

CoCoNuTs are a diverse subclass of Type IV restriction systems predicted to target RNA

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

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, United States

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 coiled-coil nuclease tandems (CoCoNuTs) for their salient features: the presence of extensive coiled-coil structures and tandem nucleases. The CoCoNuTs alone show extraordinary variety, with three 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.

***For correspondence:**

ryan.bell@nih.gov (RTB);
koonin@ncbi.nlm.nih.gov (EVK)

Competing interest: The authors declare that no competing interests exist.

Funding: See page 21

Sent for Review

01 December 2023

Preprint posted

02 December 2023

Reviewed preprint posted

02 February 2024

Reviewed preprint revised

22 April 2024

Version of Record published

13 May 2024

Reviewing Editor: Andrei N Lupas, Max Planck Institute for Developmental Biology, Germany

© This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0 public domain dedication](https://creativecommons.org/licenses/by/4.0/).

eLife assessment

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

Introduction

All organisms are subject to an incessant barrage of genetic parasites, such as viruses and transposons. Over billions of years, the continuous arms race between hosts and parasites drove the evolution of immense, intricately interconnected networks of diverse defense systems and pathways (*Burroughs et al., 2015; Gao et al., 2020; Goldfarb et al., 2015; Bell et al., 2020; Koonin and Aravind, 2002;*

eLife digest All organisms, from animals to bacteria, are subject to genetic parasites, such as viruses and transposons. Genetic parasites are pieces of nucleic acids (DNA or RNA) that can use a cell's machinery to copy themselves at the expense of their hosts. This often leads to the host's demise, so organisms evolved many types of defense mechanisms. One of the most ancient and common forms of defense against viruses and transposons is the targeted restriction of nucleic acids, that is, deployment of host enzymes that can destroy or restrict nucleic acids containing specific sequence motifs or modifications.

In bacteria, many of the restriction enzymes targeting parasitic genetic elements are formed by fusions of proteins from the so-called McrBC systems with a protein domain called EVE. EVE and other functionally similar domains are a part of proteins that recognize and bind modified bases in nucleic acids. Enzymes can use the ability of these specificity domains to bind modified bases to detect non-host nucleic acids.

Bell et al. conducted a comprehensive computational search for McrBC systems and discovered a large and highly diverse branch of this family with unusual characteristic structural and functional domains. These features include regions that form long alpha-helices (coils) that coil with other alpha-helices (known as coiled-coils), as well as several distinct enzymatic domains that break down nucleic acids (known as nucleases). They call these systems CoCoNuTs (coiled-coiled *nuclease tandems*).

All CoCoNuTs contain domains, including EVE-like ones, which are predicted to interact with components of the RNA-based systems responsible for producing proteins in the cell (translation), suggesting that the CoCoNuTs have an important impact on protein abundance and RNA metabolism.

Bell et al.'s findings will be of interest to scientists working on prokaryotic immunity and virulence. Furthermore, similarities between CoCoNuTs and components of eukaryotic RNA-degrading systems suggest evolutionary connections between this diverse family of bacterial predicted RNA restriction systems and RNA regulatory pathways of eukaryotes.

Further deciphering the mechanisms of CoCoNuTs could shed light on how certain pathways of RNA metabolism and regulation evolved, and how they may contribute to advances in biotechnology.

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 *Escherichia coli* K-12, is considered a Type IV (modification-dependent) restriction system that degrades DNA containing methylcytosine (5mC) or hydroxymethylcytosine (5hmC), with a degree of sequence context specificity (**Sutherland et al., 1992; Sukackaite et al., 2012**).

Type IV restriction enzymes contain at least two components: (1) a dedicated specificity domain that recognizes modified DNA and (2) an endonuclease domain that cleaves the target (**Weigle and Raleigh, 2016; Loenen et al., 2014**). In the well-characterized example from *E. coli* K-12, McrB harbors

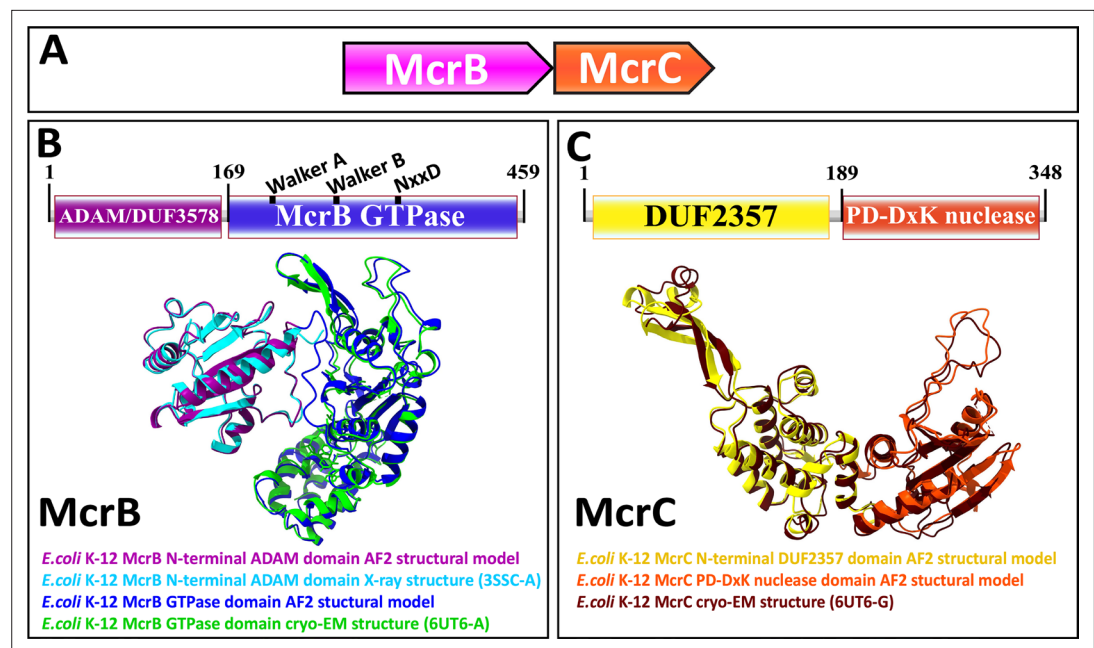


Figure 1. Genetic organization, signature sequence motifs, structural models, and phyletic distribution of McrB GTPases detected in this work. **(A)** McrBC is a two-component restriction system with each component typically (except for extremely rare gene fusions) encoded by a separate gene expressed 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-DxK architecture (Niu et al., 2020) show a high degree of similarity. The structures were visualized with ChimeraX (Pettersen et al., 2021).

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 1A and B; Sutherland et al., 1992; Iyer et al., 2004a; 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 1A and C; 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/E)xK 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, Figure 2—figure supplement 1). Unexpectedly, we

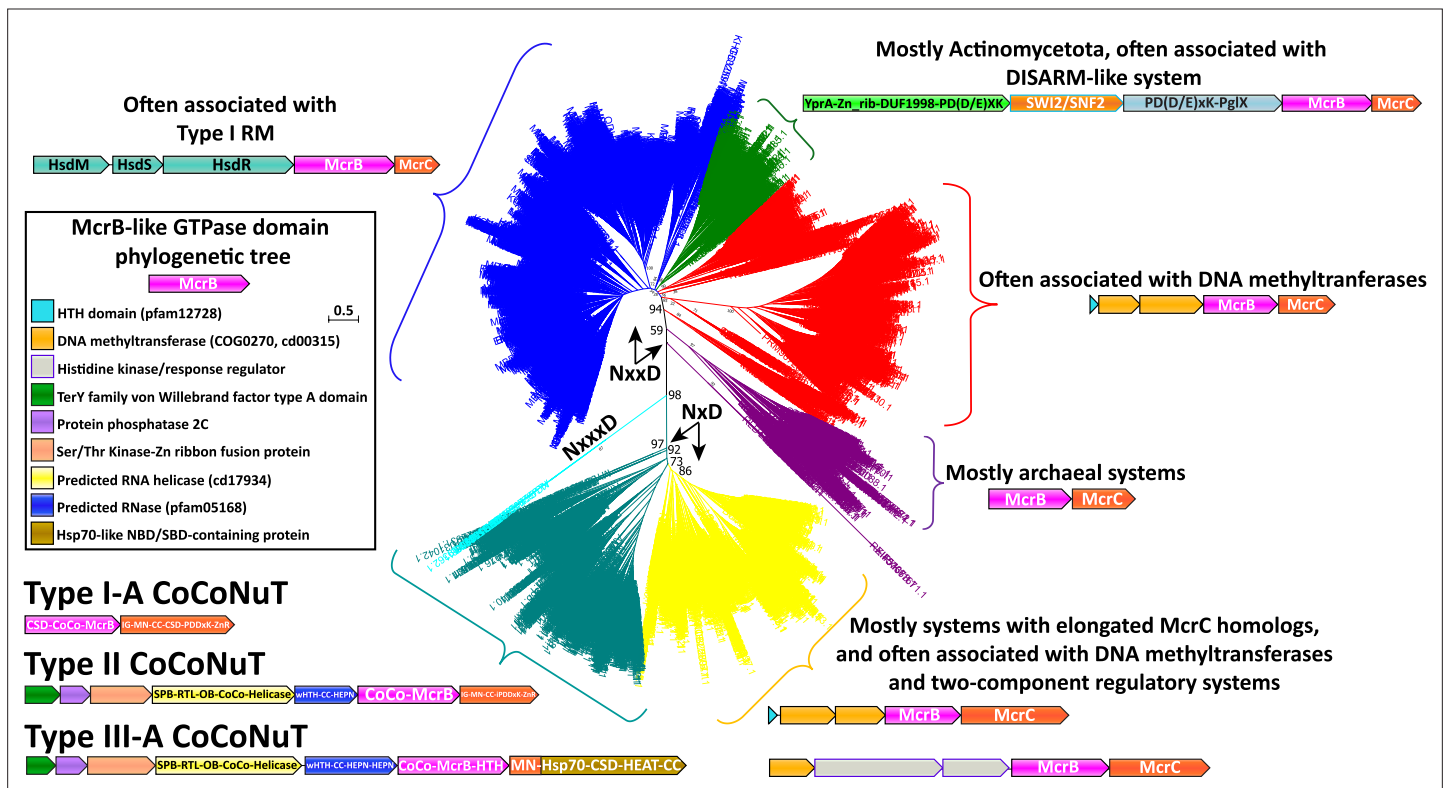


Figure 2. Phylogenetic tree of the McrB-like GTPases. The major clades in the phylogenetic tree of the McrB-like GTPases are distinguished by the distinct versions of the Nx(xx)D signature motif. The teal and yellow groups, with bootstrap support of 97%, have an NxD motif, whereas the blue, green, red, 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-like 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/E)xK, inactivated PD-(D/E)xK fold; Hsp70, 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.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Phylogenetic tree of the McrC DUF2357 domain.

