nature portfolio

https://doi.org/10.1038/s41586-024-07891-2

Supplementary information

Global marine microbial diversity and its potential in bioprospecting

In the format provided by the authors and unedited

Supplementary Information

Global marine microbial diversity and its potential in bioprospecting

Jianwei Chen[1,2,3,4,*], Yangyang Jia[2,*], Ying Sun[1,3,*,#], Kun Liu[5,*], Changhao Zhou[1], Chuan Liu[2,4], Denghui Li[1], Guilin Liu[1], Chengsong Zhang[5], Tao Yang[6,7], Lei Huang[2], Yunyun Zhuang[8], Dazhi Wang[9], Dayou Xu[1], Qiaoling Zhong[6], Yang Guo[1,16], Anduo Li[2], Inge Seim[10], Ling Jiang[11], Lushan Wang[5], Simon Ming Yuen Lee[12], Yujing Liu[1,3], Dantong Wang[1], Guoqiang Zhang[5], Shanshan Liu[1], Xiaofeng Wei[6,7], Zhen Yue[13], Shanmin Zheng[5], Xuechun Shen[2], Sen Wang[5], Chen Qi[2], Jing Chen[7], Chen Ye[2], Fang Zhao[1], Jun Wang[1], Jie Fan[1,3], Baitao Li[2], Jiahui Sun[1], Xiaodong Jia[14], Zhangyong Xia[15], He Zhang[1,2], Junian Liu[1], Yue Zheng[2], Xin Liu[1,2], Jian Wang[2], Huanming Yang[2], Karsten Kristiansen[2,3,4], Xun Xu[1,2,3,17], Thomas Mock[18,#], Shengying Li[5,19,#], Wenwei Zhang[2,17,#], Guangyi Fan[1,2,3,17,#]

[1] BGI Research, Qingdao 266555, China

[2] BGI Research, Shenzhen 518083, China

[3] Qingdao Key Laboratory of Marine Genomics, and Qingdao-Europe Advanced Institute for Life Sciences, BGI Research, Qingdao 266555, China

[4] Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, Universitetsparken 13, 2100 Copenhagen, Denmark

[5] State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, China

[6] China National GeneBank, BGI Research, Shenzhen 518120, China

[7] Guangdong Genomics Data Center, BGI Research, Shenzhen 518120, China

[8] Key Laboratory of Environment and Ecology, Ministry of Education, Ocean University of China, Qingdao 266100, China

[9] State Key Laboratory of Marine Environmental Science/College of the Environment and Ecology, Xiamen University, Xiamen 361005, China

[10] Marine Mammal and Marine Bioacoustics Laboratory, Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya 57200, China

[11] College of Food Science and Light Industry, Nanjing Tech University, Nanjing 211816, China

[12] State Key Laboratory of Chemical Biology and Drug Discovery, and Department of Food Science and Nutrition, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China
[13] BGI Research, Sanya 572025, China
[14] Joint Laboratory for Translational Medicine Research, Liaocheng People's Hospital, Liaocheng 252000, China
[15] Department of Neurology, the Second People's Hospital of Liaocheng, Liaocheng 252000, China
[16] Center of deep-Sea Research, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China
[17] State Key Laboratory of Agricultural Genomics, BGI Research, Shenzhen 518083, China
[18] School of Environmental Sciences, University of East Anglia, Norwich Research Park, NR4 7TJ Norwich, UK
[19] Laboratory for Marine Biology and Biotechnology, Qingdao Marine Science and Technology Center, Qingdao 266237, China

[*] These authors contributed equally.
[#] Correspondence: fanguangyi@genomics.cn, zhangww@genomics.cn, lishengying@sdu.edu.cn, T.Mock@uea.ac.uk and sunying6@genomics.cn

Supplementary Note 1

Calculation of metagenomic community composition

We used metagenomic data to calculate microbial community compositions across the global ocean to identify potential biogeographic patterns associated with taxonomic profiling and functional genes. Kraken2-build was used to construct a customized Kraken database using genomes constructed in this study, and then the metagenomic sequencing reads from 5,377 samples with explicit geographic location information were classified against the constructed database using Kraken2 (v2.1.2) with default parameters[1]. The read count profiles for each taxonomic level were calculated using Bracken (v2.5)[2]. Normalized species relative abundance table was imported into Seurat (v3.2.1) for the dimensionality reduction clustering using the uniform manifold approximation and projection (UMAP) method[3,4].

Microbiome analyses and visualization related to metagenomic province features were primarily carried out by the phyloseq package (v3.17)[5] under the R environment[6]. Alpha diversity was measured by Shannon index[7] and the between-community distance assessed by the Jensen-Shannon divergence (JSD) index[8] was calculated based on the relative abundance profiles generated by Bracken, using the estimate_richness and distance functions, respectively. Analysis of similarity (ANOSIM) based on the JSD measures was performed, using the anosim function of the pairwiseAdonis package (v0.4)[9], to evaluate the microbial assemblage differences among metagenomic provinces.

Geographic distance between sampling sites was calculated using the distm function with the "distHaversine" option of the geosphere package (v1.5.18)[10]. Distance-decay relationship between microbial community dissimilarity, with regard to the JSD index, and geographic distance between samples was measured by the correlation coefficients of Mantel tests, using the mantel function of the vegan package (v2.6.4)[11].

Calculation of spatial and environmental determinants of microbiome diversity

The mean annual climatology data of temperature, salinity, dissolved oxygen, silicate, phosphate, nitrate, apparent oxygen utilization, conductivity, and density at a $1^{\circ} \times 1^{\circ}$ resolution were collected from the World Ocean Atlas 2018 (WOA2018) database[12,13]. We categorized our seawater metagenomes into multiple ocean layers based on depths, including the surface (SRF; between 0 m and 10 m), the deep chlorophyll maximum (DCM; between 10 m and 200 m), the mesopelagic

(MES; between 200 m and 1,000 m), and the bathypelagic (BATH; between 1,000 m and 4,000 m) zones. Given that a large proportion of the metagenomes collected from NCBI SRA lacks clear depth information measured in meters, the environmental parameters for each sample were obtained by matching their longitude and latitude to their nearest neighbor on the $1^{\circ} \times 1^{\circ}$ grid of WOA2018, extracting all the measures across a range of depths based on the ocean layer assigned to the sample, and then calculating the mean values as an approximation.

The influence of environmental parameters on the Shannon diversity of samples was explored by multivariable linear regressions using the lm function in the stats package (v4.2.2). Before the regression analysis, spearman correlations among different environmental parameters were calculated using the cor function in the stats package (**Fig.** Environmental factors vs microbial community variation a). Among the highly correlated environmental parameters, such as temperature, conductivity and density, only one was selected as the independent predictor variable for the downstream regression analyses. As a result, four environmental parameters were retained as independent predictors, including temperature, salinity, phosphate and dissolved oxygen. The skewness of the response variable (Shannon diversity) was estimated and multiple transformation methods were applied to improve its normality (**Fig.** Environmental factors vs microbial community variation b). Further, Mantel and partial Mantel tests were implemented to evaluate the effects of different environmental parameters on the beta diversity among samples, as measured by the JSD matrix, at different depths (**Fig.** Environmental factors vs microbial community variation c), using the mantel and mantel.partial functions of the vegan package, respectively. For the partial Mantel test, geographic distance between sampling sites was considered as the third control distance matrix.

Description of global marine microbiomes and their biogeography

Different marine ecosystems appear to be home to distinct microbial communities likely because of environmental filtering (**Extended Data Fig. 3a**). For instance, whereas Proteobacteria are distributed across the water column from the surface to the deeper oceans, Cyanobacteria are enriched in the deep chlorophyll maximum (DCM) layers. Planctomycetota were relatively more abundant in the mesopelagic (MES) than in shallower waters. Chloroflexota and Marinisomatota were most abundant in meso- and bathypelagic (BATH) samples, suggesting their preference for the deeper ocean. Most widespread throughout the water column and all oceans was SAR324 with

abyssopelagic members, consistent with previously identified depth zonation patterns of its ecotypes[14]. Sediment communities were enriched with Desulfobacterota, one of the most ubiquitous sulfate-reducing bacteria in benthic environments. We identified an overall positive but weak correlation between geographical distances and the Jensen-Shannon divergence (JSD) between microbial communities from different depths (Mantel coefficients between 0.09 and 0.39, $P < 0.01$), with surface water (SRF) communities showing the weakest correlation. This result might be attributed to thermohaline mixing and surface ocean currents. Usually, both processes are less pronounced in deeper waters. Correlations with physical environmental variables (e.g., temperature, salinity) generally showed a limited predictive ability for explaining differences in the Shannon diversity of microbial communities ($P < 0.01$ but adjusted $R^2 < 0.5$) (**Fig.** Environmental factors vs microbial community variation ab). However, macronutrients appear to be more influential on the diversity of microbial communities in the DCM zone (**Fig.** Environmental factors vs microbial community variation c), which is mainly occupied by phytoplankton[15]. To shed light on drivers shaping the bacterial community composition in the global ocean will require more accurate and comprehensive datasets on environmental conditions to identify potential deterministic biotic (e.g., predator-prey interactions) and likely also abiotic factors (e.g., novel resources used by microbes) which have not been accounted for yet[16].

Despite the overarching patterns of microbial distribution observed across the water column, notable diversity persists in microbial communities within each distinct depth layer (**Extended Data Fig. 3a**). We identified 56 metagenomic provinces (MPs) using the uniform manifold approximation and projection (UMAP) method based on the relative abundance of MAGs in each sample[17] (**Extended Data Fig. 3b)**. The geographical distribution of samples within MPs is illustrated in CNGBdb website (**Fig.** UMAP biogeographical distribution). In the large figure, each MP consists of three panels, from left to right, showing the microbial community composition of each sample, the location of samples in the UMAP dimensionality reduction space, and the geographic regions covered by the MP in the global ocean. Color scheme for taxonomic groups is the same as that in **Extended Figure 3a**. In the first panel: Samples were ordered based on their distance to the center of the MP cluster in the UMAP space increasingly. Above the stacked bar of each sample, a dot with different colors is used to show the depth layer to which the sample belongs. Color scheme for depth layers is also shown on top of the figure. The transparency of each dot is proportional to the distance between the specific sample and the center of the MP

cluster in the UMAP space. The closer that sample is to the center, the opaquer the dot is. In the second and third panels, the color and transparency of each sample were the same as for the dots in the first panel. Further, we assessed the degree of separation between MPs by an ANOSIM test on the Jensen-Shannon divergence (JSD) matrix ($R = 0.61$, $P < 0.01$). Generally, we found that Bacteroidia, Alpha- and Gammaproteobacteria were prevalent in many MPs[18,19], whereas others were only abundant in particular ecosystems, such as Campylobacterota mainly occurring at deep-sea hydrothermal vents[20]. This large-scale biogeographical partitioning highlights the global influence on the distribution of marine microorganisms. The significant variation observed in the MPs across different depths suggests a complex interplay of environmental factors shaping microbial communities in the open ocean. The fact that these MPs transcend geographically clustered sampling sites indicates that the patterns we observed are not limited to specific regions but are indicative of broader processes governing oceanic microbial dynamics.

Supplementary Note 2

Comparison of GOMC large genomes with NCBI GenBank records

Genome size was further adjusted for completeness and contamination using the same methodology employed in the previous publications[21,22] with the formula *Estimated Genome Size* = *Genome Size* / *Completeness* × (*100* - *Contamination*). To gain a comprehensive view of the sizes of publicly available prokaryotic bacterial and archaeal assemblies, we acquired their summary report with metadata from the NCBI FTP site. As of July 2nd, 2023, NCBI has made available 1,608,241 latest GenBank prokaryotic bacterial and archaeal assemblies, of which 99.57% exhibit a genome size not exceeding 8 Mb (**Fig.** GOMC large bacterial genomes a). We identified 26 prokaryotic bacterial genomes exceeding 15 Mb in the NCBI GenBank archive and retrieved their genome sequences for further comparison. Among them, four were sampled from marine related environments (**Fig.** GOMC large bacterial genomes a).

We implemented the entropy-based method GUNC (v1.0.5) to evaluate genome chimerism[23] and CheckM to assess completeness and redundancy among our 3 Pirellulaceae MAGs larger than 16 Mb as well as the aforementioned 26 NCBI large genomes. The MAGs whose clade separation scores (CSS) higher than 0.45 and reference representation scores (RRS) greater than 0.5 were regarded as chimeric[23]. Additionally, the MAGs whose QS fell below 50 also failed to meet the quality criteria. Our three large Pirellulaceae MAGs successfully passed both the GUNC and

CheckM evaluations, suggesting they may constitute the largest marine prokaryotic bacterial genomes so far (**Fig.** GOMC large bacterial genomes a).

A total of 192 MAGs from the GOMC database were assigned to the Pirellulaceae family. Their phylogenetic position within the family Pirellulaceae was determined by the de_novo_wf function of the GTDB-Tk toolkit (v2.1.1)[24] with the GTDB database (v207)[25] using the family Lacipirellulaceae as the outgroup (--taxa_filter f__Pirellulaceae --outgroup_taxon f__Lacipirellulaceae) **(Fig.** GOMC large bacterial genomes b). The phylogenetic relationship of our three large MAGs and their closely related species underwent additional validation through the following steps: First, 27 genomes in the adjacency of our target large genomes, marked by a green strip, were selected according to the species tree built by GTDB-Tk, along with an outgroup genome from the family Lacipirellulaceae; Second, ORFs were annotated as mentioned above and then OrthoFinder (v2.5.4) was used to cluster orthologous proteins among the 31 genomes[26]; Third, 314 single-copy orthologous protein families (OGs) shared by at least 28 out of the 31 genomes were selected; Forth, protein sequences in each family were aligned using MAFFT (v7.407) and excessive gaps were removed by TrimAl (-gappyout). The 314 alignments were concatenated into a super alignment for phylogeny construction; Last, Maximum likelihood phylogenetic inference under the best-fit partition model was performed by IQ-Tree with 1,000 replicates (v2.1.4-beta; -m TESTMERGE -bb 1000)[27] (**Fig.** GOMC large bacterial genomes c).

Identification of functional features related to prokaryotic bacterial genome enlargement

To investigate potential genetic features that distinguish our three large Pirellulaceae MAGs from others exhibiting smaller genome sizes, we examined how the number of specific functional families in the genome relates to its size through phylogenetic regression analysis. First, we focused on the monophyletic clade of the nine genomes as shown in **Fig.** GOMC large bacterial genomes c, because the two sister clades demonstrated discernible differences in genome sizes, making them an ideal case for conducting phylogeny-controlled comparisons. Subsequent paragraphs provide detailed descriptions of the procedures employed for function assignment and statistical analyses.

Functional annotation was performed with the Pfam (v35.0) database[28] using Hidden Markov models (HMMs)-based homology searches. The hmmsearch program (HMMER v3.3.2) was carried out against the Pfam profiles with the "--cut_ga" flag. To identify a domain within a protein, specific criteria were applied: the homologous hit score had to surpass the gathering threshold

established by the curated Pfam models; the e-value had to be below 0.001; and the number of domains satisfying the inclusion thresholds had to exceed zero. Noteworthy, the proteins under consideration were permitted to comprise multiple Pfam functional domains. Then, a matrix was constructed, where rows represented Pfam families, columns represented genomes, and each cell denoted the gene count assigned to a specific family within that genome.

To mitigate the impact of evolutionary distances on correlation estimates between genetic features and genome sizes, we employed phylogenetic regression (PR) analyses[29], accounting for the likelihood of increased similarity among closely related species due to shared phylogenetic ancestry. For each Pfam family, we utilized the phylolm function in the R package phylolm[30] with a lambda model of phylogenetic covariance[31]. The lambda values measure the strength of the phylogenetic signal with 0 indicating no effect and 1 standing for a strong effect of phylogeny. We controlled for both adjusted R^2 and FDR-adjusted p-values[32] to identify functional families exhibiting significant correlations with genome sizes. The results are shown in **Extended Data Fig. 4b** as a volcano plot displays the regression coefficients of gene count against their corresponding FDR values resulted from the PR analysis. Each point in the volcano plot represents a functional family (a Pfam domain). Points are colored according to their corresponding lambda values, indicating the strength of phylogenetic signal over the regression. The size of each point represents the corresponding R^2 value. The x-axis shows the regression coefficient of the phylogenetic regression upon each Pfam domain. The y-axis displays the corresponding FDR values after a negative log10 transformation. The horizontal dashed line marks the FDR value of 0.05. The vertical dashed line marks the regression coefficient of 0. The PR approach pinpointed 604 Pfam domains displaying a notably significant positive correlation with genome size, which remains statistically robust even when accounting for phylogenetic associations ($R^2 \geq 0.5$, false discovery rate [FDR] < 0.05 and regression coefficient > 0).

Ancestral genomes were inferred with AnGST[33] based on OGs identified by OrthoFinder as mentioned above. The default penalty scores of horizontal gene transfer (HGT) (= 3), gene duplication (= 2), gene loss (= 1) and speciation (= 0) were adopted as suggested in the original paper to minimize genome size flux[33]. Only OGs presenting in all the four large genomes of our target clade (highlighted in yellow in **Extended Data Fig. 4b**) were used in this analysis. Gene trees of each OG were built using IQ-Tree with 1,000 replicates (-m TEST -bb 1000 -bnni)[27]. AnGST infers evolutionary events by the reconciliation of topological incongruences between the

gene trees and the species tree under a parsimony framework[33]. We analyzed the genome contents of three ancestral genomes, including the last common ancestor (LCA) of our target clade (LCA1), the LCA of its sister clade (LCA2) and the LCA of both clades (LCA0) (**Extended Data Fig. 4b**). To identify OGs experiencing ancient expansions, specific criteria were applied (as described in pseudocode in **Extended Data Fig. 4b**): the gene counts of the OG at LCA1 were supposed to be greater than those at both LCA0 and LCA2; and in all the extant genomes, the gene counts of the OG in each large genome of the target clade must outnumbered those in the genomes of its sister clade. A comprehensive set of 1,005 OGs potentially underwent ancient expansion in gene counts at LCA1. Further, we summarized the Pfam families annotated in each gene of the chosen OGs. We expected to see an increment of specific Pfam families alongside the expansion of OGs and defined the vote for a Pfam domain as the fraction of genes in an OG that was assigned to that domain. The votes for a Pfam family were added up across all the chosen OGs and Pfam families with a vote higher than 3 were selected. In line with this proposition, we recognized 92 expanded Pfam domains.

So far, the selection of Pfam families related to genome enlargement, using phylogenetic regression analyses and ancestral genome content inferences, was mainly based on the Pirellulaceae large genomes and their close relatives. We subsequently extended our analysis to encompass the entire GOMC dataset of 24,195 genomes. We aimed to identify Pfam families for which an increase in genome size was associated with a higher count of genes annotated within that specific Pfam family. An illustrative instance is presented in **Extended Data Fig. 4b**, considering the Pfam family denoted as "PF13360": firstly, genomes featuring a sole gene attributed to the PF13360.9 family exhibited a statistically significant increase in size compared to those lacking genes associated with PF13360 (determined by a Wilcoxon rank-sum test with the alternative set to "greater"); secondly, genomes hosting two PF13360 genes surpassed the size of genomes containing only one instance of the PF13360 gene; furthermore, genomes containing a minimum of three PF13360 genes exceeded the size of genomes harboring two copies of the PF13360 gene. A total of 2,583 Pfam families were selected according to these criteria. Finally, 77 designated Pfams involved in genome stability, cell cycle progression, signal transduction and regulation were identified as related to the genome expansion (**Extended Data Fig. 4b**). The 77 designated Pfams are PF00037 (Fer4 [24aa] (4Fe-4S binding domain)), PF00069 (Pkinase [264aa] (Protein kinase domain)), PF00072 (Response_reg [112aa] (Response regulator receiver domain)),

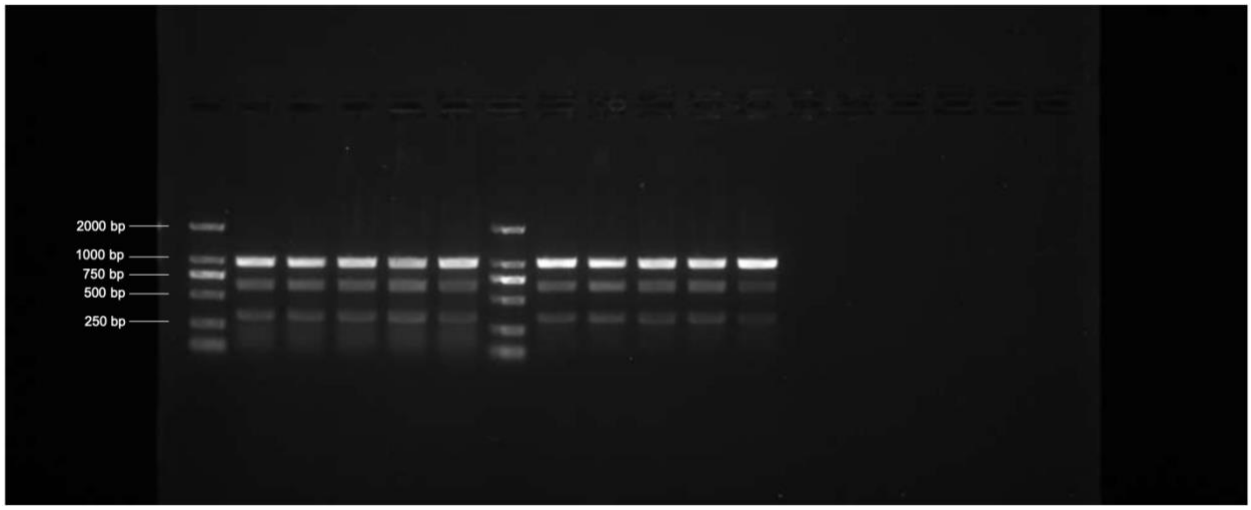
PF00092 (VWA [175aa] (von Willebrand factor type A domain)), PF00135 (COesterase [515aa] (Carboxylesterase family)), PF00149 (Metallophos [205aa] (Calcineurin-like phosphoesterase)), PF00326 (Peptidase_S9 [212aa] (Prolyl oligopeptidase family)), PF00400 (WD40 [38aa] (WD domain, G-beta repeat)), PF00474 (SSF [406aa] (Sodium:solute symporter family)), PF00571 (CBS [57aa] (CBS domain)), PF00583 (Acetyltransf_1 [117aa] (Acetyltransferase (GNAT) family)), PF00795 (CN_hydrolase [261aa] (Carbon-nitrogen hydrolase)), PF00884 (Sulfatase [309aa] (Sulfatase)), PF00890 (FAD_binding_2 [417aa] (FAD binding domain)), PF00989 (PAS [113aa] (PAS fold)), PF01011 (PQQ [38aa] (PQQ enzyme repeat)), PF01073 (3Beta_HSD [280aa] (3-beta hydroxysteroid dehydrogenase/isomerase family)), PF01120 (Alpha_L_fucos [350aa] (Alpha-L-fucosidase)), PF01134 (GIDA [392aa] (Glucose inhibited division protein A)), PF01208 (URO-D [345aa] (Uroporphyrinogen decarboxylase (URO-D))), PF01209 (Ubie_methyltran [233aa] (ubiE/COQ5 methyltransferase family)), PF01261 (AP_endonuc_2 [250aa] (Xylose isomerase-like TIM barrel)), PF01370 (Epimerase [241aa] (NAD dependent epimerase/dehydratase family)), PF01408 (GFO_IDH_MocA [120aa] (Oxidoreductase family, NAD-binding Rossmann fold)), PF01656 (CbiA [127aa] (CobQ/CobB/MinD/ParA nucleotide binding domain)), PF01663 (Phosphodiest [357aa] (Type I phosphodiesterase / nucleotide pyrophosphatase)), PF01738 (DLH [217aa] (Dienelactone hydrolase family)), PF02518 (HATPase_c [112aa] (Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase)), PF02837 (Glyco_hydro_2_N [169aa] (Glycosyl hydrolases family 2, sugar binding domain)), PF03130 (HEAT_PBS [27aa] (PBS lyase HEAT-like repeat)), PF03629 (SASA [226aa] (Carbohydrate esterase, sialic acid-specific acetylesterase)), PF03781 (FGE-sulfatase [260aa] (Sulfatase-modifying factor enzyme 1)), PF04055 (Radical_SAM [167aa] (Radical SAM superfamily)), PF04909 (Amidohydro_2 [288aa] (Amidohydrolase)), PF05048 (NosD [210aa] (Periplasmic copper-binding protein (NosD))), PF05448 (AXE1 [318aa] (Acetyl xylan esterase (AXE1))), PF05569 (Peptidase_M56 [299aa] (BlaR1 peptidase M56)), PF06283 (ThuA [210aa] (Trehalose utilisation)), PF06439 (3keto-disac_hyd [185aa] (3-keto-disaccharide hydrolase)), PF07638 (Sigma70_ECF [185aa] (ECF sigma factor)), PF07676 (PD40 [38aa] (WD40-like Beta Propeller Repeat)), PF07690 (MFS_1 [353aa] (Major Facilitator Superfamily)), PF07714 (PK_Tyr_Ser-Thr [259aa] (Protein tyrosine and serine/threonine kinase)), PF07859 (Abhydrolase_3 [211aa] (alpha/beta hydrolase fold)), PF07940 (Hepar_II_III [237aa] (Heparinase II/III-like protein)), PF07944 (Glyco_hydro_127 [508aa] (Beta-L-arabinofuranosidase, GH127)), PF08241

(Methyltransf_11 [96aa] (Methyltransferase domain)), PF08281 (Sigma70_r4_2 [54aa] (Sigma-70, region 4)), PF08448 (PAS_4 [110aa] (PAS fold)), PF12706 (Lactamase_B_2 [201aa] (Beta-lactamase superfamily domain)), PF12831 (FAD_oxidored [427aa] (FAD dependent oxidoreductase)), PF12838 (Fer4_7 [52aa] (4Fe-4S dicluster domain)), PF13088 (BNR_2 [275aa] (BNR repeat-like domain)), PF13174 (TPR_6 [33aa] (Tetratricopeptide repeat)), PF13181 (TPR_8 [34aa] (Tetratricopeptide repeat)), PF13183 (Fer4_8 [65aa] (4Fe-4S dicluster domain)), PF13188 (PAS_8 [67aa] (PAS domain)), PF13202 (EF-hand_5 [25aa] (EF hand)), PF13229 (Beta_helix [158aa] (Right handed beta helix region)), PF13360 (PQQ_2 [237aa] (PQQ-like domain)), PF13385 (Laminin_G_3 [151aa] (Concanavalin A-like lectin/glucanases superfamily)), PF13426 (PAS_9 [104aa] (PAS domain)), PF13432 (TPR_16 [68aa] (Tetratricopeptide repeat)), PF13450 (NAD_binding_8 [68aa] (NAD(P)-binding Rossmann-like domain)), PF13489 (Methyltransf_23 [165aa] (Methyltransferase domain)), PF13517 (FG-GAP_3 [61aa] (FG-GAP-like repeat)), PF13519 (VWA_2 [108aa] (von Willebrand factor type A domain)), PF13534 (Fer4_17 [61aa] (4Fe-4S dicluster domain)), PF13570 (PQQ_3 [40aa] (PQQ-like domain)), PF13646 (HEAT_2 [88aa] (HEAT repeats)), PF13649 (Methyltransf_25 [97aa] (Methyltransferase domain)), PF13768 (VWA_3 [155aa] (von Willebrand factor type A domain)), PF16347 (DUF4976 [103aa] (Domain of unknown function (DUF4976))), PF16363 (GDP_Man_Dehyd [332aa] (GDP-mannose 4,6 dehydratase)), PF17132 (Glyco_hydro_106 [742aa] (alpha-L-rhamnosidase)), PF18582 (HZS_alpha [102aa] (Hydrazine synthase alpha subunit middle domain)), PF20434 (BD-FAE [215aa] (BD-FAE)).

