Secretome of obligate intracellular Rickettsia


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

*Corresponding author: Joseph J. Gillespie, Department of Microbiology and Immunology, University of Maryland (Baltimore City), 685 W. Baltimore St., HSF 1 Suite 380, Baltimore, MD, USA. Te.: (410) 706 3815; Fax: 410 706 0282; E-mail: jgillespie@som.umaryland.edu

One Sentence Summary: A phylogenomics-based review of the Rickettsia secretome identifies poorly understood aspects of protein secretion; thus, our contribution provides a thorough and current resource to help advance our knowledge of proteins that directly engage host cells during rickettsial infection.

Editor: Christroph Dehio

ABSTRACT

The genus Rickettsia (Alphaproteobacteria, Rickettsiales, Rickettsiaceae) is comprised of obligate intracellular parasites, with virulent species of interest both as causes of emerging infectious diseases and for their potential deployment as bioterrorism agents. Currently, there are no effective commercially available vaccines, with treatment limited primarily to tetracycline antibiotics, although others (e.g. josamycin, ciprofloxacin, chloramphenicol, and azithromycin) are also effective. Much of the recent research geared toward understanding mechanisms underlying rickettsial pathogenicity has centered on characterization of secreted proteins that directly engage eukaryotic cells. Herein, we review all aspects of the Rickettsia secretome, including six secretion systems, 19 characterized secretory proteins, and potential moonlighting proteins identified on surfaces of multiple Rickettsia species. Employing bioinformatics and phylogenomics, we present novel structural and functional insight on each secretion system. Unexpectedly, our investigation revealed that the majority of characterized secretory proteins have not been assigned to their cognate secretion pathways. Furthermore, for most secretion pathways, the requisite signal sequences mediating translocation are poorly understood. As a blueprint for all known routes of protein translocation into host cells, this resource will assist research aimed at uniting characterized secreted proteins with their apposite secretion pathways. Furthermore, our work will help in the identification of novel secreted proteins involved in rickettsial ‘life on the inside’.

Key words: Sec translocon; T5SS; Sec-TolC pathway; Tat system; T1SS; T4SS

INTRODUCTION

Included in the Alphaproteobacteria, all species of Rickettsiales are Gram-negative obligate intracellular parasites of a wide range of eukaryotic hosts (Driscoll et al., 2013). Two of the well-studied families, Anaplasmataceae (e.g. Neorickettsia, Wolbachia, Ehrlichia, Anaplasma) and Rickettsiaceae (e.g. Orientia, Rickettsia), contain several species of medical and agricultural importance (Gillespie et al., 2012a). The genus Rickettsia includes virulent species of interest both as emerging infectious diseases (Walker and Ismail, 2008) and for their potential deployment as bioterrorism agents (Azad and Radulovic, 2003). There are also species of Rickettsia with unspecified pathogenicity (Felsheim et al., 2009; Gillespie et al., 2012a, b), and others having no known associations with vertebrates or blood-feeding arthropods (Perlman et al., 2006). The rickettsial life cycle, which occurs solely within cells of one or more eukaryotic hosts, poses immense research challenges, and only recently have conventional genetic methodologies been applicable (Burkhardt et al., 2011;
In Rickettsia spp., the T5SS is defined exclusively by the surface cell antigen (sca) family, which is best classified as T5SSa (Fig. 4). The Sec-TolC pathway is solely defined by the substrate RARP-1 (Rickettsia ankyrin repeat protein 1), which requires its NT secretion signal, CT ankyrin domain, and the TolC protein for extracellular secretion (Fig. 5). The components of the twin-arginine translocation (Tat) system, as well as two Rickettsia Sec-independent secretion pathways, are conserved across all genomes despite no identified substrates for these systems (right). The three components of the Tat system suggest that Rickettsia spp. translocate folded substrates across the IM (Fig. 6). The TolC protein likely couples with at least one pair of IM proteins to form a functional T1SS (Fig. 7). A P-like T4SS, named the Rickettsiales vir homolog (rvh) T4SS, is similar to the vir archetype of Agrobacterium tumefaciens, yet atypical in the duplication of several scaffold components and the lack of a pilus (Fig. 8). C, cytoplasm; IM, inner membrane; PP, periplasm; M, murein layer; OM, outer membrane; LPS, lipopolysaccharide. Note: For all additional models (Fig. 2, Figs 4–8), these abbreviations apply, although the murein layer is removed in some cases for clearly illustrating pathways.

Wood et al., 2012). Consequently, little is known about the mechanisms underlying rickettsial pathogenicity. Of critical importance is a better understanding of the factors that (1) distinguish pathogens from nonpathogens, (2) permit some Rickettsia species to invade and colonize both invertebrate and vertebrate hosts, and (3) directly engage and manipulate eukaryotic cellular pathways.

At the forefront of rickettsial research is the identification and characterization of secretory molecules (i.e. surface-attached proteins and effector proteins released into host cells) and their cognate secretion pathways. Despite considerable reduction in size and gene content as a consequence of metabolite scavenging from hosts (Andersson et al., 1998), Rickettsia genomes encode various secretion systems that are homologous to characterized protein secretion pathways in other bacteria (Fig. 1). While all of these systems are undoubtedly critical for orchestrating life inside eukaryotic cells, much of the past research has focused on only a subset, namely the Sec translocon and the Sec-dependent type V secretion system (T5SS). The latter system, also known, and now inaccurately, as autotransporters (ATs), is comprised of antigens that dominate the surface of the rickettsial cell and have a dynamic range of interactions with host cell molecules. However, the interplay of the Sec translocon with other secretion systems, as well as the functions of Sec-independent secretion systems and the twin-arginine translocation (Tat) system, remain poorly understood aspects of rickettsial biology. A better understanding of these secretion systems will illuminate fundamental processes of bacterial life within eukaryotic cells, particularly phagosome escape, host immune avoidance, inhibition of autophagy and apoptosis, drug and toxin export, and host metabolite import.

This work reviews the current knowledge of protein secretion pathways of Rickettsia species. Utilizing 55 Rickettsia genome sequences, we employ bioinformatics and phylogenomics to gain further insight on each secretion system, as well as evaluate the conservation of known secretory molecules across these genomes. While focusing specifically on Rickettsia, comparisons to the secretomes of other rickettsial genera are made wherever pertinent. We follow the terminology previously suggested for disambiguating mechanisms of protein export, secretion, and translocation in Gram-negative bacteria (Desvaux et al., 2009), and we adhere mostly to the commonly used classification system for secretion pathways (Chagnot et al., 2013). Unexpectedly, our investigation revealed that the majority of characterized Rickettsia secretory proteins have not been assigned to their cognate secretion pathways. As our work provides a blueprint for all known routes of protein translocation into host cells, this information will assist future elucidation of the Rickettsia secretome.
SEC-DEPENDENT SECRETORY PATHWAYS

Sec translocon

In Gram-negative bacteria, many proteins are inserted into the inner membrane (IM) or translocated across the IM to the periplasm (PP) or outer membrane (OM). The dominant passage pathway for such proteins is the Sec translocon, which consists of IM and cytosolic proteins that act in concert to achieve these processes (Lycklama and Driessen, 2012). The secreted proteins themselves play a role in their insertion in and translocation across the IM, as an N-terminal (NT) helical transmembrane spanning (TMS) region and/or a strongly hydrophobic signal sequence (SS) are typically required for entrance to the Sec translocon (von Heijne, 1990a,b). At its minimum, the Sec translocon consists of a protein-conducting channel (SecYEG) and an ancillary complex that facilitates the late stages of translocation (SecDF and possibly YajC) (Fig. 2a). The remaining components of the Sec pathway have more specific functions; that is, substrate recognition, SS processing, and subcellular localization of substrates. Structural interchangeability of these components allows the Sec translocon to process a diversity of substrates, which include numerous IM, periplasmic, and OM proteins, as well as lipoproteins destined for the IM and OM (Daley et al., 2005; Okuda and Tokuda, 2011; Beckwith, 2013; Kudva et al., 2013).

Two distinct pathways, co-translational translocation (CTT) and post-translational translocation (PTT), are defined early in translation of Sec substrates, wherein the signal recognition particle (SRP) and trigger factor (TF) competitively bind NT SSs emerging from the ribosome (Beck et al., 2000; Ullers et al., 2006) (Fig. 2a). Although both the CTT and PTT pathways can deliver proteins to the IM, most IM proteins are CTT substrates (Ulbrandt et al., 1997; du Plessis et al., 2011). In CTT (Fig. 2a, left), SRP tightly binds a strongly hydrophobic NT sequence of the ribosome nascent chain (RNC) and arrests translation. The SRP/polypeptide/ribosome complex is then targeted to the IM receptor FtsY, which itself binds the SecYEG complex. Driven by GTP hydrolysis [of both SRP and FtsY (Egea et al., 2004; Focia et al., 2004; Akopian et al., 2013)], the polypeptide/ribosome complex is transferred to SecYEG, where translation continues and drives the force needed to translocate the preprotein through the Sec translocon. IM proteins are then inserted into the lipid bilayer predominantly by YidC (du Plessis et al., 2006; van Bloois et al., 2006; Kol et al., 2009). Aside from IM proteins translocated by the Tat system (discussed below), several proteins destined for the IM bypass the Sec translocon and are inserted directly into the membrane by YidC (van Bloois et al., 2005; Kiefer and Kuhn, 2007) (Fig. 2a, center). Although YidC can be directly engaged by the SRP/polypeptide/ribosome complex (Samuelsson et al., 2000; van der Laan et al., 2004; Price and Driessen, 2008, 2010; Dalbey et al., 2011; Wang and Dalbey, 2011; Welte et al., 2012), a few IM proteins inserted by YidC do so independent of SRP (du Plessis et al., 2011; Kudva et al., 2013). It has been hypothesized that SRP-dependent preproteins destined for the IM may utilize the most available insertion site (SecYEG or YidC) (Welte et al., 2012), providing flexibility that would allow for the pool of Sec translocons to efficiently process substrates of both the CTT and PTT pathways (Kudva et al., 2013).

