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

Ryan T. Bell[1,*], Harutyun Sahakyan[1], Kira S. Makarova[1], Yuri I. Wolf[1], Eugene V. Koonin[1,*]

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

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

Abstract

A comprehensive census of McrBC systems, 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 (CRISPR-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.

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. coil* K-12, is considered a Type IV (modification-dependent) restriction system that degrades DNA containing methylcytosine (5mC) or hydroxymethylcytosine (5hmC), with a degree of sequence context specificity (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 McrB-like GTPases in prokaryotes, and indeed, these are the most frequently occurring EVE-containing fusion proteins (Bell et al., 2020). 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, 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) and McrC DUF2357 (Supplementary Figure S1) families. These two trees were generally topologically concordant and revealed several distinct branches not previously recognized. The branch containing the prototypical McrBC system from *E. coil* K-12 (blue in Figure 2) 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). This branch and others, which exhibit two particularly prevalent associations, with a DISARM-like antiphage system (green in Figure 2) (Ofir et al., 2018), and, surprisingly for a Type IV restriction system, with predicted DNA methyltransferases (red in Figure 2), 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) 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 NxD (Figure 1 B, Figure 2) (Neuwald et al., 1999, Iyer et al., 2004, Erzberger and Berger, 2006, Niu et al., 2020). The NxD 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 NxD motif McrB GTPase has been characterized, but the extensive study of the NxxD motif offers clues to the potential impact of the motif shortening. The asparagine (N) residue is strictly required for GTP binding and hydrolysis (Pieper et al., 1999). 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, Colicelli, 2004, Bourne et al., 1991, Niu et al., 2020, Pieper et al., 1999). 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 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 NxD branch is characterized by many McrC homologs with unusual features, such as predicted RNA-binding domains, that might not be compatible with conventional McrB GTPases from the NxxD clade, which often occur in the same genomes (see below).

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

The most notable feature of the NxD branch of GTPases is a large clade characterized by fusion to long coiled-coil domains (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 NxD 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 NxD 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).

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, 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 Melderen, 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 3′-phosphate RNA termini to 2′,3′-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 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 (Irastortza-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.

Acknowledgments

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

AGRAWAL, N., DASARADHI, P. V., MOHMMED, A., MALHOTRA, P., BHATNAGAR, R. K. & MUKHERJEE, S. K. 2003. RNA interference: biology, mechanism, and applications. *Microbiol Mol Biol Rev,* 67**,** 657-85.

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.

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.

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.

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.

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**,** 159.

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.

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.

BITTON, L., KLAIMAN, D. & KAUFMANN, G. 2015. Phage T4-induced DNA breaks activate a tRNA repair-defying anticodon nuclease. *Mol Microbiol,* 97**,** 898-910.

BOURNE, H. R., SANDERS, D. A. & MCCORMICK, F. 1991. The GTPase superfamily: conserved structure and molecular mechanism. *Nature,* 349**,** 117-27.

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.

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.

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.

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.

CHENG, Z., MUHLRAD, D., LIM, M. K., PARKER, R. & SONG, H. 2007. Structural and functional insights into the human Upf1 helicase core. *Embo j,* 26**,** 253-64.

COLICELLI, J. 2004. Human RAS superfamily proteins and related GTPases. *Sci STKE,* 2004**,** Re13.

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.

DELORENZI, M. & SPEED, T. 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics,* 18, 617-625.

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, 4888.

DOVAS, A. & COUCHMAN, J. R. 2005. RhoGDI: multiple functions in the regulation of Rho family GTPase activities. *Biochem J,* 390, 1-9.

EDGAR, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics,* 5, 113-113.

ERZBERGER, J. P. & BERGER, J. M. 2006. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct,* 35, 93-114.

FAIRMAN-WILLIAMS, M. E., GUENTHER, U. P. & JANKOWSKY, E. 2010. SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol,* 20, 313-24.

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.

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.

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, e108.

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.

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.

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.

GOEDERS, N., DRÈZE, P. L. & VAN MELDEREN, L. 2013. Relaxed cleavage specificity within the RelE toxin family. *J Bacteriol,* 195, 2541-9.

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.

GORBALENYA, A. E. & KOONIN, E. V. 1993. Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology,* 3, 419-429.

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/RelE (BECR) Fold in Antibacterial tRNase Toxins. *Structure,* 27, 1660-1674.e5.

GUGLIELMINI, J. & VAN MELDEREN, L. 2011. Bacterial toxin-antitoxin systems: Translation inhibitors everywhere. *Mob Genet Elements,* 1, 283-290.

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.

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, 4909.

HALABY, D. M., POUPON, A. & MORNON, J. 1999. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Eng,* 12**,** 563-71.

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.

HAZRA, D., CHAPAT, C. & GRAILLE, M. 2019. m^{6}A mRNA Destiny: Chained to the rhYTHm by the YTH-Containing Proteins. *Genes,* 10**,** 49.