Validation of functional features related to bacterial genome enlargement across phyla

We assessed the persistence of the positive correlations between the 77 chosen Pfam domains and genome size across a broader taxonomic spectrum (**Extended Data Fig. 4b** and **Extended Data Fig. 5a,b**). To achieve this, we conducted supplementary analyses employing high-quality genomes (with completeness ≥ 70% and contamination ≤ 10%) sourced from 9 phyla. These 9 phyla were selected due to their possession of a minimum of 5 genomes exceeding 8 Mb in size, aligning with the context of our preliminary findings. Employing the same methodology as previously described, we performed phylogenetic regression analyses on the 77 selected Pfam domains across the genomes within each of these phyla.

Supplementary Figures



Supplementary Fig. 1| Uncropped agarose gel electrophoresis of *in vitro* dsDNA cleavage products of the *AASV1* gene fragments for the Om1Cas9 (left) and SpCas9 (right), respectively.

Supplementary Tables

Supplementary Table 2. The 36 CRISRP-Cas9 systems identified in GOMC.

Genome ID	Operon	Length	Cas9 protein sequence	DRs	Judge
GOMC.bin.6469	Cas4_0_I-II,Cas2_2_CAS-I-II-III-IV-V-VI,Cas1_0_CAS-I-II-III-IV-V-VI,Cas9_0_IIB	1609 aa	MKIISPIAIDMGAKNTGVYYAKYKEGSTFEQIEKHGQVLEYGNYTSLLKERTAKRHSRRGYQRKKLAKRLLTLVLENYFNFPAKQHSQAIGFLLNRRGFTFLDEEYSKERLNEFPDDAWKMLPEEVRSILENKKNIVQIFDECVKKEPEKIGEILKCIESSDGVQRVKKLKSDMVYFTYINKIVDKCQYRADGNMVTEKGGDRNKLSQTSKWIVERLVKETTIPLLTFTQNAKQTAFYDLKEHVNGLGAEQAVKFKKQLPDVSQSIKKAKAALKESNWDFDFQKFTFDDKHEEGLMGGDVKIHLQHFCYAIDKINTEITSGARHRSKFFKEIEGDLEKFQSPEYQNNKHAHKYLKEFAKEVVQNSKLGRENLTRLVCHTSNLELKPLRAYFNDSAPILRSKNYEKSNQTKSNFISHKNDGGEFSKEKLSRVASTWFLKHWMVNEEKDGKDKVTSHTELKRLWEQHSNKDDVIAFWLKTCPTFTIPPYQSMTNRKPPGCQSLMLNGNYLEEYYPQWKDWLAKLQPNEEYKEKLQSLQSVKGKRKNKDGRLIDDESIQLRQLQFILDTAKQDDSYKLNEIWSIHHKLQQLKKGKDQWESKLAAAVNDSKLPREITDDLDFIKQDSFWHFVNKYYQTRRKARDGRYFLIQEKKEKWIAEGMLLILCPHKPKQKKRQWDIDLAAVLGIDKNELKDVLATESNSNTPEEYFHKIKSTCNRATKAQKEHRGELKYKISMALTHSNSSEAKDLLNLANKCADLKDELNKATRIANFLVDKKTHEIDQKIALVFKFAQIDNIIFKERSGFSKTCPVCSTDNAFRMQENERGAFASRLPAMSIRLIDGVVMRICDAISRQVARTKWTDIENDVEQGKKVTIPLILEQNRFEFEPSLAKIKGRTLITTDNRETAYNKKKTRIEKAAKGICAYKGTPLGENGEIDHIIPQSSQYGTLNDEANLIYVSKMANQGKSNHTYDLSGSDGLHKNYKNIVFDKKSDDEIKQFIYEQLEGEDAEVGVANVELPFCFGKYLSFVNLNDAQKNAFRHALFLADTDPLKQKVLKSLQNRNRTIVNGTQRYLAQCIADKIYRIAKNVGKEKQLKFDYFEYTAKVDDPKSTYNLRKMYGLIKPKKQPLYSHLIDAQLAFLIAVQDHQAAQERKNGGSMGIKFENNQRVWEFEYDKETGEVLPGKLFNVCNVGEDNFSAIKLDRKKANVREHRSFHRPNFYAQNYIPCFFGKKGEEIEIKAGFSWENSATIKQQKGTEKNILDILHFAKNEFVANSKFESLEDLYRKFNKKTKQGEHDVFYVDWDKTKIQKYLIDEYSTKDMAAGKKWSQEITFLMEFSYRAQKQTIKNIQDIRDALNGMEIGKAKPKPKSYFSIKLFSKNIELPVRSEWVNLLNAWEQKKKSNEGYDDFLRNYFNKSNTAEHKHQKVRKVFSLPVANELGHFLQKRQSWNGDDSYQISADSDSRKDNNKFSRIVLMNNGELKKVINHSFVSPKIFKLKFEDVETDQNYREIDPNQWIKIHQKDGDFPCGVESIEYKIDNATRPQVKIYLSNNFKKDIQVNSENILKSPLTKAKVNSELKKKLTSEDVNNTIEYTASGFNKEIKGKLSEVLS	GTTTCAGTTGCTGGCTTACTAGGCTATGATGTGTTGC	May be truncated
GCA_003711265.1_ASM371126v1_genomic	Cas9_0_CAS-II-B,Cas1_0_CAS-I-II-III-IV-V-VI,Cas2_2_CAS-I-II-III-IV-V-VI,Cas4_0_I-II	1603 aa	MSKIISPIAIDMGAKNTGVYYAHYQRNSTFQEVDKKGQVLVYGNYTPLLVSRTANRHTRRGYTRKKLAKRLLSVVLKEYFDFPAEKHTQALGFLMNRRGFSFLEEEYSKEYLNNLPNDSWEALSEEVQKMLGGQENIADRLMNLATTNPSEVTKLLEAVQNIKGYKDFQEEDKEIKADLVYFDYIEKISKACKLVKKGESIEDDKVTKSKKDKQNLSKTANWVVARLNNENALNLTVATESYQTNLIEEITQINYQQIQNNLPDFEKEREVIKQRKKDNNTSDWKFNYSEFKFNTPNEKGLEEDVARTHLHHFCYAVYKINNEIISGGRHRSKFFDEIKDDLDNFTTNSFLNNKHNPEYLKDFILAINENKGLDVDKLYKLVCHISNFELKPLRKYFNTEAYKSGDKLDNKNLSKITSRWFLQHWIVTKEKDKEAKVEDYEKLKKAWRAHQKCQDDNKNNIISFWLKTDPKLTIPPYQAMLNRHPPKCQSLLLNADYLNKHYPHWEDWLDDLNSNEEYQTKLQSLQRGKGRRKAKNGKEIDGRLLDDEKIKTRQLQFILDTAKKQDFYKLNEIWSVYHKLQKSDVDKIKWQEKLGNLIDASKLPKDLKQDLDFSQEESFGHFLNKYYQTRRKARDGRYFLIQEKKEKWLTNNKLLILCTHKPRQKKHQWQLDLAAVLGVNADDLKAKIVGNPEDYFKSIKGFASNCEKAAKAQKEHRGELKHKIYTELTSNPIDKLGKLANKCSNLKQELIKEIKIPELLADKQTNEVNQKIALVFKFAQIHNIVFKDRSGFSKTCPVCSTDNAFRMQENDKGITKASRLPALSIRLIDGVVMRICDAIARQVAVTKWNDIKDDLQNGTKVSVPLILEQNRFEFEPNLKKIKTAPKPGVKAAEKADEIKDNRTNKYKGQKVSLYSGEGLGSDGEIDHIIPRTGPYGILNDEANLIYISKSDNQNKGNDIKFLKDLHKNYKTEIFGKKSDAEIKEFIYKRLGKDIESKDFAFDKYLSFINLGEKDKKAFRHALFLQEGDPLREKVIRALNNRNRAIVNGTQRYLAQCIADKIHRIAKKENKHNLIEFDYLEYTARWDDPKSTYNLRKKYDLVKSKKQPLYSHLIDAQLAFLLASEDHQNDGTMGIKFDNNQTIWEYETNKETGEISPSKSFNAIDVGESDLKKIDLSPKDSNQKIIDIEKNNSNKKQNLSKIFKRRMFKANAIGERYKPIVKFKDKWYLGYPMTIKGGIYNCDEYCQEVASKNDIKKIEDVVNDDKYYQLTTDHKQIKIWTIKIVDKKYKKINKDSHRYFSQLNPEYNNDEKKEVDQIKFILDKCRYYVAKTDVINAPKIFDKQGYPFYKNWVDFDNAWRKEIGDDYSVKTGEYDLSNNTTKKKWDIFCKNRFSPPENQHKNKHQVKGKKYTMISSGTPSGNVFRVKRNNQNIYQALPLDNNIILKKKSNFLIKHSKNLTLSSAVADKDLAKPIDIEEKIELKNCVIPANNFFKDELKGVKVFLNNTSVDIKDFPLDKFREYFNKEIDVKKSIKIEILTEEQAEIDLTENPNLLYAHKCKVDASIRYTTRDGSVTNIRVGVKERLVSFSLPFRSCEKIYKDSQG	['CCAGCACATCATAACCTAAAAAAACCAGCAACTGAAAC' 'GTTTCAACTGCCGTATTCGAATAGGCGGGTCTCACA' 'GTTTCAACTGCCGTATTCGAATAGGCGGGTCTCAC']	May inactive
MARD_SAMN09104480_REFG_MMP09104480	Cas9_1_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_5_CAS-I-II-III-IV-V-VI,Csn2_1_CAS-II-A	1368 aa	MRKPYSIGLDIGTNSVGWAVITDDYKVPSKKMRIQGTTDRTSIKKNLIGALLFDNGETAEATRLKRTTRRRYTRRKYRIKELQKIFSSEMNELDIAFFPRLSESFLVSDDKEFENHPIFGNLKDEITYHNDYPTIYHLRQTLADSDQKADLRLIYLALAHIIKFRGHFLIEGNLDSENTDVHVLFLNLVNIYNNLFEEDIVETASIDAEKILTSKTSKSRRLENLIAEIPNQKRNMLFGNLVSLALGLTPNFKTNFELLEDAKLQISKDSYEEDLDNLLAQIGDQYADLFIAAKKLSDAILLSDIITVKGASTKAPLSASMVQRYEEHQQDLALLKNLVKKQIPEKYKEIFDNKEKNGYAGYIDGKTSQEEFYKYIKPILLKLDGTEKLISKLEREDFLRKQRTFDNGSIPHQIHLNELKAIIRRQEKFYPFLKENQKKIEKLFTFKIPYYVGPLANGQSSFAWLKRQSNESITPWNFEEVVDQEASARAFIERMTNFDTYLPEEKVLPKHSPLYEMFMVYNELTKVKYQTEGMKRPVFLSSEDKEEIVNLLFKKERKVTVKQLKEEYFSKMKCFHTVTILGVEDRFNASLGTYHDLLKIFKDKAFLDDEANQDILEEIVWTLTLFEDQAMIERRLVKYADVFEKSVLKKLKKRHYTGWGRLSQKLINGIKDKQTGKTILGFLKDDGVANRNFMQLINDSSLDFAKIIKNEQEKTIKNESLEETIANLAGSPAIKKGILQSIKIVDEIVKIMGQNPDNIVIEMARENQSTMQGIKNSRQRLRKLEEVHKNTGSKILKEYNVSNTQLQSDRLYLYLLQDGKDMYTGKELDYDNLSQYDIDHIIPQSFIKDNSIDNTVLTTQASNRGKSDNVPNIETVNKMKSFWYKQLKSGAISQRKFDHLTKAERGALSDFDKAGFIKRQLVETRQITKHVAQILDSRFNSNLTEDSKSNRNVKIITLKSKMVSDFRKDFGFYKLREVNDYHHAQDAYLNAVVGTALLKKYPKLEAEFVYGDYKHYDLAKLMIQPDSSLGKATTRMFFYSNLMNFFKKEIKLADDTIFTRPQIEVNTETGEIVWDKVKDMQTIRKVMSYPQVNIVMKTEVQTGGFSKESIWPKGDSDKLIARKKSWDPKKYGGFDSPIIAYSVLVVAKIAKGKTQKLKTIKELVGIKIMEQDEFEKDPIAFLEKKGYQDIQTSSIIKLPKYSLFELENGRKRLLASAKELQKGNELALPNKYVKFLYLASHYTKFTGKEEDREKKRSYVESHLYYFDEIMQIIVEYSNRYILADSNLIKIQNLYKEKDNFSIEEQAINMLNLFTFTDLGAPSAFKFFNGDIDRKRYSSTNEIINSTLIYQSPTGLYETRIDLSKLGGK	GTTTTAGAGCTGTGTTATTTTGAATGGTCCCAAAAC	Qualified
GCA_003967135.1_ASM396713v1_genomic	Cas9_1_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_5_CAS-I-II-III-IV-V-VI,Csn2_1_CAS-II-A	1367 aa	MDKKYSIGLDIGTNSVGWSVITDDYKVPSKKMKVFGNTGKQTIKKNLLGVILFDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFSSEMSKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINESGVDAKAILSARLSKSRRLENLIAQVSGEKRNGLFGNLIGLSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITQAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFSDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAEAFIKKMTNFDEHLPTKKVLPKHSLLYEYFTVYNELTKVKYVTEQGEAHYFDGHVKQEIFDGVFKKDRKVTKKRLLNFLDNEFDEFRIVDIIGLDESKNAFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDKEMIEERLKKYANLFDDKVMKQLKRRHYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLINDDSLTFKEAIQKAQVSGQGRSLHEQIANLAGSPAIKKGILQTVKVVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENMQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDDVPSEEVVKKMRNYWRQLLNANLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTECDDKCKTIRDVKIITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEITLANGEIRKRPLIETNEETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILAKGNSDKLIARKKGWDTKKYGGFDSPTVAYSVLVIAKVEKGKSKKLKSVKELVGITIMERSSFEKDPVGFLETKGYKEVRKDLIIKLPKYSLFELENGRKRMLASAIELQKANELVLPQHLVRLLYSAQNISGITRSDNLEYIVEHRKEFKEIFEKIIDFSENFILKNKVNSNLKTSFAEQFEVSDAVSLSNSFISLLKYTSFGAPGGFKFLDLDVKQGNLRYQTVTEVLDATLIHQSITGLYETRIDLSQLGGD	GTTTTAGAGCTATGTTGTTTTGAATGGTCCCAAAAC	Qualified
GCA_003363435.1_ASM336343v1_genomic	Cas9_0_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_5_CAS-I-II-III-IV-V-VI,Csn2_0_CAS-II-A	1361 aa	MKSVSKDYYVGLDIGTESIGWAVTDENYNLLKKNGKSLWGIRLFDKADTAEDRRLHRTTRRRGQRKIQRIQLLQELFAEEISEVDPGFYQRLKDSKFLKGDKEEKQTNTLFYEADFNDKKYHESYPTIYHLRKSLLVNNQKNYDVRLVYLALHHIVKHRGHFLFEGNLGNITSFENVFKQLIQYLEDEYDIRLECNSIIEVENVLKSHNINSTEKKLKLINLFSVGKADIQLKAIFGLLAGLKGKLSEVFQDNELKNYEIEKISFAEADYEKQREFLEDELQDKVYLLDKLKAIYDWAILSEILEGGEVDGKCYLSVAKVKLYDDHKKDLKLLKKVIRENCSSEDYKKVFSSNEENANYCAYIGMNKKNGKKQSVKRCKKEEFYGYIKGILKKINIDDNSIKEIEEKMIHGEFLRLQVNRENGVIPYQVHKMELEIILENASTYLPFLNEIDDEGITVKEKIIKIFKFRIPYYVGPLNDFHKNQGSNCWIVRKDPGRILPWNFERKVDIESSAEAFIERMTNKCTYLVNKDVIPKYSLLYMEFMVWNELNNVKIRQEKLPIELKIKVFEDVFKKNKKVTRKKLRDYLRSEGCEFKDDELTGFDGDFKASLTSYIDFGKTEIFGNSIKNKSIQVMIEQIILWITLFGDDRELLQRKIKNIYRKELSDSQIKNICKLKYNGWGRLSREFLQEVEGMDRRTGVVQTLIQALRNTNDNLMELLSLNYTYCVKIEELNAKEQNITELISYETMIRDLYISPKVKRSLWQAVRVMEEIKKVCGKEPTKIFIEMARGTEEKKRTVSRKDKLIELYKNCKNEEREWIKELEQREERDFQDRKLYLYYTQMGKCMYTGKSITLDELYNKTTYDKDHIYPQSKTKDDSIQNNLVLVDKTANWQKDIDIVPLEIQQKMEPFWRQLKHKGLITEEKYNRLTRKTNLTDEELAGFINRQLVETRQATKVTASLLKSLYSTAEIVYVKAKAVSDFRKDELDMLKVRDINDYHHAKDAYLNIVVGNVYHTKFTSNPLKWLKFSKDKKYSLNKMFSYDVMDKEKLAWKRGKDGTISLVRKTMKKNNILFTRHALSVKSGQNGGLFNQQMVGKLKGATVQIKKDNRLKVNIYGGYKTKTNKYFMLVESLDKRGNIQRTIETVPLYLAVEFEKDIEKLLNYCREEFKLQEPKIIIPKIKKNSYIVMNGFPMHLKGSTGSQLIFQGAVQLCLDEDNTKYLKQVYKYLEKNSKRKDKRKNLMISEFDKIHFEKNILIYDMLIQKHKYTIYKYRPAGQIEKLEKGRETFIKLSLEEQCIVIGEIMKLFACKPETKANLKSIGGDSNAGGMKTSKKISNCNMVRLINQSPTGIFEQEIDLLTV	GTTTGAGAATGGTGTAAATTCATATGGTAGTCAAAC	Qualified
GCA_007991355.1_ASM799135v1_genomic	Cas9_1_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_5_CAS-I-II-III-IV-V-VI,Csn2_1_CAS-II-A	1338 aa	MKKQYSIGLDIGTNSVGWAVTSDNHDLIRRRMKINGDYDKGSMKKNFWGVRLFDEGVTAEERRVTRTTRRRLRRRKNRLHYLQSFFMDEMKKVDENFFYRLTESFIVPMDKDHDRHPVFGTMEEEVAYHKQFPTIYHLRKYLADSNEQADLRLVYLALAHIIKFRGHFLIDGELSTEHASNDAVGETFKAFLHHYNHAFHLQPDEAYINLIPEDLSVEAELTAKVSKSRKAESVLHKVETEKSNGTFHQFLKLIVGNQGNFKKTFSLDEDRKLDIGKEEYEEDVEALLEEIGEDYRELFLSAKHVHDAIELSGILEGADARSKAPLSSSMIARYDNHKQDLHALKTLVKAQLPDRYFDIFRNSAKDGYAGYIEGKTSEEDFYKFLKKVLNKVDGAESFIKKIDEERFLRKQRTYDNGVIPHQIHLHELKTIIEKQGQYYAFLLENKEKIASLLTFRIPYYVGPLAKGQSKFAWIEREGHEAVRPWNIKEEVNLVSSAEKFIEKMTNDDTYLPEQKVLPKHSLLLQKFNVFNELTKVQYLDSQGQRRNFSGKEKQEIYHNLFKSEKNRTVSIITFEEYLKNEYQLDNPKVQGIEKKFNASLKTYHDFLKLGMSRDVLDNELYAERLEDLVKILTVFEDREMIRKQLESFSDLFDQAILKKLERRHYTGWGRLSAKLIHGIKDERTQKTILDYLIEDDRPKNNINRNFMQLINDDDLSFKKTIEEMRLKLESNNLKDAVDKIPGSPAIKKGILQSLKIVDEIVEIMGYEPESIVIEMARENQTTSQGQRNAKERLNNIEGALKELKSDLLKKYPVDQEALKNDRLYLYYLQNGKDMYTNQELDINKLSNYDIDHIIPRSFTTDNSIDNRVLVSSKLNRGKSNDVPSEEVVRNMKSFWSLLYRSKLISKRKFDNLTKLKLTDDDKAGFIRRQLVETRQITKHVASILHQRYNNPELDEPAVKIVTLKSALTSQFRQAFDLYKVREINDYHHAHDAYLNAVVARLLMSVYPQLIPDFVYGEYRKTSAFKENKATAKKQFYTNLIERFSKEDKLPDDNGEIAWDKSRDLAKINRVLNYHQMNIVKKVEVQTGNFSKVTVLQKGTSNKLIPRKKDWDPKKYGGFDSPNVAYSIIFTHEKGKQKKVMKDIIGITIMDQQLFEADQIKYLENKGYSNPKINFKLAKYTVFELENNRRRLLASANEAQKGNQMVLPSHLVLLLYHCRRYIDSSFKNANHETYILENKEKFNELLDYVLQFSRQYTLAEVNDRKISAVFEGAFNQAEASDLATSFINLMDYNKMGAPSAFTFFETKIDRKRYTSLKELLNAKIIYQSVTGLYETRKEV	GTTTTAGAGCTGTGTTGAATTGAATGATTCCAAAAC	Qualified
MARD_SAMN05421734_REFG_MMP05421734	Csn2_1_CAS-II-A,Cas2_5_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_1_II	1337 aa	MKKKYTIGLDIGTNSVGWAVIRDDYDLIRKRMKINGDHEKGKMKKNFWGVRLFDEGETAEDRRIRRTTRRRLRRRRNRVQYLQSILKDELNRVDENFLHRLNESFYVPIDKEHNRHPIFGTLDEEHAYHKQFPTIYHLREYLANSNEQVDIRLVYLALAHIIKYRGHFLIEGELNTENSSVEETFADFLNHYNQMFSLQDDGSYINPLSEDISVEYDLTEKTSRSLRAENVLNKFGTEKRNGMLHQFLKLIVGNQGSFKKVFSLTEDRKLNMTKETYEEYLEELLVDIGEDYRELFLAAKNAHDAIELSDILSVSDDKTNASLSSSMVKRYEEHKQDLSALKNFIKVNIPDRYYDLFRNPKKNGYAGYIEGKTTEEDFYKYLKKELDNVEGAGMFLQKIDEERFLRKQRTYDNGVIPHQIHLEELRAIIEKQSAYYPDLLNHQEKIESLLKFRIPYYVGPLAKEQSKFAWIKRLDDTSIRPWNIEEKVDLIGSAESFINEMTNYDTYLPNEKVLPKHSFLIQKFNVYNELTKVQYVNDQRKLCNFSSHEKKEIYENLFKEKRTVSQKDLEDFLINEYQLENPSIKGIENKFNASLKTYHDFIKVGVPRAVLDNDNYENELEEIVKILTVFEDRKMIRQQLQKFSHVLDYDILKKLQRRHYTGWGRFSQKLIDGIRDEHTQKTILDYLIDDDAPNKNINRNFMQLINDLDLSFKKIIETAGLQTESDKLKDIVDQIPGSPAIKKGILQSLKIVDEIVYIMGYEPEAIVVEMARENQTTTQGQKNAKERLKRVEESLKDMKSDLIKKQPVTQDALQNDRLYLYYLQNGRDMYRNQELDINNLSNYDIDHIIPRSFTTDHSIDNRVLVSSKDNRGKSDDVPTLEVVKNMKPFWSSLYRSNLISKRKFDNLTKAERGRLTADDKAGFIKRQLVETRQITKHVSSILNERYNEQNTEENHNVKIIPLKSALTNQFRKAFDLYKVREINDYHHAHDAYLNAVIANLLLKVYPHLEPGFVYGEYRKTNMFKENKATAKKQFYSNLIERFKHEGQIIDENGEIVWDMNMHVEKIKRVLSYNQINIVKKVEIQTGNFSKETVQPKGESSSLIPRKKGWHPKKYGGFDSPNIAYSIIFTHEKGKKKKEVKEMIGITIMDQSIFEADETKYLEEKGYVNSILKFKLPKYSLFELENGRQRLLASANEAQKGNQLVIPTHLMKLLHYCNRYIDSGFKNVDYEIYLNDNRHHFSELLNHVLNFSAAYTLASANEEKIKKLYIEKSDQAEVGELATSFINLMDLNRMGAPSTFKFFGTNIERKRYTSIKELLDATIVFQSITGLHETRKEV	['GTTTTGGGATCATTCAAAACAACATAGCTCTAAAAC']	Qualified
MARD_SAMN05421668_REFG_MMP05421668	Cas9_1_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_5_CAS-I-II-III-IV-V-VI,Csn2_1_CAS-II-A	1336 aa	MKKAYSIGLDIGTNSVGWAVTSDDHDLIRRRMKINGDHDKGSMKKNFWGVRLFDEGVTAEERRVNRTTRRRLRRRKNRVHYLQSFFMDEMKKVDENFFHRLTESFIVPMDKDHDRHPVFGTMEEDVAYHNQFPTIYHLRKYLADTDQQVDLRLVYLALAHIIKFRGHFLIEGELSTEHASNDAVRETFKSFLHQYNQAFHLQPDDTYINRIPEDLSVEAELTAKVSKSRKAESVLHKVETEKSNGTFHQFLKLMVGNQGNFKKPFSLDEDRKLDMGKEDYEEDVESLLAEIGEDYRELFLAAKDVHDAIELSGILDGAEPHTKAPLSSSMIARYNEHKQDLQALKTLVKAERPDRYFDIFRNKAKDGYAGYIEGKTTEESFYKFMKKELSKVDGAEAFIKKIDEERFLRKQRTYDNGVIPHQIHLHELKTIIENQGQYYPFLLDNKEKIESLLTFRIPYYVGPLAKGQSEFAWIEREGGEAVRPWNIKNEVNLVGSAEKFIEKMTNSDTYLPDEKVLPKHSMLLQKFNVFNELTKVKYSDEQGQVCNFSSKEKRDIYQNLFKLHRTVSLKVFEDYLRNDYQLENPKVQGIEKKFNASLKTYHDLLKLGISRDTLDNEAYADRLEDLVKILTVFEDREMIRKQLEGFSDLFDQATLKKLERRHYTGWGRLSAKLINSIKDERTQKTILDYLIEDDGPKKNINRNFMQLINDDNLSFKKTIEEARLNLESNDLKDAVDQIPGSPAIKKGILQSLKIVDEIVGIMGYEPESIVIEMARENQTTSQGQQNAKERLNNVEGALKELKSDLLKKHPVDQDMLKNDRLYLYYLQNGRDMYTNQELDINKLSSYDIDHIIPRSFTTDNSIDNRVLVSSKSNRGKSDDVPNKDVVQKMKPYWTSLYRSKLISKRKFDNLTKIKLTADDKAGFIRRQLVETRQITKHVASILHQRYNNPELDEPAVKIVTLKSALTSQFRQAFGLYKVREINDYHHAHDAYLNAVVARLLMSLYPQLIPDFVYGEYRKTSAFKENKATAKKEFYTNLIERFSKEDKLPDENGEIAWDKARDLAKINRVLSYHQMNVVKKTETNSGQFFDINIKAKGANNKLIPIKDHLDVTKYGGYGSPNIAYSIIFTHEKGKKAKVVNEILGINIMDRTVFEKDREKYLKDRGYVNPIFKYKVPLNTLFEFEDGKRRILTGASELQKGNQFILPKHLVNLIYYCNKYVNSQFKNVDYENYIYDHKHQFSELLNCVLEFGQKYVLSLSKVEKIRNAYKESFNQAEIKEIVASFITLMDLNKMKPSADFNFFGVKIGETRYRSTKEVLTSKIIYQSVTGLYETRKEV	GTTTTAGAGCTGTGTTGATTTGAATGATTCCGAAAC	Qualified

