# Astonishing diversity and multifaceted biological connections of Type IV restriction-modification systems

Ryan T. Bell[1,*], Jonas Juozapaitis[2], Inga Songailiene[2], Harutyun Sahakyan[1], Toma Liegute[2], Kira S. Makarova[1], Yuri I. Wolf[1], Virginijus Siksnys[2], Eugene V. Koonin[1,*]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

[2]Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania

*Correspondence to: Ryan T. Bell (ryan.bell@nih.gov) or Eugene V. Koonin (koonin@ncbi.nlm.nih.gov)

## Abstract

A comprehensive census of McrBC systems, a major form of prokaryotic Type IV restriction-modification (RM), followed by phylogenetic analysis, reveals their enormous abundance in diverse bacteria and archaea and a plethora of genomic associations. We focus on a previously uncharacterized branch, which we denote CoCoNuTs (coiled-coil nuclease tandems) for their salient features, the presence of extensive coiled-coil structures and tandem nucleases. The CoCoNuTs alone show extraordinary variety, with 3 distinct types and multiple subtypes. All CoCoNuTs contain domains predicted to interact with translation system components, such as OB-folds resembling the SmpB protein that binds bacterial tmRNA, YTH-like domains that might recognize methylated tmRNA, tRNA, or rRNA, and RNA-binding Hsp70 chaperone homologs, along with RNases, such as HEPN domains, all suggesting that the CoCoNuTs target RNA. Many CoCoNuTs may also target DNA, via McrC nuclease homologs. Additional restriction systems, such as Type I RM and Druantia Type III, are frequently encoded in the same predicted superoperons. In many of these superoperons, CoCoNuTs are likely regulated by cyclic nucleotides, possibly, RNA fragments with cyclic termini, that bind associated CARF (CRISPR-Associated Rossmann Fold) domains. We experimentally demonstrate the ability of Type II CoCoNuTs to protect *E. coli* against specific phages and plasmids.

## Introduction

All organisms are subject to an incessant barrage of genetic parasites, such as viruses and transposons. Over billions of years, the continuous arms race between hosts and parasites drove the evolution of immense, intricately interconnected networks of diverse defense systems and pathways [1-6]. In particular, in the last few years, targeted searches for defense systems in prokaryotes, typically capitalizing on the presence of variable genomic defense islands, have dramatically expanded their known diversity and led to the discovery of a plethora of biological conflict strategies and mechanisms [2, 4, 7, 8].

One of the most ancient and common forms of defense against mobile genetic elements (MGE) is the targeted restriction of nucleic acids. Since the initial discovery of this activity among strains of bacteria resistant to certain viruses, myriad forms of recognition and degradation of nucleic acids have been described, in virtually all life forms. Characterization of the most prominent of these systems, such as restriction-modification (RM), RNA interference (RNAi), and particularly CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated genes), has led to the development of a profusion of highly effective experimental and therapeutic techniques, in particular, genome editing and engineering [9-15].

Historically, the two-component McrBC (modified cytosine restriction) system was the first form of restriction to be described, although the mechanism remained obscure for decades, and for a time, this system was referred to as RglB (restriction of glucoseless phages) due to its ability to restrict T-even phage DNA which contained hydroxymethylcytosine, but not glucosylated bases [16-19]. Today, the prototypical McrBC system, native to *E. coil* K-12, is considered a Type IV (modification-dependent) restriction system that degrades DNA containing methylcytosine (5mC) or hydroxymethylcytosine (5hmC), with a degree of sequence context specificity [20, 21].

Type IV restriction enzymes contain at least two components: 1) a dedicated specificity domain that recognizes modified DNA, and 2) an endonuclease domain that cleaves the target [9, 22]. In the well-characterized example from *E. coli* K-12, McrB harbors an N-terminal DUF(domain of unknown function)3578 that recognizes methylcytosine. We denote this domain, as its function is not unknown, as ADAM (a domain with an affinity for methylcytosine). The ADAM domain is fused to a GTPase domain of the AAA+ ATPase superfamily, the only known GTPase in this clade, which is believed to translocate DNA (Figure 1 A, B, C) [20, 23-25]. McrC, encoded by a separate gene in the same operon, contains an N-terminal DUF2357 domain, which interacts with the GTPase domain in McrB and stimulates its activity, and is fused to a PD-(D/E)xK superfamily endonuclease (Figure 1 A, B, D) [20, 26].

Our recent analysis of the modified base-binding EVE (named for Protein Data Bank (PDB) structural identifier 2eve) domain superfamily demonstrated how the distribution of the EVE-like domains connects the elaborate eukaryotic RNA regulation and RNA interference-related epigenetic silencing pathways to largely uncharacterized prokaryotic antiphage restriction systems [4]. EVE superfamily domains, which in eukaryotes recognize modified DNA or RNA as part of mRNA maturation or epigenetic silencing functions, are often fused to McrB-like GTPases in prokaryotes, and indeed, these are the most frequently occurring EVE-containing fusion proteins [4]. These observations motivated us to conduct a comprehensive computational search for McrBC systems, followed by a census of all associated domains, to chart the vast and diverse population of antiviral specificity modules, vital for

prokaryotic defense, that also provided important source material during the evolution of central signature features of eukaryotic cells. Here we present the results of this census and describe an extraordinary, not previously appreciated variety of domain architectures of the McrBC family of Type IV restriction systems. In particular, we focus on a major McrBC branch that we denote coiled-coil nuclease tandem (CoCoNuT) systems, which we explore in detail and report proof of principle experiments demonstrating the anti-MGE defense mediated by the CoCoNuTs.

## Materials and Methods

*Comprehensive identification and phylogenetic and genomic neighborhood analysis of McrB and McrC proteins*

The comprehensive search for McrB and McrC proteins was seeded with publicly available multiple sequence alignments COG1401 (McrB GTPase domain), AAA_5 (The branch of AAA+ ATPases containing the McrB GTPase), COG4268 (McrC), PF10117 (McrBC), COG1700 (McrC PD-D/ExK nuclease domain), PF04411 (McrC PD-D/ExK nuclease domain), PF09823 (McrC N-terminal DUF2357). Additional alignments and individual queries were derived from data from our previous work on the EVE domain family [4]. All alignments were clustered, and each sub-alignment or individual query sequence was used to produce a position-specific scoring matrix (PSSM). These PSSMs were used as PSI-BLAST queries against the non-redundant (nr) NCBI database (*E*-value ≤ 10) [27]. Genome neighborhoods for the hits were generated by downloading their gene sequence, coordinates, and directional information from GenBank, as well as for 10 genes on each side of the hit. Domains in these genes were identified using PSI-BLAST against alignments of domains in the NCBI Conserved Domain Database (CDD) and Pfam (*E*-value 0.001). Some genes were additionally analyzed with HHpred [28]. Then, these neighborhoods were filtered for the presence of a COG1401 hit (McrB GTPase domain), or hits to both an McrB "alias" (MoxR, AAA_5, COG4127, DUF4357, Smc, WEMBL, Myosin_tail_1, DUF3578, EVE, Mrr_N, pfam01878) and an McrC "alias" (McrBC, McrC, PF09823, DUF2357, COG1700, PDDEXK_7, RE_LlaJI). These aliases were determined from a preliminary manual investigation of the data using HHpred [29]. Several of the McrB aliases are not specific to McrB, and instead are domains commonly fused to McrB GTPase homologs, or are larger families, such as AAA_5, that contain McrB homologs. We found that many *bona fide* McrB homologs, validated by HHpred, produced hits not to COG1401 but rather to these other domains, so we made an effort to retain them.

The McrB candidates identified by this filtering process were clustered to a similarity threshold of 0.5 with MMseqs2 [30], after which the sequences in each cluster were aligned with MUSCLE [31]. Next, profile-to-profile similarity scores between all clusters were calculated with HHsearch [32]. Clusters with high similarity, defined as a pairwise score to self-score ratio > 0.1, were aligned to each other with HHalign [33]. This procedure was performed for a total of 3 iterations. The alignments resulting from this protocol were analyzed with HHpred to remove false positives, after which the GTPase domain sequences were extracted manually using HHpred, and the alignments were used as queries for a second round of PSI-BLAST against the nr NCBI database as described above. At this stage, the abundance of the CoCoNuT and CoCoPALM (coiled-coil and pilus assembly linked to McrBC, see below) types of McrB GTPases had become apparent, therefore, results of targeted searches for these subtypes were included in the pool of hits from the second round of PSI-BLAST. Genome neighborhoods were generated for these hits and filtered for aliases as described above. Further filtering of the data, which did not pass this initial filter, involved relaxing the criteria to include neighborhoods with only one hit to an McrB or McrC alias, but with a gene adjacent to the hit (within 90 nucleotides), oriented in the

114 same direction as the hit, and encoding a protein of sufficient size (> 200 aa for McrB, > 150 aa for McrC)
115 to be the undetected McrB or McrC component. Afterward, we filtered the remaining data to retain
116 genome islands with no McrBC aliases but with PSI-BLAST hits in operonic association with genes of
117 sufficient size, as described above, to be the other McrBC component. These data, which contained
118 many false positives but captured many rare variants, were then clustered and analyzed as described
119 above to remove false positives. Next, an automated procedure was developed to excise the GTPase
120 domain sequences using the manual alignments generated during the first phase of the search as a
121 reference. These GTPase sequences were further analyzed by clustering and HHpred to remove false
122 positives. Definitive validation by pairing McrB homologs with their respective McrC homologs was also
123 used to corroborate their identification. Occasionally, the McrB and McrC homologs were separated by
124 intervening genes, or the operon order was reversed, and consideration of those possibilities allowed
125 the validation of many additional systems. The pairing process was complicated by and drew our
126 attention to the frequent occurrence of multiple copies of McrBC systems in the same islands that may
127 function cooperatively. Finally, the rigorously validated set of McrBC pairs, supplemented only with
128 orphans manually annotated as McrBC components using HHpred, were used for our phylogenetics. The
129 final alignments of GTPase and DUF2357 domains were produced using the iterative alignment
130 procedure described above for 10 iterations. Approximately-maximum-likelihood trees were built with
131 the FastTree program [34] from representative sequences following clustering to a 0.9 similarity
132 threshold with MMseqs2. The phyletic distribution was calculated by assigning a weight to the lineage of
133 each sequence that was inversely proportional to its cluster size (1/n, where n=the cluster size).
134
135 *Preliminary phylogenetic analysis of CoCoNuT CnuH helicases*
136
137 A comprehensive search and phylogenetic analysis of this family was beyond the scope of this
138 work, but to determine the relationships between the CoCoNuTs and the rest of the UPF1-like helicases,
139 we used the following procedure. We retrieved the best 2000 hits in each of two searches with UPF1
140 and CoCoNuT helicases as queries against both a database containing predicted proteins from 24,757
141 completely sequenced prokaryotic genomes downloaded from the NCBI GenBank in November 2021
142 and a database containing 72 representative eukaryotic genomes that were downloaded from the NCBI
143 GenBank in June 2020. Next, we combined all proteins in four searches, made a nonredundant set, and
144 annotated them using CDD profiles, as described above. Then, we aligned them with MUSCLE v5 [35],
145 constructed an approximately-maximum-likelihood tree with FastTree, and mapped the annotations
146 onto the tree. Genome neighborhoods were generated for these hits, as described above.

147 *Structural modeling with AlphaFold2 and searches for related folds*
148
149 We predicted the structures described in this work using AF2 v2.2.0 with local installations of
150 complete databases required for AF2 [36]. We used only protein models with average predicted local
151 distance difference test (pLDDT) scores $\geq$ 80 for further analysis [37]. We used many of these models as
152 queries to search for structurally similar proteins using DALI v5 against the Protein Data Bank (PDB) and
153 using FoldSeek against the AlphaFold/UniProt50 v4, AlphaFold/Swiss-Prot v4, AlphaFold/Proteome v4,
154 and PDB100 2201222 databases [38, 39]. Structure visualizations and comparisons were performed with
155 ChimeraX [40] and the RCSB PDB website [41].

156 *Gene cloning*

157 Synthetic gene fragments coding CoCoNuT gene operons with 100 nt upstream regions were
158 ordered commercially from "Twist Bioscience." Gene fragments were designed with ~30 nucleotide

159 complementary ends. pACYC184 vector was amplified by PCR, generating ~30 nucleotide ends
160 complementary to the operon ends. Fragments were assembled using Gibson assembly and
161 transformed into *E. coli* DH5α. The Type II CoCoNuT TerY-PP2C-STK only variant was constructed by
162 whole plasmid PCR retaining only the gene triad of interest, followed by blunt-end ligation and
163 transformation. Plasmids were purified by silica spin columns and sequenced.

164 *Phage plaque assay*

165 Bacteriophages were amplified in the *E. coli* MG1655 strain. 0.2-1 ml of *E. coli* night cultures
166 (depending on culture density) were diluted with Luria-Bertani (LB) medium to 2 ml volume and mixed
167 with 2 ml of hot agar-LB (1.5% agar, 60 °C). 4 ml of soft agar with bacteria was immediately spread on
168 LB-agar Petri dishes supplemented with 35 µg/ml chloramphenicol and air-dried for 20 min. 3.5 µl of 10-
169 fold phage dilutions were plated on soft agar, air-dried for 20 min, and incubated for 16 h at 30-37° C,
170 depending on the phage used.

171 *Phage challenge in liquid medium*

172 100 µl of *E. coli* night cultures grown with 1 % glucose were added to 4 ml of fresh LB medium
173 and incubated at 37 °C in a shaker for 5 h. Bacteria were diluted to $OD_{600}$ 0.5 and kept on ice while
174 preparing a 96-well flat-bottom plate. To limit condensation, the inside of the 96-well plate lid was
175 impregnated with 20% ethanol 0.05% Triton X-100 solution and dried on a paper towel. In the plate, 70
176 µl of LB medium and 20 µl of diluted bacteria were mixed. Just before measuring, 10 µl of phage or LB
177 medium for control was added to 90 µl of diluted cells. The final $OD_{600}$ was 0.1, and phage was added at
178 MOI 0.05. The dried lid was replaced, sides of the plate were sealed with parafilm tape and immediately
179 transferred to the plate reader. Cultures were incubated at 37°C at 200 rpm and $OD_{600}$ was measured
180 every 5 min. For all dilutions and cultures, the same stock of LB medium supplemented with 35 µg/ml
181 chloramphenicol was used. For phage λ-vir, 5 mM of $CaCl_2$ was added to soft agar.

182 *CFU measurement after phage infection*

183 100 µl of *E. coli* night cultures grown with 1% glucose were added to 4 ml of fresh LB medium
184 supplemented with 35 µg/ml chloramphenicol and incubated at 37 °C in a shaker for 5 h. Bacteria were
185 diluted to $OD_{600}$ 0.1, supplemented with 5 mM of $CaCl_2$, phage λ-vir added at MOI 0.05, and incubated at
186 37°C at 200 rpm. After 4 hours of incubation, an equal volume of cultures was transferred to a 1.5 ml
187 tube and centrifuged for 5 minutes at 12000 rpm on a table centrifuge. The supernatant containing
188 phages was removed by pipette and bacteria were resuspended in the same volume of LB medium
189 without $CaCl_2$. Serial 10-fold dilutions were performed and 3.5 µl of diluted cultures were plated on LB
190 agar, air-dried for 20 minutes, and incubated for 16 h at 37° C. CFU/ml was determined by counting the
191 colonies at the lowest visible dilution, taking into account the dilution and 3.5 µl volume of the plated
192 drop. All LB media were supplemented with chloramphenicol.

193 *Plasmid interference assay*

194 *E. coli* MG1655 strain cells were pre-transformed with pACYC184 plasmids (chloramphenicol
195 resistance, Cm) harboring each wild-type CoCoNuT defense system variant or control empty pACYC184
196 vector. CoCoNuT cells were cultivated in LB medium overnight supplemented with 2% glucose and
197 chloramphenicol 30 µg/ml at 37°C, 200 rpm. The next day, flasks with 50 ml of liquid LB medium and 30
198 µl/ml chloramphenicol were inoculated with cells from overnight culture at 37°C, 200 rpm until they

199 reached $OD_{600}$~ 0.6. The same number of cells for each variant (based on OD) were taken and
200 centrifuged for 10 min, 3000 rpm, next washed using 10 ml deionized water 4 times, resuspended in
201 1500 µl of deionized water, and used for electroporation using Thermo Scientific™ Electroporation
202 Cuvettes, 1mm. 100 ng of different potential target plasmids bearing distinct *ori* sequences
203 (Supplementary Table S3) were electroporated into the prepared cells with or without CoCoNuT. Cells
204 were immediately transferred to Eppendorfs with fresh LB medium and left for 60 min at 37°C 200 rpm
205 for recovery. Next, the cells were serially diluted in a 96-well plate with LB medium and 5 µl were
206 spotted on a selection medium with chloramphenicol and a selection medium containing additional
207 antibiotic for the selection of co-electroporated plasmid. Plates were incubated at 37°C for 16h.