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 (**Figure 2—figure supplement 1**) families. These two trees were generally topologically concordant and revealed several distinct branches not previously recognized. The branch containing the prototypical McrBC system from *E. coli* K-12 (**Figure 2**, blue) 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 (**Figure 2**, green; **Ofir et al., 2018**), and, surprisingly for a Type IV restriction system, with predicted DNA methyltransferases (**Figure 2**, red), 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 (**Figure 2**, teal and yellow) of the McrBC family is characterized by a conserved deletion within the NxxD GTPase signature motif (where x is any amino acid), found in all McrB-like GTPases, reducing it to Nx_D (**Figures 1B and 2; Neuwald et al., 1999; Iyer et al., 2004a; Erzberger and Berger, 2006; Niu et al., 2020**). The Nx_D variant of the motif is strictly conserved in these homologs and is usually, but not invariably, followed by a glutamate (E) or a second aspartate (D). No Nx_D motif McrB GTPase has been characterized, but the extensive study of the NxxD motif offers clues to the potential impact of the motif shortening. The asparagine (N) residue is strictly required for GTP binding and hydrolysis (**Pieper et al., 1999**). 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 Nx_D branch is characterized by many McrC homologs with unusual features, such as predicted RNA-binding domains, that might not be compatible with conventional McrB GTPases from the NxxD clade, which often occur in the same genomes (see below).

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

The most notable feature of the Nx_D branch of GTPases is a large clade characterized by fusion to long coiled-coil domains (**Figure 2**, teal; **Supplementary file 1**), 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-like GTPase fusions also often lack the PD-(D/E)xK 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 file 1; Jumper et al., 2021**). 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, Figure 3—figure supplement 1; Pillon et al., 2021**). Based on these features of domain architecture and genomic context, we denote these uncharacterized McrBC-containing operons CoCoNuT systems.

The deepest branching group of the Nx_D GTPase systems is typified by a fusion of an Hsp70-like ATPase nucleotide-binding domain and substrate-binding domain (NBD/SBD) to the McrB GTPase

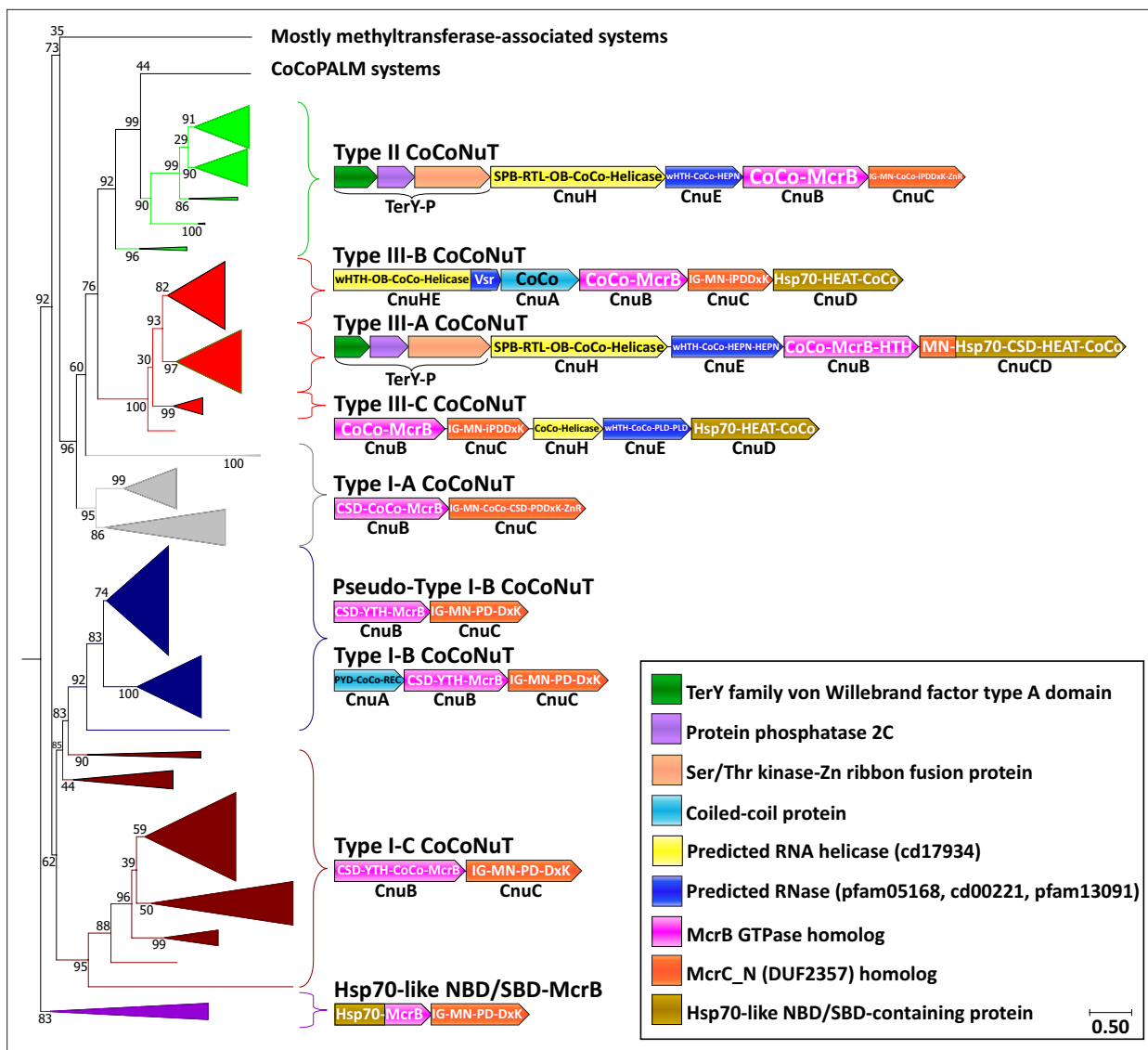


Figure 3. Coiled-coil nuclease tandem (CoCoNuT) system phylogeny and classification. The figure shows the detailed phylogeny of McrB-like GTPases from CoCoNuT systems and their close relatives. All these GTPases possess an NxD GTPase motif rather than NxxD. This tree was built from the representatives of 90% clustering of all validated homologs. Abbreviations of domains: McrB, McrB-like 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/E)xK, inactivated PD-(D/E)xK fold; REC, phosphoacceptor receiver-like domain; PLD, phospholipase D-like nuclease domain; Vsr, very-short-patch-repair PD-(D/E)xK nuclease-like domain. Underneath each gene is a proposed protein name, with Cnu as an abbreviation for CoCoNuT.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Phylogenetic tree of McrB family GTPases containing the NxD motif.

Figure supplement 2. Pseudo-Type I-B coiled-coil nuclease tandem (CoCoNuT) genomic context in *Bacillus*.

domain (Figure 3, purple, Figure 3—figure supplement 1, purple, Supplementary file 1). 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).

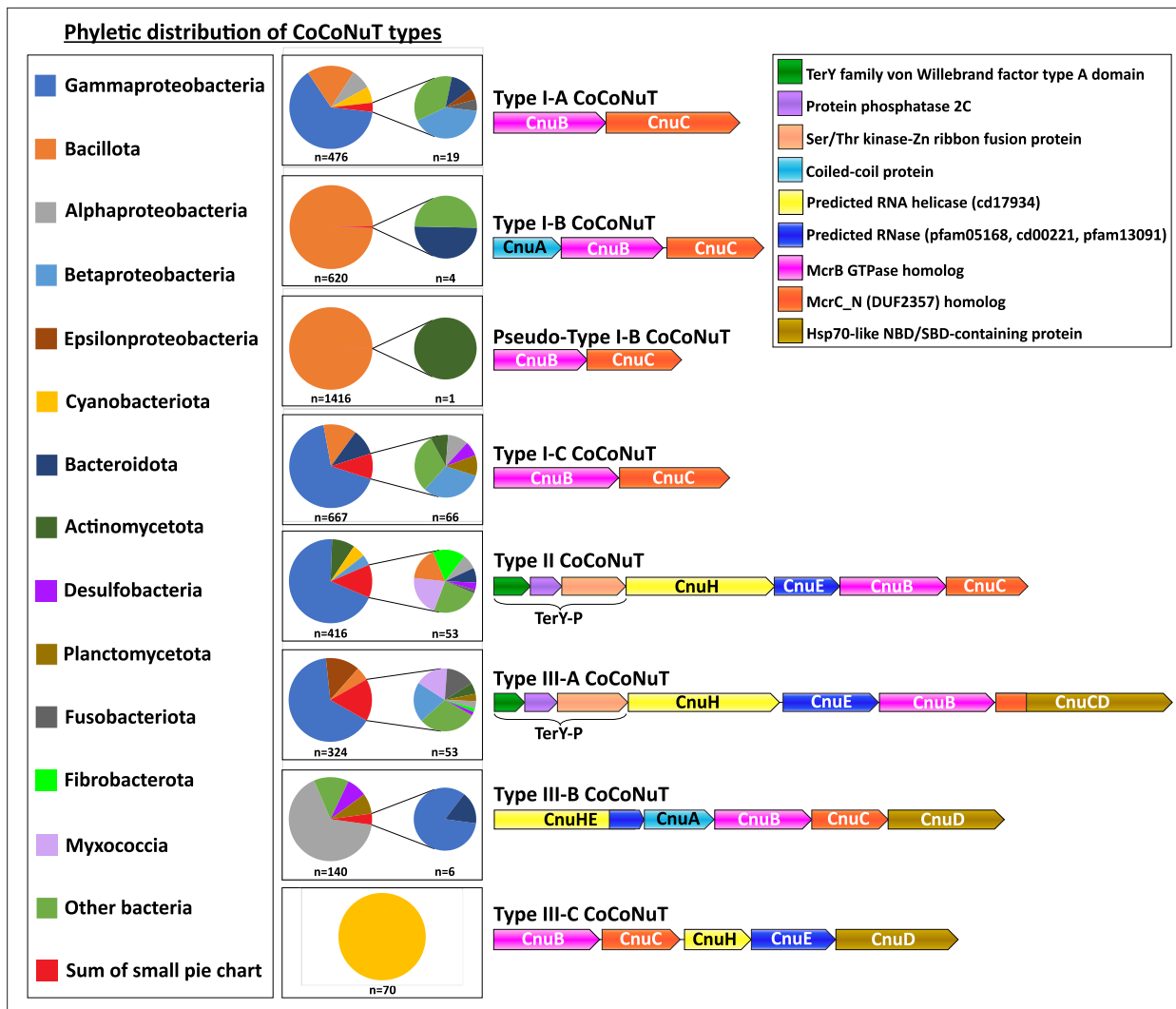


Figure 4. Phyletic distribution of coiled-coil nuclease tandems (CoCoNuTs). The phyletic distribution of CnuB/McrB-like GTPases in CoCoNuT systems found in genomic islands with distinct domain compositions. Most CoCoNuTs are found in either Bacillota or Pseudomonadota, with particular abundance in Gammaproteobacteria. 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. Types III-B and III-C are primarily found in Alphaproteobacteria and Cyanobacteriota, respectively.

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 (Figure 3—figure supplement 1, Supplementary file 1; Khadria and Senes, 2013). They are associated with signal peptidase family proteins likely to function as pilus assembly factors (Supplementary file 1; 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 (Figure 3—figure supplement 1).