GCA_007990305.1_ASM799030v1_genomic	Csn2_1_CAS-II-A,Cas2_5_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_1_II	1336 aa	MKKAYSIGLDIGTNSVGWAVTSDDHDLIRRRMKINGDHDKGSMKKNFWGVRLFDEGVTAEERRVNRTTRRRLRRRKNRVHYLQSFFMDEMKKVDENFFHRLTESFIVPMDKDHDRHPVFGTMEEDVAYHNQFPTIYHLRKYLADTDQQVDLRLVYLALAHIIKFRGHFLIEGELSTEHASNDAVRETFKSFLHQYNQAFHLQPDDTYINRIPEDLSVEAELTAKVSKSRKAESVLHKVETEKSNGTFHQFLKLMVGNQGNFKKPFSLDEDRKLDMGKEDYEEDVESLLAEIGEDYRELFLAAKDVHDAIELSGILDGAEPHTKAPLSSSMIARYNEHKQDLQALKTLVKAERPDRYFDIFRNKAKDGYAGYIEGKTTEESFYKFMKKELSKVDGAEAFIKKIDEERFLRKQRTYDNGVIPHQIHLHELKTIIENQGQYYPFLLDNKEKIESLLTFRIPYYVGPLAKGQSEFAWIEREGGEAVRPWNIKNEVNLVGSAEKFIEKMTNSDTYLPDEKVLPKHSMLLQKFNVFNELTKVKYSDEQGQVCNFSSKEKRDIYQNLFKLHRTVSLKVFEDYLRNDYQLENPKVQGIEKKFNASLKTYHDLLKLGISRDTLDNEAYADRLEDLVKILTVFEDREMIRKQLEGFSDLFDQATLKKLERRHYTGWGRLSAKLINSIKDERTQKTILDYLIEDDGPKKNINRNFMQLINDDNLSFKKTIEEARLNLESNDLKDAVDQIPGSPAIKKGILQSLKIVDEIVGIMGYEPESIVIEMARENQTTSQGQQNAKERLNNVEGALKELKSDLLKKHPVDQDMLKNDRLYLYYLQNGRDMYTNQELDINKLSSYDIDHIIPRSFTTDNSIDNRVLVSSKSNRGKSDDVPNKDVVQKMKPYWTSLYRSKLISKRKFDNLTKIKLTADDKAGFIRRQLVETRQITKHVASILHQRYNNPELDEPAVKIVTLKSALTSQFRQAFGLYKVREINDYHHAHDAYLNAVVARLLMSLYPQLIPDFVYGEYRKTSAFKENKATAKKEFYTNLIERFSKEDKLPDENGEIAWDKARDLAKINRVLSYHQMNVVKKTETNSGQFFDINIKAKGANNKLIPIKDHLDVTKYGGYGSPNIAYSIIFTHEKGKKAKVVNEILGINIMDRTVFEKDREKYLKDRGYVNPIFKYKVPLNTLFEFEDGKRRILTGASELQKGNQFILPKHLVNLIYYCNKYVNSQFKNVDYENYIYDHKHQFSELLNCVLEFGQKYVLSLSKVEKIRNAYKESFNQAEIKEIVASFITLMDLNKMKPSADFNFFGVKIGETRYRSTKEVLTSKIIYQSVTGLYETRKEV	GTTTCGGAATCATTCAAATCAACACAGCTCTAAAAC	Qualified
GCA_001710555.1_ASM171055v1_genomic	Cas9_1_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_5_CAS-I-II-III-IV-V-VI,Csn2_1_CAS-II-A	1337 aa	MKNMKNPYTIGLDIGTNSVGWAVLTNQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDDGQTAVDRRMNRTARRRIERRRNRISYLQEIFAVEMANIDANFFCRLNDSFYVDSEKRNSRHPFFATIEEEVAYHDNYRTIYHLREKLVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDEVYKQFIETYNQVFMSNIEEGALAKVEENIEVANILAGKFTRREKFERILQLYPGEKSTGMFAQFISLIVGSKGNFQKVFDLIEKTDIECAKDSYEEDLETLLAIIGDEYAELFVAAKNTYNAVVLSSIITVTDTETNAKLSASMIERFDAHEKDLVELKAFIKLNLPKQYEEIFSNAAIDGYAGYIDGKTKQVDFYKYLKTILENIEGSDYFIAKIEEENFLRKQRTFDNGAIPHQLHLEELEAIIHQQAKYYPFLKEDYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKADGEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYMVYNELTKIRYIDDQGKTNYFSGREKQQVFNDLFKQKRKVKKKDLELFLRNINHIESPTIEGLEDSFNASYATYHDLLKVGMKQEILDNPLNTEMLEDIVKILTVFEDKPMIKEQLQQFSDVLDGGVLKKLERRHYTGWGRLSAKLLVGIREKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEKEQVSTTDKDLQSIVAELAGSPAIKKGILQSLKIVDELVSIMGYPPQTIVVEMARENQTTGKGKNNSKPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMYTGQELDIHNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGGDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILYQRFNKETDNHGNTMEQVRIVTLKSALVSQFRKQFQLYKVREVNGYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFAQKERIIDENGEILWDKKYLETIKKVLDYRQMNIVKKTEIQKGEFSKATIKPKGNSSKLIPRKENWDPMKYGGLDSPNMAYAVIIEHAKGKKKIVIEKKLIQINIMERKMFEKDEEAFLEEKGYRHPKVLTKLPKYTLYECEKGRRRMLASANEAQKGNQLVLSNHLVSLLYHAKNCEASDGKSLKYIEAHRETFSELLAQVSEFATRYTLADANLSKINNLFEQNKEGDIKAIAQSFVDLMAFNAMGAPASFKFFEATIDRKRYTNLKELLSSTIIYQSITGLYESRKRLDD	GTTTTAGAGCTATGTTATTTTGAATGCTACCAAAAC	Qualified
GCA_001400055.1_ASM140005v1_genomic	Cas9_1_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_5_CAS-I-II-III-IV-V-VI,Csn2_1_CAS-II-A,AbiEii_0_CAS-III	1337 aa	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTALDENFFARLQESFLVPEDKKWHRHPIFAKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVKEKFQQFMIIYNQTFVNGEGRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQEKANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSYAKLSSSMIVRFTEHQEDLKKFKRFIRENCPDEYDNLFKNEQKDGYAGYIAHAGKVSQLKFYQYVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDASTFAWLKRQSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQFNASFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGVSKHYNRNFMQLINDSQLSFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDMYTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNAKSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQTFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKELNYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPIVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRRLLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARYTSIKEIFDATIIYQSTTGLYETRRKVVD	GTTTTAGAGTCATGTTGTTTAGAATGGTACCAAAAC	Qualified
MARD_SAMN07201002_REFG_MMP07201002	Csn2_11_CAS-II-A,Cas2_5_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_0_II	1334 aa	MAQKPYVLSLDIGTGSVGYACMDKDFNILKYYDKDAIGVYLFDGAQTAQDRRQFRTARRRNNRRIKRLGLLQEILAPLVQNPNFYQFQRQFTWKNNNMDFKNKSLSEVLNFLGYAPKKYPTIYHLQKALLLKDEKFDPELIYVALYHLVKYRGHFLFNHLNVENLSDNDSMDDFVELIELYESINHISLNLDYDKIKAIYEILQDNEMTKNDREKAIVKIEKALKHFSKMLLGLKFKEADLFSLADNAEELKEAKQSHSFADNYEENLTSFLTVEQLELIERANKIYLALTLQDILNGNKSLAISKVAAYDKFKNELKLVKDIVYQADPTKAQFKKIFVSSKESLKQYEAIPNKENLSQLCLFDQYLIRPKDQYKKLMDELKKFIPQESPLYFEAQNDTLLKVLNTTDNASIPMQINLYEAETILRNQQKYHAEITDDMIEKVLSLIQFRIPYYVGPLVNDQTIAKFGWMSRKSNKPIKPWNFDEVVDRSKSATQFIRRMTNKCTYLMHEDVLPKNSLLYQEMELLNELNAIQVRLQTDPKNRKYRLTPQIKSFAIEHIFKKYKTVSHSKFLEIMLNSNHRETFMNHGEKLSIFGTQDDKKFVSKLSSYQDMTKIFGNIDDKRGQIEEIIQWITIFEDKKILVQKLKENYPELTSQQINQLKKLNYSGWGRLSERLLTHKYHDHSIIELLRSSDENFMEIITNDAYGFQNFIKEENQVQSNTIQYQDIANLATSPALKKGIWSTIKLVRELTSIFGEPEKIIMEFATEDQQKGKKQKSRKQLWDDNIKKNKLKSVDEYKYIIDVANKLNNEQLQQEKLWLYLSQNGKCMYSGQSIDLDALLSPNATKYYEVDHIIPRSFIKDDSIDNKVLVIKAMNQTKGDQVPLQFIQRPYERIAYWKSLNKAGLISDSKLHKLMKPEFTAMDKEGFIQRQLVETRQISVHVRDFLIEEYPNSKVIPMKAKMVSEFRKKFEIPKIRQINDAHHAIDAYLNGVVYHGAQMAYPNVDLFDFNFKWEKVREKWKALGEFNTNQKARELFFFKKLEKMEVSQGERLISKIKLDMNHFKINFSKKLANTTQQFYNQMPESPKIAEIKYESNKSADAVYTSLKPHQTFVIAIKTATKKGKVKMQYQMIDYYVIDFYKFKNTDEKELALYLARRENKDEVLDAQIVYKLNKGDLLYLNNHPCYFVSSKEVINAKQFELTVEEQLSLYKVMHNKETSVEKLIKVYDLISEKVINDYRHYLDSELKIERVRTLFSESHQTHEDFIKALGELFKVVTASATRSEKIGSRKSGMGHRAFLGKGKDVKIAYTSISGLKTTKPKSLFKLAESRNEL	GTTTTACTTCTTTCTAAATAAACATAGTGCTAAAAC	Qualified
GCA_001998885.1_ASM199888v1_genomic	Csn2_1_CAS-II-A,Cas2_5_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_1_II	1332 aa	MRANYTIGLDIGTNSVGWSVIKEDYQLVRRKMPIYGDTTVKAQKKNFWGVRLFDEGQTAESRRIKRTTRRRYLRRRNRLNYLQTIFKDDIHQLDAHFFHRLEDSFLVKDAKRHTKYPIFGTLDEEINYHNEYPTIYHLRKELADSTQKADIRLVYLAIAHIVKYRGHFLIEGQLNSANTSINNTFQQFLNNYNEKFITQVPGLALSPVDTTVLVESIVTAKTSRSRKSDAILALFPSEKSTGTFAQFIKLIVGNQGNFKKTFALEADAKLQFSKIEYEEELGELLAEVGDDFADVFVAAKNVYDAIELSGILSTKDTETNAKLSASMIERYTQHQNDLKLFKKYVREYLPKQYDTIFKDSSKKGYAGYIDGEATEEEFYKFVKKTLDNTPDAAYFIDKIDQENFLRKQRTYDNGVIPHQIHLDELNAILENQSKYYTSIEKNKDKIIDIMTFRIPYYVGPLSNHNSSFGWLTRKSPGEIRPWNFTERVDTYQSSVDFIERMTNNDSYLPTEKVLPKNSFLYQKYMIFNELTKLTYTNEKNERLNLSGNEKREIVNHLFKKHRKVTRKLLEDYFANYYQLDSTQIDGIETAFNAKYTVYHDFVKLGVPQEFLDDEANVDTLEDVVKLLTVFEDRKMIREQLKKYDTCFSQETLKKMERRHYTGWGRLSQKLLVGLKDKTTGKSIMDYLIEDDGMPKNINRNFMQLINDVSLSFKEDIAAAQPTVQEEDIRQVIHGLAGSPAIKKGILQSLTIVNEIVGIMGHPPQSIVIEMARSNEISKKTNSRLKALEKMLADFQSDLLKEYPTSNDKLKNDKLFLYYLQAGKDMYTGADLNIHALSNYDVDHIIPQSFIKDDSLDNCVLVSSKANRGKSDDVPSSEIVKKQKYFWKKLLDAKLISKRKYDNLTKSERGGLTPADKEGFIRRQLVETRQITKHVAQILDSQFNDGQEDNQAVQIVTLKSSLTNQFRKQFNLYKIRELNDYHHAHDAYLNAVVGSLLLRKYPQLKPEFVYGKFVKGRLINRFKATQKKDLYTNIMKFLKTDDRVADANGEIIWSPTTIKTIKTIKRTLSSKQMSIVKKVEKQTDKLTNETIYPKAAQGTLIPRKNGLDSTVYGGFNSPTIVYSVLISHVKGKKKALTYDVIGISLLKQKTYEADKIAYLESLGFISPTLHFELPKFSLFLQEDGTKRFLASADELQKGNQLVLSDKMMALIAHSKKAIAGNLDSLDYVTTHREDYDRLLNDILVFARKYLAADNVLKKIEAAYEINRQNTIIETAEAFINLLKFTKIGAAMEFNFYDCKIARHRYRTKADFTAIFDGCVVNQSVTGLYESRWKIKR	['GTTTTGGAAGCATTCAAAACAGCATAGCTCTAAAAC' 'GTTTTTGTACTCTTAAAGATTTAGTAACGGTAAAAC']	Qualified
GCA_003987515.1_ASM398751v1_genomic	Csn2_1_CAS-II-A,Cas2_5_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_1_II	1332 aa	MRANYTIGLDIGTNSVGWSVIKEDYQLVRRKMPIYGDTTVKAQKKNFWGVRLFDEGQTAESRRIKRTTRRRYLRRRNRLNYLQTIFKDDIHQLDAHFFHRLEDSFLVKDAKRHTKYPIFGTLDEEINYHNEYPTIYHLRKELADSTQKADIRLVYLAIAHIVKYRGHFLIEGQLNSANTSINNTFQQFLNNYNEKFITQVPGLALSPVDTTVLVESIVTAKTSRSRKSDAILALFPSEKSTGTFAQFIKLIVGNQGNFKKTFALEADAKLQFSKIEYEEELGELLAEVGDDFADVFVAAKNVYDAIELSGILSTKDTETNAKLSASMIERYTQHQNDLKLFKKYVREYLPKQYDTIFKDSSKKGYAGYIDGEATEEEFYKFVKKTLDNTPDAAYFIDKIDQENFLRKQRTYDNGVIPHQIHLDELNAILENQSKYYTSIEKNKDKIIDIMTFRIPYYVGPLSNHNSSFGWLTRKSPGEIRPWNFTERVDTYQSSVDFIERMTNNDSYLPTEKVLPKNSFLYQKYMIFNELTKLTYTNEKNERLNLSGNEKREIVNHLFKKHRKVTRKLLEDYFANYYQLDSTQIDGIETAFNAKYTVYHDFVKLGVPQEFLDDEANVDTLEDVVKLLTVFEDRKMIREQLKKYDTCFSQETLKKMERRHYTGWGRLSQKLLVGLKDKTTGKSIMDYLIEDDGMPKNINRNFMQLINDVSLSFKEDIAAAQPTVQEEDIRQVIHGLAGSPAIKKGILQSLTIVNEIVGIMGHPPQSIVIEMARSNEISKKTNSRLKALEKMLADFQSDLLKEYPTSNDKLKNDKLFLYYLQAGKDMYTGADLNIHALSNYDVDHIIPQSFIKDDSLDNCVLVSSKANRGKSDDVPSSEIVKKQKYFWKKLLDAKLISKRKYDNLTKSERGGLTPADKEGFIRRQLVETRQITKHVAQILDSQFNDGQEDNQAVQIVTLKSSLTNQFRKQFNLYKIRELNDYHHAHDAYLNAVVGSLLLRKYPQLKPEFVYGKFVKGRLINRFKATQKKDLYTNIMKFLKTDDRVADANGEIIWSPTTIKTIKTIKRTLSSKQMSIVKKVEKQTDKLTNETIYPKAAQGTLIPRKNGLDSTVYGGFNSPTIVYSVLISHVKGKKKALTYDVIGISLLKQKTYEADKIAYLESLGFISPTLHFELPKFSLFLQEDGTKRFLASADELQKGNQLVLSDKMMALIAHSKKAIAGNLDSLDYVTTHREDYDRLLNDILVFARKYLAADNVLKKIEAAYEINRQNTIIETAEAFINLLKFTKIGAAMEFNFYDCKIARHRYRTKADFTAIFDGCVVNQSVTGLYESRWKIKR	GTTTTGGAAGCATTCAAAACAGCATAGCTCTAAAAC	Qualified
TARA_SAMEA2621779_METAG_AKANJMCB	Cas9_0_CAS-II-B,Cas1_8_CAS-I-II-III-IV-V-VI,Cas1_0_CAS-I-II-III-IV-V-VI,Cas2_7_CAS-I-II-III-IV-V-VI,Cas4_0_I-II	1332 aa	MNKLTISIDMGAKNNGVFIVKTQNKKVLDKKATCIIVNKGSINFSKKSRRENRHKDRNYKRRKLSKRLISELIDLSKYDNKQTEQIFGLLNNRGYTFISTSTEFEKLNDITSEFINMYFEDLKGLSSREDFENKISEFEDLENLEDFIENINEKVDFECKKKKAEFYKNYQDSFSNFIKKDLQNIKKLFSSILNEIQIGSKPRKKYFDEIKDEIEKDFDFIENFSKVEFFNIISNISNLQLRVLRKYFNNKYDDVLDTKKLDIKIRAYFKAFHYKSEKERTQRVELFEVLNKEKDILNFLKNTNPMLTIPPYEDMNNRNTYKCNSMLIKPSLITKDLENTIDLLLKSDYFSTLLISNDGEFKKESLIKIKPATNNTYIDTDFTYSKYLQRILDTKIDDDNKIYHPRVVFKNKNPKAIELFKKLFGTKTYDNLLDIATSYYEEETKIHNGIYKENTSIFEKCNTNTSYKNNVKELLLKPLYSHNFTHDETKEFISEIENYEEKINRTTLLGFLKFVEEEYKKYQNSFYHIITHNFELENIKDVKDKDIKKLILGLDSCVDSLKNILNKLNIKTYLNDIKKVEKENLSKVLNILKQTYNILFYDLSGFSKTCKSCTKENALRSDENFSIGKRLLSDVAKPIDGMLDMMIDRLAYEITQQINKKDIENCESLEILLEQNRFEFEENLNEIKRANDSSIKRYKNDTKNICPYLDTTFEKGEYDHILPQSKEVYNSKANMIYCSSKGNREKSNKIYKLEDLNIKHLKEIFGLEELEKIKEFISKYIKTINEKDFKNFDNLKLNQQKAFRYALFMPTNSEEYKKALNLVKIDKLKTFSNGTQKRLARFIYEKLVNKYPREFKNKQINIDSKTVDSMLVSATRKSLSSSNKELEKEKYQESHSHCIDAMIVFYLANSKIEGRADRKKDNLTTLIPRYNFEDIYLKESGINNLSKNVTFINSPKKEVGSYKLFDDTIYAENYKHITKESLKKDEFECLIYHSILFKNQKSKKVFINNIEELKDESIYKIDVQKVSNLIYTLFNQRDKSTLLKLKFLDKLQYFTSRKEIESIFFDEKRTKLRKFNEIKNIPYFSVNLYKAVYKKLENKRDLFIVENDKSVLNSKVLNDLLIEMFESKQKDENKEKRKRAKKRHKYTLPILGSPKFKIKRANTWQVLGNNDIATKNYIINGDIKPIAFFGKNTYPVKVKDLLDCLLVNEDTKSVYHVPINISEISKYVSKLEYLVTEAKRCTLLVTFNKNSFKDINFNEIKVFDGTKDEIFKLLINNYIDNKDLNISNYIGSIRDGLKAKAMLIDSNSTTITLQYKAAINVSKKNIILENLK	CTAACAAATTATAACTAGACACAACTTAGTTTGAAAC	Qualified
GOMC.bin.32054	Cas9_0_CAS-II-B,Cas1_0_CAS-I-II-III-IV-V-VI,Cas2_2_CAS-I-II-III-IV-V-VI,Cas4_0_I-II	1559 aa	MNKIISPIAIDMGAKNTGVYLNHFEQGEYPTSSGNSQGKTIVIDSTNITWSQADRTGKRHQIRANKRRKLSKRLLKLVLNEYDLQPNQAQSEYLSGLLNRRGYNRIEITEEHEGILQSKLVSDFFSLIIFNNSNNNNILDGLLKELTIKNDEDKIKWENVLKNKIFSMSENKNSEFKKEYIKNKYEKDEVTELSKAFTAIKNTIEKQIDNYATGHLHRKNYLENIQKDIEVELLKPLLNNSLTVEKLAYLIGNISNLQLRVLRKYFNEKSMQTDDVWLPEKLHLVFFKWIRSWHTKQQNEKDNRIEILALKDNKDILDIFTSLDPKKTIPPYEDQNNRRPPKDLTLRLNPKALDGQFSGWESITTSLANNYILPPINNHITERIDIQKGLEDNVSPQKNKGEIQDTKERQILADTLHRILDRTIILDPYKLRWLAQGDTSRGATTAKALLNQHSNNNQAGIIVRLAKKYYAEVEVAKQGLWAEAGSLFFRCNTNPPHKGKLTHKLVGHILREEFNDERIKEFRQECWEVKTGRSTIKSMAGKIEATRKKYGNGFNYIVQTIKRRNYVLANNSPAQAQKDRWVKYAEDHKDVIEAINNTKLVAEKINDFLDSNNQEKYNNPYSIAQLYNHLESEITGFSKTDKFNTEENAWRDKTQSIEILNKEGGVIKVTRSNAVRLTADSIRPFDGLLDRIISRQAIEIARAKIAQFESLNIDKNDILFVPIFMEQNRFKFEQDMHEIKGAKNVKKKAKDNAQKGLEEQEKEWQDKNTRVKKNRHCPYTGEIITSGEIDHIVPQSESKRNQDVVFNCEANLIYCSSVGNHSKSNNRYTFEQLNHNYLSEIFAGETDVKQTIINFVNGLDKNDPISFHNLDDKEQNYLRHALFIPELDSQTFPILNTRYKTFVNGTQGYLGKKIRKLLQKEYPNIKVKTYQIPAQEVSQLRTVLGEQDKAFEKQERQGAFSHVVDAALVLAAALQKPKIAEELTTIHLTDELINKTEHGQWLKDLLPKNASVQHIKRKPKYRKKLESTQIFKEGLYGERFMPILLDDEKLYYGFSLDNCHQIEPLKHPKKTVKDFDKKLQTAKQNQQVKHSEYFELLKPFLYTGKKNAKEPVTGDLFDHWQYQYLSIDKTKALTHLQKCAKEVCGEIEIKQAQQLEKLRYSVEKKEIKSVLLEGQGKRTFIKQLDPKKFTVAKLTLPAKNEWTALIKHPIKDCTGKQTTLEACFGELEQVEIDNPALCGFWETLAAESSLEIDYLKDNLLKKNDKGELSKKSGVIPKTVFNIKIKTDEQLKGLEYSPLLRKSFGVDFKVMTDLIPQSSWGGLFKAFFRTDKIKNSSHHKQVRKDYSLPIVSAPSGGFRIKRKNPLTGETVYQVSSIADGGYACGFEKDEKGNINMTKTGVTFHAGLKESPNVAYIGKGGKPNIKKQSNGNTTEPFGSTREVNLNVSAIKSSYLTLAEARRIDIRILIEKQLFETLCEFSDIDANNPPSNINITFDKKTEIEKHKEFKKLMSDQQIPMPRDSGKNRWAYKLSVQVLNQDIELFYRIGLTDPMRDAYQNGTA	CCAGCAAATCATAACCTAATAAACCAGCAACTGAAAC	May inactive
MARD_SAMN07739585_REFG_MMP07739585	Csn2_0_CAS-II-A,Cas2_5_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_0_II	1331 aa	MLDKKSYYLGLDIGTNSIGWAVSDEFYNILKIKGQKTWGVRLFDDAMPAAERRGFRSSRRRGDRVKWRLGLLRNLFSSEIDEVDNKFFKRLEDSKFFADDKGENQTNTLFNDIKFKDRDFHKKYNSIYHLRYDLITKDEKFDLRLVYLAIHHIVKKRGHFLFDGEIGDNKNNFSELTQNLIDELLALDINVDFLDNFQEVLKDESLRLGDKKAKLKEIITVSSEKFDEKKVLTQIVNLLCGKGNLKIFFPESEEKKIEISLLSDSFDENFDNLLAKFGEDANFIKPLKDIFDYILLQKIMKDSSTISESKIKSYKKHKKDLKKLKDLILNEYGKDKFNEVFKSSTKDNYPAYVGHFKTKKLKDEHKKTTKDDFYAYIKKIINKSDSQIVKTIKDDIELGNFLPRLIDGNSVIPHQIHKSELIKIIDNQTKFYPFLKDIKDKIISIFEFRIPFYVGPLNDYHKDKGGNAWIVKNSNEKIYPWNFDKVVNIKQSAQNFITKMTNKCSYLLGEDVLPKNSLLFSKYCVLNELNNLKVNDEDISVELKKELYENLFKKQKSVSKSSIKKHLINNGYYDKNLIIEGVDDKLTSNLKSYIELKEIVGNSLSDEEKEEIIKYIALFGYDKKMLKSYILDISPSLDEKTLDRLCTLKYSGFGSLSKKLLVGIKSNDKTLIQILEDTNKNLMQLVSDEFKNEIEKHNNEYKKQNNLLDVGLSYGLLDNLYVGPSVKRGIWQSLRLIDEIISAFKYPPKKIFIEMAREEGRKERTKSRKDQLLELYKSIKKDGVKEFGIEFDKVKTNLEKEEQSRLNSKKLYLYYMQFGKDIYAGNNIDISKLFDENIYDIDHIIPRSIKKDDSFDNLVLTNKISNKDKTDVFPVSFEVQNKMRSFWKMLKDKGLMSSKKLHSLTRTTPLSEDEKAEFIARQLVETRQSTKAVATLLKDLYKNSEIVYVKAGIVSEFRQKFDLIKCRDLNDLHHAKDAYLNIAVGNAYNVKFTKNPRNFFKKYKGEHYNLKIEKLLERDIIRGQEFGWDAKKSIKIVKDTMINKNPYLVTYMSYEANGGLFNQQPVKAKEGLIPLKADPRLKNTSRYGGYNGKSMAYFFIFEYKKGKNRQLRIDTIPIYLKKDIENGKLTLKEYCQNNLGFEDVVIKLPKLLKNSLIKVDGYLMKMTGNNEDKILCRNAVQLILNSKEELLLKKVLKIVDKINEAKRFKKEYFINKYQFDIEDKELLELYDRLTKKHETAYKNKPASQIETLKDGREKFISLSLEDKCLLLSELLKLFDTTAQNANLSLIGGSPNAGSVKFVKNLSDNKTYKIIFESISGFYKKEIDLNKL	GTTTTATTACCATCAAATTTTACACAACTCTCAAAC	Qualified
GOMC.bin.1060	Cas9_0_CAS-II-B,Cas1_0_CAS-I-II-III-IV-V-VI,Cas2_7_CAS-I-	1148 aa	RRYFNDENMAYGDIWDFERLHKVFYRNVSGMHYKDNTQDKQNQVTLFSMRAQDIISTFEALDPIVSIPPFEDQNNRRPPKCQSLKLSPTTMIKYLPKWHIIVPQLADDLEARSILVKDGLELNSIKNTNDWAQCLQRILELSYDHDPYLLRRWIGRKDQVGEHTKRAISNFERVIQQQHHADFIVLAELFYAETQAARNSIWFKENSQLLEVCAQKTPHKANQFKELLSGTVNVNLTEDQKEQFEKLWFDNPRIGKRGIKGWCGFAAEMQKEYGNALKEQINRNIWLNENDKKVENKKLIELNKAVESVAKIIADSMNIESKRFANVFDLAKLHNLVEKDQKGFSNNCRCCTKDNAWRGNIVENDEGKDAAFGLRLPADTTRPFDGLLARLIDKQAYQIAQAKVAQLLSEPTNEMVSVPLLIEENKFRFTSSLAEVKKNTKKRKDNEDKAKKREGLWLEKNGRIKEASKNICPYTDKPLYNGGEIDHIISRAESKGRKQTVFNHEANLIYCSVTGNGIKDRQSYSLADLKPKYLTKQFETSNREKIKQRLIALCEPIIAKRELGSFGDLNEDTRQAIRHGLFVDELRPKLIPLLHQSNKTRVNGTQAYLAKLILQKIKLLYKGPIEFSVHYVEAELASRLRVQLGNQYPEFAKPEHQPVASHVIDAAMVMAAGLQQPHTQACLQTAGLDNNGDWLKAILPTEIDVKRIEPKLKFNKKQVHSQTLFKDGIYAEHVIPLVVNDEKFGVGFTADNVAWLKNQEEAPVIWWNALQLYLNKSPETLDDVRKNIGNQTKVYSFNNTAVFELLYKVAKQPCSDEELLSADLLETLI	GTATCACTAGCTTATCAAACTTACGTAGTAATGAAAC	May inactive

