Cryptic Genes for Interbacterial Antagonism
distinguish *Rickettsia* species Infecting Blacklegged
Tick from other *Rickettsia* pathogens

Victoria I. Verhoeve ¹, Tyesha D. Fauntleroy ¹, Riley G. Ristine ¹, Timothy P. Driscoll ², Joseph J. Gillespie ¹

¹ Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA
² Department of Biology, West Virginia University, Morgantown, West Virginia, USA

Corresponding author:
Joe Gillespie, Department of Microbiology and Immunology
University of Maryland School of Medicine
685 W. Baltimore St., HSF I Suite 380 Baltimore, MD 21201
Jgillespie@som.umaryland.edu

Running head: A novel *Rickettsia buchneri* toxin-antidote module
Abstract

Background. Genus Rickettsia (Alphaproteobacteria: Rickettsiales) encompasses numerous obligate intracellular species with predominantly ciliate and arthropod hosts. Notable species are pathogens transmitted to mammals by blood-feeding arthropods. Mammalian pathogenicity evolved from basal, non-pathogenic host-associations; however, some non-pathogens are closely related to pathogens. One such species, Rickettsia buchneri, is prevalent in the blacklegged tick, Ixodes scapularis. While I. scapularis transmits several pathogens to humans, it does not transmit Rickettsia pathogens. We hypothesize that R. buchneri established a mutualism with I. scapularis, blocking tick superinfection with Rickettsia pathogens.

Methods. To improve estimates for assessing R. buchneri infection frequency in blacklegged tick populations, we used comparative genomics to identify an R. buchneri gene (REIS_1424) not present in other Rickettsia species present throughout the I. scapularis geographic range. Bioinformatic and phylogenomics approaches were employed to propose a function for the hypothetical protein (263 aa) encoded by REIS_1424.

Results. REIS_1424 has few analogs in other Rickettsiales genomes and greatest similarity to non-Proteobacteria proteins. This cohort of proteins varies greatly in size and domain composition, possessing characteristics of Recombination hotspot (Rhs) and contact dependent growth inhibition (CDI) toxins, with similarity limited to proximal C-termini (~145 aa). This domain was named CDI-like/Rhs-like CT toxin (CRCT). As such proteins are often found as toxin-antidote (TA) modules, we interrogated REIS_1423 (151 aa) as a putative antidote. Indeed, REIS_1423 is similar to proteins...
A novel *Rickettsia buchneri* toxin-antidote module

encoded upstream of CRCT domain-containing proteins. Accordingly, we named these proteins CDI-like/Rhs-like C-terminal toxin antidotes (CRCA). *R. buchneri* expressed both *REIS_1423* and *REIS_1424* in tick cell culture, and PCR assays showed specificity for *R. buchneri* over other rickettsiae and utility for positive detection in three tick populations. Finally, phylogenomics analyses uncovered divergent CRCT/CRCA modules in varying states of conservation; however, only *R. buchneri* and related Tamurae/Ixodes Group rickettsiae carry complete TA modules.

**Conclusion.** We hypothesize that *Rickettsia* CRCT/CRCA modules circulate in the *Rickettsia* mobile gene pool, arming rickettsiae for battle over arthropod colonization. While its functional significance remains to be tested, *R. buchneri* CRCT/CRCA serves as a marker to positively identify infection and begin deciphering the role this endosymbiont plays in the biology of the blacklegged tick.

**Key words:** *Rickettsia buchneri, Ixodes scapularis*, blacklegged tick, Lyme Disease, toxin-antidote, contact-dependent growth inhibition systems, recombination hot spot, lateral gene transfer
Introduction

The blacklegged tick (*Ixodes scapularis*), more commonly referred to as deer tick, is of vital importance to human health as a vector of several infectious disease agents: e.g., *Borrelia* species (Lyme disease), *Anaplasma phagocytophilum* (anaplasmosis), *Babesia* and *Theileria* parasites (babesiosis, theileriosis), and Powassan Flavivirus (Powassan disease) (Wormser et al., 2006; Madison-Antenucci et al., 2020). Curiously, blacklegged tick does not infect humans with *Rickettsia* pathogens, despite overlapping in geographic range with other tick species that do; e.g., American dog tick (*Dermacentor variabilis*), Brown dog tick (*Rhipicephalus sanguineus*), Gulf Coast tick (*Amblyomma maculatum*), and Lone Star tick (*Amblyomma americanum*) (Walker and Ismail, 2008; Lee et al., 2014; Sanchez-Vicente et al., 2019). However, *I. scapularis* is predominantly infected with a species of *Rickettsia*, *R. buchneri*, that is considered a non-pathogen of humans and has not been detected in vertebrates (Kurtti et al., 2015). The presence of *R. buchneri* in tick ovaries (Munderloh et al., 2005), high infection rate in ticks regardless of co-infection with other intracellular bacteria (Billings et al.; Magnarelli et al., 1991; Benson et al., 2004; Swanson and Norris, 2007; Troughton and Levin, 2007; Steiner et al., 2008) or composition of other microbiota (Moreno et al., 2006; Narasimhan et al., 2014; van Treuren et al., 2015; Abraham et al., 2017; Ross et al., 2018; Thapa et al., 2019; Tokarz et al., 2019) hint at an underappreciated host-microbe relationship in need of further investigation.

A decade ago, we reported the first genome sequence of *R. buchneri* by assembling bacterial-like sequencing reads generated by the *I. scapularis* genome project (Gillespie et al., 2012a; Gulia-Nuss et al., 2016). Prior to this, phylogenomics analyses of diverse
A novel *Rickettsia buchneri* toxin-antidote module

*Rickettsia* genomes indicated high conservation in synteny, moderate pseudogenization, one or zero plasmids, and relatively few mobile genetic elements (MGEs) (Darby et al., 2007; Fuxelius et al., 2008; Gillespie et al., 2008, 2012b). Several attributes emerged from our analyses highlighting the oddity of the *R. buchneri* genome, including 1) pronounced pseudogenization relative to other rickettsiae, especially for genes in other

**Spotted Fever Group (SFG)** rickettsiae with characterized functions in vertebrate pathogenesis, 2) a substantial number of transposases (~30% of total CDS), 3) four novel plasmids (pREIS1-4), and 4) nine copies (seven chromosomal and two plasmid) of the

**Rickettsiales Amplified Genetic Element (RAGE)**, a conjugative transposon found as single-copy in certain other *Rickettsia* genomes (Ogata et al., 2006; Blanc et al., 2007). Estimated phylogenies placed *R. buchneri* basal to all SFG rickettsiae, indicating a substantially different evolutionary track relative to derived SFG rickettsiae lineages, as well as species in the **Typhus Group (TG)** and **Transitional Group (TRG)** rickettsiae, yet inordinate lateral gene transfer (LGT) with the ancestral *R. bellii* and other intracellular bacteria. This was exemplified by *R. buchneri*’s RAGEs, which encode numerous genes with functions critical for obligate intracellular life, leading to our hypothesis that RAGEs and other MGEs are vehicles for gene acquisitions that offset high rates of pseudogenization (Gillespie et al., 2012a) (Fig. 1A).

The *R. buchneri* genome was also found to encode many genes absent from other rickettsiae. Noteworthy are those encoding aminoglycoside antibiotic (AGAB) synthesis/resistance genes and polyketide synthase (PKS)-containing nonribosomal protein synthases (NRPS), hinting at a defense arsenal of antibiotics and 2° metabolites. Furthermore, plasmid pREIS2 of *R. buchneri* carries identical duplications of the biotin...
synthesis operon of obligate microbes (BOOM), (Fig. 1B), which is only found in a limited range of obligate intracellular species, including certain wolbachiae (Nikoh et al., 2014; Gerth and Bleidorn, 2017; Balvín et al., 2018; Driscoll et al., 2020), Cardinium (Penz et al., 2012; Zeng et al., 2018) and Legionella (Ríhová et al., 2017) species. As some wolbachiae provide biotin to their insect hosts (Nikoh et al., 2014; Ju et al., 2019), *R. buchneri* may supplement biotin to the blacklegged tick bloodmeal, which is a poor source of B vitamins (Lehane and Lehan, 2010; Manzano-Marín et al., 2015). This possibility, coupled with potential to provision blacklegged tick with antibiotics and 2°C metabolites, indicates *R. buchneri* has characteristics of a mutualistic bacterium unlike any rickettsiae heretofore analyzed from a genomic perspective.

If *R. buchneri* is a mutualist of blacklegged tick, infection frequency should be very high in nature. Reported *R. buchneri* infection rates in *I. scapularis* populations range from under 20% to over 80%; however, this disparity may reflect different sampling strategies not aiming to directly detect *R. buchneri* or distinguish this species from other rickettsiae (i.e., using general *Rickettsia* PCR primers). Differences in tissue sampling across studies could also distort accuracy in detecting *R. buchneri* infection if ovaries are not sampled. Furthermore, infection rates are higher in females and nymphs (Cross et al., 2018; Hagen et al., 2018; Gil et al., 2020), indicating a reduction in males after molting to adulthood; thus, sexing is important to understand true infection frequency.