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 three types and seven subtypes based on the GTPase domain phylogeny and conserved genomic context (Figure 3). Type I-A and Type I-C systems consist of McrB

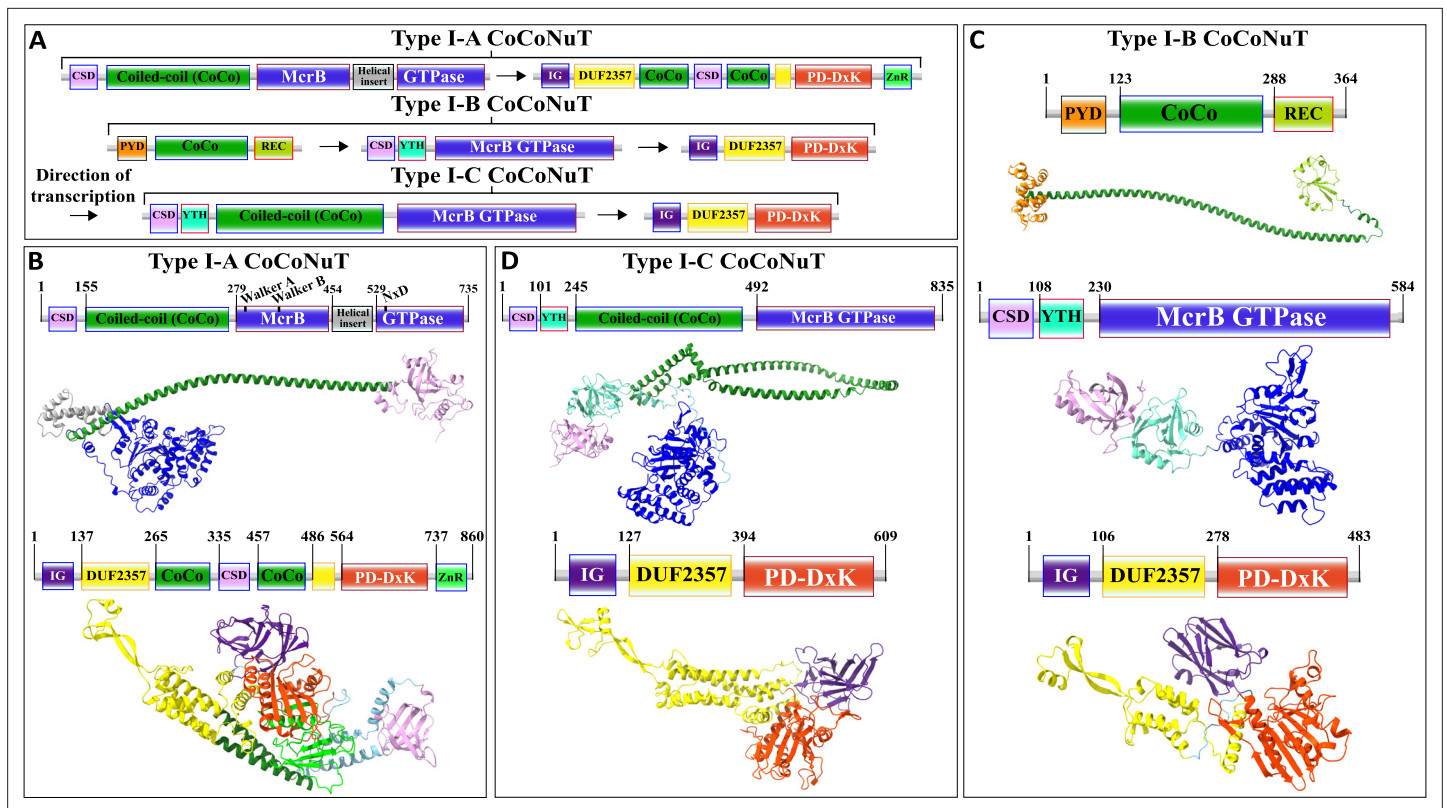


Figure 5. Domain composition, operon organization, and AlphaFold2 structural predictions of components of the Type I coiled-coil nuclease tandem (CoCoNuT) systems. **(A)** Type I CoCoNuT domain composition and operon organization. The arrows indicate the direction of transcription. **(B–D)** High-quality (average predicted local distance difference test [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 CoCoNuT systems (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 file 3** for sequences and locus tags): ROR86958.1 (Type I-A CoCoNuT CnuB), APL73566.1 (Type I-A CoCoNuT CnuC), TKH01449.1 (Type I-B CoCoNuT CnuA), GED20858.1 (Type I-B CoCoNuT CnuB), GED20857.1 (Type I-B CoCoNuT CnuC), GFD85286.1 (Type I-C CoCoNuT CnuB), and MBV0932851.1 (Type I-C CoCoNuT 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).

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Domain composition and AlphaFold2 structural predictions of components of the Type I coiled-coil nuclease tandem (CoCoNuT) systems colored by pLDDT.

Figure supplement 2. AlphaFold2 prediction of Type I-A coiled-coil nuclease tandem (CoCoNuT) CnuB GTPase hexamer and CnuC monomer complex.

Figure supplement 3. AlphaFold2 prediction of Type I-B coiled-coil nuclease tandem (CoCoNuT) CnuB hexamer and CnuC monomer complex.

Figure supplement 4. AlphaFold2 prediction of Type I-C coiled-coil nuclease tandem (CoCoNuT) CnuB GTPase hexamer and CnuC monomer complex.

and McrC homologs only, which we denote CnuB and CnuC (**Figure 3** and **Figure 5**, **Figure 5—figure supplement 1** and **2**, **Figure 5—figure supplement 4**). Type I-A is distinguished from all other CoCoNuTs by a helical insert into the CnuB GTPase domain, between the Walker B and NxD motifs (**Figure 5**, **Figure 5—figure supplement 2**).

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 GTPase domain in CnuB (**Figures 3** and **5**, **Figure 5—figure supplement 3**). 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**, **Figure 5—figure supplement 1**, **Supplementary file 1**). 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 GTPase homologs, which contain only coiled-coils fused at their N-termini (Figure 6, Figure 6—figure supplement 1, Supplementary file 1), 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 Type I-B, where the coiled-coils are encoded separately) (Figure 5, Figure 5—figure supplement 1, Supplementary file 1; 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, Figure 5—figure supplement 1, Supplementary file 1; 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 Types 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, Figure 5—figure supplements 1–4, Supplementary file 1; Anonymous, 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 (Figure 3). We hypothesize that Pseudo-Type I-B CoCoNuTs play a role in overcrowding-induced stress responses. We inferred this functional prediction from the fact that the islands in which they occur typically also encode a quorum-sensing hormone synthase, a mechanosensitive ion channel, various transporters, antibiotic resistance and synthesis factors, and cell wall-related proteins (Figure 3—figure supplement 2). In many cases, we detected Type I-B CoCoNuT homologs in the extended neighborhoods of the Pseudo-Type I-B CoCoNuTs, although only rarely in the immediate vicinity. Thus, Pseudo-Type I-B CoCoNuTs might be derived duplicates of Type I-B CoCoNuTs that acquired a specialized but likely related functionality, perhaps still using the coiled-coil protein encoded by the Type I-B CoCoNuT.

Type II and III CoCoNuT systems

Type II and III CoCoNuT CnuB/McrB GTPase domains branch from within Type I-A and are encoded in a nearly completely conserved genomic association with a Superfamily 1 (SF1) helicase of the UPF1-like clade, which we denote CnuH (Figure 3, Supplementary file 1; 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 four 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, Figure 6—figure supplement 1). 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, Figure 6—figure supplement 1). 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,

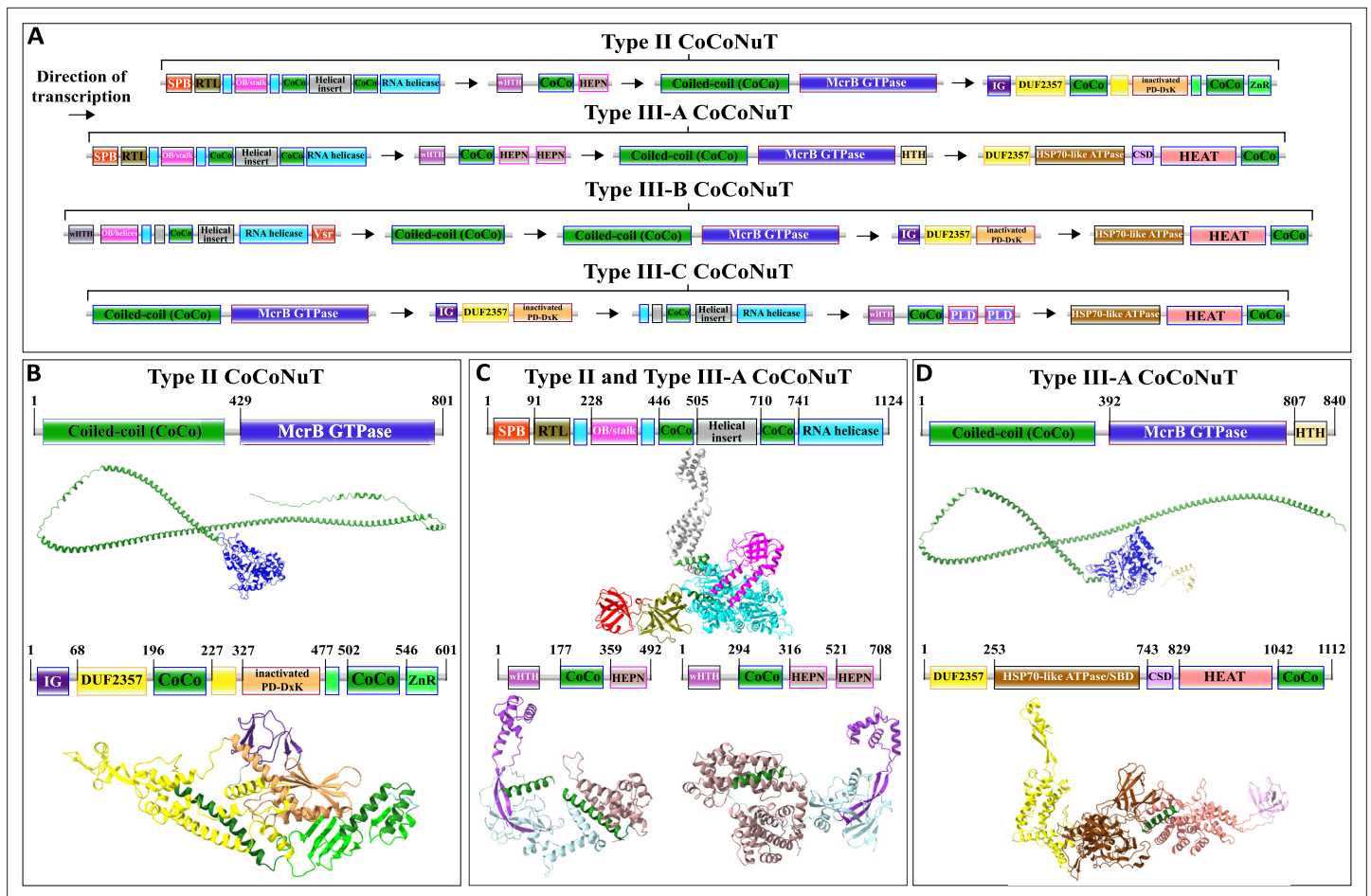


Figure 6. Domain composition, operon organization, and AlphaFold2 structural predictions for core protein components of Type II and III coiled-coil nuclease tandem (CoCoNuT) systems. **(A)** Type II and III CoCoNuT domain composition and operon organization. The arrows indicate the direction of transcription. Type II and III-A CoCoNuT 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 predicted local distance difference test [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 file 3** for sequences and locus tags): AMO81401.1 (Type II CoCoNuT CnuB), AVE71177.1 (Type II CoCoNuT CnuC), AMO81399.1 (Type II and III-A CoCoNuT CnuH), AVE71179.1 (Type II CoCoNuT CnuE), ATV59464.1 (Type III-A CnuE), PNG83940.1 (Type III-A CoCoNuT CnuB), and NMY00740.1 (Type III-A CoCoNuT 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/E)xK nuclease-like domain; PLD, phospholipase D family nuclease domain; HEAT, HEAT-like helical repeats. These structures were visualized with ChimeraX (Pettersen et al., 2021).