	II-III-IV-V-VI,Cas4_0_I-II		YTSQKVKLQSALYDENNKTYKKITDIVNMKTTDNKGDFTIKLDVKSRYLKTTMRAFESSLIVYPAIASWSNLLEHDFVKAKAGKKEAFDSDSYNRVIKGLFSTIKENTRQHKKVRKNWSLPKLSSPSGGYRARRKNADGSIIWQLFAVEGLSASGFYMIDGKIDFS EAGVAGIKAIDDSKNLTSVKSRYKNIQLPVEPFDNWRRVTIPDKLKGKIKGILLSPNSKSRFRVRIEFSYGQFKENVLPLTDSIELVQAWSQIPAELKMAKGKEWKASFGDLIGTPRSNLFTLVIGTSITLEYIVGSSNTVMKQAYQNGSLID		
GCA_001998885.1_ASM199888v1_genomic	Cas9_0_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_5_CAS-I-II-III-IV-V-VI,Csn2_2_CAS-II-A	1132 aa	MIKDISVNLGFDIGIASVGWAVLNNANGKILESGVSIFPSASAERNAERRSFRQSRRLVRRRNVRLHDMKDLLEKAGFPYPTKQHIYPNPYELRVKGLTEQLTKEELALALYHLVKRRGISYDLKDVEDETGKNSSAYQESLQKNQTLLKRLTPGEIQLNRLKNEGK VRGQIVTEDEGTLLNVFPTADYRDEAMTILSQQQAFYPNITDEFIKQVSEIITRKREYFKGPGSLKSPTDYGIYKKDGTVLDNLFEILIGKDKIFPDEYRAASNSYTAQLFNLLNDLNNLRIDTTEDGKLTTEHKEEIIHYLKTTDKKKINMLALISQVTGAEQTHIKGFR IDRDDKPEWHSLAIYRKICRNFAELGIDVATWPTTLLDDLGRILTLNTENGEIRRALTTEKFKEKYPFLTESLVDTMIQNKQWFHTSSNQKWHSFSLKTMSLLIPEMLVTSKEQMTLLYDLGLIKEKDNIYETETQINLPHILADLYNPVVRKSVKQTMKIFNALYQK YPNIQYIAIEMPRDTNEEDEKKQIKEFQKKNEQEKDKAYLEFLSLSGVSETELNNKIYKNKKLRTIMRLWYQQQGICPYSGKPMPAEDILSTPNNYEIDHIIPISVSFDDSMNNKVLCFANMNQEKGKSTPYEMMQNGHGQSFASLKAMVAKNKRMDNTKKKNLL FTEKLSDIDVRKRFIARNLVDTRYASRVVLNELQQFVKGNQLDTKVTVVRGKFTSVLRKRWSINKTRDTHHHHAIDAAVIAVTPTLRIWKKENAVIIPKQVGEHLIDEPTGEILDKETFEREAYTPPFTHFLEDVRQLAPTIKFTHQVDKKMNRKISDATIYSTRQVQ LGKDKQPADYVLDKIKDIYSVDGYKKFVETYKKDKAKFMMYRIDPKTFELLEQIVATYPDYEEVSLPNGKVKRQEVSPFELYRRQNGWVKKYSKKQNGPAVKQLKYYNKKLGSKIDITPKNVTKGKKVVLQSLSPWRTDIYYNDVTSEYEIMGIKYSDLTFGKC GQYGIKRERYEEIKTDEKVSAESCFMFSLYRNDRIKVIDTVNDETVELLFGSRTLPNQKGYVELKPIDKAKFDSKEVVSFYGSVTPNGQFIKRFLKKNYLLYKINTDVLGNPFYVKKEGENPRDIIDNN	['GTTTTGGAAGCATTCAAAACAGCATAGCTCTAAAAC' 'GTTTTTGTACTCTTAAAGATTTAGTAACGGTAAAAC']	Qualified
GCA_003987515.1_ASM398751v1_genomic	Csn2_2_CAS-II-A,Cas2_5_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_0_II	1132 aa	MIKDISVNLGFDIGIASVGWAVLNNANGKILESGVSIFPSASAERNAERRSFRQSRRLVRRRNVRLHDMKDLLEKAGFPYPTKQHIYPNPYELRVKGLTEQLTKEELALALYHLVKRRGISYDLKDVEDETGKNSSAYQESLQKNQTLLKRLTPGEIQLNRLKNEGK VRGQIVTEDEGTLLNVFPTADYRDEAMTILSQQQAFYPNITDEFIKQVSEIITRKREYFKGPGSLKSPTDYGIYKKDGTVLDNLFEILIGKDKIFPDEYRAASNSYTAQLFNLLNDLNNLRIDTTEDGKLTTEHKEEIIHYLKTTDKKKINMLALISQVTGAEQTHIKGFR IDRDDKPEWHSLAIYRKICRNFAELGIDVATWPTTLLDDLGRILTLNTENGEIRRALTTEKFKEKYPFLTESLVDTMIQNKQWFHTSSNQKWHSFSLKTMSLLIPEMLVTSKEQMTLLYDLGLIKEKDNIYETETQINLPHILADLYNPVVRKSVKQTMKIFNALYQK YPNIQYIAIEMPRDTNEEDEKKQIKEFQKKNEQEKDKAYLEFLSLSGVSETELNNKIYKNKKLRTIMRLWYQQQGICPYSGKPMPAEDILSTPNNYEIDHIIPISVSFDDSMNNKVLCFANMNQEKGKSTPYEMMQNGHGQSFASLKAMVAKNKRMDNTKKKNLL FTEKLSDIDVRKRFIARNLVDTRYASRVVLNELQQFVKGNQLDTKVTVVRGKFTSVLRKRWSINKTRDTHHHHAIDAAVIAVTPTLRIWKKENAVIIPKQVGEHLIDEPTGEILDKETFEREAYTPPFTHFLEDVRQLAPTIKFTHQVDKKMNRKISDATIYSTRQVQ LGKDKQPADYVLDKIKDIYSVDGYKKFVETYKKDKAKFMMYRIDPKTFELLEQIVATYPDYEEVSLPNGKVKRQEVSPFELYRRQNGWVKKYSKKQNGPAVKQLKYYNKKLGSKIDITPKNVTKGKKVVLQSLSPWRTDIYYNDVTSEYEIMGIKYSDLTFGKC GQYGIKRERYEEIKTDEKVSAESCFMFSLYRNDRIKVIDTVNDETVELLFGSRTLPNQKGYVELKPIDKAKFDSKEVVSFYGSVTPNGQFIKRFLKKNYLLYKINTDVLGNPFYVKKEGENPRDIIDNN	GTTTTACCGTTACTAAATCTTTAAGAGTACAAAAAC	Qualified
GCA_000165505.1_ASM16550v1_genomic	Cas2_5_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_1_II,CasR_1_CAS-I-II-III-IV-V-VI,Cas3_1_I	1092 aa	MKYSIGLDIGIASVGWSVINKDKERIEDMGVRIFQKAENPKDGSSLASSRREKRGSRRRNRRKKHRLDRIKNILCESGLVKKNEIEKIYKNAYLKSPWELRAKSLEAKISNKEIAQILLHIAKRRGFKSFRKTDRNADDTGKLLSGIQENKKIMEEKGYLTIGDMVAK DPKFNTHVRNKAGSYLFSFSRKLLEDEVRKIQAKQKELGNTHFTDDVLEKYIEVFNSQRNFDEGPSKPSPYYSEIGQIAKMIGNCTFESSEKRTAKNTWSGERFVFLQKLNNFRIVGLSGKRPLTEEERDIVEKEVYLKKEVRYEKLRKILYLKEEERFGDLNYSKDE KQDKKTEKTKFISLIGNYTIKKLNLSEKLKSEIEEDKSKLDKIIEILTFNKSDKTIESNLKKLELSREDIEILLSEEFSGTLNLSLKAIKKILPYLEKGLSYNEACEKADYDYKNNGIKFKRGELLPVVDKDLIANPVVLRAISQTRKVVNAIIRKYGTPHTIHVEVARDLAK SYDDRQTIIKENKKRELENEKTKKFISEEFGIKNVKGKLLLKYRLYQEQEGRCAYSRKELSLSEVILDESMTDIDHIIPYSRSMDDSYSNKVLVLSGENRKKSNLLPKEYFDRQGRDWDTFVLNVKAMKIHPRKKSNLLKEKFTREDNKDWKSRALNDTRYISRFVA NYLENALEYRDDSPKKRVFMIPGQLTAQLRARWRLNKVRENGDLHHALDAAVVAVTDQKAINNISNISRYKELKNCKDVIPSIEYHADEETGEVYFEEVKDTRFPMPWSGFDLELQKRLESENPREEFYNLLSDKRYLGWFNYEEGFIEKLRPVFVSRMPNRGVK GQAHQETIRSSKKISNQIAVSKKPLNSIKLKDLEKMQGRDTDRKLYEALKNRLEEYDDKPEKAFAEPFYKPTNSGKRGPLVRGIKVEEKQNVGVYVNGGQASNGSMVRIDVFRKNGKFYTVPIYVHQTLLKELPNRAINGKPYKDWDLIDGSFEFLYSFYPNDLIEI EFGKSKSIKNDNKLTKTEIPEVNLSEVLGYYRGMDTSTGAATIDTQDGKIQMRIGIKTVKNIKKYQVDVLGNVYKVKREKRQTF	GTTGTACTTCCCTAATTATTTTAGCTATGTTACAAT	Qualified
GOMC.bin.20170	Cas2_5_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_1_II,CasR_1_CAS-I-II-III-IV-V-VI,Cas3_1_I	1092 aa	MKYSIGLDIGIASVGWSVINKDKERIEDMGVRIFQKAENPKDGSSLASSRREKRGSRRRNRRKKHRLDRIKNILCESGLVKKNEIEKIYKNAYLKSPWELRAKSLEAKISNKEIAQILLHIAKRRGFKSFRKTDRNADDTGKLLSGIQENKKIMEEKGYLTIGDMVAK DPKFNTHVRNKAGSYLFSFSRKLLEDEVRKIQAKQKELGNTHFTDDVLEKYIEVFNSQRNFDEGPSKPSPYYSEIGQIAKMIGNCTFESSEKRTAKNTWSGERFVFLQKLNNFRIVGLSGKRPLTEEERDIVEKEVYLKKEVRYEKLRKILYLKEEERFGDLNYSKDE KQDKKTEKTKFISLIGNYTIKKLNLSEKLKSEIEEDKSKLDKIIEILTFNKSDKTIESNLKKLELSREDIEILLSEEFSGTLNLSLKAIKKILPYLEKGLSYNEACEKADYDYKNNGIKFKRGELLPVVDKDLIANPVVLRAISQTRKVVNAIIRKYGTPHTIHVEVARDLAK SYDDRQTIIKENKKRELENEKTKKFISEEFGIKNVKGKLLLKYRLYQEQEGRCAYSRKELSLSEVILDESMTDIDHIIPYSRSMDDSYSNKVLVLSGENRKKSNLLPKEYFDRQGRDWDTFVLNVKAMKIHPRKKSNLLKEKFTREDNKDWKSRALNDTRYISRFVA NYLENALEYRDDSPKKRVFMIPGQLTAQLRARWRLNKVRENGDLHHALDAAVVAVTDQKAINNISNISRYKELKNCKDVIPSIEYHADEETGEVYFEEVKDTRFPMPWSGFDLELQKRLESENPREEFYNLLSDKRYLGWFNYEEGFIEKLRPVFVSRMPNRGVK GQAHQETIRSSKKISNQIAVSKKPLNSIKLKDLEKMQGRDTDRKLYEALKNRLEEYDDKPEKAFAEPFYKPTNSGKRGPLVRGIKVEEKQNVGVYVNGGQASNGSMVRIDVFRKNGKFYTVPIYVHQTLLKELPNRAINGKPYKDWDLIDGSFEFLYSFYPNDLIEI EFGKSKSIKNDNKLTKTEIPEVNLSEVLGYYRGMDTSTGAATIDTQDGKIQMRIGIKTVKNIKKYQVDVLGNVYKVKREKRQTF	GTTGTACTTCCCTAATTATTTTAGCTATGTTACAAT	Qualified
GOMC.bin.4712	Cas3_1_I,CasR_1_CAS-I-II-III-IV-V-VI,Cas9_1_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_5_CAS-I-II-III-IV-V-VI	1092 aa	MKYSIGLDIGIASVGWSVINKDKERIEDMGVRIFQKAENPKDGSSLASSRREKRGSRRRNRRKKHRLDRIKNILCESGLVKKNEIEKIYKNAYLKSPWELRAKSLEAKISNKEIAQILLHIAKRRGFKSFRKTDRNADDTGKLLSGIQENKKIMEEKGYLTIGDMVAK DPKFNTHVRNKAGSYLFSFSRKLLEDEVRKIQAKQKELGNTHFTDDVLEKYIEVFNSQRNFDEGPSKPSPYYSEIGQIAKMIGNCTFESSEKRTAKNTWSGERFVFLQKLNNFRIVGLSGKRPLTEEERDIVEKEVYLKKEVRYEKLRKILYLKEEERFGDLNYSKDE KQDKKTEKTKFISLIGNYTIKKLNLSEKLKSEIEEDKSKLDKIIEILTFNKSDKTIESNLKKLELSREDIEILLSEEFSGTLNLSLKAIKKILPYLEKGLSYNEACEKADYDYKNNGIKFKRGELLPVVDKDLIANPVVLRAISQTRKVVNAIIRKYGTPHTIHVEVARDLAK SYDDRQTIIKENKKRELENEKTKKFISEEFGIKNVKGKLLLKYRLYQEQEGRCAYSRKELSLSEVILDESMTDIDHIIPYSRSMDDSYSNKVLVLSGENRKKSNLLPKEYFDRQGRDWDTFVLNVKAMKIHPRKKSNLLKEKFTREDNKDWKSRALNDTRYISRFVA NYLENALEYRDDSPKKRVFMIPGQLTAQLRARWRLNKVRENGDLHHALDAAVVAVTDQKAINNISNISRYKELKNCKDVIPSIEYHADEETGEVYFEEVKDTRFPMPWSGFDLELQKRLESENPREEFYNLLSDKRYLGWFNYEEGFIEKLRPVFVSRMPNRGVK GQAHQETIRSSKKISNQIAVSKKPLNSIKLKDLEKMQGRDTDRKLYEALKNRLEEYDDKPEKAFAEPFYKPTNSGKRGPLVRGIKVEEKQNVGVYVNGGQASNGSMVRIDVFRKNGKFYTVPIYVHQTLLKELPNRAINGKPYKDWDLIDGSFEFLYSFYPNDLIEI EFGKSKSIKNDNKLTKTEIPEVNLSEVLGYYRGMDTSTGAATIDTQDGKIQMRIGIKTVKNIKKYQVDVLGNVYKVKREKRQTF	ATTGTAACATAGCTAAAATAATTAGGGAAGTACAAC	Qualified
MARD_SAMN04488102_REFG_MMP04488102	Csn2_10_CAS-II-A,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_0_II	1080 aa	MANLVLGLDIGVSSVGWGIIDKDTNEVIDAGVRLFEEATRNANEDRRNFRGARRLKRRRKHRLERAEDFLKSNGFPVDGIGSIDPYRARYVGLSEELSKTELAAALYHLIKRRGTTIDTPEEDEKTSGSELSTKEQLKRNAKKLEDKNVVEIQYEKLVNNEVVRDH KNRFKTSDYIKEAKALLETQREYHAEIDEEFIEGIVALINNRRQYYDGPGSKKSPTPYGEWYFDENGELQHETMINKMRGKGTYFTDELRIPKKAYTAELFNLLNDLNNLRFPRDETMDGLTTEEKETLIKNHIEKGSKVTLKKIAKLVGISNELDIKGARLNLKTNK PIFSEFTGLKKLKTDLKGVDTPEGFFENIEVLDQIAEILTAEKSIARREEQLGQLFEQAYGSPLSDVVAGLVNDTSIKEYHALSKKAMDLILPELWETNKNQMQLFTEHGLGKSRLEQLQKGSKIQFDDEAILSTVARRAHREAIKVTNAVRERYGELSHVVVEMARE KNSDEAKKSYKDYQKKMGAFEKKMAKLLDVKELKDLKLNGKQMLALKLMDSQDYKCIYTGKQILPHDIVNQEYQFEIDHIIPLSYSFDDSQQNKVLCYRQENQDKGQMTPFQYFQSGKAKRTFDEFKAECTNLFKSKKISKCKLNYLLEQRDVQHDPEIQKNFI NRNLVDTRYAMRSFSSNLRTFYMNNDIDTKVLSIRGSFTSALRRRARLHKDRDHYAHHAIDALIVAAIGTTPILKQFKSTDFNREGIAVNTETGEIVGEKDIFDSPTRDFIHWLRNMESEIKYSHKVDRKPNRTMTNQTLYSTREKDGDQYVVGKVKNIYELDKKGF DSLKKRIDKDPDSFLMAQHDPKTWELVLKVMDEHNHADNPFQDFYKENGYIKKDGKVPVKGLKYLDYSPGIYVDITHKYPDAKRNVLLLKRKSLRIDVYKNEEGKYKYIGVPYNWFKKDGDHYILDKEKYYGEEGLAAPYKQIDDSYVFQYSFYKNDRISFFR EEKTRDIDTGKNKKILVQYKNKIFRGDNDPRRSTIELKDKSFSDKKQQKFTIGTLRDLKKYNVDILGNEYPIDKETFQDKIPSI	['GTTCCGTTATACCAAAGGATGAGAGAGTACCAAAAT']	Qualified
GCA_007988845.1_ASM798884v1_genomic	Csn2_10_CAS-II-A,Cas2_5_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_0_II	1080 aa	MANRVLGLDIGVSSVGWGIIDKDTNEVIDAGVRLFEEATRNANEERRSFRGARRLKRRRKHRLERAEDFLQSNGFPVDSIGSIDPYRARYVGLKEELSKKELAAALYHLIKRRGTTIDTPAEDDKTSGNELSTKEQLKRNAKKLEDKHVVEIQYEKLVNNEVVRDH TNRFKTSDYINEAKALLETQSKYHAEIDASFIEGIIELINNRRQYYDGPGSEKSPTPYGQWYYDEEGILQYETMINKMRGKGTYFTDELRIPKKAYTAELFNVLNDLNNLHFPRDETLDSLTTEEKETLIKTHIDKGSKITLKKIAKLVGISDELDIKGARLNLKTNKPIF SEFTGLKKLKKDLEGIDVPEGFFENREVLDKIAEILTAEKSIARREEQLGQLFEQAYSSPLSDVVAGLVNDTSFKEYHALSKKAMDLILPELWETNKNQMQLFTEHGLGKSRLEQLQKESKIQFDDEAILSTVARRAHREAIKVTNAVREKHGELSHVVVEMAREKN SDEAKKSYNDYQKKIGAFEKKMAKLLDVKELKDLKLNGKQMLALKLMDSQDYKCIYTGKQILPHDIVNQEYQFEIDHIIPLSYSFDDSQQNKVLCYRQENQDKGQLTPFQYFQSGKAKRTFNEFKAECTNLFNSKKISKRKLNYLLEQRDVQHDPEIQKNFINRNL VDTRYAMRSLSSNLRTFYMNNDIDTKVLSIRGSFTSALRRRARLHKDRDESHAHHAIDALIVSAIGTTPILKQFKAVDFNREGIAVNTETGEIISEEEIFASPTLKFIHWLRNMESEIKYSHKVDRKPNRTMTNQTLYSTREKDGDKYVVGKVKNIYELDKKGFDSLK KKIDKDPDSFLMAQHDPKTWELVLKVMDEHSHADNPFQDFKKQHGHIKKDGKVPVKGLKYLDNKIGIHVDISHKFPESKRDVVLLSRKSIRIDVYQNDEGKYKYLGVPYNWFTKEGDHYVLDKEKYYGKEGLAAPYKQIDDSYTFQYSFYKNDRISYDKYEKM ENPETGEKEKVMVHHDKIFRGDTNPRQNQIEIDDIAFKFNGRQFITLGKLINLEKYNIDILGNEYLIEKEELQDKIPAI	['GTTCCGTTATACCAAAGGATGAGAGAGTACCAAAAT']	Qualified
MARD_SAMEA44539918_REFG_MMP44539918	Cas9_1_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_5_CAS-I-II-III-IV-V-VI,Cas3_0_I	1059 aa	MSTMSYILGLDLGVASVGWSVIEIDENEMPIRLVDLGVRTFEKAEVPKTGESLAKSRREARSIRRLIARRVRRLLLAKKALLEEGILTQNDFLTKSIIKDLPINAWELRAKGLDHKLTSKEWATVLLHLVKHRGYLSQRKNESDTADKQLGALLSGVKENHTLLQEN KYRTPAELAICKFEKEEGHIRNQNGEYIHTFDRLDLQAELHLLFEKQRQFNNPYASAKLEEKMDYLLMYQKPALSGEAILKMLGKCSFETQEYKCAKNTYSAERFIWLTKLNNLRILTQGSERGLTESERALLLNQPYQKAKLTYTQVRNILELSESEIFKGLRYGN VEDKKMIEKETLMELKAFHQIRKVLEKANLKTEWQGLATKPELLDEIGTIFSLYKTDADITAQLQDKISAPILNALLSGINFDKFINLSLVALSKILPLMEQGQRYDEACTSIYGDHYGVKTKEEHLFLPSIDKDEIRNPVVLRCLSQTRKIINAVIRLYGSPARIHIETGR EVGKSFKERQKIEKNQLENKAKKERAIKEFKDLFPYFIGEPKSKDILKMRLYNEQQGKCLYSGKTIDINRLLEKGYVEIDHALPFSRTWDDSFNNKVLVLANENQNKANQTPYEWFDGKNNTTKWQQFVALVWGSHFNYHKKQHLLLKNLPEGFMKRNLNDTR YVARYLMNFINDHIQLTGKGKRRVFASNGQITALLRARWGLMKVRGDNDRHHALDAIVVACTTPSMQNKITDFVRSKEMNIFSGEFIDRNTGEIKQIHFPEPWNYFRNEVMIRVFSHDPLLELEQQLPSRPEARHQYVTPLFVSRMPTRKMTGQGHLETIKSAKRL AENESVQRIALTSLKLNTLENMANKDKEKALYQALKERLEQFNNDPVKAFAEPFYKKGGQQVKKIRVKQIQKTGVIVHQGNGVADNASMVRVDVFEKANKFYLVPIYTWQVAKGILPNKAAIANKNEDEWQEMDESFTFKFSIHPNDLIKVTTKKEVYFGYYV AMDRHTAAINIREPDLERSKGKDGLHRGIGVKTALSFEKYDIDPLGKNIRPCQPKARQPIR	['GTAGAAATACACGCCCTGTTTCAAAGGGATTAAGAC' 'GTAGAAATACACGCCCTGTTTTAAAGGGATTGAGAC' 'ATTGTAGCATAGCGAAATGAGAAAGGGAGCTACAAC']	Qualified
MARD_SAMN10362902_REFG_MMP10362902	Csn2_8_CAS-II-A,Cas2_9_CAS-I-II-III-IV-V-VI,Cas1_4_CAS-I-II-III-IV-V-VI,Cas9_0_II,Csx19_2_CAS-III-D	1059 aa	MKHSGQYILGLDIGIASVGYGLINNDTKEIIDAGVRLFPEGKSDENDGRRSKRGSRRLLRRRKHRIANVKSLLAEYGISIENNSNMNPYEIRVKGLSEKLSDEELAAALIHLSKRRGIHNVDVTDDSEESSELSTKDQINRNAKLLTDKHVAELQLERLNNEGSVRGHS NRFKTNDYIKEAEALLDKQIEYNQKIDGEFKRKYIHLINRRRAYYEGPGEGSPFGWNADIKKWYETLMGKCTYYPEELRSVRYAYSASLFNVLNDLNNLVIMRDENDKLTKDEKEHLINNLFKKQKSAPTLNKIAKELGVKEEDIKGYRVDKKDKPEFSTIKIYHDI KGITNNPEILENPDVLDKIAEILTIWQNEKHILEELEKLQLNLNEQERVKVANLTGYTGTHSLSLKLIKQLNSELWESTLNHMGIISNMNLRPKKMEFKGLDKIPVDFVDEWILSPAVKRSLIQSIKVVNEIIKVYGEPKDIIIELARESNSDDKKKFLQKLQKDSQKINE QVRAKISEINGKDNFSSSLFEKVKLWHLQDGLCMYSTESIKIDDLIQNPNAYEVDHIIPRSVSFDDSLNNKVLVKLEENQKKSNMTPYQYLSSGQGSISYEKFKANVLTLAKDGKRISKKKRGYLLEERDINKFEVQKEFINRNLVDTRYTTRELMNLLKLYFKDNE MKTKVKAINGGFTHALRKIWGFNKDRSEDYKHHAEDALIIAMADYIFANKDIFKNQNILMRDNQAIDLETGEILSEKDFVDRFFEDEVFSNIRAIKNFDRYKYSKRVDKKPNRQLMNETIYSSRTFDDGEYIIGKLKNLYDPKNKDVKKKINKNPQAFLMYHHDPQ TFEKLQTILDQYSDADNPLAKYYEETGEFLKKYSKEGNGPVVKSLKYKYRKLGAHHDITDKYQRSRNKVFTLSLKPFRIDIYKDGGKYKFVNINYIDVKDVGSEYNVDENSYNEKKNRKNITEQAEFIGSFFKYDLIKIDDQLYSLISIHSESSNKLEVDLVNKSYKE YAEKNKLNKQKYISIGSKNNITKYTTDILGNLYPVYNEKLKMKFPKYKGNE	ATCTCACTTATACTAAATTTCGCAGAGTACTAAAAC	Qualified

