

Metagenome diversity illuminates the origins of pathogen effectors

Stephanie S. Lehman,¹ Victoria I. Verhoeve,² Timothy P. Driscoll,³ John F. Beckmann,⁴ Joseph J. Gillespie²

AUTHOR AFFILIATIONS See affiliation list on p. 22.

ABSTRACT Recent metagenome-assembled genome (MAG) analyses have profoundly impacted Rickettsiology systematics. The discovery of basal lineages (novel families Mitibacteraceae and Athabascaceae) with predicted extracellular lifestyles exposed an evolutionary timepoint for the transition to host dependency, which seemingly occurred independent of mitochondrial evolution. Notably, these basal rickettsiae carry the Rickettsiales *vir* homolog (*rvh*) type IV secretion system and purportedly use *rvh* to kill congener microbes rather than parasitize host cells as described for later-evolving rickettsial pathogens. MAG analysis also substantially increased diversity for the genus *Rickettsia* and delineated a sister lineage (the novel genus *Tisiphia*) that stands to inform on the emergence of human pathogens from protist and invertebrate endosymbionts. Herein, we probed Rickettsiales MAG and genomic diversity for the distribution of *Rickettsia rvh* effectors to ascertain their origins. A sparse distribution of most *Rickettsia rvh* effectors outside of Rickettsiaceae lineages illuminates unique *rvh* evolution from basal extracellular species and other rickettsial families. Remarkably, nearly every effector was found in multiple divergent forms with variable architectures, indicating profound roles for gene duplication and recombination in shaping effector repertoires in *Rickettsia* pathogens. Lateral gene transfer plays a prominent role in shaping the *rvh* effector landscape, as evinced by the discovery of many effectors on plasmids and conjugative transposons, as well as pervasive effector gene exchange between *Rickettsia* and *Legionella* species. Our study exemplifies how MAGs can yield insight into pathogen effector origins, particularly how effector architectures might become tailored to the discrete host cell functions of different eukaryotic hosts.

IMPORTANCE While rickettsioses are deadly vector-borne human diseases, factors distinguishing *Rickettsia* pathogens from the innumerable bevy of environmental rickettsial endosymbionts remain lacking. Recent metagenome-assembled genome (MAG) studies revealed evolutionary timepoints for rickettsial transitions to host dependency. The *rvh* type IV secretion system was likely repurposed from congener killing in basal extracellular species to parasitizing host cells in later-evolving pathogens. Our analysis of MAG diversity for over two dozen *rvh* effectors unearthed their presence in some non-pathogens. However, most effectors were found in multiple divergent forms with variable architectures, indicating gene duplication and recombination-fashioned effector repertoires of *Rickettsia* pathogens. Lateral gene transfer substantially shaped pathogen effector arsenals, evinced by the discovery of effectors on plasmids and conjugative transposons, as well as pervasive effector gene exchanges between *Rickettsia* and *Legionella* species. Our study exemplifies how MAGs yield insight into pathogen effector origins and evolutionary processes tailoring effectors to eukaryotic host cell biology.

KEYWORDS *Rickettsia*, metagenome, type IV secretion system, effector, evolution

Editor Vaughn S. Cooper, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Address correspondence to Joseph J. Gillespie, Jgillespie@som.umaryland.edu.

The authors declare no conflict of interest.

See the funding table on p. 22.

Received 27 March 2023

Accepted 12 February 2024

Published 2 April 2024

Copyright © 2024 Lehman et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Until recently, Order Rickettsiales (*Alphaproteobacteria*) contained three families harboring diverse obligate intracellular parasites (1). Rickettsiaceae and Anaplasmataceae are best studied and harbor invertebrate endosymbionts, human pathogens, and reproductive parasites (2–7). Midichloriaceae contains some arthropod-associated bacteria of unknown vertebrate pathogenicity (8), but most species are described from protists (9–14). Remarkably, Castelli and colleagues (15) described the first extracellular rickettsial species, “*Candidatus Deianiraea vastatrix*,” as a bacterium dependent on *Paramecia* and sharing many characteristics of the intracellular lifestyle. A new family, Deianiraeaceae, was proposed, calling into question the specific timepoint in rickettsial evolution wherein obligate intracellularity emerged from an obligate extracellular or facultative intracellular lifestyle.

Historically, Rickettsiales were widely considered a sister lineage to the mitochondrial progenitor, with this assemblage representing a basal branch of the *Alphaproteobacteria* (16–20). Pioneering work on rickettsial genomes identified decreased genome size and pseudogenization of genes within many metabolic pathways, processes termed “reductive genome evolution” that coincide with addiction to the eukaryotic cytosol (18, 21–25). This dogma for shared evolutionary history and convergence in genome reduction between mitochondria and Rickettsiales held for two decades while hundreds of diverse Rickettsiales genomes were sequenced (6, 26). However, more recent phylogenetic analysis of deep marine metagenome-assembled genomes (MAGs) illustrated that mitochondria likely originated outside of all described *Alphaproteobacteria* (27). Furthermore, a recent phylogenomic description of certain novel MAGs established two basal rickettsial lineages, families Mitibacteraceae and Athabascaceae, with features indicating an extracellular lifestyle not dependent on eukaryotic hosts (28). These landmark findings bolstered the growing trend for identifying mostly aquatic, protist-associated rickettsial species with traits (e.g., flagella, larger genome size, greater metabolic capacity, etc.) more characteristic of free-living and facultative intracellular bacteria but absent from the numerous genomes of well-characterized invertebrate- and vertebrate-associated rickettsial species (13, 15, 29–33). Importantly, a revised Rickettsiales phylogenetic framework now allows for assessing the evolutionary trajectories within five later-evolving rickettsial families for innovations that emerged from transitions to host dependency (34).

Estimated to have arisen ~1.9 billion years ago (35), *Alphaproteobacteria* are highly diversified in form and function yet rife with convergence in morphology and lifestyle through common adaptation to numerous environments, including eukaryotic cells (36). *Alphaproteobacteria* have few Order-level signatures, yet the Rickettsiales *rvh* homolog (*rvh*) type IV secretion system (T4SS) is a *bona fide* rickettsial signature that functions in colonizing eukaryotic cells (37–41). The *rvh* T4SS is odd in its design (42, 43), with specialized duplications of some components hypothesized to autoregulate effector secretion (44, 45). Effectors have been experimentally characterized for species of *Ehrlichia* (46–49), *Anaplasma* (50–55), and *Rickettsia* (56–58). As the *rvh* T4SS is present in purportedly free-living Mitibacteraceae and Athabascaceae, Schön et al. (28) proposed that these rickettsiae utilize the *rvh* T4SS for killing congener microbes, provided their genomes harbor candidate *rvh* effectors with characteristics similar to effectors in other T4SS and type VI secretion system (T6SS) killing machines (59, 60). Thus, the five later-evolving families likely repurposed the *rvh* T4SS to secrete effectors that commandeer host cellular processes to support intracellular replication (or epicellular parasitism in the case of “*Candidatus Deianiraea vastatrix*” and likely other Deianiraeaceae species).

The existence of an ancient secretion machine (*rvh*), yet independent gain of its effectors later in evolution, prompted us to poll the ever-growing MAG diversity for clues on *rvh* effector origins. We focus on known or candidate effectors from the genus *Rickettsia*, as recent studies have considerably expanded Rickettsiaceae diversity. Genome sequences from “environmental” Rickettsiaceae species (i.e., those from protists, apicomplexans, diplomonads, crustaceans, and insects) have illuminated basal lineages of Rickettsiaceae that are critical for inferring the emergence of genomic traits in *Orientia*

and *Rickettsia* pathogens (29, 61–65). Furthermore, phylogenetic analysis of genome sequences from novel genera “*Candidatus Sarmatiella*” (paramecium symbiont) (66) and “*Candidatus Megaira*” (symbionts of algae and ciliates) (31, 67) indicates that *Orientia* and *Rickettsia* species are more divergent than previously appreciated. Finally, a long-standing recognized basal lineage of *Rickettsia* termed “Torix Group,” which is highly diverse and widely present in non-blood-feeding arthropods (68–72), was recently classified as a new genus, “*Candidatus Tisiphia*,” in a study that identified many new provisional *Rickettsia* (and *Tisiphia*) species from MAG analyses of diverse arthropods (73).

We present phylogenomic and other *in silico* analyses that effectively demonstrate the utility of MAG data for not only inferring the origins of pathogen effectors but also for better understanding effector architectures (i.e., protein structure, domain composition, and organization) through enhanced predictive power from greater sequence diversity. Provided that many MAGs come from environmental sampling or eukaryotic microbes with no known human association, our approach stands to inform on the evolution of vertebrate pathogenesis not only for Rickettsiales but also any bacterial taxon wherein human pathogens evolved from non-pathogenic relatives.

RESULTS AND DISCUSSION

Mapping the acquisition of rickettsial effectors

We hypothesize that the transition to an intracellular lifestyle necessitated the acquisition of a more diverse effector repertoire. Thus, to gain an appreciation of the origins and conservation of *Rickettsia rvh* T4SS effectors, we performed a phylogenomics analysis encompassing the newly appreciated rickettsial diversity (Fig. 1A). This initially involved creating a matrix of taxa (depicted by genomes and metagenomes) determined to encode the *rvh* T4SS (Fig. S1) and the distribution of effectors. Six *rvh* effector molecules (REMs: RalF, Pat1, Pat2, Risk1, RT0527, and RARP-2) and 14 candidate REMs (cREMs) were evaluated based on prior studies implicating their secretion and/or interaction with the *rvh* coupling protein (RvhD4) or presence of motifs known to target either congener bacteria or eukaryotic molecules (56–58, 74–81) (Fig. 1B). Our analyses added complexity for two REMs (Pat1 and RARP-2) and four cREMs based on the identification of duplications (cREM-1, cREM-2, and cREM-4), a partner protein (cREM-5), and a domain within the surface cell antigen (*sca*) Sca4 that we demonstrate to be widespread in non-Sca4 proteins (discussed further below). Collectively, a total of 26 proteins were analyzed within the phylogenomic framework (Fig. 1C and 2). A phylogeny was estimated from concatenated alignments of RvhB4-I and RvhB4-II proteins from 153 genome assemblies (Fig. 1C and 2 see Table S1 for sequence information; see Fig. S2 for entire phylogeny and related information). This collective matrix is an effective framework for mapping the earliest occurrence of these *rvh* effectors in the rickettsial tree, additionally identifying several likely origins for lateral gene transfer (LGT).

Origins of REMs

An emerging diversity of bacterial Arf-GEFs

Bacterial mimicry of eukaryotic-like Sec7 domains (S7D) to function as guanine nucleotide exchange factors (GEFs) for host ADP-ribosylation factors (Arfs) was first described for *Legionella pneumophila*, which utilizes the *dot/icm* T4SS effector RalF to recruit and activate host Arf1 to the *Legionella*-containing vacuole (LCV) (84, 85). Certain *Rickettsia* genomes encode RalF proteins that are remarkably similar to *Legionella* counterparts across the S7D, as well as a Sec7 capping domain (SCD) that restricts access to the catalytic site (86–88). The SCD has high specificity for host membranes and differentially regulates effector subcellular localization for *Legionella* (the LCV) and *Rickettsia* (cytosolic leaflet of plasma membrane) RalF (89). *Rickettsia* RalF was the first characterized REM; its secretion during host cell invasion activates host Arf6 at the plasma membrane, a process driven by a unique C-terminal extension, termed variable with Pro-rich region (VPR), which interacts with host actin and phosphatidylinositol 4,5-bisphosphate at entry

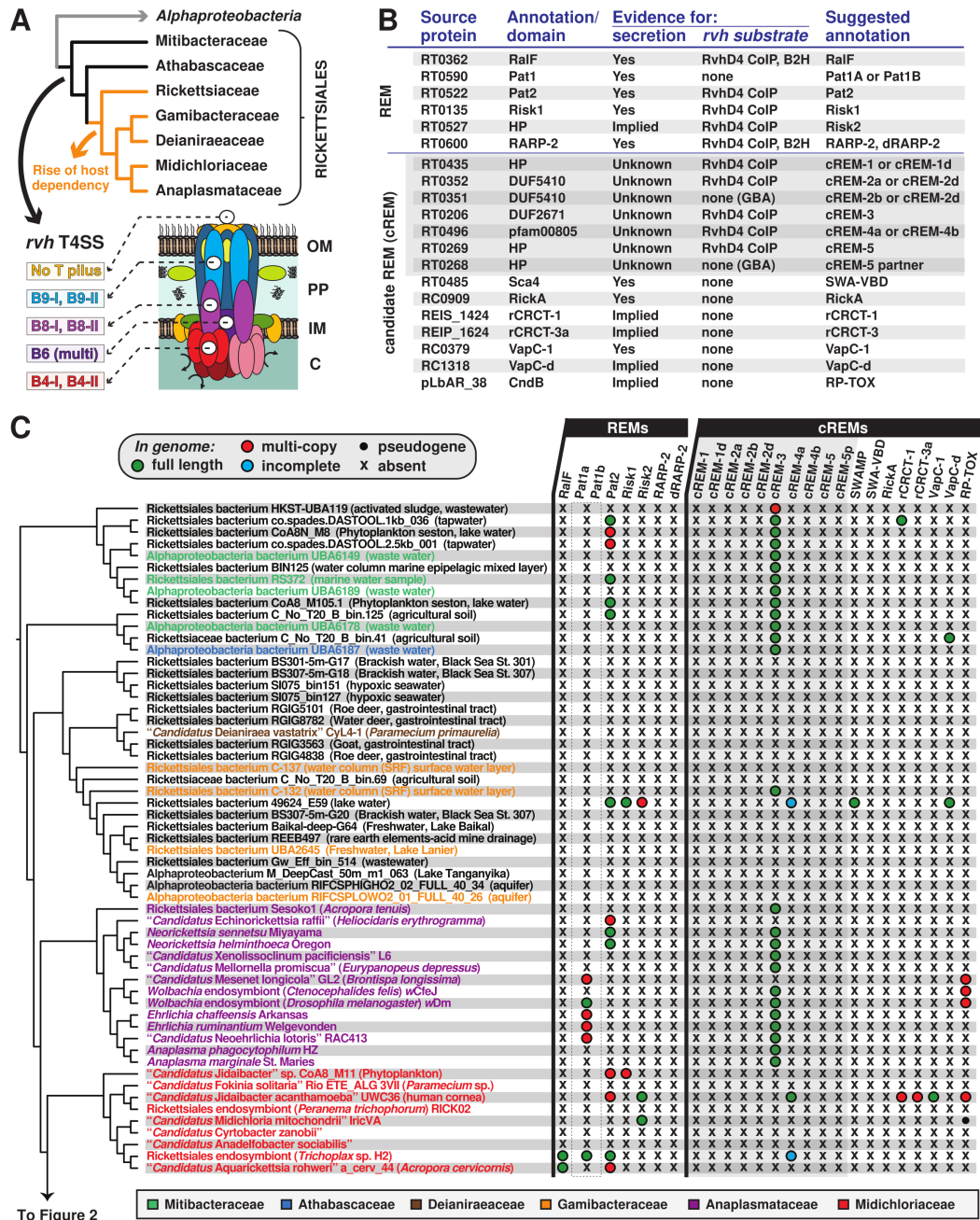


FIG 1 Probing Rickettsiales diversity for the evolution of *Rickettsia* type IV secretion system effectors. (A) The atypical Rickettsiales *vir* homolog (*rvh*) T4SS is a hallmark of Rickettsiales that was present before the origin of host dependency (orange) (28, 34). Schema depicts recent genome-based phylogeny estimation (28). Rvh characteristics (38, 45) are described at the bottom and further in Fig. S1. (B) List of *Rickettsia* *rvh* effector molecules and candidate REMs. GBA, guilty by association (meaning a tandem gene with sequence similarities has experimental support for encoding a REM). “Implied” means analogous proteins are known to be secreted by other bacteria and/or the effector has strongly predicted host cell targets. Secretion, coimmunoprecipitation (CoIP), and bacterial two-hybrid (B2H) data are compiled from prior reports (56–58, 74–81). SWA, Schuenke-Walker antigen domain. (C) Phylogenomic analysis of *Rickettsia* REMs and cREMs in non-Rickettsiaceae lineages. The cladogram summarizes a phylogeny estimated from concatenated alignments for RvhB4-I and RvhB4-II proteins from 153 rickettsial assemblies (full tree, Fig. S2; sequence information, Table S1). Non-Rickettsiaceae lineages are shown (see Fig. 2 for Rickettsiaceae). The dashed box for Pat1 proteins indicates the inability to confidently discern Pat1a and Pat1b homology outside of *Tisiphia* and *Rickettsia* species (see Fig. 2 and 4). SWAMP, SWA modular proteins. Information for all REMs and cREMs is provided in Table S2.

