

CoCoNuTs: A diverse subclass of Type IV restriction systems predicted to target RNA

Reviewed Preprint

Revised by authors after peer review.

[About eLife's process](#)

Reviewed preprint version 2

April 22, 2024 (this version)

Reviewed preprint version 1

February 2, 2024

Posted to preprint server

December 2, 2023

Sent for peer review

December 1, 2023

Ryan T. Bell , Harutyun Sahakyan, Kira S. Makarova, Yuri I. Wolf, Eugene V. Koonin 

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

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

A comprehensive census of McrBC systems, among the most common forms of prokaryotic Type IV restriction systems, followed by phylogenetic analysis, reveals their enormous abundance in diverse prokaryotes 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 transfer-messenger RNA (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 might additionally target DNA, via McrC nuclease homologs. Additional restriction systems, such as Type I RM, BREX, 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 (C RISPR-Associated Rossmann Fold) domains. We hypothesize that the CoCoNuTs, together with the ancillary restriction factors, employ an echeloned defense strategy analogous to that of Type III CRISPR-Cas systems, in which an immune response eliminating virus DNA and/or RNA is launched first, but then, if it fails, an abortive infection response leading to PCD/dormancy via host RNA cleavage takes over.

eLife assessment

This paper marks a **fundamental** advance in our understanding of prokaryotic Type IV restriction systems. The authors provide an encyclopedic overview of a hitherto uncharacterized branch of these systems, which they name CoCoNuTs, for coiled-coil nuclease tandems. They provide **compelling** evidence that these nucleases target RNA and are part of an echeloned defense response following viral infection. This article will be of great interest to scientists studying prokaryotic immunity mechanisms, as well as broadly to protein scientists engaged in the analysis, classification, and functional annotation of the proteome of life.

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 (Burroughs et al., 2015 [↗](#), Gao et al., 2020 [↗](#), Goldfarb et al., 2015 [↗](#), Bell et al., 2020 [↗](#), Koonin and Aravind, 2002 [↗](#), Swarts et al., 2014 [↗](#)). 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 (Gao et al., 2020 [↗](#), Bell et al., 2020 [↗](#), Anantharaman et al., 2012 [↗](#), Kaur et al., 2020 [↗](#)).

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 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 (Loenen et al., 2014 [↗](#), Fire et al., 1998 [↗](#), Agrawal et al., 2003 [↗](#), Makarova et al., 2006 [↗](#), Makarova et al., 2020b [↗](#), Gasiunas et al., 2012 [↗](#), Jinek et al., 2012 [↗](#)).

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 (Raleigh et al., 1989 [↗](#), Luria and Human, 1952 [↗](#), Fleischman et al., 1976 [↗](#), Dila et al., 1990 [↗](#)). Today, the prototypical McrBC system, native to *E. coli* 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 (Sutherland et al., 1992 [↗](#), Sukackaite et al., 2012 [↗](#)).

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 (Weigele and Raleigh, 2016 [↗](#), Loenen et al., 2014 [↗](#)). 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 [↗](#)) (Sutherland et al., 1992 [↗](#), Iyer et al., 2004 [↗](#), Nirwan et al., 2019 [↗](#), Panne et al., 1999 [↗](#)). 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 [↗](#)) (Niu et al., 2020 [↗](#), Sutherland et al., 1992 [↗](#)).

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 (Bell et al., 2020 [↗](#)). EVE superfamily domains, which in eukaryotes recognize modified DNA or RNA as part of mRNA maturation or epigenetic silencing functions, are often fused to

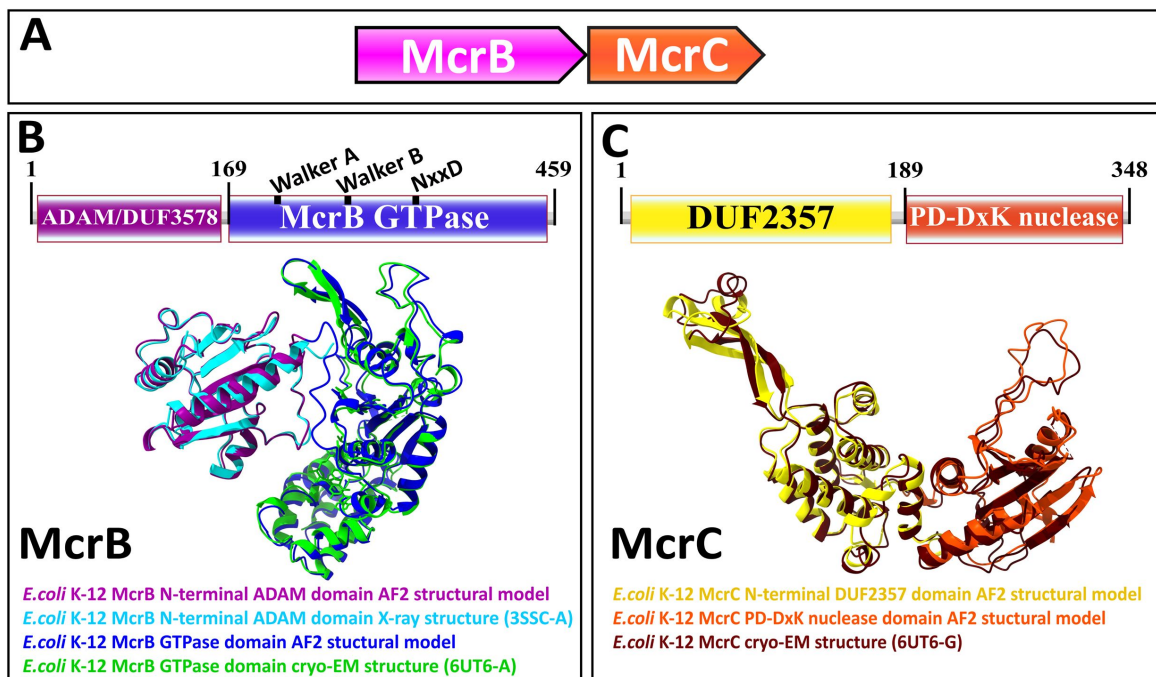


Figure 1

Genetic organization, signature sequence motifs, structural models, and phyletic distribution of McrB GTPases detected in this work

A) McrBC is a two-component restriction system with each component typically (except for extremely rare gene fusions) encoded by a separate gene expressed as a single operon, depicted here and in subsequent figures as arrows pointing in the direction of transcription. In most cases, McrB is the upstream gene in the operon. B) In the prototypical *E.coli* K-12 McrBC system, McrB contains an N-terminal methylcytosine-binding domain, ADAM/DUF3578, fused to a GTPase of the AAA+ ATPase clade (Sukackaite et al., 2012). This GTPase contains the Walker A and Walker B motifs that are conserved in P-loop NTPases as well as a signature NxxD motif, all of which are required for GTP hydrolysis (Nirwan et al., 2019; Niu et al., 2020; Pieper et al., 1999). An AlphaFold2 structural model of *E. coli* K-12 ADAM-McrB GTPase fusion protein monomer and separate X-ray diffraction and cryo-EM structures of the ADAM and GTPase domains (Niu et al., 2020; Sukackaite et al., 2012) show a high degree of similarity. C) McrC consists of a PD-DxK nuclease and an N-terminal DUF2357 domain, which comprises a helical bundle with a stalk-like extension that interacts with and activates individual McrB GTPases while they are assembled into hexamers (Niu et al., 2020; Nirwan et al., 2019). An AlphaFold2 structural model and cryo-EM structure of *E. coli* K-12 McrC monomer with DUF2357-PD-D/ExK architecture (Niu et al., 2020) show a high degree of similarity. The structures were visualized with ChimeraX (Pettersen et al., 2021).

McrB-like GTPases in prokaryotes, and indeed, these are the most frequently occurring EVE-containing fusion proteins (Bell et al., 2020 [DOI](#)). 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.

Results and Discussion

Comprehensive census of McrBC systems

The search for McrBC systems included PSI-BLAST runs against the non-redundant protein sequence database at the NCBI, followed by several filtering strategies (see Methods) to obtain a clean set of nearly 34,000, distributed broadly among prokaryotes. In the subsequent phase of analysis, GTPase domain sequences were extracted from the McrB homolog pool, and DUF2357 and PD-D/ExK nuclease domain sequences were extracted from the McrC homolog pool, leaving as a remainder the fused specificity domains that we intended to classify (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 [DOI](#), Supplementary Figure S1). Unexpectedly, we found that both the GTPase domains and DUF2357 domains frequently contained insertions into their coding sequences, likely encoding specificity domains and coiled-coils, respectively.

Removing variable inserts from the conserved McrBC domains allowed accurate, comprehensive phylogenetic analysis of the McrB GTPase (Figure 2 [DOI](#)) 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. coli* K-12 (blue in Figure 2 [DOI](#)) is characterized by frequent genomic association with Type I RM systems, usually with unidirectional gene orientations and the potential of forming a single operon (Raleigh, 1992 [DOI](#)). This branch and others, which exhibit two particularly prevalent associations, with a DISARM-like antiphage system (green in Figure 2 [DOI](#)) (Ofir et al., 2018 [DOI](#)), and, surprisingly for a Type IV restriction system, with predicted DNA methyltransferases (red in Figure 2 [DOI](#)), will be the subject of a separate, forthcoming publication.

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

A major branch (teal and yellow in Figure 2 [DOI](#)) of the McrBC family is characterized by a conserved deletion within the NxxD GTPase signature motif (where x is any amino acid), found in all McrB-like GTPases, reducing it to Nx D (Figure 1 B [DOI](#), Figure 2 [DOI](#)) (Neuwald et al., 1999 [DOI](#), Iyer et al., 2004 [DOI](#), Erzberger and Berger, 2006 [DOI](#), Niu et al., 2020 [DOI](#)). The Nx D variant of the motif is strictly conserved in these homologs and is usually, but not invariably, followed by a glutamate (E) or a second aspartate (D). No Nx D 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 (Pieper et al., 1999 [DOI](#)). 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 (Erzberger and Berger, 2006 [DOI](#), Colicelli, 2004 [DOI](#), Bourne et al., 1991 [DOI](#), Niu et al., 2020 [DOI](#), Pieper et al., 1999 [DOI](#)). A recent structural analysis has shown that the aspartate interacts with a conserved arginine/lysine residue in McrC which, via a hydrogen-bonding network, resituates the

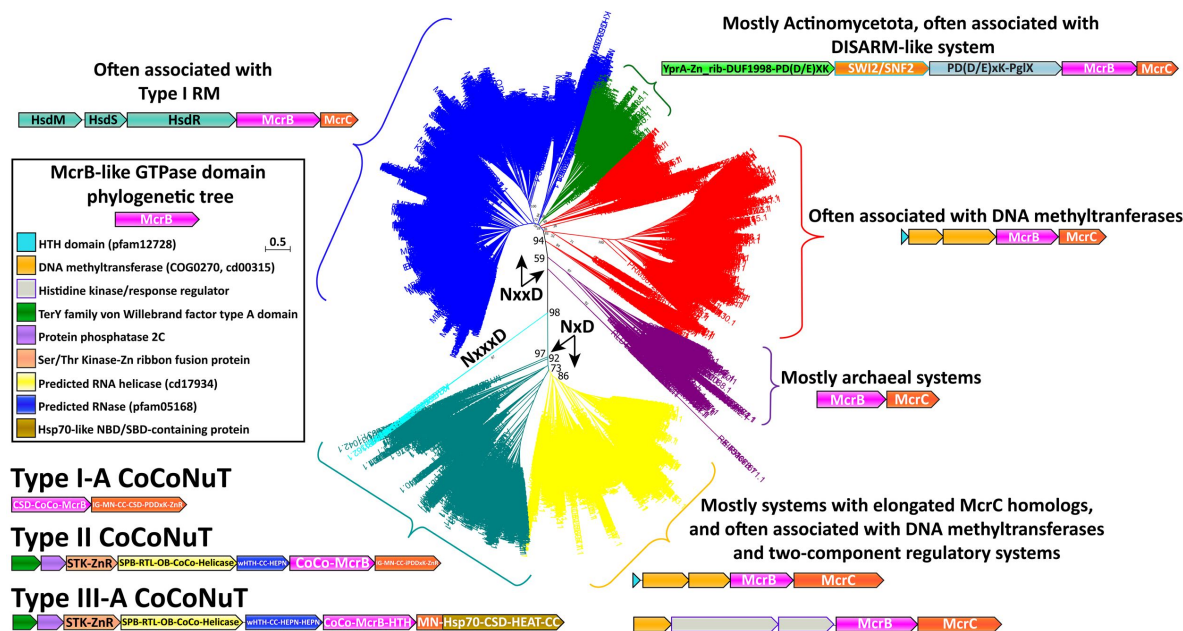


Figure 2

Phylogenetic tree of the McrB-like GTPases

The major clades in the phylogenetic tree of the McrB-like 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 Nx(x)D 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.

NxxD motif in relation to its interface with the Walker B motif such that, together, they optimally position a catalytic water to stimulate hydrolysis (Niu et al., 2020). 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. The Nx_D 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 Nx_D clade, which often occur in the same genomes (see below).