208

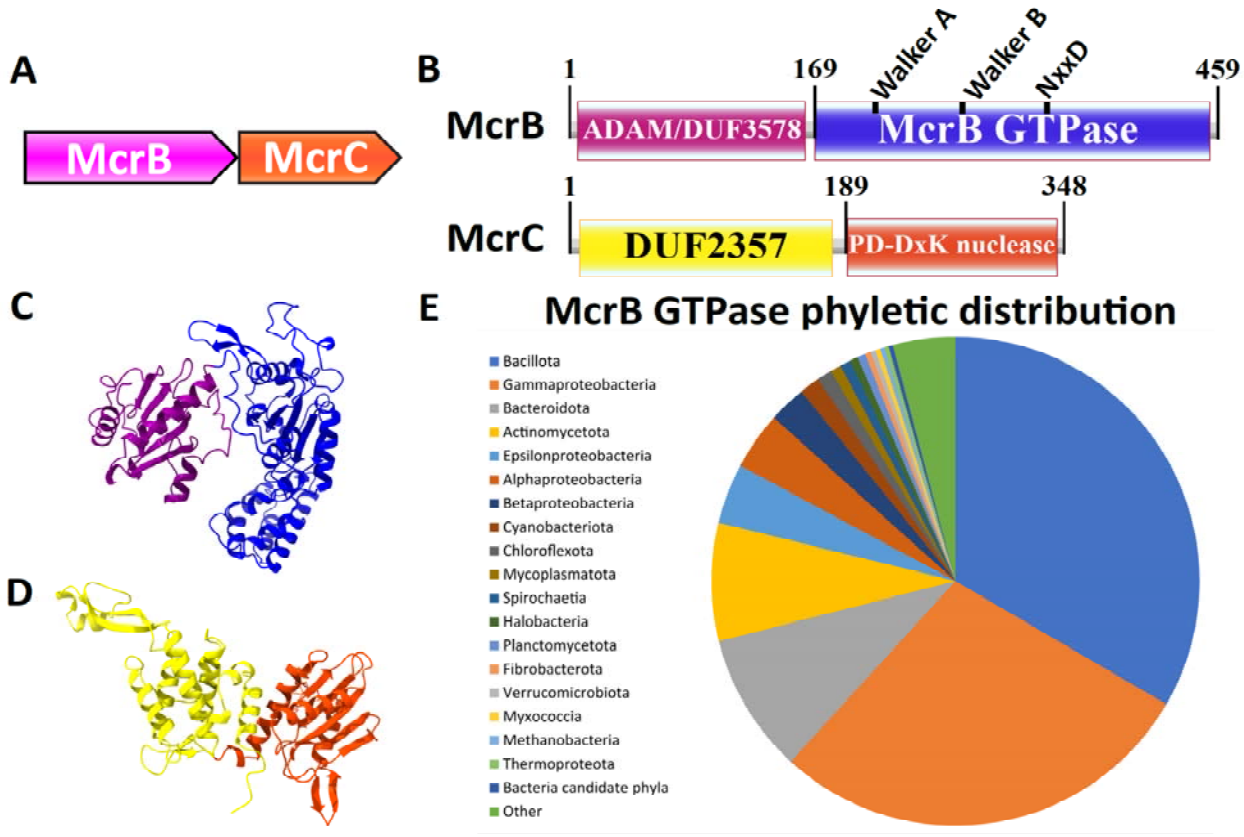
## 209 Results and Discussion

210 *Comprehensive census of McrBC systems*

211 Our initial search for McrBC systems consisted of PSI-BLAST runs against the non-redundant
212 protein sequence database at the NCBI using as queries position-specific scoring matrices (PSSMs)
213 derived from publicly available alignments of McrB and McrC homologs, or our EVE-McrB data [4].
214 Although a branch of McrB GTPase homologs has been described in animals, these are highly divergent
215 in function, and no associated McrC homologs have been reported [23]. Therefore, we excluded
216 eukaryotic sequences from our analysis to focus on the composition and contextual connections of
217 prokaryotic McrBC systems. Through this preliminary search, we obtained a set of approximately 12,500
218 McrBC homologs, which was validated using HHpred and several other criteria, including the presence of
219 the McrB GTPase signature motif NxxD (where x is any amino acid), which distinguishes McrB from
220 closely related AAA+ ATPase families such as MoxR (Figure 1 B) [23, 26, 42, 43]. Using alignments of the
221 McrB GTPase domains from this set, we conducted a second iteration of searches and validation,
222 increasing the number of McrBC systems to nearly 34,000, distributed broadly among prokaryotes
223 (Figure 1 E).

224

225




226

227 **Figure 1: Genetic organization, signature sequence motifs, structural models, and phyletic**
228 **distribution of McrB GTPases detected in this work**

229 A) McrBC is a two-component restriction system with each component typically (except for extremely rare gene fusions)
230 encoded by a separate gene expressed as a single operon, depicted here and in subsequent figures as arrows pointing in the
231 direction of transcription. In most cases, McrB is the upstream gene in the operon. B) In the prototypical *E.coli* K-12 McrBC
232 system, McrB contains an N-terminal methylcytosine-binding domain, ADAM/DUF3578, fused to a GTPase of the AAA+ ATPase
233 clade [21]. This GTPase contains the Walker A and Walker B motifs that are conserved in P-loop NTPases as well as a signature
234 NxxD motif, all of which are required for GTP hydrolysis [24, 26, 44]. McrC consists of a PD-DxK nuclease and an N-terminal
235 DUF2357 domain with a stalk-like extension comprising a helical bundle that interacts with and activates individual McrB
236 GTPases while they are assembled into hexamers [24, 26]. C) AlphaFold2 structural model of *E. coli* K-12 ADAM-McrB GTPase
237 monomer. D) AlphaFold2 structural model of *E. coli* K-12 McrC monomer with DUF2357-PD-D/ExK architecture. The structures
238 were visualized with ChimeraX [40]. E) Phyletic distribution of McrB GTPase homologs detected in this work, each found in a
239 genomic island with distinct domain composition, clustered to a similarity threshold of 0.9, then assigned a weight inversely
240 proportional to the cluster size (see Methods).

241 In the subsequent phase of analysis, we extracted GTPase domain sequences from the McrB
242 homolog pool, leaving as a remainder the fused specificity domains that we intended to classify.
243 Unexpectedly, we found that about 34% of the *mcrB* genes contained insertions into the GTPase coding
244 sequence. These insertions likely encode specificity domains, both known and uncharacterized, typically
245 located between the Walker A (P-loop) and Walker B motifs of the GTPase [23, 24, 26]. The previously
246 identified specificity domains are usually fused at the N terminus, or rarely, at the C-terminus of McrB-
247 like GTPases [21, 45].

We also observed that, although the McrC homologs are not the primary bearers of specificity modules in McrBC systems, they can be fused to various additional domains, including those of the EVE superfamily (Figure 2, Supplementary Figure S1). Many McrC homologs lack the typical PD-D/ExK nuclease domain or contain an apparently inactivated version of this domain that lacks the signature catalytic site, but the N-terminal DUF2357 domain, which interacts with McrB, is nearly always present. To isolate the likely specificity domains, we first attempted to extract the two common domain sequences from the pool of McrC homologs. Much like in the case of McrB GTPases, we identified insertions in many DUF2357 sequences. Most of these insertions consisted of predicted coiled-coils, but some contained beta-sheets predicted to fold into distinct domains. We defined the conserved regions with the maximum precision possible using HHpred, coiled-coil prediction tools, and AlphaFold2 (AF2) structural predictions [28, 36, 46-49].

Removing the variable inserts from the conserved McrBC domains allowed accurate, comprehensive phylogenetic analysis of the McrB GTPase (Figure 2) and McrC DUF2357 (Supplementary Figure S1) families. These two trees were generally topologically concordant and revealed several distinct branches not previously recognized. The branch containing the prototypical McrBC system from *E. coil* K-12 is characterized by frequent genomic association with Type I RM systems, usually with unidirectional gene orientations and the potential of forming a single operon [50]. This branch and others, which exhibit two particularly prevalent associations, with a DISARM-like antiphage system [51], and, surprisingly for a Type IV restriction system, with predicted DNA methyltransferases, will be the subject of a separate, forthcoming publication.

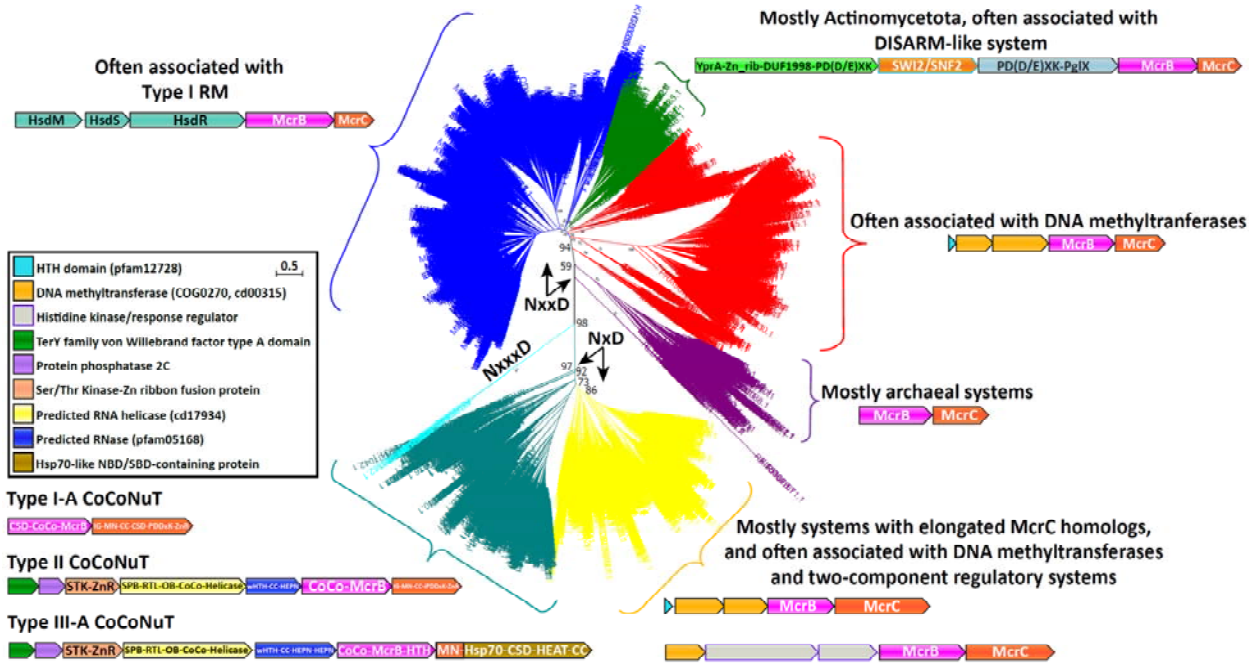



**Figure 2: Phylogenetic tree of the McrB GTPases**

The major clades in the phylogenetic tree of the McrB GTPases are distinguished by the distinct versions of the Nx(xx)D signature motif. The teal and yellow groups, with bootstrap support of 97%, have an NxD motif, whereas the blue, green, red, and purple groups, with variable bootstrap support, have an NxxD motif, indicated by the arrows; the sequences in the smaller, cyan clade, with 98% bootstrap support, have an NxxxD motif. Each of the differently colored groups is characterized by distinct conserved genomic associations that are abundant within but not completely confined to the respective groups. This tree was

built from the representatives of 90% identity clusters of all validated homologs. Abbreviations of domains: McrB – McrB GTPase domain; CoCo/CC – coiled-coil; MN – McrC N-terminal domain (DUF2357); CSD – cold shock domain; IG – Immunoglobulin (IG)-like beta-sandwich domain; ZnR – zinc ribbon domain; SPB – SmpB-like domain; RTL – RNase toxin-like domain; HEPN – HEPN family nuclease domain; OB – OB-fold domain; iPD-D/ExK – inactivated PD-D/ExK fold; Hsp70-like ATPase – Hsp70-like NBD/SBD; HEAT – HEAT-like helical repeats; YprA – YprA-like helicase domain; DUF1998 - DUF1998 is often found in or associated with helicases and contains four conserved, putatively metal ion-binding cysteine residues; SWI2/SNF2 – SWI2/SNF2-family ATPase; PglX – PglX-like DNA methyltransferase; HsdR/M/S – Type I RM system restriction, methylation, and specificity factors.

*A variant of the McrB GTPase signature motif distinguishes a large group of unusual McrBC systems*

A major branch of the McrBC family is characterized by a conserved deletion within the NxxD GTPase signature motif, reducing it to NxD (Figure 1 B, Figure 2). No NxD motif McrB GTPase has been characterized, but the extensive study of the NxxD motif offers clues to the potential impact of the motif shortening. The asparagine (N) residue is strictly required for GTP binding and hydrolysis [44]. As this residue is analogous to sensor-1 in ATP-hydrolyzing members of the AAA+ family, it can be predicted to position a catalytic water molecule for nucleophilic attack on the γ-phosphate of an NTP [26, 43, 44, 52, 53]. The role of the conserved aspartate (D), however, was initially unclear. Given the specificity of McrB for GTP, the motif was proposed to function similarly to the NKxD G4 box in conventional GTPases and confer specificity for GTP by interacting with the guanine ring [44]. However, a D336N mutation in the *E. coli* K-12 McrB GTPase NxxD motif did not alter binding specificity, unlike the corresponding mutation in EF-Tu and several other GTPases containing the G4 box [44]. Instead, this substitution unexpectedly inhibited GTP hydrolysis and prevented the stimulation of GTPase activity by McrC [44]. A recent structural analysis has shown that this aspartate interacts with a conserved arginine/lysine residue in McrC which, via a hydrogen-bonding network, resituates the NxxD motif in relation to its interface with the Walker B motif such that, together, they optimally position a catalytic water to stimulate hydrolysis [26]. Accordingly, the truncation of this motif might be expected to modulate the rate of hydrolysis, and potentially compel functional association with only a subset of McrC homologs containing compensatory mutations. Indeed, the NxD branch is characterized by many McrC homologs with unusual features, such as predicted RNA-binding domains, that might not be compatible with conventional McrB GTPases from the NxxD clade, which often occur in the same genomes (see below).

Many of these NxD GTPases are contextually associated with DNA methyltransferases, like the NxxD GTPases, but are distinguished by additional complexity in the domains fused to the McrC homologs and the frequent presence of two-component regulatory system genes in the same operon (Figure 2). We also detected a small number of GTPases with an insertion in the signature motif (Figure 2), expanding it to NxxxD, which are methyltransferase-associated as well. This association is likely to be ancestral, as it is found in all 3 branches of the McrB GTPase tree with different signature motif variations.

The most notable feature of the NxD branch of GTPases is a large clade characterized by fusion to long coiled-coil domains, a derived and distinctly different architecture from the small, modular DNA-binding specificity domains typically found in canonical McrB homologs (Figure 2). The McrC homologs associated with these coiled-coil McrB GTPase fusions also often lack the PD-D/ExK endonuclease entirely or contain an AF2-predicted endonuclease fold with inactivating replacements of catalytic residues (Supplementary Table S1). However, in the cases where the McrC nuclease is missing or likely

318 inactive, these systems are always encoded in close association, often with overlapping reading frames,
319 with HEPN (higher eukaryotes and prokaryotes nucleotide-binding) ribonuclease domains, or other
320 predicted nucleases (Figure 3, Supplementary Figure S2) [54]. Based on these features of domain
321 architecture and genomic context, we denote these uncharacterized McrBC-containing operons coiled-
322 coil nuclease tandem (CoCoNuT) systems.

323 The deepest branching group of the NxD GTPase systems is typified by a fusion of an Hsp70-like
324 nucleotide-binding domain and protein substrate-binding domain (NBD/SBD) to the McrB GTPase
325 domain (Figure 2, Supplementary Figure S2, Figure 3, Supplementary Table S1). Hsp70 is an ATP-
326 dependent protein chaperone that binds exposed hydrophobic peptides and facilitates protein folding
327 [55]. It also associates with AU-rich mRNA, an interaction involving both the ATPase and SBD [56-58].
328 These systems might be functionally related to the CoCoNuTs, given that an analogous unit is encoded
329 by a type of CoCoNuT system where similar domains are fused to the McrC homolog rather than to the
330 McrB GTPase homolog (see below). In another large clade of CoCoNuT-like systems, the GTPase is fused
331 to a domain homologous to FtsB, an essential bacterial cell division protein containing transmembrane
332 and coiled-coil helices (Supplementary Table S1) [59]. These enzymes form the core of rapidly evolving
333 protein secretion systems in Bacillota and Mycoplasmatota, likely regulated by EAL and HD-GYP domains
334 that bind cyclic di-GMP, and are presumably implicated in defense (Supplementary Figure S2) [60].
335 These unusual McrBC systems were primarily detected in metagenomic data obtained from the gut
336 microbiomes of ruminants and humans and might be involved in host-microbe interactions. They are
337 associated with signal peptidase family proteins likely to function as pilus assembly factors [47] and
338 usually contain coiled-coil domains fused to both the McrB and McrC homologs. Therefore, we denote
339 them as coiled-coil and pilus assembly linked to McrBC (CoCoPALM) systems (Supplementary Figure S2).
340 Here, we focus on the CoCoNuTs, whereas the CoCoPALMs and the rest of the NxD GTPase
341 methyltransferase-associated homologs will be explored in a separate, forthcoming publication.