In this report, we developed a *R. buchneri*-specific PCR assay to help rule out *Rickettsia* co-infections in future studies. Our strategy was guided by re-evaluating a set of genes previously determined to be unique to *R. buchneri* (Gillespie et al., 2012a). One gene, *REIS-1424* (encoding a 263 aa hypothetical protein), was shown by *in silico*
A novel *Rickettsia buchneri* toxin-antidote module

analysis to be absent in other rickettsiae known throughout the geographic range of *I. scapularis*. Unexpectedly, bioinformatic and phylogenomics analyses indicated that *REIS_1424* and its neighboring gene *REIS_1423* comprise a toxin-antidote (TA) module typical of certain bacterial systems used primarily for interbacterial warfare. As with numerous other *R. buchneri*-specific genes, *REIS_1424* and *REIS_1423* show evidence of LGT from distantly related non-proteobacteria. Similar cryptic TA modules are recurrent in rickettsiae and illuminate a potential mechanism for *Rickettsia* competition in arthropod hosts.

**Materials and Methods**

**Identifying a *R. buchneri*-specific gene**

A set of genes (739 singletons) from our prior report lacked significant similarity to genes in other *Rickettsia* genomes (Gillespie et al., 2012a); accordingly, these were used as queries in Blastn searches against the NCBI ‘Rickettsia’ databases (taxid 780). Genes with significant similarity to genes from rickettsiae present within the geographic distribution of *I. scapularis* were removed. Further, we excluded genes encoded on plasmids, those present within chromosomal RAGES, and those encoding transposases or related elements. The remaining genes were then evaluated using bioinformatics analysis to determine the likelihood that they encode functional proteins.

**Compiling toxin and antidote datasets**

For toxins, proximal (~100 aa) C-terminal sequences of *R. buchneri* protein *REIS_1424* (EER22217) and another rickettsial species “*Candidatus Jidaibacter acanthamoeba*” str.
A novel *Rickettsia buchneri* toxin-antidote module

UWC36 protein NF27_IC00050 (KIE04387) were used as queries in Blastp searches to compile and analyze diverse proteins harboring significant similarity across complementary proximal C-terminal sequences. For antidotes, entire sequences for *R. buchneri* protein REIS_1423 (EER22217 with an adjusted start site adding 41 aa at the N-terminus) and “*Candidatus Jidaibacter acanthamoeba*” str. UWC36 protein NF27_IC00040 (KIE04386) were used as queries in Blastp searches to compile and analyze diverse proteins harboring significant similarity across the entire lengths of the queries. Analyses utilized our HaloBlast method, which is a combinatorial Blastp-based approach for interrogating proteins for LGT (Driscoll et al., 2013). Individual Blastp searches were conducted against five distinct taxonomic databases: 1) “*Rickettsia*” (NCBI taxid 780”), 2) “*Rickettsiales*” (taxid: 766) excluding “*Rickettsia*”, 3) “*Alphaproteobacteria*” (taxid: 28211) excluding “*Rickettsiales*”, 4) “*Proteobacteria*” (taxid: 1224) excluding “*Alphaproteobacteria*”, 5) “*Bacteria*” (taxid: 2) excluding *Proteobacteria*, and 6) “minus bacteria”). All subjects from each search were ranked by *Sm* score (= *b* *I* *Q*, where *b* is the bitscore of the match, *I* is the percent identity, and *Q* is the percent length of the query that aligned), a comparative sequence similarity score designed to de-emphasize highly significant matches to short stretches of the query in favor of longer stretches of similarity (Driscoll et al., 2013). The “halos” or separate database searches were then compared to one another to determine the taxon with the strongest similarity to the query sequences.

**Toxin characterization**
A novel *Rickettsia buchneri* toxin-antidote module

HaloBlast subjects from the searches with REIS_1424 and NF27_IC00050 as queries were analyzed in two ways. First, only sequences matching the proximal (~100 aa) C-terminal sequences of the query were compiled and aligned with MUSCLE using default parameters (Edgar, 2004). The entire alignment was then visualized as sequence logos using WebLogo (Crooks et al., 2004). Second, two representative sequences per halo were selected for domain predictions across the entire protein. EMBL’s Simple Modular Architecture Research Tool (SMART) (Letunic and Bork, 2017) and/or the Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) (Kelley and Sternberg, 2009) were used to predict and evaluate the following domains: UBA (ubiquitin-associated) (Mueller and Feigon, 2002); haemagglutination activity site (Kajava et al., 2001); hemagglutinin repeats (Pfam ID PF13332); Peptidase M43 domain (Rawlings and Barrett, 1995); endonuclease III (Bruner et al., 2000); RHS repeat (Busby et al., 2013); VENN motif (Aoki et al., 2010; Zhang et al., 2011); DUF637: hemagglutinin-/hemolysin-associated domain (PF04830); alanine-rich-conserved phenylalanine (ALF) motif (Yeats et al., 2003); Laminin_G_3 (PF13385); LamG-like jellyroll fold domain (Liu et al., 2007; Weyer et al., 2007); HintN domain (Perler, 1998). Individual protein schemas were generated using Illustrator of Biological Sequences (Liu et al., 2015) with manual adjustment.

**Antidote characterization**

HaloBlast subjects from the searches with REIS_1423 and NF27_IC00040 as queries were compiled and aligned with MUSCLE (default parameters), with the entire alignment visualized as sequence logos using WebLogo. Additionally, CDS flanking
A novel *Rickettsia buchneri* toxin-antidote module
certain HaloBlast subjects (i.e. those with NCBI reference protein accession numbers)
from the searches with REIS_1424 and NF27_IC00050 were evaluated for their
predicted size and presence of a conserved N-terminal sequence motif
(LS/ADXE/DXQXXXW) determined to be highly conserved in subjects retrieved in
Blastp searches with REIS_1423 and NF27_IC00040 as queries. Finally, HMMER (Finn et al., 2011) searches using NF27_IC00040 or NF27_IC00050 were utilized to evaluate
our Blastp-based identification and compilation of both toxin and antidote datasets.

**Phylogeny estimation**

*Antidote phylogeny*. Selected antidotes were aligned using MUSCLE (default
parameters). A phylogeny was estimated with the WAG substitution model (gamma
model of rate heterogeneity) using RAxML v8.2.4 (50). Branch support was assessed
with 1,000 pseudo-replications.

*Rickettsia phylogeny*. Protein sequences (*n* = 121,310) from 92 sequenced genomes
were used to estimate a genus-level *Rickettsia* phylogeny. *Rickettsia* genomes were
retrieved from the NCBI Assembly database (*n* = 92). The Rapid Annotation using
Subsystem Technology (RAST) v 2.0 server (Aziz et al., 2008) was used to annotate
three *Rickettsia* assemblies that were not previously annotated. A total of 3,707
orthologous gene families were constructed from this data set using *fastortho*, a
modified version of OrthoMCL (Feris et al., 2003), at an inflation of 1.5 and a percent
identity threshold of 40%. A subset of 263 single-copy families conserved across all 92
taxa was independently aligned with MUSCLE (Edgar, 2004) using default parameters,
and regions of poor alignment were masked using Gblocks (Talavera and Castresana,
A novel *Rickettsia buchneri* toxin-antidote module

2007). All modified alignments were concatenated into a single data set (74,799 positions) for phylogeny estimation using RAxML v8.2.4 (Stamatakis et al., 2005), using a gamma model of rate heterogeneity and estimation of the proportion of invariable sites. Branch support was assessed with 1,000 pseudo-replications.

**DNA and RNA extraction, PCR**

For *R. buchneri* analysis, ticks from New York, New Hampshire, and Pennsylvania (see Fig. 6 for locality information) were stored in 100% ethanol at -20°C until isolation. DNA was extracted from *I. scapularis* adults and nymphs and rickettsiae infecting ISE6 cells and using the DNeasy kit (Qiagen) as per manufacturer's protocols for cell culture and tissue extraction, respectively. Briefly, ticks were surface sterilized with 5 min washes (1% bleach, 70% ethanol, and 1xPBS), cut into quarters with a sterile scalpel blade, incubated with kit-provided digestion buffer with proteinase K at 56°C overnight, and extracted using the tissue protocol with a final elution of 50µl of molecular grade water. Rickettsiae grown in culture were collected in their host cells, and DNA extracted using the cell culture protocol with a 50µl elution with molecular grade water. For analysis of non-target rickettsiae, all bacteria were grown in cell culture prior to DNA extraction and PCR analysis. DNA was qPCR amplified using PowerUp Sybr Mastermix (Thermo) in 20µl reactions containing 400nm of each primer and 50-100ng of DNA. Primers pairs are as follows for *R. buchneri*-specific targets: Rb-1424-120-F-5'-acaggcgtaaactagacatct-3' with Rb-1424-120-R-5'-aggaaatccagcctttcaggta-3' for the amplification of *rCRC* and Rb-1423-116-F-5'-gcataaggtttatatagcgggtgc-3' with Rb-1423-116-R-5'-ccataaggtttctctattgtgctt-3' for the amplification of *rCRCA*. *Rickettsia gltA* was
A novel *Rickettsia buchneri* toxin-antidote module

amplified for all *rickettsia* using the following primers: CSRT-F-5'-
tcgcaaatgcgcggtacttt-3' and CSRT-R-5'-tcgtgcaattctttccattg-3' (Stenos et al., 2005).