The majority of Sec substrates targeting the PP and OM are products of PTT (Ulbrandt et al., 1997; du Plessis et al., 2011). In PTT (Fig. 2a, right), less hydrophobic polypeptides emerging from the ribosome are bound by TF in a process that does not arrest translation. Upon termination of translation, the SecB chaperone binds the unfolded polypeptide and targets it to the SecA ATPase located on the cytosolic side of the Sec translocon. Using energy from ATP hydrolysis, SecA then threads the substrate through the SecYEG channel, with the SecDF complex likely regulating secretion into the PP (Matsuyama et al., 1993; Duong and Wickner, 1997; Nouwen et al., 2005). SecA also plays a role in the insertion of large periplasmic loops of IM proteins of the CTT pathway (Saaf et al., 1995; Qi and Bernstein, 1999; Neumann-Haefelin et al., 2000; Deitermann et al., 2005), although the mechanics of this process are unclear (Welte et al., 2012; Kudva et al., 2013). Like the YidC insertase, this bifunctional role of SecA demonstrates the functional flexibility of components of the CTT and PTT pathways for processing a large array of diverse Sec substrates.

After passing through the Sec translocon, preproteins targeting the PP and OM are processed based on the characteristics of their NT SSs (most IM proteins are not processed at their N-termini) (du Plessis et al., 2011). The canonical Sec SS of nonlipoproteins for Gram-negative bacteria is c. 20–30 aa and contains an NT basic region, a hydrophobic patch, and a three-residue cleavage motif (typically AxA/x) (von Heijne, 1990; Paetzel et al., 2002). This SS is recognized and cleaved by signal peptidase (SPase) I (LepB) (Sung and Dalbey, 1992; Paetzel et al., 2002), with proteins subsequently trafficked to their subcellular destinations. Lipoproteins, which contain a ‘lipobox’ (typically Lxx/C) as a cleavage site motif within the SS, are processed on the periplasmic side of the IM by three enzymes (Lgt, LspA, and Lnt), resulting in the fusion of diacylglycerol and fatty acid moieties to the NT Cys residue (Sankaran and Wu, 1994) (Fig. 2b). Mature lipoproteins are then either anchored to the IM or transferred to the localization of lipoproteins (Lol) pathway for delivery to the OM (Okuda and Tokuda, 2011).

Comparative genomic analysis indicates that nearly all core and ancillary components of the Sec pathway are conserved across Rickettsia species (data not shown). Two exceptions are found within the Lol pathway, wherein genes encoding LolE and LolB are not present within Rickettsia genomes. For Escherichia coli, LolC and LolE are very similar proteins at the sequence and structural levels, with both proteins present at an equal ratio within the LolCDE transporter (1:2:1) (Yakushi et al., 2000). Despite reported functional and topological differences (Yasuda et al., 2009) and the essentiality of both proteins in E. coli (Narita et al., 2002), LolE is not present within genomes outside of Gammaproteobacteria (Narita, 2011). Accordingly, Rickettsia spp. probably assemble Lol transporters consisting of LolC homodimers with LolD homodimers (Fig. 2b). The absence of LolB from Rickettsia genomes is also unsurprising, as genes encoding LolB are unknown from Alphaproteobacteria (Narita, 2011; Okuda and Tokuda, 2011). While deletion of LolB in E. coli is lethal and results in the accumulation of OM lipoproteins in the PP (Tanaka et al., 2001), it is likely that lipoprotein insertion into the OM in alphaproteobacterial species is achieved solely by LolA or by an as yet unidentified protein distinct from LolB (Narita, 2011).

Previous studies from our laboratory have characterized several Rickettsia Sec pathway components. SecA was demonstrated to be expressed in R. rickettsii and R. typhi, yet the recombinant proteins could not complement the SecA mutant strain MM52 of E. coli (Rahman et al., 2005). However, chimeric SecA constructs (NT 408 aa of R. rickettsii fused with CT 480 aa of E. coli) did restore SecA function in E. coli strain MM52, implying that Rickettsia SecA proteins are functional but with species specificity in the CT domain (CTD). Entire LepB proteins from R. rickettsii and R. typhi were demonstrated to possess SPase I activity by restoring preprotein processing in an E. coli lepB mutant strain (Rahman et al., 2003). Subsequently, transcriptional
Figure 2. Overview of the Sec translocation pathway. (a) Via the YidC insertase, some membrane proteins localize to the IM independent of the Sec translocon, either dependent or independent of the signal recognition particle (SRP) (center, dashed arrows). Competitive binding of the ribosome nascent chain (RNC) between SRP (left, co-translational translocation, or CTT) and trigger factor (TF) (right, post-translational translocation, or PTT) leads to diverse translocation mechanisms in most Gram-negative bacteria, including Rickettsia spp. Steps for these diverse pathways are superficially similar: (1) in binding, SRP (Ffh plus the 4.5S rRNA gene) binds strongly hydrophobic NT sequences of nascent polypeptides and halts further translation (CTT), or less hydrophobic polypeptides emerging from the ribosome are bound by TF until termination of translation, whereby the SecB chaperone binds the unfolded polypeptide and keeps it in a translocation competent state (PTT); (2) in targeting, the SRP/RNC complex is localized to the IM receptor FtsY, which itself binds the SecYEG complex (CTT), or SecB-bound preproteins are delivered to the SecA ATPase located on the cytosolic side of the Sec translocon (PTT); (3) in translocation, the RNC complex is transferred (with GTP hydrolysis) to SecYEG, where translation provides the force needed to translocate the preprotein to the PP (CTT), or SecA (with ATP hydrolysis) threads the substrate through the SecYEG channel, with the SecDF complex providing a regulatory role (PTT); (4) in SS processing, nonlipoproteins from both pathways are typically processed at their NT secretion signals (SS) by SPase I (LepB), although SSs of IM proteins are rarely processed, and (5) in subcellular localization, nonlipoproteins from both pathways are delivered to their cell envelope destinations (IM, PP, OM) or potentially further translocated out of the bacterial cell. NOTE: Dominant pathways for IM (CTT, orange arrows) and PP/OM (PTT, green arrows) proteins are distinguished. SecA-mediated insertion of large periplasmic loops of IM proteins of the CTT pathway is illustrated (blue arrow). (b) After translocation via the Sec translocon, the ‘lipobox’ within the SS of lipoproteins mediates the following pathway: (1) addition by lipoprotein diacylglycerol transferase (Lgt) of a diacylglycerol moiety to a conserved Cys residue within the lipobox; (2) cleavage by SPase II (Lap) of the NT residues adjacent to the modified Cys; (3) addition by lipoprotein N-acyl transferase (Lnt) of a second amide-linked fatty acid to the NT Cys residue; (4) insertion of some lipoproteins into the IM or transfer of OM lipoproteins to the Lol system; (5) release of OM lipoproteins from the LolCDE complex (with ATP hydrolysis) upon binding by the chaperone LolA; and (6) OM lipoprotein insertion into the OM, which is mediated in some bacteria by the assembly factor LolB. NOTE: LolE and LolB (colored black) homologs are not encoded within Rickettsia genomes; all other illustrated components of the Sec translocon and lipoprotein pathway are present in Rickettsia spp. (Table S1).
analysis in *R. typhi* of *lepB*, along with *lspA* and *lgt*, determined that all three genes are highly expressed during infection, with *lepB* expression greater than the lipoprotein-processing genes (Rahman et al., 2007). Evidence for SPase II activity of *R. typhi* LspA was garnered by overexpression in *E. coli*, which increased globomycin resistance, as well as by genetic complementation, wherein recombinant LspA restored growth of temperature-sensitive *E. coli* Y815 at the nonpermissive temperature (Rahman et al., 2007). Collectively, these studies indicate a functional Sec pathway in obligate intracellular *Rickettsia* species.

A genomewide screen for Sec substrates of *R. typhi* further demonstrated a functional Sec pathway in *Rickettsia* species (Ammerman et al., 2008). Using a robust bioinformatics analysis, coupled with an *E. coli*-based alkaline phosphatase (PhoA) gene fusion system, SSs from 84 *R. typhi* proteins exhibited signal peptide activity. With a wide range of predicted cellular functions, many of these proteins were expected to be Sec substrates based on comparison with other well-studied bacterial species, although several proteins unique to *Rickettsia* species were also identified (Fig. 3a). An unexpected finding was a lack of strong SS predictions for *R. typhi* proteins homologous to Sec-dependent proteins from other Gram-negative bacteria. Furthermore, SS prediction was shown to be variable across implemented programs, with SPase cleavage motif predictions also variable across a subset of proteins (Fig. 3b). Undoubtedly, dozens of likely Sec-dependent proteins were not identified in this study, possibly due to characteristics of the Rickettsia SPase cleavage motif that escape strong in silico prediction. In support of this, PpcE, an S9 serine protease of *R. typhi* that lacks a strongly predicted Sec SS, is a likely Sec substrate as its cognate SS was able to mediate translocation in the *E. coli* PhoA gene fusion assay (Ammerman et al., 2009).

Contrary to the dogma of a universally conserved SPase cleavage motif within the SS of Gram-negative Sec-dependent proteins (von Heijne, 1990; Paetzl et al., 2002), a previous report demonstrated that some bacterial genomes possessing a strong base compositional bias encode Sec-dependent proteins with noncanonical SSs (Payne et al., 2012). In particular, a close relative of *Rickettsia* species, *Ehrlichia chaffeensis*, was shown to encode SPase cleavage motifs with conserved Ala residues replaced in many instances with Ser residues at positions −3 and −1, with a propensity for Phe at the −2 position. As nearly all *Rickettsiales* genomes show a strong base compositional bias, averaging about 30% GC, it is likely that this observation for *E. chaffeensis* applies to *Rickettsia* species. Our reanalysis of the 84 *R. typhi* proteins that exhibited signal peptide activity in the *E. coli*
PhoA gene fusion assay revealed two subsets of predicted SPase cleavage motifs for nonlipoproteins (Fig. 3c). One subset contains the conserved Ala residue at position −1, but with a lack of Ala at position −3, with Ser and Phe residues common at sites −3 and −2, respectively (Fig. 3c, left). A second subset showed minimal conservation for any residues within predicted SPase cleavage motifs (Fig. 3c, center). For 17 lipoproteins, the canonical lipobox motif is mostly conserved, but with a propensity for Ser residues at positions −2 and −1 (Fig. 3c, right). These data reveal that the inherent base compositional bias of Rickettsia genomes has impacted the nature of Sec SSs and that caution should be used in the interpretation of results from in silico predictions for future studies aiming to identify Sec-dependent secretory proteins.

**TSSS (autotransporters)**

TSSSs are defined by the ATs, proteins originally classified as orchestrating their own secretion out of the bacterial cell (Jose et al., 1995). Despite extraordinary diversity in gene structure and protein domains, as well as an emerging general consensus that they do not function alone in their translocation out of the bacterial cell (discussed below), ATs can be largely defined by the following characteristics: (1) Sec-dependent translocation to the bacterial PP; (2) a CT autotransport domain, or β domain (in some cases, a separate protein), that targets the protein to the OM; and (3) an NT passenger domain that is translocated to the extracellular surface by (at minimum) the β domain, with no energy requirement from cytosolic ATP (Leyton et al., 2012).