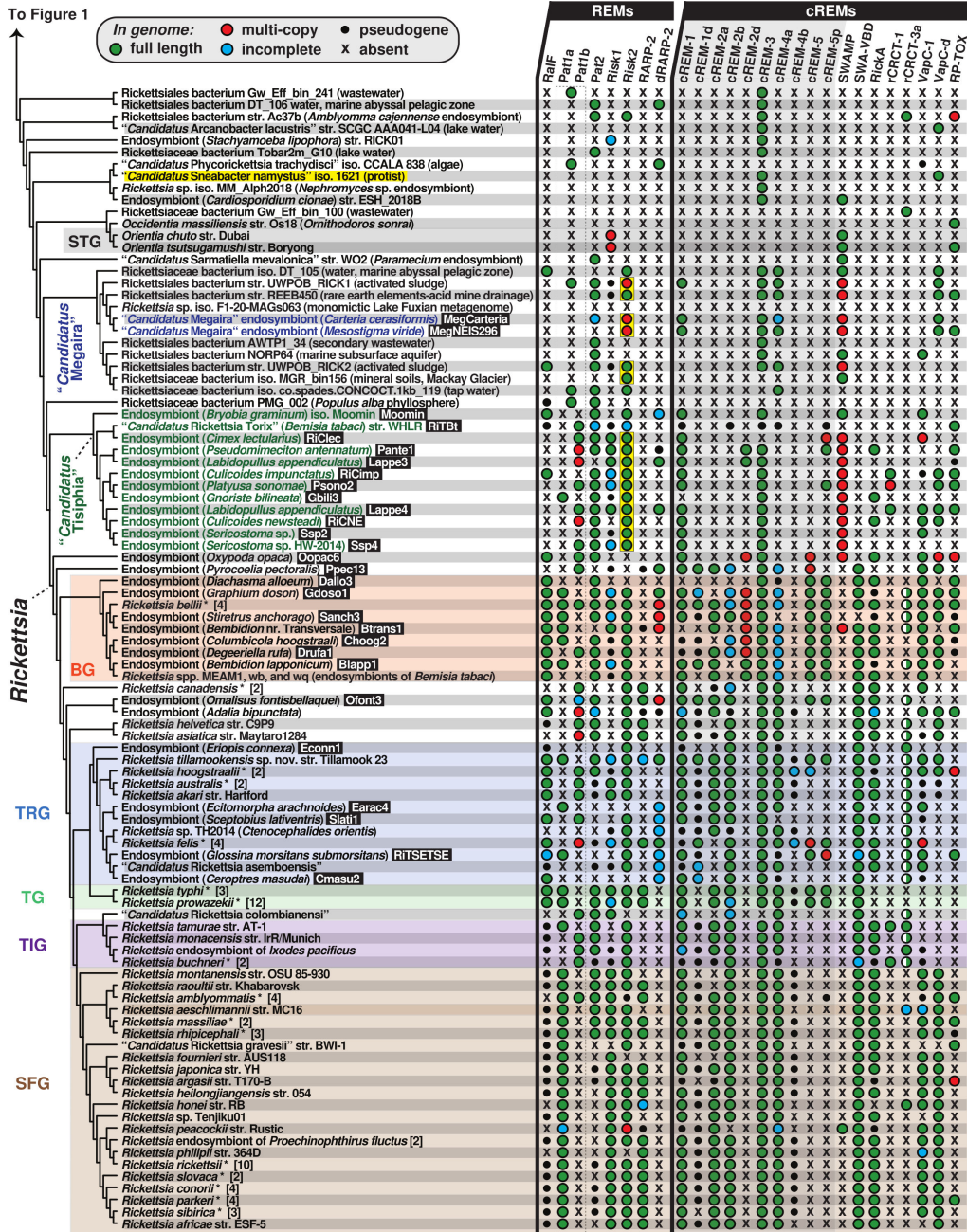


FIG 2 Phylogenomic analysis of *Rickettsia* REMs and cREMs in Rickettsiaceae. Cladogram (continued from Fig. 1C) summarizes a phylogeny estimated from concatenated alignments for RvhB4-I and RvhB4-II proteins from 153 rickettsial assemblies (full tree, Fig. S2; sequence information, Table S1). "*Candidatus Sneabacter namustus*" (highlighted yellow) was manually added to the cladogram based on prior phylogeny estimation (82, 83) as this species lacks *rvh* genes but carries a T6SS (see Fig. S3). Black boxes provide short names for 29 MAGs from Davison et al. (73) (Note: the clade colored green comprises genus *Tisiphia* though genus name *Rickettsia* reflects NCBI taxonomy as of 26 February 2023). Asterisks depict multiple genome assemblies for a species. BG, Bellii Group; TRG, Transitional Group; TG, Typhus Group; TIG, Tamurae-Ixodes Group; and SFG, Spotted Fever Group. The dashed box for Pat1 proteins indicates the inability to discern Pat1a and Pat1b homology outside of *Tisiphia* and *Rickettsia* species (see Fig. 4). Yellow boxes denote Risk2 proteins that are appended to C-terminal Schuenke-Walker antigen (SWA) domains (see Fig. 5). SWAMP, SWA modular proteins; all other REMs and cREMs are described in Fig. 1B and Table S2). Half circles for rCRCT-3a depict the presence of one or more antidotes but no toxin.

foci (56, 90). The presence of *ralF* in the genomes of some *Rickettsia* pathogens but its absence in non-pathogenic species led to speculation that this REM may be a lineage-specific virulence factor (56, 90, 91). Furthermore, while species of *Rickettsia* and *Legionella* exchange genes in common intracellular environments (92, 93), the absence of *ralF* in any other known bacteria precluded insight into the origin of RalF and specifically the nature of *Legionella*- and *Rickettsia*-specific C-terminal architectures.

Our analyses provide clarity on RalF evolution by unearthing numerous bacterial analogs with novel S7D-containing architectures (Fig. 3; Fig. S4). First, an unusual *Legionella* RalF from *Legionella clemsonensis* was found to carry a conserved domain at its C-terminus that was also detected in a large ankyrin (ANK) repeat-containing protein of the *Rickettsia* endosymbiont of *Graphium doson* (Gdoso1) genome (Fig. 3A; Fig. S4A and B). This Gdoso1 protein also contains another conserved domain at its C-terminus that is widespread in *Rickettsia* genomes but lacks any associated annotation in public databases. These observations indicate frequent recombination in conjunction with the LGT of these diverse genes. Second, while the most basal *Rickettsia* species (endosymbiont of *Oxypoda opaca*, “Oopac6”) harbors a RalF with the *Rickettsia*-like C-terminal VPR, two *Tisiphia* species (endosymbionts of *Bryobia graminum* or “Moomin,” and *Culicoides impunctatus* or “RiCimp”) instead exhibit *Legionella*-like C-terminal domains (Fig. 3A and C; Fig. S4C through E). RiCimp RalF is encoded on a plasmid (pRiCimp001), which is unique among all other known RalF genes (Fig. 3D), supporting original speculation for RalF exchange between *Legionella* and *Rickettsia* species (84). Remarkably, pRiCimp001 also carries a toxin-antidote (TA) module highly similar to the plasmid-encoded TA module of *Rickettsia felis* str. LSU-Lb (94) (Fig. 3D), which we previously described as part of the mobilome shuttling reproductive parasitism (RP) genes across *Rickettsia* and *Wolbachia* (94) (discussed in section cREMs with characterized function).

Finally, RalF proteins from three additional rickettsial species and two putative gammaproteobacterial species carry both the S7D and SCD but no C-terminal extensions (Fig. 3B; Fig. S4E and F). Four other novel S7D-containing proteins from non-rickettsial bacteria lack SCDs; however, two contain eukaryotic domains found in RHO GTPase-activating proteins (RHOGAP) that also target Arfs (Fig. 3B and C; Fig. S4F). All these discovered proteins have a highly conserved S7D and SCD (if present) and include most of the structural features that define RalF proteins (Fig. S4G). These collective characteristics attest to LGT disseminating the S7D-SCD architecture across divergent bacteria, with recurrent gains of additional domains tailored to eukaryotic cell functions (e.g., VPR, ANK, and RHOGAP). Our phylogenomics results indicate the acquisition of the *Rickettsia*-unique C-terminal VPR occurred early in *Rickettsia* evolution after divergence from *Tisiphia* spp., with multiple losses of RalF in more than half of the sequenced species (Fig. 2).

Patatins: divergent phospholipases are recurrent in Rickettsiales

Rickettsiae (and other Rickettsiales species exiting the phagosome and/or lysing host cells) require membranolytic effectors throughout the intracellular lifestyle. Phospholipase D (PLD) is a highly conserved enzyme with demonstrated membranolytic activity in a surrogate expression system (97), though its function during *Rickettsia* infection of host cells remains unresolved (98). PLD contains a N-terminal Sec signal (91), yet other phospholipase A₂ (PLA₂) enzymes (patatins Pat1 and Pat2) have sequence characteristics of *rvh* substrates (74), and Pat2 binds RvhD4 in coimmunoprecipitation assays (57) (Fig. 1B). Studies on *R. typhi* have shown that both Pat1 and Pat2 are secreted during host cell infection, require host cofactors for activation, and function early in infection by facilitating phagosome escape (74, 75). Recent work on *Rickettsia parkeri* Pat1 also demonstrated a role in phagosome escape in addition to facilitating avoidance of host polyubiquitination and autophagosome maturation, as well as promotion of actin-based motility and intercellular spread (99). *R. parkeri* lacks Pat2, which is slightly more restricted in *Rickettsia* genomes and possibly provides a function in host cell lysis for rickettsiae that do not spread intercellularly without host cell lysis (e.g., TG rickettsiae).

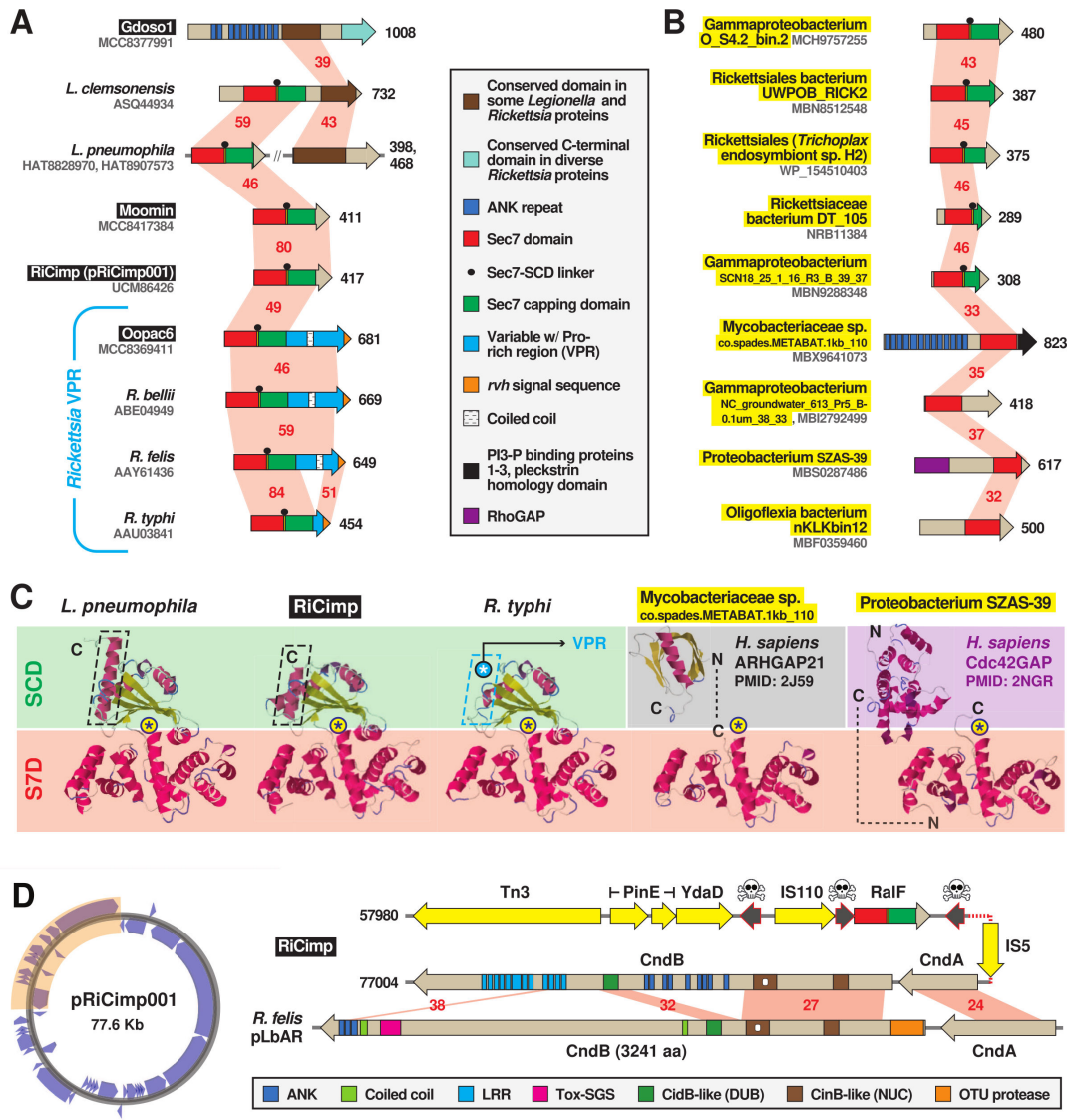


FIG 3 MAG analysis divulges novel diversity of bacterial Sec7-domain-containing proteins. Black boxes provide short names for MAGs from Davison et al. (73). These and additional newly discovered RalF-like proteins (highlighted yellow) substantially expand the prior recognized RalF diversity (56, 84, 88, 90). Structural models for proteins are found in Fig. S4D through F. (A and B) Insight from (A) novel *Legionella* and rickettsial architectures and (B) diverse RalF-like proteins discovered in MAGs. Red shading and numbers indicate percent aa identity across pairwise alignments (sequence information in Table S2). All protein domains are described in the gray inset. (C) Comparison of the *Legionella pneumophila* RalF structure (PDB 4C7P) (88) with predicted structures of S7D-SCD regions of RiCimp RalF (LF885_07310) and *Rickettsia typhi* RalF (RT0362), and S7Ds of Mycobacteriaceae sp. co.spades.METABAT.1kb_110 (K2 × 97_15435) and Proteobacterium SZAS-39 (JSR17_09325). The delineation of the Sec7 domain (S7D, red) and Sec7-capping domain (SCD, green if present) is shown with an approximation of the active site Glu (asterisk). Additional eukaryotic-like domains for the non-rickettsial proteins are noted. Modeling was done with Phyre2 (95). More detailed structural explanation can be found in Fig. S4C. (D) RiCimp plasmid pRiCimp001 carries RalF and a CindB/A toxin-antidote module similar to those characterized or implicated in reproductive parasitism (94). Gene region drawn to scale using the PATRIC compare region viewer tool (96). Yellow, transposases and other mobile elements; skull-and-crossbones, pseudogenes; other domains are described in the gray inset at the bottom. Plasmid map created with Proksee (<https://proksee.ca/>).

All patatins share a common active site architecture that is critical for PLA₂ activity (Fig. 4A). Despite this, Pat1 and Pat2 are highly divergent outside of the patatin domain and have different origins based on phylogeny estimation (74). Furthermore, Pat1 proteins form two distinct groups, Pat1A and Pat1B, with *pat1B* found on plasmids and often recombining with chromosomal *pat1* loci (74). Utilizing newly discovered rickettsial patatins from MAGs, we show that all three enzymes (Pat1A, Pat1B, and Pat2) have distinct sequence profiles, with Pat1B proteins having a high length variable C-terminal

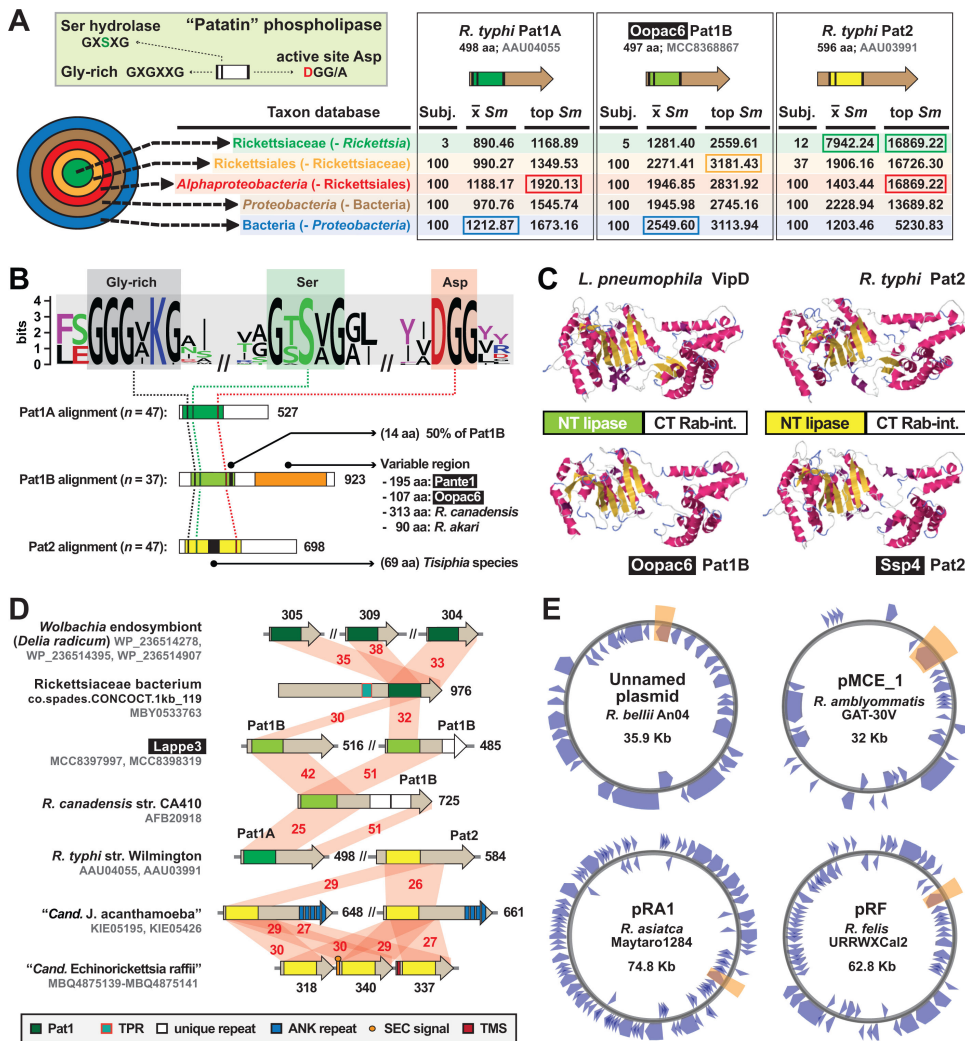


FIG 4 Divergent patatin phospholipases are recurrent in rickettsial evolution. Black boxes provide short names for MAGs from Davison et al. (73). (A) PLA₂ active site characteristics and divergent patatin forms. Green inset describes general patatin domain and active site architecture. HaloBlast results for Pat1A, Pat1B, and Pat2 (query sequences described at the top) are shown, with top-scoring halos boxed (full results in Table S4). (B) Sequence logo (103) showing conservation of the PLA₂ active site motifs across *Tisiphia* and *Rickettsia* patatins (sequence information provided in Table S2). Pat1A, Pat1B, and Pat2 sequences were aligned separately with MUSCLE (104) (default parameters) with active site motifs compiled for conservation assessment. Features unique to each patatin are noted. (C) *Legionella pneumophila* VipD structure (PDBID: 4AKF) and modeling of three rickettsial patatins to VipD using Phyre2 (95). (D) Diverse architectures for select patatins. Red shading and numbers indicate percent aa identity across pairwise alignments (sequence information in Table S2). All protein domains are described in the gray inset. Dark green indicates Pat1 domains not grouped into A or B. (E) Four *Rickettsia* plasmids carry *pat1B* (shaded orange). Plasmid maps were created with Proksee (<https://proksee.ca/>).

region relative to Pat1A and Pat2 enzymes (Fig. 4A and B). Despite this, Pat1B and Pat2 proteins have cryptic similarity across their C-terminal regions to support robust modeling to the crystal structure of *L. pneumophila dot/icm* T4SS effector VipD (100) (Fig. 4C). During *L. pneumophila* host cell infection, secreted VipD localizes to host endosomes, catalyzing the removal of phosphatidylinositol 3-phosphate from endosomal membranes (N-terminal patatin domain) and binding Rab5 or Rab25 (C-terminal domain), ultimately blocking endosome-LCV fusion (100–102). As with RalF, it is likely that *Rickettsia* Pat1 and Pat2 proteins have rudimentary analogous functions to VipD (targeting host membranes and binding host Rabs), but spatiotemporal and biochemical

differences provided that rickettsiae lyse the phagosome and seemingly do not engage early endosome trafficking on par with *Legionella* species.

We detected Pat1 and Pat2 proteins in several non-Rickettsiaceae genomes (Fig. 1C), with some genomes [e.g., novel sea urchin and cabbage root fly endosymbionts in the Anaplasmataceae (105, 106)] harboring duplications (Fig. 4D). Pat1 proteins from Rickettsiales species outside of the genera *Tisiphia* and *Rickettsia* could not be confidently assigned to either Pat1A or Pat1B (Fig. 1C and 2; dark green domains in Fig. 4D). Most species of *Tisiphia* and *Rickettsia* carry either Pat1A or Pat1B and/or Pat2 (Fig. 2). The only two species carrying all three distinct enzymes (*Rickettsia bellii* and *Rickettsia amblyommatis*) have Pat1B encoded on a plasmid (Fig. 4E). Overall, the patchwork distribution of these divergent enzymes, evidence for modular domain diversification, and presence on plasmids indicate that PLA₂ activities for rickettsiae are lineage-specific and subject to continual patatin gene gain and loss throughout evolution. Furthermore, certain *pat* gene profiles may confer advantages in particular hosts.