Reactions were amplified under the following conditions: 1 cycle for 2 min at 95°C, 45 cycles at 95°C for 15 sec and 60°C for 30 sec, followed by a melt curve analysis. All primer sets were considered positive if the cycle threshold was 37 cycles or less. All primer sets were validated for range and efficiency of amplification using pCR4-TOPO plasmid standard curves with ligated amplicons. Primer sets described in this manuscript only amplify their intended products as verified by sanger sequencing and melt curve analyses of each reaction. For visualization of qPCR products (Fig. 6) gel electrophoresis was performed using a 2% agarose gel with ethidium bromide straining and visualization using a gel imaging station. For transcriptional analysis, *R. buchneri* growing during log growth in ISE6 cells were collected in 600μl TRIzol (Invitrogen) and RNA extracted using the DirectZol (Zymo) kit using manufacturer’s instructions and on-column DNase treatment. RNA was further DNase treated using the RQ1 DNase (Promega) prior to enzyme removal and concentration with the Zymo Clean and Concentrator-5 kit. The iScript Select cDNA synthesis kit (Bio-Rad) was used for cDNA synthesis using random hexamers in 20μl reactions with 200ng of DNase-free RNA. RNA was determined to be free of DNA with no reverse transcriptase reactions with 200ng of RNA which resulted in no detectable DNA by qPCR. cDNA was analyzed for transcription of rCRCT, rCRC and *I. scapularis* β-actin (primers IsActin-95-F-5'-
aatcggcaacgagaggttc-3' and IsActin-95-R-5'-agttgtacggtgtcgcggtc-3') using the same qPCR parameters as above.
Results and Discussion

Identifying an *R. buchneri*-specific gene

Our prior analysis of the first sequenced *R. buchneri* genome (*Wikel I. scapularis* colony) indicated 32% of CDS were absent from other *Rickettsia* genomes (Gillespie et al., 2012a). Given that dozens of new *Rickettsia* genomes have been sequenced and assembled since 2012, we revisited this list and allowed for candidate genes to also be present in the *R. buchneri* str. ISO7 assembly (Kurtti et al., 2015). Further, to increase the likelihood of a stable gene present in all *R. buchneri* populations, we excluded genes 1) encoded on plasmids, 2) flanked by transposases, 3) containing annotations reflecting an association with MGEs, and 4) containing Blastp profiles indicating pseudogenization (e.g., gene fragments, split genes, etc.). These collective constraints yielded a small list of candidate genes, one of which (*REIS_1424*, NCBI accession no. EER22217) was selected for further analysis.

*REIS_1424* encodes a hypothetical protein of 263 aa; the homolog in the *R. buchneri* str. ISO7 assembly (KDO03356) is identical yet has a different predicted start site that adds 35 aa to the N-terminus. The only remaining significant Blastp matches are from *R. tamurae* str. AT-1, with two CDS spanning the entirety *REIS_1424* indicating a split gene (WP_215426163 and WP_032138795). This is consistent with the close phylogenetic position of *R. tamurae* and *R. buchneri* (Hagen et al., 2018); however, since *R. tamurae* has not been reported from the Western hemisphere, these CDS are not a concern for utilization of *REIS_1424* as a diagnostic for *R. buchneri* infection.
REIS_1424 carries a cryptic toxin domain

REIS_1424-based Blastp searches outside of the Rickettsia taxon database yielded only two significant matches: a 2192 aa protein (hypothetical protein NF27_IC00050, KIE04387) from a rickettsial amoeba-associated endosymbiont, "Candidatus Jidaibacter acanthamoeba" str. UWC36 (Rickettsiales: Midichloriaceae) (Schulz et al., 2016), and a 97 aa protein (hypothetical protein E1266_17330, TDB94289) from the actinobacterium Actinomadura sp. 7K534 (Streptosporangiales; Thermomonosporaceae). Both alignments indicated that 1) REIS_1424, NF27_IC00050, and E1266_17330 share over a dozen conserved residues, 2) the REIS_1424 and NF27_IC00050 match aligns the proximal C-terminal sequences of both proteins, and 3) E1266_17330 is truncated and lacks N-terminal sequence outside of the matches. Unlike REIS_1424, NF27_IC00050- or E1266_17330-based Blastp searches yielded many significant matches to diverse bacteria (discussed further below). However, no functional domains for the region shared across these proteins could be predicted with searches against the NCBI Conserved Domains Database or using SMART.

Given that the REIS_1424-NF27_IC00050 match spanned greater sequence in each protein (~137 aa) and "Cand. J. acanthamoeba" is another rickettsial taxon, NF27_IC00050 (aa residues 2068-2192) was used as a proxy for further Blastp searches and in silico characterization. NF27_IC00050-based HaloBlast analysis revealed strongest similarity to certain non-proteobacterial proteins (Fig. 2A). All obtained sequences matched NF27_IC000502068-2192 at their proximal C-termini (Fig. 2B). Intriguingly, this cohort of proteins (n = 155) varied greatly in size across regions outside of the conserved C-terminal sequence (Fig. 2C). A wide assortment of domains...
A novel *Rickettsia buchneri* toxin-antidote module was predicted for these proteins, with many having modular architectures and other characteristics of **contact dependent growth inhibition (CDI)** and/or **Recombination hotspot (Rhs)** toxins (**Fig. 2D**). While no functional domain could be predicted for any of the analogous C-terminal regions, sufficient conservation was found to strongly indicate a unifying functional role (**Fig. 2E, Fig. S1A**). We hereafter refer to these analogous regions as **CDI-like/Rhs-like CT toxin (CRCT)** domains, and to larger proteins possessing them as **CRCT domain-containing proteins (CRCT-DCP)**.

The fusion of small, **toxin-antidote (TA)** pairs to the C-termini of CDI and Rhs toxins has previously been described and is thought to expand the diversity of toxic activities deployed by both CDI and Rhs systems (Aoki et al., 2010; Poole et al., 2011; Zhang et al., 2011; Ruhe et al., 2018). The extreme polymorphic nature of these TA modules indicates bacterial arms races, with selection operating on species- and strain-level recognition that shapes communities (Willett et al., 2015). For instance, many of the CRCT domain-containing proteins we identified in diverse bacteria lack the CRCT domain in closely related strains (data not shown). Furthermore, we found many small proteins containing only the CRCT domain (**Fig. 2C, Table S1**), as well as larger proteins carrying truncated CRCT domains (data not shown), suggesting the mobile nature of these toxins and a rapid “birth and death” process. These observations, combined with small size and limited sequence conservation, collectively challenge computational approaches for identifying these polymorphic toxins. This is evinced by HaloBlast profiles for REIS_1424 that mirror those for NF27_IC00050\_2068-2192 once a *Rickettsia*-specific insertion is removed from the query in Blastp and HMMR searches (**Fig. 2E, Fig. S1B**).
A novel *Rickettsia buchneri* toxin-antidote module

**REIS_1423 is a cryptic immunity antidote to REIS_1424**

As CRCTs are often found as TA modules, we interrogated genes up- and downstream of *REIS_1424, NF27_IC00050*, and the genes encoding the 153 other identified CRCT-DCPs. This revealed probable antidotes, hereafter named **CDI-like/Rhs-like C-terminal toxin antidotes** (*CRCA*), adjacent to 37% of the 155 CRCT-DCPs (*Fig. 3A*).

NF27_IC00040-based HaloBlast analysis mirrored that for NF27蒋000502068-2192, revealing strongest similarity to non-proteobacterial proteins (*Fig. 3B*). Taxonomic breakdown of these non-proteobacterial proteins for both CRCT-DCPs and CRCAs revealed a majority from Actinomycetia and Cyanobacteria genomes (*Fig. 3C*). While all CRCAs are strongly constrained in length (~140 aa), with only a few proteins fused to partial genes with unrelated functions (*Fig. 3D*), Blastp profiles for the best scoring matches parsed by taxonomy strongly indicate **Rickettsia CRCT/CRCA** (*rCRCT/CRCA*) modules were acquired from distant non-proteobacteria via LGT (*Fig. 3E*).

Like analyzed CRCTs, predicted CRCAs possess enough conservation to indicate a common function (*Fig. 3F, Fig. S2*). Still, efforts to thread either CRCTs or CRIAs to solved structures of CdiA-CT/CdiI toxin/immunity complexes (Morse et al., 2012; Beck et al., 2014; Johnson et al., 2016; Michalska et al., 2018) were futile. Many of the characterized CdiA-CT toxins are from proteobacterial species and function as Rnases, specifically targeting tRNAs or rRNAs (Willett et al., 2015). Using SMART or searching against the NCBI CDD did not indicate CRCTs harbor nuclease activities, and no similarity of CRCAs to CdiI domains (e.g., cd20694: CdiI_Ct-like) or members of the
SUHK superfamily of immunity proteins (Zhang et al., 2011) could be made. Cdi antidotes have been hypothesized to drive CdiA-CT/Cdi module diversification since they evolve faster than CdiA-CT toxins (Nikolakakis et al., 2012). While we observed greater conservation in CRCTs (Fig. 2E) versus CRCAs (Fig. 3F), HaloBlast and HMMER (data not shown) searches recovered more CRCA (Fig. 3B) versus CRCT-DCPs (Fig. 2A). Still, the presence of CRCT/CRCA modules across diverse bacterial phyla, with some drastic differences in lifestyle (i.e. obligate intracellular versus extracellular, eukaryote-dependent versus environmental, etc.) indicates a common universal cellular target of CRCTs such as membrane, DNA, or RNA previously characterized for all other studied Cdi-A-CT/Cdi modules (Aoki et al., 2009, 2010).

A rickettsial CRCT/CRCA module mobilized to a eukaryote?