HEINEMANN, U. & ROSKE, Y. 2021. Cold-Shock Domains-Abundance, Structure, Properties, and Nucleic-Acid Binding. *Cancers (Basel),* 13.

HIMENO, H., KURITA, D. & MUTO, A. 2014. tmRNA-mediated trans-translation as the major ribosome rescue system in a bacterial cell. *Front Genet,* 5**,** 66.

HOLM, L. 2020. DALI and the persistence of protein shape. 29**,** 128-140.

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**,** 108527.

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.

IRASTORTZA-OLAZIREGI, M. & AMSTER-CHODER, O. 2020. Coupled Transcription-Translation in Prokaryotes: An Old Couple With New Surprises. *Front Microbiol,* 11**,** 624830.

ISHITA, J., MATVEY, K., LEONID, M., NATALIA, M., ANNA, S., ALEXANDR, K., SOFIA, M., KONSTANTIN, K., SERGEI, B., KIRA, S. M., EUGENE, V. K., KONSTANTIN, S. & EKATERINA, S. 2021. tRNA anticodon cleavage by target-activated CRISPR-Cas13a effector. *bioRxiv***,** 2021.11.10.468108.

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.

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.

JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ŽÍDEK, A., POTAPENKO, A., BRIDGLAND, A., MEYER, C., KOHL, S. A. A., BALLARD, A. J., COWIE, A., ROMERA-PAREDES, B., NIKOLOV, S., JAIN, R., ADLER, J., BACK, T., PETERSEN, S., REIMAN, D., CLANCY, E., ZIELINSKI, M., STEINEGGER, M., PACHOLSKA, M., BERGHAMMER, T., BODENSTEIN, S., SILVER, D., VINYALS, O., SENIOR, A. W., KAVUKCUOGLU, K., KOHLI, P. & HASSABIS, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature,* 596**,** 583-589.

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.

KAUFMANN, G. 2000. Anticodon nucleases. *Trends Biochem Sci,* 25**,** 70-4.

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**,** e52696.

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.

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.

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.

KOONIN, E. V. & ARAVIND, L. 2002. Origin and evolution of eukaryotic apoptosis: the bacterial connection. *Cell Death & Differentiation,* 9**,** 394-404.

KOONIN, E. V. & KRUPOVIC, M. 2019. Origin of programmed cell death from antiviral defense? *Proceedings of the National Academy of Sciences,* 116**,** 16167.

KOONIN, E. V. & ZHANG, F. 2017. Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. *BioEssays,* 39**,** e201600186.

KOTTA-LOIZOU, I., GIULIANO, M. G., JOVANOVIC, M., SCHAEFER, J., YE, F., ZHANG, N., IRAKLEIDI, D. A., LIU, X., ZHANG, X., BUCK, M. & ENGL, C. 2022. The RNA repair proteins RtcAB regulate transcription activator RtcR via its CRISPR-associated Rossmann fold domain. *iScience,* 25**,** 105425.

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.

LIAO, S., SUN, H. & XU, C. 2018. YTH Domain: A Family of N6-methyladenosine (m6A) Readers. *Genomics, Proteomics & Bioinformatics,* 16**,** 99-107.

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.

LUPAS, A., VAN DYKE, M. & STOCK, J. 1991. Predicting coiled coils from protein sequences. *Science,* 252**,** 1162-4.

LURIA, S. E. & HUMAN, M. L. 1952. A nonhereditary, host-induced variation of bacterial viruses. *J Bacteriol,* 64**,** 557-69.

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. 2022.11.04.515146.

MAKAROVA, K. S., ANANTHARAMAN, V., ARAVIND, L. & KOONIN, E. V. 2012. Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. *Biology direct,* 7**,** 40-40.

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**,** 7.

MAKAROVA, K. S., TIMINSKAS, A., WOLF, Y. I., GUSSOW, A. B., SIKSNYS, V., VENCLOVAS, Č. & KOONIN, E. V. 2020a. Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic Acids Res,* 48**,** 8828-8847.

MAKAROVA, K. S., WOLF, Y. I., IRANZO, J., SHMAKOV, S. A., ALKHNBASHI, O. S., BROUNS, S. J. J., CHARPENTIER, E., CHENG, D., HAFT, D. H., HORVATH, P., MOINEAU, S., MOJICA, F. J. M., SCOTT, D., SHAH, S. A., SIKSNYS, V., TERNS, M. P., VENCLOVAS, Č., WHITE, M. F., YAKUNIN, A. F., YAN, W., ZHANG, F., GARRETT, R. A., BACKOFEN, R., VAN DER OOST, J., BARRANGOU, R. & KOONIN, E. V. 2020b. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology,* 18**,** 67-83.

MARIANI, V., BIASINI, M., BARBATO, A. & SCHWEDE, T. 2013. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics,* 29**,** 2722-2728.