Domain repurposing is risky business

Bacterial pathogens can directly modify host membrane phosphatidylinositol (PI) composition by mimicking eukaryotic kinases, phosphatases, and phosphotransferases (107–110). Secreted PI kinases from intracellular pathogens *R. typhi* (Risk1), *L. pneumophila* (LegA and LepB), and *Francisella tularensis* (OpiA) alter the PI composition on phagosomes to prohibit maturation and fusion with lysosomes (57, 111, 112). Characterized as either PI3 (Risk1, LegA, and OpiA) or PI4 (LepB) kinases, these enzymes possess a similar PI3/4 active site architecture (pfam00454) analogous to eukaryotic PI kinases, as well as certain protein kinases, that function in a myriad of membrane-associated functions, including intracellular signaling and trafficking (113) (Fig. 5A and B). Subverting these host cell functions is highly advantageous to intracellular parasitism; thus, the dearth of identified PI3/4 kinase effectors likely reflects the cryptic nature of the PI3/4 active site within these proteins, which lack similarity outside of the PI3/4 active site domain (57, 112).

BlastP and HMMER (118) analyses using only the Risk1 PI3/4 active site unearthed nearly 300 Rickettsiales proteins, with many genomes having multiple divergent kinases. Further inspection revealed the presence of a second conserved protein harboring the PI3/4 active site, which we named *Rickettsia* intracellular secreted kinase-2 (Risk2) (Fig. 5A). Notably, the *R. typhi* Risk2 protein (RT0527) was captured in the same RvhD4 coimmunoprecipitation assay that identified Risk1 as a REM (57) (Fig. 1B). HaloBlast analyses of full-length Risk1 and Risk2 proteins indicate distinct profiles, with Risk2 sharing low similarity to Legionellales kinases (Fig. 5C). Structural analyses corroborated this result, with limited regions of Risk1 and Risk2 modeling best to structures of *Helicobacter pylori* CtkA (114) and *L. pneumophila* LepB (111), respectively (Fig. 5E and F). Comparison of Risk1 and Risk2 PI3/4 active sites revealed (i) juxtapositioned aromatic residues in their catalytic loops, (ii) the presence of a positively charged residue in the activation loops of most Wortmannin-sensitive kinases (Risk1 and human class 1 and 2 PI3 kinases), and (iii) greater sequence length between catalytic and activation loops in Risk2 proteins (Fig. 5F). Furthermore, only LepB and some rickettsial Risk2 proteins have the catalytic loop Asp replaced by Ser (Fig. 5A). These collective observations indicate two divergent PI3/4 kinases encoded in most rickettsial genomes (Fig. 2), leading us to posit that Risk2 is a PI4 kinase that complements the PI3 kinase activity of Risk1 (57).

We determined a remarkable connection between Risk2 and another *Rickettsia* effector, Sca4, which is highly conserved in *Rickettsia* species and implicated in intercellular spread by reducing mechanotransduction at cell-cell junctions (76, 119). The Sca4 C-terminal region has eukaryotic-like vinculin-binding sites (VBSs) that reduce vinculin- α -catenin interactions, which facilitates neighboring cell engulfment of *Rickettsia*-induced protrusions. The N-terminal region, shown by Schuenke and Walker (120) to elicit anti-rickettsia antibodies (Pfam: 120_Rick_ant).

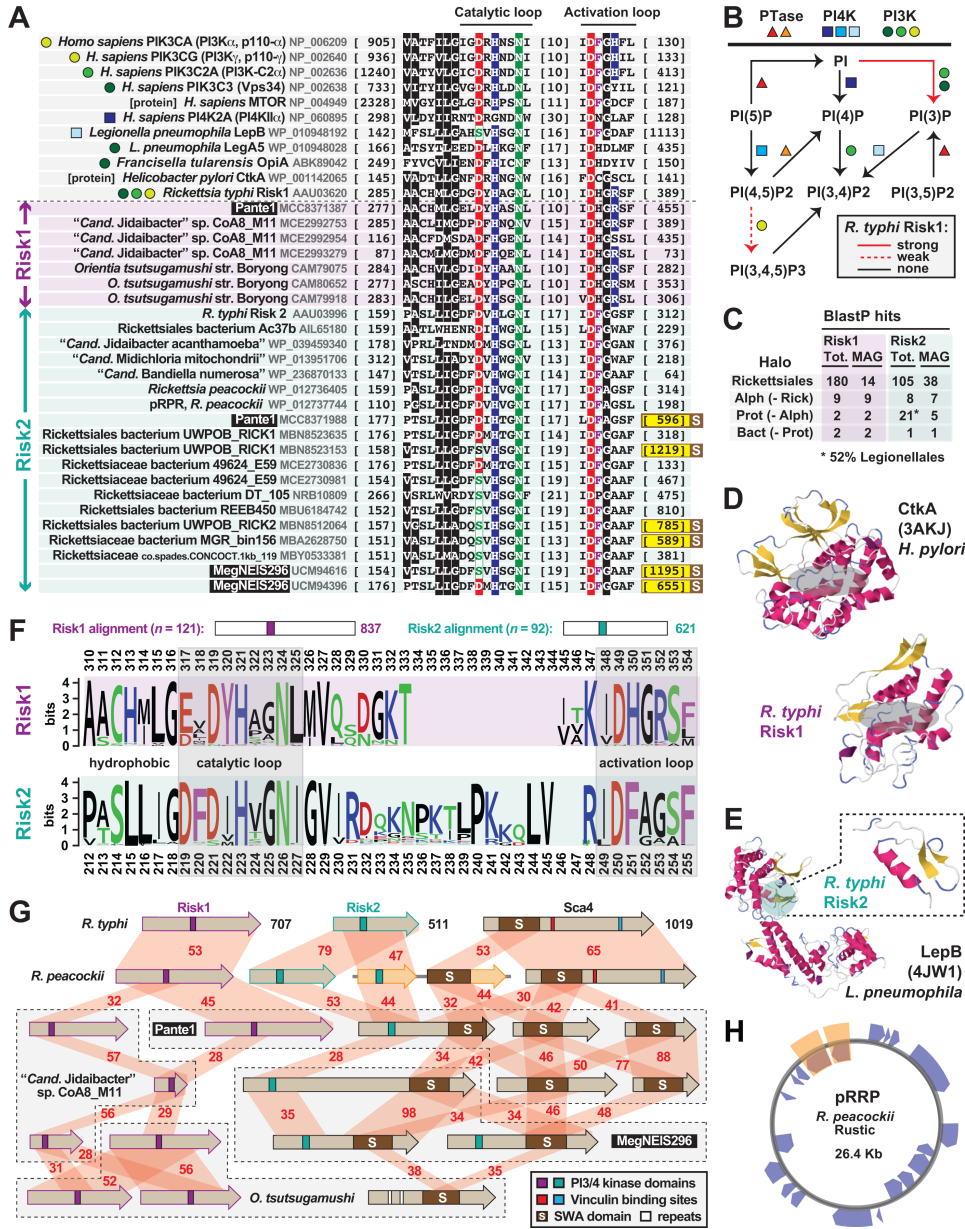


FIG 5 Discovery of a novel cryptic rickettsial PI kinase exposes a widespread rickettsial surface antigen. Black boxes provide short names for MAGs from Davison et al. (73). Amino acid coloring is described in Fig. 3 legend. (A) Previous work (above dashed line) identified a *Rickettsia* PI kinase (Risk1) with a cryptic active site like human and other bacterial PI3/PI4 kinases, as well as related protein kinases. Colored shapes depict characterized substrate specificity (see panel B). Our study (below dashed line; select proteins shown) identified new rickettsial Risk1 proteins, as well as a second distinct PI kinase (Risk2) also prevalent in rickettsial genomes and MAGs (Fig. 1C and 2). All PI3/PI4 kinase domains were aligned using MUSCLE (104) (default parameters). Sequence information is provided in Table S2. Yellow highlighting on end coordinates denotes Risk2 proteins fused to a C-terminal SWA domain (see panel G and Fig. S5 for a full description of the SWA domain). (B) Mechanisms of phosphorylation on the PI inositol ring at 3', 4', and 5' positions. Data for *R. typhi* Risk1 are shown by red arrows (57). (C) HaloBlast results (*R. typhi* Risk1 and Risk2 as queries) broken down to illustrate the presence of *Rickettsia*-like PI kinases in MAGs and the similarity between Risk2 and *Legionella* PI kinases (full data in Table S2). (D) Risk1 threads with high confidence [90.7%, 72% coverage; Phyre2 (95)] to the *Helicobacter pylori* proinflammatory kinase CtkA (114). (E) Risk2 threads with high confidence [85.1%, 9% coverage; Phyre2 (95)] to a limited region of LepB, a Rab GTPase-activating protein effector from *L. pneumophila* (115). (F) Risk1 and Risk2 proteins have cryptic and distinct PI3/4 active sites yet lack similarity outside of these regions. Logos depict individual alignments, which are summarized at the top and were performed with MUSCLE (104), (Continued on next page)

FIG 5 (Continued)

default settings. (G) Many rickettsiae carry a diverse arsenal of PIK effectors, some of which are tethered to SWA domains. Six select species are shown with their full complement of PI3/4 kinase and SWA architectures. Red shading and numbers indicate percent aa identity across pairwise alignments (sequence information in Table S2). *Rickettsia peacockii* pRRP proteins are shaded orange (see panel H). (H) Plasmid pRRP of *R. peacockii* str. Rustic (116, 117) carries a divergent Risk2 gene that is adjacent to an ORF encoding a SWAMP (orange highlighting). Plasmid map created with Proksee (<https://proksee.ca/>).

From many forms, one descendent

Early in host infection, pathogens *R. typhi* and *R. rickettsii* secrete the REM RARP-2, which traffics to the endoplasmic reticulum and Golgi apparatus, leading to trans-Golgi network (TGN) fragmentation and ultimately perturbed protein transport to the host cell surface (58, 121). Like RalF, RARP-2 has a C-terminal tail that binds RvhD4 (Fig. 1B); furthermore, the protein has well-delineated N-terminal protease and C-terminal ANK repeat domains (Fig. 6A). The protease domain has minimal analogy to clan CD protease families (C13 legumain, C14 caspase 1, C11 clostripain, and C25 gingipain R), which share a common fold that arranges a His and Cys catalytic dyad (122). This active site is essential for RARP-2 fragmentation of TGN (121) and also contributes to the lytic plaque phenotype of virulent *R. rickettsii* strains (58). The ANK repeat domain is atypical among most ANK repeat-containing proteins (123), as the composition of each repeat is highly similar in length and identity (Fig. 6B; Fig. S6A through C) despite a highly variable repeat number across orthologs, even at the strain level in most cases (Fig. 6C). RARP-2 active site mutants still traffic to perinuclear membranes, indicating that the ANK domain drives subcellular localization. However, shorter repeats (four in attenuated strains Iowa *R. rickettsii*) do not disrupt TGN fragmentation relative to those of pathogenic *R. rickettsii* (121), suggesting a larger ANK domain is required for proper localization.

Probing recently sequenced genomes and MAGs did not reveal RARP-2 sequences in the Bellii Group (BG) rickettsiae or other Rickettsiaceae genomes (Fig. 2), consistent with prior observations that RARP-2 is unique to later-evolving *Rickettsia* lineages (91). Yet, by focusing on the N-terminal protease domain, we discovered 56 divergent RARP-2 (dRARP-2) proteins that possess the clan CD active site architecture (Fig. 6D). These proteins were binned into six groups (Fig. S6D) that have very different ANK repeat domain identities (data not shown); furthermore, several could be modeled to structures of eukaryotic (124, 125) and prokaryotic (126) clan CD members (Fig. 6E). dRARP-2 proteins are predominantly found in BG rickettsiae and *Tisiphia* genomes but likely shuttle in the intracellular mobilome given that one is carried by a Midichloriaceae species [endosymbiont of *Acanthamoeba* str. UWC8 (129)]. Based on the discordant genomic distribution of RARP-2 and dRARP-2 (Fig. 1C and 2) and the strong bias of RARP-2 in vertebrate-associated species, we speculate that RARP-2 and dRARP-2 may be tailored for similar functions related to TGN fragmentation yet well diverged to allow recognition of targets specific to disparate eukaryotic hosts. This is reminiscent of the recent discovery that *R. parkeri* utilizes different factors for apoptosis induction in ticks versus mammals (130).

cREMs with unknown function

For five *R. typhi* hypothetical proteins previously shown to interact with RvhD4 (cREM-1–5; Fig. 1B), MAG analyses provided substantial clarity on the mechanisms of evolution shaping their architectures. Four of these proteins are described below in light of newfound gene fission/fusion and duplication events (cREM-1, cREM-2, and cREM-4), as well as a greater role of conjugative transposons shaping *Rickettsia* evolution (cREM-5). Unexpectedly, the small cREM-3 (~93 aa) was determined to have widespread conservation in Rickettsiales yet also exist in certain other *Proteobacteria* (Fig. 1C and 2). While likely not a REM, our analyses revealed a potential structure associated with this curious protein (Fig. S7C).

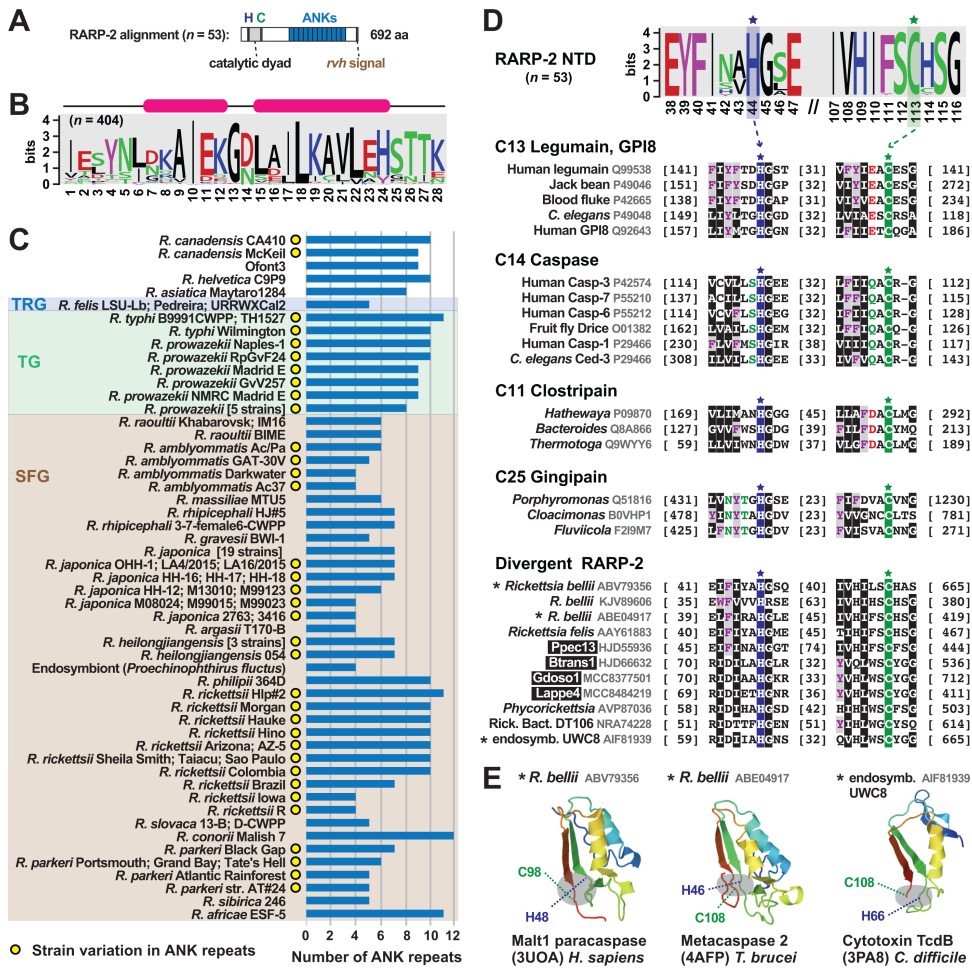


FIG 6 RARP-2 architecture is derived from multiple divergent forms. Black boxes provide short names for MAGs from Davison et al. (73). Amino acid coloring is described in Fig. 3 legend. Sequence logos constructed with WebLogo 3 (103). Sequence information in Table S2. (A) General architecture of RARP-2 proteins deduced from an alignment of 53 non-redundant RARP-2 proteins using MUSCLE (104) (default parameters). (B) Consensus sequence for the RARP-2 ANK repeat deduced from 404 repeats. (C) Depiction of the 53 non-redundant RARP-2 proteins with ANK domain repeat number provided. For brevity, some strain names are not shown for *Rickettsia prowazekii*: Chernikova, Katsinyan, Dachau, BuV67-CWPP, and Rp22; *Rickettsia japonica*: YH, DT-1, HH-1, HH06154, HH07124, HH07167, MZ08014, Nakase, PO-1, Tsuneishi, HH-13, HH06116, HH06125, LON-151, M11012, M14012, M14024, SR1567, and YH_M; *Rickettsia heilongjiangensis*: HCN-13, Sendai-29, and Sendai-58. (D) RARP-2 and dRARP-2 proteins possess N-terminal domain clan CD cysteine protease-like active sites (122). Sequences were manually aligned to illustrate the conservation across all diverse protein groups. “Rick. endo. UWC8,” endosymbiont of *Acanthamoeba* str. UWC8 (not shown in Fig. 1C but closely related to endosymbiont of *Acanthamoeba* str. UWC36 in the Midichloriaceae). (E) Insight on RARP-2/dRARP-2 structure. Asterisks indicate proteins from panel D that were used in Phyre2 (95) searches to identify template structures for modeling (124–126). A complete structure of *R. typhi* RARP-2 predicted with AlphaFold (127, 128) corroborates these dRARP-2 models and indicates deviations on a common effector architecture (Fig. S6C).