Many of these Nx_D GTPases are contextually associated with DNA methyltransferases, like the Nx_D 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 (yellow in **Figure 2**). We also detected a small number of GTPases with an insertion in the signature motif (cyan in **Figure 2**), expanding it to Nx_{xxx}D, 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 Nx_D branch of GTPases is a large clade characterized by fusion to long coiled-coil domains (teal in **Figure 2**, Supplementary Table S1), a derived and distinctly different architecture from the small, modular DNA-binding specificity domains typically found in canonical McrB homologs. The McrC homologs associated with these coiled-coil McrB GTPase fusions also often lack the PD-D/ExK endonuclease entirely or contain a region AlphaFold2 (AF2) predicted to adopt the PD-(D/E)xK endonuclease fold, but with inactivating replacements of catalytic residues (Supplementary Table S1). However, in the cases where the McrC nuclease is missing or likely inactive, these systems are always encoded in close association, often with overlapping reading frames, with HEPN (higher eukaryotes and prokaryotes nucleotide-binding) ribonuclease domains, or other predicted nucleases (**Figure 3**, Supplementary Figure S2) (Pillon et al., 2021). Based on these features of domain architecture and genomic context, we denote these uncharacterized McrBC-containing operons coiled-coil nuclease tandem (CoCoNuT) systems.

The deepest branching group of the Nx_D GTPase systems is typified by a fusion of an Hsp70-like ATPase nucleotide-binding domain and substrate-binding domain (NBD/SBD) to the McrB GTPase domain (purple in **Figure 3**, Supplementary Figure S2, Supplementary Table S1). Hsp70 is an ATP-dependent protein chaperone that binds exposed hydrophobic peptides and facilitates protein folding (Mayer, 2021). It also associates with AU-rich mRNA, and in some cases, such as the bacterial homolog DnaK, C-rich RNA, an interaction involving both the NBD and SBD (Kishor et al., 2017, Zimmer et al., 2001, Kishor et al., 2013). These systems might be functionally related to the CoCoNuTs, given that an analogous unit is encoded by a type of CoCoNuT system where similar domains are fused to the McrC homolog rather than to the McrB GTPase homolog (see below). In another large clade of CoCoNuT-like systems, the GTPase is fused to a domain homologous to FtsB, an essential bacterial cell division protein containing transmembrane and coiled-coil helices (Supplementary Figure S2, Supplementary Table S1) (Khadria and Senes, 2013). They are associated with signal peptidase family proteins likely to function as pilus assembly factors (Supplementary Table S1) (Colicelli, 2004) and usually contain coiled-coil domains fused to both the McrB and McrC homologs. Therefore, we denote them coiled-coil and pilus assembly linked to McrBC (CoCoPALM) systems (Supplementary Figure S2).

Here, we focus on the CoCoNuTs, whereas the CoCoPALMs and the rest of the Nx_D GTPase methyltransferase-associated homologs will be explored in a separate, forthcoming publication. 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**).

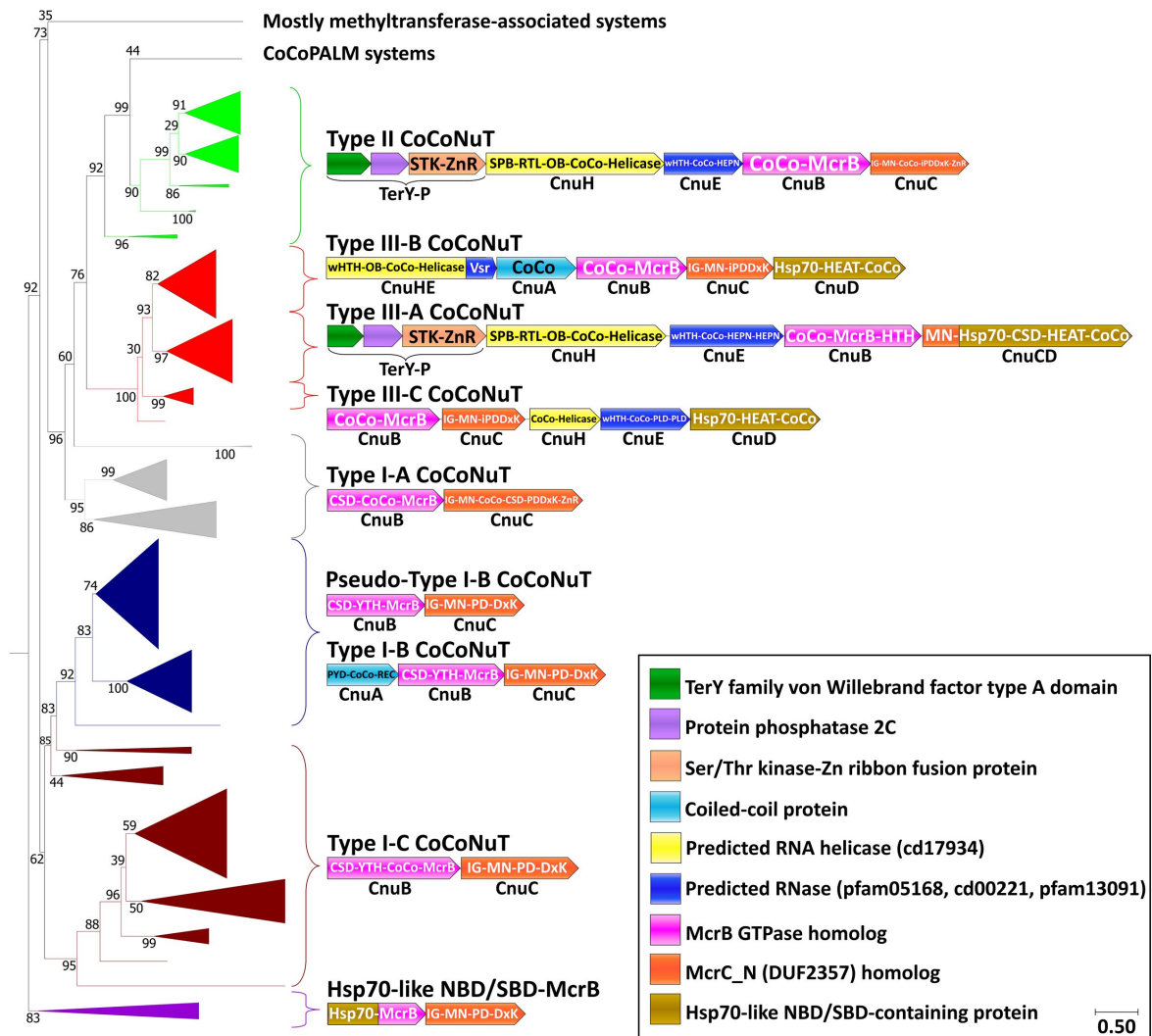


Figure 3

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 – 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.

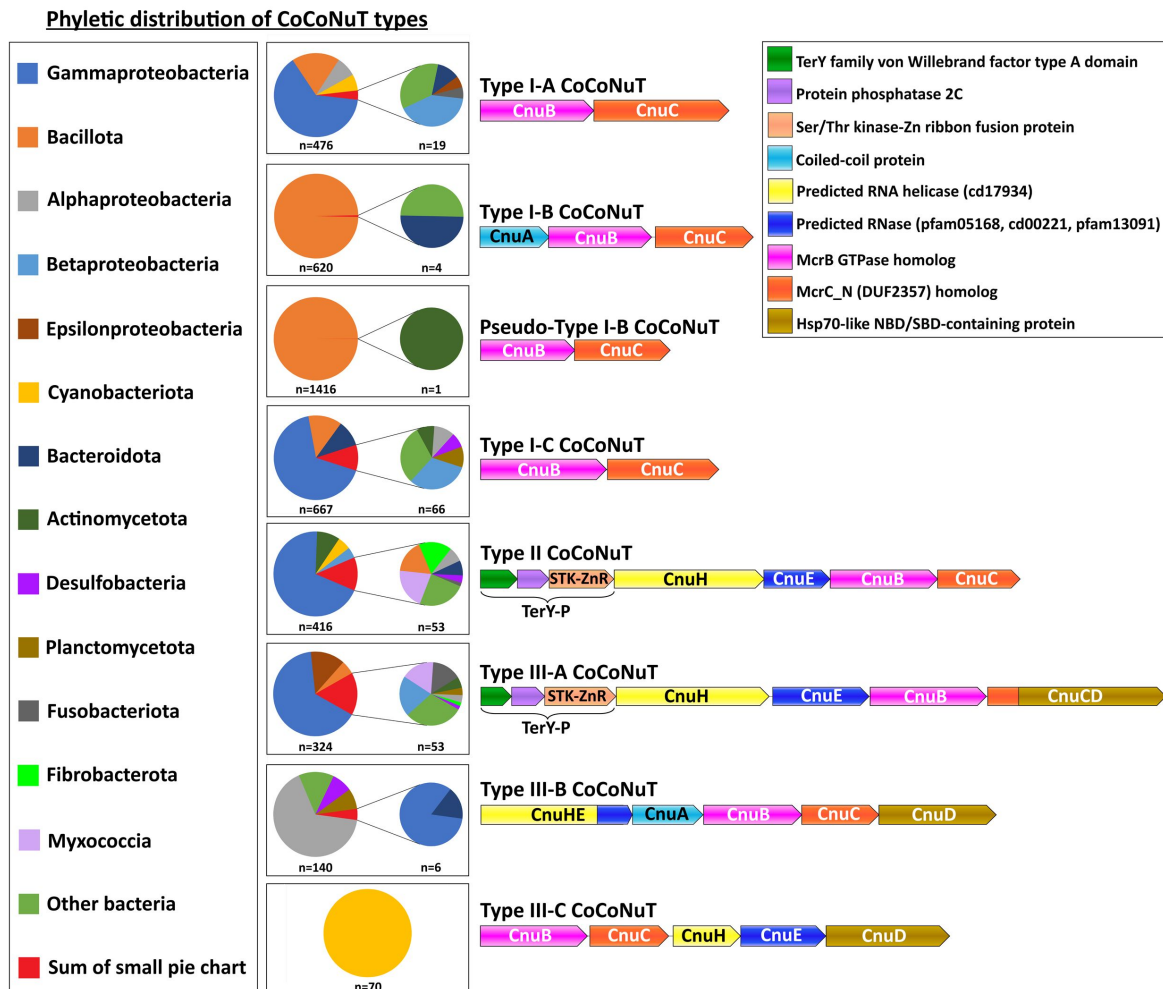


Figure 4

Phyletic distribution of CoCoNuTs

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

Type I CoCoNuT systems

We classified the CoCoNuTs into 3 types and 7 subtypes based on the GTPase domain phylogeny and conserved genomic context (**Figure 3**). Type I-A and Type I-C systems consist of McrB and McrC homologs only, which we denote CnuB and CnuC (**Figure 3**, **Figure 5**, Supplementary Figure S7, Supplementary Figure S9). Type I-A is distinguished from all other CoCoNuT types by a helical insert into the CnuB GTPase domain, between the Walker B and NxD motifs (**Figure 5**, Supplementary Figure S7).

Type I-B systems usually encode a separate coiled-coil protein, which we denote CnuA, in addition to CnuB/McrB and CnuC/McrC, with no coiled-coil fused to the McrB-like GTPase domain in CnuB (**Figure 3**, **Figure 5**, Supplementary Figure S8). CnuA is fused at the N-terminus to a pyrin (PYD)/CARD (caspase activation and recruitment domain)-like helical domain and at the C-terminus to a phosphoacceptor receiver (REC) domain (**Figure 5**, Supplementary Table S1). The association of the PYD/CARD-like domains with CoCoNuTs suggests involvement in a programmed cell death (PCD)/abortive infection-type response, as they belong to the DEATH domain superfamily and are best characterized in the context of innate immunity, inflammasome formation, and PCD (Park et al., 2007). The REC domains constitute one of the components of two-component regulatory systems. They are targeted for phosphorylation by histidine kinases (Stock et al., 2000), which could be a mechanism of Type I-B CoCoNuT regulation.

In contrast to Type II and Type III CnuB homologs, which contain only coiled-coils fused at their N-termini (**Figure 6**), cold-shock domain (CSD)-like OB-folds are usually fused at the N-termini of Type I CoCoNuT CnuB/McrB proteins, in addition to the coiled-coils (except for the majority of Type I-B, where the coiled-coils are encoded separately) (**Figure 5**, Supplementary Table S1) (Amir et al., 2018). Often, in Type I-B and I-C, but not Type I-A, a YTH-like domain, a member of the modified base-binding EVE superfamily, is present in CnuB as well, between the CSD and coiled-coil, or in Type I-B, between the CSD and the GTPase domain (**Figure 5**, Supplementary Table S1) (Bell et al., 2020, Hazra et al., 2019, Liao et al., 2018).

The CnuC/McrC proteins in Type I CoCoNuTs, as well as those in Type II and Type III-B and III-C, all contain an immunoglobulin-like N-terminal beta-sandwich domain of unknown function, not present in the *E. coli* K-12 McrC homolog, similar to a wide range of folds from this superfamily with diverse roles (**Figure 5**, Supplementary Table S1) (Natarajan et al., 2015, Halaby et al., 1999). It is also present in non-CoCoNuT McrC homologs associated with McrB GTPase homologs in the NxD clade, implying its function is not specific to the CoCoNuTs. In the Type I-B and I-C CoCoNuT CnuC homologs, these domains most closely resemble Rho GDP-dissociation inhibitor 1, suggesting that they may be involved in the regulation of CnuB/McrB GTPase activity (Dovas and Couchman, 2005).

We also detected a close relative of Type I-B CoCoNuT in many *Bacillus* species, which we denoted Pseudo-Type I-B CoCoNuT because it lost the separate coiled-coil protein CnuA. Pseudo-Type I-B CoCoNuT occurs in genomic contexts suggesting a role in overcrowding-induced stress responses (**Figure 3**, Supplementary Figure S4, Supplementary Note S1).