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Domain composition and AlphaFold2 structural predictions for core protein components of Type II and III-A coiled-coil nuclease tandem (CoCoNuT) systems colored by predicted local distance difference test (pLDDT).

Figure supplement 2. AlphaFold2 prediction of Type II coiled-coil nuclease tandem (CoCoNuT) CnuB GTPase hexamer and CnuC monomer complex.

Figure supplement 3. AlphaFold2 prediction of Type II coiled-coil nuclease tandem (CoCoNuT) CnuH helicase and CnuE effector complex.

Figure supplement 4. AlphaFold2 prediction of Type III-A coiled-coil nuclease tandem (CoCoNuT) CnuB GTPase hexamer and CnuC monomer complex.

Figure supplement 5. Comparisons of Type II and Type III-A coiled-coil nuclease tandem (CoCoNuT) N-terminal SPB domains, SmpB, and prokaryotic HIRAN domains.

Figure supplement 6. Alignment of Hsp70-like nucleotide-binding domains (NBDs) from Type III coiled-coil nuclease tandem (CoCoNuT) and related McrB homolog with *E. coli* Hsp70 (DnaK) and mammalian Hsp70 cognate protein *H. sapiens* HSC70.

in these systems, the best DALI hits are to translation factor components such as EF-Tu domain II (**Supplementary file 1; 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 file 1**).

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, Figure 6—figure supplements 1 and 5, Supplementary file 1**), 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 (**Figure 6—figure supplement 5; Bessho et al., 2007**). However, two other structural elements of SmpB involved in binding tmRNA are present (**Figure 6—figure supplement 5; 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 (**Figure 6—figure supplement 5; Chavez et al., 2018**). These HIRAN domains, however, do not overlay with the CoCoNuT domains any better than SmpB (**Figure 6—figure supplement 5**), 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, Figure 6—figure supplement 1, Supplementary file 1**). 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 Melderren, 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 file 1**).

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 (**Figures 3 and 6, Figure 6—figure supplement 1**). 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, Figure 6—figure supplement 1, Supplementary file 1**). 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 file 1**). 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 AF2 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 (**Figure 6—figure supplement 3**). 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 (**Figures 3, 5 and 6, Figure 5—figure supplement 1**). However, it is distinguished from Type I-B CnuA in containing no recognizable

domains other than the coiled-coil (**Figures 3, 5, and 6, Figure 5—figure supplement 1**). 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 (ZnR) (**Figure 3, Supplementary file 1**). TerY-P triads are involved in tellurite resistance, associated with various predicted DNA restriction and processing systems, and are hypothesized 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 file 3**). However, the complete absence of the TerY-P module in Type III-B and Type III-C systems suggests 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, Figure 6—figure supplements 1 and 4**). 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/E)xK 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 5—figure supplements 1 and 2, Figure 6, Figure 6—figure supplements 1 and 2, Supplementary file 1**). 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 5—figure supplements 1 and 2, Figure 6, Figure 6—figure supplements 1 and 2, Supplementary file 1**). 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/E)xK nuclease, and zinc ribbon found in Type I-A and Type II, but gained an Hsp70-like NBD/SBD unit similar to those fused to the McrB-like GTPase domain 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 (**Figures 3 and 5, Figure 5—figure supplement 1, Figure 6, Figure 6—figure supplements 1 and 4, Supplementary file 1**). 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 5—figure supplement 1, Figure 6, Figure 6—figure supplements 1 and 4**).

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 Types 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 Types III-B and III-C implies that fusion is not required for their functional interaction with the CnuBC/McrBC systems (**Figures 3 and 6A, Supplementary file 1**). CnuD proteins associated with both Types III-B and III-C usually contain predicted coiled-coils, suggesting that they might interact with the large coiled-coil in the CnuB homologs (**Figure 6A**).

In the CnuD homologs found in the CoCoNuTs and fused to NxMcrB 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 (**Figure 6—figure supplement 6**). 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 viral/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/E)xK catalytic motif that is required for nuclease activity, although for Type II and Types 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 file 1**). 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, Figure 6—figure supplements 1 and 4, Supplementary file 1**). 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, Figure 6—figure supplements 1 and 4, Supplementary file 1; Heinemann and Roske, 2021**). RNA targeting capability of Types 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 these hypotheses:

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-(D/E)xK 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-(D/E)xK domains might also bind DNA from which the target RNA is transcribed.

Extension of Type III-A CoCoNuT systems with ATPases and virulence factors

We observed even greater levels of complexity in Type III-A CoCoNuTs, which might ultimately beget another level of classification, where the TerY-related VWA domains were duplicated. In each of the three subtypes of Type III-A, a different ATPase was inserted between these VWA domains into the predicted operon. In these systems, the Mg²⁺-coordinating MIDAS motif (DxSxS...T...D) is perfectly conserved in the VWA domain encoded at the 5' end of the operon, which resembles the domains found in systems with only one VWA domain, whereas the internal VWA domain lacks the middle threonine and is slightly shorter (**Figure 7, Supplementary file 1; Lacy et al., 2004**). In the first two of these Type III-A subtypes shown in **Figure 7**, a CARF (CRISPR-Associated Rossmann Fold) domain-containing protein is encoded, with two distinct CARF domains encompassing different RING nuclease motifs. These CARF domains are fused at the C-terminus to a D-ExK nuclease domain, an architecture suggestive of a PCD/dormancy-eliciting antiphage effector (**Supplementary file 1; Makarova et al., 2014; Makarova et al., 2020a**). Indeed, homologs of this protein, Can1 and Can2, have been

Type III-A CoCoNuT variants

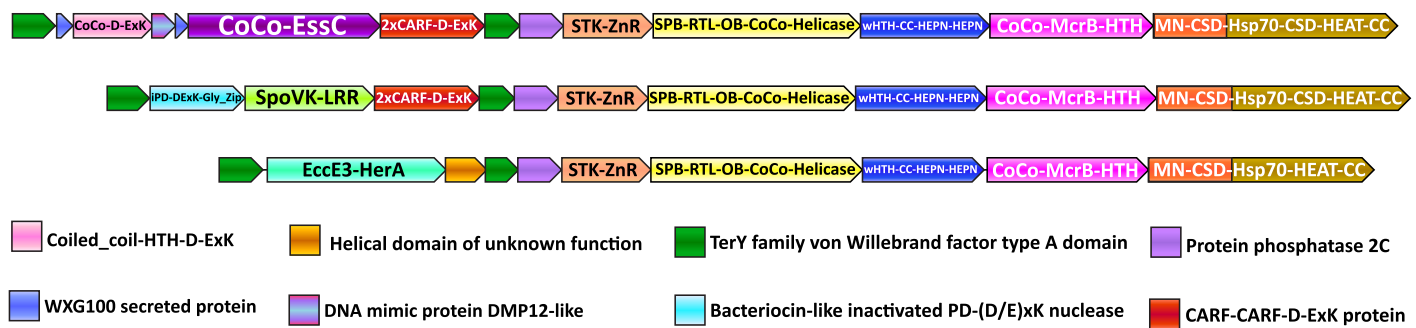


Figure 7. Compound Type III-A coiled-coil nuclease tandem (CoCoNuT) operons with 5' extensions. Additional genes that may be present in Type III-A CoCoNuT operons. Abbreviations of domains: McrB, McrB-like GTPase domain; CoCo/CC, coiled-coil; STK, serine/threonine kinase; 2xCARF, 2 CARF domains; D-ExK, D-ExK nuclease motif; MN, McrC N-terminal domain (DUF2357); CSD, cold shock domain; ZnR, zinc ribbon domain; SPB, SmpB-like domain; RTL, RNase toxin-like domain; OB, OB-fold domain attached to helical stalk-like extension of ATPase; HEPN, HEPN family nuclease domain; Hsp70, Hsp70-like NBD/SBD; HEAT, HEAT-like helical repeats; LRR, leucine-rich repeat; Gly_zip, glycine zipper domain; SpoVK, EssC, EccE3-HerA – see text.

characterized as CRISPR ancillary nucleases (McMahon *et al.*, 2020; Zhu *et al.*, 2021), and Can2 has been reported to cleave both DNA and RNA (Zhu *et al.*, 2021). In these two CoCoNuT varieties, the CnuC/McrC homologs typically contain two CSDs, whereas in the third type, where the CARF proteins are not encoded, there is a single CSD (Figure 7, Supplementary file 1).

The first of the three variable regions flanked by the VWA domain-encoding genes includes an FtsK-like ATPase homologous to the Type VII secretion system factor EssC (Figure 7, Supplementary file 1). EssC contains three tandem ATPase domains (D1, D2, and D3), with the Walker A/B motifs required for ATP hydrolysis present only in D1 and D2 (Warne *et al.*, 2016; Bobrovskyy *et al.*, 2022). The CoCoNuT-associated homologs also possess three ATPase domains but differ in that only the central D2 domain homolog is predicted to be active. They are also distinguished from EssC by the presence of a coiled-coil that can exceed 200 residues in length and is fused at their N-terminus, whereas forkhead-associated domains and transmembrane helices are found in this position in EssC (Supplementary file 1; Bobrovskyy *et al.*, 2022; Warne *et al.*, 2016). In addition, the region codes for two WXG100 proteins, one with the characteristic WxG motif and the other without (but with a similar predicted structure), as well as a coiled-coil fused to a restriction endonuclease-like domain with a D-ExK catalytic motif (Figure 7, Supplementary file 1; Poulsen *et al.*, 2014). Finally, another protein similar to DNA mimics that bind the HU histone-like factor is encoded following the coiled-coil-nuclease fusion (Figure 7, Supplementary file 1; Wang *et al.*, 2013).

It appears likely that some or all of these factors are secreted, especially the WXG100 proteins, which are known to be secreted, with the prototypical example, ESAT-6, being a T-cell antigen diagnostic of *Mycobacterium tuberculosis* infection (Poulsen *et al.*, 2014). These associated WXG100 proteins implicate the ATPases in defensive protein secretion, but as they lack transmembrane domains present in their EssC homologs, a different mechanism appears likely. The presence of coiled-coils at the N-termini of these proteins suggests that they might interact with the coiled-coils in the core CoCoNuT factors and/or with the associated coiled-coil-nuclease fusion. The FtsK superfamily ATPases form hexamers (Bobrovskyy *et al.*, 2022), so should such an interaction occur, they are likely compatible with the CnuB/McrB GTPase hexamer. A prior study that tangentially examined these operons in the context of the TerY-P triad pointed out that this WXG100/FtsK-like ATPase operon is likely a mobile element that can be found as a stand-alone secretion system in other genomes (Anantharaman *et al.*, 2012).