Cryptic gene fission and duplication obscured by rapid divergence

For cREM-1 and cREM-2, we utilized phylogeny estimation in conjunction with sequence analysis to predict gene fission (cREM-1) and duplication (cREM-2) events behind the evolution of these proteins. Neither protein was found outside of Rickettsiaceae (Fig. 1C and 2). cREM-1 proteins are streamlined from larger modular *Tisiphia* proteins that harbor the entire cREM-1 sequence as a domain; accordingly, we named these divergent cREM-1 (cREM-1d) (Fig. 7A). Some rickettsiae carry cREM-1 tandem duplications, though

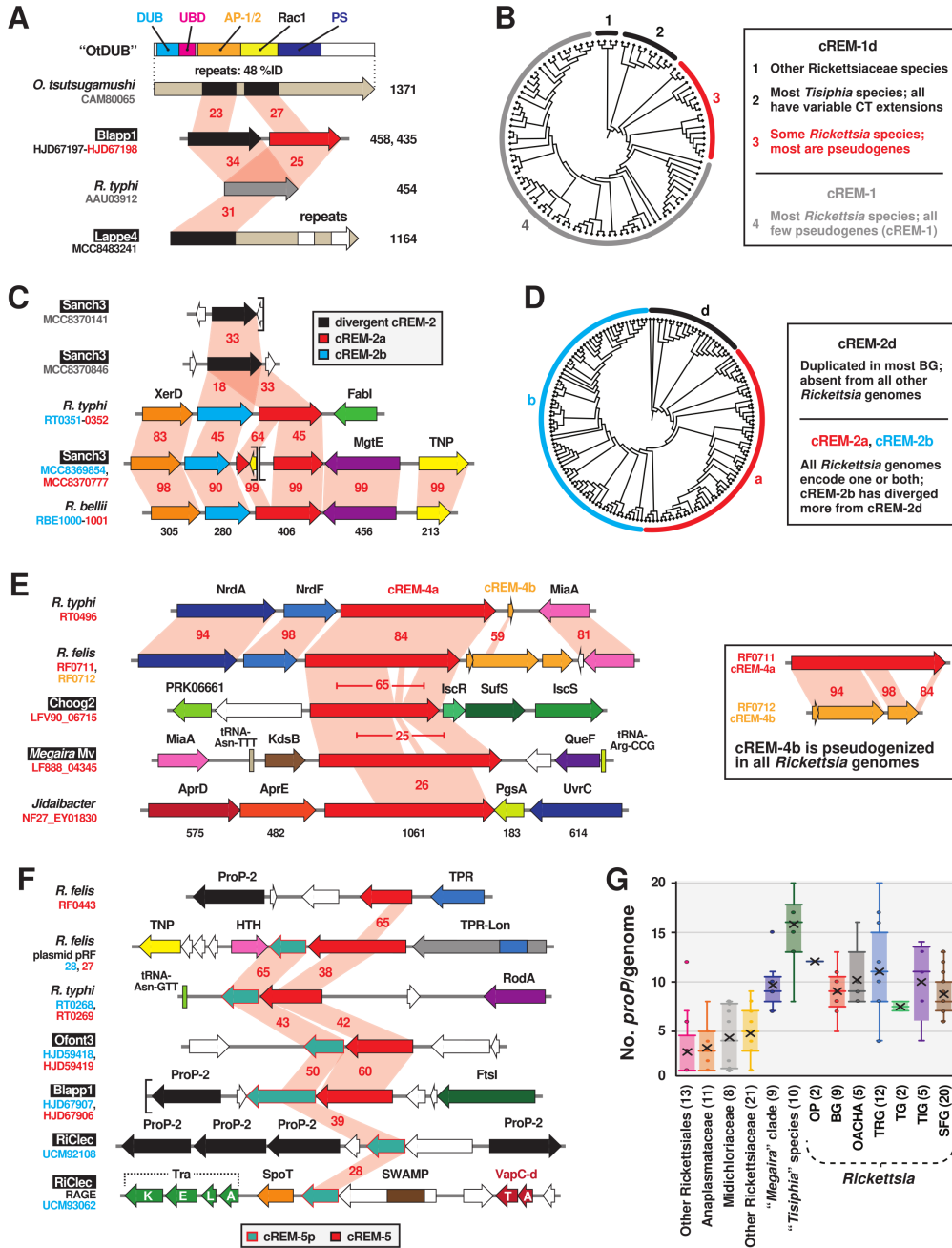


FIG 7 Four candidate REMs are characterized by gene fission and duplication. Black boxes provide short names for 29 MAGs from Davison et al. (73). Gene regions were drawn to scale using the PATRIC compare region viewer tool (96). Sequence information is provided in Table S2. (A) Similarity between *O. tsutsugamushi* effector OtDUB (CAM80065), divergent cREM-1 (cREM-1d), and cREM-1 proteins. OtDUB characterized domains: deubiquitinase (light blue), ubiquitin-binding (pink), cryptic Rac 1-like guanine nucleotide exchange factor (yellow), clathrin adaptor-protein complexes AP-1 and AP-2 (orange), and phosphatidyserine-binding (gray) (5, 131, 134). Red shading and numbers indicate percent aa identity across pairwise alignments. (B) Cladogram depicting phylogeny estimation of 102 cREM-1d and cREM-1 proteins (see Fig. S7A for phylogram and methods). Inset describes cREM-1 proteins, with clade colors matching the protein colors in the schema in panel A. (C) Diversification of cREM-2 proteins via duplication. (D) Cladogram depicting phylogeny estimation of 158 cREM-2d, cREM-2a, and cREM-2b proteins (see Fig. S7B for phylogram and methods). Inset describes cREM-2 proteins, with clade colors matching the protein colors in the schema in panel C. (E) Ancient gene duplication of cREM-4 and location of cREM-4 genes in select Rickettsiales species. The cREM-4 pentapeptide repeat domain is illustrated in Fig. S7D. (F) cREM-5/5p loci occur in variable genomic regions, including plasmids and Rickettsiales amplified genetic elements (RAGEs). The complete RAGE for RiClec is (Continued on next page)

FIG 7 (Continued)

illustrated in Fig. S7F. (G) Distribution of ProP genes in Rickettsiales genomes. OP, *Rickettsia* endosymbionts of *Oxyopoda opaca* (Oopac6) and *Pyrocoelia pectoralis* (Ppec13); OACHA, *Rickettsia* endosymbionts of *Omalisus fontisbellaquei* (Ofont3) and *Adalia bipunctata*, *Rickettsia canadensis*, *Rickettsia helvetica*, and *Rickettsia asiatica*.

most genomes have one conserved gene and the second pseudogenized (Fig. S7B; red clade). Curiously, cREM-1 proteins have similarities to a repeat region within the *Orientia tsutsugamushi* effector OtDUB (Fig. 7A; Fig. S7A). This region in OtDUB binds clathrin adaptor-protein complexes AP-1 and AP-2 and harbors a cryptic Rac1 GEF domain (131–133). This indicates that cREM-1 proteins may have evolved from larger modular proteins with functions tailored to the eukaryotic cytosol, with repeat regions of these large effectors streamlining to smaller proteins encoded by tandem genes (Fig. 7B; Fig. S7A).

cREM-2 proteins belong to pfam17422 (DUF5410: specific to *Rickettsia* species). Our analyses identified a second DUF5410-like protein encoded adjacent to cREM-2 proteins in many *Rickettsia* genomes (Fig. 2). Neither of these tandem duplicates (designated cREM-2a and cREM-2b) contain Sec signal sequences or other predictable features (Fig. 7C). Furthermore, some BG rickettsiae and *Tisiphia* species harbor a third divergent cREM-2 (cREM-2d) that is absent from later-evolving *Rickettsia* lineages. With the assumption that cREM-2d is an ancestral form, phylogeny estimation indicates cREM-2b proteins are more divergent than cREM-2a proteins (Fig. 7D), though all three protein architectures share high conservation in dozens of residues within the central region of these proteins (Fig. S7B).

cREM-4 proteins also show evidence of an ancestral duplication (Fig. 7E), though no genomes contain a complete duplicate gene, indicating a consistent pseudogenization event that rapidly followed cREM-4 duplication (Fig. 2). Despite their large size (~950 aa), these proteins contain only one observable feature, a small internal pentapeptide repeat (PR). While widespread in diverse bacterial proteins, PR function is generally unknown, though some bacterial PR-containing proteins can interact with DNA-binding proteins (135) and contribute to virulence (136) (Fig. S7D). cREM-4 proteins are encoded in certain other Rickettsiaceae genomes, and like dRARP-2, a single Midichloriaceae species [*Candidatus Jidaibacter acanthamoeba* (30)] encodes a cREM-4 protein. While cREM-4 of BG rickettsiae lack the PR (Fig. S7D), nearly all later-evolving *Rickettsia* genomes encode a complete cREM-4 protein, indicating retention of a conserved function after an ancestral duplication.

LGT of cREM-5 as a two-gene module across select species

cREM-5 proteins were previously noted for their restricted distribution in TG rickettsiae and *R. felis*, which carry copies on the chromosome and plasmid pRF (137). Our analyses yielded several novel findings. First, while absent from any Spotted Fever Group (SFG) or Tamurae-Ixodes Group (TIG) rickettsiae, cREM-5 proteins are highly conserved in all BG rickettsiae genomes, as well as in a few *Tisiphia* genomes (Fig. 2). Second, most cREM-5 genes have an associated protein, cREM-5 partner (cREM-5p), encoded immediately downstream (Fig. 7F). Despite conserved regions (Fig. S7E), neither protein has detectable domains or similarity to proteins in other Rickettsiales (Fig. 2). Third, cREM-5/5p genes have a strong co-occurrence with PropP-2 genes (black, Fig. 7F). PropP (proline betaine transporters of the major facilitator superfamily) function in osmoregulation (138, 139) are proliferated in *Rickettsia* genomes, with seven conserved groups (PropP1-7) containing species-specific duplications (93, 140). Why specifically PropP-2 genes cluster near certain cREM-5/5p loci is unclear, though insight from MAGs illuminated a previously unrealized point in Rickettsiales evolution where PropP proliferation occurred (Fig. 7G).

Finally, cREM-5 modules are found in recombination hotspots and other less conserved genomic regions, indicating LGT behind their evolution in *Rickettsia* genomes. Aside from cREM-5/5p on plasmid pRF, one copy of cREM-5p from the RiClec (endosymbiont of *Cimex lectularius*) genome is found on a conjugative transposon termed

Rickettsiales amplified genetic elements (RAGEs) (Fig. 7F). RAGEs are integrative and conjugative elements present on certain *Rickettsia* plasmids and chromosomes (93, 94, 141), as well as proliferated and scattershot in *O. tsutsugamushi* genomes (142, 143). Cargo genes, or those piggybacking on RAGEs at indiscriminate insertion sites, have functions mostly related to the stringent response and metabolism, defense and resistance, and adaptation to host cells (e.g., ProP genes are shuttled by RAGE). The addition of cREM-5p and SWAMPs, as well as a myriad of TA modules, to the list of RAGE cargo (see Fig. S7F) indicates that these mobile elements play a role in disseminating pathogenicity factors, which was previously unappreciated.

cREMs with characterized function

Several *Rickettsia* proteins that lack N-terminal Sec signals have either been well characterized for their roles in subverting host cell processes (e.g., Sca4 and RickA) or possess features that implicate them in targeting host molecules (e.g., VapC and other toxins within TA modules). Until secretory pathways for these molecules are characterized, we consider them cREMs (Fig. 1B). MAG analyses of these proteins have generated novel insight into the structure and evolution of domains targeting the host actin cytoskeleton. Furthermore, a greater appreciation for toxin architecture and distribution indicates the *rvh* T4SS may still function in congener killing despite the host-dependent lifestyle of the later-evolved Rickettsiales species.

Insight on *Rickettsia* interactions with host actin

With their RalF proteins lacking VPRs and their SWAMPs lacking VBDs (Sca4), *Tisiphia* species may interact with host cell actin cytoskeleton differently than *Rickettsia* species (Fig. 8A). We analyzed another host actin-associated protein, RickA, which some *Rickettsia* species use for intracellular motility and possibly intercellular spread (144–147). While no association with the *rvh* T4SS has been characterized, RickA localizes to the bacterial surface in the absence of a Sec signal peptide (147). RickA directly activates host Arp2/3 complexes through an architecture that mimics host nucleation-promoting factors (NPF) (78, 148). We discovered several RickA proteins from *Tisiphia* MAGs that differ in their C-terminal architectures relative to SFG rickettsiae RickA proteins characterized in actin polymerization (Fig. 8B; Fig. S8A). The functional relevance of these differences is unclear. Surprisingly, we gained novel insight into the N-terminal structure of all RickA proteins. The substantial increase in diversity from MAGs illuminated a large (~95 aa) repeat region enclosing the G actin-binding domain, with each repeat highly conserved in hydrophobicity and predicted structure (Fig. 8B; Fig. S8B). We envisage that this conserved region may facilitate the docking of G actin to the G actin-binding domain and the overall positioning of the N-terminus to Wiskott-Aldrich syndrome protein homology 2 (WH2) motifs at the C-terminal region.

Despite activation by RickA, the specific role of host Arp2/3 complexes during rickettsial infection is unclear, perhaps due to different species utilized across studies garnering contrasting results (146, 147, 152). Specifically, actin-based motility in certain *Rickettsia* species is carried out by a second NPF, Sca2, that polymerizes actin independent of Arp2/3 complexes (150, 151). The passenger domains of Sca2 mimic eukaryotic formins by elaborating ring-like structures to elongate actin, with intervening Pro-rich regions and WH2 domains incorporating profilin-actin for elongation and recruiting actin monomers for nucleation, respectively (149). Thus, at least for species carrying both RickA and Sca2 (most SFG rickettsiae), RickA-mediated Arp2/3 activation may play a greater role early in infection, possibly for inducing host cell filopodia formation during invasion (79, 151, 153, 154). Still, few *Rickettsia* species outside of SFG rickettsiae encode Sca2 proteins with intact formin-like passenger domains (Fig. 8B), and some of these species lack RickA genes as well (93, 116, 155) (Fig. 2 and 8A). This implies that host actin polymerization for motility is an expendable trait for most *Rickettsia* species, and that

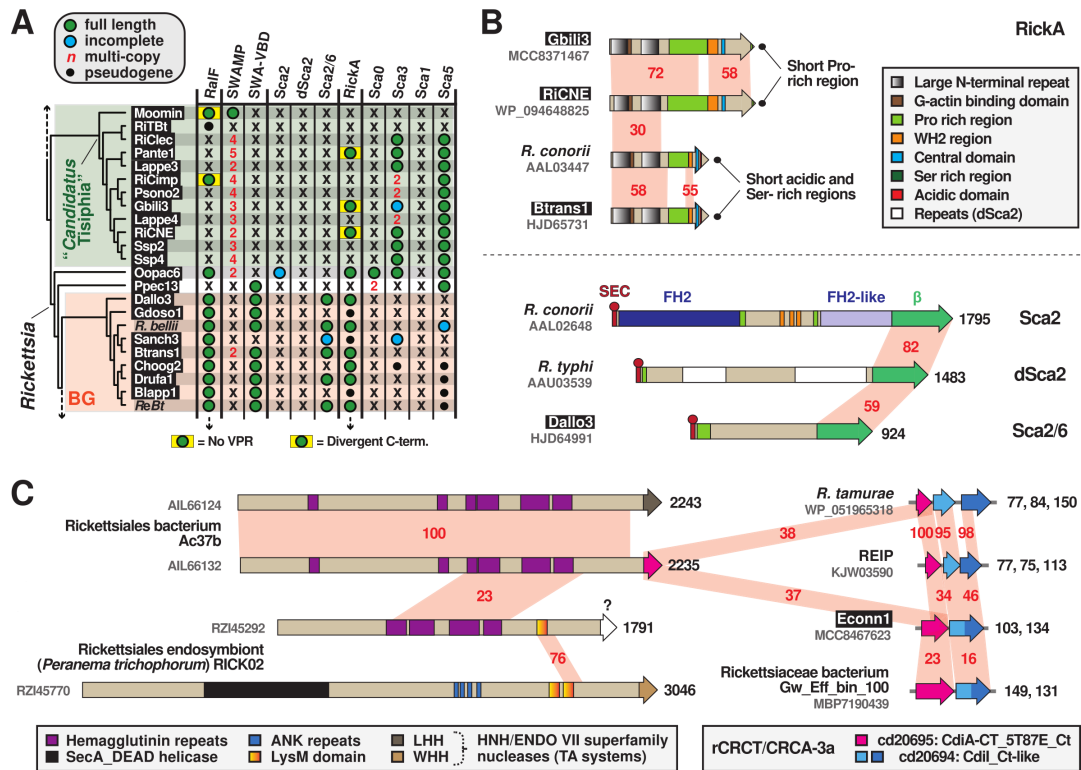


FIG 8 MAG analyses lend insight on *Rickettsia* interactions with host actin cytoskeleton and *rvh* T45S function. Black boxes provide short names for 29 MAGs from Davison et al. (73). (A) MAGs shed light on the evolution of *Rickettsia* factors behind host actin polymerization and invasion. *Tisiphia* and BG rickettsiae taxa, as well as SWAMP, SWA-VBD, RickA, and RalF info, are from Fig. 2. The passenger domains of *Rickettsia conorii* Sca2 (AAL02648), *R. typhi* dSca2 (AAU03539), *R. bellii* Sca2-6 (ABE05361), *R. conorii* Sca0 (AAL03811), *R. typhi* Sca3 (AAU03915), *R. typhi* Sca1 (AAU03504), and *R. typhi* Sca5 (AAU04158) were used in BlastP searches directly against *Tisiphia* and BG rickettsiae genomes. Passenger domains and linker sequences were delineated as previously shown (91). *ReBt*, *Rickettsia* spp. MEAM1, wb, and wq (endosymbionts of *Bemisia tabaci*). (B) Some *Rickettsia* genomes encode one or more host actin nucleation proteins. Top: *Tisiphia* and *Rickettsia* RickA proteins share a large N-terminal repeat domain but diverge at their C-termini. Further details on RickA architecture are provided in Fig. S8. Bottom: Sca2, d-Sca2, and Sca2-6 proteins have a common autotransporter domain (β) but divergent passenger domains. FH2, formin homology 2. Sca2 mimics host formin actin nucleators (149) to recruit and polymerize actin for intracellular motility and intercellular spread (150, 151). The functions of dSca2 and Sca2/6 are unknown. (C) The mobile nature of CDI-like/Rhs-like C-terminal toxin/antidote (CRCT/CRCA) modules across diverse rickettsial genomes. Schema shows the integration of CRCT/CRCA modules into larger polymorphic toxins (hemagglutinin-like toxins, LysM-like peptidoglycan/chitin-targeting toxins, etc.), as well as CRCT/CRCA modules independent of larger toxins. The toxin warhead for RZI45292 is unknown. Further details are provided in Fig. S9.

Arp2/3 recruitment and activation during invasion can be instigated by other bacteria-driven processes, i.e., Arf recruitment to the plasma membrane for inducing PI shifts required for filopodia formation (90).

MAG analyses indicate that, barring acquisitions via LGT, RickA and RalF were likely present before the diversification of major *Rickettsia* lineages, whereas Sca2 appeared later in *Rickettsia* evolution (Fig. 8A). Furthermore, we polled MAGs for the presence of genes encoding the four major autotransporters (Sca0, Sca1, Sca3, and Sca5) with known (or anticipated) functions in host cell binding and/or invasion (156–160). Remarkably, Sca3 is predominant in *Tisiphia* genomes despite a very limited distribution in *Rickettsia* species [restricted mostly to TG and TRG rickettsiae (91)]. Furthermore, BG rickettsiae are counter to most other rickettsiae in lacking both Sca0 and Sca5, the dominant proteins of the characterized *Rickettsia* S-layer (161). Collectively, these analyses show that *Rickettsia* factors described in host cell invasion and actin cytoskeleton subversion are sporadically encoded across genomes, indicating host specificity and/or expendability in their contribution to the intracellular lifestyle.

A repurposed or multi-purposed *rvh* T4SS?