342

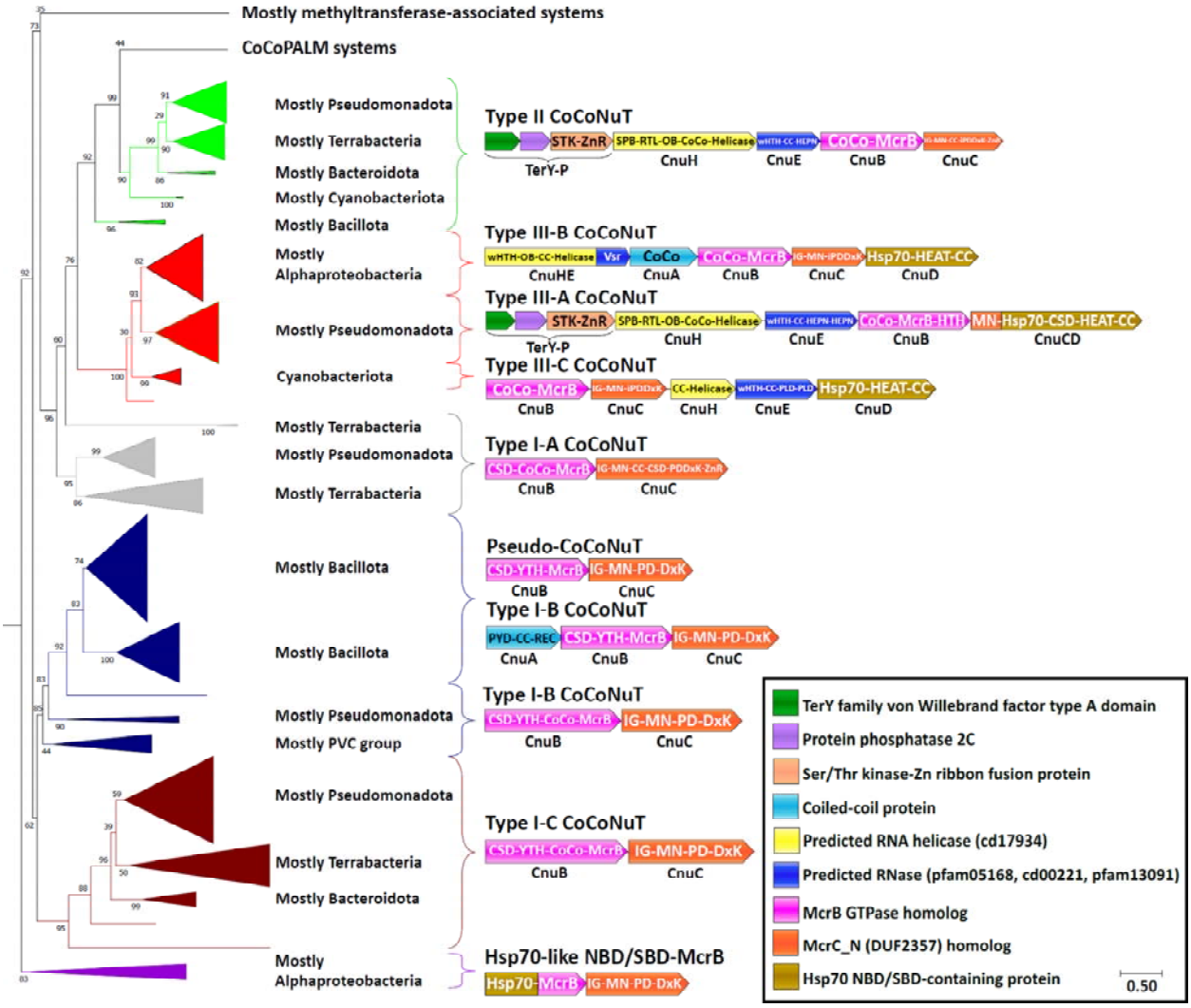
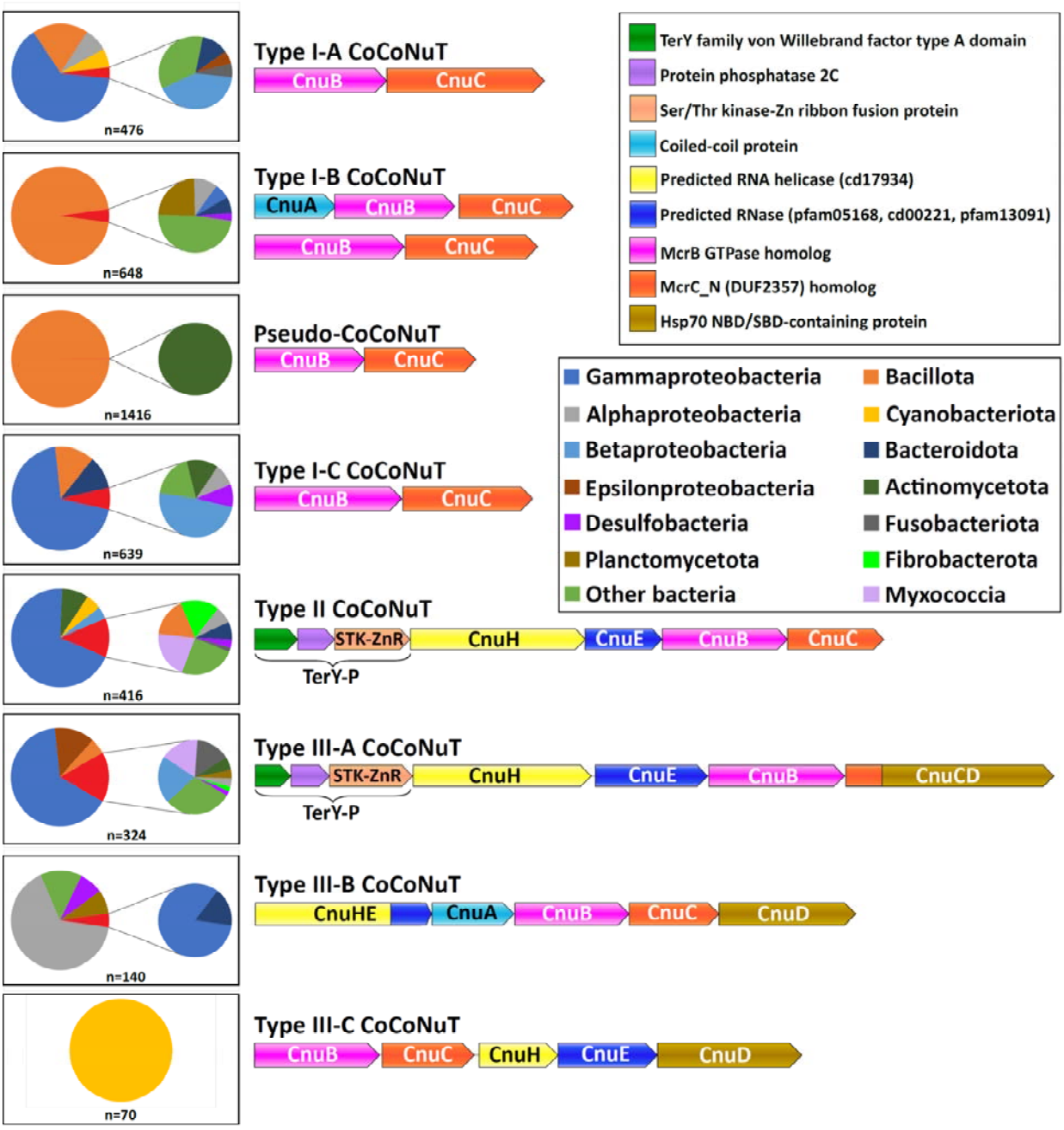



**Figure 3: CoCoNuT system phylogeny and classification**

The figure shows the detailed phylogeny of McrB GTPases from CoCoNuT systems and their close relatives. All these GTPases possess an NxD GTPase motif rather than NxxD. CoCoNuT systems are abundant in Pseudomonadota and Terrabacteria, with the various subtypes generally concentrated in a more restricted range of species, as indicated. This tree was built from the representatives of 90% clustering of all validated homologs. Abbreviations of domains: McrB – McrB GTPase domain; CoCo/CC – coiled-coil; MN – McrC N-terminal domain (DUF2357); CSD – cold shock domain; YTH – YTH-like domain; IG – Immunoglobulin (IG)-like beta-sandwich domain; Hsp70 – Hsp70-like NBD/SBD; HEAT – HEAT-like helical repeats; ZnR – zinc ribbon domain; PYD – pyrin/CARD-like domain; SPB – SmpB-like domain; RTL – RNase toxin-like domain; HEPN – HEPN family nuclease domain; OB – OB-fold domain; iPD-D/ExK – inactivated PD-D/ExK fold; REC – Phosphoacceptor receiver-like domain; PLD – Phospholipase D-like nuclease domain; Vsr – very-short-patch-repair PD-D/ExK nuclease-like domain. Underneath each gene is a proposed protein name, with Cnu as an abbreviation for CoCoNuT.

CoCoNuT systems are extremely diverse and represented in a wide variety of bacteria, particularly Pseudomonadota and Bacillota, but are nearly absent in archaea (Figure 4).

358




359

360 **Figure 4: Phyletic distribution of CoCoNuTs**

361 The phyletic distribution of CnuB/McrB GTPases in CoCoNuT systems found in genomic islands with distinct domain
362 compositions. Most CoCoNuTs are found in either Bacillota or Pseudomonadota, with particular abundance in
363 Gammaproteobacteria. Type I-B and the related Pseudo-CoCoNuTs are restricted mainly to Bacillota. In contrast, the other
364 types are more common in Pseudomonadota, but can be found in a wide variety of bacteria. Type III-B and III-C are primarily
365 found in Alphaproteobacteria and Cyanobacteriota, respectively.

366 *Type I CoCoNuT systems*

367 We classified the CoCoNuTs into 3 types and 7 subtypes based on the GTPase domain phylogeny
368 and conserved genomic context (Supplementary Figure S2, Figure 3, Figure 4). Type I-A and Type I-C
369 systems consist of McrB and McrC homologs only (Figure 5), which we denote CnuB and CnuC (Figure 3,
370 Figure 4). Type I-B systems usually encode a separate coiled-coil protein, which we denote CnuA, in
371 addition to CnuB/McrB and CnuC/McrC, with no coiled-coil fused to the McrB GTPase domain in CnuB
372 (Figure 3, Figure 4, Supplementary Figure S2, Figure 5). Deeper branching Type I-B groups lack CnuA and
373 more closely resemble Type I-C in architecture, but the GTPase domains and N-terminal
374 oligonucleotide/oligosaccharide-binding (OB)-folds are more similar to the rest of Type I-B. CnuA is fused
375 at the N-terminus to a pyrin (PYD)/CARD-like helical domain and at the C-terminus to a REC domain
376 (Figure 5, Supplementary Table S1). Notably, in animals, the prototypical PYD domain of the PYRIN
377 protein is also fused to a long coiled-coil domain [61]. PYD and CARD (caspase activation and
378 recruitment domain) folds belong to the DEATH domain superfamily and are best characterized in the
379 context of innate immunity, inflammasome formation, and programmed cell death (PCD), where these
380 domains mediate the formation of multisubunit complexes via homotypic interactions [62]. An extensive
381 network of (putative) prokaryotic defense systems containing DEATH superfamily domains has recently
382 been discovered [8, 63]. The association of the PYD domain with CoCoNuTs suggests involvement in a
383 PCD/abortive infection-type response. The REC domains constitute one of the components of two-
384 component regulatory systems. They are targeted for phosphorylation by histidine kinases [64], which
385 could be a mechanism of Type I-B CoCoNuT regulation.

386 In contrast to Type II and Type III CnuB homologs, which contain only coiled-coils fused at their
387 N-termini, cold-shock domain (CSD)-like OB-folds are usually fused at the N-termini of Type I CoCoNuT
388 CnuB/McrB proteins, in addition to coiled-coils (except for the majority of Type I-B, where the coiled-
389 coils are encoded separately) (Supplementary Figure S2, Figure 3, Figure 5, Supplementary Table S1)
390 [65]. There is also considerable variation in the length of the coiled-coils in Type I-C that was not
391 observed in the other CoCoNuTs. Given that Type I-C is one of the deepest CoCoNuT branches, this
392 greater variation was not unexpected, mirroring what we observed in the other basal clade, Type I-B.
393 Often, in Type I-B and I-C, but not Type I-A, a YTH-like domain, a member of the modified base-binding
394 EVE superfamily, is present in CnuB as well, between the CSD and coiled-coil, or in Type I-B, between the
395 CSD and the GTPase domain (Supplementary Figure S2, Figure 3, Figure 5, Supplementary Table S1) [4,
396 66, 67].

397 The CnuC/McrC proteins in Type I CoCoNuTs all contain an immunoglobulin-like N-terminal
398 beta-sandwich domain of unknown function, not present in the *E. coli* K-12 McrC homolog, similar to a
399 wide range of folds from this superfamily with diverse roles (Figure 5) [68, 69]. It is also present in non-
400 CoCoNuT McrC homologs associated with McrB GTPase homologs in the NxD clade, implying its function
401 is not specific to the CoCoNuTs.

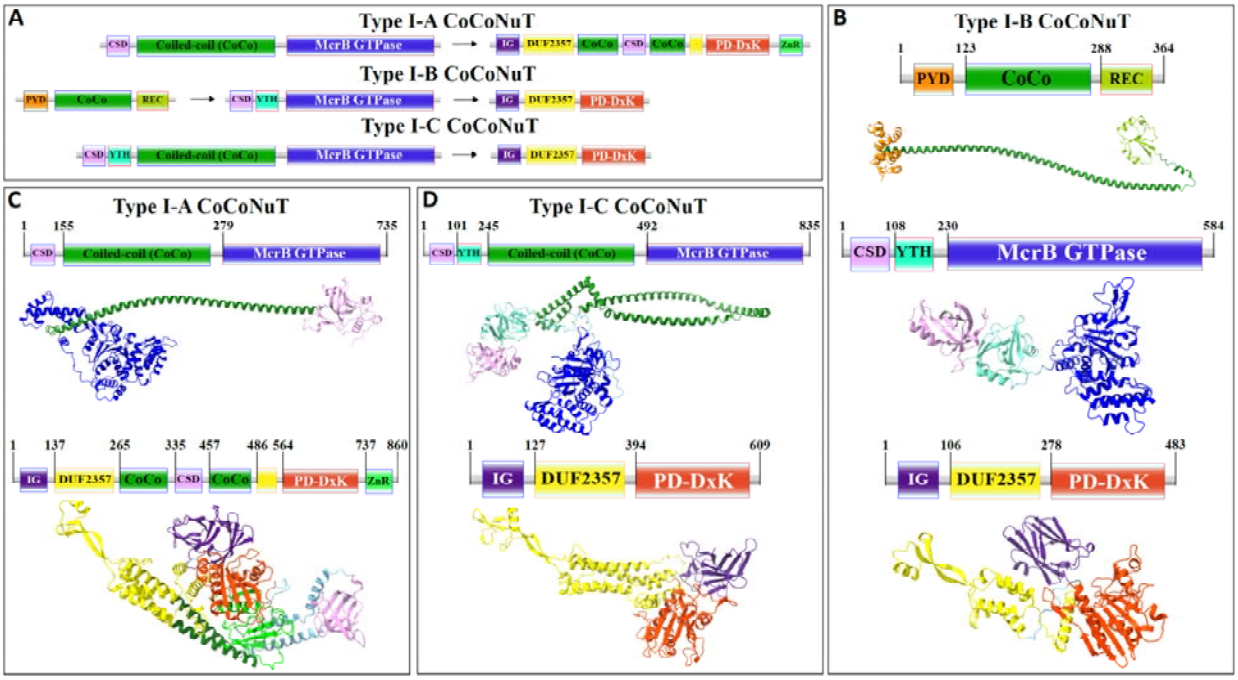



**Figure 5: Domain composition, operon organization, and AlphaFold2 structural predictions of components of the Type I CoCoNuT systems**

A) Type I CoCoNuT domain composition and operon organization. The arrows indicate the direction of transcription. B-D) High quality (Average pLDDT > 80), representative AlphaFold2 structural predictions for proteins in B) Type I-B CoCoNuT systems, C) Type I-A, and D) Type I-C CoCoNuT systems. Models were generated from representative sequences with the following GenBank accessions (see Supplementary Data for sequences and locus tags): ROR86958.1 (Type I-A CnuB), APL73566.1 (Type I-A CnuC), GED20858.1 (Type I-B CnuB), GED20857.1 (Type I-B CnuC), TKH01449.1 (Type I-B CnuA), GFD85286.1 (Type I-C CnuB), MBV0932851.1 (Type I-C CnuC). Abbreviations of domains: CSD – cold shock domain; YTH – YTH-like domain; CoCo – coiled-coil; IG – Immunoglobulin (IG)-like beta-sandwich domain; ZnR – zinc ribbon domain; PYD – pyrin/CARD-like domain; REC – Phosphoacceptor receiver-like domain. These structures were visualized with ChimeraX [40].

We also detected a close relative of Type I-B CoCoNuT in many *Bacillus* species, which we denoted Pseudo-CoCoNuT because it lost the separate coiled-coil protein CnuA. Pseudo-CoCoNuT occurs in genomic contexts suggesting a role in overcrowding-induced stress responses (Supplementary Figure S2, Figure 3, Supplementary Figure S3). We inferred this functional prediction from the fact that these islands encode a quorum-sensing hormone synthase, mechanosensitive ion channel, various transporters, antibiotic resistance and synthesis factors, and cell wall-related proteins. In many cases, we detected Type I-B CoCoNuT homologs in the extended neighborhoods of the Pseudo-CoCoNuTs, although only rarely in the immediate vicinity. Thus, Pseudo-CoCoNuTs might be derived duplicates of Type I-B CoCoNuTs that acquired a specialized but likely related functionality, perhaps still using the coiled-coil protein encoded by the Type I-B CoCoNuT.

*Type II and III CoCoNuT systems*

Type II and III CoCoNuT CnuB/McrB GTPase domains branch from within Type I-A and are encoded in a nearly completely conserved genomic association with a Superfamily 1 (SF1) helicase of the

428 UPF1-like clade, which we denote CnuH (Supplementary Figure S2, Figure 3, Figure 4, Supplementary
429 Table S1) [70, 71]. The UPF1-like family encompasses helicases with diverse functions acting on RNA and
430 single-stranded DNA (ssDNA) substrates, and notably, the prototypical UPF1 RNA helicase and its closest
431 relatives are highly conserved in eukaryotes, where they play a critical role in the nonsense-mediated
432 decay (NMD) RNA surveillance pathway [72, 73]. SF1 helicases are composed of two RecA-like domains
433 which together harbor a series of signature motifs required for the ATPase and helicase activities,
434 including the Walker A and Walker B motifs conserved in P-loop NTPases, which are located in the N-
435 terminal RecA-like domain [71]. Following the Walker A motif in all 4 Type II and Type III CoCoNuT
436 subtypes, the CnuH helicases contain a large helical insertion, with some of the helices predicted to form
437 coiled-coils (Figure 6). The Type II, Type III-A, and Type III-B CoCoNuT CnuH helicases contain an OB-fold
438 domain that, in Type II and Type III-A, is flanked by helices predicted to form a stalk-like helical extension
439 of the N-terminal RecA-like domain, a structural feature characteristic of the entire UPF1/DNA2-like
440 helicase family within SF1 [73-75] (Figure 6). DALI comparisons show that the CnuH OB-fold domain in
441 Type II CoCoNuTs is similar to the OB-fold domain in UPF1 and related RNA helicases SMUBP-2 and
442 SEN1, and this holds for Type III-A as well, although, in these systems, the best DALI hits are to
443 translation factor components such as EF-Tu domain II (Supplementary Table S1) [73, 76]. The Type III-B
444 CnuH OB-folds also match that of UPF1, albeit with lower statistical support (Supplementary Table S1).

445 In Type II and Type III-A CoCoNuTs, CnuH is fused at the N-terminus to a second OB-fold domain
446 similar to that of SmpB (small protein B) (Figure 6, Supplementary Figure S4, Supplementary Table S1),
447 which we denote SPB (SmpB-like). SmpB binds to SsrA RNA, also known as transfer-messenger RNA
448 (tmRNA), and is required for tmRNA to associate with and rescue stalled ribosomes [77-79]. These OB-
449 folds also show significant similarity to the tracrRNA/crRNA-interacting oligonucleotide-binding domains
450 (OBDs) of the Cas12 family CRISPR-associated endonucleases, although typically with weaker support
451 than to SmpB (Supplementary Figure S4) [80-82]. Conceivably, SmpB-like OB-folds were independently
452 recruited during the evolution of the complex CoCoNuTs and the transposon-encoded TnpB nucleases,
453 which also possess a similar OB-fold domain (WED) and are the ancestors of the Cas12 endonucleases
454 found in Type V CRISPR-Cas systems [83-85]. The N-terminal OB-folds in Type II and III-A CnuH homologs
455 also resemble prokaryotic HIRAN domains, which are uncharacterized, but have eukaryotic homologs
456 fused to helicases that bind ssDNA (Supplementary Figure S4) [86]. None of these comparisons is
457 without caveats. The SPB domain is around 20 amino acids shorter on average than SmpB itself, and a
458 helix that is conserved in SmpB orthologs and interacts with the tmRNA is absent in the CoCoNuT OB
459 folds (Supplementary Figure S4) [87]. However, two other structural elements of SmpB involved in
460 binding tmRNA are present (Supplementary Figure S4) [88]. The Cas12 family endonuclease OBDs are
461 interrupted by helical inserts, and thus, appear unlikely to share binding specificity and function with the
462 CoCoNuT domains [80-82]. Nevertheless, the structural similarity suggests an ancient evolutionary
463 relationship between the tmRNA system, the CoCoNuTs, and TnpBs/Type V CRISPR-Cas systems. The
464 tmRNA system itself is a type of RNA-guided restriction machinery, in that it recruits RNA degradation
465 factors such as RNase R to mRNAs bound to stalled ribosomes [89]. The CoCoNuT OB-folds are also
466 significantly structurally similar to HIRAN domains that, however, do not overlay with the CoCoNuT
467 domains any better than SmpB (Supplementary Figure S4), lack a small helix that SmpB and the
468 CoCoNuT domains share, and being DNA-binding, do not fit the other pieces of evidence we gathered
469 that all suggest an RNA-binding role for this domain in the CoCoNuTs (see below).