MARD_SAMN05421734_REFG_MMP05421734	Cas9_0_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_9_CAS-I-II-III-IV-V-VI,Csn2_8_CAS-II-A	1055 aa	MSNKNDFFILGLDIGIGSVGWGLIDKDRNVIDAGVRLFPEADPSNNEGRRSFRSTRRLLRRRQHRLDRLKELLQQSNIIQSREIPTSTKPYTIRSKGLYHPLTKDEIAIALIHLAKRRGIHNIEVEEDDSSGNELSTKDQILKNEKLLKDRYVCELQLDRFKEKGEIRGHSNRFKTEDYVKEAKKLLDTQQKHHNEITTDFIDKYIELLKTRREYYEGPGKGSPFGWDEDIEKWYQQMMGRCSYFPEELRAVKESYSAQLFNLLNDLNNLSLNREENTKLTYDEKIHLIDQVFKKYKNPTLKRVAKSLDIEEHDIKGYRITKNGKPTFTTLKVYHDIKSINDSSQLLENVDQLDEIAEIVTMYQSPSDKFDSLEKLELKIPKSDLLQIAQLNYTQTHSLSLKMIHLLLPELLETSKNQMEIITEQGLKPKKVNFEGRTYIPTKQIDDFILSPVVKRAFTQTIRVINAIIKKYGYPSEIVIELAREKNSDDRKKFLNDLQNKNEKMNKEIREKLEALDVDASKGLFNMLKLWHLQDGLCMYSNNSIKINTLVNKPYLFEIDHIIPRSVSFDDSMNNKVLVYKDENQKKSNQTPFQYLSSGRGVISYDHYKSSVLQLAKSPNKMSKKKKEYLLEERDINKYDVQQEFINRNLVDTRYATREIMTTLKTFFTANEKSVKVKSINGSFTNYLRKLWDLNKDRGADFKHHAEDALICAMANHIFETKEVFKSQNLIFNQGHMVDVETGEILQEGEFNAFFTEKINKVKQIKNYSDYKYSHRVDMKPNRQLMDDTLYSTRKRNDEEYVINKLKNIYDKNSKDAENLVKKYKKDPSQFLMYHHDPQTFKKLESIIEKYSEAKNPLHQYWQEEGGYLKKYSKKGNGPIVKSLKYYESKLKEHRDQSYKYPVYDKKVVKKSLKPYRMDVYSNGSQYKFVTVRYNDLKEQGNEYVINIEDYQNKLKSKKIDESFKFLFSLYKNDVLRLEGEYYRLIGVNNDSTNRVELNTIKYNYKDYCKKEKIKNDRIAPTIKKSVTDFQKFNVDVLGQYFINENEKLKMKYSK	GTTTTAGTACTCTGTGATTTTAGGTATTAATGAGAC	Qualified
GOMC.bin.16150	Cas9_0_II,Cas1_0_II,Cas2_5_CAS-I-II-III-IV-V-VI,Csn2_8_CAS-II-A	1054 aa	MKEKYILGLDLGITSVGYGIINFETKKIIDAGVRLFPEANVDNNEGRRSKRGSRRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIALLHLAKRRGIHNINVSSEDEDASNELSTKEQINRNNKLLKDKYVCEVQLQRLKEGQIRGEKNRFKTTDILKEIDQLLKVQKDYHNLDIDFINQYKEIVETRREYFEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALNDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYRITKSGTPQFTEFKLYHDLKSIVFDKSILENEAILDQIAEILTIYQDEQSIKEELNKLPEILNEQDKAEIAKLIGYNGTHRLSLKCIHLINEELWQTSRNQMEIFNYLNIKPNKVDLSEQNKIPKDMVNDFILSPVVKRTFIQSINVINKVIEKYGIPEDIIIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQQEGKCLYSLESIALMDLLNNPQNYEVDHIIPRSVAFDNSIHNKVLVKQIENSKKGNRTPYQYLNSSDAKLSYNQFKQHILNLSKSKDRISKKKKDYLLEERDINKFEVQKEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHLRKVWRFDKYRNHGYKHHAEDALIIANADFLFKENKKLQNANKILEKPTIENNTKKVTVEKEEDYNNVFETPKLVEDIKQYRDYKFSHRVDKKPNRQLINDTLYSTRMKDEHDYIVQTITDIYGKDNTNLKKQFNKNPEKFLMYQNDPKTFEKLSIIMKQYSDEKNPLAKYYEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTNKYENSTKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELKEKKKIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNIKGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	[GTTTTAGTACTCTGTAATTTTAGGTATGAGTGAAAC]	Qualified
MARD_SAMN04570274_REFG_MMP04570274	Cas9_0_II,Cas1_0_II,Cas2_5_CAS-I-II-III-IV-V-VI,Csm2gr11_7_CAS-III-A	1537 aa	MKKILGLDLGTNSIGWALTTNDFDTKKGTIEGLGSRIIPMSQDILGKFDSGVSLSQTAQRTEYRGKRRLVQRFLLRRERLHRVLHILELLPKQYSEAIDFDKKLGQFKPEREQKISFYKNESGKNQFIFTKSHQEMLTEFQTYYPTIKNIPYDWTIYYLRKKALTEKISKEELAWILLNFNQKRGYYQLRGEEKEENKNKKEEFHTLLVTDIIERDKGKSGTWYNVILENGWIYKRESKISLADWIGKTKSFIVTTEYNDDGSVKLTKDNEEKRSFRSPKEDDWGLIKIRTEKTINQSDKSVGAYIYDALLKNPNQKIKGELVRTIERKYYKDELREILKKQLEFHSDVLNDKILYKKCIDELYKNNDSHKNSIKDRGFDYLFLDDIIFYQRPLKSKTSLISDCSLEFRPIKNKNGVLVKDTDGKQIAKALKCIARSNPIFQEFRLLQFVKNLKIYKRGVVDDVEVTNDYLKSENDVEKLFNWLNDKAEINQKQFLVNFKLKEDKHRWNFVEDKGYPCNETRNQFLNAIAKIDVDKSFFNTKNTQDLWHILYSVTDKIEITKAITTFAKRHNLPNEFIINFSKIKPYKKEYGSFSEKAIKKLLPLMRFGSYWDAKNIDIKTLDRIDKIQTGEFDEKIRNRVREKAIDLHKIDDFKDLPLWLASYIVYDKHSEASDVSKWKTPADIEKFLKHEFKQHSLRNPIVEQVITETLRVVKDIWLQYGEGKENFFNEIHIELGREMKNDKATREKISKKVSENENTNLRIKAILTELKNDGVSDIRPYSPSQQEILKIYEEGVYNNESRKEKLEEVDKIRKNAKPTTSEIIKYKLWLEQGYASPYTGKLIPMSELFTTKYQIEHIIPQSRFFDDSLSNKVICESEVNELKSNMTALEFILTNENRIVTLTGGETVKIFTKEAYNQHIISYYSKNKNKQRLLTEEIPEQFISRQLNDTKYISKIVKNLLSKIVREEDEQEVTSKHVVSLNGSITSRMKQDWGLNDVWNTIVSPRFERLNTLTNSNDFGQFELKKDAYGNEGKRVFQTTVPDAIAKGFTKKRIDHRHHALDALVIACVNRTHINYLNNLNAKDTDDKAVKHELRNKLCFKTKPDSNGNYKWEFHKPWEKFTKEAQDKLNTTIVSFKQNTRVINKTVNKYVSYKDENGNLNVDANGLPQKKVITQTKGDNWAIRKPLHAETVSGKVFLKRIKQSHITIANAIEQVELIVDKEVKKQLKAKIKLYPNNIAGLKKHLKAFPVMINDKAIDKIQIYETLEATATRKILDITFDEKKIQKITDTGIQKILLNHLKQEKYQNTVDENGKKLAPHEIAFSENGLEDLNKNITFLNDGKNHQPIKKVRIFEEGNKFPLGEVGNKINKFVETAKGTNLFFAVYENDDKKRNYKTIPFNEVIERQKQGLNAALEIDENGNRLLFTLSPNDLVYIPLEDEIGYLNGVYFKNLNKEQTVRIYKFVSCTGGEGHFVPNNYSKEIISNENGSNNKNERMLNFNNFNIIFDGKEKPVMIKSICWKLEVDRLGKVIKIIK	GTTGGGAATTATAAGTAAAAATACAATTTTGAAAGCAATTCACAAC	May inactive
GCA_007988985.1_ASM798898v1_genomic	Cas9_0_II,Cas1_4_CAS-I-II-III-IV-V-VI,Cas2_9_CAS-I-II-III-IV-V-VI,Csn2_8_CAS-II-A	1037 aa	MANYTFGLDIGIGSVGFGVIDEDMNIIESGVRLFPEADVSNNDGRRAKRGARRLQRRRKHRKERLVRLLQNYGFDTEQVSEENPYELRAKGLYEPLNQNQLIAALYHLIKRRGVHNVTADLDDEKAADDSLSTKDQIKNNEKQLQEKYVAELQLERLQEDGEVRGHKNRFKTEDYVEEAKKLLETQQTTHGKIDDAFIKQYIELLETRRAYYEGPGFGSPFGWEQDTEKWFRQMMGKCSYFPEETRAVKESYSAQLFNVLNDLNNLSITRDENDRLTEEEKAELLEKVFKAKGNPTLKKIAKIVKVKEEDIKGYRIDKKGKPEFTVMKIAQELNKIDANLMQEPAETLDQLAEIVTVWQDFDNKLERIKTLDLNYSEHTLAKLAELNFGGTHALSLKAIQMLLPDLWATNKNQMELFTEYGLKPKKIELKGKNKIPHDYINDLILSPVVKRAFVQSIRIVNALIEKHGEPREIVIELAREKNSSDRKTFLNKMQKQHEAINKQVREKLDAMDLVERGGMFDKLRLWHFQDGQCLYSLKPIELEDLINNPSRYEIDHIIPRSVSFDDSQANKVLVHTEENQKKGNLTPYQYLKSGNGEITYEKYKSHVLQLAKSSEKMSKKKREYLLEERDLTKYDVQKEFINRNLVDTRYATREIMTLLKAYFKENEKDVIVKPINGSFTNYLRKLWGLNKDRDEDFKHHAQDALVVSMAGYLLEHKDDFKQQNLMIDEERIIDKETGEIMNEKDFAASFTERMNKVKEIKEYPHYKYSHRTDQKPNRQLMNDTIYSTRETDEGEYIIGKMTDLYDPDNDKIKKQFHKSPEAFLMFEHDEKTFEKLTTIMKQYDEAKNPLAAYYKETGNMLTKYSKKDNGPPVKSLKYRSKKLGEHKDLSHKFDGNDKRVVTLSLKPYRMDVFYEDNRYKFVTVRFDDLHELKNKYFIPEKIYQEKMAEKSIEDWGSFQFSVFKNDILRLSGQVYKFIGVNEDKRNVIECDLINQKSDKRLRMTVKKEIADFRKFNTNVLGDHYVSNNEKLKLEIQK	GTTTTAGTACCCTGCAGTTTTAGGTAGATATGAGAT	Qualified
GOMC.bin.6855	Cas2_7_CAS-I-II-III-IV-V-VI,Cas1_0_CAS-I-II-III-IV-V-VI,Cas9_0_IIB	1495 aa	MDWLYPIALDLGAKNTGVFNLSYAANTKLADFDKKNKVLKKAFVAQVPNIMNGGYQFLQINRTATRHAIRCRKRKKQAKKLFLFLLEHVHNFKTQGHQEAINHFLNRRGYNYHESEIDFTAVKQVEQEHWIALQAFLIEANASPVFDQLNFSEIEESLLQLRVFELEELQQHLESIDNKKKNSPQALLLKVVKDLKNEKAQVTGAKHRRDYFKAIHHDIEHMSKSEDKKFKRLFNALTHAKGYKQGEFIRVFWSLLCHVNNFDLKLLNRIHKSLDDKLTYDQIDAILNQHISHWVLKLWQGTDANTPERKKFKQTLKAYTQANPKKILQFLLQTNPEHTIAPYESHTNRKPPYCQTLLLNTETLTSQYPEWQKWLVQLSSSCEKMQQALSNYQTRLRSKISKEGNVLVESEKQNGVLELQARSLQFVFDTSAKQCPYKLSEIWGKLKQINQLKRKGKSTDKAEANLASLLEKSWLPSNLKVECTSQPEQSEFWHLINLYYQTRRRTKQGRYFIHRDNKQNSELNKWQKNRKLLQVCQHKPKQLKHQALGDVLQLLGLEQEQLFTISTNDEAISLEKLAEQIKSAAKGIITLAKDAHKAQKDYGIDLQAAIVSEPKLQKLASKAQEVTQKLAVYFFQDTSQQNYFFNRNKSLFTYAQLHSLLFNIERSGFGKTCPLCSADNAARMQISNGFAQASRLSTMSLRLIDGGLKRLLNHQAHHIANYLWPELEGILKQGHRVSLPLVLEQNRFDFSANLHALKGQGKNKQQRLPEDEKYFTQKQERIFNAGKQLCPYTGEVINFSKSDIDHIIPRSGVFGVLNDEANLIASSVTGNRKVKQNRLLSLKDLSKEYLQAHFNTTDPELIKAHLYATLWDSNKAKFVFGQYRQFIALTPVQQVAFRHALFLDEFDELKQGVISAIQHGSKSRVNGTQRFMAQLLADAFWQKARIIQCENQLNFDYFEYSANSQDELSTTSLRRVLTQALVKTDWDLTPFCKNKGQSQRPYSHVIDAMCAMALAVEQHQGEGALSIDVGDSSIWPKVDAQTGEYKASNFTSLAIKAEQLQPELIVKPTSLANKIVKNGPISKPHTVISRRIFGENALGIRFYDLAVDNSNVVKGYYKSEIEAGEFVRATAKVADKNVEILQWLTTHAYYQLKLKQGIKFYTPNSYAILSLMFSVLMTAKKSNLDKNSQEYKAVKWLFGKGDGQLFYYTSKAPLQAAPNIVERLKSNNPLMQKWLHFYNGWQAICGGSAKVNLGQFVIESGQLNQWKNYCQAFLGQTEKFNKHSPVRHYTMFNETTARGNVALQFKQGKFTHYPVNNNEVPKSQSMHILLNSPTLVAFESSVLTKGYLVELEEKVSFKDAQINVKTLINEKKAKVLSLNIDLLNIKVLSNTQINIEGVSKQWFEKNLLLKRAIEAEAWKKRSSLNFDKYDTEKCVIVTDTLKSILNTAIRAERNKTVKVQHIDETVSLGLPVKAEHIAKLVAKQNLEEAKSQKAP	['GTTTCAGTAGCTGATTTGTTGCGATAGGCTATGAAGG' 'GTTTCAGTAGCTGATTTGTTGCGATAGGCTATGAAG']	May inactive
GCA_005886345.1_ASM588634v1_genomic	Cas9_0_IIB	1490 aa	MKWLYPIALDLGAKNTGVFNLSYAAGTKLESFINANQVTKNAFVAQVPETKQGGYQFLQTNRTATRHSIRCRKRRKQAKKLFLFLLERVYRFQTQGHTEAINHFLNRRGYNYHESQIDFETLEQFEQEDWIALQTFLIEQNASSVLEQLDFSEVAESLLQLQAFELEELQQHLECIEKPKKNSPAAALLKVTKELKQEKIKEQTSGAKHRRDYFDAVRHDIEEMAKSKDKQFKKLYNALKHAKGFKQGEFVNVFWQLLCHVNNFDLKLLNRIHKALNDSTTSEQVDATLAKYVSHWVLKLWQGSDANTAERKQFKQTLKSHIKTNPNQILQFLLQTSPEHTIAPYESHSNRKPPYCQTLILNAQTLNNQYPAWQNWLQLLANGCEQTQQAVEQYQTKLRCKTSKEGNILVEIEKQNGLLELQARTLQFVFDASVKTCPYKLSEIWGKLKQINQLKRNSKETKKVQASLTNLLAQSQLPDTLKLACNSNPEEGSFWHLINKYYQARRRTKQGRYFIHRDNSQNSELNKWQSRDQLFQICQHKPKQLKHQALTDVLQLLGLEQNKFQVLNGSNEAVLLEQLATEIKGSAKGIVGVAKNAHKAQKDYGIDLKAAIATESKLQSLADKAKVLTQKLAEHFFTNKIQQDYFFERNKSLFTFAQLHSLLFNVERSGFAKTCPVCSADNAARMQMNQGVAQASRLSTMSLRLIDGGLKRLLNHQAHHIVNRLWPELEHHLKQGHKVDIPLILEQNRFDFSAHLHALKGQGKNKQQRPPEDERYFEQKQTRIFDAGKQLCPYTGDALSLNQADIDHIIPRSGAYGVLNDEANLIAASTVGNRKVKQDRILHFEDLSKAYLQAQFDTANVEQIKAHLYTTLWDANKKQFVFEQYKQFIALTEIQQTAFRHTLFLENTDELKQAVISAIQHSSKSRVNGTQRFMAQLLADIFWQKAKVIQCENLLHFDYFEYSANSQDELSTTSLRRTLTQTFANTEWDLTPFNKQKGHSQVPYSHVIDAMCAMALAVEQHQGEGALAVDVGDTSVWPKIDKQTGEYQASNFNMLAIKAEQLAPEVSVKPTSLATKIKQQGLTSKPHTVISRRIFGENALGIRFYDLAIDNNQVVKGYYHEQNGHQEFVRATTKVVDKTTAMLEWLTTNQYYQLKQKQAVEFYVPNLQAILSLMFRVLSAAKQTKFDKKSDEYKAVKWLFGKGDGQLFYYTNKAALQAAPSIIDKLKSSNPLMKKWQHFYAGWQAICGGSAKVNQGQFVIEAEQLNQWQNYCQAFLGQSVERNKHSPVRNYTMFNETVARGNVALQFKRDKFTLYPFNNNDVPKAQSMHIMLNSPNLVAFESSVLTKGYVVELEEKTNYKGTHIKVKTLIDENKADALKLNLEQLSIKVPSNTQINISGVCKTWFEQNLLLKRAVEADAWKKRTSLGFDKYDDEKCAVITDTLKSVLKTAIRAERNKSLKLLHTEAALTIGLPVKAEHIAKLLAKQNSEG	GTTTCAGTAGCTGATTTGTTGCGATAGGCTATGAAGC	May inactive
GCA_002836475.1_ASM283647v1_genomic	Cas3_1_I,Cas9_0_II	1455 aa	MAKILGLDLGTNSIGWAIVDKDGSDFSLVDQGVRIFPEGVKTEKGIESSKAAERTAYRNARKIKFRRKLRKYETLRVLSENNMCPLSVEEIIEWRKSGFKKYPLNPEFISWLRTDDDKNINPYYFRDKSSKQKATLLELGRAFYHIAQRRGFLSNRLDQSLEGIIEQHRSGLEDIIDNASSFTELQIELTDYFQDRDLIERKASNLDEGEKKLKTTYNSFLKIITDNIKDFDKTKSLLIERLNKKEDLGVVKQKIGEISLAMQSGNFTTLGQYFFSLYNKGKIRNQYTAREEHYLSEFEIICLAQGIEGIIASEKLPEKKYFGLAKELYRAIFFQRPLKSQKASIGKCTFEKNKSRCAVSHPDFEEYRMLQYINTIKIKTPNDERLRLFSNAERLKIMPKFLRKKDNFNFEELVKELVPSESKYAFYKSSSAPTANYLFNFNLNSSVSGCVVSASFKNVLGEDWKTKEYSYQTINKKAEIVTRNVDYNDLWHLISVSTSALYLQEFAKDKLKLDDKRAISFSKIKLKKDFASLSLKAINKINPFLREGLLYSHAVFMANIENIVDVDIWNDPNDRKYIQIKIAELIATNSFDNNLLDIINDLIKECKTEGCYYSKEAEATYKSDLDKKLSSFLRSNKIEDQKEAITNELLPLFIEQLKKYEFITIKRLDEKVFGFLNGQNDDGQVFCNNKNKLKKLYHPSDIDVFRKEITKDGNGNEKLVLGSPLISSIKNPMAMRALHQLRKVLNALIIEGQVDEKTRIHIEMARELNVANKRKGIQDYQRGIEKNRKEYIEEIKKDYKAASGLDIEPTEDDILRYQLWIEQDRKTIYEEECTNIRISQIIGSDPDYDIEHTIPRSRSLDNSKMNKTLCSKLYNRKIKKNKIPVELSNYSDILDRIKPWQKKYLALEYEIISVSRAIKAATTKEQKDRKIRLRHKLNLQKDYFKGKHERFTYTEPKVGFKNSQIPDAGIITKYANDYLKSYFGRVEAVKGGMVAEFRKAWGIQESYIKDGKKYYEVKDCSKHTHHCIDAITIACMTKDKYDVMAHAWDLEDDVEKNQDKLMKSVLEQSKPWKTFTEDLKKIEDEILVSHHTPDNIKKQSKKIIRIRGKKQFVAEMEIDVNGKKIPKTDSNNKIIFKKGFKGEKIARFQQGDTIRGSLHQDSIYGAIKNPLNNEEIRYVIRKDLESIKTSDVEYIVDEIVKQKVKEAILNKVLLLSSNAQQQNKFADNQKVWMNEEKGIAINKVRLYANSVKNPLHVKEHALLSKSKQEHKQKVYAQNDENYCMAIYEKDGRKDFELINNFNLAKLEKQGKGYYSLFKEKTNNGKTIQMPIAKRNNRDLVLKRGQQVIFYDLGTENPDDVNDLIDLKGRIYIIEGLSIQRIVRPSGKVDEYGIIMFRYFKEARKADDLKSDKYKPDGVFKLEEIKPTRKMNHNQFTAFVEGIDFKVMPSGKFEKL	AATATCGTTTTTTCTAATCTTTAATCAAACCACAAC	May inactive
GOMC.bin.8971	Cas4_0_I-II,Cas2_2_CAS-I-II-III-IV-V-VI,Cas1_0_CAS-I-II-III-IV-V-VI,Cas9_0_IIB	1417 aa	MQFNVVSPIAIDFGAANTGVNISHLSDNNIIKEGFTIIDDGKSIKWLQKERTQNRHQRRSIKRNKLAKRLLWQILTEEFKINNVEDDLCDFINGLLNRRGYTYFTDDQVDYEVFDDIYLSGLSESVLPILDEIEGTSFQEKIINITSDIDKCKKINQQFSESHVKEEGKRTKKFQRFFDELKCDSKDEKDTYSKIFTEMEKLVSSTIKSKEDGHYYRASYFENIYEDIKNNEKLLRFLSKSNVSADEFYKLLANISNFQLKLLRKYFNENSKWDPKKLNNILDKWVKSWRCPPNSKEADAKKLLLDSLKTKRNIIDFLTETDPNKTIPPYEDQDNRKPPRDKTLYIDPVKIDKIFPSWRAAVKNILKHSENKNLTEKLDEVISAATFLDINGNKRQNLDHDALVLQRFFDRAKEFDPYKIRILATYNEDKTLNFLQHYDLLKNHLNGNKEYADNLLNIAQRYYEEVEKAKKGLWSSRAGNILFVYDQKNSKKKNLINYNLSNILGVDFNDEKLKKFHEFISINEVSSGLKGKKARKIIKWAESAAILQKRLGNALNYHILDIERKGAYSKNLSKEEKNIYAVILGCLNSAEKIAKEFKIIKSDHFSSIYSFAQIFNSIEGDIHGFSSISKASMIENSWRSKTNKIEINGITKYHANAVRLPKDSVRPFDGMLDRIIKRQAFEIACSKIDQIEKNVIDKREKIFIPIVIEQNNFAFTYDLATIKQDNKKKKNSEGEYKRLSSEYYNKYERLKNSSGNICPYSGDYIRNGDIDHIIPRSITRKFSSTVYNHEANLIYCTTSGNVKKGNSRYTLKNLNKKYLQNQFGTDNIKQIADQVTEKLENIPGRKREFFSGLDNEEQKVIRHALFIPELDYITFPILSTQTKARVNGTQAWLCKNIKQIIEDELSKKGYNIEFDIVKVKSEDVHRNRFLLGEHDSRFKKPVNISQSPVSHVLDASIALADAFNAELSSESAEEDNVDYIKKFIPDNFEIKKLGKKKIWDKNKPQSEKLFKDTLYGINFLPILYTNKEEVYIGFSLDNSVEIKDSSSFFDLMLPFLKYKKQTFNNRNELRKLLEGRKFLPLNIDKDVAMEHLHYVSKNSVSAKDNEIANVLDALFFTTLKKDIKNVLSDSSSKLKFDKIKEKDFQIKVEKIGKFKKKGNSITHPSYQEWLQVVKTTKELLDKGLENNDTEFWSEVVKNSGLFKRYKPDRNHKKVRKQFSLPLLEAPSGGYRIKQQAGNGNNVWQIVAIQNYSSIGFYKKENGDIDFNNPAGVQHFRKSPNVSKIGGRYSLENEYVPFGSVRNVDLEIGLPCNYSVYLKLAETGRVDIGLTLNKEDFVEYVLPGLETEETDYLKIPSFLKTRDNFVKSDNNTNGIFPGYMKPRSQINILEIGEEVTIEYRTEGFTNVLKEAFNKGKEV	GTTTCAGAAGCTGGCACTTAAGAATGATTTATGTTAC	May inactive
MALA_SAMN05422149_METAG_FHIEBIKM	WYL_0_CAS-I-II-III-IV-V-VI,Cas9_0_CAS-II-B,Cas4_0_I-II	1415 aa	MTDSKIISPIAIDLGGLYTGLYMAQYPYGESLAQHHGKLTTLVMPEEGGKMTWSQAARTATRHRIRSNKRRKLAKRLLKVIFTQGLSYKLNDAELEALNGLLNRRGYNRVETDIDRQPLETVPPSWIANEFPEFFDSSSPISEQLEELLEDIPKLRDLAADEKFNLNNREAKKQLFTDAEDEQADILKAYGILRESVTELLNSIDFGHQHRRKYLEAVGKEILKDSRLAELRAKLGASALKNLIGNISNLQLRALRWYFNDTAMKSADYYDGERFKSVYVRWLQNWHVDEETRVNRNNALKVIHPSEDILTALMTIDAELTIPPYENQNNRRPPKDQTLWLNPQSLNEQYGSQWRVWVNNLRKRNPELAEGIAPVLEFTDRRSRLKTNKQRLTDTSLYADSYFLQRLLDRSRIHDTYALRAISKNFKSEKLEKFRWNLKQDLGDQHTEKFLDFTQKYYQEVERAKSGIWEQESASLLEKADINPPSKRKILNRLVGNILGVEMSAEELTTFLSDVWTKKVHGPSTVKSICKAAEDLRKQNGNLFNMYIQSVRYRASVEAGAEKKFKGLDKEIWQLLVKVEKAADVISGSISATETQKQRFSNTFSLAQLYTLIETDRNGFSKVSAAAHMELAWRMTMIPVDDKEAARCSRLPADSSRPFDGVLRRLLERQAYEIATLKSGQIKESGVRDSEISIPVVVEENRFTFSIGLADVKKDGSKRKKLLSLGEKQLEKWQEKNERIRQGGSGLCPYKGTILDKDAELDHIIPRSYSRDASGTIFNSEANLIYASRRGNQEKKDQRYGISDLAPNYLKAVYGTSDSAEITRIIEASVASIKKDFLFSELEGDQQRDVRHALFLERDSKAFEKVFRRLATQQSARVNGTQAWLARKLISVLKSQLKSWLSETGNTIVFNLCKADAQQTSFVRTGLAVTKPETTKGERQGVGSHAIDAYCAFSVAAAEGMHETLAQGAGLMEDTEWLATALPVDVDVNWVKRLPLYRKSELSTRPIFKKEGIYAESFLTTWISEKGIESGFSLGSNSVQWQGKDTRGLLELLAPFATKNLNVDTALASPRPVQLDISKEKAFKHFEMVAKAPHSEVDLHQADILDALRYCTAKKDISSRLIDVQKKTLKKPEEVLKAKDIYVKVDVDVKGLGKLKGKVTHPAIEQWKRLFLIPDIETRIGEKAVGINWQQIGEKYFGEGSSRPHSKTRRVYSLPIVDSPDGGYRIKRIAPDGSTIWQLTSIKDTGSSGFTVNDKGVNWKQPVAIPALRDSSVVGVGERYQPVPEQQVYFDDWKCVDITDERVLSVSVAPGTKDRRYVSVSQTLSDFKAWLTESGIKLHSPFEIPSRVKVDAAKFQSAHGIDVLGKPKSNLFISKVGDVVNYWYIRESSPVVINQYYDKALLAEREAVNETEPCDI	GTTGCAAAGCCTATCCCAATAAACCTGCAACTGAAAC	May inactive