In the second of these Type III-A CoCoNuT extensions with two VWA domains, a SpoVK family of AAA+ATPases homologous to p97/CDC48 is encoded adjacent to a protein containing a C-terminal bacteriocin-like glycine zipper motif (Figure 7, Supplementary file 1). CDC48 is involved in eukaryotic protein quality control, particularly the degradation of proteins synthesized from non-stop mRNA,

where it is required to release nascent polypeptides from stalled ribosomes to enable proteolysis (Verma *et al.*, 2013). CDC48 contains two tandem ATPase domains, both of which are active; the CoCoNuT-associated homologs also contain tandem ATPases, but the Walker A/B motifs required to bind and hydrolyze ATP are conserved only in the C-terminal domain (Baek *et al.*, 2013; Wolf and Stolz, 2012). As members of the AAA+ superfamily, these ATPases assemble into hexamers (Wolf and Stolz, 2012), similarly to the McrB family GTPases. At their C-termini, these ATPases are fused to domains of unknown function, namely, a leucine-rich repeat element and a beta-barrel domain structurally similar to biotin carrier proteins, suggesting that these proteins might be biotinylated (Figure 7, Supplementary file 1; Choi-Rhee and Cronan, 2003). The conserved association of the CDC48-like ATPases with these Type III-A CoCoNuTs, which encode the potentially tmRNA-binding SPB domain, seems to provide support for the scenario of tmRNA interaction. Structural analysis of the bacteriocin-like protein encoded in these loci indicates that it adopts an inactivated PD-(D/E)xK-type restriction endonuclease fold, potentially nucleic acid-binding (Supplementary file 1). These proteins might be analogous to the restriction endonuclease-like factors in the FtsK/EssC homolog neighborhoods (Figure 7).

The third variant of these extended Type III-A CoCoNuTs encodes a distinct member of the FtsK/HerA superfamily, which is also likely assembled into hexamers (Figure 7, Supplementary file 1; Iyer *et al.*, 2004b). AF2 structural modeling suggests these enzymes are homologs of the ESX-3 Type VII secretion system factor EccE3 (Poweleit *et al.*, 2019), albeit containing a unique beta-strand insertion of variable length. This EccE3-like domain is fused to an ATPase domain similar to the Type IV secretion system protein VirB4, which is involved in bacterial conjugation (Figure 7, Supplementary file 1; Wallden *et al.*, 2010). These systems also encode a small helical domain of unknown function in operonic association with the ATPase (Figure 7). These genes might be involved in the mobilization of the locus via conjugation or instead play a similar role in secretion as predicted for the EssC-like ATPases. However, WXG100 homologs, like those that strongly imply a secretion-related function for the EssC-like ATPases, are not encoded near these VirB4-like ATPases. Lastly, we observed that, unlike the EssC-like ATPases and the SpoVK-like proteins, these enzymes are not always encoded between VWA genes at the 5' end of the operons, but in some cases, migrated to the 3' end; in these cases, however, duplicated VWA domains are present at the 5' end, a potential vestige of an extension that was recently lost or relocated.

Overall, these elaborations of Type III-A CoCoNuT systems resemble the TerY-P triads in that they could be stand-alone defensive cassettes that augment the effectiveness of the core CoCoNuT systems. It is unclear, however, why these types of factors are flanked by TerY-like VWA domains, as opposed to restriction systems such as Type I RM, GmrSD, and Druantia Type III, which are commonly associated with Type III-A CoCoNuTs as well, but are never so tightly integrated into the operon (Loenen *et al.*, 2014; Weigle and Raleigh, 2016; Doron *et al.*, 2018). Type III-A systems embedded in these extended operons are annotated in Supplementary file 3.

While investigating this additional diversity of Type III-A CoCoNuTs, we observed that Type III-A CoCoNuTs in *Helicobacter* appeared to be translated using an alternate genetic code because gene predictions with the standard code divided the expected open-reading frames into many small fragments. We were unable to identify a known alternative code that would yield the expected CoCoNuT gene products. Thus, a novel type of conditional or otherwise complex translation regulation likely occurs in these species, perhaps triggered by phage infection (for the accessions of identifiable Type III-A CoCoNuT factors in *Helicobacter*, see Supplementary file 3).

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 domain-containing proteins (Figures 7 and 8, Figure 8—figure supplement 1). 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

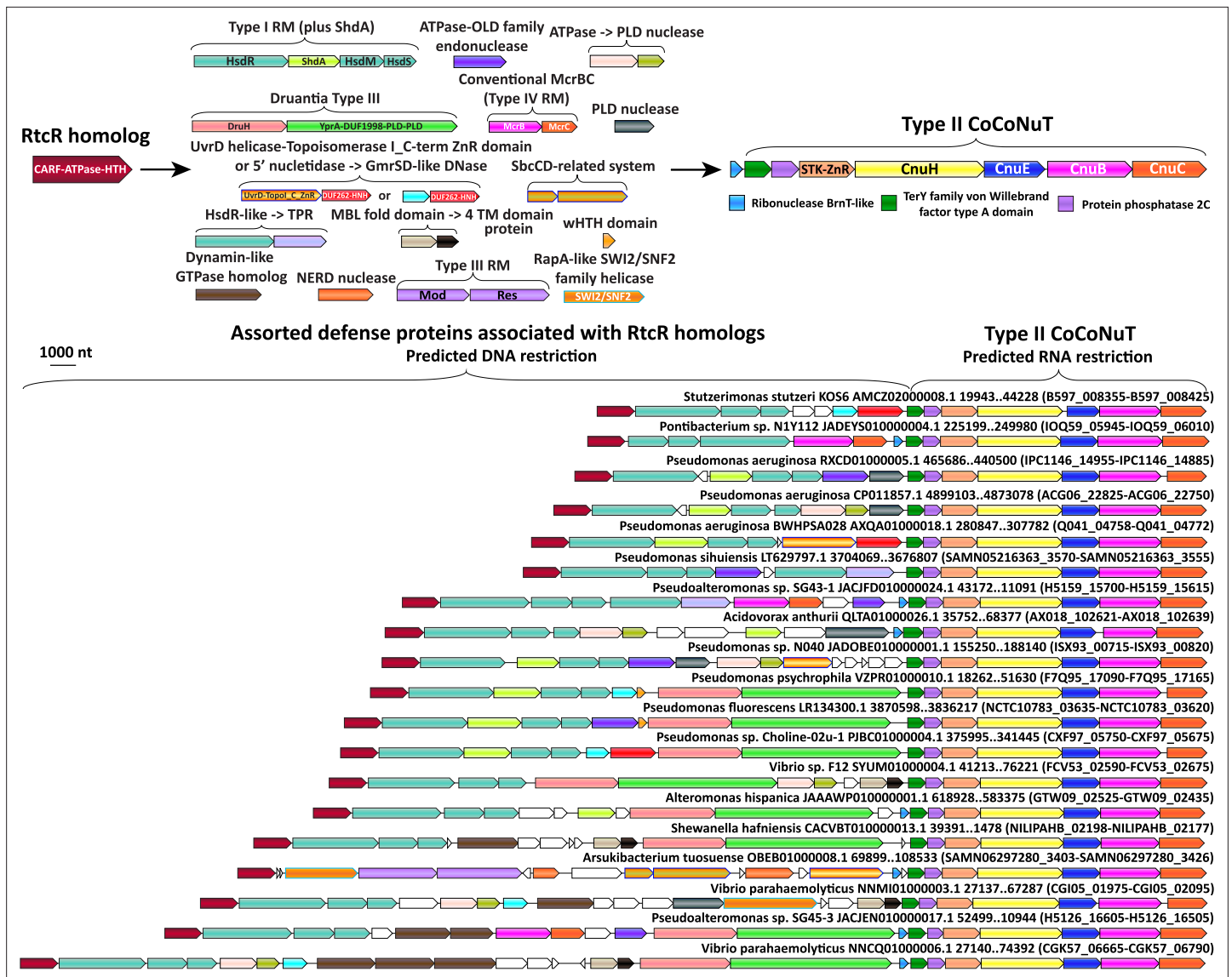


Figure 8. Complex operonic associations of Type II coiled-coil nuclease tandems (CoCoNuTs). Type II CoCoNuTs are frequently associated with RtcR homologs, and in many cases, ancillary defense genes are located between the RtcR gene and the CoCoNuT, almost always oriented in the same direction in an apparent superoperon. Abbreviations of domains: STK, serine/threonine kinase; ZnR, zinc ribbon domain; 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 four predicted transmembrane helices; Mod/Res, Type III RM modification and restriction factors.

The online version of this article includes the following figure supplement(s) for figure 8:

Figure supplement 1. Type II coiled-coil nuclease tandems (CoCoNuTs) are associated with RtcR homologs in a variety of species.

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 (**Figure 8, Figure 8—figure supplement 1, Supplementary file 1**). 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 (**Figure 8, Figure 8—figure supplement 1; 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) signal molecules by Cas10, and these bind the CARF domain of PCD effectors, such as Csm6, activating their nuclease moieties, typically HEPN domains that function as promiscuous RNases (**Pillon et al., 2021; Millman et al., 2020**). One of the two outcomes can result from this cascade: infection is either eradicated quickly by 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 else, 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 (**Figure 8—figure supplement 1**), there are also numerous cases where several types of DNA restriction systems are encoded between the CARF gene and the CoCoNuT (**Figure 8**). In these cases, nearly all genes are in an apparent operonic organization that can extend upward of 40 kb (**Figure 8**). 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 by the HEPN RNase(s) of the CoCoNuT. Another noteworthy observation consistent with PCD induction is the frequent presence of

BrnT-like RelE family RNase toxins encoded in the same direction and immediately upstream of RtcR-associated CoCoNuTs (**Figure 8**, **Figure 8—figure supplement 1**, **Supplementary file 1**). A comparison of these sequences with BrnT shows notable conservation in the RNA-binding regions, but not all residues required for toxicity of BrnT are conserved (**Heaton et al., 2012**). However, as described above in reference to the RelE-like RTL domain in CnuH, nucleases of this family can vary considerably in sequence while remaining active toxins (**Guglielmini and Van Melderer, 2011**; **Goeders et al., 2013**).

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 8**). 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 8**).

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 gene 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 8**), 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**).

Conclusion

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 three 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 system 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 potentially endowed with DNA-targeting activity as well, either by factors integral to the system, such as the McrC-like nuclease, or more loosely associated, such as Type I RM and *Druantia* Type III systems that are encoded in the same predicted superoperons 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/E)xK nuclease domain), PF04411 (McrC PD-(D/E)xK nuclease domain), and 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.*, 2004a). 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_LlaJl). These aliases were determined from a preliminary manual investigation of the data using HHpred (Zimmermann *et al.*, 2018). 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, 2005). 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 three 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 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.