MAYER, M. P. 2021. The Hsp70-Chaperone Machines in Bacteria. *Front Mol Biosci,* 8**,** 694012.

MCDONNELL, A. V., JIANG, T., KEATING, A. E. & BERGER, B. 2006. Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics,* 22**,** 356-358.

MCMAHON, S. A., ZHU, W., GRAHAM, S., RAMBO, R., WHITE, M. F. & GLOSTER, T. M. 2020. Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nature Communications,* 11**,** 500.

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.

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.

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.

NATARAJAN, K., MAGE, M. G. & MARGULIES, D. H. 2015. Immunoglobulin Superfamily. *Encyclopedia of Life Sciences.*

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.

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 McrB by the endonuclease McrC. *Nature Communications,* 10**,** 3058.

NIU, Y., SUZUKI, H., HOSFORD, C. J., WALZ, T. & CHAPPIE, J. S. 2020. Structural asymmetry governs the assembly and GTPase activity of McrBC restriction complexes. *Nat Commun,* 11**,** 5907.

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.

PANNE, D., RALEIGH, E. A. & BICKLE, T. A. 1999. The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *J Mol Biol,* 290**,** 49-60.

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.

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.

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.

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.

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.

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.

PRICE, M. N., DEHAL, P. S. & ARKIN, A. P. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE,* 5**,** e9490.

RALEIGH, E. A. 1992. Organization and function of the mcrBC genes of Escherichia coli K-12. *Mol Microbiol,* 6**,** 1079-86.

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.

ROBERT, C. E. 2022. High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *bioRxiv***,** 2021.06.20.449169.

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**,** 562.

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.

SÖDING, J. 2004. Protein homology detection by HMM–HMM comparison. *Bioinformatics,* 21, 951-960.

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.

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.e7.

STOCK, A. M., ROBINSON, V. L. & GOUDREAU, P. N. 2000. Two-component signal transduction. *Annu Rev Biochem,* 69, 183-215.

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.

SUTHERLAND, E., COE, L. & RALEIGH, E. A. 1992. McrBC: a multisubunit GTP-dependent restriction endonuclease. *J Mol Biol,* 225, 327-48.

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.

TANAKA, N. & SHUMAN, S. 2011. RtcB is the RNA ligase component of an Escherichia coli RNA repair operon. *J Biol Chem,* 286, 7727-7731.

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, e23519.

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.

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, e85183.

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*.

WEIGELE, P. & RALEIGH, E. 2016. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chemical reviews,* 116.

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.

YAMAMOTO, S., SUBEDI, G. P., HANASHIMA, S., SATOH, T., OTAKA, M., WAKUI, H., SAWADA, K.-I., YOKOTA, S.-I., YAMAGUCHI, Y., KUBOTA, H. & ITOH, H. 2014. ATPase Activity and ATP-dependent Conformational Change in the Co-chaperone HSP70/HSP90-organizing Protein (HOP)*. *Journal of Biological Chemistry,* 289, 9880-9886.

ZHOU, C., POURMAL, S. & PAVLETICH, N. P. 2015. Dna2 nuclease-helicase structure, mechanism and regulation by Rpa. *eLife,* 4, e09832.

ZHU, W., MCQUARRIE, S., GRÜSCHOW, 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.

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.

ZIMMERMANN, L., STEPHENS, A., NAM, S.-Z., RAU, D., KÜBLER, J., LOZAJIC, M., GABLER, F., SÖDING, J., LUPAS, A. N. & ALVA, V. 2018a. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology,* 430**,** 2237-2243.

ZIMMERMANN, L., STEPHENS, A., NAM, S. Z., RAU, D., KUBLER, J., LOZAJIC, M., GABLER, F., SODING, J., LUPAS, A. N. & ALVA, V. 2018b. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol,* 430**,** 2237-2243.

Figure legends

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).

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 NxD motif, whereas the blue, green, red, and purple groups, with variable bootstrap support, have an NxxD motif, indicated by the arrows; the sequences in the smaller, cyan clade, with 98% bootstrap support, have an NxxxD motif. Each of the differently colored groups is characterized by distinct conserved genomic associations that are abundant within but not completely confined to the respective groups. This tree was built from the representatives of 90% identity clusters of all validated homologs. Abbreviations of domains: McrB – McrB GTPase domain; CoCo/CC – coiled-coil; MN – McrC N-terminal domain (DUF2357); CSD – cold shock domain; IG – Immunoglobulin (IG)-like beta-sandwich domain; ZnR – zinc ribbon domain; SPB – SmpB-like domain; RTL – RNase toxin-like domain; HEPN – HEPN family nuclease domain; OB – OB-fold domain; iPD-D/ExK – inactivated PD-D/ExK fold; Hsp70-like ATPase – Hsp70-like NBD/SBD; HEAT – HEAT-like helical repeats; YprA – YprA-like helicase domain; DUF1998 - DUF1998 is often found in or associated with helicases and contains four conserved, putatively metal ion-binding cysteine residues; SWI2/SNF2 – SWI2/SNF2-family

ATPase; PglX – PglX-like DNA methyltransferase; HsdR/M/S – Type I RM system restriction, methylation, and specificity factors.