We identified a eukaryotic genome harboring a possible LGT of a rickettsial CRCT-DCP/CRCA system. The genome of the smooth cauliflower coral (Stylophora pistillata) contains a gene encoding a large Rhs-like toxin (YbeQ) that was assembled on an unincorporated scaffold with other bacterial-like genes (Fig. 4). YebQ has highest similarity to a few smaller rickettsial proteins that are remnants of degraded Rhs-like toxins (see next section), yet consistent similarity to larger toxins from non-proteobacterial genomes (Table S3). While YebQ does not contain a CRCT domain, smaller S. pistillata genes were found encoding the CRCT and CRCA domains with higher similarity to rickettsial equivalents than other bacteria (Table S3). This attests to the mobile nature of CRCT/CRCA modules and their tendency to incorporate into larger bacterial toxins. It also resonates on our prior work showing another aquatic animal, the
placozoan *Trichoplax adhaerens*, contains LGTs from bacteria (particularly rickettsiae) (Driscoll et al., 2013). While the presence of introns in many of these bacterial-like *S. pistillata* gene models supports integration, mis-assembly of reads from a rickettsial endosymbiont is also possible.

**Diverse CRIA/CRCT modules are recurrent in *Rickettsia* genomes**

Inspection of the genomic region where rCRCT/CRCA modules have inserted revealed two interesting findings, both of which further attest to the mobile nature of these polymorphic TA modules and their rapid birth and death process. First, the rCRCT/CRCA loci of *R. buchneri* and *R. tamurae* occur in a recombination hotspot adjacent to the SecA gene ([Fig. 5A](#)). This region is highly variable across *Rickettsia* genomes ([Fig. S3A](#)) and contains small CDS with matches to the *S. pistillata* YebQ Rhs toxin described above ([Fig. S4A](#)). Despite extraordinary variability in the number and size of CDS in this region across *Rickettsia* genomes, a conserved tRNA-Ala^TGC^ locus is always present, corroborating prior observations for CidA-CT/CdiAI modules often inserting near tRNA genes in bacterial genomes.

Second, Blastp searches with a concatenated query (REIS\_1427-REIS\_1423) revealed matches to additional rCRCA proteins, indicating other rCRCT/CRCA modules elsewhere in *Rickettsia* genomes ([Fig. 5A](#), dashed box). We designated these divergent rCRCA proteins as components of “rCRCT/CRCA-2” modules, with the above-described system named “rCRCT/CRCA-1” modules. Like the rCRCA-1 protein of *R. tamurae*, most rCRCA-2 genes are truncated or fragmented, yet a complete protein was found for the *Rickettsia* endosymbiont of *Ixodes pacificus* (hereafter REIP) ([Fig. S4B](#)). Despite
A novel *Rickettsia buchneri* toxin-antidote module

limited similarity (~30 %ID), an estimated phylogeny grouped rCRCA-1 and rCRCA-2 proteins together to the exclusion of other CRCA proteins (Fig. 5B; Fig. S4B). rCRCA-2 loci all mapped to a second recombination hotspot in *Rickettsia* genomes adjacent to the conserved BamA and tRNA-Thr<sup>CGT</sup> genes (Fig. 5C; Fig. S3B). As with rCRCA-1, we could not identify similarity between rCRCA-2 proteins and domains of CdiI proteins or SUKH immunity proteins.

Searching upstream of rCRCA-2 for potential cognate toxins instead yielded additional genes encoding CRCA antidotes that are duplicated and highly divergent from rCRCA-1 and rCRCA-2 proteins (Fig. 5C). This arrangement of arrayed immunity antidotes is more characteristic of cdiA-CT/cdiI loci in many proteobacterial genomes (Aoki et al., 2010). Indeed, we were able to identify CdiI-like domains in these proteins using the NCBI Conserved Domains Database (cd20694). Accordingly, we named these antidotes rCRCA-3 proteins. Further inspection of rCRCA-3 genes identified eight genomes with additional copies found in a third recombination hotspot between the genes encoding the cytochrome oxidase subunits 2 (CyoA) and 1 (CyoB) (Fig. 5C; Fig. S3C). Seven *Rickettsia* genomes have rCRCA-3 genes in both the BamA and CyoA/B recombination hotspots, indicating recent recombination between these loci (Fig. S4C).

We designated rCRCA-3a and rCRCA-3b proteins to distinguish between those located at the BamA or CyoA/B recombination hotspots, respectively.

At the BamA recombination hotspot, we identified cognate rCRCT-3a toxins with CdiA-CT-like domains (cd20695) in the *R. tamurae* and REIP genomes (Fig. 5C). These two toxins have strongest similarity to counterparts in proteobacterial genomes, particularly *Pseudomonas* and *Moraxella* species (Fig. S4D,E). We modeled the *R.*
A novel Rickettsia buchneri toxin-antidote module

A novel Rickettsia buchneri toxin-antidote module

tamurae rCRCT-3a toxin to the CdiA-CT structure of Cupriavidus taiwanensis (Kryshtafovych et al., 2018) with high confidence, indicating rCRCT-3 toxins are unrelated to rCRCT-1 toxins (Fig. S4F,G). The rCRCA-3 antidotes could not be modeled to the CdiI structure of C. taiwanensis or any other CdiI structures, making the association of rCRCT-3 with rCRCA-3 supported by genome proximity alone. Collectively, this analysis of rCRCT/CRCA genes in Rickettsia genomes illuminates recurrent genome integration, possibly in larger Rhs toxins that have degraded over time. The presence of complete, yet divergent rCRCT/CRCA modules in different species of the Tamurae/Ixodes Group (TIG) rickettsiae (Fig. 5D,E) indicates weaponry for interbacterial antagonism may be functional for these species and implicates a previously unrealized mechanism for rickettsial competition in the same arthropod host.

Probing rCRCT/CRIA-1 for a role in R. buchneri biology

Consideration for another factor behind a putative mutualism

Factors that that distinguish parasitic rickettsiae from species exhibiting other host associations are sorely needed for Rickettsiology. Previously, we searched for genes underlying potential mutualism within the intriguing R. felis system, wherein typical strains infect blood-feeding arthropods (mostly the cat flea, Ctenocephalides felis) yet another has developed a tight host association with a non-blood-feeding insect (the booklouse Liposcelis bostrychophila) (Gillespie et al., 2015a). Only the L. bostrychophila-infecting strain harbored the unique plasmid, pLbAR, which we postulated encoded factors inducing parthenogenesis in booklice since sexually reproducing populations are only observed in the absence of R. felis (Yang et al., 2015).
A novel *Rickettsia buchneri* toxin-antidote module

A TA module on pLbAR was found to have similarity to gene pairs in *Wolbachia* reproductive parasites that were later characterized as the factors underpinning cytoplasmic incompatibility (or male sterilization) (Beckmann et al., 2017; LePage et al., 2017). We later reported on the widespread occurrence in intracellular species of highly diverse TA modules with rudimentary similarity to the *Wolbachia* and *R. felis* TA modules (Gillespie et al., 2018). Despite its high frequency of infection in *I. scapularis* populations, we found no evidence for these TA modules in *R. buchneri*, although other different factors inducing reproductive parasitism could still be present.

The presence of genes in *R. buchneri* that encode AGAB synthesis/resistance proteins and PKS-containing NRPS modules hint at arsenals of antibiotics and 2° metabolites possibly used for defense against certain microbes also infecting blacklegged tick. Furthermore, two copies of BOOM suggest this species supplements the blacklegged tick diet with biotin. Absence (AGAB synthesis/resistance and PKS-containing NRPS) (Hagen et al., 2018) or scarce (BOOM) (Driscoll et al., 2020) distribution of these genes in other *Rickettsia* genomes suggests they are at least utilized for functions generally not employed by other rickettsiae. For instance, a similar type I PKS of a honey bee endosymbiont has recently been shown to suppress growth of fungal pathogens and protect bee brood from infection (Miller et al., 2020). We previously showed that, unlike the conserved AGAB synthesis/resistance genes, the PKS NRPS module is variable in gene content across *R. buchneri* strains from different populations (Hagen et al., 2018). More recently, the PKS NRPS module and the AGAB synthesis/resistance gene array were investigated for their possible roles in limiting superinfection of pathogenic rickettsia in tick cells infected with *R. buchneri* str. ISO7.
A novel *Rickettsia buchneri* toxin-antidote module

(Cull et al., 2022). Despite demonstrating that *R. buchneri* substantially reduced superinfection by pathogenic *R. parkeri* in cell culture, no anti-bacterial activity against either *R. parkeri* or extracellular bacteria (*Escherichia coli* and *Staphylococcus aureus*) was shown by the unknown product(s) of these loci, leaving their function in *R. buchneri* unclear.

**rCRCT/CRCA modules are tailored for interspecific antagonism**

The sequence profiles that strongly indicate rCRCT/CRCA modules were acquired from distant non-proteobacteria ([Fig. 2A; Fig. 3A-C,E](#)) are similar to those we previously reported for AGAB synthesis/resistance proteins (Gillespie et al., 2012a) and PKS-containing NRPS modules (Hagen et al., 2018). This implicates LGT in shaping these factors that likely underpin interbacterial antagonism. While the target of the product of the PKS-containing NRPS module is hard to predict, it is likely that the aminoglycoside synthesized by the AGAB gene array does not target rickettsiae, as this class of antibiotics is ineffective against *Rickettsia* species (Rolain et al., 1998) and overall generally poor for destroying intracellular bacteria (Maurin and Raoult, 2001). We find the rCRCT/CRCA-1 module a more appealing candidate for intrageneric antagonism given most described CdiA-CT/Cdil and Rhs TA modules arm bacteria for battle with self or closely related species (Aoki et al., 2010; Poole et al., 2011; Zhang et al., 2011).