Protein domain detection and annotation

Protein domains in McrBC homologs and 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 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., 2018*). The CoCoNuT system components were subjected to additional scrutiny using the coiled-coil detection and visualization tool Waggawagga, which employs 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 from the four searches, made a non-redundant set, and annotated them using CDD profiles, as described above. Then, we aligned them with MUSCLE v5 (*Edgar, 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 structures

Protein structures were predicted using AlphaFold2 (AF2) v2.2.0 with local installations of complete databases required for AF2 (*Jumper et al., 2021*). Only single protein models with average predicted local distance difference test (pLDDT) scores ≥ 80 were retained for further analysis, and among the multimer models analyzed, all had average pLDDT scores ≥ 76 , with only two scores <80 (*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., 2024*). Structure visualizations and comparisons were performed with ChimeraX (*Pettersen et al., 2021*) and the RCSB PDB website (*Berman et al., 2000*).

Acknowledgements

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

Additional information

Funding

Funder	Grant reference number	Author
National Institutes of Health	Intramural Research Program	Eugene V Koonin

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Ryan T Bell, Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review and editing; Harutyun Sahakyan, Yuri I Wolf, Investigation, Methodology, Writing – review and editing; Kira S Makarova, Data curation, Investigation, Writing – review and editing; Eugene V Koonin, Conceptualization, Supervision, Investigation, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Ryan T Bell  <http://orcid.org/0000-0003-1249-8398>

Eugene V Koonin  <http://orcid.org/0000-0003-3943-8299>

Peer review material

Reviewer #1 (Public review): <https://doi.org/10.7554/eLife.94800.3.sa1>

Reviewer #2 (Public review): <https://doi.org/10.7554/eLife.94800.3.sa2>

Author response <https://doi.org/10.7554/eLife.94800.3.sa3>

Additional files

Supplementary files

- Supplementary file 1. Protein structure prediction and analysis for CoCoNuT systems components.
- Supplementary file 2. List of AlphaFold 2 models for CoCoNuT protein components and their complexes, with modelarchive accession numbers.
- Supplementary file 3. GenBank accession numbers and protein sequences for protein components of the CoCoNuT systems.
- MDAR checklist

Data availability

All data generated and analyzed in this study are included in the manuscript and supporting files. Domain identification statistics are listed in Supplementary file 1. The AlphaFold2 structural models generated and presented in the figures are available at <https://modelarchive.org/> with accessions

listed in Supplementary file 2. The CoCoNuT systems are documented in detail in Supplementary file 3. The code generated during this work is available at <https://doi.org/10.5281/zenodo.10971641>.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Bell et al.	2024	CoCoNuTs are a diverse subclass of Type IV restriction systems predicted to target RNA	https://doi.org/10.5281/zenodo.10971641	Zenodo, 10.5281/zenodo.10971641

References

- Agrawal N**, Dasaradhi PVN, Mohammed A, Malhotra P, Bhatnagar RK, Mukherjee SK. 2003. RNA interference: biology, mechanism, and applications. *Microbiology and Molecular Biology Reviews* **67**:657–685. DOI: <https://doi.org/10.1128/MMBR.67.4.657-685.2003>, PMID: 14665679
- Altschul SF**, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389–3402. DOI: <https://doi.org/10.1093/nar/25.17.3389>, PMID: 9254694
- Amir M**, Kumar V, Dohare R, Islam A, Ahmad F, Hassan M. 2018. Sequence, structure and evolutionary analysis of cold shock domain proteins, a member of OB fold family. *Journal of Evolutionary Biology* **31**:1903–1917. DOI: <https://doi.org/10.1111/jeb.13382>, PMID: 30267552
- Anantharaman V**, Iyer LM, 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. DOI: <https://doi.org/10.1039/c2mb25239b>, PMID: 23044854
- Anonymous**. 2015. Encyclopedia of life sciences. *Immunoglobulin Superfamily*. Encyclopedia of Life Sciences. DOI: <https://doi.org/10.1002/047001590X>
- Baek GH**, Cheng H, Choe V, Bao X, Shao J, Luo S, Rao H. 2013. Cdc48: a swiss army knife of cell biology. *Journal of Amino Acids* **2013**:183421. DOI: <https://doi.org/10.1155/2013/183421>, PMID: 24167726
- Barends S**, Karzai AW, Sauer RT, Wower J, Kraal B. 2001. Simultaneous and functional binding of SmpB and EF-Tu-GTP to the alanyl acceptor arm of tmRNA. *Journal of Molecular Biology* **314**:9–21. DOI: <https://doi.org/10.1006/jmbi.2001.5114>
- Bell RT**, Wolf YI, Koonin EV. 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 Biology* **18**:159. DOI: <https://doi.org/10.1186/s12915-020-00885-2>, PMID: 33148243
- Berman HM**, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Research* **28**:235–242. DOI: <https://doi.org/10.1093/nar/28.1.235>, PMID: 10592235
- Bessho Y**, Shibata R, Sekine S, 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. *PNAS* **104**:8293–8298. DOI: <https://doi.org/10.1073/pnas.0700402104>, PMID: 17488812
- Bitton L**, Klaiman D, Kaufmann G. 2015. Phage T4-induced DNA breaks activate a tRNA repair-defying anticodon nuclease. *Molecular Microbiology* **97**:898–910. DOI: <https://doi.org/10.1111/mmi.13074>, PMID: 26031711
- Bobrovskyy M**, Oh SY, Missiakas D. 2022. Contribution of the EssC ATPase to the assembly of the type 7b secretion system in *Staphylococcus aureus*. *The Journal of Biological Chemistry* **298**:102318. DOI: <https://doi.org/10.1016/j.jbc.2022.102318>, PMID: 35921891
- Bourne HR**, Sanders DA, McCormick F. 1991. The GTPase superfamily: conserved structure and molecular mechanism. *Nature* **349**:117–127. DOI: <https://doi.org/10.1038/349117a0>
- Burroughs AM**, Zhang D, Schäffer DE, Iyer LM, 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. DOI: <https://doi.org/10.1093/nar/gkv1267>, PMID: 26590262
- 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. *Molecular Cell* **41**:693–703. DOI: <https://doi.org/10.1016/j.molcel.2011.02.010>
- 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. *The Journal of Biological Chemistry* **283**:15502–15511. DOI: <https://doi.org/10.1074/jbc.M708992200>, PMID: 18400763
- Chavez DA**, Greer BH, Eichman BF. 2018. The HIRAN domain of helicase-like transcription factor positions the DNA translocase motor to drive efficient DNA fork regression. *The Journal of Biological Chemistry* **293**:8484–8494. DOI: <https://doi.org/10.1074/jbc.RA118.002905>, PMID: 29643183
- Cheng Z**, Muhlrad D, Lim MK, Parker R, Song H. 2007. Structural and functional insights into the human Upf1 helicase core. *The EMBO Journal* **26**:253–264. DOI: <https://doi.org/10.1038/sj.emboj.7601464>, PMID: 17159905
- Choi-Rhee E**, Cronan JE. 2003. The biotin carboxylase-biotin carboxyl carrier protein complex of *Escherichia coli* acetyl-coa carboxylase. *Journal of Biological Chemistry* **278**:30806–30812. DOI: <https://doi.org/10.1074/jbc.M302507200>

- Colicelli J.** 2004. Human RAS superfamily proteins and related GTPases. *Science's STKE* **2004**:RE13. DOI: <https://doi.org/10.1126/stke.2502004re13>, PMID: 15367757
- 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**:1355–1362. DOI: <https://doi.org/10.1261/rna.039917.113>, PMID: 23945037
- Delorenzi M, Speed T.** 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**:617–625. DOI: <https://doi.org/10.1093/bioinformatics/18.4.617>, PMID: 12016059
- Dila D, Sutherland E, Moran L, Slatko B, Raleigh EA.** 1990. Genetic and sequence organization of the mcrBC locus of *Escherichia coli* K-12. *Journal of Bacteriology* **172**:4888–4900. DOI: <https://doi.org/10.1128/jb.172.9.4888-4900.1990>
- Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R.** 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**:eaar4120. DOI: <https://doi.org/10.1126/science.aar4120>, PMID: 29371424
- Dovas A, Couchman JR.** 2005. RhoGDI: multiple functions in the regulation of Rho family GTPase activities. *Biochemical Journal* **390**:1–9. DOI: <https://doi.org/10.1042/BJ20050104>
- Edgar RC.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**:113. DOI: <https://doi.org/10.1186/1471-2105-5-113>, PMID: 15318951
- Edgar RC.** 2022. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature Communications* **13**:6968. DOI: <https://doi.org/10.1038/s41467-022-34630-w>, PMID: 36379955
- Erzberger JP, Berger JM.** 2006. Evolutionary relationships and structural mechanisms of aaa+ proteins. *Annual Review of Biophysics and Biomolecular Structure* **35**:93–114. DOI: <https://doi.org/10.1146/annurev.biophys.35.040405.101933>
- Fairman-Williams ME, Guenther U-P, Jankowsky E.** 2010. SF1 and SF2 helicases: family matters. *Current Opinion in Structural Biology* **20**:313–324. DOI: <https://doi.org/10.1016/j.sbi.2010.03.011>, PMID: 20456941
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC.** 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**:806–811. DOI: <https://doi.org/10.1038/35888>
- Fleischman RA, Cambell JL, Richardson CC.** 1976. Modification and restriction of T-even bacteriophages: in vitro degradation of deoxyribonucleic acid containing 5-hydroxymethylctosine. *Journal of Biological Chemistry* **251**:1561–1570. DOI: [https://doi.org/10.1016/S0021-9258\(17\)33685-2](https://doi.org/10.1016/S0021-9258(17)33685-2)
- Gabler F, Nam S, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V.** 2020. Protein sequence analysis using the mpi bioinformatics toolkit. *Current Protocols in Bioinformatics* **72**:e108. DOI: <https://doi.org/10.1002/cpbi.108>
- Gao L, Altae-Tran H, Böhning F, Makarova KS, Segel M, Schmid-Burgk JL, Koob J, Wolf YI, Koonin EV, Zhang F.** 2020. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**:1077–1084. DOI: <https://doi.org/10.1126/science.aba0372>, PMID: 32855333
- Gasiunas G, Barrangou R, Horvath P, Siksnys V.** 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS* **109**:E2579–E2586. DOI: <https://doi.org/10.1073/pnas.1208507109>, PMID: 22949671
- Genschik P, Drabikowski K, Filipowicz W.** 1998. Characterization of the *Escherichia coli* RNA 3'-terminal phosphate cyclase and its ζ 54-regulated operon. *Journal of Biological Chemistry* **273**:25516–25526. DOI: <https://doi.org/10.1074/jbc.273.39.25516>
- Goeders N, Drèze P-L, Van Melderen L.** 2013. Relaxed cleavage specificity within the RelE toxin family. *Journal of Bacteriology* **195**:2541–2549. DOI: <https://doi.org/10.1128/JB.02266-12>, PMID: 23543711
- 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. *The EMBO Journal* **34**:169–183. DOI: <https://doi.org/10.15252/embj.201489455>, PMID: 25452498
- Gorbalenya AE, Koonin EV.** 1993. Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology* **3**:419–429. DOI: [https://doi.org/10.1016/S0959-440X\(05\)80116-2](https://doi.org/10.1016/S0959-440X(05)80116-2)
- Gucinski GC, Michalska K, Garza-Sánchez F, Eschenfeldt WH, Stols L, Nguyen JY, Goulding CW, Joachimiak A, Hayes CS.** 2019. Convergent evolution of the barnase/endou/colicin/rele (becr) fold in antibacterial trnase toxins. *Structure* **27**:1660–1674. DOI: <https://doi.org/10.1016/j.str.2019.08.010>, PMID: 31515004
- Guglielmini J, Van Melderen L.** 2011. Bacterial toxin-antitoxin systems: Translation inhibitors everywhere. *Mobile Genetic Elements* **1**:283–290. DOI: <https://doi.org/10.4161/mge.18477>, PMID: 22545240
- Gutmann S, Haebel PW, 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. DOI: <https://doi.org/10.1038/nature01831>
- Guyomar C, D'Urso G, Chat S, Giudice E, Gillet R.** 2021. Structures of tmRNA and SmpB as they transit through the ribosome. *Nature Communications* **12**:4909. DOI: <https://doi.org/10.1038/s41467-021-24881-4>, PMID: 34389707
- Halaby DM, Poupon A, Mornon J-P.** 1999. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Engineering, Design and Selection* **12**:563–571. DOI: <https://doi.org/10.1093/protein/12.7.563>