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.

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).

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

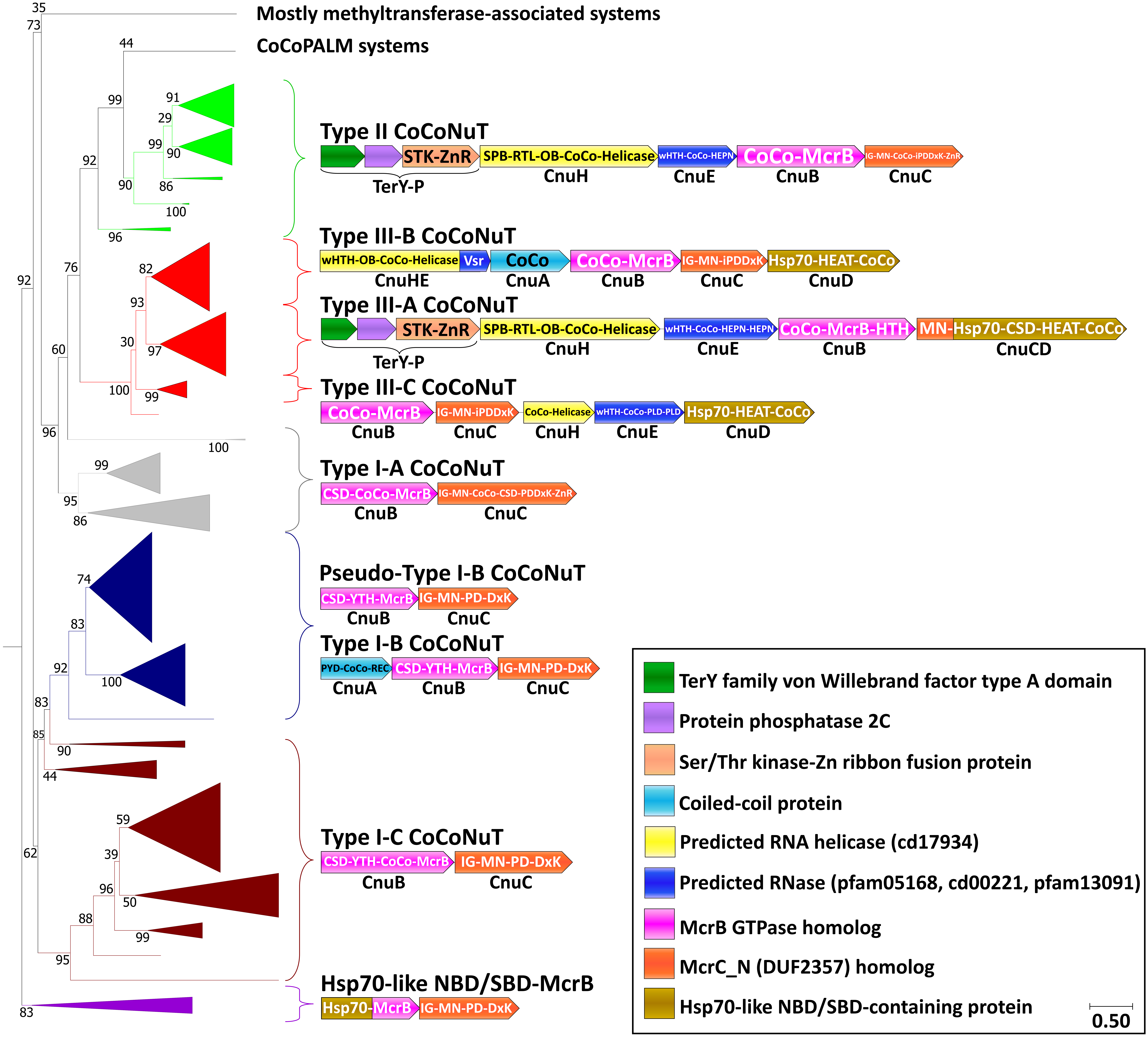
A) Type II and III CoCoNuT domain composition and operon organization. The arrows indicate the direction of transcription. Type II and III-A systems very frequently contain TerY-P systems as well, but not invariably, and these are never found in Type III-B or III-C, thus, we do not consider them core 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).