ATs have been classified into five groups (TSSSa-e) based primarily on their divergent gene structures, protein domain organizations, and overall functions (Leo et al., 2012; Grijpstra et al., 2013). Three well-known groups were previously classified (Henderson et al., 2000, 2004), namely the classical monomeric ATs (TSSSa), two-partner secretion systems (TPSS or TSSSb), and the trimeric ATs (TSSSc). Recently, genes encoding ‘fused TPSSs’, which have a passenger domain and β domain separated by a polypeptide transport associated (POTRA) domain, have been proposed as a fourth autotransporter (AT) group (TSSSd) (Salacha et al., 2010). Similar to the TSSSb β domain-like proteins (e.g. TpsB, FhaC, etc.), which also contain POTRA domains (Delattre et al., 2011), this modified β domain likely functions both in targeting TSSSd proteins to the OM and in initiating extracellular translocation, as the passenger domains of several TSSSa ATs are known to interact with the POTRA domains of BamA (Bodelon et al., 2009; Ieva and Bernstein, 2009; Ruiz-Perez et al., 2009; Sauri et al., 2009; Leyton et al., 2011). Finally, evaluation of a topology model for proteins of the intimin/invasin family (Tsai et al., 2010), coupled with in silico and experimental evidence for an AT-like arrangement of these proteins in the OM (Oberhettinger et al., 2012), led to a recent proposal for these proteins as TSSSe (Leo et al., 2012). The major distinguishing feature of TSSSe ATs is the reverse arrangement of the β (NT) and passenger (CT) domains in relation to classical TSSSa ATs, thus the mechanism for folding/translocation occurs in reverse order.

The original model for the mode of passenger domain translocation across the OM posits that a hairpin structure formed by the proximal CT of the passenger domain loops out of the pore formed by the β domain, with passenger domain folding (from CT to NT), driving the translocation of the remaining protein (Pohlner et al., 1987). While experimental data strongly support this ‘hairpin model’ (Oliver et al., 2003; Junker et al., 2006; Ieva and Bernstein, 2009; Junker et al., 2009; Peterson et al., 2010; Sopova et al., 2010), the growing number of AT crystal structures collectively reveals for the β domain a narrow 12-stranded β-barrel (Oomen et al., 2004; Barnard et al., 2007; van den Berg, 2010), leaving minimal space for two peptide strands (the alpha helix and proximal passenger domain) (Sauri et al., 2012). An initial observation that β domains of ATs contain recognition signals for the β-barrel assembly machine (BAM) (Jose et al., 1995) was followed by experimental evidence linking autotransport with the BAM complex (Jain and Goldberg, 2007; Ieva and Bernstein, 2009; Sauri et al., 2009). The exact contribution of the BAM complex to AT secretion, particularly the manner in which the passenger domain is exported across the OM, is still the subject of intense debate (Bernstein, 2007; Jain and Goldberg, 2007; Ieva et al., 2008, 2011; Rossiter et al., 2011). However, the importance of each ATs β domain to passenger domain export should not be dismissed, as replacement of structurally equivalent β-barrel domains or mutation of residues affecting β domain diameter has negative impacts on autotransport (Sauri et al., 2011). It is now apparent that for many (and maybe all) ATs, assistance in the process of passenger domain translocation is provided by one of the following: (1) the BAM complex (e.g. TSSSa, c, e), (2) AT-associated components with POTRA domains (e.g. TSSSb, d), or (3) the recently characterized archetypal translocation and assembly module (TAM) (Selkirk et al., 2012). This leaves the term ’autotransporter’, an inaccurate description for TSSS proteins (Leyton et al., 2012). Given this, it is important to consider the genomic makeup of these accessory systems in conjunction with the gene structure and protein domain organization of ATs for determining the mechanisms of type 5 secretion in each organism.

A recent model for the secretion pathway of a classical AT, EspP of E. coli, illustrates the contribution of the BAM complex and several periplasmic chaperones to type 5a secretion (Pavlova et al., 2013). Considering several lines of evidence, it is probable that a similar model defines the TSSS in Rickettsia (Fig. 4a). First, Rickettsia ATs typically contain a Sec SS, NT passenger domain, predicted linker (described below), and CT β domain, with AT genes occurring as single transcriptional units or within operons that include genes not associated with type 5 secretion (Sears et al., 2012). POTRA domains are unknown from Rickettsia ATs, and there are no homologs to TSSSb TpsB-like proteins encoded in Rickettsia genomes (data not shown). Thus, Rickettsia ATs are best described as classical, monomeric TSSSa proteins. Second, Rickettsia genomes contain genes for nearly a complete BAM complex, missing only bmc (Supporting Information, Table S1). Given that only BamA and BamD are likely required for AT biogenesis (Leyton et al., 2012), β domain insertion into the OM via a reduced BAM complex is plausible. For example, AT assembly in the OM for Neisseria spp. occurs even though bamB is absent from these genomes (Volokhina et al., 2011). Furthermore, genes encoding the TAM module (tama and tamb), which assists OM assembly and translocation for some ATs (Selkirk et al., 2012), are unknown from Rickettsia genomes (data not shown). Finally, Rickettsia genomes encode homologs to chaperones from other bacterial species that assist AT transport to the OM and subsequent assembly, namely SurA and Skp (Table S1). A homolog to DegP, which has been shown to be important for efficient secretion of EspP (Ruiz-Perez et al., 2009), is also encoded in Rickettsia genomes. Periplasmic DegP likely stabilizes and degrades proteins destined for the OM (Krojer et al., 2008), yet an extracellular function has also been described for E. chaffeensis degradation of the surface protein TRP120 (Kumagai et al., 2010). The OM foldase PrsA potentially assists passenger domain folding at the cell surface.
Figure 4. The Rickettsia TS55a is comprised of autotransporters of the surface cell antigen (Sca) family. (a) Model for the Rickettsia TS55a based largely on the recently proposed secretion pathway for Escherichia coli EspP (Pavlova et al., 2013). This model applies to all rickettsial Scas containing NT passenger (P) domains, linker (L) domains, and CT autotransporter (β) domains: (1) Sca engagement of the Sec translocon and subsequent translocation to the PP. Cleavage of NT Sec signal sequences (SS) is predicted to be delayed until the β domain has entered the PP, with misfolded proteins degraded by DegP; (2) folding of the β domain and insertion of the α-helical linker in a hairpin-like fashion within the β domain. The chaperone Skp is predicted to assist these processes (Ieva et al., 2011); (3) targeting of the Sca/Skp complex to the OM as the β domain contacts BamA, BamB, and BamD (light orange), with potential interactions of the passenger domain with BamA POTRA domains and the chaperone SurA; (4) conformational change in the β domain that initiates passenger domain translocation, with increased contacts between the passenger domain and BamA POTRA domains facilitating entry into an OM channel putatively comprised at least of the β domain and BamA; (5) a second conformational change in the β domain as the passenger domain accumulates and folds on the cell surface, possibly assisted by the foldase PrsA; (6) termination of passenger domain folding, triggering disassembly of the Bam complex, and release of the β domain into the OM; and (7) potential release of the passenger domain either by an intrabarrel cleavage reaction or assistance by an OM-associated protease. Information for each protein within the model is provided in Table S1. Inset: Immunogold electron microscopy of Sca2 expression on intact, negative-stained R. typhi (see Sears et al. (Sears et al., 2012) for further information). Scale bar = 0. 25 µm. (b) Conserved sequence spanning the L (red) and β (blue) domains of 12 ATs from Gram-negative bacterial species (above line) and a representative Sca family from five Rickettsia species (below line). Full species names and sequence information is provided in Table S2. Yellow depicts residues conserved across all sequences, with the dashed red box depicting the sequence shown in (c). Sequences aligned with muscle v3.6 (Edgar, 2004) using default parameters. Numbers in red depict regions of the alignment not shown. (c) Predicted structure of the β domain (12 β-strands in yellow) and partial L domain (intrabarrel α-helix in purple) of R. rickettsii Sca0. Modeling done using phyre v.2.0 (Kelley and Sternberg, 2009), with the Pseudomonas aeruginosa EstA AT (PDB no. c3kvnA) selected for comparison (97% coverage, 100% confidence). The acidic patch (GDED) proximal to the first β-strand is within a red dashed box in the ‘bottom view’ at right. Structure visualized with jmol v.13.0.15 (Herraez, 2006). (d) Conservation of the sequence illustrated in (b) across 336 Rickettsia Scas. Individual Sca families are shown (Sca0-3, Sca5) as well as all families together (all). Sequence logos generated using weblogo v.3.3 (Crooks et al., 2004).
The TSSSa of <i>Rickettsia</i> spp. is comprised entirely of the superfamiliy of surface cell antigens (Scas). While a similar nomenclature has recently been applied to ATs of the closely related scrub typhus agents (<i>Orientia tsutsugamushi</i>) (Ha et al., 2011), there is no phylogenetic basis for common ancestry of these two TSSSa, with neither of their evolutionary origins currently known. Nearly a decade ago, phylogenomics was employed to name 17 distinct Sca families encoded across nine <i>Rickettsia</i> genomes (Blanc et al., 2005b). Recent analyses based on dozens more genomes suggest that only five families are conserved in most rickettsial genomes: Sca0 (formerly Rompa), Sca1, Sca2, Sca4 (formerly PS 120 or protein D), and Sca5 (formerly RompB) (Gillespie et al., 2012b; Sears et al., 2012).

Interestingly, Sca3 proteins are encoded only within the genomes of typhus group (TG) rickettsiae and <i>R. felis</i>, all of which are primarily associated with insect vectors (to the exclusion of other rickettsial species primarily associated with ticks or mites). Sca3 proteins are among the largest Scas, ranging from 2050 aa (<i>R. typhi</i>) to 2841 aa (<i>R. felis</i>), yet other than their expression and surface localization in <i>R. typhi</i> (Sears et al., 2012), nothing is known regarding their function.