470 The Type II and Type III-A CoCoNuT CnuH helicases contain an additional domain, structurally
471 similar to the RelE/Colicin D RNase fold [90], which we denote RNase toxin-like (RTL), located between
472 the SPB OB-fold and the helicase (Figure 6, Supplementary Table S1). The proteins of this family are
473 ribosome-dependent toxins that cleave either mRNA or tRNA in the ribosomal A-site [91]. This domain
474 was identified by structural similarity search with the AF2 models, but no sequence conservation with
475 characterized members of this family was detected, leaving it uncertain whether the RTL domain is an
476 active nuclease. However, considerable divergence in sequence is not unusual in this toxin family [92,
477 93]. Notably, Type III-A CoCoNuTs contain another domain with the potential to associate with
478 ribosomes, the Hsp70-like NBD/SBD (Figure 6, Supplementary Table S1) [55, 94]. In Type III-B CoCoNuTs,
479 a small potential RNA-binding wHTH domain resembling the archaeal ssDNA-binding protein Sul7s is
480 fused to the CnuH N-termini (Figure 6, Supplementary Table S1). We did not detect additional domains
481 fused to CnuH in Type III-C systems (Figure 6).

482 We searched for homologs of CnuH (see Methods). Our observations of the contextual
483 associations, both of the CoCoNuTs and their relatives, showed that helicases of this large family are
484 typically encoded in operons with downstream genes coding for an elongated wHTH domain fused at its
485 C-terminus to a variety of effectors, generally, nucleases, which we denote CnuE (Figure 3, Figure 4). In
486 the CoCoNuTs, except for Types III-B and III-C, these CnuE effectors are HEPN ribonucleases, with one
487 HEPN domain in Type II and two in Type III-A (Figure 6, Supplementary Table S1). In Type III-B, the
488 effector is a Vsr (very-short-patch repair)-like PD-(D/E)xK family endonuclease [95] fused directly to the
489 helicase, with no wHTH domain present, whereas in Type III-C, a distorted version of the elongated
490 wHTH domain is fused to two Phospholipase D (PLD) family endonucleases (Figure 3, Figure, 4, Figure 6,
491 Supplementary Table S1). All these nucleases can degrade RNA, and some, such as HEPN, have been
492 found to cleave RNA exclusively [54, 96-99]. We also detected coiled-coils in the region between the
493 wHTH and effector domains in Type II, Type III-A, and Type III-C CoCoNuT CnuE homologs. Multimer
494 structural modeling with AlphaFold2 suggests that the wHTH domain in CnuE might interact with the
495 coiled-coil-containing helical insertion of CnuH, perhaps mediated by the coiled-coils in each protein, to
496 couple the ATP-driven helicase activity to the various nuclease effectors (Supplementary Figure S5). The
497 accuracy of this model notwithstanding, the fusion of the Vsr-like effector to CnuH in Type III-B
498 CoCoNuTs and the similarity of this system to the other CoCoNuT types strongly suggests that the CnuE
499 effector proteins in these systems form complexes with their respective CnuH helicases. An additional
500 factor in potential complexing by these proteins is the presence of coiled-coils in the associated
501 CnuB/McrB and CnuC/McrC homologs, which may interact with the coiled-coils in CnuH and CnuE. Type
502 III-B CoCoNuTs also code for a separate coiled-coil protein, which we denote CnuA, as it resembles the
503 CnuA protein encoded in Type I-B CoCoNuTs (Figure 3, Figure 4). However, it is distinguished from Type
504 I-B CnuA in containing no recognizable domains other than the coiled-coil (Figure 3, Figure 4). This also
505 can potentially interact with other coiled-coil proteins in the system.

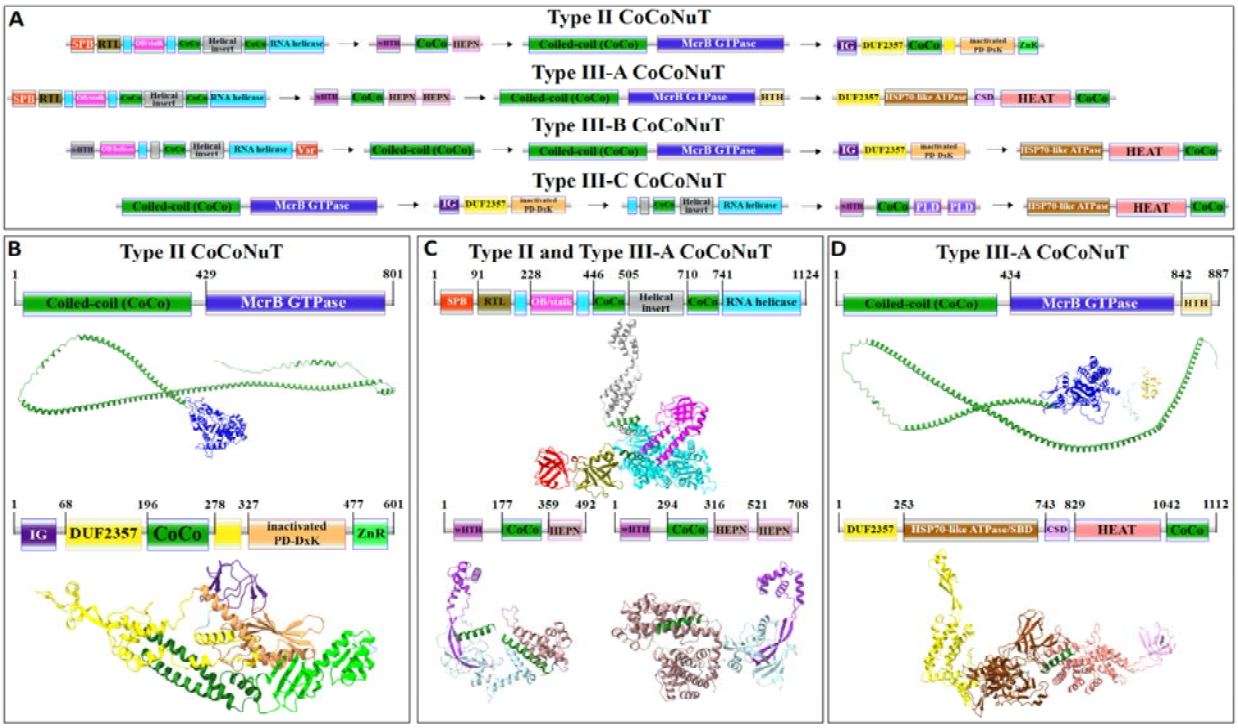



**Figure 6: Domain composition, operon organization, and AlphaFold2 structural predictions for core protein components of Type II and III CoCoNuT systems**

A) Type II and III CoCoNuT domain composition and operon organization. The arrows indicate the direction of transcription. Type II and III-A systems very frequently contain TerY-P systems as well, but not invariably, and these are never found in Type III-B or III-C, thus, we do not consider them core components. B-D) High quality (Average pLDDT > 80), representative AlphaFold2 structural predictions for proteins in B) Type II CoCoNuT systems, C) Type II and III-A CoCoNuT systems, and D) Type III-A CoCoNuT systems. Models were generated from representative sequences with the following GenBank accessions (see Supplementary Data for sequences and locus tags): AMO81401.1 (Type II CnuB), AVE71177.1 (Type II CnuC), AMO81399.1 (Type II and III-A CoCoNuT CnuH), AVE71179.1 (Type II CnuE), ATV59464.1 (Type III-A CnuE), PSR44427.1 (Type III-A CnuB), NMY00740.1 (Type III-A CnuC). Abbreviations of domains: CSD – cold shock domain; CoCo – coiled-coil; IG – Immunoglobulin (IG)-like beta-sandwich domain; ZnR – zinc ribbon domain; SPB – SmpB-like domain; RTL – RNase toxin-like domain; HEPN – HEPN family nuclease domain; OB/stalk – OB-fold domain attached to a helical stalk-like extension of ATPase; Vsr – very-short-patch-repair PD-D/ExK nuclease-like domain; PLD – Phospholipase D family nuclease domain; HEAT – HEAT-like helical repeats. These structures were visualized with ChimeraX [40].

Type II and Type III-A CoCoNuTs, the most widespread varieties apart from Type I, also include conserved genes coding for a "TerY-P" triad. TerY-P consists of a TerY-like von Willebrand factor type A (VWA) domain, a protein phosphatase 2C-like enzyme, and a serine/threonine kinase (STK) fused at the C-terminus to a zinc ribbon (Figure 3, Figure 4, Supplementary Figure S2, Figure 6, Supplementary Table S1). TerY-P triads are involved in tellurite resistance, associated with various predicted DNA restriction and processing systems, and are predicted to function as a metal-sensing phosphorylation-dependent signaling switch [7]. In addition, TerY-P-like modules, in which the kinase is fused at the C-terminus to an OB-fold rather than a zinc ribbon, have been recently shown to function as stand-alone antiphage defense systems [2]. The OB-fold fusion suggests that this kinase interacts with an oligonucleotide and raises the possibility that the zinc ribbon, which occupies the same position in the CoCoNuTs, is also nucleic acid-binding. Almost all CoCoNuT systems containing *cnuHE* operons also encompass TerY-P,

533 with a few exceptions among Terrabacterial Type II systems and Myxococcal Type III-A systems, implying
534 an important contribution to their function (Supplementary Data). However, the complete absence of
535 the TerY-P module in Type III-B and Type III-C systems implies that when different nucleases and other
536 putative effectors fused to the helicase are present, TerY-P is dispensable for the CoCoNuT activity.
537 Therefore, we do not consider them to be core components of these systems. Below, we describe
538 experimental results showing that TerY-P in Type II CoCoNuT can indeed function independently to
539 protect *E. coli* from λ-vir infection partially, but that the other factors are required for full-scale survival.

540 Type II and Type III-A CoCoNuTs are generally similar, but a more detailed comparison reveals
541 substantial differences. The CnuE proteins in Type II contain one HEPN domain with the typical RxxxxH
542 RNase motif conserved in most cases, whereas those in Type III-A contain two HEPN domains, one with
543 the RxxxxH motif, and the other, closest to the C-terminus, with a shortened RxH motif. Furthermore,
544 Type III-A CnuB/McrB GTPase homologs often contain C-terminal HTH-domain fusions absent in Type II
545 (Figure 6). Finally, Type II CnuC/McrC homologs resemble Type I-A CnuCs, which contain N-terminal
546 immunoglobulin-like beta-sandwich domains, insertions of coiled-coils and a CSD-like OBD into
547 DUF2357, PD-D/ExK nucleases, and zinc ribbon domains, except that in Type II CnuCs, this domain
548 architecture underwent reductive evolution (Figure 5, Figure 6, Supplementary Table S1). In particular,
549 the beta-sandwich domain and coiled-coils are shorter, the CSD was lost, the nuclease domain was
550 inactivated, and in many cases, the number of Zn-binding CPxC motifs was reduced from three to two
551 (Figure 5, Figure 6, Supplementary Table S1). This degeneration pattern could indicate functional
552 replacement by the associated CnuH and CnuE proteins, often encoded in reading frames overlapping
553 with the start of the *cnuBC* operon.

554 By contrast, Type III-A CnuC/McrC homologs entirely lost the beta-sandwich domain, PD-D/ExK
555 nuclease, and zinc ribbon, but gained an Hsp70-like NBD/SBD unit similar to those that are fused to the
556 McrB GTPase in early branching members of the NxD clade, as well as a helical domain similar to the
557 HEAT repeat family, and in some cases, a second CSD (Figure 5, Figure 6, Supplementary Table S1). Most
558 Type III-A CnuCs contain a CSD and coiled-coils, and thus, resemble Type I-A, but the positioning of these
559 domains, which in Type I-A and Type II are inserted into the DUF2357 helix bundle, is not conserved in
560 Type III-A, where these domains are located outside DUF2357 (Figure 5, Figure 6). The further addition
561 of the Hsp70-like NBD/SBD and HEAT-like repeat domains accentuates the distinct organization of Type
562 III-A (Figure 5, Figure 6, Supplementary Table S1). Moreover, some *Pseudomonas* genomes encode both
563 Type II and Type III-A CoCoNuTs, implying distinct targets. As mentioned above, the Type II CnuH
564 helicases also differ from Type III-A in that their stalk-attached OBDs more closely resemble UPF1-like
565 helicases, whereas the OBDs in this position in Type III-A are more similar to EF-Tu domain II, a
566 divergence potentially linked to the number of HEPN domains (1 vs. 2) in the closely associated CnuE
567 proteins (Figure 6, Supplementary Table S1).

568 Hsp70-like NBD/SBD domains and HEAT-like repeats are fused to the CnuC/McrC N-terminal
569 DUF2357 domain in Type III-A CoCoNuT, but their homologs in Type III-B and III-C are encoded by a
570 separate gene. We denote these proteins CnuD and CnuCD, the latter for the CnuC-CnuD fusions in Type
571 III-A. The separation in Type III-B and III-C implies that fusion is not required for their functional
572 interaction with the CnuBC/McrBC systems (Figure 6, Supplementary Table S1). CnuD proteins
573 associated with both Type III-B and III-C contain predicted coiled-coils, suggesting that they might
574 interact with the large coiled-coil in the CnuB homologs, which would not be unexpected, given that
575 similar domains are fused to closely related McrB GTPases (Figure 3, Supplementary Figure S2). Because

576 Hsp70 homologs have an affinity for AU-rich RNA, and in some cases, such as the bacterial DnaK, C-rich
577 RNA, and given the strong predictions of RNA interactions for the other components of the CoCoNuTs,
578 we hypothesize that the CnuD homologs present in all Type III CoCoNuT subtypes and fused to the NxD
579 McrB GTPase also bind RNA [56, 57]. As cited above, many Hsp70 homologs have been reported to
580 associate with ribosomes [55, 94], which could also be true for the CoCoNuT CnuDs. Hsp70 ATP binding
581 and hydrolysis allosterically regulate the affinity of Hsp70s for peptide and protein substrates, such that
582 the ATP-bound state has low affinity for misfolded proteins, but the ADP-bound protein has high affinity,
583 and ATP hydrolysis "traps" protein substrates by generating a high affinity ADP-bound complex, after
584 initially contacting the substrates in the ATP-bound form [55]. To our knowledge, the effects of ATP
585 binding and hydrolysis on RNA-binding by Hsp70 homologs were investigated in one study only, which
586 reported that RNA binding was independent of nucleotide binding, ATP hydrolysis, and unfolded peptide
587 interaction [56]. In the CnuD homologs found in the CoCoNuTs and fused to NxD McrB GTPases, Walker
588 B-like motifs [100] are usually, but not invariably, conserved, whereas the sequences of the helical
589 domains just adjacent to the motifs are more strongly constrained. Walker A-like motifs [101] are
590 present but degenerate, with the DxGT(T/S) signature motif mutated, although the initial aspartate is
591 conserved in all except the Type III-C CnuD homologs (Supplementary Figure S6). It appears likely that
592 the CoCoNuT CnuD homologs bind ATP/ADP but hydrolyze ATP with extremely low efficiency, at best.
593 Such properties are better compatible with RNA binding than unfolded protein binding or remodeling,
594 suggesting that these domains target the respective systems to aberrant RNA.

595 Consistent with this prediction, Type II and III-A CoCoNuTs generally appear likely to target RNA
596 rather than DNA, given that the respective operons encode HEPN RNases, typically the only recognizable
597 nuclease in these systems. All CnuC/McrC homologs in Type II or III CoCoNuT systems lack the PD-D/ExK
598 catalytic motif that is required for nuclease activity, although for Type II and Type III-B and III-C, but not
599 Type III-A, structural modeling indicates that the inactivated nuclease domain was retained, likely for a
600 nucleic acid-binding role (Figure 6, Supplementary Table S1). In Type III-A CoCoNuT, the nuclease
601 domain was lost entirely and replaced by the Hsp70-like NBD/SBD domain with RNA-binding potential
602 (Figure 6, Supplementary Table S1). Often, one or two CSDs, generally RNA-binding domains, although
603 capable of binding ssDNA, are fused to Type III-A CnuC homologs as well (Figure 6, Supplementary Table
604 S1) [102]. RNA targeting capability of Type III-B and III-C can perhaps be inferred from the similarity to
605 Type III-A. Moreover, higher-order associations of Type II and Type III-A CoCoNuT systems with various
606 DNA restriction systems suggest a two-pronged DNA and RNA restriction strategy reminiscent of Type III
607 CRISPR-Cas (see below).

608 We suspect that RNA targeting is an ancestral feature of the CoCoNuT systems. Several
609 observations are compatible with this hypothesis:

610 1) CSD-like OB-folds are pervasive in these systems, being present in the CnuB/McrB homologs
611 of all Type I subtypes and in Type I-A and Type III-A CnuC/McrC homologs. As previously noted, these
612 domains typically bind RNA, although they could bind ssDNA as well.

613 2) YTH-like domains are present in most Type I systems, particularly, in almost all early
614 branching Type I-B and I-C systems, suggesting that the common ancestor of the CoCoNuTs contained
615 such a domain. YTH domains in eukaryotes sense internal N6-methyladenosine (m6A) in mRNA [66, 67,
616 103].

617 3) Type II and III-B/C CoCoNuTs, which are likely to target RNA, given their genomic context
618 encoding HEPN proteins and Hsp70 NBD/SBD1 homologs, retain inactivated PD-DxK nuclease domains,
619 which suggests these domains contribute an affinity for RNA inherited from Type I-A CoCoNuTs. PD-
620 (D/E)xK nucleases are generally DNA-specific, however, some examples of RNase activity have been
621 reported [97, 98]. The inactivated PD-DxK domains might also bind DNA from which the target RNA is
622 transcribed.