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 UPF1-like clade, which we denote CnuH (**Figure 3**, Supplementary Table S1) (Gorbalenya and Koonin, 1993, Fairman-Williams et al., 2010). The UPF1-like family encompasses helicases with diverse functions acting on RNA and single-stranded DNA (ssDNA) substrates, and notably,

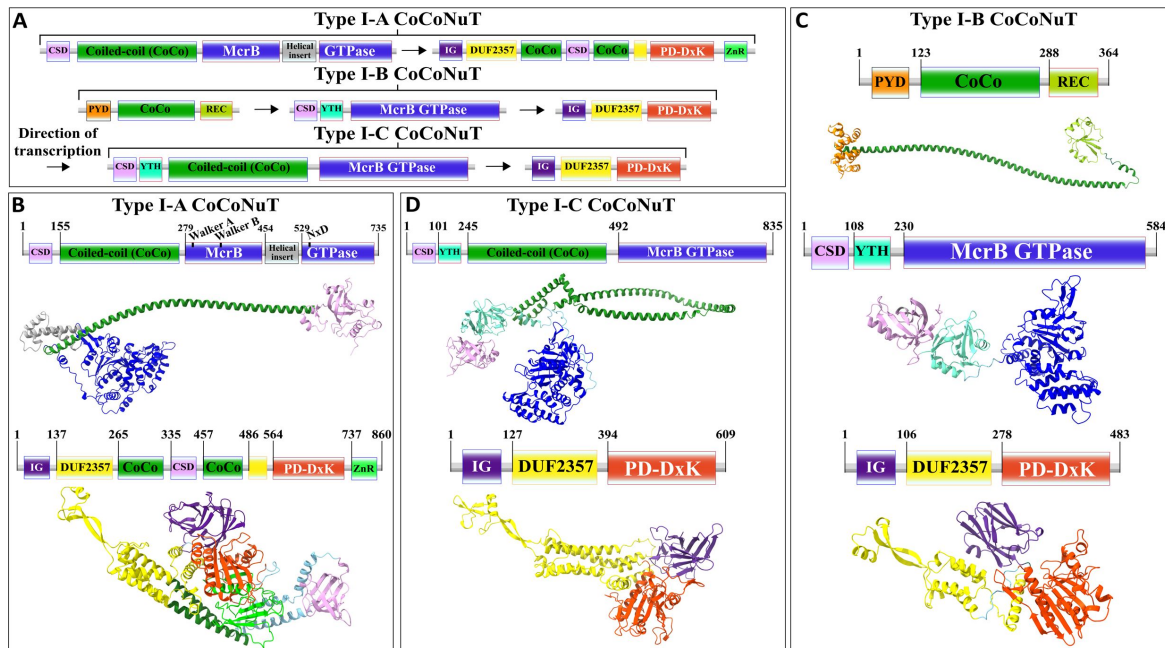


Figure 5

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 protein monomers in B) Type I-A CoCoNuT systems (CnuB and CnuC, from top to bottom), C) Type I-B (CnuA, CnuB, and CnuC, from top to bottom), and D) Type I-C CoCoNuT systems (CnuB and CnuC, from top to bottom). 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), TKH01449.1 (Type I-B CnuA), GED20858.1 (Type I-B CnuB), GED20857.1 (Type I-B CnuC), 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 (Pettersen et al., 2021 [DOI](#)).

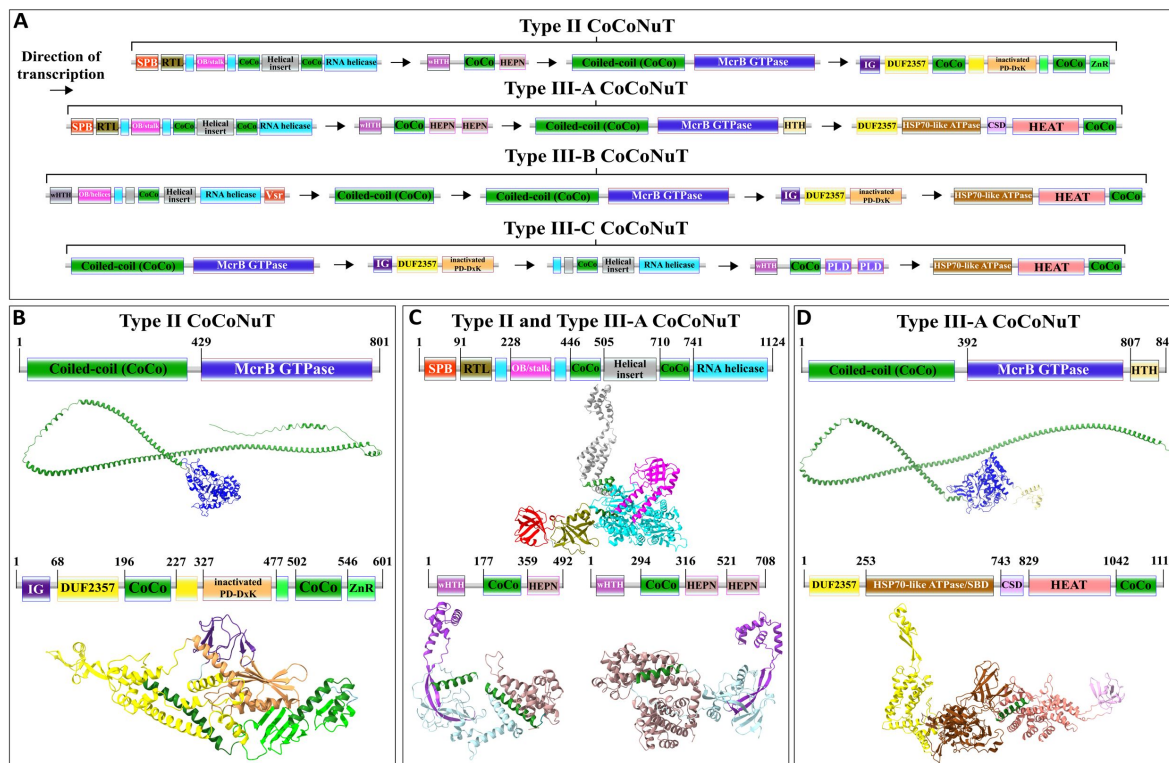


Figure 6

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 protein monomers in B) Type II CoCoNuT systems (CnuB and CnuC, from top to bottom), C) Type II and III-A CoCoNuT systems (CnuH at the top, Type II CnuE on the bottom left, Type III-A CnuE on the bottom right), and D) Type III-A CoCoNuT systems (CnuB and CnuC, from top to bottom). 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), PNG83940.1 (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 (Pettersen et al., 2021 [link](#)).

the prototypical UPF1 RNA helicase and its closest relatives are highly conserved in eukaryotes, where they play a critical role in the nonsense-mediated decay (NMD) RNA surveillance pathway (Cheng et al., 2007 [Chakrabarti et al., 2011](#)).

SF1 helicases are composed of two RecA-like domains which together harbor a series of signature motifs required for the ATPase and helicase activities, including the Walker A and Walker B motifs conserved in P-loop NTPases, which are located in the N-terminal RecA-like domain (Fairman-Williams et al., 2010). In all 4 Type II and Type III CoCoNuT subtypes, following the Walker A motif, the CnuH helicases contain a large helical insertion, with some of the helices predicted to form coiled-coils (**Figure 6**). The Type II, Type III-A, and Type III-B CoCoNuT CnuH helicases contain an OB-fold domain that, in Type II and Type III-A, is flanked by helices predicted to form a stalk-like helical extension of the N-terminal RecA-like domain, a structural feature characteristic of the entire UPF1/DNA2-like helicase family within SF1 (Chakrabarti et al., 2011, Zhou et al., 2015, Kalathiya et al., 2019) (**Figure 6**). DALI comparisons show that the CnuH predicted OB-fold domain in Type II CoCoNuTs is similar to the OB-fold domain in UPF1 and related RNA helicases SMUBP-2 and SEN1, and this holds for Type III-A as well, although, in these systems, the best DALI hits are to translation factor components such as EF-Tu domain II (Supplementary Table S1) (Chakrabarti et al., 2011, Morse et al., 2020). The Type III-B CnuH predicted OB-folds also match that of UPF1, albeit with lower statistical support (Supplementary Table S1).

In Type II and Type III-A CoCoNuTs, CnuH is fused at the N-terminus to a second OB-fold domain similar to that of SmpB (small protein B) (**Figure 6**, Supplementary Figure S5, Supplementary Table S1), which we denote SPB (SmpB-like). SmpB binds to SsrA RNA, also known as transfer-messenger RNA (tmRNA), and is required for tmRNA to rescue stalled ribosomes, via entry into their A-sites with its alanine-charged tRNA-like domain (Barends et al., 2001, Himeno et al., 2014, Guyomar et al., 2021). The SPB domain is around 20 amino acids shorter on average than SmpB itself, and a helix that is conserved in SmpB orthologs and interacts with the tmRNA is absent in the CoCoNuT OB folds (Supplementary Figure S5) (Bessho et al., 2007). However, two other structural elements of SmpB involved in binding tmRNA are present (Supplementary Figure S5) (Gutmann et al., 2003). The N-terminal OB-folds in Type II and III-A CnuH homologs also resemble prokaryotic HIRAN domains, which are uncharacterized, but have eukaryotic homologs fused to helicases that bind ssDNA (Supplementary Figure S5) (Chavez et al., 2018). These HIRAN domains, however, do not overlay with the CoCoNuT domains any better than SmpB (Supplementary Figure S5), lack a small helix that SmpB and the CoCoNuT domains share, and likely being DNA-binding, do not fit the other pieces of evidence we gathered that all suggest an RNA-binding role for this domain in the CoCoNuTs (see below).

The Type II and Type III-A CoCoNuT CnuH helicases contain an additional domain, structurally similar to the RelE/Colicin D RNase fold (Gucinski et al., 2019), which we denote RNase toxin-like (RTL), located between the SPB OB-fold and the helicase (**Figure 6**, Supplementary Table S1). The proteins of this family are ribosome-dependent toxins that cleave either mRNA or tRNA in the ribosomal A-site (Pedersen et al., 2003). This domain was identified by structural similarity search with the AF2 models, but no sequence conservation with characterized members of this family was detected, leaving it uncertain whether the RTL domain is an active nuclease. However, considerable divergence in sequence is not unusual in this toxin family (Guglielmini and Van Melder, 2011, Goeders et al., 2013). In Type III-B CoCoNuTs, a WHTH domain resembling the archaeal ssDNA-binding protein Sul7s is fused to the CnuH N-termini (**Figure 6**, Supplementary Table S1).

We searched for additional homologs of CnuH (see Methods). Our observations of the contextual associations, both of the CoCoNuTs and their relatives, showed that helicases of this large family are typically encoded in operons with downstream genes coding for an elongated WHTH domain fused at its C-terminus to a variety of effectors, generally, nucleases, which we denote CnuE (**Figure 3**, **Figure 6**). In the CoCoNuTs, except for Types III-B and III-C, these CnuE effectors

are HEPN ribonucleases, with one HEPN domain in Type II and two in Type III-A (**Figure 6**, Supplementary Table S1). In Type III-B, the effector is a Vsr (very-short-patch repair)-like PD-(D/E)xK family endonuclease (Tsutakawa et al., 1999) fused directly to the helicase, with no wHTH domain present, whereas in Type III-C, a distorted version of the elongated wHTH domain is fused to two Phospholipase D (PLD) family endonuclease domains (**Figure 6**, Supplementary Table S1). All these nucleases can degrade RNA, and some, such as HEPN, have been found to cleave RNA exclusively (Pillon et al., 2021, Ipsaro et al., 2012, Mendez et al., 2018, Laganeckas et al., 2011, Songailiene et al., 2020). We also detected coiled-coils in the region between the wHTH and effector domains in Type II, Type III-A, and Type III-C CoCoNuT CnuE homologs. Multimer structural modeling with AlphaFold2 suggests that the wHTH domain in CnuE might interact with the coiled-coil-containing helical insertion of CnuH, perhaps mediated by the coiled-coils in each protein, to couple the ATP-driven helicase activity to the various nuclease effectors (Supplementary Figure S11). The accuracy of this model notwithstanding, the fusion of the Vsr-like effector to CnuH in Type III-B CoCoNuTs and the similarity of this system to the other CoCoNuT types strongly suggests that the CnuE effector proteins in these systems form complexes with their respective CnuH helicases. An additional factor in potential complexing by these proteins is the presence of coiled-coils in the associated CnuB/McrB and CnuC/McrC homologs, which may interact with the coiled-coils in CnuH and CnuE. Type III-B CoCoNuTs also code for a separate coiled-coil protein, which we denote CnuA, as it resembles the CnuA protein encoded in Type I-B CoCoNuTs (**Figure 3**, **Figure 5**, **Figure 6**). However, it is distinguished from Type I-B CnuA in containing no recognizable domains other than the coiled-coil (**Figure 3**, **Figure 5**, **Figure 6**). This also could potentially interact with other coiled-coil proteins in the system.

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 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 (Anantharaman et al., 2012). 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 (Gao et al., 2020). 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, with a few exceptions among Terrabacterial Type II systems and Myxococcal Type III-A systems, implying an important contribution to their function (Supplementary Table S3). However, the complete absence of the TerY-P module in Type III-B and Type III-C systems implies that when different nucleases and other putative effectors fused to the helicase are present, TerY-P is dispensable for the CoCoNuT activity. Therefore, we do not consider them to be core components of these systems.