Of the five conserved Sca proteins, four have been determined to function in adherence to host cells. Sca0 of <i>R. rickettsii</i> was previously demonstrated to be essential for host cell adhesion (Li and Walker, 1998), and more recently, Sca0 of <i>R. conorii</i> was shown to interact with α2β1 integrins of endothelial cells (Hillman et al., 2013). Sca1 of <i>R. conorii</i> also mediates adhesion to various types of epithelial and endothelial cells, although it alone cannot facilitate host cell entry (Riley et al., 2010). By contrast, Sca2 of <i>R. conorii</i> mediates both adhesion and invasion of cultured mammalian cells (Cardwell and Martinez, 2009). Sca5 is the predominant protein of the rickettsial S-layer, comprising 15% of the total protein mass (Ching et al., 1990; Hahn and Chang, 1996). Sca5 of <i>R. conorii</i> mediates both adhesion and invasion of mammalian cells (Uchiyama et al., 2006; Chan et al., 2009), with the Ku70 subunit of nuclear DNA-dependent protein kinase identified as the primary host receptor (Martinez et al., 2005). The downstream internalization process resulting from this interaction was shown to be clathrin and caveolin 2 dependent and also involves Ku70 ubiquitination (by the ubiquitin ligase c-Cbl) and host actin polymerization (HAP) (Chan et al., 2009).

Conserved in all rickettsial genomes, Sca5 is likely an essential rickettsial protein. This is supported by <i>R. conorii</i> death in cells pretreated with anti-Sca5 antibodies (Feng et al., 2004a, b) and also through the demonstration that Sca5 antibody-mediated killing confers immunity to <i>R. conorii</i> infection (Chan et al., 2011). Recently, <i>R. conorii</i> Sca5 binding of a host complement inhibitor (factor H) has been implicated in rickettsial evasion of host complement-mediated clearance from the hematogenous circulation, with the β domain alone sufficient to facilitate this resistance (Riley et al., 2012).

Rickettsial invasion of host cells involves HAP. For instance, during <i>R. conorii</i> host cell entry, activation of the Arp2/3 complex results from a complexity of host signaling interactions involving (at least) the GTPase Cdc42, phosphoinositide 3-kinase, and c-Src (Martinez and Cossart, 2004; Rydkina et al., 2008). As noted above, Sca5 binding of host receptor Ku70 triggers this signaling cascade (Chan et al., 2009), yet other Scas are now known to directly engage host actin and other cytoskeletal elements, although in Arp2/3-independent mechanisms. Scas2 proteins of <i>R. parkeri</i> (Haglund et al., 2010) and <i>R. rickettsii</i> (Kleba et al., 2010) have been shown to play a role in HAP, with sequence analysis suggesting these proteins mimic eukaryotic formins. For <i>R. parkeri</i>, Sca2 was demonstrated to nucleate unbranched actin filaments and associate with extending barbed ends of actin, while requiring host profilin and inhibiting host capping protein (Haglund et al., 2010). The basis for this formin mimicry was recently discovered, as a crystal structure of the <i>R. conorii</i> formin homology 2 (FH2) domain revealed a crescent-like fold similar to half of the dimeric structure of formin (Madasu et al., 2013). Together with a similar structure in the CT region of the Sca2 passenger, a doughnut-shaped structure analogous to formins is predicted to elongate actin, with intervening Pro-rich regions (PRRs) and Wiskott-Aldrich syndrome protein (WASP) homology 2 (WH2) domains incorporating profilin-actin for elongation and recruiting actin monomers for nucleation, respectively (Madasu et al., 2013). Importantly, while conserved in most spotted fever group (SFG) rickettsiae Sca2 proteins, the FH2 domains contain deletions or are entirely absent in Sca2 proteins of rickettsial species that either form minimal actin tails (<i>R. typhi</i>) or entirely lack actin-based motility (ABM) (<i>R. canadensis</i>, <i>R. peacockii</i>). Regarding <i>R. bellii</i>, which does generate actin tails similar to most SFG rickettsiae (Oliver et al., 2014), its Sca2 protein also lacks an FH2 domain, suggesting it either directly promotes nucleation of the Arp2/3 complex or is not the factor responsible for ABM.

Recently, <i>R. rickettsii</i> Sca4 was shown to bind and activate host vinculin (Park et al., 2011), a membrane-cytoskeletal protein associated with focal adhesions that anchors F-actin to the membrane. Specifically, generated structures for Sca4 revealed two small domains that structurally resemble vinculin-binding sites of talin, a protein that activates vinculin (Park et al., 2011; Lee et al., 2013). These two vinculin-binding sites are conserved across most rickettsial Sca4 proteins from both pathogenic and nonpathogenic <i>Rickettsia</i> spp. (Supporting Information, Fig. S1), with several exceptions found in the CT vinculin-binding sites of <i>R. bellii</i> strains (highly divergent), <i>R. typhi</i> strains (missing several residues), and <i>R. canadensis</i> strains (degraded). This suggests that this mode of host cytoskeletal interaction is a highly conserved process for most <i>Rickettsia</i> spp. It is unclear how such an important surface protein is translocated out of the rickettsial cell, as all known Sca4 proteins lack the β domain. This, coupled with no recognizable homology between Sca4 proteins and other Sca families, makes inclusion of Sca4 proteins within the Sca superfamily tenuous.

Despite a plethora of research on the extracellular localization and interaction with eukaryotic molecules, knowledge on the method of Sca secretion is very limited. While expected to encode Sec-processed SSs, in silico algorithms do not predict strong SSs for most of the Scas; however, the NT sequences do contain hydrophobic patches, which in some cases are adjacent to charged regions (data not shown). None of the Scas exhibit NT sequences similar to the extended SSs identified for a subset of TSSSa proteins that have been suggested to delay SS cleavage to allow chaperone contact with the passenger, hence preventing premature folding (Sijbrandi et al., 2003; Szabady et al., 2005; Desvaux et al., 2006, 2007). In support of Sec-processing of these unknown Sca SSs, we recently demonstrated that the proximal 65 NT residues of four <i>R. typhi</i> Scas (Sca1-Sca3, Sca5) were capable of translocating E. coli alkaline phosphatase (PhoA) to the PP in the PhoA fusion assay (Ammerman et al., 2008). Thus, while the nature of this assay did not reveal the exact cleavage sites, somewhere within these proximal NT residues reside sufficient Sec SSs. Further studies are needed to identify the Sca SSs, as well as to determine the chaperones that engage translocated Scas in the PP before subsequent localization to the OM.

Evolutionary analyses indicate that Scas are evolving under positive selection, with evidence of recombination.
suggestions that these antigens are adapting to host selective pressures (Blanc et al., 2005b; Jiggins, 2006; Sears et al., 2012b). The presence of Scas-encoding genes on some rickettsial plasmids indicates a source for recombination across species (Baldridge et al., 2008, 2010; Gillespie et al., 2012), potentially equipping plasmid recipient strains with novel Scas repertoires (Sears et al., 2012). Consistent with numerous studies showing either anti-Sca antibody inhibition of infection or Sca secretion by heterologous systems, recent reports have demonstrated Scas on the surface of rickettsial cells (Haglund et al., 2010; Sears et al., 2012). However, while some well-studied ATs are processed (by a variety of means) to release the passenger domain from the bacterial surface (Leyton et al., 2012), little is known regarding Sca secretion into the extracellular milieu (host cytoplasm). To our knowledge, only Scas of some Rickettsia species has been shown to be processed into two polypeptides resembling the sizes of the passenger and β domains (Gilmore et al., 1991; Renesto et al., 2006), although the processed proteins remain covalently associated with the bacterial cell surface (Hackstadt et al., 1992). Our recent work on surface localization of R. typhi Scas revealed the immunogold labeling of host cell cytoplasm and chromatin (see Fig. 4a, inset), suggesting that Scas may indeed be released into the host cytosol and potentially localized to specific cellular compartments (Sears et al., 2012).

We evaluated the possibility that, despite the extraordinary divergence of their passenger domains, all conserved Scas families may contain a signature in the β domain that mediates their processing. Comparison of Sca sequences spanning the region between the passenger and β domains revealed a linker domain (L) that shares characteristics with the autocatalytic region of serine protease autotransporters of the Enterobacteriaceae (SPATEs), the pertactin family of ATs, and some other ATs (Fig. 4b). In SPATEs, this L domain comprises the α-helix that is inserted within the β domain, with intrabarrel autoproteolytic cleavage of the α-helix occurring midway between the PP and OM boundaries of the β-barrel (Kostakioti and Stathopoulos, 2006; Dautin et al., 2007; Yen et al., 2008). Mutagenesis and structural studies have revealed the participation of residues within both the α-helix and β-barrel lumen that function in proteolysis (Barnard et al., 2007; Tajima et al., 2010; Yen et al., 2010; Barnard et al., 2012), and mutation of several of these residues suggests they are essential for cleavage, but not secretion (Dautin and Bernstein, 2011). Related to SPATEs and other similar ATs, this region within Sca proteins contains an acidic patch adjacent to the first strand of the β-barrel, forming a loop in the PP that may be important for proper orientation of the α-helix within the β-barrel (Fig. 4c). An analysis of 336 Sca proteins from five families (Sca0, Sca1-3, Sca5) indicates that this acidic patch is highly conserved (Fig. 4d). Additionally, three Sca families (Sca0, Sca1, Sca3) have one or more Asn residues at the autocleavage site within most SPATEs and related ATs that are cleaved at a conserved site typically containing Asn residues. The lack of Asn conservation in Sca2 and Sca5 families may indicate these ATs are not processed (Sca2) or processed by another mechanism (Sca2, Sca5). However, cleavage of Scas may have a negligible effect on extracellular localization and function if processed passenger domains remain closely associated with their cognate β domains in the rickettsial S-layer, as has been shown for R. rickettsii (Hackstadt et al., 1992). Such a case is well characterized for the AIDA-I AT of E. coli (Charbonneau et al., 2009).

Determining those Scas that are processed, and if so, the nature of the underlying mechanism(s), is an important area of future research. Specifically, it is critical to determine whether the β domains of processed Scas can be used to translocate the passenger domains of other Scas lacking β domains (e.g. Sca4, and other less conserved Scas, such as Sca9), or even serve as OM channels for secretion of other proteins. Such stable β-barrel proteins are likely prevalent in the rickettsial OM, given the abundance of Scas at the bacterial surface (Ching et al., 1990; Hahn and Chang, 1996), and probably would only turn over during cell growth and division (Grijpstra et al., 2013). Indeed, families of proteins comprising solitary AT-like β domains have been identified in some proteobacterial genomes (Prakash et al., 2011). However, no genes encoding these proteins are present in Rickettsia genomes; thus, it is likely that Sca4 is secreted via another secretion pathway, with other less conserved Scas lacking β domains probable pseudogenes. Supporting this is the fact that autoproteolysis of EspP (and probably other SPATEs and related ATs depicted in Fig. 4b) results in the β-helix occluding the β-domain channel (Barnard et al., 2007). Furthermore, secondary functions of Sca β domains, such as binding of host serum factor H by R. conorii (Riley et al., 2012), may be compromised by additional roles in protein translocation across the OM. Nonetheless, given the minimal size and gene number of Rickettsia genomes, it is necessary to assume multiple functions for conserved proteins in the face of a lack of latent redundancy that is typical of larger bacterial genomes.