Supplementary Table 3. The Om1Cas9 description, activity prediction and verification.

Supplementary Table 3a. Om1Cas9 ID and human codon-optimized Cas9 gene.	
Assession	Genome: CNA0069409
Host strain	*Staphylococcus warneri*
Nuclease name	Om1Cas9
Human codon optimized Cas9 Gene	ATGAAGGAGAAGTACATCCTGGGCCTGGACCTGGGCATTACCAGCGTGGGATATGGCATTATCAACTTCGAG ACAAAGAAGATCATCGACGCCGGCGTGAGACTGTTCCCCGAAGCTAATGTGGATAACAACGAGGGCAGAAG AAGCAAGAGAGGCAGCAGGAGACTGAAGAGAAGAAGAATCCACAGACTGGAGAGAGTGAAGCTGCTGCT GACCGAGTACGACCTGATCAACAAGGAGCAGATCCCCACCAGCAACAACCCCTACCAGATCAGAGTGAAGG GCCTGAGCGAGATCCTGAGCAAGGACGAGCTGGCCATTGCCCTGCTTCATCTGGCTAAAAGAAGAGGCATC CACAACATCAACGTGAGCAGCGAGGACGAGGACGCCAGCAATGAACTGAGCACCAAGGAGCAGATCAACA GAAACAACAAGCTGCTGAAGGACAAGTACGTGTGCGAGGTGCAGCTGCAGAGACTGAAGGAGGGCCAAAT TAGAGGCGAGAAGAACAGATTCAAGACCACCGACATCCTGAAGGAGATCGACCAGCTGCTGAAGGTGCAG AAGGACTACCACAACCTGGACATCGACTTCATCAACCAGTACAAGGAGATCGTGGAGACAAGAAGAGAGTA CTTCGAGGGCCCCGGCCAAGGCAGCCCTTTTGGATGGAATGGAGATCTGAAGAAGTGGTACGAGATGCTGA TGGGCCACTGCACCTACTTCCCCCAGGAACTGAGAAGCGTGAAGTACGCCTATAGCGCCGACCTGTTCAAC GCCCTGAACGACCTGAATAACCTGATCATCCAGAGAGACAACAGCGAGAAGCTGGAGTACCACGAGAAGTA CCACATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAGGAGATCGGC GTGAACCCCGAAGATATCAAGGGCTACAGAATCACCAAGAGCGGCACCCCCCAATTCACCGAATTTAAGCTG TACCACGACCTGAAGAGCATCGTGTTCGACAAGAGCATCCTGGAGAACGAGGCCATCCTGGACCAGATCGC CGAAATCCTGACCATCTACCAGGACGAGCAGAGCATCAAGGAGGAGCTGAACAAGCTGCCCGAGATCCTGA ACGAGCAGGACAAGGCCGAAATCGCCAAGCTGATCGGCTACAACGGCACCCACAGACTGAGCCTGAAGTG CATCCACCTGATCAACGAGGAGCTGTGGCAGACCAGCAGAAACCAGATGGAGATCTTCAACTACCTGAACA TCAAGCCCAACAAGGTGGACCTGAGCGAGCAGAACAAGATCCCCAAGGACATGGTGAACGACTTCATCCTG AGCCCCGTGGTGAAGAGAACCTTCATCCAGAGCATCAACGTGATCAACAAGGTGATCGAGAAGTACGGCAT CCCCGAGGACATCATCATCGAGCTGGCCAGAGAGAACAACAGCGACGACAGAAAGAAGTTCATCAACAAC CTGCAGAAGAAGAACGAGGCCACCAGAAAGAGAATCAACGAGATCATCGGCCAGACCGGCAACCAGAAACG CCAAAAGAATCGTGGAGAAGATCAGACTGCACGACCAGCAGGAGGGCAAGTGCCTGTACAGCCTTGAAAG CATTGCCCTGATGGACCTGCTGAACAACCCCCAGAATTACGAGGTGGACCACATCATCCCCAGAAGCGTGGC CTTTGACAACAGCATCCACAACAAGGTGCTGGTGAAGCAGATCGAGAACAGCAAGAAGGGCAACAGAACC CCCTACCAGTACCTGAACAGCAGCGACGCCAAACTGAGCTACAACCAGTTCAAGCAGCACATCCTGAACCT GAGCAAGAGCAAGGACAGAATCAGCAAGAAGAAGAAGGACTACCTGCTGGAGGAGAGAGACATCAACAA GTTCGAGGTGCAGAAGGAGTTCATCAACAGAAACCTGGTGGACACCAGATACGCCACCAGAGAGCTGACC AGCTACCTGAAGGCCTACTTCAGCGCCAACAACATGGACGTGAAGGTCAAGACCATCAACGGCAGCTTCAC CAACCACCTGAGAAAGGTGTGGAGATTCGACAAGTACAGAAACCACGGCTACAAGCACCACGCCGAGGAC GCTCTGATTATTGCCAATGCCGATTTCCTGTTCAAGGAGAACAAGAAGCTGCAGAACGCCAACAAGATCCTG GAGAAGCCCACCATCGAGAACAACACCAAGAAGGTGACCGTGGAGAAGGAGGAGGACTACAACAACGTG TTCGAAACCCCCAAGCTGGTGGAGGACATTAAGCAGTACAGAGACTACAAGTTCAGCCACAGAGTGGACAA GAAGCCCAACAGACAGCTGATCAACGACACCCTGTACAGCACCAGAATGAAGGACGAGCACGACTACATC GTGCAGACCATCACCGACATCTACGGCAAGGACAACACCAACCTGAAGAAGCAGTTCAACAAGAACCCCG AGAAGTTCCTGATGTACCAGAACGACCCCAAGACCTTCGAGAAGCTGAGCATCATCATGAAGCAGTACAGC GACGAGAAGAACCCCCTGGCCAAATACTACGAGGAAACCGGCGAATACCTGACCAAGTACAGCAAGAAGA ACAACGGCCCCATCGTGAAGAAGATCAAGCTGCTGGGCAACAAGGTGGGCAACCACCTGGATGTGACAAA CAAGTACGAGAACAGCACCAAGAAGCTGGTGAAGCTGAGCATCAAGAACTACAGATTCGACGTGTACCTGA CCGAGAAGGGCTACAAGTTCGTGACCATCGCCTACCTGAACGTGTTCAAGAAGGACAACTACTACTACATCC CCAAGGACAAGTACCAGGAGCTGAAGGAGAAGAAGAAGATCAAGGACACCGACCAGTTCATCGCCAGCTT CTACAAGAACGACCTGATCAAGCTGAACGGCGACCTGTACAAGATCATCGGCGTGAACAGCGACGACAGAA ACATCATCGAGCTGGACTACTACGACATCAAGTACAAGGACTACTGCGAGATCAACAACATCAAGGGCGAG CCCAGAATCAAGAAAACCATCGGCAAAAAGACCGAGAGCATCGAGAAGTTCACCACCGACGTGCTGGGCA ACCTGTACCTGCATAGCACCGAAAAGGCCCCCCAGCTGATTTTCAAGAGAGGCCTG
Cas9 Amino Acid	MKEKYILGLDLGITSVGYGIINFETKKIIDAGVRLFPEANVDNNEGRRSKRGSRRLKRRRIHRLERVKLLLTEYDLI NKEQIPTSNNPYQIRVKGLSEILSKDELAIALLHLAKRRGIHNINVSSEDEDASNELSTKEQINRNNKLLKDKYVC EVQLQRLKEGQIRGEKNRFKTTDILKEIDQLLKVQKDYHNLDIDFINQYKEIVETRREYFEGPGQGSPFGWNGDL KKWYEMLMGHCTYFPQELRSVKYAYSADLFNALNDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIA KEIGVNPEDIKGYRITKSGTPQFTEFKLYHDLKSIVFDKSILENEAILDQIAEILTIYQDEQSIKEELNKLPEILNEQD KAEIAKLIGYNGTHRLSLKCIHLINEELWQTSRNQMEIFNYLNIKPNKVDLSEQNKIPKDMVNDFILSPVVKRTFI QSINVINKVIEKYGIPEDIIIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQQEGK CLYSLESIALMDLLNNPQNYEVDHIIPRSVAFDNSIHNKVLVKQIENSKKGNRTPYQYLNSSDAKLSYNQFKQHIL NLSKSKDRISKKKKDYLLEERDINKFEVQKEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHL RKVWRFDKYRNHGYKHHAEDALIIANADFLFKENKKLQNANKILEKPTIENNTKKVTVEKEEDYNNVFETPKL VEDIKQYRDYKFSHRVDKKPNRQLINDTLYSTRMKDEHDYIVQTITDIYGKDNTNLKKQFNKNPEKFLMYQNDP KTFEKLSIIMKQYSDEKNPLAKYYEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTNKYENSTKKLVKLSI KNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELKEKKKIKDTDQFIASFYKNDLIKLNGDLYKIIGV

	NSDDRNIIELDYYDIKYKDYCEINNIKGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL
Lengh (aa)	1054
Repeat	GTTTTAGTACTCTGTAATTTTAGGTATGAGTGAAAC
mature crRNA	GUUUUAGUACUCUGUA
tracrRNA	UAAAAUUACAGAAUCUACUGAAACAAGACUAUAUGUCGUGUUUUAUCCCACUAAUUUUAUUAGUGGGGAUUUUUU

Supplementary Table 3b. Plasmids used in this study.