Figure 7: Complex operonic associations of Type II CoCoNuTs

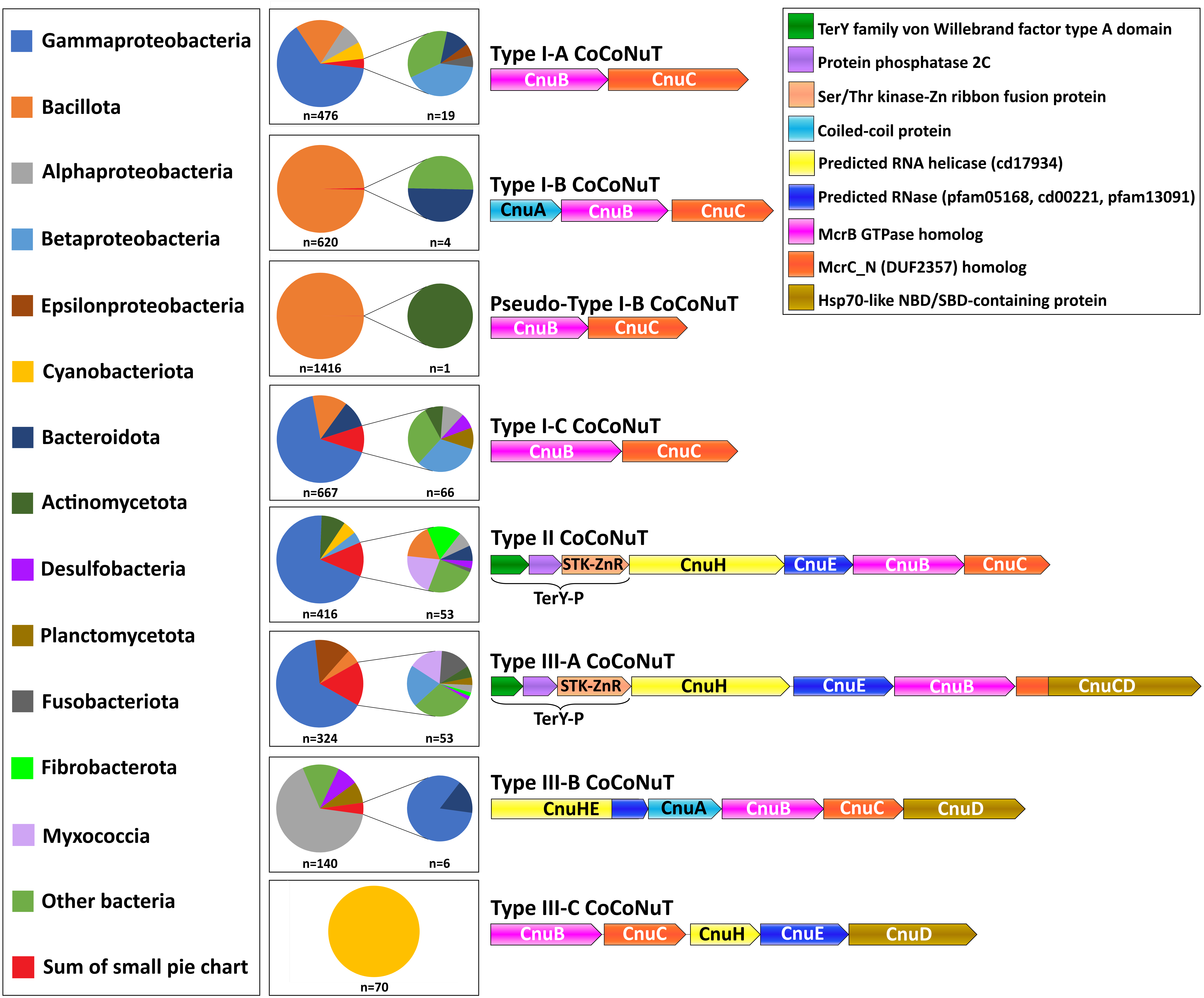
Type II CoCoNuTs are frequently associated with RtcR homologs, and in many cases, ancillary defense genes are located between the RtcR gene and the CoCoNuT, almost always oriented in the same direction in an apparent superoperon. Abbreviations of domains: YprA – YprA-like helicase domain; DUF1998 - DUF1998 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.

A
McrB
McrC
B
1
169
459
Walker A
Walker B
NxxD
ADAM/DUF3578
McrB GTPase
McrB
E.coli K-12 McrB N-terminal ADAM domain AF2 structural model
E.coli K-12 McrB N-terminal ADAM domain X-ray structure (3SSC-A)
E.coli K-12 McrB GTPase domain AF2 stuctural model
E.coli K-12 McrB GTPase domain cryo-EM structure (6UT6-A)
C
1
189
348
DUF2357
PD-DxK nuclease
McrC
E.coli K-12 McrC N-terminal DUF2357 domain AF2 structural model
E.coli K-12 McrC PD-DxK nuclease domain AF2 stuctural model
E.coli K-12 McrC cryo-EM structure (6UT6-G)