**Sec-TolC pathway**

Aside from the T5SS, four other Gram-negative bacterial secretion systems also rely on the Sec translocon for export of proteins to the PP. For the T2SS, energy from an IM ATPase extends pseudopili structures that possibly push Sec-processed protein substrates out of the PP through an OM channel complex (Korotkov et al., 2012). The T7SS (chaperone-usher pathway) requires a periplasmic chaperone (e.g. PapD, FimC) and an OM usher protein (e.g. PapC, FimD) for the assembly and secretion of surface pili (Dodson et al., 1993; Thanassi et al., 1998). The T8SS, previously referred to as the extracellular nucleation-precipitation pathway (Stathopoulos et al., 2000; Kostakioti et al., 2005), is defined by Sec-processed curli subunits (CsgA, CsgB) that are translocated across the OM in a poorly understood process, but one strictly dependent on an associated lipoprotein (CsgG) (Loferer et al., 1997). Finally, the T9SS (Por secretion system) secretes adhesins involved in gliding motility (Sato et al., 2010; McBride and Zhu, 2013) as well as other substrates with a wide range of functions (Sato et al., 2013). Collectively, as opposed to ‘one-step’ pathways that secrete substrates directly from the cytoplasm to the extracellular milieu (i.e. T1SS, T3SS, T4SS, T6SS), these secretion mechanisms can be considered ‘two-step’ pathways that are conserved in their dependency on the Sec translocon for IM translocation of substrates (Rego et al., 2010).

Examples within two Sec-independent secretion systems illustrate ‘exceptions to the rule’ of one- and two-step mechanisms defining entire secretion system groups. Perhaps, the epitome of this is the specialized ptl T4SS of Bordetella pertussis, which secretes a sole substrate, the pertussis toxin (Carbonetti, 2010). The ptl T4SS lacks a gene encoding the cytosolic substrate recognition particle (TraG/TrwB/VirD4 family) present in most other T4SSs, and the pertussis toxin subunits are exported to the PP via the Sec translocon (Locht and Keith, 1986; Nicosia et al., 1986). It is predicted that holotoxin formation in the PP drives the assembly of the core PtI scaffold components around pertussis toxin prior to secretion, with ATP hydrolysis providing the energy for secretion (Verma and Burns, 1986).
Like T4SSs, the substrates of T1SSs typically lack NT SSs and are secreted directly across the cell envelope (discussed below). However, there are a few cases in which Sec-translocated proteins in the PP are secreted extracellularly via the T1SS. Heat-stable enterotoxins STI and STII produced by enteropathogenic E. coli (EAEC) are exported to the PP via the Sec translocon, yet are secreted from the bacterial cell through a channel formed by TolC (Foreman et al., 1995; Yamanaka et al., 1997, 1998). STII secretion was further shown to involve an ATP-binding cassette (ABC) transporter and membrane fusion protein (MFP) (MacA/B) that, together with TolC, form a canonical T1SS (Yamanaka et al., 2008). Another EAEC protein, named dispersin (Aap), was also shown to be exported to the PP by the Sec translocon, yet extracellular secretion was strongly dependent on a cluster of genes (aat) encoding the regulatory components of a T1SS (Nishi et al., 2003).

Similar to the secretion pathways described above for EAEC STI, STII, and Aap, we recently demonstrated that the extracellular secretion of a protein by R. typhi requires both the Sec translocon and TolC (Kaur et al., 2012). Secrecion of Rickettsia ankyrin repeat protein 1 (RARP-1) into the host cytoplasm was shown during R. typhi in vitro infection of mammalian cells (Fig. 5a). Recombinant wild-type RARP-1 was also secreted extracellularly in a surrogate host, E. coli str. C500. To understand the RARP-1 secretion mechanism, recombinant proteins truncated at their N- (RARP-1NT) and C- (RARP-1CT) termini were monitored for their secretion in E. coli str. C600; surprisingly, both truncated proteins were not observed in the extracellular milieu. The lack of RARP-1NT secretion is consistent with our previous report showing a functional Sec SS within the NT of RARP-1 (Ammerman et al., 2008). The contribution to the secretion process of the RARP-1 CT, which contains three contiguous ankyrin (ANK) repeats and no clear T1SS SS, is currently unknown. Furthermore, secretion of RARP-1 was abolished in an E. coli str. C600 isogenic tolC insertion mutant. Importantly, expression of R. typhi tolC in the E. coli tolC mutant restored wild-type RARP-1 secretion, bolstering the role of TolC in RARP-1 translocation across the OM.

While evidence supports the requirement of TolC, as well the complete N- and C-termini of RARP-1 for its efficient secretion, the specific mechanics of the secretion pathway remain tenuous. After being exported by the Sec translocon, RARP-1 likely engages the tips of the α-helical domain of TolC in the PP. From this point, two possible secretion routes can be considered. Firstly, RARP-1 might exit the trimeric TolC channel independently of any additional factors, thus making OM translocation an ATP-independent process (Fig. 5b-1). Support of this model comes from the requirement for the CT ANK repeats in the secretion process. As RARP-1 is already in the PP and processed at its required NT, the CTD may be required not only for docking the protein to TolC, but also for facilitating its extracellular secretion. Interestingly, the ANK domains are highly conserved across all Rickettsia RARP-1 homologs, with other regions of the protein exhibiting extreme levels of divergence and additional repeats (Kaur et al., 2012). Thus, the RARP-1 ANK repeats may have a conserved role in secretion, with more variable regions of the protein interacting with host cell molecules. Notwithstanding, ruling against this model of RARP-1 secretion is the fact that, to our knowledge, secretion of protein substrates through TolC has never been demonstrated to occur independently of cytosolic ATP hydrolysis. Furthermore, of the two-step secretion pathways outlined above, only T5SSs, the T7SS, and the T9SS function independent of ATP hydrolysis in the OM translocation of substrates (Thanassi et al., 2005; Leo et al., 2012). While largely unknown, the energy-independent translocation mechanics of these systems [i.e. passenger domain folding for T5SSs (Pohlner et al., 1987), pilin subunit assembly on an OM platform for the T7SS (Remaut et al., 2008), curli subunit nucleation for the T8SS (Hammer et al., 2007), adhesin secretion for the T9SS (Sato et al., 2010)] seem to share a common theme in that secretion products are anchored to the OM (although some T9SS substrates are released from the OM (Sato et al., 2013)). Furthermore, none of these secretion machines have channel tunnels extending into the PP such as TolC. As we primarily detected RARP-1 of R. typhi in the cytoplasm of host cells, it is probable...
that a source of energy is required to push the substrate through the TolC channel and disassociate the protein from the bacterial OM.

Alternatively, like the secretion pathways for EAEC STII and Aap, as well as PT, RARP-1 localization to the PP may induce the assembly of a secretion channel that engulfs the protein and bridges ATP hydrolysis at the bacterial cytoplasm with an OM translocation pore. Specifically, this mechanism of secretion would require an ABC transporter and MFP to provide energy to push RARP-1 through the TolC channel (Fig. Sb-2). As Rickettsia genomes encode components of a T1SS (discussed below), this model of secretion is reasonable and invokes the most probable source of energy required to translocate RARP-1 across the OM. However, unpublished data from our laboratory indicate that expression of R. typhi T1SS genes (aprE and aprD), along with RARP-1, fails to increase the amount of RARP-1 secretion in the wild-type E. coli strain C600. This may indicate that E. coli strain C600 is not a suitable expression system for assembly of the Rickettsia T1SS. Furthermore, while we clearly showed the ability of R. typhi TolC to functionally complement an E. coli tolC mutant strain, growth was abolished in SDS-containing medium, suggesting R. typhi TolC cannot interact with the AcrA/AcrB efflux complex of E. coli (Kaur et al., 2012). In vivo studies, as well as more direct experiments designed to complement an apposite T1SS mutant strain, are needed to determine the energy requirements (if any) for RARP-1 secretion.

Collectively, our work implies that TolC is involved in the secretion process of a Sec-processed substrate, RARP-1. Given the numerous hypothetical proteins detected by the PhoA fusion assay (Fig. 3a), it is likely that additional Sec-processed proteins are secreted extracellularly via the TolC channel. The contribution of T1SS ABC transporter and MFP components to the secretion of RARP-1, as well as other potential substrates of the Sec-TolC pathway, remains unclear and will be critical to determine in future studies.

**SEC-INDEPENDENT SECRETORY PATHWAYS**

**Tat system**

In contrast to the export of unstructured polypeptides by the Sec translocon and or YidC, the Tat system in Gram-negative bacteria translocates folded substrates from the cytoplasm to the IM or PP (Frobel et al., 2012a, b; Palmer and Berks, 2012). Tat systems are encoded in the genomes of most bacteria and some Archaea and are also present in many plastid-containing eukaryotic genomes, wherein primary Tat function is in export across the thylakoid membrane (Palmer and Berks, 2012). At its core, the Tat system across these diverse phylogenetic lineages is comprised of two proteins: (1) TatA, a small protein with an NT TMS helix and an amphipathic helix closely associated with the lipid bilayer (Hu et al., 2010; Walther et al., 2010, 2013); and (2) TatC, a polytopic protein with six TMS regions arranged in curved-wall conformation capped by two periplasmic loops (Punginelli et al., 2007; Rollauer et al., 2012). Many Tat systems of Gram-negative bacteria, high-%GC Gram-positive bacteria, and plant chloroplasts contain a third protein, TatB, that is homologous to TatA in domain architecture (Bogsch et al., 1998; Sargent et al., 1998, 1999). Despite negligible sequence identity, the structural similarities shared by TatA and TatB were suggested to have arisen from multiple ancient gene duplication events (Yen et al., 2002). More evolutionarily, recent gene duplications of TatA [e.g. TatE of E. coli (Sargent et al., 1998)], as well as TatC (van Dijl et al., 2002; Jongbloed et al., 2004), along with considerable flexibility in genomic arrangement of Tat genes, account for substantial diversity of Tat gene repertoires across organisms encoding these secretion systems (Palmer and Berks, 2012). TatD, once considered a Tat system component, is a widespread deoxyribonuclease that probably functions independent of Tat systems (Wexler et al., 2000).

