Supplementary Figures

Biochemically diverse CRISPR-Cas9 orthologs


Figure S1. Cas9 protein size distribution according subtype of CRISPR-Cas system.
Figure S2. Schematic representation of CRISPR loci of the Type II CRISPR-Cas systems used in this study. Cas9 and tracrRNA encoding sequences are colored in blue and red accordingly. Predicted transcription directions of tracrRNA and CRISPR loci are indicated by arrows. Clade membership is indicated by roman numerals.
Figure S3. The secondary structure diagrams of the guide RNA molecules identified for selected Cas9 orthologs. Clade membership is indicated by roman numerals. The sgRNA structures can be classified into at least 6 groups based on structural prediction and exemplified by Tde, Seq1 (Spy-like), Sth1A, Nme2, Nsa, and Cca1 (names are underlined).
Figure S4. Verification of PAM determination methods. (A) PAM preferences for Spy, Sth3 and Sth1 Cas9 proteins determined using different dilutions of crude IVT RNP mixtures. (B) PAM confirmation using purified Cas9 proteins.
Figure S5 HHpred results of searches using representative members from clusters 7 (A), 8 (B), 9 (C) and 10 (D).
Figure S6. Phylogenetic tree of PI domains from clusters 1 (A), 2 (B), 3 (C), 4 (D) and 5&6 (E). SH-aLRT bootstrap values are shown at the branches.
Figure S7. Weblogo of SpyCas9 homolog sequences. Only positions having Jalview conservation score of 1 and higher are shown. Average conservation score is shown in parenthesis. Sequences having all domains characteristic to the SpyCas9 from Cluster 1 (Supplementary table S3) were used to build MAFFT alignment.
Figure S8. Spacer length preferences for purified Cas9 orthologs. (A) The cleavage activity of selected Cas9 orthologs was examined using sgRNAs with 20 or 24 nt spacers. In vitro DNA cleavage reactions were performed in different buffers at 37°C (see Methods). Spy and Ssi Cas9 do not have a preference for 20 or 24 nt spacers. Nsa Cas9 has a preference for 24 nt spacers whereas Pac Cas9 has a preference for 20 nt spacers. Replicate experiments are shown as points and error bars represent standard deviation. (B) The in vitro DNA cleavage activity of purified Cas9 proteins was screened using 20 and 24 nt spacers and different buffers as shown. Clade membership is indicated by roman numerals.
Figure S9. Thermostability of purified Cas9 proteins. Melting temperatures of purified Cas9 proteins was determined by nanoDSF (see Methods). Tm indicated in the histograms represent the average of 3 independent determinations plotted as open circles. Error bars indicate standard deviation.
Figure S10. PAM discovery and cleavage site mapping using purified Cas9 RNP and dsDNA minicircles. Diagram of workflow to make small circular substrates for PAM and cut site determination assays. Pools of single stranded DNA oligonucleotides (blue) containing a 5’ PO4 ends and 10 nt of randomized sequence (red) were circularized and converted to double stranded circles after annealing a second strand primer (green) and extending it with polymerase in the presence of a ligase. DNA circles are then exposed to Cas9 RNPs containing guide RNA complementary to the DNA sequence flanking the 10 nt randomized sequence region. DNA circles that are cut by the Cas9 are substrates for ligation of adapters (orange) for Illumina sequencing. Custom bioinformatic analysis was used to determine cut site pattern and PAM preference. Control reactions were carried out where each pool of circles was cut with the restriction enzyme BstXI to determine the baseline frequency of nucleotides at every position in the randomized region.
Figure S11. Target-dependent cleavage patterns. Diverse cleavage sites over multiple targets for a single Cas9 ortholog are depicted as heatmaps depicting mapped cleavage ends at each position in a single DNA target. Sequence of the DNA non-target strand and PAM is depicted above the heatmaps. Intensity of the blue color indicates the proportion of mapped cleavage ends. NTS indicates non-target strand; TS indicates target strand. (A) Spy Cas9 cleavage patterns. (B) Sau Cas9 cleavage patterns. (C) Khu Cas9 cleavage patterns. (D) Cpe Cas9 cleavage patterns. (E) Lpn cleavage patterns. (F) Tsp Cas9 cleavage patterns. (G) Nsa cleavage patterns. (H) Esp1 Cas9 cleavage patterns. (I) Lan Cas9 cleavage patterns. (J) Seq1 Cas9 cleavage patterns.
**Figure S12.** Comparison of PI domain structures from 3 major clusters (figure 2): Spy (4zto), Nmen (6ke8), Sau (5axw). There are no solved structures of Cas9 proteins which PI domain belongs to cluster 2.