- 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. DOI: <https://doi.org/10.1093/bioinformatics/btw006>
- Hazra D**, Chapat C, Graille M. 2019. m. *Genes* **10**:49. DOI: <https://doi.org/10.3390/genes10010049>, PMID: 30650668
- Heaton BE**, Herrou J, Blackwell AE, Wysocki VH, Crosson S. 2012. Molecular structure and function of the novel BrnT/BrnA toxin-antitoxin system of *Brucella abortus*. *The Journal of Biological Chemistry* **287**:12098–12110. DOI: <https://doi.org/10.1074/jbc.M111.332163>, PMID: 22334680
- Heinemann U**, Roske Y. 2021. Cold-shock domains-abundance, structure, properties, and nucleic-acid binding. *Cancers* **13**:190. DOI: <https://doi.org/10.3390/cancers13020190>, PMID: 33430354
- Himeno H**, Kurita D, Muto A. 2014. tmRNA-mediated trans-translation as the major ribosome rescue system in a bacterial cell. *Frontiers in Genetics* **5**:66. DOI: <https://doi.org/10.3389/fgene.2014.00066>, PMID: 24778639
- Holm L**. 2020. DALI and the persistence of protein shape. *Protein Science* **29**:128–140. DOI: <https://doi.org/10.1002/pro.3749>, PMID: 31606894
- Hughes KJ**, Chen X, Burroughs AM, Aravind L, Wolin SL. 2020. An rna repair operon regulated by damaged trnas. *Cell Reports* **33**:108527. DOI: <https://doi.org/10.1016/j.celrep.2020.108527>, PMID: 33357439
- Ipsaro JJ**, Haase AD, Knott SR, Joshua-Tor L, Hannon GJ. 2012. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* **491**:279–283. DOI: <https://doi.org/10.1038/nature11502>, PMID: 23064227
- Irastortza-olaziregi M**, Amster-choder O. 2020. Coupled transcription-translation in prokaryotes: an old couple with new surprises. *Frontiers in Microbiology* **11**:624830. DOI: <https://doi.org/10.3389/fmicb.2020.624830>
- Ishita J**, Matvey K, Leonid M, Natalia M, Alexandr K, Sofia M, Konstantin K, Sergei B, Kira SM, Eugene VK, Konstantin S, Ekaterina S. 2021. tRNA anticodon cleavage by target-activated crispr-cas13a effector. *Science Advances* **10**:eadl0164. DOI: <https://doi.org/10.1126/sciadv.adl0164>
- Iyer LM**, Leipe DD, Koonin EV, Aravind L. 2004a. Evolutionary history and higher order classification of AAA+ ATPases. *Journal of Structural Biology* **146**:11–31. DOI: <https://doi.org/10.1016/j.jsb.2003.10.010>
- Iyer LM**, Makarova KS, Koonin EV, Aravind L. 2004b. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Research* **32**:5260–5279. DOI: <https://doi.org/10.1093/nar/gkh828>, PMID: 15466593
- Jinek M**, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**:816–821. DOI: <https://doi.org/10.1126/science.1225829>, PMID: 22745249
- Jumper J**, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. DOI: <https://doi.org/10.1038/s41586-021-03819-2>, PMID: 34265844
- Kalathiya U**, Padariya M, Pawlicka K, Verma CS, Houston D, Hupp TR, Alfaro JA. 2019. Insights into the effects of cancer associated mutations at the upf2 and atp-binding sites of nmd master regulator: upf1. *International Journal of Molecular Sciences* **20**:5644. DOI: <https://doi.org/10.3390/ijms20225644>, PMID: 31718065
- Kaufmann G**. 2000. Anticodon nucleases. *Trends in Biochemical Sciences* **25**:70–74. DOI: [https://doi.org/10.1016/S0968-0004\(99\)01525-X](https://doi.org/10.1016/S0968-0004(99)01525-X)
- Kaur G**, Burroughs AM, Iyer LM, Aravind L. 2020. Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. *eLife* **9**:e52696. DOI: <https://doi.org/10.7554/eLife.52696>, PMID: 32101166
- Khadria AS**, Senes A. 2013. The transmembrane domains of the bacterial cell division proteins ftsb and ftsl form a stable high-order oligomer. *Biochemistry* **52**:7542–7550. DOI: <https://doi.org/10.1021/bi4009837>
- Kishor A**, Tandukar B, Ly YV, Toth EA, Suarez Y, Brewer G, Wilson GM. 2013. Hsp70 is a novel posttranscriptional regulator of gene expression that binds and stabilizes selected mrnas containing au-rich elements. *Molecular and Cellular Biology* **33**:71–84. DOI: <https://doi.org/10.1128/MCB.01275-12>
- Kishor A**, White EJF, Matsangos AE, Yan Z, Tandukar B, Wilson GM. 2017. Hsp70's RNA-binding and mRNA-stabilizing activities are independent of its protein chaperone functions. *The Journal of Biological Chemistry* **292**:14122–14133. DOI: <https://doi.org/10.1074/jbc.M117.785394>, PMID: 28679534
- Koonin EV**, Aravind L. 2002. Origin and evolution of eukaryotic apoptosis: the bacterial connection. *Cell Death & Differentiation* **9**:394–404. DOI: <https://doi.org/10.1038/sj.cdd.4400991>
- Koonin EV**, Zhang F. 2017. Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. *BioEssays* **39**:1–9. DOI: <https://doi.org/10.1002/bies.201600186>, PMID: 27896818
- Koonin EV**, Krupovic M. 2019. Origin of programmed cell death from antiviral defense? *PNAS* **116**:16167–16169. DOI: <https://doi.org/10.1073/pnas.1910303116>, PMID: 31289224
- Kotta-Loizou I**, Giuliano MG, Jovanovic M, Schaefer J, Ye F, Zhang N, Irakleidi DA, 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. DOI: <https://doi.org/10.1016/j.isci.2022.105425>, PMID: 36388977
- Lacy DB**, Wigelsworth DJ, Scobie HM, Young JAT, Collier RJ. 2004. Crystal structure of the von Willebrand factor A domain of human capillary morphogenesis protein 2: an anthrax toxin receptor. *PNAS* **101**:6367–6372. DOI: <https://doi.org/10.1073/pnas.0401506101>, PMID: 15079089
- 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 Research* **39**:1187–1196. DOI: <https://doi.org/10.1093/nar/gkq958>, PMID: 20961958

- Liao S, Sun H, Xu C. 2018. Yth domain: a family of *m* 6 -methyladenosine (*m* 6 a) readers. *Genomics, Proteomics & Bioinformatics* **16**:99–107. DOI: <https://doi.org/10.1016/j.gpb.2018.04.002>
- Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG, Murray NE. 2014. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Research* **42**:3–19. DOI: <https://doi.org/10.1093/nar/gkt990>, PMID: 24141096
- Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research* **48**:D265–D268. DOI: <https://doi.org/10.1093/nar/gkz991>, PMID: 31777944
- Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. *Science* **252**:1162–1164. DOI: <https://doi.org/10.1126/science.252.5009.1162>, PMID: 2031185
- Luria SE, Human ML. 1952. A nonhereditary, host-induced variation of bacterial viruses. *Journal of Bacteriology* **64**:557–569. DOI: <https://doi.org/10.1128/jb.64.4.557-569.1952>, PMID: 12999684
- Macdonald E, Strahl H, Blower TR, Palmer T, Mariano G. 2022. Shield Co-Opto an RmuC Domain to Mediate Phage Defence across *Pseudomonas* Species. *bioRxiv*. DOI: <https://doi.org/10.1101/2022.11.04.515146>
- Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. 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. DOI: <https://doi.org/10.1186/1745-6150-1-7>, PMID: 16545108
- Makarova KS, Anantharaman V, Aravind L, Koonin EV. 2012. Live virus-free or die: coupling of antiviral immunity and programmed suicide or dormancy in prokaryotes. *Biology Direct* **7**:40. DOI: <https://doi.org/10.1186/1745-6150-7-40>, PMID: 23151069
- Makarova KS, Anantharaman V, Grishin NV, Koonin EV, Aravind L. 2014. CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Frontiers in Genetics* **5**:102. DOI: <https://doi.org/10.3389/fgene.2014.00102>, PMID: 24817877
- Makarova KS, Timinskas A, Wolf YI, Gussow AB, Siksnys V, Venclovas Č, Koonin EV. 2020a. Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antiviral defense. *Nucleic Acids Research* **48**:8828–8847. DOI: <https://doi.org/10.1093/nar/gkaa635>, PMID: 32735657
- Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, Moineau S, Mojica FJM, Scott D, Shah SA, Siksnys V, Terns MP, Venclovas Č, White MF, Yakunin AF, Yan W, et al. 2020b. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nature Reviews. Microbiology* **18**:67–83. DOI: <https://doi.org/10.1038/s41579-019-0299-x>, PMID: 31857715
- Mariani V, Biasini M, Barbato A, Schwede T. 2013. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**:2722–2728. DOI: <https://doi.org/10.1093/bioinformatics/btt473>, PMID: 23986568
- Mayer MP. 2021. The hsp70-chaperone machines in bacteria. *Frontiers in Molecular Biosciences* **8**:694012. DOI: <https://doi.org/10.3389/fmolb.2021.694012>, PMID: 34164436
- McDonnell AV, Jiang T, Keating AE, Berger B. 2006. Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics* **22**:356–358. DOI: <https://doi.org/10.1093/bioinformatics/bti797>
- McMahon SA, Zhu W, Graham S, Rambo R, White MF, Gloster TM. 2020. Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nature Communications* **11**:500. DOI: <https://doi.org/10.1038/s41467-019-14222-x>, PMID: 31980625
- Mendez AS, Vogt C, Bohne J, Glaunsinger BA. 2018. Site specific target binding controls RNA cleavage efficiency by the Kaposi's sarcoma-associated herpesvirus endonuclease SOX. *Nucleic Acids Research* **46**:11968–11979. DOI: <https://doi.org/10.1093/nar/gky932>, PMID: 30321376
- 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. DOI: <https://doi.org/10.1038/s41564-020-0777-y>, PMID: 32839535
- Morse JC, Girodat D, Burnett BJ, Holm M, Altman RB, Sanbonmatsu KY, Wieden HJ, Blanchard SC. 2020. Elongation factor-Tu can repetitively engage aminoacyl-tRNA within the ribosome during the proofreading stage of tRNA selection. *PNAS* **117**:3610–3620. DOI: <https://doi.org/10.1073/pnas.1904469117>, PMID: 32024753
- Neuwald AF, Aravind L, Spouge JL, Koonin EV. 1999. AAA⁺: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Research* **9**:27–43. DOI: <https://doi.org/10.1101/gr.9.1.27>
- Nirwan N, Itoh Y, Singh P, Bandyopadhyay S, Vinothkumar KR, Amunts A, Saikrishnan K. 2019. Structure-based mechanism for activation of the AAA+ GTPase McrB by the endonuclease McrC. *Nature Communications* **10**:3058. DOI: <https://doi.org/10.1038/s41467-019-11084-1>, PMID: 31296862
- Niu Y, Suzuki H, Hosford CJ, Walz T, Chappie JS. 2020. Structural asymmetry governs the assembly and GTPase activity of McrBC restriction complexes. *Nature Communications* **11**:5907. DOI: <https://doi.org/10.1038/s41467-020-19735-4>, PMID: 33219217
- 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. DOI: <https://doi.org/10.1038/s41564-017-0051-0>, PMID: 29085076
- Panne D, Raleigh EA, Bickle TA. 1999. The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis 1 Edited by J. Karn. *Journal of Molecular Biology* **290**:49–60. DOI: <https://doi.org/10.1006/jmbi.1999.2894>