Most characterized Tat system substrates have an NT SS with a tripartite structure similar to the Sec SS, yet are distinguishable by three characteristics: (1) a consensus ‘twin-arginine’ motif, (S/T)-R-R-x-F-L-K, at the distal end of the basic region (Berks, 1996); (2) an overall lower hydrophobicity across the entire signal peptide (Cristobal et al., 1999; Ize et al., 2002); and (3) a propensity for positively charged residues flanking the signal cleavage motif (Bogsch et al., 1997; Ize et al., 2002; Blaudeck et al., 2003; Tullman-Ercek et al., 2007). The Tat SS of substrates is recognized by TatC (Holzapfel et al., 2007; Rollauer et al., 2012; Zoufaly et al., 2012), with the dynamic process of translocation in Gram-negative bacteria involving multiple TatCB complexes in conjunction with pore-forming TatA oligomers and the proton motive force (Palmer and Berks, 2012; Kudva et al., 2013). The number and type of Tat system substrates vary extensively across organisms (Rose et al., 2002; van Dijl et al., 2002; Dilsk et al., 2003; Palmer and Berks, 2012). Many characterized Tat system substrates incorporate cofactors prior to translocation, with some substrates translocated as oligomers (Halbig et al., 1999; Sanders et al., 2001; Delisa et al., 2003). Single components of translocated oligomers may lack SSs, and thus ‘hitchhike’ with their SS-containing partners in the translocation process (Rodrigue et al., 1999; Sambasivarao et al., 2000). Still, some monomeric proteins lacking cofactors are translocated by the Tat system due to their rapid folding kinetics, a prevalent condition characteristic of halophilic Archaea (Rose et al., 2002; Hutcheon and Bolhuis, 2003).

Previous genomic analyses of Tat systems have reported that, relative to most other Gram-negative bacteria, Rickettsia spp. lack a gene encoding the TatB protein (Dilsk et al., 2003; Nunez et al., 2012). A study on the Tat system of Anaplasma marginale revealed the expression of tatA, tatB, and tatC, as well as the ability of recombinant TatA and TatB (but not TatC) to restore Tat function in E. coli tat mutant strains (Nunez et al., 2012). TatB homologs were also reported for Anaplasma phagocytophilum and Ehrlichia spp., but not for the remaining genera of Rickettsiales (Wolbachia, Neorickettsia, Orientia, Rickettsia). Bacterial TatB proteins are extraordinarily variable; for instance, the CT domain has been shown to be dispensable (Lee et al., 2002), and no mutations thus far attempted have a negative effect on TatB function (Barrett et al., 2003; Hicks et al., 2003; Lee et al., 2006; Maldonado et al., 2011). Despite this, deletion of tatB in E. coli abolishes Tat function (Sargent et al., 1998, 1999). Given this, we hypothesized that tatB genes are probably encoded within all Rickettsiales genomes.

Our analysis identified putative TatB proteins encoded within all sequenced Rickettsiales genomes, with the Rickettsia proteins previously assigned to DUF2672 at NCBI (Fig. 6a). Like other bacterial TatB homologs (Hicks et al., 2003; Lee et al., 2006), these rickettsial TatB proteins are (1) small in length (88–146 aa), (2) strongly hydrophobic in the NT region, and (3) most conserved in the ‘hinge’ between the NT TMS region and the amphipathic helix. This suggests that all members of the Rickettsiales likely encode complete Tat systems consisting of single genes encoding TatA, TatB, and TatC (Fig. 6b). Interestingly, the Tat genes are arrayed contiguously in ‘Candidatus Odyssella thessalonicensis’ (Holosporaceae, hereafter Odyssella), with TatB and TatC genes adjacent in ‘Candidatus
Figure 6. Characteristics of the Rickettsia twin-arginine translocation (Tat) system. Annotated Rickettsia genomes typically include genes predicted to encode TatA and TatC, but not TatB. Prior analyses focusing on bacteria (Dilks et al., 2003) and Alphaproteobacteria (Nunez et al., 2012) reported Tat systems of Rickettsia spp. lacking TatB proteins. (a) In silico prediction of Rickettsia TatB homologs. A series of blastp searches against Rickettsia genomes using diverse TatB sequences as queries resulted in the identification of putative TatB proteins. These proteins were previously assigned to DUF2672 at NCBI. Additionally, putative TatB homologs from other Rickettsiales genera also lacking TatB annotations were identified, with all rickettsial sequences compiled with 33 bacterial TatB proteins and aligned with muscle v3.6 (Edgar, 2004) using default parameters. The domain architecture of TatB is shown at top (NT TMS region in red, amphipathic helix in blue). Numbers in red depict regions of the alignment not shown, with black numbers at right depicting total protein sizes (aa). Sequences with predicted TMS regions [tMHMM v.2.0 (Krogh et al., 2001)] are colored light blue. Conservation across the short hinge region as follows: conserved Gly, green; polar residues, gray; acidic residues, red; and basic residues, blue. Taxon names are colored as follows: four nonproteobacterial species (Dehalococcoides sp. str. CBDB1, Chloroflexi; Thermobifida fusca strain YX and Dietzia alimentaria strain 72, Actinobacteria; and Thermodesulfovibrio yellowstonii strain ATCC 51303, Nitrospirae), gray; Epsilonproteobacteria, light blue; Deltaproteobacteria, purple; Betaproteobacteria, green; Gammaproteobacteria, dark blue; and Alphaproteobacteria, black. Rickettsiales taxa are within the shaded box. Full species names and sequence information is provided in Table S3. Red circles denote taxa further analyzed in (b) and (c). (b) Genomic arrangement of genes encoding Tat system components in genera of Rickettsiales. Phylogeny is redrawn from Driscoll et al. (Driscoll et al., 2013), with select species noted in (a) by red circles. Gene models are colored according to the Tat system illustrated in Fig. 1. Sequence information is provided in Table S4. (c) Characteristics of a putative rickettsial Tat substrate. Sequences are from species shown in (b). Aside from 'Candidatus Midichloria mitochondrii', the NT sequences of ubiquinol-cytochrome c reductase iron-sulfur subunit (PetA) across Rickettsiales genera possess a predicted Tat signal peptide. The ‘twin-arginine’ motif (shaded purple) is merged with a predicted TMS region (light blue), consistent with other characterized PetA proteins that contain a noncleavable leader sequence within a TMS region (Molik et al., 2001; Bachmann et al., 2006; De Buck et al., 2007; Aldridge et al., 2008). Numbers in red depict regions of the alignment not shown. Residues colored yellow are conserved across all rickettsial PetA proteins, including the Midichloria homolog lacking the Tat signal peptide. Sequence information is provided in Table S5.
Midichloria mitochondrii’ (Midichloriaceae, hereafter Midichloria).

The Tat genes in the derived rickettsial families Anaplasmataceae
and Rickettsiaceae are scattered throughout the genome, reflectively
of reductive genome evolution eliminating operon structure
(Andersson et al., 1998) and reminiscent of the scattered distri-
bution of T4SS genes within these genomes (Gillespie et al., 2009,
2010).

Despite the presence of Tat systems in Rickettsiales, previous
bioinformatics analyses have identified very few predicted Tat
substrates (Dilks et al., 2003; Nunez et al., 2012). In support of this,
our own examination of in silico approaches (Rose et al., 2002;
Bendtsen et al., 2005) across Rickettsia genomes identified only
one conserved protein containing a putative Tat signal pep-
tide, the ubiquinol-cytochrome c reductase (cytochrome b-c1
complex), or Rieske protein (PetA). PetA is translocated by the
Tat system in plastids and some bacteria, with noncleavable
leader sequences part of an NT TMS region that anchors PetA
to membranes (Molik et al., 2001; Bachmann et al., 2006; De
Buck et al., 2007; Aldridge et al., 2008). Aside from M. mitochondrii,
the NT sequences of PetA across Rickettsiales genera possess a
predicted Tat signal peptide that is merged with a probable TMS
region (Fig. 6c). The lack of an NT TMS region or predicted Tat
signal peptide for M. mitochondrii PetA may coincide with an
alternative respiration (e.g. heme copper oxidases of type cb3,
but not cbbb) employed by this mitochondria-associated organism
(Sassera et al., 2011). As TatC proteins are encoded within mito-
ochondrial genomes of plants and most protists (Yen et al., 2002),
it has been posited that TatC plays a role in assembly of PetA in
these organisms, reflecting the ancient role of Tat in assembly of
respiration complexes (Hinsley et al., 2001). For Rickettsia spp.,
it appears that its Tat system has been retained solely for the
assembly of PetA into the IM.

The paucity of identifiable Tat substrates in Rickettsia may be a
consequence of reductive evolution pruning many genes
involved in processes that utilize Tat transport in other bacte-
ria. While hard to envision the retention of an entire secretion
system for the translocation of a single substrate (PetA), it has
previously been shown that Tat substrates vary substantially in
number across bacteria, with minimal substrates predicted for
other obligate intracellular bacterial species (Dilks et al., 2003).
Still, as discussed above for Sec translocon substrates, it is pos-
sible that the base compositional bias of Rickettsia genomes has
obscured the Tat signal peptide motif and that the repertoire of
Tat substrates is larger. For many bacterial pathogens, the Tat
pathway has been implicated in virulence (De Buck et al., 2008),
with direct roles in the export of virulence factors known for
some species, for example, Agrobacterium tumefaciens (Ding and
Christie, 2003), Legionella pneumophila (Rossier and Cianciozzo,
2005), Mycobacterium smegmatis (McDonough et al., 2005), Pseu-
domonas aeruginosa (Ochsner et al., 2002), and enterohemorrhagic
Escherichia coli O157:H7 (Pradel et al., 2003). For F. aeruginosa
(Voulhoux et al., 2001) and possibly Vibrio cholerae during anae-
robic growth (Lee et al., 2012a), some T2SS effectors are secreted
in a Tat-dependent manner. For Rickettsia spp., such a pathway
would seemingly involve TolC or some other outer membrane
porin large enough to support the export of folded substrates
(Fig. 1). Future research is needed to understand the limits of the
Tat system for protein secretion in these bacteria.