While often tethered to larger N-terminal sequences with toxic activity, CdiA-CT domains themselves are toxins (Nikolakakis et al., 2012). Curiously, the N-terminal sequence of REIS_1424 contains a putative **ubiquitin-associated (UBA)** domain that is separated from the C-terminal CRCT domain by two “VENN” motifs, which typically
demarcate the CdiA-CT domain from the remaining protein in most other CDI systems (Aoki et al., 2010; Zhang et al., 2011) (Fig. S6). If functional, this UBA domain may recruit host ubiquitin to the target cell and render it vulnerable to proteasomal destruction. Combined with toxin effect of the C-terminal domain, rCRCT-1 would be highly effective for targeting conspecific bacteria in the intracellular environment.

**rCRCT/CRCA-1 expression and possible routes for secretion**

During infection of ISE6 cells, which were originally derived from *I. scapularis* embryos (Munderloh et al., 1999), *R. buchneri* expresses the REIS_1424/REIS_1423 module (Fig. 6A). The same primers did not amplify product in PCR reactions using DNA template from 10 diverse rickettsiae, illustrating their specificity and efficacy for future surveys of blacklegged tick populations. This is demonstrated by testing the primers on a small sampling of blacklegged ticks from three different populations throughout the *I. scapularis* range (Fig. 6B). Future sequencing of these loci will provide resolution on the conservation rCRCT/CRCA-1 and whether there is evidence for an arms race between *R. buchneri* that have diverged throughout the *I. scapularis* geographic range.

As *R. buchneri* REIS_1424 lacks a predicted signal sequence, it is likely secreted via one of two Sec-independent pathways: **type I secretion system (T1SS)** or **type IV secretion system (T4SS)**, which are both conserved in rickettsiae (Gillespie et al., 2015b). Inspection of the neighborhood loci around select CRCT/CRCA modules indicates several diverse secretion pathways for some CRCT-DCPs, including T4SS (*Sphingobium chlorophenolicum*) and **type VI secretion system (T6SS)** (*Sphaerisporangium krabiense*), as well as typical CDI systems (*Pseudomonas*...
A novel *Rickettsia buchneri* toxin-antidote module

orientedis and *Enterobacter* sp. BIGb0383) with nearby cdB loci, which encode the outer membrane CdIB that translocates the large CdiA protein as a **type Vb secretion system** (T5bSS) (**Fig. S7A-G**). Using CdIB as a query in Blastp searches against the major taxa harboring CRCT/CRCA modules revealed their scarcity in Rickettsiales (none detected in rickettsiae), Actinobacteria, Chloroflexi, and Spirochaetes genomes, but widespread distribution in other proteobacterial and Cyanobacteria genomes (**Fig. S7H,I**). This indicates that bacteria employing the CRCT/CRCA modules we describe here for warfare utilize a variety of secretions systems (e.g., T1SS, T4SS, T5bSS, T6SS, and likely others) consistent with the plethora of secretory pathways now characterized for diverse TA modules involved in interbacterial antagonism (Ruhe et al., 2018; Lin et al., 2020). The lack of cdB genes and evidence that rCRCT toxins were originally appended to larger proteins indicates these modular Rhs toxins were once widespread in *Rickettsia* genomes. This is reminiscent of the large modular toxins we identified across numerous intracellular bacteria that encode a myriad of eukaryotic-like domains, some of which function in commandeering host reproduction (Gillespie et al., 2018). Like rCRCT/CRCA modules, many of these variable toxins are found adjacent to genes encoding probable antidotes, indicating a recapitulating theme for toxin architecture that persists evolutionarily and drives innovative strategies for colonizing eukaryotic hosts.

**Conclusion**

Since our initial report on its genome (Gillespie et al., 2012a) and its subsequent formal species description (Kurtti et al., 2015), the appreciation for the oddity of *R. buchneri*
A novel *Rickettsia buchneri* toxin-antidote module

relative to other rickettsiae has grown. Numerous reports on the *I. scapularis* microbiome now attest to the high infection rate of this microbe, particularly in females, throughout the blacklegged tick geographic range. While few reports indicate tick salivary gland infection (Steiner et al., 2008; Zolnik et al., 2016; Al-Khafaji et al., 2020), ovaries are the predominant tissue infected, consistent with the lack of reports on vertebrate infection or presence in other arthropods that co-feed on blacklegged tick hosts. This indicates a unique endosymbiosis, the intricacies of which stand to be illuminated in light of the powerful tools created by Kurtti, Munderloh and colleagues for studying this system (Munderloh et al., 1999; Simser et al., 2002; Kurtti et al., 2016; Oliver et al., 2021). Ultimately, the fitness of blacklegged ticks uninfected with *R. buchneri* (Oliver et al., 2021) needs to be evaluated, as does the possibility that other endosymbionts may replace *R. buchneri* in a mutualistic capacity.

While previously observed in ticks (Burgdorfer et al., 1981; Macaluso et al., 2002; Wright et al., 2015; Levin et al., 2018) and cell culture (Cull et al., 2022), the ability of *Rickettsia* infection to block *Rickettsia* superinfection remains a sorely understudied aspect of vector biology. Our identification and characterization of rCRCT/CRCA modules adds to a short list of factors, namely BOOM, AGAB and PKS-NRPS, that have been hypothesized to underpin a mutualism between *R. buchneri* and blacklegged tick. Future characterization of these factors will determine their contribution to blocking superinfection of *I. scapularis* by *Rickettsia* pathogens (Cull et al., 2022). When mutagenesis is someday an efficacious tool for bioengineering Rickettsiae, this line of research will offer a gene drive tool (*R. buchneri*) ready to disseminate into blacklegged tick populations to combat the spread of human disease agents,
A novel Rickettsia buchneri toxin-antidote module

Acknowledgments

We thank Alison Luce-Fedrow (Shippensburg University), Richard Ostfeld (Cary Institute of Ecosystem Studies), and Kathryn Cottingham (Dartmouth College) for providing tick specimens, and Kevin Macaluso (University of South Alabama) and Sean Riley (University of Maryland College Park). We are grateful to Uli Munderloh and Tim Kurtti (University of Minnesota) for sharing the ISE6 cell line and R. buchneri strain IS07.

This work was supported with funds from the National Institute of Health/National Institute of Allergy and Infectious Diseases grant R21AI156762 (JJG and TPD). TDF was supported by the Biotechnical Institute of Maryland. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References


A novel *Rickettsia buchneri* toxin-antidote module


593 doi:10.1038/nature09490.


603 doi:10.1038/s41598-018-25545-y.


A novel *Rickettsia buchneri* toxin-antidote module


A novel *Rickettsia buchneri* toxin-antidote module


Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and
A novel *Rickettsia buchneri* toxin-antidote module


Gillespie, J. J., Driscoll, T. P., Verhoeve, V. I., Utsuki, T., Husseneder, C., Chouljenko,
A novel *Rickettsia buchneri* toxin-antidote module


A novel *Rickettsia buchneri* toxin-antidote module


704 Johnson, P. M., Gucinski, G. C., Garza-Sánchez, F., Wong, T., Hung, L. W., Hayes, C.

705 S., et al. (2016). Functional Diversity of Cytotoxic tRNase/Immunity Protein


709 *Wolbachia* supplement biotin and riboflavin to enhance reproduction in


712 (2001). Beta-helix model for the filamentous haemagglutinin adhesin of *Bordetella*


714 doi:10.1046/J.1365-2958.2001.02598.X.


717 doi:10.1038/nprot.2009.2.


719 (2018). Target highlights from the first post-PSI CASP experiment (CASP12, May–


723 Protein Expressing *Rickettsia buchneri* and *Rickettsia peacockii* for Tracking


A novel *Rickettsia buchneri* toxin-antidote module


A novel *Rickettsia buchneri* toxin-antidote module


A novel *Rickettsia buchneri* toxin-antidote module


doi:10.1073/PNAS.1216238110/-/DCSUPPLEMENTAL.


doi:10.1128/JCM.37.8.2518-2524.1999/ASSET/225E7731-33D8-4A77-AB6E-
A novel *Rickettsia buchneri* toxin-antidote module


A novel *Rickettsia buchneri* toxin-antidote module


A novel *Rickettsia buchneri* toxin-antidote module


A novel *Rickettsia buchneri* toxin-antidote module


van Treuren, W., Ponnusamy, L., Brinkerhoff, R. J., Gonzalez, A., Parobek, C. M.,
A novel *Rickettsia buchneri* toxin-antidote module


FIGURE LEGENDS

FIGURE 1. Two extraordinary features of the R. buchneri genome include RAGEs and four unique plasmids (pREIS1-4) (Gillespie et al., 2012a). (A) General schema of RAGEs: INT, integrase; MOB, mobilization genes (traT and traAI). RAGEs typically insert in tRNA genes near cytosine kinase (CK)-encoding loci. Cargo genes (black)
A novel *Rickettsia buchneri* toxin-antidote module

Occur in variable clusters. Grey inset lists recurring cargo genes grouped by class. (B)

Schema of pREIS plasmids: pink, full-length RAGE; pink w/ asterisk, MOB genes;
yellow, BOOM (illustrated in yellow inset). BOOM genes are uniquely arrayed relative to
other bacteria that carry biotin synthesis genes (Gillespie et al., 2012a)

**FIGURE 2.** REIS_1424 of *Rickettsia buchneri* contains a C-terminal toxin domain
characteristic of some bacterial contact-dependent growth inhibition (CDI) and
rearrangement hotspot (Rhs) toxins. This domain was named CDI-like/Rhs-like C-
terminal toxin (CRCT). (A) HaloBlast analysis for “Candidatus Jidaibacter
acanthamoeba” CRCT of NF27_IC00050 (NCBI acc. no. KIE04387; aa residues 2068-
2192). Concentric halos depict hierarchical taxonomic databases increasing in
divergence from the center. Average Sm score (see text for details) for all subjects and
top ten subjects are provided, with highest score per database highlighted. ‘na’, not
applicable. All corresponding information for proteins from HaloBLAST analyses are
provided in **Table S1.** (B-E) For compiled bacterial CRCT-containing proteins (*n* = 155):
(B) number of aa residues between CRCT and C-terminus, (C) lengths for associated
N-terminal regions parsed by taxonomic group (the two largest proteins for
*Proteobacteria* (AZE30872, *Pseudomonas chlororaphis* subsp. Aureofaciens and
TVR95235, Wenzhouxiangellaceae bacterium are not shown), (D) NCBI protein
annotations, and (E) conservation within the CRCT with sequence logo illustrating
alignment of 155 CRCTs (see full alignment in **Fig. S1**) and amino acid coloring as
follows: black, hydrophobic; red, negatively charged; green, hydrophilic; purple,
aromatic; blue, positively charged; star depicts unique *Rickettsia* insertion and triangles
other insertions). (F) Diverse CRCT-containing proteins from select genomes.