Aside from secreting effectors that target host cellular processes, evidence is mounting for intracellular bacteria utilizing large contact-dependent growth inhibition (CDI) and recombination hotspot (Rhs) toxins for interbacterial antagonism (162, 163). We recently identified a few rickettsial genomes encoding specialized TA modules that some bacteria integrate into CDI and Rhs toxins to expand toxic activities (164–167). Widespread in bacteria, these CDI-like/Rhs-like C-terminal toxin and antidote (CRCT/A) modules are extremely polymorphic, variable at the species- and strain-levels, and found either associated with larger toxins or alone as small TA modules (168). The two types of *Rickettsia* CRCT/A (rCRCT/A) modules we identified, rCRCT/A-1 and rCRCT/A-3a, were once associated with large Rhs-like toxins that have mostly degraded (81). rCRCT/A-1 modules are highly divergent from other characterized CRCT/A modules and predominantly occur in Actinomycetia and Cyanobacteria genomes. Only two *Rickettsia* species, *Rickettsia tamurae* and *Rickettsia buchneri*, harbor rCRCT/A-1 modules; however, the “*Cand. J. acanthamoeba*” genome encodes one as an independent module and one integrated into a large modular hemagglutinin toxin with nuclease and peptidase domains. MAG analyses herein discovered several more rCRCT/A-1 modules mostly in *Tisiphia* genomes associated with pseudogenized hemagglutinin-like toxins (Fig. 1C and 2).

rCRCT/A-3a modules resemble the prototype CDI TA module (CdiA-CT/CdiI), wherein the nuclease CdiA-CT targets tRNAs in recipient cytosol (164). CdiA-CT/CdiI is associated with a large modular protein (CdiA) that joins with a second protein (CdiB) as a type Vb secretion system to deliver the toxin into neighboring bacteria (166, 168, 169). However, rCRCT/A-3a modules (and all Rickettsiales genomes) lack CdiB genes. This type of CRCT/A module is widespread in proteobacterial genomes (166). MAG analyses also revealed more rCRCT/A-3a modules in *Rickettsia* genomes and a much higher presence of single antidotes (Fig. 1C and 2), possibly indicating greater selection for defense against toxins versus toxin secretion. Additionally, a rCRCT/A-3a module was found integrated into a large hemagglutinin-like toxin in the Rickettsiales endosymbiont Ac37b, an early-branching Rickettsiaceae species that can co-infect ticks with SFG rickettsiae (Fig. 8B; Fig. S9). Remarkably, this species also carries an identical hemagglutinin-like toxin but with a divergent warhead of the HNH/ENDO VII nuclease superfamily, illustrating the integrative nature of diverse CRCT/A modules. Furthermore, two toxins in the genome of the Rickettsiales endosymbiont of *Peranema trichophorum* (Midichloriaceae) carry different C-terminal toxins, as well as lysin motifs (LysM) that often occur in cell wall-degrading enzymes (170). Collectively, our analyses illuminate diverse CRCT/A modules in the Rickettsiales mobilome that equip bacteria with weapons for interbacterial antagonism.

Rickettsiales species may also utilize filamentation induced by cAMP (FIC) proteins and type II TA modules for interbacterial antagonism. Some intracellular bacteria harbor FIC domain-containing proteins (163), and several human pathogens secrete effectors with FIC domains into host cells to subvert cellular processes (171). Furthermore, a recent report illustrated that *Yersinia pseudotuberculosis* utilizes an FIC domain effector, CccR, that alters conspecific gene expression and inhibits congener growth (172). Many *Rickettsia* genomes encode multiple divergent FIC proteins (Table S2); however, to our knowledge, none of these proteins are known to be secreted by rickettsiae. Similarly, *Rickettsia* species also harbor a myriad of diverse type II TA modules, with many found on RAGE (e.g., Fig. 7F; Fig. S7F) or plasmids (94, 173). Only one module, VapBC of *R. felis*, has been characterized. Structural analysis revealed the nature of antidote (VapB) binding to toxin (VapC) (174), and VapC possesses toxic RNase activity when expressed in bacterial or eukaryotic host cells. We previously showed that *Rickettsia* genomes encode VapBC and/or a divergent module (VapBC-d) (91), and MAG analysis confirmed this observation (Fig. 2). Furthermore, in light of the new genomic diversity, more discrete VapBC loci are encoded in many genomes (data not shown), as well as other type II TA modules (e.g., those encoding ParE, BrnT, and RatA toxins) that have yet to be characterized (Table S2).

Finally, MAG analyses doubled the number of Rickettsiaceae proteins harboring domains found in RP toxins, particularly those of *Wolbachia* cytoplasmic incompatibility-inducing nucleases (CinB) and deubiquitinases (CidB), as well as the *Spiroplasma* male killer toxin deubiquitinase (Spaid) (175–179) (Fig. S10). Many of these toxins are substantially large and modular, encoding numerous domains with uncharacterized effects on host cells (5, 180, 181) (e.g., see Fig. 3D). The increasing number of RP toxins (and antidotes when present) in rickettsiae, particularly in species associated with male-killing and parthenogenesis phenotypes, indicates undiscovered molecular mechanisms underpinning these modes of RP. Like the rCRCT/A modules, FIC toxins, and type II TA modules, these RP toxins all lack characterized secretion pathways.

While the *rvh* T4SS cannot be ruled out as a secretion pathway for any of these potential effectors, a T1SS conserved in all Rickettsiales (91) (and possibly other unappreciated routes) could also be involved. It is important to consider that all putative effectors, particularly those acquired by LGT, maybe in various stages of coevolving with novel secretory pathways and secretion systems. Thus, while there is strong selection for retaining *rvh* T4SS genes in rickettsial genomes, the presence of REMs and cREMs alone does not imply their secretion given the often-complex interactions between the T4SS machine components and translocated effectors that typically co-evolve to streamline recognition and secretion (182, 183).

Power and efficacy of MAG diversity

The inclusion of diverse MAGs in the assessment of *rvh* effector evolution has provided several key insights. First, like the *rvh* T4SS (42), many REMs and cREMs are often present as duplicate genes in rickettsial genomes. However, unlike the *rvh* machine, effector duplication seems to define basal lineages (*Tisiphia*, BG rickettsiae, and other Rickettsiaceae) and tends to lead to the retention of only one protein in the later-evolving *Rickettsia* groups. Still, divergent forms arising from duplication stand to inform on effector function, particularly if derived proteins are utilized for vertebrate cell infection.

Second, the sparse distribution of polled effectors outside of *Rickettsia* genomes indicates they originated after the divergence of rickettsial families. In some cases, analyses strongly implicate LGT for the acquisition of effectors, with a particular bias from Legionellales (e.g., RalF, patatins, and PIKs) and other aquatic microbes. This supports the “intracellular arena” hypothesis for the gain of similar effectors in divergent pathogens that occupy common hosts (i.e., protists and arthropods) (184). It also corroborates our earlier observations that LGT, particularly by RAGEs and plasmids, offsets reductive genome evolution in rickettsiae (26, 93). A more recent study reached a similar conclusion via the discovery of gene gain shaping Chlamydiae genome architecture, despite the reduced size of most chlamydial genomes (185). MAGs have also provided a greater appreciation for Legionellales diversity and revealed that the major host-adaptive features (i.e., the *dot/icm* T4SS and a few conserved effectors) were established in the last common Legionellales ancestor (186). This is consistent with the discovery by Schön et al. (28) of the *rvh* T4SS in ancestral Rickettsiales; however, it is premature to conclude that strict *rvh* repurposing from congener killing to facilitation of host parasitism occurred until the secretory pathways of the numerous effectors described above are experimentally determined.

Third, MAGs help bridge the gap between research on microbial ecology and human pathogenesis, revealing genome evolutionary and architectural traits that are underappreciated due to biases of clinical isolates or more common environmental strains on public databases (72, 73). Our discovery here of REMs and cREMs on novel plasmids and RAGEs accentuates this point, indicating that such genetic elements may be underestimated for roles in rickettsial biology due to the strong bias of high passage clinical isolates on databases. This is particularly relevant considering the recent demonstration that the *Rickettsia* regulator of actin motility (*roaM*) is often pseudogenized in highly passaged laboratory strains, suggesting serial passage in cell culture can

eliminate essential genes lacking environmental selective pressure (in this case, the arthropod cytosol) (187).

Fourth, the most profound insight gained from our work shows how MAG analyses often illuminate novel architectures for well-studied virulence factors. Unearthing new effector designs provides clues on how general foundations are tailored to different hosts and host cell processes. This is epitomized by our discovery of novel RalF-like proteins with SCDs substituted for ARF-interacting domains, which not only fortifies the literature on *Legionella* and *Rickettsia* RalF-mediated host ARF recruitment (56, 84, 87, 89, 90) but also pinpoints the rise of the actin-targeting VPR regions in *Rickettsia* RalF proteins after the divergence from *Tisiphia* species. Combined with numerous other novel effector architectures identified herein, this highlights a remarkable recapitulation of mechanisms for mimicking eukaryotic functions that exist beyond *Rickettsia* and other human pathogens and are widespread in the environment. We assert that widening the comparative genomics lens will allow evolution, which has already matched effector form and function to host environments, to guide experimental designs and reinvigorate pathogen effector research.

Finally, as the landscape of *Rickettsia* pathogenesis undergoes gradual change due to virulence factor characterization and immunological studies (1, 2), the traditional designation of SFG and TG rickettsiae as the major lineages defining the genus has become grossly outdated. A substantial spike in TRG rickettsiae diversity (73, 188, 189), coupled with robust genome-based phylogeny estimations and phylogenomic analyses (28, 73), make the common ancestry of TG and TRG rickettsiae incontrovertible. Prior bias in genome sequences for SFG rickettsiae portrayed TG rickettsiae as unique by smaller genome size and greater pseudogenization relative to all other rickettsiae. However, our focus on *rvh* effectors across a highly diverse and unbiased genomic sampling shows that all the major *Rickettsia* groups (BG, TRG, TG, TIG, and SFG rickettsiae) have distinct evolutionary trajectories of gene gain, loss, and modification (Fig. 2). Thus, MAGs have exposed far greater *Rickettsia* diversity than previously realized, though long ago conjectured by environmental sampling of rickettsiae infecting non-blood-feeding eukaryotes (190). These data, as well as careful dissection of the attributes distinguishing *Tisiphia* and *Rickettsia* species, will be paramount for deciphering how human pathogens have emerged, possibly multiple times, from this veritable bevy of endosymbiont diversity. Furthermore, an understanding of environmental genomic richness, particularly in mobile element diversity, may help forecast the next serious rickettsial diseases to emerge.

Conclusion

Discovery and analyses of MAGs have greatly impacted the landscape of Rickettsiology, adding substantial diversity and dispelling the long-held dogma for an ancestral link to the mitochondrial ancestor. Despite predicted extracellular lifestyles, basal rickettsial species carry the *rvh* T4SS and likely use it as a congener killing machine. Our study coupled a robust evolutionary framework with the inspection of over two dozen known or predicted *Rickettsia rvh* T4SS effectors to provide insight on the origin of mechanisms for host cell subversion and obligate intracellular parasitism. Though focused on taxonomic scope, this experimental design is amenable to probing the origins of virulence factors in any human pathogen with representation in the diverse treasure trove of MAG data. At the bare minimum, our work demonstrates that utilizing MAGs in comparative approaches greatly enlightens dialogue on mechanisms of pathogenesis.

MATERIALS AND METHODS

Rickettsiales phylogeny estimation

Robust genome-based phylogeny estimations for Rickettsiales (28) and *Rickettsia-Tisiphia* (73) were used as benchmarks to evaluate our estimated phylogenies based on single

or concatenated *rvh* proteins. We polled the rich MAG diversity on the NCBI database for the presence of *vir*-like T4SS genes possessing *rvh* hallmarks (38, 42, 45) (i.e., RvhB8, RvhB9, and RvhB4 duplication, multicopy RvhB6, no VirB5 analog; Fig. S1). Provided that many MAGs and certain genome assemblies are likely incomplete, we limited our data set to assemblies containing both RvhB4-I and RvhB4-II, except for a few cases where strong evidence from other *rvh* genes indicated a Rickettsiales assembly. A total of 153 genome assemblies were retained for further analyses: (i) 93 Rickettsiaceae genome assemblies [including the 28 MAGs from Davison et al. (73) and another 15 previously unanalyzed MAGs], (ii) 14 and 9 genome assemblies from Anaplasmataceae and Midichloriaceae, respectively, (iii) the “*Candidatus* Deianiraea vastatrix” (Deianiraeaceae) genome assembly, and (iv) 33 environmental MAGs likely comprising Deianiraeaceae, Athabascaceae, or Mitibacteraceae [nine previously analyzed by Schön et al. (28)] (Table S1; Fig. S2).

Only RvhB4-I and RvhB4-II proteins were included in phylogeny estimation, as alignments of other Rvh proteins were extraordinarily variable across the selected taxa (data not shown). RvhB4-I and RvhB4-II proteins were separately aligned using MUSCLE (default parameters). Each alignment included *Agrobacterium tumefaciens* str. F4 VirB4, which was used as an outgroup to root estimated trees. RvhB4-I and RvhB4-II protein alignments were subsequently concatenated (1,974 total positions, “unmasked alignment”). TRIMAL (191) was used to create a second alignment with less conserved positions masked (1,613 total positions, “masked alignment”).

For both unmasked and masked alignments, a maximum likelihood-based phylogeny was estimated with PhyML (192), using the Smart Model Selection (193) tool to determine the best substitution matrix and model for rates across aa sites [LG (G + I + F) for both alignments]. Branch support was assessed with 1,000 pseudo-replications. Trees were drawn using FigTree (<https://github.com/rambaut/figtree/>) and manually modified using Adobe Illustrator. Final trees were manually adjusted to place “*Candidatus* Sneabacter namystus” (which lacks the *rvh* T4SS) in a position on the phylogram suggested by prior phylogeny estimation (82, 83). All terminal taxa were assigned names based on NCBI database taxonomy (as of 26 February 2023), with some “short names” taken from Davison et al. (73) (these are provided in black boxes throughout the figures). Rickettsial classification scheme (Scrub Typhus Group, Bellii Group, Transitional Group, Typhus Group, Tamurae-Ixodes Group, and Spotted Fever Group) follows our prior reports (81, 173).

Phylogenomics analysis

The RvhB4-based estimated phylogeny was used as a scaffold to complete a distribution matrix for REMs and cREMs. It was not our goal to assess the relative completeness of each MAG included in the matrix, but to only assess if MAGs and other genome assemblies possessing a *rvh* T4SSs also include counterparts (homologs or analogs) to *Rickettsia* REMs and cREMs. REM/cREM assignment is based on prior studies implicating their secretion and/or interaction with RvhD4 (by bacterial two-hybrid and/or coimmunoprecipitation assays) or the presence of motifs known to target either congener bacteria or eukaryotic molecules (56–58, 74–81). Analyses of some REMs and cREMs illuminated more complex gene structures (duplications, gene streamlining, and gene fusions) that prompted expansion of the total effector data set. A total of 26 proteins were analyzed within the final phylogenomic framework (Table S2).

In silico protein characterization

Analyses of each REM and cREM data set contained discrete workflows tailored to the level of effector conservation in Rickettsiales (and in some cases, other bacteria), prior studies that included bioinformatics analyses, and identification of gene duplication, streamlining, or gene fusion. All individual workflows are described in the pertinent

figure legends and/or supplemental figure legends. Only general bioinformatics analyses are described below.

Data set compilation

Rickettsia REMs and cREMs 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 originally designed to interrogate proteins for LGT (26). HaloBlasting compiles Blastp subjects from restricted taxonomic searches that theoretically decrease in similarity by sampling lower levels of bacterial classification. A general search strategy for rickettsiae entails individual Blastp searches against six distinct taxonomic databases: (i) “*Rickettsia*” (NCBI taxid: 780); (ii) “*Rickettsiales*” (taxid: 766) excluding “*Rickettsia*”; (iii) “*Alphaproteobacteria*” (taxid: 28211) excluding “*Rickettsiales*”; (iv) “*Proteobacteria*” (taxid: 1224) excluding “*Alphaproteobacteria*”; (v) “*Bacteria*” (taxid: 2) excluding *Proteobacteria*; and (vi) “minus bacteria.” Data subsets were constructed strictly using the NCBI taxid and following the NCBI taxid hierarchy to identify “daughter” taxonomic groups. Typically, 500 subjects (if available) are retained per search. All subjects from each search were separately ranked by *Sm* score ($= b \times l \times Q$, where *b* is the bit score of the match, *l* 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 (26). The “halos” or separate database searches are then compared to one another to determine the taxon with the strongest similarity to the query sequences. These analyses usually make LGT apparent when divergent data sets contain top-ranking proteins more similar to the *Rickettsia* queries than more closely related data sets.

Protein characterization

Select proteins or domains (again, context-dependent) are typically compiled and aligned with MUSCLE using default parameters (104). To identify conserved regions, alignments are then visualized as sequence logos using WebLogo (103). Domain analyses are performed by cross-checking predictions from the NCBI Conserved Domains Database and EMBL's Simple Modular Architecture Research Tool (194). In some cases, proteins were evaluated for N-terminal signal peptides (195) and transmembrane-spanning regions (196). Alignments shown in the figures and supplemental figures are manually assessed for conservation, typically considering 80% of a position conserved (alignment size-dependent), with amino acid coloring scheme and assignment as follows: black, hydrophobic (Ala, Val, Iso, Leu, Pro, Met, and Gly); gray, less hydrophobic (can include a minority of Try, Phe, and Tyr); red, negatively charged (Glu and Asp); green, hydrophilic (Cys, Asn, Gln, Ser, and Thr); purple, aromatic (Try, Phe, and Tyr); and blue, positively charged (His, Lys, and Arg). Individual protein schemas were generated using Illustrator of Biological Sequences (197) with manual adjustment in Adobe Illustrator.

Protein structures were predicted using the Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) (95); in some cases, published structures were retrieved from the Protein Data Bank (198) and used in one-to-one threading mode with Phyre2. For some effectors, we also evaluated structures generated with AlphaFold (127, 128). Finally, some regions of proteins were analyzed for predicted secondary structure using JPred (199).

Phylogenies were estimated for some REM and cREM data sets, which were compiled uniquely for each case and utilized HaloBlast to obtain non-rickettsial taxa (if available). Alignments were not masked since masking eliminated too many informative positions. Maximum likelihood-based phylogenies were estimated with PhyML (192), using the Smart Model Selection (193) tool to determine the best substitution matrix and model for rates across aa sites. Branch support was assessed with 1,000 pseudo-replications. Trees were visualized and drawn as described above.

ACKNOWLEDGMENTS

This work was supported with funds from the National Institute of Health/National Institute of Allergy and Infectious Diseases grant R21AI156762 and R21AI166832. 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.

AUTHOR AFFILIATIONS

¹Division of Molecular and Cellular Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA

²Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA

³Department of Biology, West Virginia University, Morgantown, West Virginia, USA

⁴Department of Microbiology and Immunology, University of South Alabama, Mobile, Alabama, USA

AUTHOR ORCIDs

Joseph J. Gillespie  <http://orcid.org/0000-0002-5447-7264>

FUNDING

Funder	Grant(s)	Author(s)
HHS NIH National Institute of Allergy and Infectious Diseases (NIAID)	R21AI15676	Joseph J. Gillespie

AUTHOR CONTRIBUTIONS

Stephanie S. Lehman, Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review and editing | Victoria I. Verhoeve, Data curation, Investigation, Writing – original draft, Writing – review and editing | Timothy P. Driscoll, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Writing – original draft, Writing – review and editing | John F. Beckmann, Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review and editing | Joseph J. Gillespie, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Fig. S1-S5 (mBio00759-23-s0001.pdf). Supplemental figures.

Fig. S6-S10 (mBio00759-23-s0002.pdf). Additional supplemental figures.

Table S1 (mBio00759-23-s0003.xlsx). RvhB4 sequences used for phylogeny estimation.

Table S2 (mBio00759-23-s0004.xlsx). Supporting information for REMs and cREMs.

Table S3 (mBio00759-23-s0005.xlsx). Supporting information for type VI secretion system analyses.

Table S4 (mBio00759-23-s0006.xlsx). Supporting information for patatin analyses.