Type II and Type III-A CoCoNuTs have similar domain compositions, but a more detailed comparison reveals substantial differences. The CnuE proteins in Type II contain one HEPN domain with the typical RxxxxH RNase motif conserved in most cases, whereas those in Type III-A contain two HEPN domains, one with the RxxxxH motif, and the other, closest to the C-terminus, with a shortened RxH motif. Furthermore, Type III-A CnuB/McrB GTPase homologs often contain C-terminal HTH-domain fusions absent in Type II (**Figure 6**). Finally, striking divergence has occurred between the Type II and III-A CnuC/McrC homologs. Type II CnuCs resemble Type I-A CnuCs, which contain N-terminal immunoglobulin-like beta-sandwich domains, PD-D/ExK nucleases, zinc ribbon domains, and insertions into DUF2357 containing coiled-coils and a CSD-like OBD. However, in Type II CnuCs, this domain architecture underwent reductive evolution (**Figure 5**, **Figure 6**, Supplementary Table S1). In particular, the beta-sandwich domain and coiled-coils are shorter, the CSD was lost, the nuclease domain was inactivated, and in many cases, the number of Zn-binding CPxC motifs was reduced from three to two (**Figure 5**, **Figure 6**,

Supplementary Table S1). This degeneration pattern could indicate functional replacement by the associated CnuH and CnuE proteins, often encoded in reading frames overlapping with the start of the *cnuBC* operon.

By contrast, Type III-A CnuC/McrC homologs entirely lost the beta-sandwich domain, PD-D/ExK nuclease, and zinc ribbon found in Type I-A and Type II, but gained an Hsp70-like NBD/SBD unit similar to those that are fused to the McrB GTPase in early branching members of the NxD clade. They have also acquired a helical domain similar to the HEAT repeat family, and in some cases, a second CSD (**Figure 3**, **Figure 5**, **Figure 6**, Supplementary Table S1). Most Type III-A CnuCs contain a CSD and coiled-coils, and thus, resemble Type I-A, but the positioning of these domains, which in Type I-A are inserted into the DUF2357 helix bundle, is not conserved in Type III-A, where these domains are located outside DUF2357 (**Figure 5**, **Figure 6**).

Hsp70-like ATPase NBD/SBD domains and HEAT-like repeats are fused to the CnuC/McrC N-terminal DUF2357 domain in Type III-A CoCoNuT, but their homologs in Type III-B and III-C are encoded by a separate gene. We denote these proteins CnuD and CnuCD, the latter for the CnuC-CnuD fusions in Type III-A. The separation of these domains in Type III-B and III-C implies that fusion is not required for their functional interaction with the CnuBC/McrBC systems (**Figure 6 A**, Supplementary Table S1). CnuD proteins associated with both Type III-B and III-C contain predicted coiled-coils, suggesting that they might interact with the large coiled-coil in the CnuB homologs (**Figure 6 A**).

In the CnuD homologs found in the CoCoNuTs and fused to NxD McrB GTPases, Walker B-like motifs (Yamamoto et al., 2014) are usually, but not invariably, conserved, whereas the sequences of the helical domains adjacent to the motifs are more strongly constrained. Walker A-like motifs (Chang et al., 2008) are present but degenerate (Supplementary Figure S13). Therefore, it appears likely that the CoCoNuT CnuD homologs bind ATP/ADP but hydrolyze ATP with extremely low efficiency, at best. Such properties in an Hsp70-like domain are better compatible with RNA binding than unfolded protein binding or remodeling, suggesting that these CnuD homologs may target the respective systems to aberrant RNA. Many Hsp70 homologs have been reported to associate with ribosomes (Mayer, 2021, Willmund et al., 2013), which could also be true for the CoCoNuT Cnu(C)Ds.

Consistent with this prediction, Type II and III-A CoCoNuTs likely target RNA rather than DNA, given that the respective operons encode HEPN RNases, typically the only recognizable nuclease in these systems. All CnuC/McrC homologs in Type II or III CoCoNuT systems lack the PD-D/ExK catalytic motif that is required for nuclease activity, although for Type II and Type III-B and III-C, but not Type III-A, structural modeling indicates that the inactivated nuclease domain was retained, likely for a nucleic acid-binding role (**Figure 6**, Supplementary Table S1). In Type

III-A CoCoNuT, the nuclease domain was lost entirely and replaced by the Hsp70-like NBD/SBD domain with RNA-binding potential described above (**Figure 6**, Supplementary Table S1). Often, one or two CSDs, generally RNA-binding domains, although capable of binding ssDNA, are fused to Type III-A CnuC homologs as well (**Figure 6**, Supplementary Table S1) (Heinemann and Roske, 2021). RNA targeting capability of Type III-B and III-C can perhaps be inferred from their similarity to Type III-A in encoding CnuD Hsp70-like proteins. Moreover, higher-order associations of Type II and Type III-A CoCoNuT systems with various DNA restriction systems suggest a two-pronged DNA and RNA restriction strategy reminiscent of Type III CRISPR-Cas (see below).

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

1. Most of the CoCoNuts encompass HEPN nucleases that appear to possess exclusive specificity for RNA.

2. CSD-like OB-folds are pervasive in these systems, being present in the CnuB/McrB homologs of all Type I subtypes and in Type I-A and Type III-A CnuC/McrC homologs. As previously noted, these domains typically bind RNA, although they could bind ssDNA as well.
3. YTH-like domains are present in most Type I CoCoNuTs, particularly, in almost all early branching Type I-B and I-C systems, suggesting that the common ancestor of the CoCoNuTs contained such a domain. YTH domains in eukaryotes sense internal N6-methyladenosine (m6A) in mRNA (Hazra et al., 2019 [↗](#), Liao et al., 2018 [↗](#), Patil et al., 2018 [↗](#)).
4. Type II and Type III-B/III-C CoCoNuTs, which likely target RNA, given the presence of HEPN domains and Hsp70 NBD/SBD homologs, retain inactivated PD-DxK nuclease domains, suggesting that these domains contribute an affinity for RNA inherited from Type I-A CoCoNuTs. PD-(D/E)xK nucleases are generally DNA-specific, however, some examples of RNase activity have been reported (Mendez et al., 2018 [↗](#), Laganeckas et al., 2011 [↗](#)). The inactivated PD-DxK domains might also bind DNA from which the target RNA is transcribed.

Complex higher-order associations between CoCoNuTs, CARF domains, and other defense systems

Genomic neighborhoods of many Type II and Type III-A CoCoNuTs encompass complex operonic associations with genes encoding several types of CARF (CRISPR-Associated Rossmann Fold) domain-containing proteins. This connection suggests multifarious regulation by cyclic (oligo)nucleotide second messengers synthesized in response to viral infection and bound by CARF domains (Makarova et al., 2020a [↗](#), McMahon et al., 2020 [↗](#), Zhu et al., 2021 [↗](#)). Activation of an effector, most often a nuclease, such as HEPN or PD-(D/E)xK, by a CARF bound to a cyclic (oligo)nucleotide is a crucial mechanism of CBASS (cyclic oligonucleotide-based antiphage signaling system) as well as Type III CRISPR-Cas systems (McMahon et al., 2020 [↗](#), Makarova et al., 2020a [↗](#), Zhu et al., 2021 [↗](#)). These CARF-regulated enzymes generally function as a fail-safe that eventually induces PCD/dormancy when other antiphage defenses fail to bring the infection under control and are deactivated, typically through cleavage of the second messenger by a RING nuclease, if other mechanisms succeed (Makarova et al., 2012 [↗](#), Koonin and Zhang, 2017 [↗](#), Makarova et al., 2020a [↗](#), Koonin and Krupovic, 2019 [↗](#)).

The presence of CARFs could implicate the HEPN domains of these systems as PCD effectors that would carry out non-specific RNA degradation in response to infection. Surprisingly, however, most of these CARF domain-containing proteins showed the highest similarity to RtcR, a sigma54 transcriptional coactivator of the RNA repair system RtcAB with a CARF-ATPase-HTH domain architecture, suggesting an alternative functional prediction (Supplementary Figure S14, Supplementary Table S1). Specifically, by analogy with RtcR, CoCoNuT-associated CARF domain-containing proteins might bind (t)RNA fragments with 2',3' cyclic phosphate ends (Kotta-Loizou et al., 2022 [↗](#), Hughes et al., 2020 [↗](#)). This interaction could promote transcription of downstream genes, in this case, genes encoding CoCoNuT components, through binding an upstream activating sequence by the HTH domain fused to the CARF-ATPase C-terminus (Hughes et al., 2020 [↗](#), Kotta-Loizou et al., 2022 [↗](#)).

The manifold biological effects of tRNA-like fragments are only beginning to be appreciated. Lately, it has been shown that bacterial anticodon nucleases, in response to infection and DNA degradation by phages, generate tRNA fragments, likely a signal of infection and a defensive strategy to slow down the translation of viral mRNA, and that phages can deploy tRNA repair enzymes and other strategies to counteract this defense mechanism (Bitton et al., 2015 [↗](#), van den Berg et al., 2023 [↗](#), Kaufmann, 2000 [↗](#), Ishita et al., 2021 [↗](#)). Moreover, the activity of the HEPN ribonucleases in the CoCoNuTs themselves would produce RNA cleavage products with cyclic phosphate ends that might be bound by the associated RtcR-like CARFs (Shigematsu et al., 2018 [↗](#), Pillon et al., 2021 [↗](#)), in a potential feedback loop.

Such a CoCoNuT mechanism could complement the function of RtcAB, as RtcA is an RNA cyclase that converts 3L-phosphate RNA termini to 2L,3L-cyclic phosphate and thus, in a feedback loop, generates 2',3'-cyclic phosphate RNA fragments that induce expression of the operon (Genschik et al., 1998 [↗](#), Das and Shuman, 2013 [↗](#), Hughes et al., 2020 [↗](#)). Acting downstream of RtcA, RtcB is an RNA ligase that joins 2',3'-cyclic phosphate RNA termini to 5'-OH RNA fragments, generating a 5'-3' bond and, in many cases, reconstituting a functional tRNA (Tanaka and Shuman, 2011 [↗](#)). Recent work has shown that, in bacteria, the most frequent target of RtcB is SsrA, the tmRNA (Kotta-Loizou et al., 2022 [↗](#)). Intriguingly, as described above, in order to rescue stalled ribosomes, tmRNA must bind to SmpB, an OB-fold protein highly similar to the predicted structures of the SPB domains fused at the N-termini of the CnuH helicases in Type II and III-A CoCoNuTs, which are the only types that frequently associate with RtcR homologs (Supplementary Figure S14, **Figure 7** [↗](#)) (Himeno et al., 2014 [↗](#), Guyomar et al., 2021 [↗](#)).

There are notable parallels between CoCoNuTs and Type III CRISPR-Cas systems, where the Cas10-Csm-crRNA effector complex binds phage RNA complementary to the spacer of the crRNA, triggering both restriction of phage DNA and indiscriminate cleavage of RNA (McMahon et al., 2020 [↗](#)). The target RNA recognition stimulates the production of cyclic oligoadenylate (cOA) signals by Cas10, which activate, via CARF domain binding, PCD effectors, typically HEPN domain-containing proteins, such as Csm6, that function as promiscuous RNases (Pillon et al., 2021 [↗](#), Millman et al., 2020 [↗](#)). One of the two outcomes can result from this cascade: the infection can either be eradicated quickly due to the restriction of the virus DNA, which inhibits cOA signaling via the depletion of viral RNA, along with the activity of RING nucleases, thus averting PCD, or the continued presence of viral RNA stimulates cOA signaling until PCD or dormancy occurs, limiting the spread of viruses to neighboring cells in the bacterial population (Makarova et al., 2012 [↗](#), Koonin and Aravind, 2002 [↗](#), Koonin and Krupovic, 2019 [↗](#)).

If CoCoNuTs associated with RtcR homologs can induce PCD, a conceptually similar but mechanistically distinct phenomenon might occur. Although many of these CoCoNuTs only contain an appended gene encoding a CARF domain-containing protein at the 5' end of the predicted operon (Supplementary Figure S14), there are also numerous cases where several types of DNA restriction systems are encoded between the CARF gene and the CoCoNuT (**Figure 7** [↗](#)). In these cases, nearly all genes are in an apparent operonic organization that can extend upwards of 40 kb (**Figure 7** [↗](#)). Although internal RtcR-independent promoters likely exist in these large loci, the consistent directionality and close spacing of the genes in these superoperons suggests coordination of expression. The complex organization of the CARF-CoCoNuT genomic regions, and by implication, the corresponding defense mechanisms, might accomplish the same effect as Type III CRISPR-Cas, contriving a no-win situation for the target virus. Under this scenario, the virus is either destroyed by the activity of the DNA restriction systems, which would inhibit signal production (likely RNA fragments with cyclic phosphate ends rather than cOA) and drive down CoCoNuT transcription, thereby preventing PCD, or as the virus replicates, signaling and CoCoNuT transcription would continue until the infection is aborted by PCD or dormancy caused by the degradation of host mRNA via the HEPN RNase(s) of the CoCoNuT.

In many species of *Pseudomonas*, where CoCoNuTs are almost always associated with RtcR and various ancillary factors, Type I RM systems often contain an additional gene that encodes a transmembrane helix and a long coiled-coil fused to an RmuC-like nuclease (**Figure 7** [↗](#)). These proteins were recently described as ShdA, the core component of the *Pseudomonas*-specific defense system Shield (Macdonald et al., 2022 [↗](#)). These can potentially interact with coiled-coils in the CoCoNuTs, perhaps, guiding them to the DNA from which RNA targeted by the CoCoNuTs is being transcribed, or vice versa (**Figure 7** [↗](#)).