**FIGURE 3.** REIS_1423 of *Rickettsia buchneri* is a predicted antidote to REIS_1424.

REIS_1423 and related proteins were named CDI-like/Rhs-like immunity antidotes (CRIA). (A) Blastp searches with REIS_1423 and NF27_IC00040 unearthed putative CRIAs adjacent to 37% of the 155 CRCT domain-containing toxins. In 61 cases, assignment of CRIAs to adjacent CRCT domain-containing toxins could not be made due to the lack of strain-specific NCBI reference protein accession numbers (non-redundant protein record (WP_) only). (B) HaloBlast analysis for “Candidatus Jidaibacter acanthamoeba” CRCT of NF27_IC00040 (NCBI acc. no. KIE04386). Concentric halos depict hierarchical taxonomic databases increasing in divergence from the center. Average Sm score (see text for details) for all subjects and top ten subjects are provided, with highest score per database highlighted. ‘na’, not applicable. All corresponding information for proteins from HaloBLAST analyses are provided in Table S2. (C) Taxonomic breakdown of non-proteobacterial hits retrieved in HaloBlast analysis of CRCT domain-containing toxins and CRIAs. (D) Lengths for CRIAs. (E) Top ten blastp subjects by Sm score (see text for details) from ‘Alph’, ‘Prot’, and ‘Bact’ searches for CRCT domain-containing toxins and CRIAs. (F) CRIA conservation; sequence logos illustrate alignment of 380 CRIAs (see full alignment in Fig. S2), with amino acid coloring as follows: black, hydrophobic; red, negatively charged; green, hydrophilic; purple, aromatic; blue, positively charged; star depicts unique *Rickettsia* insertion and triangles other insertions). (G) Diverse CRCT/CRIA modules from select genomes. White stipple on REIS_1423 indicate an adjusted start site.
A novel *Rickettsia buchneri* toxin-antidote module

**FIGURE 4.** The mobile nature of CRCT/CRCA modules captured in a eukaryotic genome assembly. In blastp searches against the NCBI “non-bacteria” database, the CRCT/CRIA module of “*Candidatus Jidaibacter acanthamoeba*” str. UWC36 consistently hit only predicted proteins from the smooth cauliflower coral (*Stylophora pistillata*) genome. *Left*, the large N-terminal region of NF27_IC00050 is similar to a large protein (PFX12133, 2285 aa) encoded on a five-gene *S. pistillata* scaffold (NW_019219349). PFX12133 domain architecture (descriptions in gray inset at bottom) is reminiscent of large, multi-domain hemagglutinin-like RHS toxins that may or may not carry CRCT domains (see Fig. 2 and Fig. 3). The HaloBlast profile of PFX12133 and adjacent proteins indicates either rampant bacterial gene incorporation into the *S. pistillata* genome or mis-assembly of bacterial sequencing reads from *S. pistillata*-associated microbes. See text and Fig. 2 legend for description of HaloBlast.

PFX12132: stomatin-like protein 2, mitochondrial; PFX12133: uncharacterized protein YbeQ; PFX12133: uncharacterized protein YqaJ; PFX12133: hypothetical protein AWC38_SpisGene23959; PFX12135: DNA polymerase I. *Right*, the complete or partial CRCT/CRCA module was detected in ten smaller predicted *S. pistillata* proteins encoded by genes on scaffolds not incorporated into the *S. pistillata* assembly.

**FIGURE 5.** Evolution of *Rickettsia* CRCT/CRCA modules. (A) Comparison of *R. tamurae* and *R. buchneri* genomic regions containing *Rickettsia* CRCT/CRCA module 1 (rCRCT/CRCA-1). Gray genes encode hypothetical proteins. Limited similarity between *R. buchneri* REIS_1425 and *Stylophora pistillata* YebQ is shown. Dashed inset:
A novel *Rickettsia buchneri* toxin-antidote module

1002 subjects retrieved from a Blastp search against the NCBI nr database using five
1003 concatenated *R. buchneri* proteins (REIS_1427-REIS_1423) as the query (further
1004 information provided in Fig. S4A). Gray box illustrates additional rCRCA-like sequences
1005 identified by this search. (B) rCRCA-2 genes are divergent from rCRCA-1 and
1006 pseudogenized in all Tamurae/Ixodes Group (TIG) rickettsiae except *Rickettsia*
1007 endosymbiont of *Ixodes pacificus*. Phylogeny of rCRCA-1, rCRCA-2 and other CRCA
1008 proteins was estimated with the WAG substitution model (gamma model of rate
1009 heterogeneity) using RAxML v8.2.4 (Stamatakis, 2014). Branch support was assessed
1010 with 1,000 pseudo-replications. Final ML optimization likelihood was -3490.9. (C)
1011 rCRCT/CRCA modules 2 and 3a are mostly clustered near the BamA gene in some
1012 *Rickettsia* genomes, while rCRCA-3b genes occur between genes encoding the
1013 cytochrome oxidase subunits 2 (CyoA) and 1 (CyoB) in certain *Rickettsia* genomes (see
1014 Fig. S3 for illustration of genomic regions near secA, bamA and cyoA/B loci). (D)
1015 *Rickettsia* phylogeny estimated from 92 genomes (see Fig. S5 for more information).
1016 Yellow highlighting depicts TIG rickettsiae as the only species harboring complete
1017 CRCT/CRCA modules. (E) Phylogenomics analysis of rCRCT/CRCA modules. Yellow
1018 highlighting indicates complete rCRCT/CRCA modules. Further information is provided
1019 for rCRCT-1 (Table S1), rCRCA-1 (Table S2), rCRCA-2 (Fig. S4B), rCRCA-3a and
1020 rCRCA-3b (Fig. S4C), and rCRCT-3a (Fig. S4D-G).

**FIGURE 6.** REIS_1424/REIS_1423 is a species-specific marker. (A) PCR assay for
1022 *rCRCT* and *rCRCA* in 11 different rickettsiae. Citrate synthase (*gltA*) was used as a
1023 positive control. (B) The same PCR assay was conducted on *I. scapularis* from three
A novel *Rickettsia buchneri* toxin-antidote module

populations.

**SUPPLEMENTARY MATERIAL**

**FIGURE S1.** Sequence analysis of 155 predicted CRCTs integrated into diverse bacterial toxins. (A) Sequence logos (Crooks et al., 2004) depict complete protein alignment generated using MUSCLE, default parameters (Edgar, 2004). Information for all proteins is provided in Table S1. Regions shown in Figure 2 are denoted with a bar above the logos. Amino acid coloring as follows: black, hydrophobic; red, negatively charged; green, hydrophilic; purple, aromatic; blue, positively charged. Unique *Rickettsia* insertion is shown in orange. Other insertions noted by triangles in Figure 2 are highlighted yellow). (B) HaloBlast profiles for REIS_1424 with and without a 23 aa insertion (see Fig. 2E and text for more details).

**Table S1.** Information pertaining to 155 predicted CRCTs integrated into diverse bacterial toxins.

**FIGURE S2.** Sequence analysis of 380 predicted CRCAs. Sequence logos (Crooks et al., 2004) depict complete protein alignment generated using MUSCLE, default parameters (Edgar, 2004). Information for all proteins is provided in Table S2. Regions shown in Figure 3 are denoted with a bar above the logos. Amino acid coloring as follows: black, hydrophobic; red, negatively charged; green, hydrophilic; purple, aromatic; blue, positively charged. Sequence spanning the adjusted start site of
REIS_1423 (*Rickettsia buchneri*) insertion is shown in blue.

**Table S2.** Information pertaining to 380 predicted CRCA modules.

**Table S3.** Evidence for CRCT/CRCA modules in the *Stylophora pistillata* genome assembly.

**FIGURE S3.** *Rickettsia* CRCT/CRCA modules occur in three recombination hotspots. (A) rCRCT/CRCA-1 occurs in a recombination hotspot near the SecA and tRNA-Ala^{TGC} genes. (B) rCRCT/CRCA-2 occurs in a recombination hotspot near the BamA and tRNA-Thr^{CGT} genes. A second TA module, rCRCT/CRCA-3a, also occurs in this region and is distinct from rCRCT/CRCA-1 and rCRCT/CRCA-2 (cd20695: CdiA-CT_5T87E_Ct, cd20694: CdiI_Ct-like). (C) rCRCT/CRCA-3b proteins are analogous to rCRCT/CRCA-3a and occur in a subset of *Rickettsia* genomes between *cyoB* and *cyoA*, which encode the cytochrome c oxidase subunits I and II, respectively.