REFERENCES

- Salje J. 2021. Cells within cells: rickettsiales and the obligate intracellular bacterial lifestyle. *Nat Rev Microbiol* 19:375–390. <https://doi.org/10.1038/s41579-020-00507-2>
- Sahni A, Fang R, Sahni SK, Walker DH. 2019. Pathogenesis of rickettsial diseases: pathogenic and immune mechanisms of an endotheliotropic infection. *Annu Rev Pathol Mech Dis* 14:127–152. <https://doi.org/10.1146/annurev-pathmechdis-012418-012800>
- Narra HP, Sahni A, Walker DH, Sahni SK. 2020. Recent research milestones in the pathogenesis of human rickettsioses and opportunities ahead. *Future Microbiol* 15:753–765. <https://doi.org/10.2217/fmb-2019-0266>
- Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 6:741–751. <https://doi.org/10.1038/nrmicro1969>
- Gillespie JJ, Driscoll TP, Verhoeve VI, Rahman MS, Macaluso KR, Azad AF. 2018. A tangled web: origins of reproductive parasitism. *Genome Biol Evol* 10:2292–2309. <https://doi.org/10.1093/gbe/evy159>
- Gillespie JJ, Nordberg EK, Azad AA, Sobral BW. 2012. Phylogeny and comparative genomics: the shifting landscape in the genomics era, p 84–141. In Azad AF, Palmer GH (ed), *Intracellular pathogens II: Rickettsiales*. American Society of Microbiology, Boston.
- Kaur R, Shropshire JD, Cross KL, Leigh B, Mansueto AJ, Stewart V, Bordenstein SR, Bordenstein SR. 2021. Living in the endosymbiotic world of *Wolbachia*: a centennial review. *Cell Host Microbe* 29:879–893. <https://doi.org/10.1016/j.chom.2021.03.006>
- Mariconti M, Epis S, Gaibani P, Dalla Valle C, Sasser D, Tomao P, Fabbri M, Castelli F, Marone P, Sambri V, Bazzocchi C, Bandi C. 2012. Humans parasitized by the hard tick *Ixodes ricinus* are seropositive to *Midichloria mitochondrii*: is *Midichloria* a novel pathogen, or just a marker of tick bite? *Pathog Glob Health* 106:391–396. <https://doi.org/10.1179/204773212Y.0000000050>
- Szokoli F, Castelli M, Sabaneyeva E, Schrollhammer M, Krenek S, Doak TG, Berendonk TU, Petroni G. 2016. Disentangling the taxonomy of Rickettsiales and description of two novel symbionts (“*Candidatus* Bealeia paramacronuclearis” and “*Candidatus* Fokinia cryptica”) sharing the cytoplasm of the ciliate protist *Paramecium bialaurelia*. *Appl Environ Microbiol* 82:7236–7247. <https://doi.org/10.1128/AEM.02284-16>
- Floriano AM, Castelli M, Krenek S, Berendonk TU, Bazzocchi C, Petroni G, Sasser D. 2018. The genome sequence of “*Candidatus* Fokinia solitaria”: insights on reductive evolution in Rickettsiales. *Genome Biol Evol* 10:1120–1126. <https://doi.org/10.1093/gbe/evy072>
- Sasser D, Beninati T, Bandi C, Bouman EAP, Sacchi L, Fabbri M, Lo N. 2006. “*Candidatus* *Midichloria mitochondrii*”, an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int J Syst Evol Microbiol* 56:2535–2540. <https://doi.org/10.1099/ijs.0.64386-0>
- Montagna M, Sasser D, Epis S, Bazzocchi C, Vannini C, Lo N, Sacchi L, Fukatsu T, Petroni G, Bandi C. 2013. “*Candidatus* midichloriaceae” fam. nov. (Rickettsiales), an ecologically: widespread clade of intracellular alphaproteobacteria. *Appl Environ Microbiol* 79:3241–3248. <https://doi.org/10.1128/AEM.03971-12>
- Giannotti D, Boscaro V, Husnik F, Vannini C, Keeling PJ. 2022. The “other” Rickettsiales: an overview of the family “*Candidatus* Midichloriaceae” *Appl Environ Microbiol* 88. <https://doi.org/10.1128/aem.02432-21>
- Klinges JG, Rosales SM, McMinds R, Shaver EC, Shantz AA, Peters EC, Eitel M, Wörheide G, Sharp KH, Burkepille DE, Silliman BR, Vega Thurber RL. 2019. Phylogenetic, genomic, and biogeographic characterization of a novel and ubiquitous marine invertebrate-associated Rickettsiales parasite, *Candidatus* *Aquarickettsia rohweri*, gen. nov., sp. nov. *ISME J* 13:2938–2953. <https://doi.org/10.1038/s41396-019-0482-0>
- Castelli M, Sabaneyeva E, Lanzoni O, Lebedeva N, Floriano AM, Gaiarsa S, Benken K, Modeo L, Bandi C, Potekhin A, Sasser D, Petroni G. 2019. Deianiraea, an extracellular bacterium associated with the ciliate *Paramecium*, suggests an alternative scenario for the evolution of Rickettsiales. *ISME J* 13:2280–2294. <https://doi.org/10.1038/s41396-019-0433-9>
- Viklund J, Ettema TJG, Andersson SGE. 2012. Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol* 29:599–615. <https://doi.org/10.1093/molbev/msr203>
- Emelyanov VV, Vyssokikh MY. 2006. On the nature of obligate intracellular symbiosis of rickettsiae—*Rickettsia prowazekii* cells import mitochondrial porin. *Biochemistry (Mosc)* 71:730–735. <https://doi.org/10.1134/s0006297906070054>
- Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, Podowski RM, Näslund AK, Eriksson A-S, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133–140. <https://doi.org/10.1038/24094>
- Andersson SG. 1998. Bioenergetics of the obligate intracellular parasite *Rickettsia prowazekii*. *Biochim Biophys Acta* 1365:105–111. [https://doi.org/10.1016/s0005-2728\(98\)00050-4](https://doi.org/10.1016/s0005-2728(98)00050-4)
- Andersson SG, Kurland CG. 1998. Ancient and recent horizontal transfer events: the origins of mitochondria. *APMIS Suppl* 84:5–14. <https://doi.org/10.1111/j.1600-0463.1998.tb05641.x>
- Andersson JO, Andersson SG. 1999. Insights into the evolutionary process of genome degradation. *Curr Opin Genet Dev* 9:664–671. [https://doi.org/10.1016/s0959-437x\(99\)00024-6](https://doi.org/10.1016/s0959-437x(99)00024-6)
- Fuxelius H-H, Darby A, Min C-K, Cho N-H, Andersson SGE. 2007. The genomic and metabolic diversity of *Rickettsia*. *Res Microbiol* 158:745–753. <https://doi.org/10.1016/j.resmic.2007.09.008>
- Andersson JO, Andersson SG. 1999. Genome degradation is an ongoing process in *Rickettsia*. *Mol Biol Evol* 16:1178–1191. <https://doi.org/10.1093/oxfordjournals.molbev.a026208>
- Darby AC, Cho N-H, Fuxelius H-H, Westberg J, Andersson SGE. 2007. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet* 23:511–520. <https://doi.org/10.1016/j.tig.2007.08.002>
- Fuxelius H-H, Darby AC, Cho N-H, Andersson SGE. 2008. Visualization of pseudogenes in intracellular bacteria reveals the different tracks to gene destruction. *Genome Biol* 9:1–15. <https://doi.org/10.1186/gb-2008-9-2-r42>
- Driscoll TP, Gillespie JJ, Nordberg EK, Azad AF, Sobral BW. 2013. Bacterial DNA sifted from the trichoplax adhaerens (Animalia: placozoa) genome project reveals a putative rickettsial endosymbiont. *Genome Biol Evol* 5:621–645. <https://doi.org/10.1093/gbe/evt036>
- Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJG. 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* 557:101–105. <https://doi.org/10.1038/s41586-018-0059-5>
- Schön ME, Martijn J, Vosseberg J, Köstlbacher S, Ettema TJG. 2022. The evolutionary origin of host association in the Rickettsiales. *Nat Microbiol* 7:1189–1199. <https://doi.org/10.1038/s41564-022-01169-x>
- Martijn J, Schulz F, Zaremba-Niedzwiedzka K, Viklund J, Stepanauskas R, Andersson SGE, Horn M, Guy L, Ettema TJG. 2015. Single-cell genomics of a rare environmental alphaproteobacterium provides unique insights into rickettsiaceae evolution. *ISME J* 9:2373–2385. <https://doi.org/10.1038/ismej.2015.46>
- Schulz F, Martijn J, Wascher F, Lagkouvardos I, Kostanjšek R, Ettema TJG, Horn M. 2016. A Rickettsiales symbiont of amoebae with ancient features. *Environ Microbiol* 18:2326–2342. <https://doi.org/10.1111/1462-2920.12881>
- Lanzoni O, Sabaneyeva E, Modeo L, Castelli M, Lebedeva N, Verni F, Schrollhammer M, Potekhin A, Petroni G. 2019. Diversity and environmental distribution of the cosmopolitan endosymbiont “*Candidatus* Megaira” *Sci Rep* 9:1179. <https://doi.org/10.1038/s41598-018-37629-w>
- Sasser D, Lo N, Epis S, D’Auria G, Montagna M, Comandatore F, Horner D, Peretó J, Luciano AM, Franciosi F, Ferri E, Crotti E, Bazzocchi C, Daffonchio D, Sacchi L, Moya A, Latorre A, Bandi C. 2011. Phylogenomic evidence for the presence of a flagellum and cbb(3) oxidase in the free-living mitochondrial ancestor. *Mol Biol Evol* 28:3285–3296. <https://doi.org/10.1093/molbev/msr159>
- Kang YJ, Diao XN, Zhao GY, Chen MH, Xiong Y, Shi M, Fu WM, Guo YJ, Pan B, Chen XP, Holmes EC, Gillespie JJ, Dumler SJ, Zhang YZ. 2014. Extensive diversity of Rickettsiales bacteria in two species of ticks from China and the evolution of the Rickettsiales. *BMC Evol Biol* 14:167. <https://doi.org/10.1186/s12862-014-0167-2>
- Verhoeve VI, Gillespie JJ. 2022. Origin of rickettsial host dependency unravelled. *Nat Microbiol* 7:1110–1111. <https://doi.org/10.1038/s41564-022-01187-9>

35. Wang S, Luo H. 2021. Dating *Alphaproteobacteria* evolution with eukaryotic fossils. *Nat Commun* 12:1–9. <https://doi.org/10.1038/s41467-021-23645-4>
36. Ettema TJG, Andersson SGE. 2009. The α -proteobacteria: the Darwin finches of the bacterial world. *Biol Lett* 5:429–432. <https://doi.org/10.1098/rsbl.2008.0793>
37. Rikihisa Y, Lin M, Niu H, Cheng Z. 2009. Type IV secretion system of *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. *Ann N Y Acad Sci* 1166:106–111. <https://doi.org/10.1111/j.1749-6632.2009.04527.x>
38. Gillespie JJ, Brayton KA, Williams KP, Diaz MAQ, Brown WC, Azad AF, Sobral BW. 2010. Phylogenomics reveals a diverse Rickettsiales type IV secretion system. *Infect Immun* 78:1809–1823. <https://doi.org/10.1128/IAI.01384-09>
39. Rancès E, Voronin D, Tran-Van V, Mavingui P. 2008. Genetic and functional characterization of the type IV secretion system in *Wolbachia*. *J Bacteriol* 190:5020–5030. <https://doi.org/10.1128/JB.00377-08>
40. Rikihisa Y. 2017. Role and function of the type IV secretion system in *Anaplasma* and *Ehrlichia* species, p. 297–321. *Current topics in microbiology and immunology*.
41. Rikihisa Y, Lin M. 2010. *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins. *Curr Opin Microbiol* 13:59–66. <https://doi.org/10.1016/j.mib.2009.12.008>
42. Gillespie JJ, Ammerman NC, Dreher-Lesnick SM, Rahman MS, Worley MJ, Setubal JC, Sobral BS, Azad AF. 2009. An anomalous type IV secretion system in *Rickettsia* is evolutionarily conserved. *PLoS ONE* 4:e4833. <https://doi.org/10.1371/journal.pone.0004833>
43. Suttel EL, Norimine J, Beare PA, Heinzen RA, Lopez JE, Morse K, Brayton KA, Gillespie JJ, Brown WC. 2010. *Anaplasma marginale* type IV secretion system proteins VirB2, VirB7, VirB11, and VirD4 are immunogenic components of a protective bacterial membrane vaccine. *Infect Immun* 78:1314–1325. <https://doi.org/10.1128/IAI.01207-09>
44. Gillespie JJ, Phan IQH, Scheib H, Subramanian S, Edwards TE, Lehman SS, Piitulainen H, Rahman MS, Rennoll-Bankert KE, Staker BL, Taira S, Stacy R, Myler PJ, Azad AF, Pulliainen AT. 2015. Structural insight into how bacteria prevent interference between multiple divergent type IV secretion systems. *mBio* 6:e01867-15. <https://doi.org/10.1128/mBio.01867-15>
45. Gillespie JJ, Phan IQH, Driscoll TP, Guillotte ML, Lehman SS, Rennoll-Bankert KE, Subramanian S, Beier-Sexton M, Myler PJ, Rahman MS, Azad AF. 2016. The *Rickettsia* type IV secretion system: unrealized complexity mired by gene family expansion. *Pathog Dis* 74:ftw058. <https://doi.org/10.1093/femspd/ftw058>
46. Rikihisa Y. 2021. The “biological weapons” of *Ehrlichia chaffeensis*: novel molecules and mechanisms to subjugate host cells. *Front Cell Infect Microbiol* 11:830180. <https://doi.org/10.3389/fcimb.2021.830180>
47. Yan Q, Zhang W, Lin M, Teymournejad O, Budachetri K, Lakritz J, Rikihisa Y. 2021. Iron robbery by intracellular pathogen via bacterial effector-induced ferritinophagy. *Proc Natl Acad Sci USA* 118:e2026598118. <https://doi.org/10.1073/pnas.2026598118>
48. Yan Q, Lin M, Huang W, Teymournejad O, Johnson JM, Hays FA, Liang Z, Li G, Rikihisa Y. 2018. *Ehrlichia* type IV secretion system effector Etf-2 binds to active RAB5 and delays endosome maturation. *Proc Natl Acad Sci USA* 115:E8977–E8986. <https://doi.org/10.1073/pnas.1806904115>
49. Liu H, Bao W, Lin M, Niu H, Rikihisa Y. 2012. *Ehrlichia* type IV secretion effector ECH0825 is translocated to mitochondria and curbs ROS and apoptosis by upregulating host MnSOD. *Cell Microbiol* 14:1037–1050. <https://doi.org/10.1111/j.1462-5822.2012.01775.x>
50. Kim Y, Wang J, Clemens EG, Grab DJ, Dumler JS. 2022. *Anaplasma phagocytophilum* ankyrin A protein (AnkA) enters the nucleus using an importin- β , RanGTP-dependent mechanism. *Front Cell Infect Microbiol* 12:828605. <https://doi.org/10.3389/fcimb.2022.828605>
51. Lockwood S, Voth DE, Brayton KA, Beare PA, Brown WC, Heinzen RA, Broschat SL. 2011. Identification of *Anaplasma marginale* type IV secretion system effector proteins. *PLoS One* 6:e27724. <https://doi.org/10.1371/journal.pone.0027724>
52. Zhu J, He M, Xu W, Li Y, Huang R, Wu S, Niu H. 2019. Development of TEM-1 β -Lactamase based protein translocation assay for identification of *Anaplasma phagocytophilum* type IV secretion system effector proteins. *Sci Rep* 9. <https://doi.org/10.1038/s41598-019-40682-8>
53. Lin M, den Dulk-Ras A, Hooykaas PJJ, Rikihisa Y. 2007. *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cell Microbiol* 9:2644–2657. <https://doi.org/10.1111/j.1462-5822.2007.00985.x>
54. Niu H, Kozjak-Pavlovic V, Rudel T, Rikihisa Y. 2010. *Anaplasma phagocytophilum* Ats-1 is imported into host cell mitochondria and interferes with apoptosis induction. *PLoS Pathog* 6:e1000774. <https://doi.org/10.1371/journal.ppat.1000774>
55. Park JM, Genera BM, Fahy D, Swallow KT, Nelson CM, Oliver JD, Shaw DK, Munderloh UG, Brayton KA. 2023. An *Anaplasma phagocytophilum* T4SS effector, AteA, is essential for tick infection. *bioRxiv:2023.02.06.527355*. <https://doi.org/10.1101/2023.02.06.527355>
56. Rennoll-Bankert KE, Rahman MS, Gillespie JJ, Guillotte ML, Kaur SJ, Lehman SS, Beier-Sexton M, Azad AF. 2015. Which way in? The rlf ARF-GEF orchestrates *Rickettsia* host cell invasion. *PLoS Pathog* 11:e1005115. <https://doi.org/10.1371/journal.ppat.1005115>
57. Voss OH, Gillespie JJ, Lehman SS, Rennoll SA, Beier-Sexton M, Rahman MS, Azad AF. 2020. Risk1, a phosphatidylinositol 3-kinase effector, promotes *Rickettsia typhi* intracellular survival. *mBio* 11:1–22. <https://doi.org/10.1128/mBio.00820-20>
58. Lehman SS, Noriea NF, Aistleitner K, Clark TR, Dooley CA, Nair V, Kaur SJ, Rahman MS, Gillespie JJ, Azad AF, Hackstadt T. 2018. The rickettsial ankyrin repeat protein 2 is a type IV secreted effector that associates with the endoplasmic reticulum. *mBio* 9:e00975-18. <https://doi.org/10.1128/mBio.00975-18>
59. Souza DP, Oka GU, Alvarez-Martinez CE, Bisson-Filho AW, Dunger G, Hobeika L, Cavalcante NS, Alegria MC, Barbosa LRS, Salinas RK, Guzzo CR, Farah CS. 2015. Bacterial killing via a type IV secretion system. *Nat Commun* 6:6453. <https://doi.org/10.1038/ncomms7453>
60. Ho BT, Dong TG, Mekalanos JJ. 2014. A view to a kill: the bacterial type 6 secretion system. *Cell Host Microbe* 15:9–21. <https://doi.org/10.1016/j.chom.2013.11.008>
61. Muñoz-Gómez SA, Hess S, Burger G, Lang BF, Susko E, Slamovits CH, Roger AJ. 2019. An updated phylogeny of the alphaproteobacteria reveals that the parasitic rickettsiales and holosporales have independent origins. *eLife* 8:eLife. <https://doi.org/10.7554/eLife.42535>
62. Yurchenko T, Ševčíková T, Příbyl P, El Karkouri K, Klimeš V, Amaral R, Zbránková V, Kim E, Raoult D, Santos LMA, Eliáš M. 2018. A gene transfer event suggests a long-term partnership between eustigmatophyte algae and a novel lineage of endosymbiotic bacteria. *ISME J* 12:2163–2175. <https://doi.org/10.1038/s41396-018-0177-y>
63. Hunter ES, Paight C, Lane CE. 2020. Metabolic contributions of an alphaproteobacterial endosymbiont in the apicomplexan *Cardiosporidium cionae*. *Front Microbiol* 11:580719. <https://doi.org/10.3389/fmicb.2020.580719>
64. Tashyreva D, Prokopchuk G, Votýpka J, Yabuki A, Horák A, Lukeš J. 2018. Life cycle, ultrastructure, and phylogeny of new diplomonids and their endosymbiotic bacteria. *mBio* 9:e02447-17. <https://doi.org/10.1128/mBio.02447-17>
65. Prokopchuk G, Tashyreva D, Yabuki A, Horák A, Masařová P, Lukeš J. 2019. Morphological, ultrastructural, motility and evolutionary characterization of two new *Hemistasiidae* species. *Protist* 170:259–282. <https://doi.org/10.1016/j.protis.2019.04.001>
66. Castelli M, Lanzoni O, Nardi T, Lometto S, Modeo L, Potekhin A, Sassera D, Petroni G. 2021. ‘*Candidatus* Sarmatiella mevalonica’ endosymbiont of the ciliate paramecium provides insights on evolutionary plasticity among rickettsiales. *Environ Microbiol* 23:1684–1701. <https://doi.org/10.1111/1462-2920.15396>
67. Schrällhammer M, Ferrantini F, Vannini C, Galati S, Schweikert M, Görtz H-D, Verni F, Petroni G. 2013. “*Candidatus* Megaira polyxenophila” gen. nov., sp. nov.: considerations on evolutionary history, host range and shift of early divergent rickettsiae. *PLoS One* 8:e72581. <https://doi.org/10.1371/journal.pone.0072581>
68. Park E, Poulin R. 2020. Widespread torix *Rickettsia* in New Zealand amphipods and the use of blocking primers to rescue host COI sequences. *Sci Rep* 10:1–13. <https://doi.org/10.1038/s41598-020-73986-1>
69. Pilgrim J, Ander M, Garros C, Baylis M, Hurst GDD, Siozios S. 2017. Torix group *Rickettsia* are widespread in *Culicoides* biting midges (Diptera: Ceratopogonidae), reach high frequency and carry unique genomic