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 viral RNA recognition by the Cas10-Csm complex, rather than binding of a second messenger to a CARF

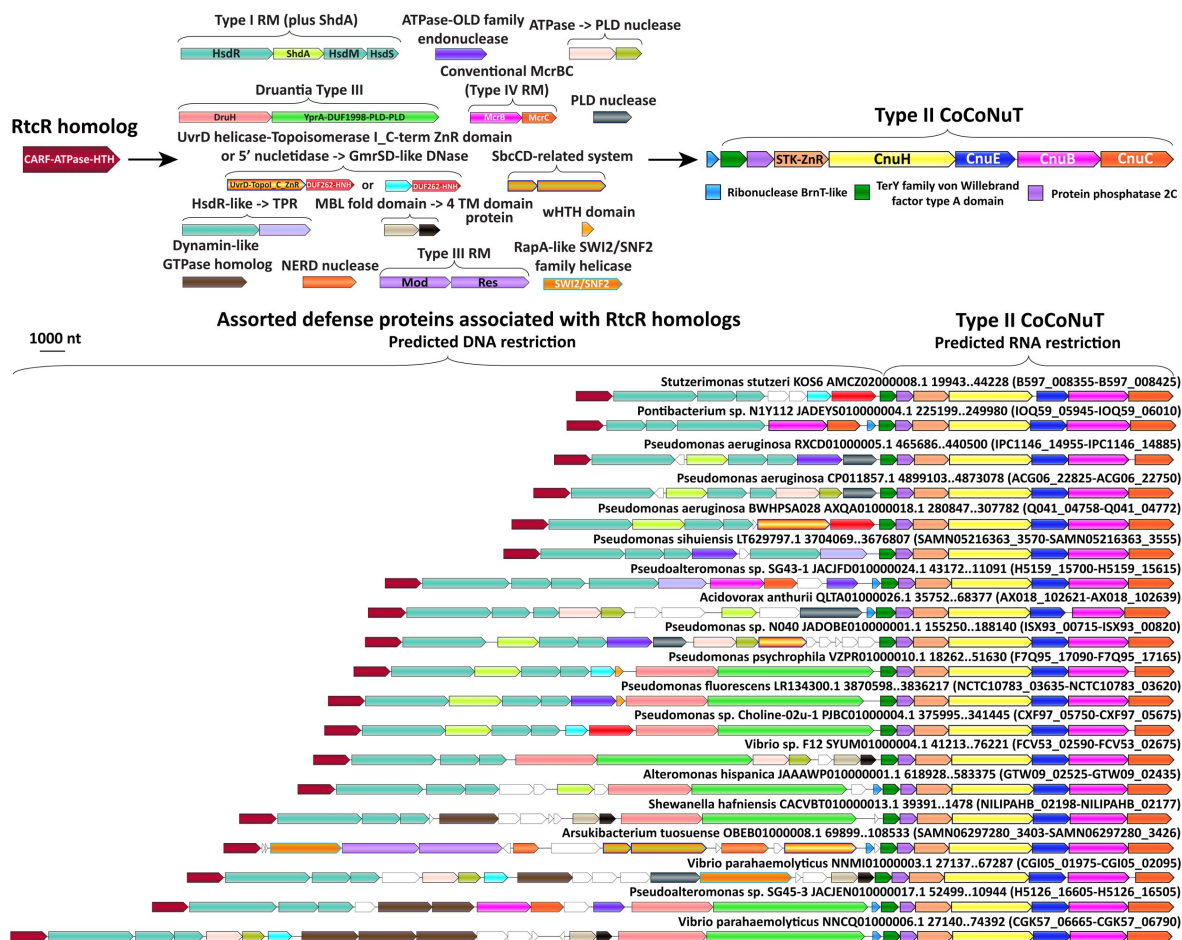


Figure 7

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; MBL fold – Metallo-beta-lactamase fold; 4 TM domain – Protein with 4 predicted transmembrane helices; Mod/Res – Type III RM modification and restriction factors.

domain, activates both DNA cleavage activity by the HD domain and production of cOA that triggers non-specific RNA cleavage. In contrast, in the CARF-containing CoCoNuTs, both the DNA and RNA restriction factors appear to be arranged such that binding of a 2',3' cyclic phosphate RNA fragment by the CARF domain would initiate the expression of the entire cluster (McMahon et al., 2020). In the case of the CoCoNuTs, signals of infection could promote transcription, first of the DNA restriction systems, and then, the CoCoNuT itself, a predicted RNA restriction system. In these complex configurations of the CoCoNuT genome neighborhoods (Figure 7), the gene order is likely to be important, with Type I RM almost always directly following CARF genes and CoCoNuTs typically coming last, although *Druantia* Type III sometimes follows the CoCoNuT. As translation in bacteria is co-transcriptional, the products of genes transcribed first would accumulate before those of the genes transcribed last, so that the full, potentially suicidal impact of the CoCoNuT predicted RNA nucleolytic engine would only be felt after the associated DNA restriction systems had ample time to act – and possibly, fail (Irastorza-Olaziregi and Amster-Choder, 2020).

Concluding remarks

In recent years, systematic searches for defense systems in prokaryotes, primarily by analysis of defense islands, revealed enormous, previously unsuspected diversity of such systems that function through a remarkable variety of molecular mechanisms. In this work, we uncovered the hidden diversity and striking hierarchical complexity of a distinct class of defense mechanisms, the Type IV (McrBC) restriction systems. We then zeroed in on a single major but previously overlooked branch of the McrBC systems, which we denoted CoCoNuTs for their salient features, namely, the presence of extensive coiled-coil structures and tandem nucleases. Astounding complexity was discovered at this level as well, with 3 distinct types and multiple subtypes of CoCoNuTs that differ by their domain compositions and genomic associations. All CoCoNuTs contain domains capable of interacting with translation systems components, such as the SmpB-like OB-fold, Hsp70 homologs, or YTH domains, along with RNases, such as HEPN, suggesting that at least one of the activities of these systems targets RNA. Most of the CoCoNuTs are likely endowed with DNA-targeting activity as well, either by factors integral to the system, such as the McrC nuclease, or more loosely associated, such as Type I RM and *Druantia* Type III systems that are encoded in the same predicted superoperon with many CoCoNuTs. Numerous CoCoNuTs are associated with proteins containing CARF domains, suggesting that cyclic (oligo)nucleotides regulate the CoCoNuT activity. Given the presence of the RtcR-like CARF domains, it appears likely that the specific second messengers involved are RNA fragments with cyclic phosphate termini. We hypothesize that the CoCoNuTs, in conjunction with ancillary restriction factors, implement an echeloned defense strategy analogous to that of Type III CRISPR-Cas systems, whereby an immune response eliminating virus DNA and/or RNA is launched first, but then, if it fails, an abortive infection response leading to PCD/dormancy via host RNA cleavage takes over.

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 (Bell et al., 2020). 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) (Altschul et al., 1997). Although a branch of McrB GTPase homologs has been described in animals, these are highly divergent in function, and no associated McrC homologs have been reported (Iyer et al., 2004). Therefore, we excluded eukaryotic sequences from our analysis to focus on the composition and contextual connections of prokaryotic McrBC systems.

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 for validation of the BLAST hits or if no hits were obtained. (Gabler et al., 2020). 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 (Zimmermann et al., 2018a). 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 (Hauser et al., 2016), after which the sequences in each cluster were aligned with MUSCLE (Edgar, 2004). Next, profile-to-profile similarity scores between all clusters were calculated with HHsearch (Söding, 2004). Clusters with high similarity, defined as a pairwise score to self-score ratio > 0.1 , were aligned to each other with HHalign (Söding et al., 2006). This procedure was performed for a total of 3 iterations. The alignments of each cluster resulting from this protocol, which included some false positive clusters consisting of other members of the AAA_5 family, were analyzed with HHpred to remove the 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 above) 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 same direction as the hit, and encoding a protein of sufficient size (> 200 aa for McrB, > 150 aa for McrC) to be the undetected McrB or McrC component. Afterward, we filtered the remaining data to retain genome islands with no McrBC aliases but with PSI-BLAST hits in operonic association with genes of sufficient size, as described above, to be the other McrBC component. These data, which contained many false positives but captured many rare variants, were then clustered and analyzed as described above to remove false positives. Next, an automated procedure was developed to excise the GTPase domain sequences using the manual alignments generated during the first phase of the search as a reference. These GTPase sequences were further analyzed by clustering and HHpred to remove false positives.

Definitive validation by pairing McrB homologs with their respective McrC homologs was also used to corroborate their identification. Occasionally, the McrB and McrC homologs were separated by intervening genes, or the operon order was reversed, and consideration of those possibilities allowed the validation of many additional systems. The pairing process was complicated by and drew our attention to the frequent occurrence of multiple copies of McrBC

systems in the same islands that may function cooperatively. Lastly, the rigorously validated set of McrBC pairs, supplemented only with orphans manually annotated as McrBC components using HHpred, were used for our phylogenetics. The final alignments of GTPase and DUF2357 domains were produced using the iterative alignment procedure described above for 10 iterations. Approximately-maximum-likelihood trees were built with the FastTree program (Price et al., 2010) from representative sequences following clustering to a 0.9 similarity threshold with MMseqs2.

Domain detection and annotation

Protein domains in McrBC homologs and in proteins encoded by neighboring genes were initially identified using the method described above, the first pass using PSI-BLAST against alignments of domains in the NCBI Conserved Domain Database (CDD) and Pfam (*E*-value 0.001). In many cases where no domains could be confidently detected with this method, or for validation of the hits from the first pass, HHpred was used for more sensitive analysis (Zimmermann et al., 2018b). The CoCoNuT system components were subjected to additional scrutiny using the coiled-coil detection and visualization tool Waggawagga, which employs a several algorithms for coiled-coil prediction, including Marcoil, Multicoil2, Ncoils, and Paircoil2 (Simm et al., 2015, Delorenzi and Speed, 2002, Trigg et al., 2011, Lupas et al., 1991, McDonnell et al., 2006). These predictions varied in their strength, with the long coiled-coils detected in CnuA and CnuB homologs having the highest likelihood (usually the maximum P-score of 100 with Marcoil and Multicoil2) and being recognized by the most of the applied tools (BLAST, HHpred, and multiple algorithms used by Waggawagga). The shorter coiled-coils in CnuC, CnuD, and CnuH homologs were less strongly, but nevertheless confidently predicted, usually being detected by HHpred and by at least one but typically, more than one, coiled-coil prediction tool. The analysis was performed on both representative individual sequences and consensus sequences. The potential coiled-coils in CnuE homologs were often only found by Ncoils and were near the limit of detection, but these regions were also reported as coiled-coils in another study (Anantharaman et al., 2012). Given the context of extensive, high-probability coiled-coils in other components of the CoCoNuT systems with which they might interact, we chose to report these CnuE regions as coiled-coils, despite the comparative weakness of these predictions. In the AF2 multimer model of CnuHE, one of these potential coiled-coils is positioned near the coiled-coils detected in CnuH, suggesting they may facilitate interaction between these two factors.

Preliminary phylogenetic analysis of CoCoNuT CnuH helicases

A comprehensive search and phylogenetic analysis of this family was beyond the scope of this work, but to determine the relationships between the CoCoNuTs and the rest of the UPF1-like helicases, we used the following procedure. We retrieved the best 2000 hits in each of two searches with UPF1 and CoCoNuT helicases as queries against both a database containing predicted proteins from 24,757 completely sequenced prokaryotic genomes downloaded from the NCBI GenBank in November 2021 and a database containing 72 representative eukaryotic genomes that were downloaded from the NCBI GenBank in June 2020. Next, we combined all proteins in four searches, made a nonredundant set, and annotated them using CDD profiles, as described above. Then, we aligned them with MUSCLE v5 (Robert, 2022), constructed an approximately-maximum-likelihood tree with FastTree, and mapped the annotations onto the tree. Genome neighborhoods were generated for these hits, as described above.

Structural modeling with AlphaFold2 and searches for related folds

Protein structures were predicted using AlphaFold2 (AF2) v2.2.0 with local installations of complete databases required for AF2 (Jumper et al., 2021). Only protein models with average predicted local distance difference test (pLDDT) scores > 80 were retained for further analysis (Mariani et al., 2013). Many of these models were used as queries to search for structurally similar proteins using DALI v5 against the Protein Data Bank (PDB) and using FoldSeek against the

AlphaFold/UniProt50 v4, AlphaFold/Swiss-Prot v4, AlphaFold/Proteome v4, and PDB100 2201222 databases (Holm, 2020 [↗](#), van Kempen et al., 2023 [↗](#)). Structure visualizations and comparisons were performed with ChimeraX (Pettersen et al., 2021 [↗](#)) and the RCSB PDB website (Berman et al., 2000 [↗](#)).

Data availability

All data generated and analyzed in this study are included in the manuscript and supporting files. The CoCoNuT systems are documented in detail in the Supplementary Data. The AlphaFold2 structures generated and presented in the figures are available at modelarchive.org [↗](#) with accessions listed in Supplementary Table S2.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

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. analyzed data; R.T.B. and E.V.K. wrote the manuscript that was edited and approved by all authors. The authors declare no competing interests.

Acknowledgements

The authors thank Becky Xu Hua Fu (University of California, San Francisco) for correspondence that led to her contribution of the name CoCoNuT, Andrew Z. Fire and Usman Enam (Stanford University) for critical reading of the manuscript and insightful comments, Joseph Bondy-Denomy (University of California, San Francisco) for bringing the Shield factor ShdA to our attention, and Koonin group members for helpful discussions. The authors' research is supported by the Intramural Research Program of the National Institutes of Health (National Library of Medicine).