**FIGURE S4.** Evidence for recurrent integration of diverse CRCT/CRCA modules in *Rickettsia* genomes at recombination hotspots. (A) Subjects (n = 40) retrieved from a Blastp search against the NCBI nr database using five concatenated *R. buchneri* proteins (REIS_1427-REIS_1423) as the query. **Bottom**, a portion of the large protein (PFX12133, 2285 aa) from the coral *Stylophora pistillata* shares similarity with these smaller *Rickettsia* proteins. (B) rCRCT-2 proteins are analogous to rCRCT-1 proteins. **Alignment performed using MUSCLE with default settings (Edgar, 2004).** (C) rCRCA-3
A novel *Rickettsia buchneri* toxin-antidote module

proteins recovered from Blastp searches using *R. bellii* rCRCA-3b as a query (the
largest rCRCA-3 protein). na, sequences not recovered in Blastp searches against the
NCBI `Rickettsia’ database but retrieved from PATRIC (asterisks denote PATRIC Local
Family IDs). Subjects are listed accordingly to their placement in the phylogeny
presented in Figure 5E. Colored boxes unite rCRCA-3 proteins recombined within the
same genome. (D) rCRCT-3a HaloBalst analysis. *R. tamurae* rCRCT-3a
(WP_051965318) was used as the query. Concentric halos depict hierarchical
taxonomic databases increasing in divergence from the center. Average Sm score (see
text for details) for all subjects and top ten subjects are provided, with highest score per
database highlighted. ‘na’, not applicable. (E) Top 25 subjects from the Blastp search
against ‘Proteobacteria’ using *R. tamurae* rCRCT-3a as the query. (F) Comparison of
the *Cupriavidus taiwanensis* str. DSM 17343 CdiA toxin (Uniprot acc. B3R1C1) to
rCRCT-3a of *R. tamurae*. Domains were predicted with SMART (Letunic and Bork,
2017). Amino acid similarity (%ID, red shading) was assessed using Blastp. (G)
Structural analysis of a rCRCT/CRCA-3a module. (top) Alignment of residues 3391-
3469 of *C. taiwanensis* CdiA, *R. tamurae* rCRCT-3a, and REIP rCRCT-3a. Structural
information from *C. taiwanensis* CdiA (PDB:5T87) (Kryshtafovych et al., 2018) is
provided at top. Alignment performed using MUSCLE with default settings (Edgar,
2004). (bottom left) Modeling with Phyre2 of *R. tamurae* rCRCT-3a to the CdiA-CT
toxin structure of *C. taiwanensis* CdiA (PDB:5T87). The threading was done with 100%
confidence. (bottom right) While the *C. taiwanensis* module was solved as a co-
complex (Kryshtafovych et al., 2018), Phyre2 modeling could not thread *R. tamurae*
rCRCA-3a to the antidote CdiI within the TA co-complex. The best model (46.3%
A novel *Rickettsia buchneri* toxin-antidote module

1094 confidence, 12% ID) for *R. tamurae* rCRCA-3a was to the structure of *Drosophila*
melanogaster MAST/Orbit N-terminal domain PDB:4G3A (De La Mora-Rey et al., 2013)

**FIGURE S5. Rickettsia genome-based phylogeny estimation.** *Rickettsia* groups

1097 follow previous classification (Gillespie et al., 2007), except that we recognize
1098 Tamurae/Ixodes Group (TIG) rickettsiae as a distinct clade from SFG rickettsiae. *R.
1099 buchneri* is highlighted. Phylogeny was estimated for 92 *Rickettsia* genomes; gray inset
described details (see “Materials and Methods” for more details). Branch support was
assessed with 1,000 pseudoreplications.

**FIGURE S6.** REIS_1424 contains a ubiquitin-associated domain demarcated from
the C-terminal toxin by two predicted VENN motifs.

1104 (Mueller and Feigon, 2002);
1107 (E) Schemas for a few. Accession nos. Described domains here or in the M&M and get
refs from SMART (Letunic and Bork, 2017).
1109 Wolbachia supergroups are within colored ellipses. Ctenocephalides felis-

**FIGURE S7.** Gene neighborhoods for CRCT/CRCA modules in diverse bacteria.

1112 (A-G) Genome neighborhoods for the CRCT/CRCA modules shown in Figures 1 and 2
for (A) “*Candidatus* Jidaibacter acanthamoeba”, (B) *Sphingobium chlorophenolicum*, (C)
Holosporales bacterium, (D) *Pseudomonas orientalis*, (E) *Enterobacter* sp. BIGb0383,
(F) *Streptomyces luteoverticillatus*, (G) *Sphaerisporangium krabiense*. (H) Results of
Blastp searches against these seven genomes (specific taxon databases at NCBI)
using the CdiB protein of *Escherichia coli* (UniProtKB-Q3YL97). (I) Occurrence of CdiB and CdiB-like proteins in the major taxa in which CRCT/CRCA modules were identified.
**A**

<table>
<thead>
<tr>
<th>METABOLISM</th>
<th>CELLULAR REGULATION</th>
<th>EFFECTORS</th>
<th>DEFENSE and RESISTANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>- F-T4SS (~16 genes)</td>
<td>- INT</td>
<td>- Histidine kinases</td>
<td>- aminoglycoside synthesis genes</td>
</tr>
<tr>
<td>- ribonucleotide translocase (Tlc)</td>
<td>- MOB</td>
<td>- stringent response regulators (SpoT)</td>
<td>- aminoglycoside resistance genes</td>
</tr>
<tr>
<td>- SAM transporter (EamA)</td>
<td>- MOB</td>
<td>- muropeptide transporters (AmpG)</td>
<td>- multi-drug exporter (MdlB)</td>
</tr>
<tr>
<td>- osmoregulatory transport (ProP)</td>
<td>- DNA methyltransferases</td>
<td>- O-antigen sugar synthesis (Udg)</td>
<td>- nonribosomal megasynthase</td>
</tr>
<tr>
<td>- muropeptide transporters (AmpG)</td>
<td>- primosomal protein DnaI</td>
<td>- glycerol-3-P importer (GlpT)</td>
<td>- phospholipase D (PLD)</td>
</tr>
<tr>
<td>- O-antigen sugar synthesis (Udg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- glycerol-3-P importer (GlpT)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**B**

- aminoglycoside synthesis genes
- aminoglycoside resistance genes
- multi-drug exporter (MdlB)
- nonribosomal megasynthase

---

Fig. 1
**Fig. 2**

**A** Taxon database

- Rickettsiales
- Alphaproteobacteria
- Proteobacteria
- Bacteria

**B** Protein length (aa)

- Rick
- Alph
- Prot
- Bact

**C** Protein length (aa)

- Rick
- Alph
- Prot
- Bact

**D** Taxon (genome)

- Rickettsia buchneri
- “Cand. Jidaibacter acanthamoeba”
- Sphingobium chlorophenolicum
- Holosporales bacterium
- Pseudomonas orientalis
- Enterobacter sp. BIGb0383
- Streptomyces luteoviridatus
- Sphaerisporangium krabiense

**E** Locus Tag

- REIS_1424
- NF27_IC00050
- Sphch_1491
- J0H87_08305
- CAJ97_0166
- EDF81_3004
- EKHH7_08965
- BJ981_004090

**F** Ubiquitin associated domain
- ENDO3c (endonuclease III)
- Carbohydrate-binding module 2
- Peptide_M43

**G** Hypothetical protein
- RHS repeat-associated core domain-containing protein
- VENN motif pre-toxin domain-containing protein
- ALF repeat-containing protein
- filamentosus hemagglutinin
- DUF637 domain-containing protein
- Putative large exoprotein, ShlA/HecA/FhaA family
- Hemolysin
- choice-of-anchor A family protein
- chemotaxis protein

**H** Present once: PAAR domain-containing protein, pre-toxin TG domain-containing protein, helix-turn-helix domain-containing protein, IPT/TIG domain-containing protein, PrsW family intramembrane metalloprotease, uncharacterized protein YukE, sugar-binding protein

**I** Unique repeat
- ALF repeat
- VENN motif pre-toxin domain-containing protein

**J** ALF repeat-containing protein
- VENN motif pre-toxin domain-containing protein

**K** VenN motif pre-toxin domain-containing protein
- Hemolysin
- Putative large exoprotein, ShlA/HecA/FhaA family

**L** Choice-of-anchor A family protein
- Chemotaxis protein

**M** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**N** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**O** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**P** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**Q** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**R** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**S** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**T** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**U** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**V** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**W** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**X** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**Y** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family

**Z** Hemagglutinin
- Putative large exoprotein, ShlA/HecA/FhaA family
**Fig. 4**

<table>
<thead>
<tr>
<th>NT domains</th>
<th>CRCT/CRCA module</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haemagglutination activity site</td>
<td>CidA-like/Rhs-like CT toxin</td>
</tr>
<tr>
<td>Hemagglutinin repeats</td>
<td>CidA-like/Rhs-like immunity antidote</td>
</tr>
<tr>
<td>Peptidase_M43</td>
<td>ENDO3c (endonuclease III)</td>
</tr>
<tr>
<td>Laminin G</td>
<td>AHH of HNH/ENDO VII nuclease</td>
</tr>
<tr>
<td>UBA (ubiquitin-associated)</td>
<td>SEL repeat</td>
</tr>
</tbody>
</table>