- Park HH**, Lo Y-C, Lin S-C, Wang L, Yang JK, Wu H. 2007. The death domain superfamily in intracellular signaling of apoptosis and inflammation. *Annual Review of Immunology* **25**:561–586. DOI: <https://doi.org/10.1146/annurev.immunol.25.022106.141656>, PMID: 17201679
- Patil DP**, Pickering BF, Jaffrey SR. 2018. Reading m6A in the Transcriptome: m6A-Binding Proteins. *Trends in Cell Biology* **28**:113–127. DOI: <https://doi.org/10.1016/j.tcb.2017.10.001>
- Pedersen K**, Zavialov AV, Pavlov MYu, Elf J, Gerdes K, Ehrenberg M. 2003. The bacterial toxin rele displays codon-specific cleavage of mrnas in the ribosomal a site. *Cell* **112**:131–140. DOI: [https://doi.org/10.1016/S0092-8674\(02\)01248-5](https://doi.org/10.1016/S0092-8674(02)01248-5)
- Pei J**, Kim B-H, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Research* **36**:2295–2300. DOI: <https://doi.org/10.1093/nar/gkn072>, PMID: 18287115
- Pettersen EF**, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**:70–82. DOI: <https://doi.org/10.1002/pro.3943>, PMID: 32881101
- Pieper U**, Schweitzer T, Groll DH, Gast F-U, Pingoud A. 1999. The gtp-binding domain of mcrb: more than just a variation on a common theme? *Journal of Molecular Biology* **292**:547–556. DOI: <https://doi.org/10.1006/jmbi.1999.3103>
- Pillon MC**, Gordon J, Frazier MN, Stanley RE. 2021. HEPN RNases - an emerging class of functionally distinct RNA processing and degradation enzymes. *Critical Reviews in Biochemistry and Molecular Biology* **56**:88–108. DOI: <https://doi.org/10.1080/10409238.2020.1856769>, PMID: 33349060
- Poulsen C**, Panjikar S, Holton SJ, Wilmanns M, Song Y-H. 2014. WXG100 protein superfamily consists of three subfamilies and exhibits an α -helical C-terminal conserved residue pattern. *PLOS ONE* **9**:e89313. DOI: <https://doi.org/10.1371/journal.pone.0089313>, PMID: 24586681
- Poweleit N**, Czudnochowski N, Nakagawa R, Trinidad DD, Murphy KC, Sasseti CM, Rosenberg OS. 2019. The structure of the endogenous ESX-3 secretion system. *eLife* **8**:e52983. DOI: <https://doi.org/10.7554/eLife.52983>, PMID: 31886769
- Price MN**, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**:e9490. DOI: <https://doi.org/10.1371/journal.pone.0009490>, PMID: 20224823
- Raleigh EA**, 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. DOI: <https://doi.org/10.1093/genetics/122.2.279>
- Raleigh EA**. 1992. Organization and function of the mcrBC genes of *Escherichia coli* K-12. *Molecular Microbiology* **6**:1079–1086. DOI: <https://doi.org/10.1111/j.1365-2958.1992.tb01546.x>, PMID: 1316984
- Shigematsu M**, Kawamura T, Kirino Y. 2018. Generation of 2',3'-cyclic phosphate-containing rnas as a hidden layer of the transcriptome. *Frontiers in Genetics* **9**:562. DOI: <https://doi.org/10.3389/fgene.2018.00562>
- 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**:767–769. DOI: <https://doi.org/10.1093/bioinformatics/btu700>
- Söding J**. 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**:951–960. DOI: <https://doi.org/10.1093/bioinformatics/bti125>, PMID: 15531603
- Söding J**, Remmert M, Biegert A, Lupas AN. 2006. HHsenser: exhaustive transitive profile search using HMM-HMM comparison. *Nucleic Acids Research* **34**:W374–W378. DOI: <https://doi.org/10.1093/nar/gkl195>, PMID: 16845029
- 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 oligoampliation. *Molecular Cell* **80**:955–970. DOI: <https://doi.org/10.1016/j.molcel.2020.11.034>
- Stock AM**, Robinson VL, Goudreau PN. 2000. Two-Component Signal Transduction. *Annual Review of Biochemistry* **69**:183–215. DOI: <https://doi.org/10.1146/annurev.biochem.69.1.183>
- 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. DOI: <https://doi.org/10.1093/nar/gks332>, PMID: 22570415
- Sutherland E**, Coe L, Raleigh EA. 1992. McrBC: a multisubunit GTP-dependent restriction endonuclease. *Journal of Molecular Biology* **225**:327–348. DOI: [https://doi.org/10.1016/0022-2836\(92\)90925-A](https://doi.org/10.1016/0022-2836(92)90925-A)
- Swarts DC**, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J. 2014. The evolutionary journey of Argonaute proteins. *Nature Structural & Molecular Biology* **21**:743–753. DOI: <https://doi.org/10.1038/nsmb.2879>, PMID: 25192263
- Tanaka N**, Shuman S. 2011. RtcB is the RNA ligase component of an *Escherichia coli* RNA repair operon. *The Journal of Biological Chemistry* **286**:7727–7731. DOI: <https://doi.org/10.1074/jbc.C111.219022>, PMID: 21224389
- Trigg J**, Gutwin K, Keating AE, Berger B. 2011. Multicoil2: predicting coiled coils and their oligomerization states from sequence in the twilight zone. *PLOS ONE* **6**:e23519. DOI: <https://doi.org/10.1371/journal.pone.0023519>, PMID: 21901122
- Tsutakawa SE**, 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. DOI: [https://doi.org/10.1016/S0092-8674\(00\)81550-0](https://doi.org/10.1016/S0092-8674(00)81550-0)
- van den Berg DF**, van der Steen BA, Costa AR, Brouns SJJ. 2023. Phage tRNAs evade tRNA-targeting host defenses through anticodon loop mutations. *eLife* **12**:e85183. DOI: <https://doi.org/10.7554/eLife.85183>, PMID: 37266569

- van Kempen M**, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. 2024. Fast and accurate protein structure search with Foldseek. *Nature Biotechnology* **42**:243–246. DOI: <https://doi.org/10.1038/s41587-023-01773-0>, PMID: 37156916
- Verma R**, Oania RS, Kolawa NJ, Deshaies RJ. 2013. Cdc48/p97 promotes degradation of aberrant nascent polypeptides bound to the ribosome. *eLife* **2**:e00308. DOI: <https://doi.org/10.7554/eLife.00308>, PMID: 23358411
- Wallden K**, Rivera-Calzada A, Waksman G. 2010. Type IV secretion systems: versatility and diversity in function. *Cellular Microbiology* **12**:1203–1212. DOI: <https://doi.org/10.1111/j.1462-5822.2010.01499.x>, PMID: 20642798
- Wang H-C**, Wu M-L, Ko T-P, Wang AH-J. 2013. Neisseria conserved hypothetical protein DMP12 is a DNA mimic that binds to histone-like HU protein. *Nucleic Acids Research* **41**:5127–5138. DOI: <https://doi.org/10.1093/nar/gkt201>, PMID: 23531546
- Warne B**, Harkins CP, Harris SR, Vatsiou A, Stanley-Wall N, Parkhill J, Peacock SJ, Palmer T, Holden MTG. 2016. The Ess/Type VII secretion system of *Staphylococcus aureus* shows unexpected genetic diversity. *BMC Genomics* **17**:222. DOI: <https://doi.org/10.1186/s12864-016-2426-7>, PMID: 26969225
- Waterhouse AM**, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:1189–1191. DOI: <https://doi.org/10.1093/bioinformatics/btp033>, PMID: 19151095
- Weigle P**, Raleigh EA. 2016. Biosynthesis and function of modified bases in bacteria and their viruses. *Chemical Reviews* **116**:12655–12687. DOI: <https://doi.org/10.1021/acs.chemrev.6b00114>, PMID: 27319741
- Willmund F**, del Alamo M, Pechmann S, Chen T, Albanèse V, Dammer EB, Peng J, Frydman J. 2013. The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis. *Cell* **152**:196–209. DOI: <https://doi.org/10.1016/j.cell.2012.12.001>, PMID: 23332755
- Wolf DH**, Stolz A. 2012. The Cdc48 machine in endoplasmic reticulum associated protein degradation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**:117–124. DOI: <https://doi.org/10.1016/j.bbamcr.2011.09.002>
- Yamamoto S**, Subedi GP, Hanashima S, Satoh T, Otaka M, Wakui H, Sawada K, Yokota S, Yamaguchi Y, Kubota H, Itoh H. 2014. ATPase activity and ATP-dependent conformational change in the co-chaperone HSP70/HSP90-organizing protein (HOP). *The Journal of Biological Chemistry* **289**:9880–9886. DOI: <https://doi.org/10.1074/jbc.M114.553255>, PMID: 24535459
- Zhou C**, Pourmal S, Pavletich NP. 2015. Dna2 nuclease-helicase structure, mechanism and regulation by Rpa. *eLife* **4**:e09832. DOI: <https://doi.org/10.7554/eLife.09832>, PMID: 26491943
- Zhu W**, McQuarrie S, Gruschow S, McMahon SA, Graham S, Gloster TM, White MF. 2021. The CRISPR ancillary effector Can2 is a dual-specificity nuclease potentiating type III CRISPR defence. *Nucleic Acids Research* **49**:2777–2789. DOI: <https://doi.org/10.1093/nar/gkab073>, PMID: 33590098
- 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**:1628–1637 PMID: 11720291.
- Zimmermann L**, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented mpi bioinformatics toolkit with a new hhpred server at its core. *Journal of Molecular Biology* **430**:2237–2243. DOI: <https://doi.org/10.1016/j.jmb.2017.12.007>