Often associated with Type I RM
HsdM
HsdS
HsdR
McrB
McrC
Mostly Actinomycetota, often associated with DISARM-like system
YprA-Zn_rib-DUF1998-PD(D/E)XK
SWI2/SNF2
PD(D/E)xK-PglX
McrB
McrC
McrB-like GTPase domain phylogenetic tree
McrB
HTH domain (pfam12728)
0.5
DNA methyltransferase (COG0270, cd00315)
Histidine kinase/response regulator
TerY family von Willebrand factor type A domain
Protein phosphatase 2C
Ser/Thr Kinase-Zn ribbon fusion protein
Predicted RNA helicase (cd17934)
Predicted RNase (pfam05168)
Hsp70-like NBD/SBD-containing protein
Often associated with DNA methyltranferases
McrB
McrC
94
59
NxxD
98
NxD
NxxxD
97
92
73
86
Mostly archaeal systems
McrB
McrC
Type I-A CoCoNuT
CSD-CoCo-McrB
IG-MN-CC-CSD-PDDxK-ZnR
Type II CoCoNuT
STK-ZnR
SPB-RTL-OB-CoCo-Helicase
wHTH-CC-HEPN
CoCo-McrB
IG-MN-CC-iPDDxK-ZnR
Type III-A CoCoNuT
STK-ZnR
SPB-RTL-OB-CoCo-Helicase
wHTH-CC-HEPN-HEPN
CoCo-McrB-HTH
MN-Hsp70-CSD-HEAT-CC
Mostly systems with elongated McrC homologs, and often associated with DNA methyltransferases and two-component regulatory systems
McrB
McrC
McrB
McrC