Type I secretion system

T1SSs span the entire Gram-negative cell envelope and con-
sist of three protein components: an IM ABC transporter, an
IM/periplasmic MFP (also known as an adaptor protein), and
an outer membrane efflux protein (OEP) of the TolC family
(Delepelaire, 2004; Holland et al., 2005). The secretion machine
exports virulence-associated proteins (e.g. toxins, lipases, proteases, etc.)
from the cytoplasm to extracellular milieu in a so-called one-step process (Zgruskaya et al., 2011). Specif-
ically, the MFP links the ABC transporter and OEP compo-
nents, with the former providing energy (via ATP hydrolysis) to
drive substrates through the channel and the latter providing a
gaited pore that regulates the extracellular release of substrates
(Holland et al., 2005). Within the T1SS tripartite complex, the
OEP and ABC transporter function as trimers and dimers, re-
spectively, with an unknown number of MFP subunits bridg-
ing the asymmetry between OEP and ABC transporter oligomers
(Lee et al., 2012b).

While the T1SS MFP/ABC transporter complex determines substrate specificity (Akatsuka et al., 1997), the OEP compo-
nent can couple with several different IM/periplasmic proteins
to transport a variety of molecules (Delepelaire, 2004). This
is fortified by the recent identification of a TolC-binding re-
gion within diverse MFPs from T1SSs and drug efflux systems
(Lee et al., 2012b). Hence, aside from their cognate T1SS com-
plexes, OEM components may function in MFP-dependent or
MFP-independent complexes (e.g. other ABC transporters, re-
sistance nodulation division transporters, and transporters of
the major facilitator superfamily) that expel antibiotics, deter-
gents, and organic solvents from the cytoplasm to extracellular
milieu (Zgruskaya et al., 2011). For example, in uropathogenic
E. coli, TolC interacts with the AcrA/B and MacA/B com-
plexes to constitute drug efflux pumps (acriflavine, macrolides),
yet also engages the HlyB/D complex for secretion of the
α-hemolysin toxin (Koronakis, 2003; Koronakis et al., 2004;
Symmons et al., 2009). The number of OEM protein-encoding
genes, as well as genes encoding MFP/ABC transporter pairs,
varies extensively across Gram-negative bacterial genomes
(Ma et al., 2003; Barabote et al., 2007; Youm and Saier, 2012),
suggesting diverse strategies for combining the multifunctional
OEM proteins with MFP transporter complexes. Indeed, while the
genes encoding MFP/ABC transporter pairs are often within
operons, the OEM-encoding genes tend to be separate from
these loci (Thomas et al., 2014).

The mechanism by which substrates interact with the T1SS
is poorly understood (Holland et al., 2005). Substrates are gener-
ally considered to lack NT cleavable signal peptides, with poten-
tial SSs located within their CT. However, T1SS substrates
can possess either NT or CT SSs, with their cognate ABC trans-
porters (and possibly MFPs) containing domain architectures
designed to accommodate particular SSs (Kanonenberg et al.,
2013; Lenders et al., 2013; Thomas et al., 2014). The ABC transporters
of T1SSs have recently been categorized into three groups
(Kanonenberg et al., 2013). Group 1 transporters contain a C39
cysteine protease domain that processes NT SSs of small sub-
strates (e.g. bacteriocins and microcins) (Havvarstein et al., 1995;
Wu and Tai, 2004; Nishie et al., 2009; Wu et al., 2012). Group 2
transporters contain a C39-like domain, which is proteolytically
inactive due to a conserved Tyr residue replacing Cys within
the peptidase active site (Kotake et al., 2008; Ishii et al., 2010;
Lecher et al., 2012). Substrates engaging Group 2 transporters
(e.g. toxins HlyA and CyaA) contain CT SSs instead NT SSs,
yet the C39-like domain is not involved in SS processing, but
may serve as a chaperone to prevent premature folding of
large substrates (Mackman et al., 1985; Lecher et al., 2012). For
HlyA, its CT SS has been demonstrated to interact with the
nucleotide-binding domain (NBD) of the HlyB ABC transporter
(Ben Abdelhak et al., 2003). Finally, Group 3 transporters lack any
additional domains outside of the NBD and TMS domain that define ABC transporters (Davidson et al., 2008), with all characterized substrates (e.g. alkaline protease AprA and hemophore HasA) containing CT SSs (Delepeleire, 2004).

Rickettsia genomes encode OEP, MFP, and ABC transporter proteins that likely assemble into a functional T1SS (Fig. 7a). The OEP component, annotated as ToIC, has a predicted structure similar to E. coli TolC, with strict conservation of essential residues involved in channel formation and regulation (Kaur et al., 2012). As no other OEP proteins are encoded within Rickettsia genomes, TolC is likely bifunctional, participating in protein secretion (Kaur et al., 2012) and drug efflux pumps. This is supported by the presence of dozens of genes encoding components of non-T1SS transporters (Gillespie et al., 2008). A sole module encoding a T1SS MFP/ABC transporter complex is also highly conserved across Rickettsia genomes, suggesting that Rickettsia spp. contain only one T1SS. These MFP and ABC transporter proteins are most similar to alkaline protease (Apr)-like T1SS components, for which the AprD proteins are Group 3 ABC transporters (Fig. 7b). Thus, Rickettsia AprD proteins lack NT appendices, with structural arrangement of the TMS region and NBD consistent with classical ABC transporters (Fig. 7c). The AprE proteins contain unique NT sequences flanking their predicted TMS regions that are longer than most Group 3 ABC transporter-associated MFPs (data not shown). Comparison with AprE homologs from other Rickettsiae genomes reveals that these proximal NT sequences are highly conserved within rickettsial genera, but extraordinarily divergent across genera (Fig. 7d, Fig. S2). While studies exist in favor of (Letoffe et al., 1996) and against (Balakrishnan et al., 2001) the interaction of MFPs with substrates, these genera-specific NT sequences of rickettsial AprE proteins may play a role, together with AprD, in substrate recognition/processing in the bacterial cytosol.

Phylogeny estimation of rickettsial MFP/ABC transporter pairs suggests a common origin for these genes, with vertical inheritance from an alphaproteobacterial ancestor (Fig. 7e). An exception is seen for Odyssella, which encodes different T1SS components (data not shown). The aprE/D genes are adjacent in many bacterial genomes and most rickettsial genomes, with a break in operon structure arising in Anaplasmataceae genomes. Despite conservation of the rickettsial T1SS components, there is very limited data supporting its function. As described above by our previous work, Rickettsia ToIC is involved in protein secretion, although the AprE/D complex has not been characterized in this process (Kaur et al., 2012). Four secretory proteins of E. chaffeensis were previously demonstrated to be secreted in an E. coli heterologous system containing the hly T1SS, with a TolC mutant diminishing effector secretion (Wakeel et al., 2011). Potential T1SS SSs characteristic of the repeats-in-toxins family were predicted within the CT sequences of these effectors, despite these SSs predominantly described from substrates specific for Group 2 ABC transporters. As E. chaffeensis only encodes the rickettsial-type AprD protein (Group 3 ABC transporter), the secretion of its effectors by the hly T1SS suggests an extreme lack of substrate specificity across systems.

Despite no known T1SS substrates for Rickettsia spp., it is probable that the AprD/E complex joins with TolC for protein export and that this T1SS is a conserved secretion pathway utilized by most species within Rickettsiales. Future studies on secretory proteins lacking Sec SSs will undoubtedly identify this pathway for delivery of effector molecules. Our observations suggest that surrogate expression systems encoding Apr-like T1SSs will be invaluable for identifying effectors via characterization of the T1SS substrate recognition mechanism.

**Type IV secretion**

Type IV secretion (T4SSs) are macromolecular machines that secrete substrates (protein, DNA, and nucleoprotein) across the bacterial cell envelope in both Gram-negative and Gram-positive species, as well as some wall-less bacteria and archaea (Christie et al., 2005; Alvarez-Martinez and Christie, 2009). T4SSs function in conjugation (nucleoprotein secretion) (Lawley et al., 2003), naked DNA uptake and release (Hofreuter et al., 2001; Hamilton et al., 2005), and the dissemination of genomic islands (Juhas et al., 2007a); thus, they are major factors in bacterial diversification and are responsible for the horizontal spread of antimicrobial resistance and virulence genes. T4SSs are also used by some bacterial species to deliver effector molecules (DNA and/or protein) into eukaryotic host cells (Cascales and Christie, 2003), wherein they engage signaling and regulatory molecules for their survival. In an extreme case, the cag T4SS of Helicobacter pylori secretes peptidoglycan into mammalian epithelial cells, wherein host pathogen recognition protein Nod1 detects PG and triggers an immune response to the pathogen (Viala et al., 2004). Thus, from a perspective of the extraordinary diversity of known substrates (Walden et al., 2010; Voth et al., 2012), the T4SSs might be considered the most plastic secretion systems known from prokaryotes.

In general, T4SSs form a channel that spans the cell envelope, often culminating in a surface-exposed structure (i.e. a pilus or sheath-like structure). Nearly all described T4SSs contain a membrane-bound ATPase (VirD4 or TcpA families), termed the type IV coupling protein (T4CP), which functions as a gatekeeper that directs substrates to the secretion channel (Llosa et al., 2002; Gomis-Ruth et al., 2004). The substrate for conjugative T4SSs is a relaxase bound to single-stranded DNA (Moncalan et al., 1999), while protein and DNA substrates of other T4SSs directly engage the T4CP in most cases (Zechner et al., 2012). Diverse T4SSs typically encode other membrane-bound ATPases (e.g. VirB4 and VirB11 families) that provide energy for processes such as substrate transfer, channel regulation, and pilin dislocation from the IM (Berger and Christie, 1993; Stephens et al., 1995; Kerr and Christie, 2010). The components forming the T4SS channel are quite variable across systems, often reflective of different cell envelope architectures from diverse bacteria and archaea (Alvarez-Martinez and Christie, 2009; Bhatt et al., 2013). T4SSs of Gram-negative bacteria form a pore at the OM, with the structure of one system (E. coli plasmid pKM101) determined to include 14 copies each of three proteins (VirB10, VirB5, and VirB7 families) (Chandran et al., 2009; Fronzes et al., 2009; Rivera-Calzada et al., 2013). Such important structures are needed for other T4SSs to understand the manner in which components from diverse systems collectively assemble cell envelope spanning secretion channels.