Plasmid	Description	Insert sequence
pACYC184-Om1Cas9	Expression Om1Cas9 intact locus in *E. coli*	AATACAATATAAAAGCACTCGCTAGTACGGCGAGTGCTTATTTGTGGTTATAAAAGTAGATTTTTATTTGTTTAAGCTATATATTTAGATTATCAAGTAGATAATACGAAGTGCTTATT TGTGGTTATAAAAGTAGATTTTTATTTGTTTAAGATATATATTTAGATTATCAAGTAGATAATACGAAGTGCTTATTTGTGGTTATAAAAGTAGATTTTTATTTGTTTAAGATATATATTT AGATTATCAAGTAGATAATACGAAGTGCTTATTTGTGGTTATAAAAGTAGATTTTTATTTGTTTAAGATATATATTTAGATTATCAAGTAGATAATACGAAGTATGTATTACTAAAATTT TTAGTAGTAAAGATTTATACGCTAGAGGTTTTAGTACTCTGTAATTTTAGGTATGAGTGAAACTGGTCAAATGGAATTATGGTCATCAGTGTAGTTTTAGTACTCTGTAATTTTAGGT ATGAGTGAAACACCACTACCTGATTTCTTTAAGTTTTCAGCGTTTTAGTACTCTGTAATTTTAGGTATGAGTGAAACGAGTTATTAATTCTTAATACTTACCATCTAGTTTTAGTACT CTGTAATTTTAGGTATGAGTGAAACGAGTTATTAATTCTTAATACTTACCATCTAGTTTTAGTACTCTATAATTTTAGGTATGAGTGAAACATTATACACAATTTCGGATACATCCAG TAGTTTTAGTACTCTGTAATTTTAGGTATGAGTGAAACTGAATAAGAAAGCATTCCTAAGCCAGTATTGTTTTAGTACTCTGTAATTTGAATATTGAAAGTGTAATATTTTAGTTGCA ATGCTATTGAATAACAATTATAATTTAATTACAGACAACAAGTTTAGCTAACTTGTACTCATACCTAAAATTACAGAATCTACTGAACAAGACTATATGTCGTTTATCCCACTAA TTTATTAGTGGGATTTTTTTGTTTTAGAACAACTTATGTAGAAAAATATAAAAAGTAGTGCCACGTTTATTGGTTTTAAATTATAAAATAAAACAAATAAGGTGGGTGTAAAC TATGAAAGAAAAATATATTTTGGGTTTAGATTTAGGTATTACATCAGTAGGATATGGAATTATTAATTTTGAAACTAAAAAAATTATAGATGCAGGAGTTAGGCTTTTTCCAGAAGC AAACGTAGATATAATGAAGGTCGTAGAAGTAAAAGAGGATCTAGACGATTAAAAAGACGAAGAATACATAGATTAGAACGAGTAAAATTATTACTCACAGAATATGACCTAATA AATAAAGAACAAATACCAACATCAAATAATCCTTATCAAATTAGAGTTAAAGGACTAAGTGAAATATTAAGTAAAGATGAACTTGCTATTGCTTTATTACATTTAGCGAAAGAAG GGGAATTCACAACATTAATGTTAGTAGTGAAGATGAGGATGCCTCCAATGAGTTATCTACAAAGGAACAGATTAATAGAAATAATAAACTATTAAAAGATAAAATATGTTTGTGAAG TACAGTTACAAAGATTAAAAGAAGGACAAATAAGGGGAGAAAAAAACAGGTTCAAAACAACAGATATATTAAAGGAAATTGATCAACTCCTTAAAGTGCAAAAAGATTATCAC AACTTAGATATAGATTTCATTAACCAATATAAAGAAATAGTTGAAACACGTCGAGAGTATTTTGAAGGACCAGGACAAGGTAGTCCTTTTGGATGGAATGGTGATTTGAAGAAAT GGTATGAAATGTTAATGGGACACTGCACATATTTCCCCAAGAATTACGAAGTGTAAAATATGCATATTCTGCTGATTTATTTAACGCATTAAATGATTTAAATATCTCATAATACA AAGAGATAATTCTGAAAAATTAGAATATCATGAAAAATATCATATTATTGAAAATGTATTTAAACAAAAGAAAAAGCCAACTTTAAAACAAATTGCTAAAGAAATAGGGGTAAATC CAGAAGATTATAAGGTTATAGAATAACAAAATCTGGGACACCGCAATTTACAGAATTTAAATTATATCATGATCTAAAGTCAATTGTGTTTGATAAAAGTATTTAGAAAATGAAG CTATTTTAGACCAAATTGCTGAAATATTAACAATCTATCAAGATGAGCAGAGCATAAAAGAAGAGCTAAACAAATTACCAGAAATCTTAAATGAACAAGATAAAGCTGAAATAGC CTATTTAGACCAAATTGCTGAAATATTAACAATCTATCAAGATGAGCAGAGCATAAAAGAAGAGCTAAACAAATTACCAGAAATCTTAAATGAACAAGATAAAGCTGAAATAGC AAACCGAACAAAGTCGACTTATCCGAACAAAATAAAATACCAAAAGATATGGTGAACGACTTTATTTTATCTCCTGTAGTTAAGAGGACTTTTATTCAATCAATCAATGTTATCAAT AAAGTGATTGAAAAATACGGTATACCTGAAGATATCATTATAGAATTAGCTCGAGAAAATAATAGTGATGATAGAAAAAAATTTATTAATAACCTACAAAAGAAAAACGAAGCAA CTCGTAAGAGAATAAATGAAATAATTGGCCAAACTGGAAATCAAAATGCCAAAAGAATTGTAGAGAAGATAAGATTACATGATCAACAAGAAGGTAAATGTTTGTATTCATTAGA AAGTATAGCTTTAATGGATTTGTTAAATAATCCTCAAAATTATGAAGTCGATCATATTATTCCTCGTTCTGTAGCATTTGATAACTCAATTCATAATAAAGTGTTAGTTAAACAAATAG AAAATAGTAAAAAAGGTAATAGAACACCATACCAATACCTTAATTCTAGTGATGCAAAGTTATCATACAATCAATTTAAGCAACATATCTTAAATTTAAGTAAGTCGAAAGATAGA ATTTCTAAGAAAAAGAAAGACTATTTATTGGAAGAACGGGATATTAATAAATTCGAAGTACAAAAGGAATTTATCAATCGAAATTTAGTAGATACAAGATATGCGACTAGAGAACT TACTAGCTATCGAAAGCATATTTAGTGCTAATAACATGGATGTAAAAGTTAAAACAATTAATGGGAGTTTTACCAATCATTTAAGAAAAGTATGGAGATTTGATAAATATAGAAAT CATGGTTACAAACATCATGCTGAAGACGCTTTAATAATCGCAAATGCAGATTTTTATTTAAAGAAAATAAAAAATTACAAAATGCAAATAAAATTTTAGAAAAACCAACAATTGA AAATAATACTAAAAAGTAACAGTTGAAAAAGAAGAAGATTATAACAACGTGTTTGAGACACCTAAGCTTGTTGAAGACATTAAACAATACAGAGATTATAAATTTTCTCATAGA GTAGATAAGAAACCGAATAGACAACTTATTAATGACACTTTATATTCTACAAGAATGAAAGCGAACATGATTACATTGTTCAAACTATTACGGATATTTATGGGAAAGATAATACT AATTTAAAAAGCAGTTTAATAAAATCCAGAAAAATTTTAATGTACCAAAATGATCCAAAAACTTTTGAAAAATTATCAATTATATGAAACAGTATTCTGATGAAAAAACCC GCTAGCTATAGTATTATGAAGACCGGAGAATATTTAACAAAGTATAGTAAAAAAAATATGGTCCAATTGTTAAAAGATAAAGCTTTAGGAAATAAAGTTATGGAAATCATTTAG ATGTAACAAATAAATACGAAAATTCTACAAAAAAATTAGTGAAATTATCAATTAAAAATTATAGGTTTGATGTTTATTTAACTGAAAAAGGATATAAATTCGTAACAATTGCTTATTT AAATGTATTTAAGAAAGATAATTATTATTATATCCCTAAAGATAAGTATCAAGAATTAAAAGAAAAGAAAAAAATTAAAGATACAGATCAATTTATTGCCAGTTTTATAAGAAGTAT ATCATTGAATTAGATTATTACGATATTAAATATAAAGATTACTGTGAGATAAATAATATTAAAGGTGAACCACGTATAAAGAAAACAATTGGTAAAAAGACGGAATCAATTG AAAAATTTACAACAGATGTTTTAGGTAATTTGTATTTGCATAGTACTGAAAAAGCACCACAATTAATATTTAAACGAGGTCTTaaacgcccggcggcgactaaaaagcaggccaggcgaaaaaaagaa aggattacaaggacgacgatgacaagtaa

pX458-Om1Cas9-NC	Expresses SpCas9, GFP, and sgRNA in human HEK 293T cell (non-target plasmid)	gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaac ttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccggggtcttcgagaagacctGTTTTAGTACTCTGTAcgcTAAAATTACAGAATCTACTGAAACAAGACTATATGTCGTGTTTATCCCACTAAT TTATTAGTGGGATTTTTTtttttgttttagagctagaaatagcaagttaaaataaggctagtccgtttttagcgcgtgcgccaattctgcagacaaatggctctagaggtacccgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgac gtcaatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattgtgcccagtacatgaccttatgggactttcctacttgg cagtacatctacgtattagtcatcgctattaccatggtcgaggtgagccccacgttctgcttcactctccccatctcccccccctccccacccccaattttgtatttatttattttttaattattttgtgcagcgatgggggcgggggggggggggggcgcgcgccaggcggggcggggcgggg cgaggggcggggcggggcgaggcggagaggtgcggcggcagccaatcagagcggcgcgctccgaaagtttccttttatggcgaggcggcggcggcggcggccctataaaaagcgaagcgcgcggcgggcgggagtcgctgcgcgctgccttcgccccgtgccccgctccgcc gccgcctcgcgccgcccgccccggctctgactgaccgcgttactcccacaggtgagcgggcgggacggcccttctcctccgggctgtaattagctgagcaagaggtaagggtttaagggatggttggttggtggggtattaatgtttaattacctggagcacctgcctgaaatcacttttttc aggttggaccggtgccaccatggactataaggaccacgacggagactacaaggatcatgatattgattacaaagacgatgacgataagatggccccaaagaagaagcggaaggtcggtatccacggagtcccagcagccATGAAGGAGAAGTACATCCTGGGC CTGGACCTGGGCATTACCAGCGTGGGATATGGCATTATCAACTTCGAGACAAAGAAGATCATCGACGCCGGCGTGAGACTGTTCCCCGAAGCTAATGTGGATAACAACGAGGGC AGAAGAAGCAAGAGAGGCAGCAGGAGACTGAAGAGAAGAAGAATCCACAGACTGGAGAGAGTGAAGCTGCTGCTGACCGAGTACGACCTGATCAACAAGGAGCAGATCCCC ACCAGCAACAACCCCTACCAGATCAGAGTGAAGGGCCTGAGCGAGATCCTGAGCAAGGACGAGCTGGCCATTGCCCTGCTTCATCTGGCTAAAAGAAGAGGCATCCACAACAT CAACGTGAGCAGCGAGGACGAGGACGCCAGCAATGAACTGAGCACCAAGGAGCAGATCAACAGAAACAACAAGCTGCTGAAGGACAAGTACGTGTGCGAGGTGCAGCTGCA GAGACTGAAGGAGGGCCAAATTAGAGGCGAGAAGAACAGATTCAAGACCACCGACATCCTGAAGGAGATCGACCAGCTGCTGAAGGTGCAGAAGGACTACCACAACCTGGA CATCGACTTCATCAACCAGTACAAGGAGATCGTGGAGACAAGAAGAGAGTACTTCGAGGGCCCCGGCCAAGGCAGCCCTTTTGGATGGAATGGAGATCTGAAGAAGTGGTACG AGATGCTGATGGGCCACTGCACCTACTTCCCCCAGGAACTGAGAAGCGTGAAGTACGCCTATAGCGCCGACCTGTTCAACGCCCTGAACGACCTGAATAACCTGATCATCCAGA GAGACAACAGCGAGAAGCTGGAGTACCACGAGAAGTACCACATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAGGAGATCGGCGTGA ACCCCGAAGATATCAAGGGCTACAGAATCACCAAGAGCGGCACCCCCCAATTCACCGAATTTAAGCTGTACCACGACCTGAAGAGCATCGTGTTCGACAAGAGCATCCTGGAG AACGAGGCCATCCTGGACCAGATCGCCGAAATCCTGACCATCTACCAGGACGAGCAGAGCATCAAGGAGGAGCTGAACAAGCTGCCCGAGATCCTGAACGAGCAGGACAAGG CCGAAATCGCCAAGCTGATCGGCTACAACGGCACCCACAGACTGAGCCTGAAGTGCATCCACCTGATCAACGAGGAGCTGTGGCAGACCAGCAGAAACCAGATGGAGATCTTC AACTACCTGAACATCAAGCCCAACAAGGTGGACCTGAGCGAGCAGAACAAGATCCCCAAGGACATGGTGAACGACTTCATCCTGAGCCCCGTGGTGAAGAGAACCTTCATCCA GAGCATCAACGTGATCAACAAGGTGATCGAGAAGTACGGCATCCCCGAGGACATCATCATCGAGCTGGCCAGAGAGAACAACAGCGACGACAGAAAGAAGTTCATCAACAAC CTGCAGAAGAAGAACGAGGCCACCAGAAAGAGAATCAACGAGATCATCGGCCAGACCGGCAACCAGAACGCCAAAAGAATCGTGGAGAAGATCAGACTGCACGACCAGCAG GAGGGCAAGTGCCTGTACAGCCTTGAAAGCATTGCCCTGATGGACCTGCTGAACAACCCCCAGAATTACGAGGTGGACCACATCATCCCCAGAAGCGTGGCCTTTGACAACAG CATCCACAACAAGGTGCTGGTGAAGCAGATCGAGAACAGCAAGAAGGGCAACAGAACCCCCTACCAGTACCTGAACAGCAGCGACGCCAAACTGAGCTACAACCAGTTCAAG CAGCACATCCTGAACCTGAGCAAGAGCAAGGACAGAATCAGCAAGAAGAAGAAGGACTACCTGCTGGAGGAGAGAGACATCAACAAGTTCGAGGTGCAGAAGGAGTTCATC AACAGAAACCTGGTGGACACCAGATACGCCACCAGAGAGCTGACCAGCTACCTGAAGGCCTACTTCAGCGCCAACAACATGGACGTGAAGGTCAAGACCATCAACGGCAGCT TCACCAACCACCTGAGAAAGGTGTGGAGATTCGACAAGTACAGAAACCACGGCTACAAGCACCACGCCGAGGACGCTCTGATTATTGCCAATGCCGATTTCCTGTTCAAGGAG AACAAGAAGCTGCAGAACGCCAACAAGATCCTGGAGAAGCCCACCATCGAGAACAACACCAAGAAGGTGACCGTGGAGAAGGAGGAGGACTACAACAACGTGTTCGAAACC CCCAAGCTGGTGGAGGACATTAAGCAGTACAGAGACTACAAGTTCAGCCACAGAGTGGACAAGAAGCCCAACAGACAGCTGATCAACGACACCCTGTACAGCACCAGAATGA AGGACGAGCACGACTACATCGTGCAGACCATCACCGACATCTACGGCAAGGACAACACCAACCTGAAGAAGCAGTTCAACAAGAACCCCGAGAAGTTCCTGATGTACCAGAA CGACCCCAAGACCTTCGAGAAGCTGAGCATCATCATGAAGCAGTACAGCGACGAGAAGAACCCCCTGGCCAAATACTACGAGGAAACCGGCGAATACCTGACCAAGTACAGC AAGAAGAACAACGGCCCCATCGTGAAGAAGATCAAGCTGCTGGGCAACAAGGTGGGCAACCACCTGGATGTGACAAACAAGTACGAGAACAGCACCAAGAAGCTGGTGAAG CTGAGCATCAAGAACTACAGATTCGACGTGTACCTGACCGAGAAGGGCTACAAGTTCGTGACCATCGCCTACCTGAACGTGTTCAAGAAGGACAACTACTACTACATCCCCAAG GACAAGTACCAGGAGCTGAAGGAGAAGAAGAAGATCAAGGACACCGACCAGTTCATCGCCAGCTTCTACAAGAACGACCTGATCAAGCTGAACGGCGACCTGTACAAGATCA TCGGCGTGAACAGCGACGACAGAAACATCATCGAGCTGGACTACTACGACATCAAGTACAAGGACTACTGCGAGATCAACAACATCAAGGGCGAGCCCAGAATCAAGAAAAC CATCGGCAAAAAGACCGAGAGCATCGAGAAGTTCACCACCGACGTGCTGGGCAACCTGTACCTGCATAGCACCGAAAAGGCCCCCCAGCTGATTTTCAAGAGAGGCCTGaaaagg ccggcggccacgaaaaaggccggccaggcaaaaaagaaaaaggaattcggcagtggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccaatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccagggccgtacg caccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcgatccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccg gagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggagtctcgcccgaccaccagggcaa gggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgc aagcccggtgccgaattcggcagtggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccagtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcg agggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaagg ctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacg tctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagca aagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaaggaattctaa

Supplementary Table 3c: Guide RNA and single-string DNA (ssDNA) template used for DocMF.

Guide RNA	CUCAAAAGGGAACUGCUACCGAAGUUUUAGUACUCUGUAcgcUAAAAUUACAGAAUCUACUGAAACAAG ACUAUAUGUCGUGUUUAUCCCACUAAUUUAUUAGUGGGAUUUUUU
ssDNA	TGTGAGCCAAGGAGTTGGCCTAGGCAATTGTCTTCCTAAGACCGCTTGGCCTCCGACTTNNNNNNNNNNNN NNNCTCAAAAGGGAACTGCTACCGAANNNNNNNNNNNNNNNNNAAGTCGGATCGTAGCCATGTCGTTC

Supplementary Table 3d. The PCR primers of *AAVS1* gene

AAVS1F1	TCCCTTTTCCTTCTCCTTCT
AAVS1R1	TCACAAAGGGAGTTTTCCAC

Supplementary Table 3e. The endogenous target sites of human and their downstream PAM sequences

Gene	Editing sites name	<target sites> <PAM>
BCL11a	Om1Cas9 B1	<ATCGGTGGCCGTTTGCCCAG><GGGG>
	Om1Cas9 B2	<CCCTGGGCAAACGGCCACCGATG><GAGA>
	Om1Cas9 B3	<CTGGGCAAACGGCCACCGATGGA><GAGG>
	Om1Cas9 B4	<CTGATTAGGGTGGGGGCGTGGGT><GGGG>
	Om1Cas9 B5	<TCTACTCTTAGACATAACACACC><AGGG>
HBG	Om1Cas9 H1	<CATTGAGATAGTGTGGGGAA><GGGG>
	Om1Cas9 H2	<TTGCATTGAGATAGTGTGGGGAA><GGGG>
	Om1Cas9 H3	<CCTCACTGGATACTCTAAGACTA><TTGG>
	Om1Cas9 H4	<TTGCCTTGTCAAGGCTATTGGTC><AAGG>
	Om1Cas9 H5	<TCAAGTTTGCCTTGTCAAGGCTA><TTGG>

Supplementary Table 3f. A list of PCR primers used for deep sequencing.

Editing Sites	F_primer	R_primer
Om1Cas9 B1 T	AAGACCTCTAacacaccagggtcaatacaact	gggaagcttcacctcctttaca
Om1Cas9 B1 M	CGTGCGATCCacacaccagggtcaatacaact	gggaagcttcacctcctttaca
Om1Cas9 B2 T	AGTTGCCATAacacaccagggtcaatacaact	gggaagcttcacctcctttaca
Om1Cas9 B2 M	CGTGCGATCCacacaccagggtcaatacaact	gggaagcttcacctcctttaca
Om1Cas9 B3 T	ATTCAACGGAacacaccagggtcaatacaact	gggaagcttcacctcctttaca
Om1Cas9 B3 M	CGTGCGATCCacacaccagggtcaatacaact	gggaagcttcacctcctttaca
Om1Cas9 B4 T	AACTGTACTGccccaggtgtgcataagtaaga	ctccatcaccaagagagcct
Om1Cas9 B4 M	TGAAGCGTTGccccaggtgtgcataagtaaga	ctccatcaccaagagagcct
Om1Cas9 B5 T	GTACCTCAATtgtagtgagatggctgaaaagc	gtctgccagtcctcttctacc
Om1Cas9 B5 M	GACTTCTAATtgtagtgagatggctgaaaagc	gtctgccagtcctcttctacc
Om1Cas9 H1 T	TTAGATGCATtggaactgctgaagggtgc	atgactgaatcggaacaaggca
Om1Cas9 H1 M	CCGATGTCGCtggaactgctgaagggtgc	atgactgaatcggaacaaggca
Om1Cas9 H2 T	GTCCAGAGCTtggaactgctgaagggtgc	atgactgaatcggaacaaggca
Om1Cas9 H2 M	CCGATGTCGCtggaactgctgaagggtgc	atgactgaatcggaacaaggca
Om1Cas9 H3 T	CACGTGATAGgccttgtcctcctctgtgaaat	cccttccccacactatctcaat
Om1Cas9 H3 M	TCGGAAGGCAgccttgtcctcctctgtgaaat	cccttccccacactatctcaat
Om1Cas9 H4 T	CCACTAGTCCgccttgtcctcctctgtgaaat	cccttccccacactatctcaat
Om1Cas9 H4 M	TCGGAAGGCAgccttgtcctcctctgtgaaat	cccttccccacactatctcaat
Om1Cas9 H5 T	TGGACTTGGCgccttgtcctcctctgtgaaat	cccttccccacactatctcaat
Om1Cas9 H5 M	TCGGAAGGCAgccttgtcctcctctgtgaaat	cccttccccacactatctcaat

Supplementary Table 4. Information of the 121 candidate antimicrobial peptides.

ID	cAMP sequence	Genome ID	Taxonomy	Verification*
APD_25	SCTTCVCTCSCCTT	GCA_002005165.1_ASM200516v1_genomic	Novibacillus thermophilus	APD3
APD_43	WKSESLCTPGCVTGALQTCFLQTLTCNCKISK	GCA_003667885.1_ASM366788v1_genomic	Bacillus vallismortis	APD3
APD_47	TTWACATVTLTVTVCSPTGTLCGSCSMGTRGCC	GCA_003846185.1_ASM384618v1_genomic	Streptomyces	APD3
APD_94	ITSVSWCTPGCTSEGGGSGCSHCC	MARD_SAMN03999371_REFG_MMP03999371	Streptomyces sp001905385	APD3
cAMP_1	ITSKSLCTPGCITGILMCLTQNSCVSCNSCIRC	GCA_000009785.1_ASM978v1_genomic	Geobacillus thermoleovorans	Fail
cAMP_2	SVRPARAPRVNKTFASTLKWMASVPVCFHFGCWSAFCSA	GCA_000219105.1_ASM21910v1_genomic	Myxococcus macrosporus	Fail
cAMP_3	GRCGTGCSGSQTRLCC	GCA_000219105.1_ASM21910v1_genomic	Myxococcus macrosporus	Non-effective
cAMP_4	ACGTGDGCKSTCASSCASAV	GCA_000220705.2_ASM22070v2_genomic	Streptomyces xinghaiensis	Non-effective
cAMP_5	GCATCSIGAACLVDGPIPDFEIAGATGLFGLWG	GCA_000264395.1_C89_version_1_genomic	Bacillus atrophaeus	Fail
cAMP_6	TTWACATITLTVTVCSPTGTLCGSCSMGTRGCC	GCA_000743295.1_ASM74329v1_genomic	Streptomyces globisporus_C	Fail
cAMP_7	GSGVLGTLGCCSCLPWYSGWTVCGLACNPGKPCKN	GCA_000743295.1_ASM74329v1_genomic	Streptomyces globisporus_C	Non-effective
cAMP_8	FDTNTLTTSTSILISMTLGNNGWVCTATKECMPSCN	GCA_000756615.1_ASM75661v1_genomic	Paenibacillus durus	Fail
cAMP_9	DGGCGSTCGTSCVSNAA	GCA_000829715.2_ASM82971v2_genomic	Streptomyces platensis	Fail
cAMP_10	AICSPCPPRHCL	GCA_000948975.1_ASM94897v1_genomic	Thalassomonas actiniarum	Non-effective
cAMP_11	IGFRNVLKKKNKFAIAWVGVSLAACLLASVLVVATGFGTP	GCA_001015055.1_ASM101505v1_genomic	Aneurinibacillus_A tyrosinisolvens	Fail
cAMP_12	VTPSVVVSVITGFISTNTCPSTACTRAC	GCA_001307105.1_ASM130710v1_genomic	Bacillus australimaris	Non-effective
cAMP_13	CSTNTFSLSDYWGNNGGWCTISHECMSWCK	GCA_001307105.1_ASM130710v1_genomic	Bacillus australimaris	Fail
cAMP_14	HYFETFWRSCVLSGGYWKWGIQEIC	GCA_001307105.1_ASM130710v1_genomic	Bacillus australimaris	Fail
cAMP_15	GDVDPQTTPSSVACGIAVSALFCPTTKCTSQC	GCA_001307805.2_ASM130780v2_genomic	Lentibacillus amyloliquefaciens	Fail
cAMP_16	SDVDPRWTPVISLVARSIVKSTKACIGAGASAISGIVSHNKDCLG	GCA_001456895.1_ASM145689v1_genomic	Exiguobacterium_A indicum	Non-effective
cAMP_17	THRDFFLLTQCSSCTASSPIQCC	GCA_001482195.1_ASM148219v1_genomic	Luteimonas abyssi	Non-effective
cAMP_18	SSSCSCSSCCSCSTSSCCSSSTSTGCGGG	GCA_001507315.1_ASM150731v1_genomic	Micromonospora maris	Fail
cAMP_19	NNGGCETSRTSLCTDTLSTCC	GCA_001753705.1_ASM175370v1_genomic	Streptomyces sp000424765	Effective
cAMP_20	ASWWPWPICIPTSCREVSCLRKTFINAVK	GCA_001767235.1_ASM176723v1_genomic	Moorea producens_A	Non-effective
cAMP_21	ITSISACTPGCGNTGSFNSFCC	GCA_001884135.1_ASM188413v1_genomic	Bacillus_A cereus_P	Non-effective
cAMP_22	RTENLATFGCCFSPQTFTYTSPILCSVTLSVC	GCA_001941645.1_ASM194164v1_genomic	Actinoalloteichus sp001941625	Fail
cAMP_23	TTWPCFGITVSIAACQGTKGYGSVGCC	GCA_001941645.1_ASM194164v1_genomic	Actinoalloteichus sp001941625	Non-effective
cAMP_24	CAWYDISCKLGNKGAWCTLTVECQSSCN	GCA_001999205.1_ASM199920v1_genomic	Bacillus velezensis	Fail
cAMP_26	IPKISHNLQCT	GCA_002154725.1_ASM215472v1_genomic	Aulosira sp002154725	Non-effective
cAMP_27	GQGWGTLVSSLSCYGASYVLGNNGNFCTATAECQNNCK	GCA_002382885.1_WZ.A104_genomic	Streptomyces sp002382885	Non-effective
cAMP_28	LDTISIIKTTTKLATKVTCKSTCTCNCTNYCSILC	GCA_002744245.1_ASM274424v1_genomic	Bacillus pumilus_M	Fail