- features. *Environ Microbiol* 19:4238–4255. <https://doi.org/10.1111/1462-2920.13887>
70. Thongprem P, Davison HR, Thompson DJ, Lorenzo-Carballa MO, Hurst GDD. 2021. Incidence and diversity of torix *Rickettsia*-Odonata symbioses. *Microb Ecol* 81:203–212. <https://doi.org/10.1007/s00248-020-01568-9>
 71. Wang H-L, Lei T, Wang X-W, Maruthi MN, Zhu D-T, Cameron SL, Rao Q, Shan H-W, Colvin J, Liu Y-Q, Liu S-S. 2020. A newly recorded *Rickettsia* of the Torix group is a recent intruder and an endosymbiont in the whitefly *Bemisia tabaci*. *Environ Microbiol* 22:1207–1221. <https://doi.org/10.1111/1462-2920.14927>
 72. Pilgrim J, Thongprem P, Davison HR, Siozios S, Baylis M, Zakharov EV, Ratnasingham S, deWaard JR, Macadam CR, Smith MA, Hurst GDD. 2021. Torix *Rickettsia* are widespread in arthropods and reflect a neglected symbiosis. *Gigascience* 10:giab021. <https://doi.org/10.1093/gigascience/giab021>
 73. Davison HR, Pilgrim J, Wybouw N, Parker J, Pirro S, Hunter-Barnett S, Campbell PM, Blow F, Darby AC, Hurst GDD, Siozios S. 2022. Genomic diversity across the *Rickettsia* and '*Candidatus* Megaira' genera and proposal of genus status for the torix group. *Nat Commun* 13:2630. <https://doi.org/10.1038/s41467-022-30385-6>
 74. Rahman MS, Gillespie JJ, Kaur SJ, Sears KT, Ceraul SM, Beier-Sexton M, Azad AF. 2013. *Rickettsia typhi* possesses phospholipase A2 enzymes that are involved in infection of host cells. *PLoS Pathog* 9:e1003399. <https://doi.org/10.1371/journal.ppat.1003399>
 75. Rahman MS, Ammerman NC, Sears KT, Ceraul SM, Azad AF. 2010. Functional characterization of a phospholipase A2 homolog from *Rickettsia typhi*. *J Bacteriol* 192:3294–3303. <https://doi.org/10.1128/JB.00155-10>
 76. Park H, Lee JH, Gouin E, Cossart P, Izard T. 2011. The *Rickettsia* surface cell antigen 4 applies mimicry to bind to and activate vinculin. *J Biol Chem* 286:35096–35103. <https://doi.org/10.1074/jbc.M111.263855>
 77. Sears KT, Ceraul SM, Gillespie JJ, Allen ED, Popov VL, Ammerman NC, Rahman MS, Azad AF. 2012. Surface proteome analysis and characterization of surface cell antigen (Sca) or autotransporter family of *Rickettsia typhi*. *PLoS Pathog* 8:e1002856. <https://doi.org/10.1371/journal.ppat.1002856>
 78. Gouin E, Egile C, Dehoux P, Villiers V, Adams J, Gertler F, Li R, Cossart P. 2004. The RickA protein of *Rickettsia conorii* activates the Arp2/3 complex. *Nature* 427:457–461. <https://doi.org/10.1038/nature02318>
 79. Reed SCO, Lamason RL, Risca VI, Abernathy E, Welch MD. 2014. *Rickettsia* actin-based motility occurs in distinct phases mediated by different actin nucleators. *Curr Biol* 24:98–103. <https://doi.org/10.1016/j.cub.2013.11.025>
 80. Audoly G, Vincentelli R, Edouard S, Georgiades K, Mediannikov O, Gimenez G, Socolovschi C, Mège JL, Cambillau C, Raoult D. 2011. Effect of rickettsial toxin VapC on its eukaryotic host. *PLoS One* 6:e26528. <https://doi.org/10.1371/journal.pone.0026528>
 81. Verhoeve VI, Fauntleroy TD, Risteen RG, Driscoll TP, Gillespie JJ. 2022. Cryptic genes for interbacterial antagonism distinguish *Rickettsia* species infecting blacklegged ticks from other *Rickettsia* pathogens. *Front Cell Infect Microbiol* 12:880813. <https://doi.org/10.3389/fcimb.2022.880813>
 82. George EE, Husnik F, Tashyreva D, Prokopchuk G, Horák A, Kwong WK, Lukeš J, Keeling PJ. 2020. Highly reduced genomes of protist endosymbionts show evolutionary convergence. *Curr Biol* 30:925–933. <https://doi.org/10.1016/j.cub.2019.12.070>
 83. George EE, Tashyreva D, Kwong WK, Okamoto N, Horák A, Husnik F, Lukeš J, Keeling PJ. 2022. Gene transfer agents in bacterial endosymbionts of microbial eukaryotes. *Genome Biol Evol* 14:evac099. <https://doi.org/10.1093/gbe/evac099>
 84. Nagai H, Kagan JC, Zhu X, Kahn RA, Roy CR. 2002. A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. *Science* 295:679–682. <https://doi.org/10.1126/science.1067025>
 85. Nagai H, Cambronne ED, Kagan JC, Amor JC, Kahn RA, Roy CR. 2005. A C-terminal translocation signal required for Dot/Icm-dependent delivery of the *Legionella* RalF protein to host cells. *Proc Natl Acad Sci U S A* 102:826–831. <https://doi.org/10.1073/pnas.0406239101>
 86. Folly-Klan M, Sancerne B, Alix E, Roy CR, Cherfils J, Campanacci V. 2015. On the use of *Legionella/Rickettsia* chimeras to investigate the structure and regulation of *Rickettsia* effector RalF. *J Struct Biol* 189:98–104. <https://doi.org/10.1016/j.jsb.2014.12.001>
 87. Alix E, Chesnel L, Bowzard BJ, Tucker AM, Delprato A, Cherfils J, Wood DO, Kahn RA, Roy CR. 2012. The capping domain in RalF regulates effector functions. *PLoS Pathog* 8:e1003012. <https://doi.org/10.1371/journal.ppat.1003012>
 88. Amor JC, Swails J, Zhu X, Roy CR, Nagai H, Ingmundson A, Cheng X, Kahn RA. 2005. The structure of RalF, an ADP-ribosylation factor guanine nucleotide exchange factor from *Legionella pneumophila*, reveals the presence of a cap over the active site. *J Biol Chem* 280:1392–1400. <https://doi.org/10.1074/jbc.M410820200>
 89. Folly-Klan M, Alix E, Stalder D, Ray P, Duarte LV, Delprato A, Zeghouf M, Antonny B, Campanacci V, Roy CR, Cherfils J. 2013. A novel membrane sensor controls the localization and argef activity of bacterial RalF. *PLoS Pathog* 9:e1003747. <https://doi.org/10.1371/journal.ppat.1003747>
 90. Rennoll-Bankert KE, Rahman MS, Guillotte ML, Lehman SS, Beier-Sexton M, Gillespie JJ, Azad AF. 2016. RalF-mediated activation of Arf6 controls *Rickettsia typhi* invasion by co-opting phosphoinositol metabolism. *Infect Immun* 84:3496–3506. <https://doi.org/10.1128/IAI.00638-16>
 91. Gillespie JJ, Kaur SJ, Rahman MS, Rennoll-Bankert K, Sears KT, Beier-Sexton M, Azad AF. 2015. Secretome of obligate intracellular *Rickettsia*. *FEMS Microbiol Rev* 39:47–80. <https://doi.org/10.1111/1574-6976.12084>
 92. Ogata H, La Scola B, Audic S, Renesto P, Blanc G, Robert C, Fournier P-E, Claverie J-M, Raoult D. 2006. Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet*. 2:e76. <https://doi.org/10.1371/journal.pgen.0020076>
 93. Gillespie JJ, Joardar V, Williams KP, Driscoll TP, Hostetler JB, Nordberg E, Shukla M, Walenz B, Hill CA, Nene VM, Azad AF, Sobral BW, Caler E. 2012. A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. *J Bacteriol* 194:376–394. <https://doi.org/10.1128/JB.06244-11>
 94. Gillespie JJ, Driscoll TP, Verhoeve VI, Utsuki T, Husseneder C, Chouljenko VN, Azad AF, Macaluso KR. 2015. Genomic diversification in strains of *Rickettsia felis* isolated from different arthropods. *Genome Biol Evol* 7:35–56. <https://doi.org/10.1093/gbe/evu262>
 95. Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the web: a case study using the phyre server. *Nat Protoc* 4:363–371. <https://doi.org/10.1038/nprot.2009.2>
 96. Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL, Shukla MP, Dalay O, Driscoll T, Hix D, Mane SP, Mao C, Nordberg EK, Scott M, Schulman JR, Snyder EE, Sullivan DE, Wang C, Warren A, Williams KP, Xue T, Seung Yoo H, Zhang C, Zhang Y, Will R, Kenyon RW, Sobral BW. 2011. Patric: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect Immun* 79:4286–4298. <https://doi.org/10.1128/IAI.00207-11>
 97. Whitworth T, Popov VL, Yu XJ, Walker DH, Bouyer DH. 2005. Expression of the *Rickettsia prowazekii* *pld* or *tlyC* gene in *Salmonella enterica* serovar typhimurium mediates phagosomal escape. *Infect Immun* 73:6668–6673. <https://doi.org/10.1128/IAI.73.10.6668-6673.2005>
 98. Driskell LO, Yu X, Zhang L, Liu Y, Popov VL, Walker DH, Tucker AM, Wood DO. 2009. Directed mutagenesis of the *Rickettsia prowazekii* *pld* gene encoding phospholipase D. *Infect Immun* 77:3244–3248. <https://doi.org/10.1128/IAI.00395-09>
 99. Borgo GM, Burke TP, Tran CJ, Lo NTN, Engström P, Welch MD. 2022. A patatin-like phospholipase mediates *Rickettsia parkeri* escape from host membranes. *Nat Commun* 13:3656. <https://doi.org/10.1038/s41467-022-31351-y>
 100. Ku B, Lee K-H, Park WS, Yang C-S, Ge J, Lee S-G, Cha S-S, Shao F, Heo WD, Jung JU, Oh B-H. 2012. VipD of *Legionella pneumophila* targets activated Rab5 and Rab22 to interfere with endosomal trafficking in macrophages. *PLoS Pathog* 8:e1003082. <https://doi.org/10.1371/journal.ppat.1003082>
 101. Gaspar AH, Machner MP. 2014. VipD is a Rab5-activated phospholipase A1 that protects *Legionella pneumophila* from endosomal fusion. *Proc Natl Acad Sci USA* 111:4560–4565. <https://doi.org/10.1073/pnas.1316376111>
 102. Lucas M, Gaspar AH, Pallara C, Rojas AL, Fernández-Recio J, Machner MP, Hierro A. 2014. Structural basis for the recruitment and activation of the *Legionella phospholipase* VipD by the host GTPase Rab5. *Proc Natl*

- Acad Sci USA 111:E3514–E3523. <https://doi.org/10.1073/pnas.1405391111>
103. Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190. <https://doi.org/10.1101/gr.849004>
 104. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
 105. Carrier TJ, Leigh BA, Deaker DJ, Devens HR, Wray GA, Bordenstein SR, Byrne M, Reitzel AM. 2021. Microbiome reduction and endosymbiont gain from a switch in sea urchin life history. *Proc Natl Acad Sci USA* 118:e2022023118. <https://doi.org/10.1073/pnas.2022023118>
 106. Sontowski R, Poeschl Y, Okamura Y, Vogel H, Guyomar C, Cortesero A-M, van Dam NM. 2022. A high-quality functional genome assembly of *Delia radicum* L. (Diptera: Anthomyiidae) annotated from egg to adult. *Mol Ecol Resour* 22:1954–1971. <https://doi.org/10.1111/1755-0998.13594>
 107. Walpole GFW, Pacheco J, Chauhan N, Clark J, Anderson KE, Abbas YM, Brabant-Kirwan D, Montaña-Rendón F, Liu Z, Zhu H, Brumell JH, Deiters A, Stephens LR, Hawkins PT, Hammond GRV, Grinstein S, Fairn GD. 2022. Kinase-independent synthesis of 3-phosphorylated phosphoinositides by a phosphotransferase. *Nat Cell Biol* 24:708–722. <https://doi.org/10.1038/s41556-022-00895-y>
 108. Pizarro-Cerdá J, Kühbacher A, Cossart P. 2015. Phosphoinositides and host–pathogen interactions. *Biochim Biophys Acta* 1851:911–918. <https://doi.org/10.1016/j.bbailip.2014.09.011>
 109. Pizarro-Cerdá J, Cossart P. 2004. Subversion of phosphoinositide metabolism by intracellular bacterial pathogens. *Nat Cell Biol* 6:1026–1033. <https://doi.org/10.1038/ncb1104-1026>
 110. Steiner B, Weber S, Hilbi H. 2018. Formation of the *Legionella*-containing vacuole: phosphoinositide conversion, GTPase modulation and ER dynamics. *Int J Med Microbiol* 308:49–57. <https://doi.org/10.1016/j.ijmm.2017.08.004>
 111. Dong N, Niu M, Hu L, Yao Q, Zhou R, Shao F. 2016. Modulation of membrane phosphoinositide dynamics by the phosphatidylinositol 4-kinase activity of the *Legionella* LepB effector. *Nat Microbiol* 2:16236. <https://doi.org/10.1038/nmicrobiol.2016.236>
 112. Ledvina HE, Kelly KA, Eshraghi A, Plemel RL, Peterson SB, Lee B, Steele S, Adler M, Kawula TH, Merz AJ, Skerrett SJ, Celli J, Mougous JD. 2018. A phosphatidylinositol 3-kinase effector alters phagosomal maturation to promote intracellular growth of *Francisella*. *Cell Host Microbe* 24:285–295. <https://doi.org/10.1016/j.chom.2018.07.003>
 113. Li X, Anderson DE, Chang Y-Y, Jarnik M, Machner MP. 2022. VpdC is a ubiquitin-activated phospholipase effector that regulates *Legionella* vacuole expansion during infection. *Proc Natl Acad Sci USA* 119:e2209149119. <https://doi.org/10.1073/pnas.2209149119>
 114. Kim DJ, Park K-S, Kim J-H, Yang S-H, Yoon JY, Han B-G, Kim HS, Lee SJ, Jang JY, Kim KH, Kim MJ, Song J-S, Kim H-J, Park C-M, Lee S-K, Lee BI, Suh SW. 2010. *Helicobacter pylori* proinflammatory protein up-regulates NF-kappaB as a cell-translocating Ser/Thr kinase. *Proc Natl Acad Sci USA* 107:21418–21423. <https://doi.org/10.1073/pnas.1010153107>
 115. Yu Q, Hu L, Yao Q, Zhu Y, Dong N, Wang DC, Shao F. 2013. Structural analyses of *Legionella* LepB reveal a new GAP fold that catalytically mimics eukaryotic RasGAP. *Cell Res* 23:775–787. <https://doi.org/10.1038/cr.2013.54>
 116. Felsheim RF, Kurtti TJ, Munderloh UG. 2009. Genome sequence of the endosymbiont *Rickettsia peacockii* and comparison with virulent *Rickettsia rickettsii*: identification of virulence factors. *PLoS One* 4:e8361. <https://doi.org/10.1371/journal.pone.0008361>
 117. Baldridge GD, Burkhardt NY, Felsheim RF, Kurtti TJ, Munderloh UG. 2008. Plasmids of the pRM/pRF family occur in diverse *Rickettsia* species. *Appl Environ Microbiol* 74:645–652. <https://doi.org/10.1128/AEM.02262-07>
 118. Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37. <https://doi.org/10.1093/nar/gkr367>
 119. Lamason RL, Bastounis E, Kafai NM, Serrano R, Del Álamo JC, Theriot JA, Welch MD. 2016. *Rickettsia* Sca4 reduces vinculin-mediated intercellular tension to promote spread. *Cell* 167:670–683. <https://doi.org/10.1016/j.cell.2016.09.023>
 120. Schuenke KW, Walker DH. 1994. Cloning, sequencing, and expression of the gene coding for an antigenic 120-kilodalton protein of *Rickettsia conorii*. *Infect Immun* 62:904–909. <https://doi.org/10.1128/iai.62.3.904-909.1994>
 121. Aistleitner K, Clark T, Dooley C, Hackstadt T. 2020. Selective fragmentation of the trans-golgi apparatus by *Rickettsia rickettsii*. *PLoS Pathog* 16:e1008582. <https://doi.org/10.1371/journal.ppat.1008582>
 122. Barrett AJ, Rawlings ND. 2001. Evolutionary lines of cysteine peptidases. *Biol Chem* 382:727–733. <https://doi.org/10.1515/BC.2001.088>
 123. Mosavi LK, Minor DL, Peng Z-Y. 2002. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc Natl Acad Sci USA* 99:16029–16034. <https://doi.org/10.1073/pnas.252537899>
 124. Yu JW, Jeffrey PD, Ha JY, Yang X, Shi Y. 2011. Crystal structure of the mucosa-associated lymphoid tissue lymphoma translocation 1 (MALT1) paracaspase region. *Proc Natl Acad Sci USA* 108:21004–21009. <https://doi.org/10.1073/pnas.1111708108>
 125. McLuskey K, Rudolf J, Proto WR, Isaacs NW, Coombs GH, Moss CX, Mottram JC. 2012. Crystal structure of a *Trypanosoma brucei* metacaspase. *Proc Natl Acad Sci USA* 109:7469–7474. <https://doi.org/10.1073/pnas.1200885109>
 126. Puri AW, Lupardus PJ, Deu E, Albrow VE, Garcia KC, Bogoy M, Shen A. 2010. Rational design of inhibitors and activity-based probes targeting *Clostridium difficile* virulence factor TcdB. *Chem Biol* 17:1201–1211. <https://doi.org/10.1016/j.chembiol.2010.09.011>
 127. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
 128. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. 2022. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 50:D439–D444. <https://doi.org/10.1093/nar/gkab1061>
 129. Wang Z, Wu M. 2014. Complete genome sequence of the endosymbiont of *Acanthamoeba* strain UWC8, an amoeba endosymbiont belonging to the “*Candidatus* Midichloriaceae” family in Rickettsiales. *Genome Announc* 2:e00791-14. <https://doi.org/10.1128/genomeA.00791-14>
 130. Wang X-R, Burkhardt NY, Kurtti TJ, Oliver JD, Price LD, Cull B, Thorpe CJ, Thiel MS, Munderloh UG. 2021. Mitochondrion-dependent apoptosis is essential for *Rickettsia parkeri* replication and replication in vector cells. *mSystems* 6:e01209-20. <https://doi.org/10.1128/mSystems.01209-20>
 131. Berk JM, Lee MJ, Zhang M, Lim C, Hochstrasser M. 2022. OTDUB from the human pathogen *Orientia tsutsugamushi* modulates host membrane trafficking by multiple mechanisms. *Mol Cell Biol* 42:e0007122. <https://doi.org/10.1128/mcb.00071-22>
 132. Lim C, Berk JM, Blaise A, Bircher J, Koleske AJ, Hochstrasser M, Xiong Y. 2020. Crystal structure of a guanine nucleotide exchange factor encoded by the scrub typhus pathogen *Orientia tsutsugamushi*. *Proc Natl Acad Sci USA* 117:30380–30390. <https://doi.org/10.1073/pnas.2018163117>
 133. Adcox HE, Berk JM, Hochstrasser M, Carlyon JA. 2022. *Orientia tsutsugamushi* OTDUB is expressed and interacts with adaptor protein complexes during infection. *Infect Immun* 90:e0046922. <https://doi.org/10.1128/iai.00469-22>
 134. Berk JM, Lim C, Ronau JA, Chaudhuri A, Chen H, Beckmann JF, Loria JP, Xiong Y, Hochstrasser M. 2020. A deubiquitylase with an unusually high-affinity ubiquitin-binding domain from the scrub typhus pathogen *Orientia tsutsugamushi*. *Nat Commun* 11:2343. <https://doi.org/10.1038/s41467-020-15985-4>
 135. Mérens A, Matrat S, Aubry A, Lascols C, Jarlier V, Soussy CJ, Cavallo JD, Cambau E. 2009. The Pentapeptide repeat proteins MfpA_{mt} and QnrB4 exhibit opposite effects on DNA gyrase catalytic reactions and on the ternary gyrase-DNA-quinolone complex. *J Bacteriol* 191:1587–1594. <https://doi.org/10.1128/JB.01205-08>
 136. Kim N, Han G, Jung H, Lee HH, Park J, Seo YS. 2021. T6SS accessory proteins, including DUF2169 domain-containing protein and pentapeptide repeats protein, contribute to bacterial virulence in T6SS droup_5 of *Burkholderia glumae* BGR1. *Plants (Basel)* 11:34. <https://doi.org/10.3390/plants11010034>
 137. Ammerman NC, Gillespie JJ, Neuwald AF, Sobral BW, Azad AF. 2009. A typhus group-specific protease defies reductive evolution in rickettsiae. *J Bacteriol* 191:7609–7613. <https://doi.org/10.1128/JB.01077-09>
 138. van der Heide T, Stuart MC, Poolman B. 2001. On the osmotic signal and osmosensing mechanism of an ABC transport system for glycine betaine. *EMBO J* 20:7022–7032. <https://doi.org/10.1093/emboj/20.24.7022>