References

1. Agrawal N., Dasaradhi P. V., Mohammed A., Malhotra P., Bhatnagar R. K., Mukherjee S. K. (2003) **Rna interference: biology, mechanism, and applications** *Microbiol Mol Biol Rev* **67**:657–85
2. Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J. (1997) **Gapped Blast and Psi-Blast: a new generation of protein database search programs** *Nucleic acids research* **25**:3389–3402
3. Amir M., Kumar V., Dohare R., Islam A., Ahmad F., Hassan M. I. (2018) **Sequence, structure and evolutionary analysis of cold shock domain proteins, a member of Ob fold family** *Journal of Evolutionary Biology* **31**:1903–1917

4. Anantharaman V., Iyer L. M., Aravind L. (2012) **Ter-dependent stress response systems: novel pathways related to metal sensing, production of a nucleoside-like metabolite, and Dna-processing** *Molecular bioSystems* **8**:3142–3165
5. Barends S., Karzai A. W., Sauer R. T., Wower J., Kraal B. (2001) **Simultaneous and functional binding of Smpb and Ef-Tu-Tp to the alanyl acceptor arm of tmrna** *J Mol Biol* **314**:9–21
6. Bell R. T., Wolf Y. I., Koonin E. V. (2020) **Modified base-binding Eve and Dcd domains: striking diversity of genomic contexts in prokaryotes and predicted involvement in a variety of cellular processes** *Bmc Biol* **18**
7. Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E. (2000) **The Protein Data Bank** *Nucleic Acids Res* **28**:235–42
8. Bessho Y., Shibata R., Sekine S.-I., Murayama K., Higashijima K., Hori-Takemoto C., Shirouzu M., Kuramitsu S., Yokoyama S. (2007) **Structural basis for functional mimicry of long-variable-arm trna by transfer-messenger Rna** *Proceedings of the National Academy of Sciences* **104**:8293–8298
9. Bitton L., Klaiman D., Kaufmann G. (2015) **Phage T4-induced Dna breaks activate a trna repair-defying anticodon nuclease** *Mol Microbiol* **97**:898–910
10. Bourne H. R., Sanders D. A., McCormick F. (1991) **The GTPase superfamily: conserved structure and molecular mechanism** *Nature* **349**:117–27
11. Burroughs A. M., Zhang D., Schäffer D. E., Iyer L. M., Aravind L. (2015) **Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling** *Nucleic Acids Research* **43**:10633–10654
12. Chakrabarti S., Jayachandran U., Bonneau F., Fiorini F., Basquin C., Domcke S., Le Hir H., Conti E. (2011) **Molecular mechanisms for the Rna-dependent ATPase activity of Upf1 and its regulation by Upf2** *Mol Cell* **41**:693–703
13. Chang Y. W., Sun Y. J., Wang C., Hsiao C. D. (2008) **Crystal structures of the 70-kDa heat shock proteins in domain disjoining conformation** *J Biol Chem* **283**:15502–11
14. Chavez D. A., Greer B. H., Eichman B. F. (2018) **The Hiran domain of helicase-like transcription factor positions the Dna translocase motor to drive efficient Dna fork regression** *J Biol Chem* **293**:8484–8494
15. Cheng Z., Muhlrads D., Lim M. K., Parker R., Song H. (2007) **Structural and functional insights into the human Upf1 helicase core** *Embo j* **26**:253–64
16. Colicelli J. (2004) **Human Ras superfamily proteins and related GTPases** *Sci Stke* **2004**
17. Das U., Shuman S. (2013) **2'-Phosphate cyclase activity of Rtca: a potential rationale for the operon organization of Rtca with an Rna repair ligase Rtcb in Escherichia coli and other bacterial taxa** *Rna* **19**:1355–62
18. Delorenzi M., Speed T. (2002) **An Hmm model for coiled-coil domains and a comparison with Pssm-based predictions** *Bioinformatics* **18**:617–625
19. Dila D., Sutherland E., Moran L., Slatko B., Raleigh E. A. (1990) **Genetic and sequence organization of the mcrbc locus of Escherichia coli K-12** *Journal of Bacteriology* **172**

20. Dovas A., Couchman J. R. (2005) **Rhogdi: multiple functions in the regulation of Rho family GTPase activities** *Biochem J* **390**:1–9
21. Edgar R. C. (2004) **Muscle: a multiple sequence alignment method with reduced time and space complexity** *Bmc bioinformatics* **5**:113–113
22. Erzberger J. P., Berger J. M. (2006) **Evolutionary relationships and structural mechanisms of Aaa+ proteins** *Annu Rev Biophys Biomol Struct* **35**:93–114
23. Fairman-Williams M. E., Guenther U. P., Jankowsky E. (2010) **SF1 and SF2 helicases: family matters** *Curr Opin Struct Biol* **20**:313–24
24. Fire A., Xu S., Montgomery M. K., Kostas S. A., Driver S. E., Mello C. C. (1998) **Potent and specific genetic interference by double-stranded Rna in *Caenorhabditis elegans*** *Nature* **391**:806–11
25. Fleischman R. A., Cambell J. L., Richardson C. C. (1976) **Modification and restriction of T-even bacteriophages. In vitro degradation of deoxyribonucleic acid containing 5-hydroxymethylctosine** *J Biol Chem* **251**:1561–70
26. Gabler F., Nam S. Z., Till S., Mirdita M., Steinegger M., Söding J., Lupas A. N., Alva V. (2020) **Protein Sequence Analysis Using the Mpi Bioinformatics Toolkit** *Curr Protoc Bioinformatics* **72**
27. Gao L., Altae-Tran H., Bohning F., Makarova K. S., Segel M., Schmid-Burgk J. L., Koob J., Wolf Y. I., Koonin E. V., Zhang F. (2020) **Diverse enzymatic activities mediate antiviral immunity in prokaryotes** *Science* **369**:1077–1084
28. Gasiunas G., Barrangou R., Horvath P., Siksnys V. (2012) **Cas9-crrna ribonucleoprotein complex mediates specific Dna cleavage for adaptive immunity in bacteria** *Proc Natl Acad Sci U S A* **109**:E2579–86
29. Genschik P., Drabikowski K., Filipowicz W. (1998) **Characterization of the *Escherichia coli* Rna 3'-terminal phosphate cyclase and its sigma54-regulated operon** *J Biol Chem* **273**:25516–26
30. Goeders N., Drèze P. L., Van Melderen L. (2013) **Relaxed cleavage specificity within the Rel toxin family** *J Bacteriol* **195**:2541–9
31. Goldfarb T., Sberro H., Weinstock E., Cohen O., Doron S., Charpak-Amikam Y., Afik S., Ofir G., Sorek R. (2015) **Brex is a novel phage resistance system widespread in microbial genomes** *Embo j* **34**:169–83
32. Gorbalenya A. E., Koonin E. V. (1993) **Helicases: amino acid sequence comparisons and structure-function relationships** *Current Opinion in Structural Biology* **3**:419–429
33. Gucinski G. C., Michalska K., Garza-Sánchez F., Eschenfeldt W. H., Stols L., Nguyen J. Y., Goulding C. W., Joachimiak A., Hayes C. S. (2019) **Convergent Evolution of the Barnase/Endou/Colicin/Rel (Becr) Fold in Antibacterial tRNase Toxins** *Structure* **27**:1660–1674
34. Guglielmini J., Van Melderen L. (2011) **Bacterial toxin-antitoxin systems: Translation inhibitors everywhere** *Mob Genet Elements* **1**:283–290

35. Gutmann S., Haebel P. W., Metzinger L., Sutter M., Felden B., Ban N. (2003) **Crystal structure of the transfer-Rna domain of transfer-messenger Rna in complex with Smpb** *Nature* **424**:699–703
36. Guyomar C., D'urso G., Chat S., Giudice E., Gillet R. (2021) **Structures of tmrna and Smpb as they transit through the ribosome** *Nat Commun* **12**
37. Halaby D. M., Poupon A., Mornon J. (1999) **The immunoglobulin fold family: sequence analysis and 3d structure comparisons** *Protein Eng* **12**:563–71
38. Hauser M., Steinegger M., Söding J. (2016) **MMseqs software suite for fast and deep clustering and searching of large protein sequence sets** *Bioinformatics* **32**:1323–1330
39. Hazra D., Chapat C., Graille M. (2019) . **m⁶A mrna Destiny: Chained to the rhYTHm by the Yth-Containing Proteins** *Genes* **10**
40. Heinemann U., Roske Y. (2021) **Cold-Shock Domains-Abundance, Structure Properties, and Nucleic-Acid Binding.** *Cancers (Basel)* **13**
41. Himeno H., Kurita D., Muto A. (2014) . **tmrna-mediated trans-translation as the major ribosome rescue system in a bacterial cell** *Front Genet* **5**
42. Holm L. (2020) **Holm, L. 2020. Dali and the persistence of protein shape.** **29, 128–140.** *Dali and the persistence of protein shape* **29**:128–140
43. Hughes K. J., Chen X., Burroughs A. M., Aravind L., Wolin S. L. (2020) **An Rna Repair Operon Regulated by Damaged tRNAs** *Cell Rep* **33**
44. Ipsaro J. J., Haase A. D., Knott S. R., Joshua-Tor L., Hannon G. J. (2012) **The structural biochemistry of Zucchini implicates it as a nuclease in pirna biogenesis** *Nature* **491**:279–283
45. Irastortza-Olaziregi M., Amster-Choder O. (2020) **Coupled Transcription-Translation in Prokaryotes: An Old Couple With New Surprises** *Front Microbiol* **11**
46. Ishita J. *et al.* (2021) . **trna anticodon cleavage by target-activated Caspr-Cas13a effector** *bioRxiv*
47. Iyer L. M., Leipe D. D., Koonin E. V., Aravind L. (2004) **Evolutionary history and higher order classification of Aaa+ ATPases** *Journal of Structural Biology* **146**:11–31
48. Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. (2012) **A Programmable Dual-Rna-Guided Dna Endonuclease in Adaptive Bacterial Immunity** *Science* **337**:816–821
49. Jumper J. *et al.* (2021) **Highly accurate protein structure prediction with AlphaFold** *Nature* **596**:583–589
50. Kalathiya U., Padariya M., Pawlicka K., Verma C. S., Houston D., Hupp T. R., Alfaro J. A. (2019) **Insights into the Effects of Cancer Associated Mutations at the UPF2 and Atp-Binding Sites of Nmd Master Regulator: UPF1** *Int J Mol Sci* **20**
51. Kaufmann G. (2000) **Anticodon nucleases** *Trends Biochem Sci* **25**:70–4

52. Kaur G., Burroughs A. M., Iyer L. M., Aravind L. (2020) **Highly regulated, diversifying Ntp-dependent biological conflict systems with implications for the emergence of multicellularity** *eLife* **9**
53. Khadria A. S., Senes A. (2013) **The transmembrane domains of the bacterial cell division proteins Ftsb and Ftsl form a stable high-order oligomer** *Biochemistry* **52**:7542–50
54. Kishor A., Tandukar B., Ly Y. V., Toth E. A., Suarez Y., Brewer G., Wilson G. M. (2013) **Hsp70 is a novel posttranscriptional regulator of gene expression that binds and stabilizes selected mRNAs containing Au-rich elements** *Mol Cell Biol* **33**:71–84
55. Kishor A., White E. J. F., Matsangos A. E., Yan Z., Tandukar B., Wilson G. M. (2017) **Hsp70's Rna-binding and mrna-stabilizing activities are independent of its protein chaperone functions** *J Biol Chem* **292**:14122–14133
56. Koonin E. V., Aravind L. (2002) **Origin and evolution of eukaryotic apoptosis: the bacterial connection** *Cell Death & Differentiation* **9**:394–404
57. Koonin E. V., Krupovic M. (2019) **Origin of programmed cell death from antiviral defense?** *Proceedings of the National Academy of Sciences* **116**
58. Koonin E. V., Zhang F. (2017) **Coupling immunity and programmed cell suicide in prokaryotes: Life- or-death choices** *BioEssays* **39**
59. Kotta-Loizou I. *et al.* (2022) **The Rna repair proteins Rtcab regulate transcription activator Rtcr via its Crispr-associated Rossmann fold domain** *iScience* **25**
60. Laganeckas M., Margelevicius M., Venclovas C. (2011) **Identification of new homologs of Pd-(D/E)Xk nucleases by support vector machines trained on data derived from profile-profile alignments** *Nucleic Acids Res* **39**:1187–96
61. Liao S., Sun H., Xu C. (2018) **Yth Domain: A Family of N6-methyladenosine (m6a) Readers** *Genomics, Proteomics & Bioinformatics* **16**:99–107
62. Loenen W. A., Dryden D. T., Raleigh E. A., Wilson G. G., Murray N. E. (2014) **Highlights of the Dna cutters: a short history of the restriction enzymes** *Nucleic Acids Res* **42**:3–19
63. Lupas A., Van Dyke M., Stock J. (1991) **Predicting coiled coils from protein sequences** *Science* **252**:1162–4
64. Luria S. E., Human M. L. (1952) **A nonhereditary, host-induced variation of bacterial viruses** *J Bacteriol* **64**:557–69
65. Macdonald E., Strahl H., Blower T. R., Palmer T., Mariano G. (2022) **Shield co-opts an Rmuc domain to mediate phage defence across Pseudomonas species**
66. Makarova K. S., Anantharaman V., Aravind L., Koonin E. V. (2012) **Live virus-free or die: coupling of antiviral immunity and programmed suicide or dormancy in prokaryotes** *Biology direct* **7**:40–40
67. Makarova K. S., Grishin N. V., Shabalina S. A., Wolf Y. I., Koonin E. V. (2006) **A putative Rna-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action** *Biology Direct* **1**