cAMP_29	DNGGCATSRTSLCTDTLSGCC	GCA_002754675.1_ASM275467v1_genomic	Streptomyces sp002754675	Fail
cAMP_30	GQGWGTIVSSLSCYGASYVLGNNGNFCTATAECQNNCK	GCA_002832675.1_ASM283267v1_genomic	Streptomyces filamentosus	Non-effective
cAMP_31	RRGSPNWPPRRACG	GCA_002846525.1_ASM284652v1_genomic	Micromonospora echinofusca	Effective
cAMP_32	SSEPQGTPATIIPVSLAICPTTKCASVVKPCND	GCA_002849835.1_ASM284983v1_genomic	Virgibacillus dokdonensis	Fail
cAMP_33	SCTTCECCCSCSS	GCA_002895405.1_ASM289540v1_genomic	Actinoalloteichus cyanogriseus	Fail
cAMP_34	SKKSKPGDGKFGRGVKRG	GCA_002993925.1_ASM299392v1_genomic	Bacillus paralicheniformis	Non-effective
cAMP_35	GLYTSTCYTSSCYTSTCYTSTCYSSDCYTGQNMCGYTHTSSC	GCA_003313305.1_ASM331330v1_genomic	Vallitalea guaymasensis	Fail
cAMP_36	NNGGCATSRTSLCTDTLSGCC	GCA_003323715.1_ASM332371v1_genomic	Streptomyces	Fail
cAMP_37	LPTRVPSCFHP	GCA_003326795.1_ASM332679v1_genomic	Thalassospira	Non-effective
cAMP_38	ERFGWLSLALRC	GCA_003410355.1_ASM341035v1_genomic	Bacillus_A anthracis	Non-effective
cAMP_39	GCSTSSCSCSSSTTSCTSTASCA	GCA_003432485.1_ASM343248v1_genomic	Spirillospora	Fail
cAMP_40	STLSLLSCVSAASVTLCL	GCA_003443735.1_ASM344373v1_genomic	Streptomyces sviceus	Fail
cAMP_41	GWWGHWWDGFNGWRDW	GCA_003626645.1_ASM362664v1_genomic	Streptomyces	Non-effective
cAMP_42	ADCANVCTWTKDCSICPSWSCWSWSC	GCA_003667765.1_ASM366776v1_genomic	Bacillus_AZ	Non-effective
cAMP_44	ASSGNICTATTECTYWSAICC	GCA_003833015.1_ASM383301v1_genomic	Hungatella	Fail
cAMP_45	VTSVSLCTPGCITGVIMTCTIKTATCGCHVAGK	GCA_003833015.1_ASM383301v1_genomic	Hungatella	Non-effective
cAMP_46	CTATCAWTCAITSIEQQ	GCA_003846185.1_ASM384618v1_genomic	Streptomyces	Fail
cAMP_48	GSGVLGTLGCCSCLPWYSGWTVCGLACDPGKPCKN	GCA_003846185.1_ASM384618v1_genomic	Streptomyces	Non-effective
cAMP_49	CTRTCTWTCSITSIEQQ	GCA_003846255.1_ASM384625v1_genomic	Streptomyces anulatus_A	Fail
cAMP_50	GEVRPQSGPVCAVTLAVCVPVSLAFCPTTACTSDCR	GCA_003856495.1_ASM385649v1_genomic	Thermoactinomycetaceae CDF	Fail
cAMP_51	CDWWNISCHLGNTGRFCTLTKECQPNCNY	GCA_003856495.1_ASM385649v1_genomic	Thermoactinomycetaceae CDF	Non-effective
cAMP_52	CDWWNISCHLGNTGQFCTLTKECQPSCNR	GCA_003856495.1_ASM385649v1_genomic	Thermoactinomycetaceae CDF	Effective
cAMP_53	AATRPNWRTSEPRWSTACRWCWGWSCWSPWSCSS	GCA_004000405.1_ASM400040v1_genomic	Streptomyces bacillaris	Fail
cAMP_54	NYTINCPTSSCYTSRCHSSTCYSSRCYTGQNRCGYTSRC	GCA_004115085.1_ASM411508v1_genomic	Paenibacillales	Non-effective
cAMP_55	KIAAYTSTCYTSTCYTSRCYTSNCYTSKCYTGQNMCGYTHSSSC	GCA_004115085.1_ASM411508v1_genomic	Paenibacillales	Fail
cAMP_56	QGRSGGATMGPLCKMPSLFCTYIMCL	GCA_004209775.1_ASM420977v1_genomic	Synechococcus_C	Fail
cAMP_57	RIQCQWLPVLATHRA	GCA_004765815.2_ASM476581v2_genomic	Haloadaptaceae SW-6-65-15	Effective
cAMP_58	RGPYLSLHYRCY	GCA_004765815.2_ASM476581v2_genomic	Haloadaptaceae SW-6-65-15	Non-effective
cAMP_59	GDVHAQTTWPCATVGVSVALCPTTKCTSQC	GCA_005671335.1_ASM567133v1_genomic	Bacillus_M okuhidensis	Fail
cAMP_60	VNGACAWYNISCRLGNKGAYCTLTVECMPSCN	GCA_005671335.1_ASM567133v1_genomic	Bacillus_M okuhidensis	Fail
cAMP_61	CGTLGGTSCSYNGGCLCC	GCA_005954665.1_ASM595466v1_genomic	Lysobacter enzymogenes_B	Fail
cAMP_62	QAMKPRSRPCTPCVRAGACRGRQPRPR	GCA_005954665.1_ASM595466v1_genomic	Lysobacter enzymogenes_B	Non-effective
cAMP_63	VTGFTAPGQSCCAGSNGTYTSCC	GCA_006149045.1_ASM614904v1_genomic	Dyadobacter	Non-effective
cAMP_64	PEVNKEVVRGGAEDSDSAGWICTVSGECWGFSCNPFSRK	GCA_006383075.2_ASM638307v2_genomic	Kordia	Fail

cAMP_65	LQEDQLSNVKGGVRGVTCANASCANSCNQNSCNAQADQVLER	GCA_006491645.1_ASM649164v1_genomic	Flavobacterium	Non-effective
cAMP_66	GRCSLCHC	GCA_007004655.1_ASM700465v1_genomic	Cellvibrionaceae	Non-effective
cAMP_67	GEQAPPPSISTIATRTVCTRVGCQVTKTCACTSMCTIFCVGGK	GCA_008312835.1_ASM831283v1_genomic	Streptomyces	Fail
cAMP_68	NDVNPETTPATTSSWTCITAGVTVSASLCPTTKCTSRC	GCA_009905595.1_ASM990559v1_genomic	Bacillus licheniformis	Fail
cAMP_69	TITLSTCAILSKPLGNNGYLCTVTKECMPSCN	GCA_009905595.1_ASM990559v1_genomic	Bacillus licheniformis	Fail
cAMP_70	LFPVACHSLGFGGRAFCT	GCA_900473935.1_UW105_genomic	Synechococcus_C sp002171995	Non-effective
cAMP_71	LRCSKFCLNPNCFITERSTIRCKIPR	GCA_902623005.1_AG-893-M05_genomic	UBA9148	Non-effective
cAMP_72	CDRGPAHISQRC	GOMC.bin.11461	Pseudophaeobacter_A	Non-effective
cAMP_73	AGPGWVETLTKDCPWKQPGACVIIMGQKICKKCY	GOMC.bin.11526	Bacillus_A toyonensis	Non-effective
cAMP_74	GAQRGPGTAVQNCGWWPLLTGNPT	GOMC.bin.12791	Burkholderiaceae SCGC-AAA027-K21	Effective
cAMP_75	PEVNKDVVRGGLEDSDSAGWFCSVSGECWGFSCNPF	GOMC.bin.13238	Kordia	Fail
cAMP_76	GQWAYTYKEG	GOMC.bin.14155	Desulfocapsaceae	Non-effective
cAMP_77	AAHFGRTISYCGACP	GOMC.bin.14222	Massilia	Non-effective
cAMP_78	PSQAVCGKKRCFSISGGKWFIRKR	GOMC.bin.16027	Rhizobiales Im1	Fail
cAMP_79	SCTTCECCCSCS	GOMC.bin.2032	Myxococcota GCA-2862545	Fail
cAMP_80	SAETQDITLPIDLCWLSRLGGNQGWFCTITKECQVNCNVG	GOMC.bin.2337	Oceanobacillus kimchii	Fail
cAMP_81	TWCLWQTCETICINTCADTCVCSVGTGMCC	GOMC.bin.4741	Delongbacteria UBA4055	Fail
cAMP_82	AASSGWVCTVSGECNGGSSCNPFKDILPSFNEQR	GOMC.bin.5307	Pseudoalteromonas	Non-effective
cAMP_83	SGSGDAIVPATTYCSIVATCHAR	GOMC.bin.5307	Pseudoalteromonas	Non-effective
cAMP_84	NSNWGTCTLDCSLRCPTNGGTCGCPSGYGTCTVRGTGTGC	GOMC.bin.6444	Ktedonobacteraceae	Non-effective
cAMP_85	ARPIVTTAITCTQYTWLNWKACCY	GOMC.bin.6501	Holophagae GCA-2747255	Fail
cAMP_86	WRPVGVCRTP	GOMC.bin.7566	Salinibacteraceae	Non-effective
cAMP_87	SLFRSIRSFRVRWVCLSHRRIK	GOMC.bin.7566	Salinibacteraceae	Effective
cAMP_88	IPTGTVTWTPVCTNYGCL	GOMC.bin.8506	Nostocaceae	Non-effective
cAMP_89	PLSAPPGSRCC	GOMC.bin.8611	Roseitalea	Effective
cAMP_90	RLFHFPDCIRCVSIEIPPVAGCLCRERKCSASTI	GOMC.bin.8622	Tepidamorphus	Non-effective
cAMP_91	QVLAKARCLPKSFNSKDCSKNSGYGWAVGEAARV	GOMC.bin.8998	Pirellulaceae ARS98	Non-effective
cAMP_92	NILGGRPKTNGGGCWSFNATCSESQAVCCSWQC	MARD_SAMEA2272395_REFG_MMP2272395	Fibrisoma limi	Fail
cAMP_93	SFCFFLQSISAVKCQC	MARD_SAMEA4707921_REFG_MMP4707921	Synechococcus_C sp900473965	Fail
cAMP_95	ERWGRRPHAPVPSPPPRWTRAARPGRPGTGRC	MARD_SAMN03999384_REFG_MMP03999384	Streptomyces sp001905905	Effective
cAMP_96	LPGSCCSGLGGLVTILTCFGCTFGTN	MARD_SAMN03999393_REFG_MMP03999393	Streptomyces sp001905725	Fail
cAMP_97	CTTNTFSLSDALGNQGGWCTLTVECQPNCN	MARD_SAMN04026265_REFG_MMP04026265	Streptomyces nanshensis_A	Fail
cAMP_98	NLRGGTWTFPTFQTGCNTGNTCTCDGGTIHNPSDCACP	MARD_SAMN04488116_REFG_MMP04488116	Muricauda flava	Non-effective

cAMP_99	CTKTCTWTCRITSLEQQ	MARD_SAMN04571751_REFG_MMP04571751	Streptomyces sp001687325	Non-effective
cAMP_100	KNLRVTSFILCTPGTCNNKCPNTNWLCSNVCVTKTCWTCA	MARD_SAMN05421791_REFG_MMP05421791	Facklamia miroungae	Non-effective
cAMP_101	SLFWCTPGTCNNTCKGDSTLKSNCCGGSIICSLGGC	MARD_SAMN05421791_REFG_MMP05421791	Facklamia miroungae	Non-effective
cAMP_102	DAAPNTISITNPLIACALLSRNNTGRFCTVTVECNVPGPVC	MARD_SAMN05421803_REFG_MMP05421803	Nocardiopsis flavescens	Fail
cAMP_103	RRFCGKTTCCGRRM	MARD_SAMN05421843_REFG_MMP05421843	Litoreibacter meonggei	Effective
cAMP_104	LQDSQMSSFKGGVREAEGLTCANGSCIKSCNRNSCKDQVIN	MARD_SAMN05444344_REFG_MMP05444344	Tenacibaculum mesophilum	Non-effective
cAMP_105	ITSVSLCTPGCGETGSFNSYCC	MARD_SAMN05444487_REFG_MMP05444487	Marininema mesophilum	Non-effective
cAMP_106	GAKPGDGWISTITDDCPNSIIVCC	MARD_SAMN05526266_REFG_MMP05526266	Bacillus_AV solimangrovi	Non-effective
cAMP_107	GDGWISTITDDCPNSIIVCC	MARD_SAMN05526266_REFG_MMP05526266	Bacillus_AV solimangrovi	Non-effective
cAMP_108	KCVTGGLCE	MARD_SAMN05660649_REFG_MMP05660649	Desulfotruncus arcticus	Non-effective
cAMP_109	CGCPPGAGSVLCTPCPPLNCGA	MARD_SAMN06265352_REFG_MMP06265352	Eilatimonas milleporae	Non-effective
cAMP_110	SITGGAEATLYDYTCSEGCPTDTTNNNCTSGTSIVISCC	MARD_SAMN10146989_REFG_MMP10146989	Ulvibacterium marinum	Fail
cAMP_111	GDVSGEFTTSPACVYVSVAATRISSQKCAGAGSAVFSAISGAVASAVKC	OceanDNA-b17186	Paraclostridium benzoelyticum	Fail
cAMP_112	GDVEPNSLTAISAALTSAAATWAFSKDVVKCTKGNC	OceanDNA-b17186	Paraclostridium benzoelyticum	Fail
cAMP_113	GQVITASLVVPRHLTALCMPQTTLCCFLTSNQRAVVCVLLAARHRRWC	OceanDNA-b33543	Algicoccus	Non-effective
cAMP_114	WSGWWCTVSGECNGGVCCNPFGSDIKTLEQ	OceanDNA-b3938	Cyclobacteriaceae	Fail
cAMP_115	NPHPWPVAPSRAPLAPCP	OceanDNA-b43	Thermoanaerobaculia UBA5704	Non-effective
cAMP_116	RPSTPAKRHWVSKLLRCCWAARSAPSWPAGWGIAWGAAAY	OceanDNA-b44694	Stenotrophomonas maltophilia_AJ	Effective
cAMP_117	QMKNINGGLIEIPICFSSCPNVGTKKSKSCCGSKCTKELEQF	OceanDNA-b6831	Flavobacterium	Non-effective
cAMP_118	RVPCWSSHCKINVVSVNVVPSGNSKAGTRPVGLCGNTTALGSRSAVTVSLICADKSRSRRAILQMRA	OceanDNA-b33555	Algicoccus sp017857895	NA
cAMP_119	PEQRDVVPRRTARIRGQVSLQVGARGGGRPADTECCSGGTVTCRLPRWSSLFSCSEPLSDRALRNE	GOMC.bin.8625	Haliangiaceae	NA
cAMP_120	WDETELVELSEADVHGGTGPACVATGLILGITAIQAGVATAVSAAFCPTTKCSSRC	GCA_002224125.1_ASM222412v1_genomic	Streptomyces sp002224125	NA
cAMP_121	LVKGGTGNATLHSCPGQGNDPGGTCYTAGCGGGGTNGCNPSNPCSGTSVFITC	GOMC.bin.2081	Flavobacteriaceae	NA

* Note: “APD3” means the peptides are verified in APD3 database[34], “NA” means the peptides are longer than 50 aa and not subjected to synthesize, “Fail” means the peptides are synthesis failed, “Non-effective” means the peptides are non effective against multiple species, and “Effective” means the the peptides are non effective against at least one strain.

Supplementary Table 5. Amino acid sequences of reference *Is*PETase and recovered candidate dsPETases.

Enzyme	Accession No.	Amino acid sequence*
*Is*PETase	WP_054022242	MNFPRASRLMQAAVLGGLMAVSAAATAQTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYG AGTVYYPTNAGGTVGAIAIVPGYTARQSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQMA ALRQVASLNGTSSSPIYGKVDTARMGVMGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSS VTVPTLIFACENDSIAPVNSSALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMK RFMDNDTRYSTFACENPNSTRVSDFRTANCS
dsPETase01	SAMN07748057 Mariana Trench: 4,000m	MKIIRTLGAGIAAVALGFGLLAPSASVAAENDYERGPDPTSSSIEASRGPYAVSTKSISRFAARGFG GGTIHYPTTTADGTFGVVAVSPGYTASESTIRWLGPRLASFGFVVITFDTNSRYDQPRARGTQLLA AIDQAIGDSTVGSRIDPSRQAVVGHSMGGGGTLEAAKTRPSIEAAVGLTPWNLDKTWPEVEAAAL QIGAQNDSVAPPRSHAVPFYGSLTNAERRAYLELRGASHFAPNTSNTTIAKYTLAWLKRYVDDDT RYEQFLAPGPSTGFGSAVSDYRIQ
dsPETase02	SAMEA4473311 Pacific Ocean: North Su vent: 2,107m	MNLLRIKTLSSALIGLVILAGSFTTPSTVAQDNSDGRYRPEGRTFARNAANMFDDRVFEARTYETG TNQEAFASATIFYPLTLSFDPPNGAVIMVPGYRGTPPVYDWWGPMLASIGVITMIIETNTPEDSLEA RKNAFIAGVEFLRGENNNADSPIRDKLDTGNIAIMGHSLGGGASLRAAEELASQIKAVIPLTPYCCE LGQPFEGDLSGVSVPTLIIASAEDAIAPPGDHARMLYDSVNASTKKVYLEFATGNHMLPSNSGQD LETLGTYVVYAFIKENFTDNPRYTDFIFGDGEEQFSIYETNQ
dsPETase03	SAMN07631021 Kermadec Trench: 9,177m	MLLALLFAACSTSVTNRTGDPNISIASLSADGPYAVRAYTSFPEVGEFADATIYYPLDAETPVGGV AVSPGFTELQRAINWWGPRLASHGFAVLILDTNEPRDSPELRAEALISAVRLLKAENSRPDSPLNG RIDVNKMAIMGHSMGGGGTLIAANNYSDEIQAAIPFTPWQPEGDFSQITVPTLVIAGAADRIAAVS DHAWPHYQSLPHSTTRVYLEVAGGNHFLANSSGPDLSTIGRYGIAWLKLYLDGDERYRDFIYGEA QKVDTGKFSRYIANP
dsPETase04	SAMN02727538 Atlantic Ocean: Mid Cayman Rise hydrothermal vent: 4,900m	MANPYERGPNPTDALLEASSGPFSVSEENVSRLSASGFGGGTIYYPRENNTYGAVAISPGYTMNR WLFPRALLLLFSLLLASCANSPQPTVESLSGDGEYQVMTYTDFPDVPEFGDATIYYPLDTRGSIGG VAISPGYTERQSHIEWWGPLLASHGYAVLVLDTNDRRESTDLRADALIAAVTTLRAENTRNDSPL MGRIDGGKMAIMGHSMGGGGTLIAAHEHGEEIQAAIPFTPWEPDGPFDNITVPTLIIAGSIDRIAGV DEHAWRHFQSIPESTTKVYMEIDGGNHYIADTDRGTDLATVGRYGIAWLKLYLDGDERYRDFIY GEYHTPDMEKFSRYVTNP
dsPETase05	SAMEA4473313 Pacific Ocean: North Su vent: 2,107m	MKSGQHQKKDTVMKTPLFKLAALTLGVSLSSVALATNPGGGGGGSNPDTGTGFPGVSSFSADGS FATTSGSAGLSCTVFRPSTLGANGLKHPIIVWGNGTTASPSTYSGILEHWASHGFVVIAANTSNAG TGQDMLNCVDYLTTQNNRSTGTYANKLDLNRIGAAGHSQGGGGTIMAGQDYRIKVTAPFQPYTI GLGHNSSSQSNQNGPMFLMTGSADTIASPTLNALPVYNRANVPVFWGELSGASHFEPVGSAGDFR GPSTAWFRYHLMDDASAEDTFYGSNCDLCTDNDWDVRRKGIN
dsPETase06	SAMN07748060 Mariana Trench: 10,400m	MSVSLKGALRVLPLAASVALVGCFGGGDPEPGPDGQALTNPGEYEICSYETDLENSGYASARMT YPCDLSDGPYPATTLTGGFTNTKEQMEWLAEHLTTHGYVVLTMTPNNTLGVPPGWRDAQLGGF AELADENARSNSPLKGKIDLSKRNIMGFSMGGGGVILAAEEMGDAPTSAIALAPWLGAYNVDYS QIETPMLMLGSENDELAYYTEDYYAQLPADLERGVAIYAGASHFDWYGVNNQDQKAQFRTLVT AFLEVQLKGDTSAYSYFDGAEHDEHVQEGWFSAFDYQK

*The underlined sequences indicate the N-terminal signal peptides predicted by SignalP 5.0, which were truncated in the cloining process.

Supplementary Table 6. The taxonomy and keywords used to collect marine bacterial and archaeal genomes from public databases.

NCBI_ taxonomy	NCBI_taxid	EBI/JGI Keywords
algae_metagenome	1300146	ocean
ballast_water_metagenome	1954210	seawater
beach_sand_metagenome	412757	sea-water
brine_metagenome	1981201	sea
cetacean_metagenome	1822005	marine
ciliate_metagenome	1969832	bay
cold_seep_metagenome	1583376	coast
coral_metagenome	496922	deepsea
coral_reef_metagenome	471232	deep-sea
crab_metagenome	1660082	coral
crustacean_metagenome	1681198	shore
ctenophore_metagenome	1508044	seashore
desalination_cell_metagenome	1983455	microalgal
dinoflagellate_metagenome	1579005	algal
echinoderm_metagenome	1411990	hydrothermal
estuary_metagenome	1649191	estuary
eukaryotic_plankton_metagenome	2315767	seep
flotsam_metagenome	1602165	brine
gill_metagenome	1455666	trench
glacier_metagenome	1651087	mangrove
hydrothermal_vent_metagenome	652676	pelagic
hydrozoan_metagenome	1941281	Atlantic
invertebrate_gut_metagenome	1775950	Antarctic
invertebrate_metagenome	1711999	Arctic
jellyfish_metagenome	1549733	Pacific
lagoon_metagenome	1763544	phytoplankton
macroalgae_metagenome	2015907	Mariana
mangrove_metagenome	1284368	
marine_metagenome	408172	
marine_plankton_metagenome	1874687	
marine_sediment_metagenome	412755	
microbial_mat_metagenome	527640	

mollusc_metagenome	1417798	
oyster_metagenome	1541066	
periphyton_metagenome	1825055	
sand_metagenome	1671699	
sea_anemone_metagenome	1825923	
sea_squirt_metagenome	1041057	
sea_urchin_metagenome	1873886	
seagrass_metagenome	1904484	
seawater_metagenome	1561972	
sediment_metagenome	749907	
shrimp_gut_metagenome	1588881	
sponge_metagenome	1163772	
starfish_metagenome	2053188	
surface_metagenome	1774230	
tidal_flat_metagenome	1269027	
whale_fall_metagenome	412756	
zebrafish_metagenome	1331678	

Supplementary References

1 Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biology* **20**, 257, doi:10.1186/s13059-019-1891-0 (2019).

2 Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science* **3**, e104 (2017).

3 Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e1821, doi:10.1016/j.cell.2019.05.031 (2019).

4 Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* **37**, 38-44, doi:10.1038/nbt.4314 (2019).

5 McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* **8**, e61217 (2013).

6 R Core Team. R: A language and environment for statistical computing (2013).

7 Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379-423, doi:10.1002/j.1538-7305.1948.tb01338.x (1948).

8 Nielsen, F. On the Jensen-Shannon symmetrization of distances relying on abstract means. *Entropy (Basel, Switzerland)* **21**, doi:10.3390/e21050485 (2019).

9 pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4 (2020).

10 Hijmans, R. J., Williams, E., Vennes, C. & Hijmans, M. R. J. Package 'geosphere'. *Spherical trigonometry* **1** (2017).

11 Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927-930 (2003).

12 Zweng, M. *et al.* World ocean atlas 2018, volume 2: salinity. (2019).

13 Locarnini, M. *et al.* World ocean atlas 2018, volume 1: Temperature. (2018).

14 Boeuf, D. *et al.* Metapangenomics reveals depth-dependent shifts in metabolic potential for the ubiquitous marine bacterial SAR324 lineage. *Microbiome* **9**, 172, doi:10.1186/s40168-021-01119-5 (2021).

15 Cullen, J. J. Subsurface chlorophyll maximum layers: enduring enigma or mystery solved? *Ann Rev Mar Sci* **7**, 207-239, doi:10.1146/annurev-marine-010213-135111 (2015).

16 Richter, D. J. *et al.* Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *Elife* **11**, doi:10.7554/eLife.78129 (2022).

17 McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

18 Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359, doi:10.1126/science.1261359 (2015).

19 Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology* **3**, 804-813, doi:10.1038/s41564-018-0176-9 (2018).

20 Zeng, X., Alain, K. & Shao, Z. Microorganisms from deep-sea hydrothermal vents. *Marine Life Science & Technology* **3**, 204-230, doi:10.1007/s42995-020-00086-4 (2021).

21 Zhou, Y. L., Mara, P., Cui, G. J., Edgcomb, V. P. & Wang, Y. Microbiomes in the Challenger Deep slope and bottom-axis sediments. *Nat Commun* **13**, 1515, doi:10.1038/s41467-022-29144-4 (2022).

22 Cabello-Yeves, P. J. *et al.* The microbiome of the Black Sea water column analyzed by shotgun and genome centric metagenomics. *Environ Microbiome* **16**, 5, doi:10.1186/s40793-021-00374-1 (2021).

23 Orakov, A. *et al.* GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biol* **22**, 178, doi:10.1186/s13059-021-02393-0 (2021).

24 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315-5316, doi:10.1093/bioinformatics/btac672 (2022).

25 Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* **50**, D785-D794, doi:10.1093/nar/gkab776 (2022).
26 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157, doi:10.1186/s13059-015-0721-2 (2015).
27 Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530-1534, doi:10.1093/molbev/msaa015 (2020).
28 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-D419, doi:10.1093/nar/gkaa913 (2021).
29 Grafen, A. The phylogenetic regression. *Philos Trans R Soc Lond B Biol Sci* **326**, 119-157, doi:10.1098/rstb.1989.0106 (1989).
30 Ho, L. & Ane, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst Biol* **63**, 397-408, doi:10.1093/sysbio/syu005 (2014).
31 Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877-884, doi:10.1038/44766 (1999).
32 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300 (1995).
33 David, L. A. & Alm, E. J. Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* **469**, 93-96, doi:10.1038/nature09649 (2011).
34 Wang, G., Li, X. & Wang, Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res* **44**, D1087-1093, doi:10.1093/nar/gkv1278 (2016).