623 *Type II CoCoNuT protects bacteria against plasmids with CloDF13 ori and λ-vir phage*

624 To probe whether CoCoNuTs can defend bacteria against MGE, we engineered heterologous *E.*
625 *coli* strains carrying Type I-A, Type II, or Type III-C CoCoNuT systems on a low-copy pACYC plasmid
626 (Figure 7 A) and challenged them with different phages and plasmids.

627 Initially, we monitored phage restriction in *E. coli* strains DH5α, MG1655, and BL21(DE3)
628 carrying CoCoNuT systems using a classical double-layer agar assay and a set of 22 *E. coli* phages
629 spanning five families: *Autographiviridae* T3, T7; *Demerecviridae* T5, Lau5; *Drexlerviridae* NBD2;
630 *Siphoviridae* λ-vir, Ervi20; *Myoviridae* T2, T4, T6, and other T-even-like phages (Supplementary Table S2).
631 This assay did not reveal phage restriction or plaque size decrease, indicating no phage defense activity
632 of CoCoNuT (Supplementary Figure S7 A, B). The Type I-A CoCoNuT expression in *E. coli* BL21(DE3)
633 triggered toxicity so that the bacterial lawn did not form in soft agar (Supplementary Figure S7 A, B).

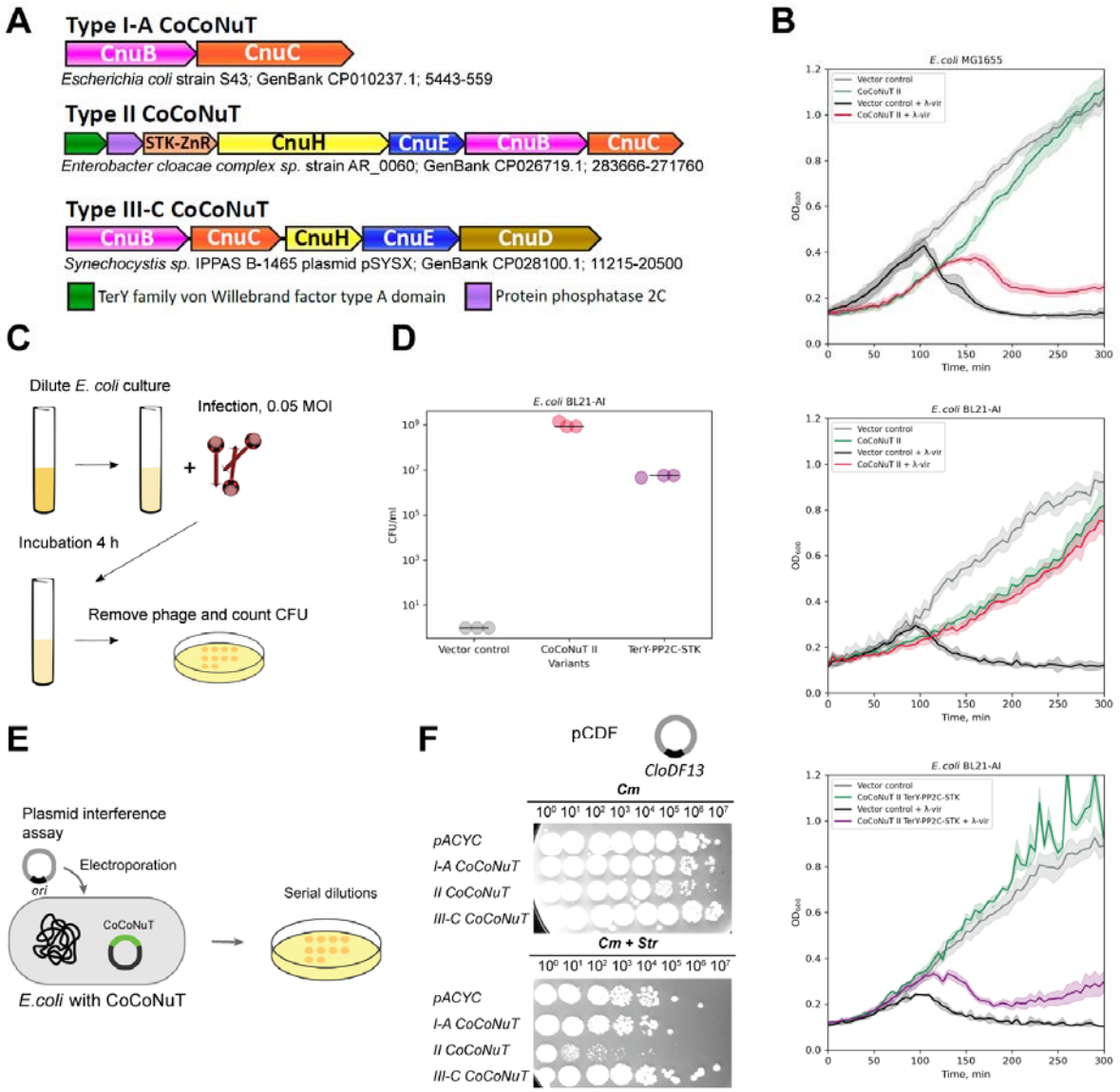
634 Next, we tested the antiphage activity of the CoCoNuTs in *E. coli* MG1655 and BL21-AI strains in
635 a liquid medium (Figure 7 B, Supplementary Figure S7 C, D). In BL21-AI, Type II CoCoNuT increased
636 bacterial survival under λ-vir infection, while for Type I-A CoCoNuT, less pronounced bacterial survival
637 was observed (Figure 7 B, Supplementary Figure S7 D). In the MG1655 strain, only Type II CoCoNuT
638 enabled bacterial survival under phage infection (Figure 7 B, Supplementary Figure S7 C). Given that the
639 phage resistance phenotype was reliably replicated for Type II CoCoNuT and λ-vir phage in both strains
640 (Figure 7 B), we further focused on this CoCoNuT type.

641 As described above, the Type II CoCoNuT locus includes three genes homologous to a stand-
642 alone TerY-P defense system that has been shown to provide resistance against T3, T7, and ϕ-V1 phages
643 [2]. To probe whether the TerY-PP2C-STK genes from Type II CoCoNuT can provide antiphage activity,
644 we cloned only these genes in the same pACYC vector and used the *E. coli* BL21-AI strain to monitor cell
645 survival upon λ-vir infection in liquid medium (Figure 7 B, C, D). The cells showed increased survival
646 compared to control, but the effect was weaker compared to the full-length Type II CoCoNuT system,
647 indicating that TerY-PP2C-STK alone is insufficient to provide full-scale survival (Figure 7 B, D).
648 Furthermore, we compared the number of colony-forming units (CFU) in cultures without CoCoNuT,
649 with wild-type Type II CoCoNuT, or with the CoCoNuT TerY-PP2C-STK variant after 4 hours of λ-vir
650 infection (Figure 7 C, D). Cells carrying an empty vector did not yield any colonies, while the cells
651 carrying Type II CoCoNuT or CoCoNuT TerY-PP2C-STK survived under λ-vir infection, albeit to different
652 extents (Figure 7 D). Under the same conditions, full-length Type II CoCoNuT cells formed 100-fold more
653 CFUs than the TerY-PP2C-STK variant (Figure 7 D). In a separate set of experiments, we found that λ-vir
654 phage formed turbid plaques on Type II CoCoNuT bacteria lawn, indicating that a fraction of the *E. coli*
655 population survives during the infection (Supplementary Figure S7 E). This finding could explain why
656 initially we did not observe a classic phage resistance phenotype in the plaque assay (Supplementary
657 Figure S7 A, B).

658 We next tested whether heterologous *E. coli* strains carrying CoCoNuTs can provide plasmid
659 interference. In this experiment, we used *E. coli* MG1655 strains with CoCoNuTs or a control plasmid and
660 electroporated a set of plasmids with different *ori* sequences (Supplementary Table S3). We plated serial
661 dilutions of the cells on a single antibiotic to select for a CoCoNuT or on two antibiotics, to select for
662 both a CoCoNuT and an invading plasmid that carried a distinct antibiotic resistance gene (Figure 7 E).
663 When Type II CoCoNuT was co-electroporated with the pCDF plasmid containing CloDF13 *ori* sequence,
664 decreased cell viability and smaller colonies were observed, indicative of plasmid interference mediated
665 by the CoCoNuT (Figure 7 F). Plasmid interference was detected only for pCDF plasmids suggesting that
666 CoCoNuT can restrict plasmids carrying a specific *ori* sequence.

667 Taken together, our experimental results show that the Type II CoCoNuT system protects *E. coli*
668 against λ-vir infection in a liquid medium and impairs the retention of pCDF plasmid. Our findings further
669 suggest that similarly to the previously characterized TerY-P system, the TerY-P module of Type II
670 CoCoNuT has an antiphage activity of its own, but that other CoCoNuT components are required for the
671 full antiphage effect. Clearly, these are only the proof of principle experiments demonstrating that at
672 least some of the CoCoNuTs are bacterial defense systems. The basis of the CoCoNuT specificity, the
673 functionalities of the other types of CoCoNuTs, and the mechanistic details of the CoCoNuT activities
674 remain to be elucidated, but the additional bioinformatic analysis we describe below allows some
675 plausible and testable hypotheses to be proposed.

676




677

678 **Figure 7: Type II CoCoNuT protects *Escherichia coli* against plasmids and λ-vir phage**

679 A) Gene organization of the CoCoNuT systems selected for experimental analysis. In the case of the Type II CoCoNuT
680 system, for optimal gene synthesis, 5 bp were changed, preserving the amino acid sequence (see Methods). B) Phage λ-vir
681 challenge in a liquid medium for Type II CoCoNuT or TerY-PP2C-STK gene module from Type II CoCoNuT in different *E. coli*
682 strains. The cells were infected at MOI 0.05. Top – infection of *E. coli* MG1655 strain with empty vector or wild-type Type II
683 CoCoNuT system (5 independent colonies selected, each picked twice, for 10 replicates in total). Middle – infection of *E.*
684 *coli* BL21-AI strain with empty vector or with wild-type Type II CoCoNuT system (in the case of the Type II CoCoNuT variant,
685 4 independent colonies, each picked twice, for 8 replicates in total; otherwise, as described for the MG1655 strain).
686 Bottom – infection of *E. coli* BL21-AI strain with empty vector or wild-type TerY-PP2C-STK gene triad from Type II CoCoNuT
687 (3 independent replicates). The shaded area shows standard deviation. C) Schematic representation of the experimental
688 setup to evaluate cell viability upon λ-vir phage infection. Diluted overnight culture with control or CoCoNuT plasmids
689 were infected with λ-vir phage at MOI 0.05. After inoculation for 4 h, the phage was removed by centrifugation, and cells
690 were plated by serial dilutions to compare the viability of different variants. D) Viability of cells carrying Type II CoCoNuT or
691 TerY-PP2C-STK gene module after infection with λ-vir phage. Colony forming units per milliliter of culture (CFU/ml) were

692 calculated after 4 hours of λ-vir phage infection in *E. coli* cultures harboring vector control (no colonies formed), wild-type
693 Type II CoCoNuT system, or Type II CoCoNuT TerY-STK gene triad only. Three independent colonies were used for three
694 replicates; the black line indicates the median. E) Schematic representation of plasmid interference assay. Heterologous *E.*
695 *coli* cells harboring Type I-A, Type II, or Type III-C CoCoNuT systems in pACYC184 plasmid or empty pACYC184 were co-
696 transformed with plasmids containing different *ori* sequences (Supplementary Table S2), and cell viability was evaluated
697 on agar plates by performing serial dilutions. F) pCDF plasmid co-electroporation into heterologous *E. coli* hosts expressing
698 CoCoNuT systems. Cell viability was evaluated by serial dilutions in the presence of chloramphenicol to select CoCoNuT
699 systems in pACYC184 or in the presence of both chloramphenicol and streptomycin (for pACYC184 and co-transformed
700 pCDF). Type II CoCoNuT showed ~2-fold interference with pCDF. Data with other plasmids are shown in Supplementary
701 Figure S8.

702

703 *CoCoNuT, CARF, and RtcR: A convergence on tmRNA?*

704 Genomic neighborhoods of many Type II and Type III-A CoCoNuTs encompass complex operonic
705 associations with genes encoding several types of CARF (CRISPR-Associated Rossmann Fold) domain-
706 containing proteins. This connection seems to suggest multifarious regulation by cyclic (oligo)nucleotide
707 second messengers synthesized in response to viral infection and bound by CARF domains [104-106].
708 Activation of an effector, most often a nuclease, such as HEPN or PD-(D/E)xK, by a CARF bound to a
709 cyclic (oligo)nucleotide is a crucial mechanism of CBASS (cyclic oligonucleotide-based antiphage signaling
710 system) as well as Type III CRISPR-Cas systems [104-106]. These CARF-regulated enzymes generally
711 function as a fail-safe that eventually induces PCD/dormancy, if other antiphage defenses fail to bring an
712 infection under control, and are deactivated, typically through cleavage of the second messenger by a
713 RING nuclease, if other mechanisms succeed [104, 107-109].

714 The presence of CARFs could implicate the HEPN domains in these systems as PCD effectors that
715 would carry out non-specific RNA degradation in response to infection. Surprisingly, however, most of
716 these CARF domains showed the highest similarity to the CARF domain of RtcR, a sigma54
717 transcriptional coactivator of the RNA repair system RtcAB with a CARF-ATPase-HTH domain
718 architecture, suggesting an alternative functional prediction (Supplementary Figure S9, Supplementary
719 Table S1). Specifically, by analogy with RtcR, CoCoNuT-associated CARF domain-containing proteins
720 might bind (t)RNA fragments with cyclic phosphate ends [110, 111]. This interaction could activate
721 transcription initiation of downstream genes, in this case, genes encoding CoCoNuT components,
722 through binding an upstream activating sequence by the HTH domain fused to the CARF-ATPase C-
723 terminus [110, 111]. The manifold effects of tRNA-like fragments are only beginning to be appreciated.
724 Lately, it has been shown that bacterial anticodon nucleases, in response to infection and DNA
725 degradation by phages, generate tRNA fragments, likely a signal of infection and a defensive strategy to
726 slow down the translation of viral mRNA, and that phages can deploy tRNA repair enzymes and other
727 strategies to counteract this defense mechanism [112-115].

728 In the RtcAB system, tRNA fragments with 2', 3'-cyclic phosphate ends, which could be produced
729 as a consequence of stress-related damage and anticodon nuclease activity, bind to the CARF domain of
730 RtcR, leading to its oligomerization and activation of *rtcAB* operon expression [111, 116]. However, a
731 recent study found that, although the CARF domain of RtcR does bind 2', 3'-cyclic phosphate tRNA
732 fragments, full-length RtcR did not stably associate with these tRNA fragments that were insufficient for
733 RtcR to initiate transcription *in vitro* [110]. Thus, the exact identity of the ligands of these CARFs remains
734 uncertain, but a consensus that they bind 2' 3' cyclic phosphate RNA is emerging [110, 111]. RtcA is an

735 RNA cyclase that converts 3′-phosphate RNA termini to 2′, 3′-cyclic phosphate and thus, in a feedback
736 loop, generates 2′, 3′-cyclic phosphate RNA fragments that induce expression of the operon [111, 116,
737 117]. RtcB is an RNA ligase that joins 2′, 3′-cyclic phosphate RNA termini to 5′-OH RNA fragments,
738 generating a new 5′-3′ bond and, in many cases, reconstituting a functional tRNA [118]. In eukaryotes,
739 RtcB homologs have important roles in joining pre-tRNA halves after intron splicing and in the unfolded
740 protein response [119, 120], and in bacteria, compelling recent work has shown that the most frequent
741 target of RtcB is SsrA, the tmRNA [110].

742 The tmRNA, a unique hybrid molecule that combines a tRNA-like domain (TLD) with an internal
743 ORF in an mRNA-like region (MLR), might unexpectedly be key to cracking the CoCoNuTs and elucidating
744 their functions [78]. In all bacteria and some mitochondria, but not archaea or eukaryotes, where this
745 system does not appear to exist, tmRNA binds to the SmpB protein [78]. This interaction is essential for
746 targeting the complex to stalled ribosomes, a type of translation stress [78]. In response to these
747 irregular translation events, often caused by mRNAs missing a stop codon, tmRNA switches translation
748 by the stalled ribosome to the ORF in the tmRNA MLR, which encodes a small peptide tag [79].
749 Translation of this ORF fuses the tag to the C-terminus of the nascent polypeptide, marking it for
750 proteolysis, and releases the ribosome when it reaches the MLR stop codon [89].

751 In both Type II and III-A CoCoNuTs, the SmpB-like SPB OB-fold (Figure 6, Supplementary Figure
752 S4) fused at the N-termini of the CnuH helicases might bind tmRNA, which is targeted for repair by RtcB,
753 itself regulated by the same family of RtcR homologs as the CoCoNuT systems. Given this close kinship,
754 these CoCoNuTs might respond to the same stimuli as *rtcAB*, with which they are often present in the
755 same genomes. If both genes in the operon are activated, RtcA could provide RNA fragments with 2′, 3′-
756 cyclic phosphate ends that would promote CoCoNuT transcription. Concurrently, RtcB could regenerate
757 tmRNA that might enable the localization of CoCoNuT RNA degradation complexes to mRNAs associated
758 with stalled ribosomes. A search against the database of predicted proteins from completely sequenced
759 prokaryotic genomes, described above, showed that 31% of those that encoded a Type II or III-A
760 CoCoNuT also contained a typical RtcRB locus (RtcA is not always present in these operons). Although
761 this co-occurrence was less common than expected, our search was not exhaustive, and furthermore,
762 we observed several cases when an RtcB homolog was encoded within or nearby the CoCoNuT operon
763 itself (Supplementary Figure S10). Regardless of the co-occurrence with RtcRAB, tmRNA is present in any
764 bacterium encoding CoCoNuTs because this RNA is (nearly) universal in bacteria, while missing in
765 archaea, which conspicuously lack CoCoNuTs as well.