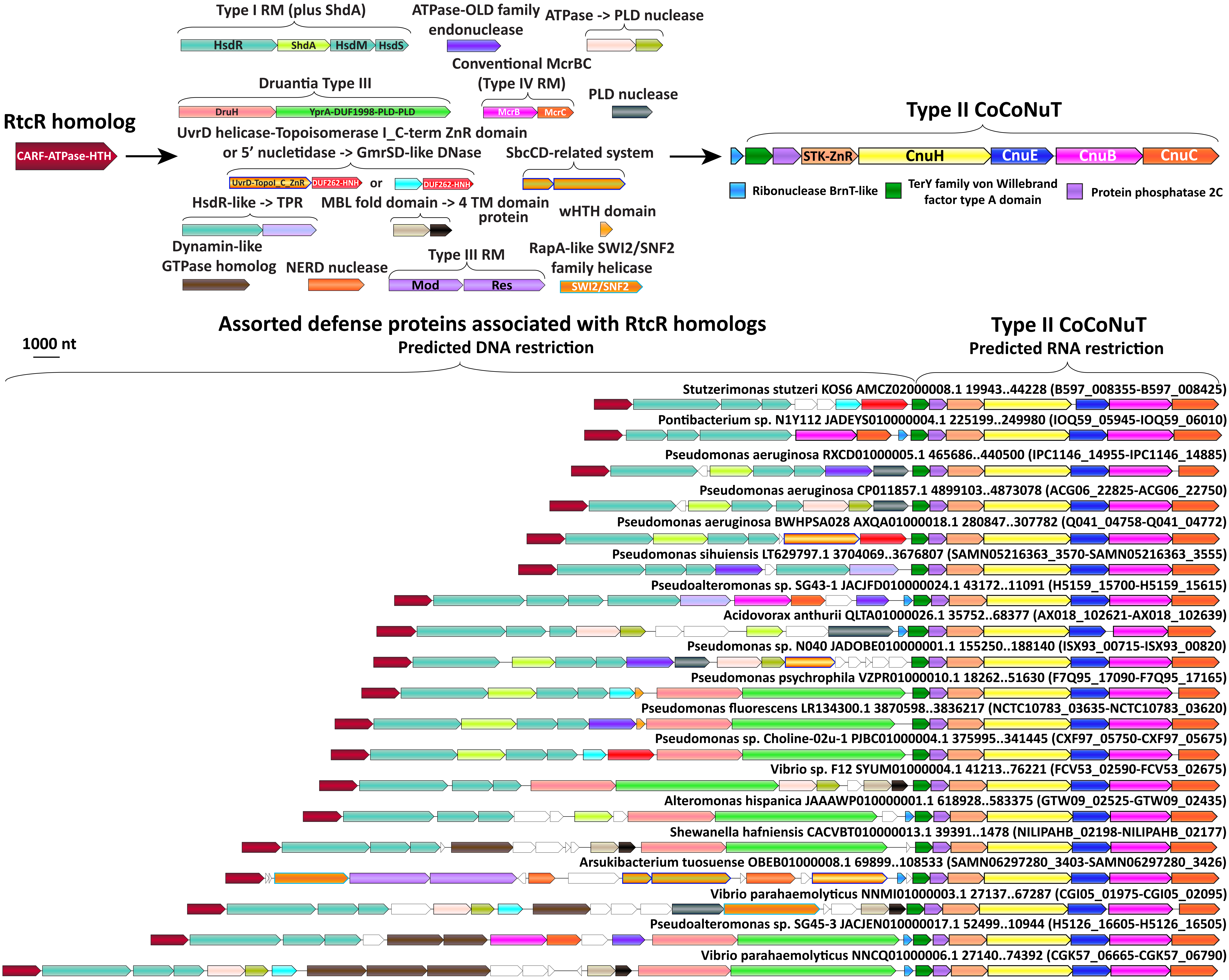

Mostly methyltransferase-associated systems
CoCoPALM systems
Type II CoCoNuT
STK-ZnR
SPB-RTL-OB-CoCo-Helicase
wHTH-CoCo-HEPN
CoCo-McrB
IG-MN-CoCo-iPDDxK-ZnR
TerY-P
CnuH
CnuE
CnuB
CnuC
Type III-B CoCoNuT
wHTH-OB-CoCo-Helicase
Vsr
CoCo
CoCo-McrB
IG-MN-iPDDxK
Hsp70-HEAT-CoCo
CnuHE
CnuA
CnuB
CnuC
CnuD
Type III-A CoCoNuT
STK-ZnR
SPB-RTL-OB-CoCo-Helicase
wHTH-CoCo-HEPN-HEPN
CoCo-McrB-HTH
MN-Hsp70-CSD-HEAT-CoCo
TerY-P
CnuH
CnuE
CnuB
CnuCD
Type III-C CoCoNuT
CoCo-McrB
IG-MN-iPDDxK
CoCo-Helicase
wHTH-CoCo-PLD-PLD
Hsp70-HEAT-CoCo
CnuB
CnuC
CnuH
CnuE
CnuD
Type I-A CoCoNuT
CSD-CoCo-McrB
IG-MN-CoCo-CSD-PDDxK-ZnR
CnuB
CnuC
Pseudo-Type I-B CoCoNuT
CSD-YTH-McrB
IG-MN-PD-DxK
CnuB
CnuC
Type I-B CoCoNuT
PYD-CoCo-REC
CSD-YTH-McrB
IG-MN-PD-DxK
CnuA
CnuB
CnuC
Type I-C CoCoNuT
CSD-YTH-CoCo-McrB
IG-MN-PD-DxK
CnuB
CnuC
Hsp70-like NBD/SBD-McrB
Hsp70-McrB
IG-MN-PD-DxK
35
73
44
99
91
29
99
90
92
90
86
100
96
92
76
82
93
30
97
60
100
99
96
100
99
95
86
74
83
92
100
83
85
90
44
59
62
39
96
50
88
99
95
83
🟩 TerY family von Willebrand factor type A domain
🟪 Protein phosphatase 2C
🟧 Ser/Thr kinase-Zn ribbon fusion protein
🟦 Coiled-coil protein
🟨 Predicted RNA helicase (cd17934)
🟦 Predicted RNase (pfam05168, cd00221, pfam13091)
🟪 McrB GTPase homolog
🟧 McrC_N (DUF2357) homolog
🟫 Hsp70-like NBD/SBD-containing protein
0.50

Phyletic distribution of CoCoNuT types



A
Type I-A CoCoNuT
CSD
Coiled-coil (CoCo)
McrB
Helical insert
GTPase
IG
DUF2357
CoCo
CSD
CoCo
PD-DxK
ZnR
Type I-B CoCoNuT
PYD
CoCo
REC
CSD
YTH
McrB GTPase
IG
DUF2357
PD-DxK
Direction of transcription
Type I-C CoCoNuT
CSD
YTH
Coiled-coil (CoCo)
McrB GTPase
IG
DUF2357
PD-DxK
B
Type I-A CoCoNuT
1
155
279
Walker A
Walker B
454
529
NxD
735
CSD
Coiled-coil (CoCo)
McrB
Helical insert
GTPase
1
137
265
335
457
486
564
737
860
IG
DUF2357
CoCo
CSD
CoCo
PD-DxK
ZnR
C
Type I-B CoCoNuT
1
123
288
364
PYD
CoCo
REC
1
108
230
584
CSD
YTH
McrB GTPase
1
106
278
483
IG
DUF2357
PD-DxK
D
Type I-C CoCoNuT
1
101
245
492
835
CSD
YTH
Coiled-coil (CoCo)
McrB GTPase
1
127
394
609
IG
DUF2357
PD-DxK