**Table:**

<table>
<thead>
<tr>
<th></th>
<th>PFX12132</th>
<th>PFX12133</th>
<th>PFX12131</th>
<th>PFX12134</th>
<th>PFX12135</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (aa)</td>
<td>717</td>
<td>1-1954</td>
<td>1955-2285</td>
<td>343</td>
<td>676</td>
</tr>
<tr>
<td>No. introns</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>x Sm of top blast hits</td>
<td>307.79</td>
<td>5442.69</td>
<td>2248.74</td>
<td>1189.57</td>
<td>1106.82</td>
</tr>
<tr>
<td>Alph</td>
<td>2798.29</td>
<td>ns</td>
<td>6196.72</td>
<td>1386.11</td>
<td>657.18</td>
</tr>
<tr>
<td>Prot</td>
<td>7520.67</td>
<td>247.62</td>
<td>7111.27</td>
<td>1506.69</td>
<td>ns</td>
</tr>
<tr>
<td>Bact</td>
<td>9970.19</td>
<td>330.01</td>
<td>6933.99</td>
<td>2252.18</td>
<td>ns</td>
</tr>
<tr>
<td>- Bact</td>
<td>9436.66</td>
<td>78.41</td>
<td>6672.25</td>
<td>3655.15</td>
<td>1 phage</td>
</tr>
</tbody>
</table>

**Figure:**

- **Stylophora pistillata**
- Scaffold NW_019219349
- Cand. Jidaibacter acanthamoeba CRCT-DCP (2192)
- CRCA (124)

**Table values:**

<table>
<thead>
<tr>
<th></th>
<th>PFX11133</th>
<th>PFX1109 (154)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x Sm of top blast hits</td>
<td>80</td>
<td>36</td>
</tr>
<tr>
<td>PFW98200 (123)</td>
<td>34</td>
<td>51</td>
</tr>
<tr>
<td>PFW97864 (225)</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>PFX11617 (308)</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>PFX11140 (482)</td>
<td>1106.82</td>
<td>6162.58</td>
</tr>
<tr>
<td>PFX97864 (225)</td>
<td>1386.11</td>
<td>ns</td>
</tr>
<tr>
<td>PFX97864 (225)</td>
<td>657.18</td>
<td>6162.58</td>
</tr>
<tr>
<td>PFX11140 (482)</td>
<td>1106.82</td>
<td>6162.58</td>
</tr>
<tr>
<td>PFX97864 (225)</td>
<td>1386.11</td>
<td>ns</td>
</tr>
<tr>
<td>PFX11140 (482)</td>
<td>657.18</td>
<td>6162.58</td>
</tr>
</tbody>
</table>
### Table B

<table>
<thead>
<tr>
<th>Locality</th>
<th>rCRCT</th>
<th>rCRCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY</td>
<td>4/5</td>
<td>5/5</td>
</tr>
<tr>
<td>NH</td>
<td>5/5</td>
<td>5/5</td>
</tr>
<tr>
<td>PA</td>
<td>9/10</td>
<td>7/10</td>
</tr>
</tbody>
</table>
removing the insertion results in greatest similarity to non-proteobacterial proteins
rCRCT/CRCA-1 is found in a region of recombination adjacent to the SecA gene.
rCRCT/CRCA-2 is found in a region of recombination adjacent to the BamA gene.

rCRCT/CRCA-3a is also found in this region, indicating multiple insertion of distinct CRCT/CRCA s.
C

rCRCT/CRCA-3a is similar to rCRCT/CRCA-3b between subunits I and II of cytochrome c oxidase.

No toxins occur with CRCA-3b and this region has little recombination relative to SecA and BamA.

part of degraded cytochrome c oxidase pathway in *R. typhi*
### A

<table>
<thead>
<tr>
<th>Rickettsia tamurae</th>
<th>Max Score</th>
<th>Query Cover</th>
<th>E value</th>
<th>Per. ident</th>
<th>Acc. Len</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>77.8</td>
<td>77.8</td>
<td>11%</td>
<td>4.00E-06</td>
<td>4.00E-06</td>
<td>87.88</td>
<td>WP_152133.1</td>
</tr>
</tbody>
</table>

**Hypothetical protein**

**Stylophora pistillata**

<table>
<thead>
<tr>
<th>Rickettsia japonica</th>
<th>Max Score</th>
<th>Query Cover</th>
<th>E value</th>
<th>Per. ident</th>
<th>Acc. Len</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>155</td>
<td>155</td>
<td>32%</td>
<td>2.00E-43</td>
<td>94.12</td>
<td>100</td>
<td>WP_0141209.1</td>
</tr>
</tbody>
</table>

**Hypothetical protein**

**Stylophora pistillata**

<table>
<thead>
<tr>
<th>Rickettsia peacockii</th>
<th>Max Score</th>
<th>Query Cover</th>
<th>E value</th>
<th>Per. ident</th>
<th>Acc. Len</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>136</td>
<td>136</td>
<td>32%</td>
<td>2.00E-43</td>
<td>94.12</td>
<td>100</td>
<td>WP_0141209.1</td>
</tr>
</tbody>
</table>

**Hypothetical protein**

**Stylophora pistillata**

<table>
<thead>
<tr>
<th>Rickettsia philippii</th>
<th>Max Score</th>
<th>Query Cover</th>
<th>E value</th>
<th>Per. ident</th>
<th>Acc. Len</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>145</td>
<td>145</td>
<td>32%</td>
<td>2.00E-43</td>
<td>94.12</td>
<td>100</td>
<td>WP_0141209.1</td>
</tr>
</tbody>
</table>

**Hypothetical protein**

**Stylophora pistillata**

### B

**rCRCA-1**

- **C** = Complete
- **S** = Split (multiple CDS)
- **F** = Fragment
- **M** = Start site of additional CDS

<table>
<thead>
<tr>
<th>Rickettsia tamurae</th>
<th>Max Score</th>
<th>Query Cover</th>
<th>E value</th>
<th>Per. ident</th>
<th>Acc. Len</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>600</td>
<td>600</td>
<td>32%</td>
<td>2.00E-43</td>
<td>94.12</td>
<td>100</td>
<td>WP_0141209.1</td>
</tr>
</tbody>
</table>

**rCRCA-2**

<table>
<thead>
<tr>
<th>Rickettsia monacensis</th>
<th>Max Score</th>
<th>Query Cover</th>
<th>E value</th>
<th>Per. ident</th>
<th>Acc. Len</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>100</td>
<td>32%</td>
<td>2.00E-43</td>
<td>94.12</td>
<td>100</td>
<td>WP_0141209.1</td>
</tr>
</tbody>
</table>

### Fig. S4

- **rCRCA-2 NCBI acc. #s:** R. tamurae (WP_032139358); R. buchneri (EER22830, EER22831); REIP (KJW03593); R. monacensis (CEO17090, CEO17089).
Further comparisons shown in Fig. S4F, G.
**Fig. S4**

**Extended Signal Peptide Region**
- C. taiwanensis
- R. tamurae
- REIP

**Haemagglutination activity site**
- C. taiwanensis
- R. tamurae
- REIP

**Hemagglutinin repeats**
- C. taiwanensis
- R. tamurae
- REIP

**VENN motif (cleavage site)**
- C. taiwanensis
- R. tamurae
- REIP

---

**rCRCT/CRCA-3a**
- R. tamurae
- C. taiwanensis

**~ 22% ID between either CRCA-3a protein and C. taiwanensis Cdil protein**

**Extended Signal Peptide Region**
- CdiA-CT_5T87E_Ct
- CdiI_Ct-like

**Haemagglutination activity site**
- CdiA-CT

**Hemagglutinin repeats**
- CdiI

**VENN motif (cleavage site)**
- CdiI

---

**rCRCT-3a NCBI acc. #s:**
- R. tamurae (WP_051965318) and REIP (KJW03590).

**Solved as a toxin-antidote complex**

**Toxins in these diverse TA modules are 34% ID**

**Antidotes in these diverse TA modules are too divergent for modeling; however, rCRCA-3a modeling to 4G3A indicates a similar helical bundle and possible similar topology as C. taiwanensis Cdil**
A. “Cand. Jidaibacter acanthamoeba” Jswe01000196 98-8417

Tandemly-arrayed CRCT-CRCA modules; one tethered to a large toxin, one comprising mostly the toxin itself.

NF27_IC00040/NF27_IC00050

B. Sphingobium chlorophenolicum 1602558-1613831

Type IV secretion system genes Sphch_1492/Sphch_1491

C. Holosporales bacterium JAFKHM010000035 1-11875

Highly divergent! Highly divergent! Highly divergent!

D. Pseudomonas orientalis 184384-1217426

Biofilm PGA synthesis CdiB CdiA CdiI C4J97_0166/C4J97_0167

Highly divergent!

E. Enterobacter sp. BIGb0383 1372977-1398714

Degraded CdiA-like EDF81_3004/EDF81_3003 CdiA

Highly divergent!

F. Streptomyces luteoverticillatus 2111023-2131620

PhzF EKH77_08965/EKH77_08960

Putative secreted effectors

G. Sphaerisporangium krabiense 4635008-4665060

LigD degraded TNP MerR GvR: T6SS, phage-baseplate injector protein

Putative T6SS

H. Taxon (genome) Subjects (with > 80% query coverage)

“I. Taxon No. No. > 80% query cov.

- Rickettsiales 16 10 (no Rickettsia species)
- Actinobacteria 12 2
- Cyanobacteria > 1000 > 1000 (widespread)
- Chloroflexi 4 1
- Spirochaetes 12 5

Widespread in all proteobacterial classes and many other non-proteobacterial groups