766 We performed phylogenetic analysis of confidently identified ATPase domains of CnuH and
767 UPF1-like helicase homologs obtained by a search against the completely sequenced prokaryotic
768 genomes described above, as well as a search against a set of 72 representative eukaryotic genomes
769 that were downloaded from the NCBI GenBank in June 2020. The topology of the resulting phylogenetic
770 tree indicated that the CoCoNuT CnuH helicases are more closely related to the UPF1-like helicases than
771 other members of this family, such as DNA2, which is fused to an AddB-like nuclease and acts on ssDNA
772 [74] (Figure 8 A). As detailed above, in addition to the ATPase domain, these helicases also share OBDs
773 attached to stalk-like helical domains (Figure 6). In UPF1, this OBD interacts with RNA 3′ ends [73]. These
774 domains in the CoCoNuT CnuHs are most structurally similar to either the corresponding domain in
775 UPF1 or to OB-fold domains of translation factors, such as EF-Tu and ribosomal protein L3
776 (Supplementary Table S1). The tmRNA forms a ribonucleoprotein complex with SmpB, EF-Tu, and
777 ribosomal protein S1 [77], and S1-like CSDs are present in the CnuC/McrC homologs of Type III-A

778 CoCoNuTs (Figure 6). Therefore, if the CoCoNuT helicases indeed bind tmRNA, the multiple OB-folds
779 structurally similar to tmRNA-binding domains could compensate for the absence of some of the
780 structural elements of SmpB in the CoCoNuT SPB domains. Additionally, the ATPase domain of UPF1
781 contains a helical insert, which is required for RNA binding, in the same position, following the Walker A
782 motif, as the much larger helical insert in the CoCoNuT CnuH helicases, which might also bind RNA
783 (Figure 6) [72].

784 Emerging from this phylogeny, and in accord with the hypothesis of tmRNA interaction with the
785 Type II and Type III-A CoCoNuT CnuH helicases, we observed clear associations between close relatives
786 of the CoCoNuTs and SlpA homologs that catalyze recombination involving tmRNA genes (Figure 8 B).
787 The SlpA proteins are P4 integrase homologs characterized as factors required for the excision of a P4-
788 like cryptic prophage, CP4-57, in *E. coli* K-12 [121]. This excision disrupts the tmRNA gene *ssrA* that
789 overlaps with the *attL* integration site and removes at least several kilobases of the downstream
790 sequence [121]. The presence of CnuH-like helicases encoded adjacent to *slpA* genes implies that
791 excision of the cryptic prophage would also delete the helicase gene while simultaneously inactivating
792 the tmRNA, establishing a direct connection between the functioning of the tmRNA system and
793 expression of the helicase. Encoded immediately upstream of these loci is the tmRNA-binding protein
794 SmpB (Figure 8 B). The CnuH-like helicases in these neighborhoods share many features with those
795 found in Type II and III-A CoCoNuTs, including the ATPase, the OBD/stalk domain, the helical insert
796 following the Walker A motif with coiled-coils, and the RelE family RNase toxin-like (RTL) domain. At the
797 N-termini, these helicases are fused to domains with a serine-threonine kinase (STK)-like fold
798 (Supplementary Table S1). These are similar to the kinase homology (KH) domains found in RNase L, a
799 eukaryotic antiviral HEPN superfamily endoribonuclease that, notably, is activated by oligoadenylate
800 second messengers [122, 123] (Figure 8 B, Supplementary Table S1). However, whereas the RNase L KH
801 domains are inactivated pseudokinases, we predict the domains in the CnuH homologs to be active
802 kinases [122]. Like their CoCoNuT CnuH relatives, these helicases occur in operons with CnuE-like
803 predicted effectors containing a wHTH domain fused to a PLD-type nuclease (Figure 8 B, Supplementary
804 Table S1). In some CnuH homologs, more distantly related to the CoCoNuTs, there is an additional N-
805 terminal NERD (nuclease-related domain) with a D-ExK nuclease motif, and two tandem kinase domains
806 instead of one (Figure 8 B). In the associated CnuE-like effector, the PLD nuclease is likely inactivated
807 (Figure 8 B, Supplementary Table S1). These CnuHE-like tandems, AlpA, and SlpA homologs are adjacent
808 to tyrosine tRNA operons that express RtT sRNAs rather than tmRNA genes (Figure 8 B). This genomic
809 context implies that these loci are distinct cryptic prophages that integrate into the RtT sRNA gene.
810 Intriguingly, RtT is another type of RNA that regulates a stress response, in this case, to starvation and
811 antibiotics, distinct from the tmRNA translation stress response mechanism [124-126]. These tmRNA
812 and RtT sRNA-associated CnuH-like helicases are among the most closely related to the CoCoNuTs in our
813 phylogenetic tree of the UPF1-like family (Figure 8 A, B).

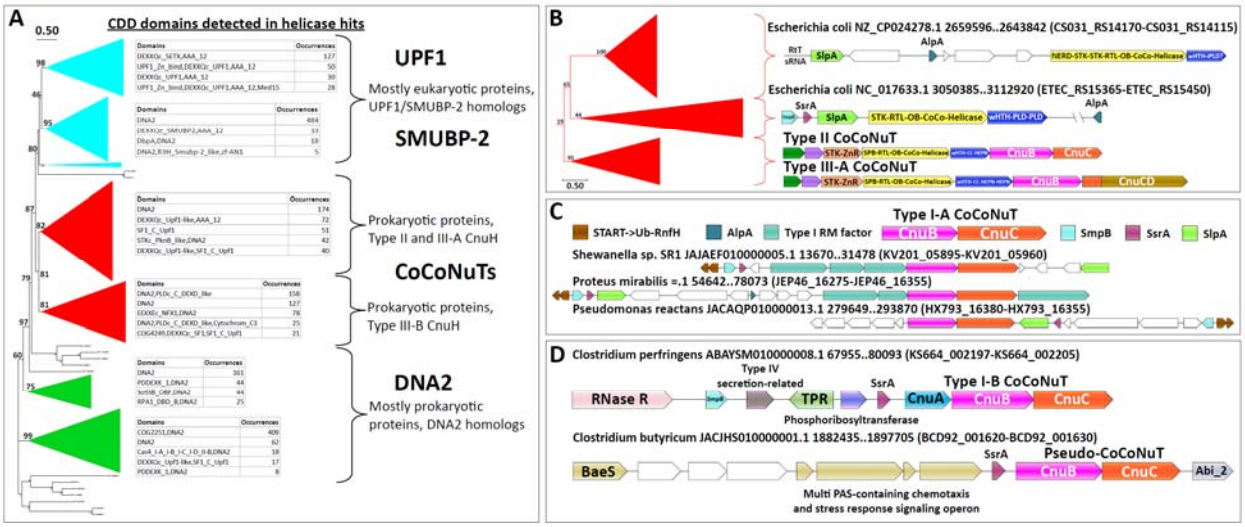



**Figure 8: Type II and III-A CoCoNuT helicases are close relatives of UPF1-like RNA helicases, and CoCoNuT homologs are associated with SmpB/SsrA(tmRNA) genes**

A) Phylogenetic analysis of the UPF1-like helicase family and CnuH. B) Neighborhood analysis of the closest relatives of the CoCoNuT helicases shows association with SlpA prophage integrase homologs that integrate into tmRNA and RtT sRNA genes in *E. coli*. The broken line for *E. coli* NC_017633.1 indicates a region omitted for clarity that contains several recombination-related genes. C) Type I-B CoCoNuTs and the closely related Pseudo-CoCoNuTs can be found adjacent to tmRNA and SmpB genes in *Clostridium*. Abbreviations of domains: SPB – SmpB-like domain; RTL – RNase toxin-like domain; OB/stalk – OB-fold domain attached to a helical stalk-like extension of ATPase; STK – serine/threonine kinase-like domain; PLD – Phospholipase D family nuclease domain; SmpB – Small protein B; SsrA – tmRNA gene; SlpA – P4 integrase homolog; RtT – tRNA-derived ncRNA; AlpA – A prophage transcription factor that stimulates expression of SlpA homologs; Abi_2 – Abi-like HEPN domain-containing protein; BaeS – Signal transduction histidine kinase; TPR – Tetratricopeptide repeat protein; PAS – PAS fold signal transduction domain; iPLD? – predicted inactivated PLD nuclease fold; Ub-RnfH – RnfH family ubiquitin; START – START domain-containing protein. Ub-RnfH and START are frequently associated with SmpB/tmRNA in bacteria, suggesting they play a role in the tagging and proteolysis of targeted proteins, similar to how ubiquitin family proteins are involved in this process in eukaryotes [127].

These observations suggest an evolutionary scenario in which SmpB became fused at the N-terminus of a common ancestor of the Type II CoCoNuT CnuH helicases and their closest relatives, which still associate with SmpB (Figure 8 B). This hypothetical fusion protein could have retained the tmRNA-binding capacity of SmpB. The operon encoding this CnuH-like helicase and its associated CnuE-like effector, which we can infer to have been a HEPN domain-containing protein, could then have become associated with a Type I-A CoCoNuT system, likely also RNA-targeting, as well as a TerY-P operon, to give rise to the Type II CoCoNuTs, after reductive evolution of Type I-A CoCoNuT (see above). Conceivably, this combination could have enabled the highly efficient degradation of both aberrant mRNAs and their nascent proteins, an effective antiviral and general stress response, but perhaps one that comes with a substantial cost to the host. Regulation by RtcR homologs could have been selected to confer additional precision in deploying this potent mechanism. Type III CoCoNuTs and CoCoPALMs seem to have evolved subsequently from Type II CoCoNuTs through a complex evolutionary process. Thus, the CoCoNuT CnuH helicases appear to be but one branch of a large helicase family associated with RNA restriction systems, including NMD and tmRNA/SmpB/RNase R, many of which likely act at the level of translation and, in prokaryotes, often mobilize via prophages which integrate into tmRNA and tRNA genes.

845 Further support for this hypothesis seems to come from the contextual associations of Type I-A
846 CoCoNuTs in Gammaproteobacteria and Type I-B CoCoNuTs and Pseudo-CoCoNuTs in *Clostridium,* which
847 are often encoded near the tmRNA gene *ssrA* and SmpB (Figure 8 C, D, Supplementary Table S1). Type I-
848 A CoCoNuTs, which possess two CSD-like OB-folds, one each in CnuB/McrB and CnuC/McrC (Figure 3,
849 Figure 5), can be found associated not only with SmpB/tmRNA, but also with SlpA and AlpA homologs,
850 implying these are also capable of mobilization via prophages that target the tmRNA gene (Figure 8 C,
851 Supplementary Table 1). In two examples, the Type I-A CoCoNuT is also tightly associated with a Type I
852 RM system (Figure 8 C). The Type I-B and Pseudo-CoCoNuTs have contextual connections to additional
853 factors predicted to be involved in defensive and stress-related signaling, such as alarmone-interacting
854 phosphoribosyltransferases, PAS domain-containing proteins, and HEPN domain-containing Abi-like
855 proteins (Figure 8 D, Supplementary Table S1) [128-130]. The genomic environments of these Type I-A
856 and I-B/Pseudo-CoCoNuT systems suggest they might be involved in the tmRNA-directed degradation of
857 aberrant and viral mRNA (Figure 8 C, D). In Type I-A, the two CSD-like OB-folds might interact with
858 tmRNA, and in Type I-B, the CSD and YTH domains fused to the CnuB/McrB GTPase domains in these
859 systems might interact with modified tmRNA.

860 In light of these observations, the origins and roles of the CnuBC systems in Type II and III
861 CoCoNuTs, which have lost the OB-fold domains, begin to come into focus. Both the Type I-A CoCoNuTs
862 and the CnuHE-like systems that are the closest relatives of the Type II/III CoCoNuTs can be found as
863 apparent cargo of P4-like prophages that integrate into tmRNA genes (Figure 8 B, C). Therefore, it can be
864 confidently presumed that the CnuHE and Type I-A CoCoNuT operons could have come into association
865 with each other, adding credence to the scenario of the origins of Type II CoCoNuTs proposed above.
866 The reductive evolution of CnuBC in Type II CoCoNuT can be explained by the observation that both
867 CnuHE-like systems and Type I-A CoCoNuT systems associate with SmpB/tmRNA. If Type I-A CoCoNuTs
868 interact with tmRNA and can associate with CnuHE homologs, then, the capture of an SmpB-like tmRNA-
869 binding domain by the associated CnuHE system would make redundant the OB-fold domains of the
870 Type I-A CoCoNuT system. Moreover, the nuclease activity of CnuC could also have become dispensable
871 if the RNase activity of a HEPN domain in CnuE was now directed by the tmRNA interaction.

872 The present role of CnuBC in Type II CoCoNuTs can perhaps be predicted from domains that
873 remain: the inactivated PD-D/ExK nuclease, potentially descending from an ancestral Rnase, and the
874 coiled-coil fused to CnuB, which is much longer than the one fused to CnuB in Type I-A CoCoNuT. The
875 retention of these domains implies an RNA-binding function for Type II CnuC and a gain of function for
876 the coiled-coil in Type II CnuB, likely, complex formation with CnuHE, which allowed the reductive
877 evolution to take place. The role of the CnuB GTPase domain remains somewhat enigmatic. This domain
878 is believed to translocate DNA in canonical McrBC systems, therefore translocation of RNA by the
879 CoCoNuTs is a possibility [25, 44]. The GTPase domain also assembles into hexamers upon binding GTP,
880 and this could be the primary reason for the conservation of CnuBC in Type II CoCoNuT [131]. This
881 trigger for complex formation could be important for a system interacting with the translation
882 apparatus, which is also GTP-dependent [132-134]. Depletion of GTP by CnuBs associated with
883 ribosomes might also contribute to slowing translation elongation and allowing time for RNA restriction
884 and proteolysis to occur.

885 *Complex higher-order associations and functions of the CoCoNuT systems*

886 Although the tmRNA-associated CoCoNuT activity proposed in the preceding section would be a
887 distinct antivirus mechanism, there are intriguing parallels with Type III CRISPR-Cas systems, where the
888 Cas10-Csm-crRNA effector complex binds phage RNA complementary to the spacer of the crRNA,
889 triggering both restriction of phage DNA and indiscriminate cleavage of RNA [105]. The target RNA
890 recognition stimulates the production of cyclic oligoadenylate (cOA) signals by Cas10, which activate, via
891 CARF domain binding, PCD effectors, typically HEPN domain-containing proteins, such as Csm6, that
892 function as promiscuous RNases [54, 135]. One of the two outcomes can result from this cascade: the
893 infection can either be abrogated quickly due to the restriction of the virus DNA, which inhibits cOA
894 signaling via the depletion of viral RNA, along with the activity of RING nucleases, thus averting PCD, or
895 the continued presence of viral RNA stimulates cOA signaling until PCD or dormancy occurs, limiting the
896 spread of viruses to neighboring cells in the bacterial population [5, 107, 109].

897 In CoCoNuTs associated with RtcR homologs, a conceptually similar but mechanistically distinct
898 phenomenon might occur. Although many of these CoCoNuTs only contain an appended gene encoding
899 a CARF domain protein at the 5' end of the predicted operon (Supplementary Figure S9), there are also
900 numerous cases where several types of DNA restriction systems are encoded between the CARF gene
901 and the CoCoNuT (Figure 9). In these cases, nearly all genes are in an apparent operonic organization
902 that can extend upwards of 40 kb, such that they could all be transcribed as a single, giant polycistronic
903 mRNA (Figure 9). Although internal RtcR-independent promoters likely exist in these large loci, the
904 consistent directionality and close spacing of the genes in these superoperons suggests coordination of
905 expression. The complex organization of the CARF-CoCoNuT genomic regions, and by implication, the
906 corresponding defense mechanisms, might accomplish the same effect as Type III CRISPR-Cas, contriving
907 a no-win situation for the target virus. Under this scenario, the virus is either destroyed by the activity of
908 the DNA restriction systems, which would inhibit signal production (likely RNA fragments with cyclic
909 phosphate ends rather than cOA) and drive down CoCoNuT transcription, thereby preventing PCD, or as
910 the virus replicates, signaling and CoCoNuT transcription would continue until the infection is aborted by
911 PCD or dormancy caused by the degradation of host mRNA via the HEPN RNase(s) of the CoCoNuT. The
912 apparent toxicity of the CoCoNuT systems we expressed in *E. coli* (Supplementary Figure S7 A, B) is
913 compatible with this hypothesis, but as these were expressed from a plasmid, it is unclear if the toxicity
914 is naturally part of the CoCoNuT function. Future experiments will titrate the expression of the
915 CoCoNuTs to investigate this possibility further.

916 In many species of *Pseudomonas*, where CoCoNuTs are almost always associated with RtcR and
917 various ancillary factors, Type I RM systems often contain an additional gene that encodes a
918 transmembrane helix and a long coiled-coil fused to an RmuC-like nuclease. These proteins were
919 recently described as ShdA, the core component of the *Pseudomonas*-specific defense system Shield
920 [136]. These can potentially interact with coiled-coils in the CoCoNuTs, perhaps, guiding them to the
921 DNA from which RNA targeted by the CoCoNuTs is being transcribed or vice versa (Figure 9).