A
Direction of transcription
Type II CoCoNuT
SPB RTL OB/stalk CoCo Helical insert CoCo RNA helicase
wHTH CoCo HEPN
Coiled-coil (CoCo) McrB GTPase
IG DUF2357 CoCo inactivated PD-DxK CoCo ZnR
Type III-A CoCoNuT
SPB RTL OB/stalk CoCo Helical insert CoCo RNA helicase
wHTH CoCo HEPN HEPN
Coiled-coil (CoCo) McrB GTPase HTH
DUF2357 HSP70-like ATPase CSD HEAT CoCo
Type III-B CoCoNuT
wHTH OB/helices CoCo Helical insert RNA helicase Vsr
Coiled-coil (CoCo)
Coiled-coil (CoCo) McrB GTPase
IG DUF2357 inactivated PD-DxK
HSP70-like ATPase HEAT CoCo
Type III-C CoCoNuT
Coiled-coil (CoCo) McrB GTPase
IG DUF2357 inactivated PD-DxK
CoCo Helical insert RNA helicase
wHTH CoCo PLD PLD
HSP70-like ATPase HEAT CoCo
B
Type II CoCoNuT
1 429 801
Coiled-coil (CoCo) McrB GTPase
1 68 196 227 327 477 502 546 601
IG DUF2357 CoCo inactivated PD-DxK CoCo ZnR
C
Type II and Type III-A CoCoNuT
1 91 228 446 505 710 741 1124
SPB RTL OB/stalk CoCo Helical insert CoCo RNA helicase
1 177 359 492
wHTH CoCo HEPN
1 294 316 521 708
wHTH CoCo HEPN HEPN
D
Type III-A CoCoNuT
1 392 807 840
Coiled-coil (CoCo) McrB GTPase HTH
1 253 743 829 1042 1112
DUF2357 HSP70-like ATPase/SBD CSD HEAT CoCo



Type I RM (plus ShdA)
HsdR
ShdA
HsdM
HsdS
ATPase-OLD family endonuclease
ATPase -> PLD nuclease
Conventional McrBC (Type IV RM)
McrB
McrC
Druantia Type III
DruH
YprA-DUF1998-PLD-PLD
PLD nuclease
RtcR homolog
CARF-ATPase-HTH
UvrD helicase-Topoisomerase I_C-term ZnR domain or 5' nucletidase -> GmrSD-like DNase
UvrD-TopoI_C_ZnR
DUF262-HNH
or
DUF262-HNH
SbcCD-related system
HsdR-like -> TPR
MBL fold domain -> 4 TM domain protein
wHTH domain
Dynamin-like GTPase homolog
NERD nuclease
Type III RM
Mod
Res
RapA-like SWI2/SNF2 family helicase
SWI2/SNF2
Type II CoCoNuT
STK-ZnR
CnuH
CnuE
CnuB
CnuC
Ribonuclease BrnT-like
TerY family von Willebrand factor type A domain
Protein phosphatase 2C
Assorted defense proteins associated with RtcR homologs
Predicted DNA restriction
Type II CoCoNuT
Predicted RNA restriction
1000 nt
Stutzerimonas stutzeri KOS6 AMCZ02000008.1 19943..44228 (B597_008355-B597_008425)
Pontibacterium sp. N1Y112 JADEYS010000004.1 225199..249980 (IOQ59_05945-IOQ59_06010)
Pseudomonas aeruginosa RXCD01000005.1 465686..440500 (IPC1146_14955-IPC1146_14885)
Pseudomonas aeruginosa CP011857.1 4899103..4873078 (ACG06_22825-ACG06_22750)
Pseudomonas aeruginosa BWHPSA028 AXQA01000018.1 280847..307782 (Q041_04758-Q041_04772)
Pseudomonas sihuiensis LT629797.1 3704069..3676807 (SAMN05216363_3570-SAMN05216363_3555)
Pseudoalteromonas sp. SG43-1 JACJFD010000024.1 43172..11091 (H5159_15700-H5159_15615)
Acidovorax anthurii QLTA01000026.1 35752..68377 (AX018_102621-AX018_102639)
Pseudomonas sp. N040 JADOBE010000001.1 155250..188140 (ISX93_00715-ISX93_00820)
Pseudomonas psychrophila VZPR01000010.1 18262..51630 (F7Q95_17090-F7Q95_17165)
Pseudomonas fluorescens LR134300.1 3870598..3836217 (NCTC10783_03635-NCTC10783_03620)
Pseudomonas sp. Choline-02u-1 PJBC01000004.1 375995..341445 (CXF97_05750-CXF97_05675)
Vibrio sp. F12 SYUM01000004.1 41213..76221 (FCV53_02590-FCV53_02675)
Alteromonas hispanica JAAAWP010000001.1 618928..583375 (GTW09_02525-GTW09_02435)
Shewanella hafniensis CACVBT010000013.1 39391..1478 (NILIPAHB_02198-NILIPAHB_02177)
Arsukibacterium tuosuense OBEB01000008.1 69899..108533 (SAMN06297280_3403-SAMN06297280_3426)
Vibrio parahaemolyticus NNMI01000003.1 27137..67287 (CGI05_01975-CGI05_02095)
Pseudoalteromonas sp. SG45-3 JACJEN010000017.1 52499..10944 (H5126_16605-H5126_16505)
Vibrio parahaemolyticus NNCQ01000006.1 27140..74392 (CGK57_06665-CGK57_06790)