139. Wood JM. 1999. Osmosensing by bacteria: signals and membrane-based sensors. *Microbiol Mol Biol Rev* 63:230–262. <https://doi.org/10.1128/MMBR.63.1.230-262.1999>
140. Gillespie JJ, Williams K, Shukla M, Snyder EE, Nordberg EK, Ceraul SM, Dharmanolla C, Rainey D, Soneja J, Shallom JM, Vishnubhat ND, Wattam R, Purkayastha A, Czar M, Crasta O, Setubal JC, Azad AF, Sobral BS. 2008. *Rickettsia* phylogenomics: unwinding the intricacies of obligate intracellular life. *PLoS ONE* 3:e2018. <https://doi.org/10.1371/journal.pone.0002018>
141. Hagen R, Verhoeve VI, Gillespie JJ, Driscoll TP. 2018. Conjugative transposons and their cargo genes vary across natural populations of *Rickettsia buchneri* infecting the tick *Ixodes scapularis*. *Genome Biol Evol* 10:3218–3229. <https://doi.org/10.1093/gbe/evy247>
142. Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, Fukuhara M, Urakami H, Ohnishi M, Uchiyama I, Ogura Y, Ooka T, Oshima K, Tamura A, Hattori M, Hayashi T. 2008. The whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. *DNA Res* 15:185–199. <https://doi.org/10.1093/dnares/dsn011>
143. Cho N-H, Kim H-R, Lee J-H, Kim S-Y, Kim J, Cha S, Kim S-Y, Darby AC, Fuxelius H-H, Yin J, Kim JH, Kim J, Lee SJ, Koh Y-S, Jang W-J, Park K-H, Andersson SGE, Choi M-S, Kim I-S. 2007. The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. *Proc Natl Acad Sci USA* 104:7981–7986. <https://doi.org/10.1073/pnas.0611553104>
144. Heinzen RA, Hayes SF, Peacock MG, Hackstadt T. 1993. Directional actin polymerization associated with spotted fever group *Rickettsia* infection of Vero cells. *Infect Immun* 61:1926–1935. <https://doi.org/10.1128/iai.61.5.1926-1935.1993>
145. Heinzen RA, Grieshaber SS, Van Kirk LS, Devin CJ. 1999. Dynamics of actin-based movement by *Rickettsia rickettsii* in Vero cells. *Infect Immun* 67:4201–4207. <https://doi.org/10.1128/IAI.67.8.4201-4207.1999>
146. Van Kirk LS, Hayes SF, Heinzen RA. 2000. Ultrastructure of *Rickettsia rickettsii* actin tails and localization of cytoskeletal proteins. *Infect Immun* 68:4706–4713. <https://doi.org/10.1128/IAI.68.8.4706-4713.2000>
147. Gouin E, Gantelet H, Egile C, Lasa I, Ohayon H, Villiers V, Gounon P, Sansonetti PJ, Cossart P. 1999. A comparative study of the actin-based motilities of the pathogenic bacteria *Listeria monocytogenes*, *Shigella flexneri* and *Rickettsia conorii*. *J Cell Sci* 112 (Pt 11):1697–1708. <https://doi.org/10.1242/jcs.112.11.1697>
148. Jeng RL, Goley ED, D'Alessio JA, Chaga OY, Svitkina TM, Borisy GG, Heinzen RA, Welch MD. 2004. A *Rickettsia* WASP-like protein activates the Arp2/3 complex and mediates actin-based motility. *Cell Microbiol* 6:761–769. <https://doi.org/10.1111/j.1462-5822.2004.00402.x>
149. Madasu Y, Suarez C, Kast DJ, Kovar DR, Dominguez R. 2013. *Rickettsia* Sca2 has evolved formin-like activity through a different molecular mechanism. *Proc Natl Acad Sci USA* 110:E2677–E2686. <https://doi.org/10.1073/pnas.1307235110>
150. Kleba B, Clark TR, Lutter EI, Ellison DW, Hackstadt T. 2010. Disruption of the *Rickettsia rickettsii* Sca2 autotransporter inhibits actin-based motility. *Infect Immun* 78:2240–2247. <https://doi.org/10.1128/IAI.00100-10>
151. Haglund CM, Choe JE, Skau CT, Kovar DR, Welch MD. 2010. *Rickettsia* Sca2 is a bacterial formin-like mediator of actin-based motility. *Nat Cell Biol* 12:1057–1063. <https://doi.org/10.1038/ncb2109>
152. Serio AW, Jeng RL, Haglund CM, Reed SC, Welch MD. 2010. Defining a core set of actin cytoskeletal proteins critical or actin-based motility of *Rickettsia*. *Cell Host Microbe* 7:388–398. <https://doi.org/10.1016/j.chom.2010.04.008>
153. Balraj P, El Karkouri K, Vestris G, Espinosa L, Raoult D, Renesto P. 2008. *RickA* expression is not sufficient to promote actin-based motility of *Rickettsia raoultii*. *PLoS ONE* 3:e2582. <https://doi.org/10.1371/journal.pone.0002582>
154. Martinez JJ, Cossart P. 2004. Early signaling events involved in the entry of *Rickettsia conorii* into mammalian cells. *J Cell Sci* 117:5097–5106. <https://doi.org/10.1242/jcs.01382>
155. Oliver JD, Burkhardt NY, Felsheim RF, Kurtti TJ, Munderloh UG. 2014. Motility characteristics are altered for *Rickettsia bellii* transformed to overexpress a heterologous *rickA* gene. *Appl Environ Microbiol* 80:1170–1176. <https://doi.org/10.1128/AEM.03352-13>
156. Chan Y-Y, Riley SP, Martinez JJ. 2010. Adherence to and invasion of host cells by spotted fever group *Rickettsia* species. *Front Microbiol* 1:139. <https://doi.org/10.3389/fmicb.2010.00139>
157. Chan YGY, Cardwell MM, Hermanas TM, Uchiyama T, Martinez JJ. 2009. Rickettsial outer-membrane protein B (rOmpB) mediates bacterial invasion through Ku70 in an actin, c-Cbl, clathrin and caveolin 2-dependent manner. *Cell Microbiol* 11:629–644. <https://doi.org/10.1111/j.1462-5822.2008.01279.x>
158. Hillman RD, Baktash YM, Martinez JJ. 2013. OmpA-mediated rickettsial adherence to and invasion of human endothelial cells is dependent upon interaction with $\alpha 2\beta 1$ integrin. *Cell Microbiol* 15:727–741. <https://doi.org/10.1111/cmi.12068>
159. Riley SP, Goh KC, Hermanas TM, Cardwell MM, Chan YGY, Martinez JJ. 2010. The *Rickettsia conorii* autotransporter protein sca1 promotes adherence to nonphagocytic mammalian cells. *Infect Immun* 78:1895–1904. <https://doi.org/10.1128/IAI.01165-09>
160. Martinez JJ, Seveau S, Veiga E, Matsuyama S, Cossart P. 2005. Ku70, a component of DNA-dependent protein kinase, is a mammalian receptor for *Rickettsia conorii*. *Cell* 123:1013–1023. <https://doi.org/10.1016/j.cell.2005.08.046>
161. Carl M, Dasch GA. 1989. The importance of the crystalline surface layer protein antigens of rickettsiae in T-cell immunity. *J Autoimmun* 2 Suppl:81–91. [https://doi.org/10.1016/0896-8411\(89\)90119-4](https://doi.org/10.1016/0896-8411(89)90119-4)
162. Steele MI, Moran NA. 2021. Evolution of interbacterial antagonism in bee gut microbiota reflects host and symbiont diversification. *mSystems* 6:e00063-21. <https://doi.org/10.1128/mSystems.00063-21>
163. Massey JH, Newton ILG. 2022. Diversity and function of arthropod endosymbiont toxins. *Trends Microbiol* 30:185–198. <https://doi.org/10.1016/j.tim.2021.06.008>
164. Poole SJ, Diner EJ, Aoki SK, Braaten BA, t'Kint de Roodenbeke C, Low DA, Hayes CS. 2011. Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems. *PLoS Genet* 7:e1002217. <https://doi.org/10.1371/journal.pgen.1002217>
165. Zhang D, Iyer LM, Aravind L. 2011. A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic Acids Res* 39:4532–4552. <https://doi.org/10.1093/nar/gkr036>
166. Aoki SK, Diner EJ, de Roodenbeke CT, Burgess BR, Poole SJ, Braaten BA, Jones AM, Webb JS, Hayes CS, Cotter PA, Low DA. 2010. A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* 468:439–442. <https://doi.org/10.1038/nature09490>
167. Ruhe ZC, Subramanian P, Song K, Nguyen JY, Stevens TA, Low DA, Jensen GJ, Hayes CS. 2018. Programmed secretion arrest and receptor-triggered toxin export during antibacterial contact-dependent growth inhibition. *Cell* 175:921–933. <https://doi.org/10.1016/j.cell.2018.10.033>
168. Willett JLE, Ruhe ZC, Goulding CW, Low DA, Hayes CS. 2015. Contact-dependent growth inhibition (CDI) and CdiB/CdiA two-partner secretion proteins. *J Mol Biol* 427:3754–3765. <https://doi.org/10.1016/j.jmb.2015.09.010>
169. Aoki SK, Webb JS, Braaten BA, Low DA. 2009. Contact-dependent growth inhibition causes reversible metabolic downregulation in *Escherichia coli*. *J Bacteriol* 191:1777–1786. <https://doi.org/10.1128/JB.01437-08>
170. Joris B, Englebert S, Chu CP, Kariyama R, Daneo-Moore L, Shockman GD, Ghuyssen JM. 1992. Modular design of the *Enterococcus hirae* muramidase-2 and *Streptococcus faecalis* autolysin. *FEMS Microbiol Lett* 70:257–264. [https://doi.org/10.1016/0378-1097\(92\)90707-u](https://doi.org/10.1016/0378-1097(92)90707-u)
171. Veyron S, Peyroche G, Cherfils J. 2018. FIC proteins: from bacteria to humans and back again. *Pathog Dis* 76:12. <https://doi.org/10.1093/femspd/fty012>
172. Wang D, Zhu L, Zhen X, Yang D, Li C, Chen Y, Wang H, Qu Y, Liu X, Yin Y, Gu H, Xu L, Wan C, Wang Y, Ouyang S, Shen X. 2022. A secreted effector with a dual role as a toxin and as a transcriptional factor. *Nat Commun* 13:1–15. <https://doi.org/10.1038/s41467-022-35522-9>
173. Gillespie JJ, Beier MS, Rahman MS, Ammerman NC, Shallom JM, Purkayastha A, Sobral BS, Azad AF. 2007. Plasmids and rickettsial evolution: insight from *Rickettsia felis*. *PLoS One* 2:e266. <https://doi.org/10.1371/journal.pone.0000266>
174. Maté MJ, Vincentelli R, Foos N, Raoult D, Cambillau C, Ortiz-Lombardía M. 2012. Crystal structure of the DNA-bound VapBC2 antitoxin/toxin pair from *Rickettsia felis*. *Nucleic Acids Res* 40:3245–3258. <https://doi.org/10.1093/nar/gkr1167>
175. Beckmann JF, Sharma GD, Mendez L, Chen H, Hochstrasser M. 2019. The *Wolbachia* cytoplasmic incompatibility enzyme CIDB targets nuclear import and protamine-histone exchange factors. *Elife* 8:e50026. <https://doi.org/10.7554/eLife.50026>

176. Chen H, Ronau JA, Beckmann JF, Hochstrasser M. 2019. A *Wolbachia* nuclease and its binding partner comprise a novel mechanism for induction of cytoplasmic incompatibility. *Proc Natl Acad Sci USA*, no. 44:22314–22321. <https://doi.org/10.1073/pnas.1914571116>
177. LePage DP, Metcalf JA, Bordenstein SR, On J, Perlmutter JI, Shropshire JD, Layton EM, Funkhouser-Jones LJ, Beckmann JF, Bordenstein SR. 2017. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature* 543:243–247. <https://doi.org/10.1038/nature21391>
178. Beckmann JF, Ronau JA, Hochstrasser M. 2017. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat Microbiol* 2:17007. <https://doi.org/10.1038/nmicrobiol.2017.7>
179. Harumoto T, Lemaitre B. 2018. Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature* 557:252–255. <https://doi.org/10.1038/s41586-018-0086-2>
180. Driscoll TP, Verhoeve VI, Brockway C, Shrewsbury DL, Plumer M, Sevdalis SE, Beckmann JF, Krueger LM, Macaluso KR, Azad AF, Gillespie JJ. 2020. Evolution of *Wolbachia* mutualism and reproductive parasitism: insight from two novel strains that co-infect cat fleas. *PeerJ* 8:e10646. <https://doi.org/10.7717/peerj.10646>
181. Beckmann JF, Bonneau M, Chen H, Hochstrasser M, Poinot D, Merçot H, Weill M, Sicard M, Charlat S. 2019. The toxin–antidote model of cytoplasmic incompatibility: genetics and evolutionary implications. *Trends Genet* 35:175–185. <https://doi.org/10.1016/j.tig.2018.12.004>
182. Cascales E, Christie PJ. 2003. The versatile bacterial type IV secretion systems. *Nat Rev Microbiol* 1:137–149. <https://doi.org/10.1038/nrmicro753>
183. Cascales E, Christie PJ. 2004. Definition of a bacterial type IV secretion pathway for a DNA substrate. *Science* 304:1170–1173. <https://doi.org/10.1126/science.1095211>
184. Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol* 21:1981–1991. <https://doi.org/10.1093/molbev/msh211>
185. Dharamshi JE, Köstlbacher S, Schön ME, Collingro A, Ettema TJG, Horn M. 2023. Gene gain facilitated endosymbiotic evolution of Chlamydiae. *Nat Microbiol* 8:40–54. <https://doi.org/10.1038/s41564-022-01284-9>
186. Hugoson E, Guliaev A, Ammunét T, Guy L. 2022. Host adaptation in *Legionellales* is 1.9 Ga, coincident with eukaryogenesis. *Mol Biol Evol* 39:msac037. <https://doi.org/10.1093/molbev/msac037>
187. Nock AM, Clark TR, Hackstadt T. 2022. Regulator of actin-based motility (RoAM) downregulates actin tail formation by *Rickettsia rickettsii* and is negatively selected in mammalian cell culture. *mBio* 13:e0035322. <https://doi.org/10.1128/mbio.00353-22>
188. Gauthier DT, Karpathy SE, Grizzard SL, Batra D, Rowe LA, Paddock CD. 2021. Characterization of a novel transitional group *Rickettsia* species (*Rickettsia tillamookensis* sp. nov.) from the western black-legged tick, *Ixodes pacificus*. *Int J Syst Evol Microbiol* 71:004880. <https://doi.org/10.1099/ijsem.0.004880>
189. Blanton LS, Quade BR, Bouyer DH. 2019. Differentiation of *Rickettsia felis* and *Rickettsia felis*-like organisms via restriction fragment length polymorphism analysis. *Vector Borne Zoonotic Dis* 19:637–639. <https://doi.org/10.1089/vbz.2018.2428>
190. Perlman SJ, Hunter MS, Zchori-Fein E. 2006. The emerging diversity of *Rickettsia*. *Proc Biol Sci* 273:2097–2106. <https://doi.org/10.1098/rspb.2006.3541>
191. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
192. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>
193. Lefort V, Longueville JE, Gascuel O. 2017. SMS: smart model selection in PhyML. *Mol Biol Evol* 34:2422–2424. <https://doi.org/10.1093/molbev/msx149>
194. Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res* 46:D493–D496. <https://doi.org/10.1093/nar/gkx922>
195. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786. <https://doi.org/10.1038/nmeth.1701>
196. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>
197. Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, Lahrmann U, Zhao Q, Zheng Y, Zhao Y, Xue Y, Ren J. 2015. IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* 31:3359–3361. <https://doi.org/10.1093/bioinformatics/btv362>
198. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The protein data bank. *Nucleic Acids Res* 28:235–242. <https://doi.org/10.1093/nar/28.1.235>
199. Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 43:W389–W394. <https://doi.org/10.1093/nar/gkv332>