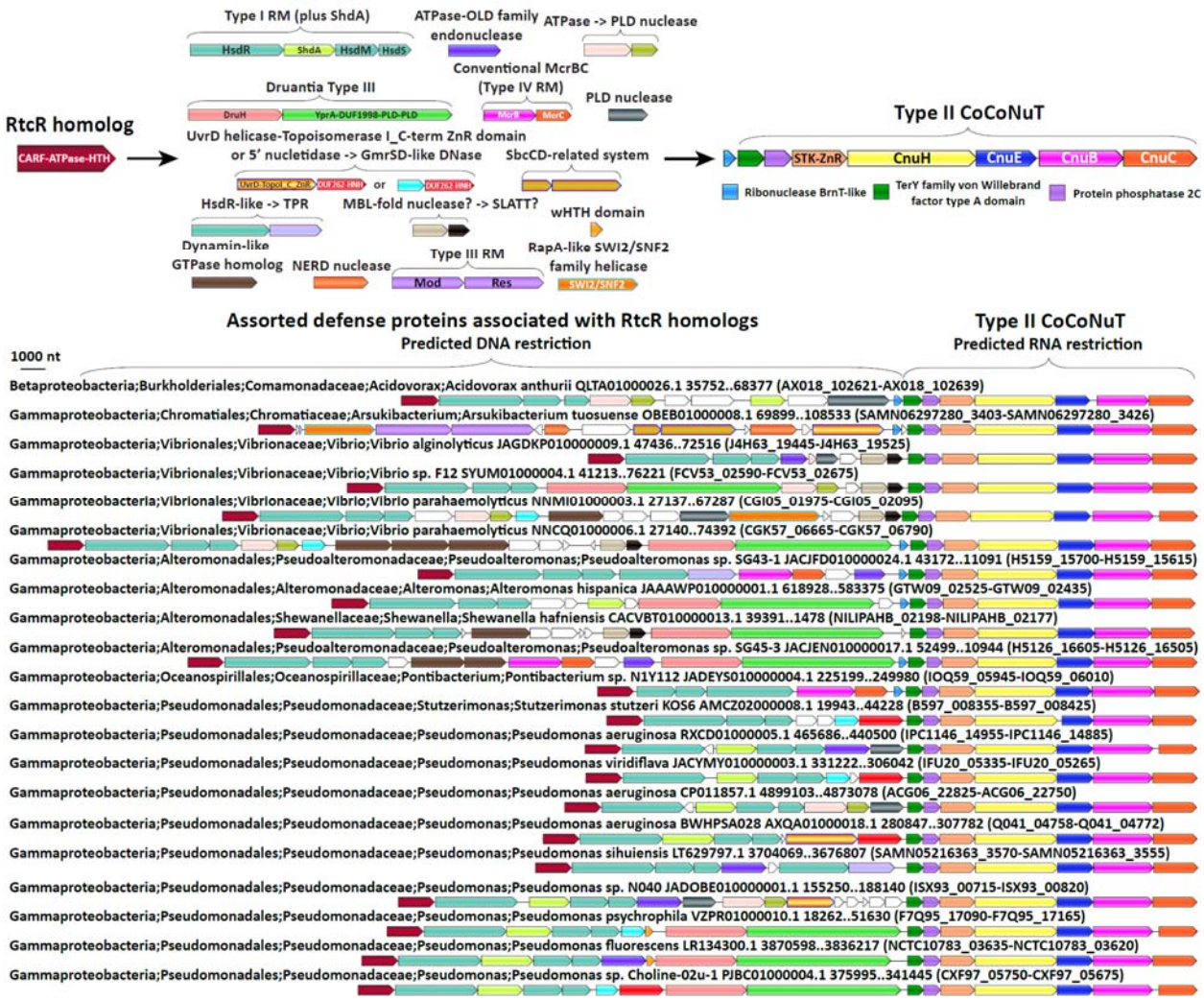



**Figure 9: Complex operonic associations of Type II CoCoNuTs**

Type II CoCoNuTs are frequently associated with RtcR homologs, and in many cases, ancillary defense genes are located between the RtcR gene and the CoCoNuT, almost always oriented in the same direction in an apparent superoperon. Abbreviations of domains: YprA – YprA-like helicase domain; DUF1998 - DUF1998 is often found in or associated with helicases and contains four conserved, putatively metal ion-binding cysteine residues; PLD – Phospholipase D family nuclease domain; SWI2/SNF2 – SWI2/SNF2-family ATPase; HsdR/M/S – Type I RM system restriction, methylation, and specificity factors; ShdA – Shield system core component ShdA; TPR – Tetratricopeptide repeat protein; SLATT – Possible SLATT domain; Mod/Res – Type III RM modification and restriction factors.

A notable difference between the generally similar, dual DNA and RNA-targeting mechanism of many Type III CRISPR-Cas systems and the proposed mechanism of the CoCoNuTs is that specific DNA cleavage activity of the Cas10-Csm complex, which in turn triggers non-specific RNA cleavage, is not activated by cOA binding to a CARF domain, but rather by viral RNA binding. In contrast, in the CARF-containing CoCoNuTs, both the DNA and RNA restriction factors appear to be arranged so that CARF binding of a 2' 3' cyclic phosphate RNA fragment would initiate the expression of the entire cluster [105]. In the case of the CoCoNuTs, signals of infection could activate transcription, first of the DNA restriction systems, and then, the CoCoNuT itself, a predicted RNA restriction system. In these complex

940 configurations of the CoCoNuT genome neighborhoods (Figure 9), the gene order is likely to be
941 important, with Type I RM almost always directly following CARF genes and CoCoNuTs generally coming
942 last, although Druantia Type III sometimes follows the CoCoNuT. As translation in bacteria is co-
943 transcriptional, the products of genes transcribed first would accumulate before those of the genes
944 transcribed last, so that the full, potentially suicidal impact of the CoCoNuT RNA nucleolytic engine
945 would only be felt after the associated DNA restriction systems had ample time to act – and possibly, fail
946 [137].

947 Considering that the CoCoNuTs are potentially involved in degrading non-self and defective
948 mRNA and are abundant in Alphaproteobacteria, with a specialized variant, Type III-B, found mainly in
949 this class, some extant CoCoNuT CnuH homologs and the UPF1-like helicases required for NMD might
950 share a CoCoNuT-like common ancestor that was captured by emerging eukaryotes from the
951 protomitochondrial endosymbionts. In the case of NMD, recognition of a premature termination codon
952 (PTC), the mark of a nonsense message, is not completely understood, but a broadly conserved indicator
953 of a PTC seems to be inefficient translation termination [138-140]. Therefore, it is conceivable, and
954 perhaps not surprising, that the core regulator of NMD, UPF1, has prokaryotic roots in a family of
955 defensive systems that might exploit the bacterial tmRNA mechanism to degrade nonstop messages,
956 which stall ribosomes and prevent translation termination [78, 141]. The sophistication of these systems
957 in bacteria presages that of the NMD pathway, reinforcing the hypothesis on the origin of NMD from the
958 CoCoNuTs.

959

## 960 Concluding remarks

961 In recent years, systematic searches for defense systems in prokaryotes, primarily by analysis of
962 defense islands, revealed enormous, previously unsuspected diversity of such systems that function
963 through a remarkable variety of molecular mechanisms. In this work, we uncovered the hidden diversity
964 and striking hierarchical complexity of a distinct class of defense mechanisms, the Type IV (McrBC)
965 restriction systems. We then zeroed in on a single major but previously overlooked branch of the McrBC
966 family, which we denoted CoCoNuTs for their salient features, namely, the presence of extensive coiled-
967 coil structures and tandem nucleases. Astounding complexity was also discovered at this level, with 3
968 distinct types and multiple subtypes of CoCoNuTs that differ by their domain compositions and genomic
969 associations. All CoCoNuTs contain domains interacting with translation systems components, such as
970 SmpB-like OBD, HSP70 homologs, or YTH domains, along with RNases, such as HEPN, suggesting that at
971 least one of the activities of these systems targets RNA. Most of the CoCoNuTs could be endowed with
972 DNA-targeting activity as well, either integral to the system, such as the McrC nuclease, or more loosely
973 associated, such as Type I RM and Druantia Type III systems that are encoded in the same predicted
974 superoperon with many CoCoNuTs. Numerous CoCoNuTs are associated with proteins containing CARF
975 domains, suggesting that cyclic (oligo)nucleotides regulate the CoCoNuT activity. Given the presence of
976 the RtcR-like CARF domains, it appears likely that the specific second messengers involved are RNA
977 fragments with cyclic phosphate termini. We hypothesize that the CoCoNuTs, in conjunction with
978 ancillary restriction factors, implement an echeloned defense strategy analogous to that of Type III
979 CRISPR-Cas systems, whereby an immune response eliminating virus DNA and/or RNA is launched first,
980 but then, if it fails, an abortive infection response leading to PCD/dormancy via host RNA cleavage takes
981 over. In a series of pilot experiments, we demonstrated that at least Type II CoCoNuTs can protect *E. coli*

982 against specific viruses and plasmids. The detailed exploration of the CoCoNuT functions and
983 mechanisms requires further, extensive experimentation that has the potential to uncover a variety of
984 currently unknown defense circuits.

## 985 Author contributions

986 R.T.B and E.V.K. initiated the project; R.T.B., Y.I.W. and E.V.K. designed research; R.T.B., H.S. and K.S.M.
987 analyzed data; J.J., I.S., T.L. and V.S. performed the experiments; R.T.B and E.V.K. wrote the manuscript
988 that was edited and approved by all authors.

## 989 Acknowledgments

990 The authors thank Dr. Becky Xu Hua Fu (University of California, San Francisco) for correspondence that
991 led to her contribution of the name CoCoNuT, Dr. Andrew Z. Fire and Usman Enam (Stanford University)
992 for critical reading of the manuscript and insightful comments, Dr. Joseph Bondy-Denomy (University of
993 California, San Francisco) for bringing the Shield factor ShdA to our attention, and Koonin group
994 members for helpful discussions. R.T.B., H.S., K.S.M., Y.I.W and E.V.K. are supported by the Intramural
995 Research Program of the National Institutes of Health (National Library of Medicine). J.J and I.S are
996 supported by Vilnius University Science Promotion fund project MSFJM-12/2021, I.S is funded by the
997 European Regional Development Fund under grant agreement number 01.2.2-CPVA-V-716-01-0001 with
998 the Lithuanian Central Project Management Agency (CPVA) and fellowship from The Lithuanian
999 Academy of Sciences.

1000

1001 # References


1002 1. Burroughs, A.M., et al., *Comparative genomic analyses reveal a vast, novel network of*
1003 *nucleotide-centric systems in biological conflicts, immunity and signaling.* Nucleic Acids
1004 Research, 2015. **43**(22): p. 10633-10654.
1005 2. Gao, L., et al., *Diverse enzymatic activities mediate antiviral immunity in prokaryotes.* Science,
1006 2020. **369**(6507): p. 1077-1084.
1007 3. Goldfarb, T., et al., *BREX is a novel phage resistance system widespread in microbial genomes.*
1008 Embo j, 2015. **34**(2): p. 169-83.
1009 4. Bell, R.T., Y.I. Wolf, and E.V. Koonin, *Modified base-binding EVE and DCD domains: striking*
1010 *diversity of genomic contexts in prokaryotes and predicted involvement in a variety of cellular*
1011 *processes.* BMC Biol, 2020. **18**(1): p. 159.
1012 5. Koonin, E.V. and L. Aravind, *Origin and evolution of eukaryotic apoptosis: the bacterial*
1013 *connection.* Cell Death & Differentiation, 2002. **9**(4): p. 394-404.
1014 6. Swarts, D.C., et al., *The evolutionary journey of Argonaute proteins.* Nature structural &
1015 molecular biology, 2014. **21**(9): p. 743-753.
1016 7. Anantharaman, V., L.M. Iyer, and L. Aravind, *Ter-dependent stress response systems: novel*
1017 *pathways related to metal sensing, production of a nucleoside-like metabolite, and DNA-*
1018 *processing.* Molecular bioSystems, 2012. **8**(12): p. 3142-3165.
1019 8. Kaur, G., et al., *Highly regulated, diversifying NTP-dependent biological conflict systems with*
1020 *implications for the emergence of multicellularity.* eLife, 2020. **9**: p. e52696.
1021 9. Loenen, W.A., et al., *Highlights of the DNA cutters: a short history of the restriction enzymes.*
1022 Nucleic Acids Res, 2014. **42**(1): p. 3-19.
1023 10. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis*
1024 *elegans.* Nature, 1998. **391**(6669): p. 806-11.
1025 11. Agrawal, N., et al., *RNA interference: biology, mechanism, and applications.* Microbiol Mol Biol
1026 Rev, 2003. **67**(4): p. 657-85.
1027 12. Makarova, K.S., et al., *A putative RNA-interference-based immune system in prokaryotes:*
1028 *computational analysis of the predicted enzymatic machinery, functional analogies with*
1029 *eukaryotic RNAi, and hypothetical mechanisms of action.* Biology Direct, 2006. **1**(1): p. 7.
1030 13. Makarova, K.S., et al., *Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and*
1031 *derived variants.* Nature Reviews Microbiology, 2020. **18**(2): p. 67-83.
1032 14. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for*
1033 *adaptive immunity in bacteria.* Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-86.
1034 15. Jinek, M., et al., *A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial*
1035 *Immunity.* Science, 2012. **337**(6096): p. 816-821.
1036 16. Raleigh, E.A., R. Trimarchi, and H. Revel, *Genetic and physical mapping of the mcrA (rglA) and*
1037 *mcrB (rglB) loci of Escherichia coli K-12.* Genetics, 1989. **122**(2): p. 279-296.
1038 17. Luria, S.E. and M.L. Human, *A nonhereditary, host-induced variation of bacterial viruses.* J
1039 Bacteriol, 1952. **64**(4): p. 557-69.
1040 18. Fleischman, R.A., J.L. Cambell, and C.C. Richardson, *Modification and restriction of T-even*
1041 *bacteriophages. In vitro degradation of deoxyribonucleic acid containing 5-*
1042 *hydroxymethylctosine.* J Biol Chem, 1976. **251**(6): p. 1561-70.
1043 19. Dila, D., et al., *Genetic and sequence organization of the mcrBC locus of Escherichia coli K-12.*
1044 Journal of Bacteriology, 1990. **172**(9): p. 4888.
1045 20. Sutherland, E., L. Coe, and E.A. Raleigh, *McrBC: a multisubunit GTP-dependent restriction*
1046 *endonuclease.* J Mol Biol, 1992. **225**(2): p. 327-48.

1047 21. Sukackaite, R., et al., *The recognition domain of the methyl-specific endonuclease McrBC flips out*
1048 *5-methylcytosine.* Nucleic acids research, 2012. **40**(15): p. 7552-7562.
1049 22. Weigele, P. and E. Raleigh, *Biosynthesis and Function of Modified Bases in Bacteria and Their*
1050 *Viruses.* Chemical reviews, 2016. **116**.
1051 23. Iyer, L.M., et al., *Evolutionary history and higher order classification of AAA+ ATPases.* Journal of
1052 Structural Biology, 2004. **146**(1): p. 11-31.
1053 24. Nirwan, N., et al., *Structure-based mechanism for activation of the AAA+ GTPase McrB by the*
1054 *endonuclease McrC.* Nature Communications, 2019. **10**(1): p. 3058.
1055 25. Panne, D., E.A. Raleigh, and T.A. Bickle, *The McrBC endonuclease translocates DNA in a reaction*
1056 *dependent on GTP hydrolysis.* J Mol Biol, 1999. **290**(1): p. 49-60.
1057 26. Niu, Y., et al., *Structural asymmetry governs the assembly and GTPase activity of McrBC*
1058 *restriction complexes.* Nat Commun, 2020. **11**(1): p. 5907.
1059 27. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search*
1060 *programs.* Nucleic acids research, 1997. **25**(17): p. 3389-3402.
1061 28. Gabler, F., et al., *Protein Sequence Analysis Using the MPI Bioinformatics Toolkit.* Curr Protoc
1062 Bioinformatics, 2020. **72**(1): p. e108.
1063 29. Zimmermann, L., et al., *A Completely Reimplemented MPI Bioinformatics Toolkit with a New*
1064 *HHpred Server at its Core.* Journal of Molecular Biology, 2018. **430**(15): p. 2237-2243.
1065 30. Hauser, M., M. Steinegger, and J. Söding, *MMseqs software suite for fast and deep clustering*
1066 *and searching of large protein sequence sets.* Bioinformatics, 2016. **32**(9): p. 1323-1330.
1067 31. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space*
1068 *complexity.* BMC bioinformatics, 2004. **5**: p. 113-113.
1069 32. Söding, J., *Protein homology detection by HMM–HMM comparison.* Bioinformatics, 2004. **21**(7):
1070 p. 951-960.
1071 33. Söding, J., et al., *HHsenser: exhaustive transitive profile search using HMM–HMM comparison.*
1072 Nucleic Acids Research, 2006. **34**(suppl_2): p. W374-W378.
1073 34. Price, M.N., P.S. Dehal, and A.P. Arkin, *FastTree 2 – Approximately Maximum-Likelihood Trees*
1074 *for Large Alignments.* PLOS ONE, 2010. **5**(3): p. e9490.
1075 35. Robert, C.E., *High-accuracy alignment ensembles enable unbiased assessments of sequence*
1076 *homology and phylogeny.* bioRxiv, 2022: p. 2021.06.20.449169.
1077 36. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021.
1078 **596**(7873): p. 583-589.
1079 37. Mariani, V., et al., *lDDT: a local superposition-free score for comparing protein structures and*
1080 *models using distance difference tests.* Bioinformatics, 2013. **29**(21): p. 2722-2728.
1081 38. Holm, L., *DALI and the persistence of protein shape.* 2020. **29**(1): p. 128-140.
1082 39. van Kempen, M., et al., *Fast and accurate protein structure search with Foldseek.* Nature
1083 Biotechnology, 2023.
1084 40. Pettersen, E.F., et al., *UCSF ChimeraX: Structure visualization for researchers, educators, and*
1085 *developers.* 2021. **30**(1): p. 70-82.
1086 41. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
1087 42. Neuwald, A.F., et al., *AAA+: A class of chaperone-like ATPases associated with the assembly,*
1088 *operation, and disassembly of protein complexes.* Genome Res, 1999. **9**(1): p. 27-43.
1089 43. Erzberger, J.P. and J.M. Berger, *Evolutionary relationships and structural mechanisms of AAA+*
1090 *proteins.* Annu Rev Biophys Biomol Struct, 2006. **35**: p. 93-114.
1091 44. Pieper, U., et al., *The GTP-binding domain of McrB: more than just a variation on a common*
1092 *theme?* J Mol Biol, 1999. **292**(3): p. 547-56.