68. Makarova K. S., Timinskas A., Wolf Y. I., Gussow A. B., Siksnys V., Venclovas Č., Koonin E. V. (2020) **Evolutionary and functional classification of the Carf domain superfamily, key sensors in prokaryotic antiviral defense** *Nucleic Acids Res* **48**:8828–8847
69. Makarova K. S. *et al.* (2020) **Evolutionary classification of Caspr-Cas systems: a burst of class 2 and derived variants** *Nature Reviews Microbiology* **18**:67–83
70. Mariani V., Biasini M., Barbato A., Schwede T. (2013) . **Iddt: a local superposition-free score for comparing protein structures and models using distance difference tests** *Bioinformatics* **29**:2722–2728
71. Mayer M. P. (2021) **The Hsp70-Chaperone Machines in Bacteria** *Front Mol Biosci* **8**
72. McDonnell A. V., Jiang T., Keating A. E., Berger B. (2006) **Paircoil2: improved prediction of coiled coils from sequence** *Bioinformatics* **22**:356–358
73. McMahon S. A., Zhu W., Graham S., Rambo R., White M. F., Gloster T. M. (2020) **Structure and mechanism of a Type Iii Caspr defence Dna nuclease activated by cyclic oligoadenylate** *Nature Communications* **11**
74. Mendez A. S., Vogt C., Bohne J., Glaunsinger B. A. (2018) **Site specific target binding controls Rna cleavage efficiency by the Kaposi's sarcoma-associated herpesvirus endonuclease Sox** *Nucleic Acids Research* **46**:11968–11979
75. Millman A., Melamed S., Amitai G., Sorek R. (2020) **Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems** *Nature Microbiology* **5**:1608–1615
76. Morse J. C., Girodat D., Burnett B. J., Holm M., Altman R. B., Sanbonmatsu K. Y., Wieden H.-J., Blanchard S. C. (2020) **Elongation factor-Tu can repetitively engage aminoacyl-trna within the ribosome during the proofreading stage of trna selection** *Proceedings of the National Academy of Sciences* **117**:3610–3620
77. Natarajan K., Mage M. G., Margulies D. H. (2015) **Immunoglobulin Superfamily**
78. Neuwald A. F., Aravind L., Spouge J. L., Koonin E. V. (1999) **Aaa+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes** *Genome Res* **9**:27–43
79. Nirwan N., Itoh Y., Singh P., Bandyopadhyay S., Vinothkumar K. R., Amunts A., Saikrishnan K. (2019) **Structure-based mechanism for activation of the Aaa+ GTPase Mrcb by the endonuclease Mrcr** *Nature Communications* **10**
80. Niu Y., Suzuki H., Hosford C. J., Walz T., Chappie J. S. (2020) **Structural asymmetry governs the assembly and GTPase activity of Mrcbc restriction complexes** *Nat Commun* **11**
81. Ofir G., Melamed S., Sberro H., Mukamel Z., Silverman S., Yaakov G., Doron S., Sorek R. (2018) **Disarm is a widespread bacterial defence system with broad anti-phage activities** *Nature Microbiology* **3**:90–98
82. Panne D., Raleigh E. A., Bickle T. A. (1999) **The Mrcbc endonuclease translocates Dna in a reaction dependent on Gtp hydrolysis** *J Mol Biol* **290**:49–60
83. Park H. H., Lo Y. C., Lin S. C., Wang L., Yang J. K., Wu H. (2007) **The death domain superfamily in intracellular signaling of apoptosis and inflammation** *Annu Rev Immunol* **25**:561–86

84. Patil D. P., Pickering B. F., Jaffrey S. R. (2018) **Reading m6a in the Transcriptome: m6a-Binding Proteins** *Trends in Cell Biology* **28**:113–127
85. Pedersen K., Zavialov A. V., Pavlov M. Y., Elf J., Gerdes K., Ehrenberg M. (2003) **The bacterial toxin Rele displays codon-specific cleavage of mRNAs in the ribosomal A site** *Cell* **112**:131–40
86. Pettersen E. F., Goddard T. D., Huang C. C., Meng E. C., Couch G. S., Croll T. I., Morris J. H., Ferrin T. E. (2021) **Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. 2021. Ucsf ChimeraX: Structure visualization for researchers, educators, and developers. 30, 70–82. Ucsf ChimeraX: Structure visualization for researchers, educators, and developers 30:70–82**
87. Pieper U., Schweitzer T., Groll D. H., Gast F. U., Pingoud A. (1999) **The Gtp-binding domain of Mcrb: more than just a variation on a common theme?** *J Mol Biol* **292**:547–56
88. Pillon M. C., Gordon J., Frazier M. N., Stanley R. E. (2021) **Hepn RNases - an emerging class of functionally distinct Rna processing and degradation enzymes** *Crit Rev Biochem Mol Biol* **56**:88–108
89. Price M. N., Dehal P. S., Arkin A. P. (2010) **FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments** *Plos One* **5**
90. Raleigh E. A. (1992) **Organization and function of the mcrbc genes of Escherichia coli K-12** *Mol Microbiol* **6**:1079–86
91. Raleigh E. A., Trimarchi R., Revel H. (1989) **Genetic and physical mapping of the mcra (rgla) and mcrb (rglb) loci of Escherichia coli K-12** *Genetics* **122**:279–296
92. Robert C. E. (2022) **High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny** *bioRxiv*
93. Shigematsu M., Kawamura T., Kirino Y. (2018) **Generation of 2',3'-Cyclic Phosphate-Containing RNAs as a Hidden Layer of the Transcriptome** *Front Genet* **9**
94. Simm D., Hatje K., Kollmar M. (2015) **Waggawagga: comparative visualization of coiled-coil predictions and detection of stable single α -helices (Sah domains)** *Bioinformatics* **31**:767–9
95. Söding J. (2004) **Protein homology detection by HMM–HMM comparison** *Bioinformatics* **21**:951–960
96. Söding J., Remmert M., Biegert A., Lupas A. N. (2006) **HHsenser: exhaustive transitive profile search using HMM–HMM comparison** *Nucleic Acids Research* **34**:W374–W378
97. Songailiene I., Juozapaitis J., Tamulaitiene G., Ruksenaite A., Šulčius S., Sasnauskas G., Venclovas Č., Siksnys V. (2020) **Hepn-Mnt Toxin-Antitoxin System: The Hepn Ribonuclease Is Neutralized by OligoAMPylation** *Molecular Cell* **80**:955–970
98. Stock A. M., Robinson V. L., Goudreau P. N. (2000) **Two-component signal transduction** *Annu Rev Biochem* **69**:183–215
99. Sukackaite R., Grazulis S., Tamulaitis G., Siksnys V. (2012) **The recognition domain of the methyl-specific endonuclease Mcrbc flips out 5-methylcytosine** *Nucleic acids research* **40**:7552–7562

100. Sutherland E., Coe L., Raleigh E. A. (1992) **Mcrbc: a multisubunit Gtp-dependent restriction endonuclease** *J Mol Biol* **225**:327–48
101. Swarts D. C., Makarova K., Wang Y., Nakanishi K., Ketting R. F., Koonin E. V., Patel D. J., Van Der Oost J. (2014) **The evolutionary journey of Argonaute proteins** *Nature structural & molecular biology* **21**:743–753
102. Tanaka N., Shuman S. (2011) **Rtcb is the Rna ligase component of an Escherichia coli Rna repair operon** *J Biol Chem* **286**:7727–7731
103. Trigg J., Gutwin K., Keating A. E., Berger B. (2011) **Multicoil2: Predicting Coiled Coils and Their Oligomerization States from Sequence in the Twilight Zone** *Plos One* **6**
104. Tsutakawa S. E., Jingami H., Morikawa K. (1999) **Recognition of a Tg Mismatch: The Crystal Structure of Very Short Patch Repair Endonuclease in Complex with a Dna Duplex** *Cell* **99**:615–623
105. Van Den Berg D. F., Van Der Steen B. A., Costa A. R., & Brouns S. J. J. (2023) **Phage tRNAs evade trna-targeting host defenses through anticodon loop mutations** *eLife* **12**
106. Van Kempen M., Kim S. S., Tumescheit C., Mirdita M., Lee J., Gilchrist C. L. M., Söding J., Steinegger M. (2023) **Fast and accurate protein structure search with Foldseek** *Nature Biotechnology*
107. Weigele P., Raleigh E. (2016) **Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses** *Chemical reviews* **116**
108. Willmund F., Del Alamo M., Pechmann S., Chen T., Albanèse V., Dammer E. B., Peng J., Frydman J. (2013) **The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis** *Cell* **152**:196–209
109. Yamamoto S. *et al.* (2014) **ATPase Activity and Atp-dependent Conformational Change in the Co-chaperone HSP70/HSP90-organizing Protein (Hop)*** *Journal of Biological Chemistry* **289**:9880–9886
110. Zhou C., Pourmal S., Pavletich N. P. (2015) **Dna2 nuclease-helicase structure, mechanism and regulation by Rpa** *eLife* **4**
111. Zhu W., Mcquarrie S., Grüşchow S., McMahon S. A., Graham S., Gloster T. M., White M. F. (2021) **The Crispr ancillary effector Can2 is a dual-specificity nuclease potentiating type Iii Crispr defence** *Nucleic Acids Research* **49**:2777–2789
112. Zimmer C., Von Gabain A., Henics T. (2001) **Analysis of sequence-specific binding of Rna to Hsp70 and its various homologs indicates the involvement of N- and C-terminal interactions** *Rna* **7**:1628–37
113. Zimmermann L., Stephens A., Nam S.-Z., Rau D., Kübler J., Lozajic M., Gabler F., Söding J., Lupas A. N., Alva V. (2018) **A Completely Reimplemented Mpi Bioinformatics Toolkit with a New HHpred Server at its Core** *Journal of Molecular Biology* **430**:2237–2243
114. Zimmermann L., Stephens A., Nam S. Z., Rau D., Kubler J., Lozajic M., Gabler F., Soding J., Lupas A. N., Alva V. (2018) **A Completely Reimplemented Mpi Bioinformatics Toolkit with a New HHpred Server at its Core** *J Mol Biol* **430**:2237–2243

Article and author information

Ryan T. Bell

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

For correspondence: ryan.bell@nih.gov

Harutyun Sahakyan

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

Kira S. Makarova

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

Yuri I. Wolf

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

Eugene V. Koonin

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

For correspondence: koonin@ncbi.nlm.nih.gov

Copyright

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0 public domain dedication](https://creativecommons.org/publicdomain/zero/1.0/).

Editors

Reviewing Editor

Andrei Lupas

Max Planck Institute for Developmental Biology, Tübingen, Germany

Senior Editor

Detlef Weigel

Max Planck Institute for Biology Tübingen, Tübingen, Germany

Reviewer #1 (Public Review):

Summary:

In this manuscript, Bell et al. provide an exhaustive and clear description of the diversity of a new class of predicted type IV restriction systems that the authors denote as CoCoNuTs, for their characteristic presence of coiled-coil segments and nuclease tandems. Along with a comprehensive analysis that includes phylogenetics, protein structure prediction, extensive protein domain annotations, and in-depth investigation of encoding genomic contexts, they also provide detailed hypothesis about the biological activity and molecular functions of the

members of this class of predicted systems. This work is highly relevant, it underscores the wide diversity of defence systems that are used by prokaryotes and demonstrates that there are still many systems to be discovered. The work is sound and backed up by a clear and reasonable bioinformatics approach.

Strengths:

The analysis provided by the authors is extensive and covers the three most important aspects that can be covered computationally when analysing a new family/superfamily: phylogenetics, genomic context analysis, and protein-structure-based domain content annotation. With this, one can directly have an idea about the superfamily of the predicted system and infer about their biological role. The bioinformatics approach is sound and makes use of the most current advances in the fields of protein evolution and structural bioinformatics.

Weaknesses:

It is not clear how coiled-coil segments were assigned if only based on AF2-predicted models or also backed by sequence analysis, as no description is provided in the methods. The structure prediction quality assessment is based solely on the average pLDDT of the obtained models (with a threshold of 80 or better). However, this is not enough, particularly when multimeric models were used. The PAE matrix should be used to evaluate relative orientations, particularly in the case where there is a prediction that parts from 2 proteins are interacting. In the case of multimers, interface quality scores, as the ipTM or pDockQ, should also be considered and, at minimum, reported.

These weaknesses were addressed during revision, and the results provided by the authors support their conclusions. The data resulting from this work will be useful for the general life sciences community, particularly the prokaryotic defense and microbiology communities. It also underscores the high range of functionally unknowns in sequenced genomes that are now much easier to find and interpret due to the success of deep-learning based methods and automated robust bioinformatics pipelines.

<https://doi.org/10.7554/eLife.94800.2.sa1>

Reviewer #2 (Public Review):

Summary:

In this work, using in-depth computational analysis, Bell et al. explore the diverse repertoire of type IV McrBC modification dependent restriction systems. The prototypical two-component McrBC system has been structurally and functionally characterised and is known to act as a defence by restricting phage and foreign DNA containing methylated cytosines. Here, the authors find previously unanticipated complexity and versatility of these systems and focus on detailed analysis and classification of a distinct branch, the so-called CoCoNut, named after its composition of coiled-coil structures and tandem nucleases. These CoCoNut systems are predicted to target RNA as well as DNA and to utilise defence mechanisms with some similarity to type III CRISPR-Cas systems.

Strengths:

This work is enriched with a plethora of ideas and a myriad of compelling hypotheses that now will await experimental verification. The study comes from the group that was amongst the first to describe, characterise, and classify CRISPR-Cas systems. By analogy, the findings described here can similarly promote ingenious experimental and conceptual research that

could further drive technological advances. It could also instigate vigorous scientific debates that will ultimately benefit the community.

