western Blot Analysis

Cells were lysed in 1 3 RIPA buffer (Cell Signaling Technology) supplemented with protease inhibitor cocktail (Roche). Equal volumes of cell lysate were run on BOLT 4%–12% Bis-Tris gradient gels (Invitrogen) and transferred to PVDF membranes (Millipore). Non-specific antigen binding was blocked with TBS-T (50 mM Tris, 150 mM NaCl and 0.05% Tween-20) with 5% BLOT-QuickBlocker Reagent (Millipore) for 1 hr. Membranes were incubated with primary antibodies (anti-HA-tag [Cell Signaling Technology C29F4] or HRP-conjugated GAPDH [Cell Signaling Technology 14C10]) for 1 hr in TBS-T with 1% BLOT-QuickBlocker. Membranes were washed for three 10 min washes and anti-HA-tag membranes were further incubated with anti-rabbit antibody (Cell Signaling Technology 7074) for 1 hr followed by six 10 min washes in TBS-T. Proteins were visualized with West Pico Chemiluminescent Substrate (Life Technology) and imaged using the ChemiDoc MP Imaging System (Bio-Rad) and processed with ImageLab software (Bio-Rad).

SURVEYOR Nuclease Assay for Genome Modification

PCR amplicons comprised of a U6 promoter driving expression of the crRNA sequence were generated using Herculase II (Agilent Technologies) and appropriate U6 reverse primers (Table S2). 400 ng of Cpf1 expression plasmids and 100 ng of the U6::crRNA expression cassettes were transfected into 24-well plates of HEK293FT cells at 75%–90% confluency using Lipofectamine 2000 (Life Technologies).

Cells were incubated at 37 °C for 72 hr post-transfection before genomic DNA extraction. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicenter) following the manufacturer's protocol. The genomic region flanking the CRISPR target site for each gene was PCR amplified, and products were purified using QiaQuick Spin Column (QIAGEN) following the manufacturer's protocol. 200–500 ng total of the purified PCR products were mixed with 1 ml 10 3 Taq DNA Polymerase PCR buffer (Enzymatics) and ultrapure water to a final volume of 10 ml and were subjected to a re-annealing process to enable heteroduplex formation: 95 °C for 10 min, 95°C to 85 °C ramping at −2°C/s, 85 °C to 25 °C at −0.25 °C/s, and 25 °C hold for 1 min. After re-annealing, products were treated with SURVEYOR nuclease and SURVEYOR enhancer S (Integrated DNA Technologies) following the manufacturer's recommended protocol and analyzed on 4%–20% Novex TBE polyacrylamide gels (Life Technologies). Gels were stained with SYBR Gold DNA stain (Life Technologies) for 10 min and imaged with a Gel Doc gel imaging system (Bio-rad). Quantification was based on relative band intensities. Indel percentage was determined by the formula, 100 3 (1 − sqrt(1 − (b + c)/(a + b + c))), where a is the integrated intensity of the undigested PCR product, and b and c are the integrated intensities of each cleavage product.

Deep Sequencing to Characterize Cpf1 Indel Patterns in 293FT Cells

HEK293FT cells were transfected and harvested as described for assessing activity of Cpf1 cleavage. The genomic-region-flanking DNMT1 targets were amplified using a two-round PCR region to add Illumina P5 adapters as well as unique sample-specific barcodes to the target amplicons. PCR products were run on 2% E-gel (Invitrogen) and gel extracted using QiaQuick Spin Column (QIAGEN) as per the manufacturer's recommended protocol. Samples were pooled and quantified by Qubit 2.0 Fluorometer (Life Technologies). The prepared cDNA libraries were sequenced on a MiSeq with a single-end 300 cycle kit (Illumina). Indels were mapped using a Python implementation of the Geneious 6.0.3 Read Mapper.

Computational Analysis of Cpf1 loci

PSI-BLAST program (Altschul et al., 1997) was used to identify Cpf1 homologs in the NCBI NR database using several known Cpf1 sequences as queries with the Cpf1 with the E-value cut-off of 0.01 and low-complexity filtering and composition-based statistics turned off. The TBLASTN program with the E-value cut-off of 0.01 and low-complexity filtering turned off was used to search the NCBI WGS database using the Cpf1 profile (Makarova et al., 2015) as the query. Results of all searches were combined (Table S3). The HHpred program was used with default parameters (Söding et al., 2006) to identify remote sequence similarity using a subset of representative Cpf1 sequences queries. Multiple sequence alignments were constructed using MUSCLE (Edgar, 2004) with manual correction based on pairwise alignments obtained using PSI-BLAST and HHpred programs. Phylogenetic analysis was performed using the FastTree program with the WAG evolutionary model and the discrete gamma model with 20 rate categories (Price et al., 2010). Protein secondary structure was predicted using Jpred 4 (Drozdetskiy et al., 2015). CRISPR repeats were identified using PILER-CR (Edgar, 2007) and CRISPRfinder (Grissa et al., 2007).

Reasons for using a particular methodology