1093 45. Hosford, C.J., A.Q. Bui, and J.S. Chappie, *The structure of the Thermococcus gammatolerans*
1094 *McrB N-terminal domain reveals a new mode of substrate recognition and specificity among*
1095 *McrB homologs.* Journal of Biological Chemistry, 2020. **295**(3): p. 743-756.
1096 46. Zimmermann, L., et al., *A Completely Reimplemented MPI Bioinformatics Toolkit with a New*
1097 *HHpred Server at its Core.* J Mol Biol, 2018. **430**(15): p. 2237-2243.
1098 47. Simm, D., K. Hatje, and M. Kollmar, *Waggawagga: comparative visualization of coiled-coil*
1099 *predictions and detection of stable single α-helices (SAH domains).* Bioinformatics, 2015. **31**(5):
1100 p. 767-9.
1101 48. Varadi, M., et al., *AlphaFold Protein Structure Database: massively expanding the structural*
1102 *coverage of protein-sequence space with high-accuracy models.* Nucleic Acids Research, 2022.
1103 **50**(D1): p. D439-D444.
1104 49. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server.* Nucleic Acids
1105 Research, 2015. **43**(W1): p. W389-W394.
1106 50. Raleigh, E.A., *Organization and function of the mcrBC genes of Escherichia coli K-12.* Mol
1107 Microbiol, 1992. **6**(9): p. 1079-86.
1108 51. Ofir, G., et al., *DISARM is a widespread bacterial defence system with broad anti-phage*
1109 *activities.* Nature Microbiology, 2018. **3**(1): p. 90-98.
1110 52. Colicelli, J., *Human RAS superfamily proteins and related GTPases.* Sci STKE, 2004. **2004**(250): p.
1111 Re13.
1112 53. Bourne, H.R., D.A. Sanders, and F. McCormick, *The GTPase superfamily: conserved structure and*
1113 *molecular mechanism.* Nature, 1991. **349**(6305): p. 117-27.
1114 54. Pillon, M.C., et al., *HEPN RNases - an emerging class of functionally distinct RNA processing and*
1115 *degradation enzymes.* Crit Rev Biochem Mol Biol, 2021. **56**(1): p. 88-108.
1116 55. Mayer, M.P., *The Hsp70-Chaperone Machines in Bacteria.* Front Mol Biosci, 2021. **8**: p. 694012.
1117 56. Kishor, A., et al., *Hsp70's RNA-binding and mRNA-stabilizing activities are independent of its*
1118 *protein chaperone functions.* J Biol Chem, 2017. **292**(34): p. 14122-14133.
1119 57. Zimmer, C., A. von Gabain, and T. Henics, *Analysis of sequence-specific binding of RNA to Hsp70*
1120 *and its various homologs indicates the involvement of N- and C-terminal interactions.* Rna, 2001.
1121 **7**(11): p. 1628-37.
1122 58. Kishor, A., et al., *Hsp70 is a novel posttranscriptional regulator of gene expression that binds and*
1123 *stabilizes selected mRNAs containing AU-rich elements.* Mol Cell Biol, 2013. **33**(1): p. 71-84.
1124 59. Khadria, A.S. and A. Senes, *The transmembrane domains of the bacterial cell division proteins*
1125 *FtsB and FtsL form a stable high-order oligomer.* Biochemistry, 2013. **52**(43): p. 7542-50.
1126 60. Chou, S.H. and M.Y. Galperin, *Diversity of Cyclic Di-GMP-Binding Proteins and Mechanisms.* J
1127 Bacteriol, 2016. **198**(1): p. 32-46.
1128 61. Schnappauf, O., et al., *The Pyrin Inflammasome in Health and Disease.* Frontiers in Immunology,
1129 2019. **10**.
1130 62. Park, H.H., et al., *The death domain superfamily in intracellular signaling of apoptosis and*
1131 *inflammation.* Annu Rev Immunol, 2007. **25**: p. 561-86.
1132 63. Kaur, G., et al., *Bacterial death and TRADD-N domains help define novel apoptosis and immunity*
1133 *mechanisms shared by prokaryotes and metazoans.* Elife, 2021. **10**.
1134 64. Stock, A.M., V.L. Robinson, and P.N. Goudreau, *Two-component signal transduction.* Annu Rev
1135 Biochem, 2000. **69**: p. 183-215.
1136 65. Amir, M., et al., *Sequence, structure and evolutionary analysis of cold shock domain proteins, a*
1137 *member of OB fold family.* Journal of Evolutionary Biology, 2018. **31**(12): p. 1903-1917.
1138 66. Hazra, D., C. Chapat, and M. Graille, *$m^6A$ mRNA Destiny: Chained to the rhYTHm by the YTH-*
1139 *Containing Proteins.* Genes, 2019. **10**(1): p. 49.

1140 67. Liao, S., H. Sun, and C. Xu, *YTH Domain: A Family of N6-methyladenosine (m6A) Readers.*
1141 Genomics, Proteomics & Bioinformatics, 2018. **16**(2): p. 99-107.
1142 68. Natarajan, K., M.G. Mage, and D.H. Margulies, *Immunoglobulin Superfamily*, in *Encyclopedia of*
1143 *Life Sciences*. 2015. p. 1-7.
1144 69. Halaby, D.M., A. Poupon, and J. Mornon, *The immunoglobulin fold family: sequence analysis and*
1145 *3D structure comparisons.* Protein Eng, 1999. **12**(7): p. 563-71.
1146 70. Gorbalenya, A.E. and E.V. Koonin, *Helicases: amino acid sequence comparisons and structure-*
1147 *function relationships.* Current Opinion in Structural Biology, 1993. **3**(3): p. 419-429.
1148 71. Fairman-Williams, M.E., U.P. Guenther, and E. Jankowsky, *SF1 and SF2 helicases: family matters.*
1149 Curr Opin Struct Biol, 2010. **20**(3): p. 313-24.
1150 72. Cheng, Z., et al., *Structural and functional insights into the human Upf1 helicase core.* Embo j,
1151 2007. **26**(1): p. 253-64.
1152 73. Chakrabarti, S., et al., *Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and*
1153 *its regulation by Upf2.* Mol Cell, 2011. **41**(6): p. 693-703.
1154 74. Zhou, C., S. Pourmal, and N.P. Pavletich, *Dna2 nuclease-helicase structure, mechanism and*
1155 *regulation by Rpa.* eLife, 2015. **4**: p. e09832.
1156 75. Kalathiya, U., et al., *Insights into the Effects of Cancer Associated Mutations at the UPF2 and*
1157 *ATP-Binding Sites of NMD Master Regulator: UPF1.* Int J Mol Sci, 2019. **20**(22).
1158 76. Morse, J.C., et al., *Elongation factor-Tu can repetitively engage aminoacyl-tRNA within the*
1159 *ribosome during the proofreading stage of tRNA selection.* Proceedings of the National Academy
1160 of Sciences, 2020. **117**(7): p. 3610-3620.
1161 77. Barends, S., et al., *Simultaneous and functional binding of SmpB and EF-Tu-TP to the alanyl*
1162 *acceptor arm of tmRNA.* J Mol Biol, 2001. **314**(1): p. 9-21.
1163 78. Himeno, H., D. Kurita, and A. Muto, *tmRNA-mediated trans-translation as the major ribosome*
1164 *rescue system in a bacterial cell.* Front Genet, 2014. **5**: p. 66.
1165 79. Guyomar, C., et al., *Structures of tmRNA and SmpB as they transit through the ribosome.* Nat
1166 Commun, 2021. **12**(1): p. 4909.
1167 80. Yang, H., et al., *PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas*
1168 *Endonuclease.* Cell, 2016. **167**(7): p. 1814-1828.e12.
1169 81. Zetsche, B., et al., *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System.*
1170 Cell, 2015. **163**(3): p. 759-771.
1171 82. Liu, J.-J., et al., *CasX enzymes comprise a distinct family of RNA-guided genome editors.* Nature,
1172 2019. **566**(7743): p. 218-223.
1173 83. Altae-Tran, H., et al., *The widespread IS200/IS605 transposon family encodes diverse*
1174 *programmable RNA-guided endonucleases.* Science, 2021. **374**(6563): p. 57-65.
1175 84. Sasnauskas, G., et al., *TnpB structure reveals minimal functional core of Cas12 nuclease family.*
1176 Nature, 2023. **616**(7956): p. 384-389.
1177 85. Karvelis, T., et al., *Transposon-associated TnpB is a programmable RNA-guided DNA*
1178 *endonuclease.* Nature, 2021. **599**(7886): p. 692-696.
1179 86. Chavez, D.A., B.H. Greer, and B.F. Eichman, *The HIRAN domain of helicase-like transcription*
1180 *factor positions the DNA translocase motor to drive efficient DNA fork regression.* J Biol Chem,
1181 2018. **293**(22): p. 8484-8494.
1182 87. Bessho, Y., et al., *Structural basis for functional mimicry of long-variable-arm tRNA by transfer-*
1183 *messenger RNA.* Proceedings of the National Academy of Sciences, 2007. **104**(20): p. 8293-8298.
1184 88. Gutmann, S., et al., *Crystal structure of the transfer-RNA domain of transfer-messenger RNA in*
1185 *complex with SmpB.* Nature, 2003. **424**(6949): p. 699-703.
1186 89. Giudice, E., K. Mace, and R. Gillet, *Trans-translation exposed: understanding the structures and*
1187 *functions of tmRNA-SmpB.* Frontiers in Microbiology, 2014. **5**.

1188 90. Gucinski, G.C., et al., *Convergent Evolution of the Barnase/EndoU/Colicin/RelE (BECR) Fold in*
1189 *Antibacterial tRNase Toxins.* Structure, 2019. **27**(11): p. 1660-1674.e5.
1190 91. Pedersen, K., et al., *The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the*
1191 *ribosomal A site.* Cell, 2003. **112**(1): p. 131-40.
1192 92. Guglielmini, J. and L. Van Melderen, *Bacterial toxin-antitoxin systems: Translation inhibitors*
1193 *everywhere.* Mob Genet Elements, 2011. **1**(4): p. 283-290.
1194 93. Goeders, N., P.L. Drèze, and L. Van Melderen, *Relaxed cleavage specificity within the RelE toxin*
1195 *family.* J Bacteriol, 2013. **195**(11): p. 2541-9.
1196 94. Willmund, F., et al., *The cotranslational function of ribosome-associated Hsp70 in eukaryotic*
1197 *protein homeostasis.* Cell, 2013. **152**(1-2): p. 196-209.
1198 95. Tsutakawa, S.E., H. Jingami, and K. Morikawa, *Recognition of a TG Mismatch: The Crystal*
1199 *Structure of Very Short Patch Repair Endonuclease in Complex with a DNA Duplex.* Cell, 1999.
1200 **99**(6): p. 615-623.
1201 96. Ipsaro, J.J., et al., *The structural biochemistry of Zucchini implicates it as a nuclease in piRNA*
1202 *biogenesis.* Nature, 2012. **491**(7423): p. 279-283.
1203 97. Mendez, A.S., et al., *Site specific target binding controls RNA cleavage efficiency by the Kaposi's*
1204 *sarcoma-associated herpesvirus endonuclease SOX.* Nucleic Acids Research, 2018. **46**(22): p.
1205 11968-11979.
1206 98. Laganeckas, M., M. Margelevicius, and C. Venclovas, *Identification of new homologs of PD-*
1207 *(D/E)XK nucleases by support vector machines trained on data derived from profile-profile*
1208 *alignments.* Nucleic Acids Res, 2011. **39**(4): p. 1187-96.
1209 99. Songailiene, I., et al., *HEPN-MNT Toxin-Antitoxin System: The HEPN Ribonuclease Is Neutralized*
1210 *by OligoAMPylation.* Molecular Cell, 2020. **80**(6): p. 955-970.e7.
1211 100. Yamamoto, S., et al., *ATPase Activity and ATP-dependent Conformational Change in the Co-*
1212 *chaperone HSP70/HSP90-organizing Protein (HOP)*.* Journal of Biological Chemistry, 2014.
1213 **289**(14): p. 9880-9886.
1214 101. Chang, Y.W., et al., *Crystal structures of the 70-kDa heat shock proteins in domain disjoining*
1215 *conformation.* J Biol Chem, 2008. **283**(22): p. 15502-11.
1216 102. Heinemann, U. and Y. Roske, *Cold-Shock Domains-Abundance, Structure, Properties, and*
1217 *Nucleic-Acid Binding.* Cancers (Basel), 2021. **13**(2).
1218 103. Patil, D.P., B.F. Pickering, and S.R. Jaffrey, *Reading m6A in the Transcriptome: m6A-Binding*
1219 *Proteins.* Trends in Cell Biology, 2018. **28**(2): p. 113-127.
1220 104. Makarova, K.S., et al., *Evolutionary and functional classification of the CARF domain superfamily,*
1221 *key sensors in prokaryotic antivirus defense.* Nucleic Acids Res, 2020. **48**(16): p. 8828-8847.
1222 105. McMahon, S.A., et al., *Structure and mechanism of a Type III CRISPR defence DNA nuclease*
1223 *activated by cyclic oligoadenylate.* Nature Communications, 2020. **11**(1): p. 500.
1224 106. Zhu, W., et al., *The CRISPR ancillary effector Can2 is a dual-specificity nuclease potentiating type*
1225 *III CRISPR defence.* Nucleic Acids Research, 2021. **49**(5): p. 2777-2789.
1226 107. Makarova, K.S., et al., *Live virus-free or die: coupling of antivirus immunity and programmed*
1227 *suicide or dormancy in prokaryotes.* Biology direct, 2012. **7**: p. 40-40.
1228 108. Koonin, E.V. and F. Zhang, *Coupling immunity and programmed cell suicide in prokaryotes: Life-*
1229 *or-death choices.* BioEssays, 2017. **39**(1): p. e201600186.
1230 109. Koonin, E.V. and M. Krupovic, *Origin of programmed cell death from antiviral defense?*
1231 Proceedings of the National Academy of Sciences, 2019. **116**(33): p. 16167.
1232 110. Kotta-Loizou, I., et al., *The RNA repair proteins RtcAB regulate transcription activator RtcR via its*
1233 *CRISPR-associated Rossmann fold domain.* iScience, 2022. **25**(11): p. 105425.
1234 111. Hughes, K.J., et al., *An RNA Repair Operon Regulated by Damaged tRNAs.* Cell Rep, 2020. **33**(12):
1235 p. 108527.

1236 112. Bitton, L., D. Klaiman, and G. Kaufmann, *Phage T4-induced DNA breaks activate a tRNA repair-*
1237 *defying anticodon nuclease.* Mol Microbiol, 2015. **97**(5): p. 898-910.
1238 113. van den Berg, D.F., et al., *Phage tRNAs evade tRNA-targeting host defenses through anticodon*
1239 *loop mutations.* eLife, 2023. **12**: p. e85183.
1240 114. Kaufmann, G., *Anticodon nucleases.* Trends Biochem Sci, 2000. **25**(2): p. 70-4.
1241 115. Ishita, J., et al., *tRNA anticodon cleavage by target-activated CRISPR-Cas13a effector.* bioRxiv,
1242 2021: p. 2021.11.10.468108.
1243 116. Genschik, P., K. Drabikowski, and W. Filipowicz, *Characterization of the Escherichia coli RNA 3'-*
1244 *terminal phosphate cyclase and its sigma54-regulated operon.* J Biol Chem, 1998. **273**(39): p.
1245 25516-26.
1246 117. Das, U. and S. Shuman, *2'-Phosphate cyclase activity of RtcA: a potential rationale for the operon*
1247 *organization of RtcA with an RNA repair ligase RtcB in Escherichia coli and other bacterial taxa.*
1248 Rna, 2013. **19**(10): p. 1355-62.
1249 118. Tanaka, N. and S. Shuman, *RtcB is the RNA ligase component of an Escherichia coli RNA repair*
1250 *operon.* J Biol Chem, 2011. **286**(10): p. 7727-7731.
1251 119. Kosmaczewski, S.G., et al., *The RtcB RNA ligase is an essential component of the metazoan*
1252 *unfolded protein response.* EMBO Rep, 2014. **15**(12): p. 1278-85.
1253 120. Popow, J., et al., *HSPC117 is the essential subunit of a human tRNA splicing ligase complex.*
1254 Science, 2011. **331**(6018): p. 760-4.
1255 121. Kirby, J.E., J.E. Trempy, and S. Gottesman, *Excision of a P4-like cryptic prophage leads to Alp*
1256 *protease expression in Escherichia coli.* J Bacteriol, 1994. **176**(7): p. 2068-81.
1257 122. Han, Y., et al., *Structure of human RNase L reveals the basis for regulated RNA decay in the IFN*
1258 *response.* Science, 2014. **343**(6176): p. 1244-8.
1259 123. Chakrabarti, A., B.K. Jha, and R.H. Silverman, *New insights into the role of RNase L in innate*
1260 *immunity.* J Interferon Cytokine Res, 2011. **31**(1): p. 49-57.
1261 124. Bösl, M. and H. Kersten, *A novel RNA product of the tyrT operon of Escherichia coli.* Nucleic Acids
1262 Res, 1991. **19**(21): p. 5863-70.
1263 125. Kohanski, M.A., et al., *A common mechanism of cellular death induced by bactericidal*
1264 *antibiotics.* Cell, 2007. **130**(5): p. 797-810.
1265 126. Rossi, C.C., et al., *A computational strategy for the search of regulatory small RNAs in*
1266 *Actinobacillus pleuropneumoniae.* Rna, 2016. **22**(9): p. 1373-85.
1267 127. Iyer, L.M., A.M. Burroughs, and L. Aravind, *The prokaryotic antecedents of the ubiquitin-*
1268 *signaling system and the early evolution of ubiquitin-like beta-grasp domains.* Genome Biol,
1269 2006. **7**(7): p. R60.
1270 128. Anderson, B.W., et al., *Molecular Mechanism of Regulation of the Purine Salvage Enzyme XPRT*
1271 *by the Alarmones pppGpp, ppGpp, and pGpp.* J Mol Biol, 2020. **432**(14): p. 4108-4126.
1272 129. Henry, J.T. and S. Crosson, *Ligand-binding PAS domains in a genomic, cellular, and structural*
1273 *context.* Annu Rev Microbiol, 2011. **65**: p. 261-86.
1274 130. Anantharaman, V., et al., *Comprehensive analysis of the HEPN superfamily: identification of*
1275 *novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing.* Biol Direct,
1276 2013. **8**: p. 15.
1277 131. Panne, D., et al., *The McrBC restriction endonuclease assembles into a ring structure in the*
1278 *presence of G nucleotides.* Embo j, 2001. **20**(12): p. 3210-7.
1279 132. Maracci, C. and M.V. Rodnina, *Review: Translational GTPases.* Biopolymers, 2016. **105**(8): p.
1280 463-75.
1281 133. Jewett, M.C., et al., *Continued protein synthesis at low [ATP] and [GTP] enables cell adaptation*
1282 *during energy limitation.* J Bacteriol, 2009. **191**(3): p. 1083-91.

1283 134. Rodnina, M.V., et al., *GTPases mechanisms and functions of translation factors on the ribosome.*
1284 Biol Chem, 2000. **381**(5-6): p. 377-87.
1285 135. Millman, A., et al., *Diversity and classification of cyclic-oligonucleotide-based anti-phage*
1286 *signalling systems.* Nature Microbiology, 2020. **5**(12): p. 1608-1615.
1287 136. Macdonald, E., et al., *Shield co-opts an RmuC domain to mediate phage defence across*
1288 *Pseudomonas species.* 2022: p. 2022.11.04.515146.
1289 137. Irastortza-Olaziregi, M. and O. Amster-Choder, *Coupled Transcription-Translation in Prokaryotes:*
1290 *An Old Couple With New Surprises.* Front Microbiol, 2020. **11**: p. 624830.
1291 138. Karousis, E.D. and O. Mühlemann, *Nonsense-Mediated mRNA Decay Begins Where Translation*
1292 *Ends.* Cold Spring Harb Perspect Biol, 2019. **11**(2).
1293 139. Serdar, L.D., et al., *Inhibition of post-termination ribosome recycling at premature termination*
1294 *codons in UPF1 ATPase mutants.* Elife, 2020. **9**.
1295 140. Kervestin, S. and A. Jacobson, *NMD: a multifaceted response to premature translational*
1296 *termination.* Nat Rev Mol Cell Biol, 2012. **13**(11): p. 700-12.
1297 141. Kim, Y.K. and L.E. Maquat, *UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA*
1298 *decay and beyond.* RNA, 2019. **25**(4): p. 407-422.

1299