Weaknesses:

The multi-component systems described here function in the context of large oligomeric complexes similarly to the prototypical McrBC system. While the AlphaFold2 (AF2) multimer predictions are provided in this work, these are not compared with the known McrBC structures. These comparisons could have been helpful not only for providing insights into these multimeric protein systems but also for giving more sound explanations of the differences observed amongst different McrBC types.

<https://doi.org/10.7554/eLife.94800.2.sa0>

Author response:

The following is the authors' response to the original reviews.

Public Reviews:

Reviewer #1 (Public Review):

Summary:

In this manuscript, Bell et al. provide an exhaustive and clear description of the diversity of a new class of predicted type IV restriction systems that the authors denote as CoCoNuTs, for their characteristic presence of coiled-coil segments and nuclease tandems. Along with a comprehensive analysis that includes phylogenetics, protein structure prediction, extensive protein domain annotations, and an in-depth investigation of encoding genomic contexts, they also provide detailed hypotheses about the biological activity and molecular functions of the members of this class of predicted systems. This work is highly relevant, it underscores the wide diversity of defence systems that are used by prokaryotes and demonstrates that there are still many systems to be discovered. The work is sound and backed-up by a clear and reasonable bioinformatics approach. I do not have any major issues with the manuscript, but only some minor comments.

Strengths:

The analysis provided by the authors is extensive and covers the three most important aspects that can be covered computationally when analysing a new family/superfamily: phylogenetics, genomic context analysis, and protein-structure-based domain content annotation. With this, one can directly have an idea about the superfamily of the predicted system and infer their biological role. The bioinformatics approach is sound and makes use of the most current advances in the fields of protein evolution and structural bioinformatics.

Weaknesses:

It is not clear how coiled-coil segments were assigned if only based on AF2-predicted models or also backed by sequence analysis, as no description is provided in the methods. The structure prediction quality assessment is based solely on the average pLDDT of the obtained models (with a threshold of 80 or better). However, this is not enough, particularly when multimeric models are used. The PAE matrix should be used to evaluate relative orientations, particularly in the case where there is a prediction that

parts from 2 proteins are interacting. In the case of multimers, interface quality scores, such as the ipTM or pDockQ, should also be considered and, at minimum, reported.

A description of the coiled-coil predictions has been added to the Methods. For multimeric models, PAE matrices and ipTM+pTM scores have been included in Supplementary Data File S1.

Reviewer #2 (Public Review):

Summary:

In this work, using in-depth computational analysis, Bell et al. explore the diverse repertoire of type IV McrBC modification-dependent restriction systems. The prototypical two-component McrBC system has been structurally and functionally characterised and is known to act as a defence by restricting phage and foreign DNA containing methylated cytosines. Here, the authors find previously unanticipated complexity and versatility of these systems and focus on detailed analysis and classification of a distinct branch, the so-called CoCoNut, named after its composition of coiled-coil structures and tandem nucleases. These CoCoNut systems are predicted to target RNA as well as DNA and to utilise defence mechanisms with some similarity to type III CRISPR-Cas systems.

Strengths:

This work is enriched with a plethora of ideas and a myriad of compelling hypotheses that now await experimental verification. The study comes from the group that was amongst the first to describe, characterize, and classify CRISPR-Cas systems. By analogy, the findings described here can similarly promote ingenious experimental and conceptual research that could further drive technological advances. It could also instigate vigorous scientific debates that will ultimately benefit the community.

Weaknesses:

The multi-component systems described here function in the context of large oligomeric complexes. Some of the single chain AF2 predictions shown in this work are not compatible, for example, with homohexameric complex formation due to incompatible orientation of domains. The recent advances in protein structure prediction, in particular AlphaFold2 (AF2) multimer, now allow us to confidently probe potential protein-protein interactions and protein complex formation. This predictive power could be exploited here to produce a better glimpse of these multimeric protein systems. It can also provide a more sound explanation for some of the observed differences amongst different McrBC types.

Hexameric CnuB complexes with CnuC stimulatory monomers for Type I-A, I-B, I-C, II, and III-A CoCoNuT systems have been modeled with AF2 and included in Supplementary Data File S1, albeit without the domains fused to the GTPase N-terminus (with the exception of Type I-B, which lacks the long coiled-coil domain fused to the GTPase and was modeled with its entire sequence). Attempts to model the other full-length CnuB hexamers did not lead to convincing results.

Recommendations for the authors:

Reviewing Editor:

The detailed recommendations by the two reviewers will help the authors to further strengthen the manuscript, but two points seem particularly worth considering: 1. The methods are barely sketched in the manuscript, but it could be useful to detail them more closely. Particularly regarding the coiled-coil segments, which are currently just

statists, useful mainly for the name of the family, more detail on their prediction, structural properties, and purpose would be very helpful. 2. Due to its encyclopedic nature, the wealth of material presented in the paper makes it hard to penetrate in one go. Any effort to make it more accessible would be very welcome. Reviewer 1 in particular has made a number of suggestions regarding the figures, which would make them provide more support for the findings described in the text.

A description of the techniques used to identify coiled-coil segments has been added to the Methods. Our predictions ranged from near certainty in the coiled-coils detected in CnuB homologs, to shorter helices at the limit of detection in other factors. We chose to report all probable coiled-coils, as the extensive coiled-coils fused to CnuB, which are often the only domain present other than the GTPase, imply involvement in mediating complex formation by interacting with coiled-coils in other factors, particularly the other CoCoNuT factors. The suggestions made by Reviewer 1 were thoughtful and we made an effort to incorporate them.

Reviewer #1 (Recommendations For The Authors):

I do not have any major issues with the manuscript. I have however some minor comments, as described below.

- The last sentence of the abstract at first reads as a fact and not a hypothesis resulting from the work described in the manuscript. After the second read, I noticed the nuances in the sentence. I would suggest a rephrasing to emphasize that the activity described is a theoretical hypothesis not backed-up by experiments.*

This sentence has been rephrased to make explicit the hypothetical nature of the statement.

- In line 64, the authors rename DUF3578 as ADAM because indeed its function is not unknown. Did the authors consider reaching out to InterPro to add this designation to this DUF? A search in interpro with DUF3578 results in "MrcB-like, N-terminal domain" and if a name is suggested, it may be worthwhile to take it to the IntrePro team.*

We will suggest this nomenclature to InterPro.

- I find Figure 1E hard to analyse and think it occupies too much space for the information it provides. The color scheme, the large amount of small slices, and the lack of numbers make its information content very small. I would suggest moving this to the supplementary and making it instead a bar plot. If removed from Figure 1, more space is made available for the other panels, particularly the structural superpositions, which in my opinion are much more important.*

We have removed Figure 1E from the paper as it adds little information beyond the abundance and phyletic distribution of sequenced prokaryotes, in which MrcBC systems are plentiful.

- In Figure 2, it is not clear due to the presence of many colorful "operon schemes" that the tree is for a single gene and not for the full operon segment. Highlighting the target gene in the operons or signalling it somehow would make the figure easy to understand even in the absence of the text and legend. The same applies to Supplementary Figure 1.*

The legend has been modified to show more clearly that this is a tree of McrB-like GTPases.

- *In line 146, the authors write "AlphaFold-predicted endonuclease fold" to say that a protein contains a region that AF2 predicts to fold like an endonuclease. This is a weird way of writing it and can be confusing to non-expert readers. I would suggest rephrasing for increased clarity.*

This sentence has been rephrased for greater clarity.

- *In line 167, there is a [47]. I believe this is probably due to a previous reference formatting.*

Indeed, this was a reference formatting error and has been fixed.

- *In most figures, the color palette and the use of very similar color palettes for taxonomy pie charts, genomic context composition schemes, and domain composition diagrams make it really hard to have a good understanding of the image at first. Legends are often close to each other, and it is not obvious at first which belong to what. I would suggest changing the layouts and maybe some color schemes to make it easier to extract the information that these figures want to convey.*

It seemed that Figure 4 was the most glaring example of these issues, and it has been rearranged for easier comprehension.

- *In the paragraph that starts at line 199, the authors mention an Ig-like domain that is often found at the N-terminus of Type I CoCoNuTs. Are they all related to each other? How conserved are these domains?*

These domains are all predicted to adopt a similar beta-sandwich fold and are found at the N-terminus of most CoCoNuT CnuC homologs, suggesting they are part of the same family, but we did not undertake a more detailed sequenced-based analysis of these regions.

We also find comparable domains in the CnuC/McrC-like partners of the abundant McrB-like NxD motif GTPases that are not part of CoCoNuT systems, and given the similarity of some of their predicted structures to Rho GDP-dissociation inhibitor 1, we suspect that they have coevolved as regulators of the non-canonical NxD motif GTPase type. Our CnuBC multimer models showing consistent proximity between these domains in CnuC and CnuB GTPase domains suggest this could indeed be the case. We plan to explore these findings further in a forthcoming publication.

- *In line 210, the authors write "suggesting a role in overcrowding-induced stress response". Why so? In >all other cases, the authors justify their hypothesis, which I really appreciated, but not here.*

A supplementary note justifying this hypothesis has been added to Supplementary Data File S1.

- *At the end of the paragraph that starts in line 264, the authors mention that they constructed AF2 multimeric models to predict if 2 proteins would interact. However, no quality scores were provided, particularly the PAE matrix. This would allow for a better judgement of this prediction, and I would suggest adding the*

PAE matrix as another panel in the figure where the 3D model of the complex is displayed.

The PAE matrix and ipTM+pTM scores for this and other multimer models have been added to Supplementary Data File S1. For this model in particular, the surface charge distribution of the model has been presented to support the role of the domains that have a higher PAE in RNA binding.

- In line 306, "(supplementary data)" refers to what part of the file?*

This file has been renamed Supplementary Table S3 and referenced as such.

- In line 464, the authors suggest that ShdA could interact with CoCoNuTs. Why not model the complex as done for other cases? what would co-folding suggest?*

As we were not able to convincingly model full-length CnuB hexamers with N-terminal coiled-coils, we did not attempt modeling of this hypothetical complex with another protein with a long coiled-coil, but it remains an interesting possibility.

- In line 528, why and how were some genes additionally analyzed with HHPred?*

Justification for this analysis has been added to the Methods, but briefly, these genes were additionally analyzed if there were no BLAST hits or to confirm the hits that were obtained.

- In the first section of the methods, the first and second (particularly the second) paragraphs are extremely long. I would suggest breaking them to facilitate reading.*

This change has been made.

- In line 545, what do the authors mean by "the alignment (...) were analyzed with HHPred"?*

A more detailed description of this step has been added to the Methods.

- The authors provide the models they produced as well as extensive supplementary tables that make their data reusable, but they do not provide the code for the automated steps, as to excise target sequence sections out of multiple sequence alignments, for example.*

The code used for these steps has been in use in our group at the NCBI for many years. It will be difficult to utilize outside of the NCBI software environment, but for full disclosure, we have included a zipped repository with the scripts and custom-code dependencies, although there are external dependencies as well such as FastTree and BLAST. In brief, it involves PSI-BLAST detection of regions with the most significant homology to one of a set of provided alignments (seals-2-master/bin/wrappers/cog_psicognitor). In this case, the reference alignments of McrB-like GTPases and DUF2357 were generated manually using HHpred to analyze alignments of clustered PSI-BLAST results. This step provided an output of coordinates defining domain footprints in each query sequence, which were then combined and/or extended using scripts based on manual analysis of many examples with HHpred (footprint_finders/get_GTPase_fragments.py and footprint_finders/get_DUF2357_fragments.py), then

these coordinates were used to excise such regions from the query amino acid sequence with a final script (seals-2-master/bin/misc/fa2frag).

Reviewer #2 (Recommendations For The Authors):

(1) Page 4, line 77 - 'PUA superfamily domains' could be more appropriate to use instead of "EVE superfamily".

While this statement could perhaps be applied to PUA superfamily domains, our previous work we refer to, which strongly supports the assertion, was restricted to the EVE-like domains and we prefer to retain the original language.

(2) Page 5, lines 128-130 - AF2 multimer prediction model could provide a more sound explanation for these differences.

Our AF2 multimer predictions added in this revision indeed show that the NxD motif McrB-like CoCoNuT GTPases interact with their respective McrC-like partners such that an immunoglobulin-like beta-sandwich domain, fused to the N-termini of the McrC homologs and similar to Rho GDP-dissociation inhibitor 1, has the potential to physically interact with the GTPase variants. However, we did not probe this in greater detail, as it is beyond the scope of this already highly complex article, but we plan to study it in the future.

(3) Page 8, line 252 - The surface charge distribution of CnuH OB fold domain looks very different from SmpB (pdb3iyr). In fact, the regions that are in contact with RNA in SmpB are highly acidic in CoCoNut CnuH. Although it looks likely that this domain is involved in RNA binding, the mode of interaction should be very different.

We did not detect a strong similarity between the CnuH SmpB-like SPB domain and PDB 3IYR, but when we compare the surface charge distribution of PDB 1WJX and the SPB domain, while there is a significant area that is positively charged in 1WJX that is negatively charged in SPB, there is much that overlaps with the same charge in both domains.

The similarity between SmpB and the SPB domain is significant, but definitely not exact. An important question for future studies is: If the domains are indeed related due to an ancient fusion of SmpB to an ancestor of CnuH, would this degree of divergence be expected?

In other words, can we say anything about how the function of a stand-alone tmRNA-binding protein could evolve after being fused to a complex predicted RNA helicase with other predicted RNA binding domains already present? Experimental validation will ultimately be necessary to resolve these kinds of questions, but for now, it may be safe to say that the presence of this domain, especially in conjunction with the neighboring RelE-like RTL domain and UPF1-like helicase domain, signals a likely interaction with the A-site of the ribosome, and perhaps restriction of aberrant/viral mRNA.