Despite the extensive diversity (in gene composition and organization) that underlies the hundreds of known T4SSs (Alvarez-Martinez and Christie, 2009), as well as tremendous ambiguity in annotation of T4SS genes (Ananiadou et al., 2011), T4SSs have been primarily classified into four groups: F, P, I, and G (Juhas et al., 2008). The widespread F-T4SSs and P-T4SSs are defined by the archetypal encoded by the F plasmid of E. coli (tra/tra) and the pTi plasmid of A. tumefaciens (vir), respectively. I-T4SSs are typified by the dot/icm system of IncI plasmids and some Legionellales (Legionella spp. and Coxiella burneti). GI-T4SSs are distinct systems that function in the transfer of genomic islands with which they are associated (Juhas et al., 2007b, 2008). The number of genes encoding analogous components between different groups is minimal.
Figure 7. Rickettsia genomes encode components of a type I secretion system. (a) Model for the Rickettsia T1SS illustrating the tripartite structure formed by the trimeric OEP (green), IM/periplasmic MFP (blue), and IM ABC transporter (red). The MFP is shown as a putative hexamer, bridging the asymmetry between the OEP and ABC transporter oligomers (Lee et al., 2012b), with the NT sequences in the cytoplasm. A theoretical T1SS substrate is shown with a predicted noncleavable SS at the CT. (b) Rickettsia typhi T1SS MFP and ABC transporter proteins are more similar to alkaline protease (Apr) T1SS components than those of α-hemolysin (Hly) T1SSs. Homology (% aa identity) determined across comparable regions only (gray): Ec, Escherichia coli str. 536 HlyD (YP_671701) and HlyB (YP_671700); Rt, R. typhi str. Wilmington AprE (YP_067266) and AprD (YP_067267); and Pa, Pseudomonas aeruginosa str. PA7 AprE (YP_001349499) and AprD (YP_001349500). For MFP proteins, variable NT sequences are colored black; for ABC transporter proteins, the NT protease domain of HlyB is colored green. (c) Predicted structure of R. typhi AprD. Modeling done using Phyre v.2.0 (Kelley and Sternberg, 2009), with the Thermotoga maritima (Thermotogae) ABC transporter (PDB no. 3QF4) selected for comparison (97% coverage, 100% confidence). The TMS region and NBD are noted. Structure visualized with jmol v.13.0.15 (Herraez, 2006). (d) Characteristics of the NT region of rickettsial AprE proteins. Predicted TMS regions are colored light blue. The asterisk denotes the variable NT sequences that are highly variable across rickettsial genera, but strongly conserved within rickettsial genera (see Fig. S2). Numbers in red depict regions of the alignment not shown. Residues colored yellow are conserved across all rickettsial Apr proteins (excluding ‘Candidatus Odyssella thessalonicensis’, which encodes different T1SSs). Generic names refer to the species shown in (e). Sequence information is provided in Table S6. (e) Phylogeny estimation of select bacterial MFP/ABC transporter pairs. MFP and ABC transporter proteins were each compiled and aligned using Muscle v3.6 (default parameters) (Edgar, 2004), with ambiguously aligned positions culled using Gblocks (Castresana, 2000; Talavera and Castresana, 2007). Phylogeny was estimated on concatenated alignments under maximum likelihood using RAxML v.7.2.8 (Stamatakis et al., 2008). A gamma model of rate heterogeneity was used with estimation of the proportion of invariant sites. Branch support was assessed with 1000 bootstrap pseudoreplications. Full species names and sequence information is provided in Table S6. All sampled MFP and ABC transporter proteins are encoded by adjacent genes, except for the Anaplasmataceae (star).
(Ananiadou et al., 2011), with one IM/cytosolic ATPase (VirB4 family) ubiquitous across all groups (Alvarez-Martinez and Christie, 2009). A recent phylogenetics-based classification scheme utilizing the VirB4 family proteins proposed an additional four groups, encompassing the diverse T4SSs encoded within genomes from Cyanobacteria, Bacteroidetes, Firmicutes, Actinobacteria, Tenericutes, and Archaea (Guglielmini et al., 2013). This global perspective on T4SSs accentuates some evolutionary constraints on T4SS architecture within certain taxonomic groups, with much of the diversity across systems having strong correlation with cell envelope structure (Alvarez-Martinez and Christie, 2009; Gillespie et al., 2010; Guglielmini et al., 2013).

Two distinct T4SSs, encoding P- and F-T4SSs, are known across Rickettsia genomes (Fig. 8A). All 55 sequenced Rickettsia genomes encode a P-T4SS similar to the vir archetype of Agrobacterium tumefaciens and related systems (Gillespie et al., 2009). Accordingly, we proposed the name Rickettsiales vir homolog (rvh) T4SS, as this T4SS is encoded across all genomes in the rickettsial families Rickettsiaceae, Anaplasmataceae, and Midichloriaceae (Gillespie et al., 2010). In contrast, the F-T4SS is known only from Rickettsiaceae, being sporadically encoded in some species of Rickettsia, but present within both sequenced genomes of the scrub typhus agent (O. tsutsugamushi) (Gillespie et al., 2012a, b). While the rvh T4SS genes are encoded solely on rickettsial chromosomes, the F-T4SS is present on both the chromosomes and plasmids of some Rickettsia spp. (Ogata et al., 2006; Blanch et al., 2007; Gillespie et al., 2012b) and is extensively proliferated throughout the chromosomes of O. tsutsugamushi (Cho et al., 2007; Nakayama et al., 2008). All available data at present suggest that the rvh T4SS functions in protein secretion, while the F-T4SS is responsible for dissemination of the integrative conjugative element, named Rickettsiales-amplified genetic element (RAGE), in which it is encoded. Accordingly, we refer to the F-T4SS hereafter as the RAGE T4SS.

rvh T4SS

Relative to the vir T4SS that typically encodes 12 scaffold proteins (VirB1-VirB11, VirD4), the rvh T4SS only lacks a homolog to VirB5, which is the minor subunit that is indispensable for pilus formation but not substrate transfer (Schmidt-Eisenlohr et al., 1999; Lai et al., 2000). A major subunit homolog (VirB2) homolog, RvhB2, is encoded within all rvh T4SSs, consistent with VirB2 proteins being essential for both pilus biogenesis and substrate transfer (Cascales and Christie, 2004; Jakubowski et al., 2005). The lack of VirB5-like genes and T-like pili on rickettsial surfaces indicates that the rvh T4SS does not extend a pili associated with the secretion channel, yet instead deposits substrates directly into the host cell environment through the OM pore (Gillespie et al., 2010). This is supported by the periplasmic localization of RvhB2 in R. typhi (Fig. 8A). Interestingly, other Rickettsiales genomes encode duplicate (Orientia, Neorickettsia, Midichloria) or proliferated (Wolbachia, Anaplasma, and Ehrlichia) rvhB2 genes (Nelson et al., 2008; Gillespie et al., 2010; Sutton et al., 2010; Al-Khedery et al., 2012). Some proliferated RvhB2 genes have become co-opted into a surface antigen family, possibly to cope for the lack of LPS biosynthesis (Gillespie et al., 2010). In support of a function at the bacterial surface, RvhB2 proteins of Neorickettsia risticii have been shown to localize to cell poles (Lin et al., 2009), and A. marginale RvhB2 proteins induce a T-cell response in cattle as part of a protective bacterial membrane vaccine (Lopez et al., 2008; Sutton et al., 2010). Aside from this RvhB2-associated variability, Rickettsia spp. alone encode a VirB1-like lytic transglycosylase (RvhB1), which may function as a local PG hydrolase that facilitates rvh channel assembly (Gillespie et al., 2009). RvhB1 would be unnecessary in other rickettsial genomes that do not synthesize PG. Collectively, these plastic attributes of RvhB1 and RvhB2 prompted us to conclude that variable features of the rvh T4SS correlate with the different cell envelope structures across Rickettsiales (Gillespie et al., 2010).

A defining feature of the rvh T4SS is the proliferation of genes encoding RvhB6 proteins, which are predicted IM channel components of the VirB6/TrbL family that are required for substrate transfer in P-T4SSs (Jakubowski et al., 2004). Encoded by five contiguously arrayed genes, RvhB6 proteins are unique among VirB6-like proteins due to large N- and/or C-terminal extensions flanking the VirB6/TrbL domain (Gillespie et al., 2009). Each family (RvhB6a-e) is highly conserved across Rickettsia spp., although extremely divergent from one another. While their functional significance is unknown, the detection of RvhB6a on the surface of R. typhi (Sears et al., 2012) may hint at surface structures elaborated by the RvhB6 extensions. In support of this, RvhB6 proteins have been detected on the surface of several strains of Wolbachia (Rances et al., 2008), which like all other Anaplasmataceae, contain four RvhB6 proteins that are contiguously arrayed, and contain large N- and/or C-terminal extensions outside of the VirB6/TrbL domain (Gillespie et al., 2010). Many RvhB6 proteins across Rickettsiales have repeat regions within the N- and/or C-terminal extensions, and a staggering degree of sequence length and variation across strains of A. phagocytophilum has been reported within a repeat region of one RvhB6 protein (Al-Khedery et al., 2012). Remarkably, all four rvhB6 genes of E. chaffeensis are coexpressed in tick and human cells, and various RvhB6-RvhB6 and RvhB6-RvhB9 interactions suggest more than one RvhB6 protein may assemble at the IM channel region of the rvh T4SS (Bao et al., 2009). The contribution of multiple divergent VirB6-like proteins to channel formation and substrate transfer and the potential for large N- and C-terminal extensions to form extracellular protrusions are fascinating areas of future research.

Three other components are encoded by duplicate genes within the rvh T4SS. First, all rvh T4SSs encode two proteins (RvhB4a and RvhB4b) homologous to VirB4 proteins, which are IM ATPases requiring an intact NBD for substrate transfer (Berger and Christie, 1993; Fullner et al., 1994; Watarai et al., 2002) and pilin dislocation from the IM (Kerr and Christie, 2010). While both proteins are full length, only RvhB4a contains the conserved RecA-like ATPase motifs that comprise the NBD (Gillespie et al., 2009). The mutations throughout the NBD of RvhB4b proteins should render them ATPase mutants, particularly the lack of the conserved Arg finger implicated in oligomerization, which is essential for ATPase activity (Wallen et al., 2012). Second, two VirB9-like proteins, RvhB9a and RvhB9b, are encoded within all rvh T4SSs. VirB9 proteins are Sec-processed, with the CTD participating with VirB7 and VirB10 in the OM pore of the T4SS (Chandran et al., 2009; Fronzes et al., 2009). Despite having a functional Sec SS (Ammerman et al., 2008), Rickettsia RvhB9b proteins lack the entire CTD relative to RvhB9a (Gillespie et al., 2009, 2010). While also truncated in Orientia genomes, RvhB9b proteins from all other rvh T4SSs contain the CTD, with both RvhB9a and RvhB9b of A. marginale able to induce IgG and stimulate CD4+ T cells from OM-vaccinated cattle (Morse et al., 2012). Finally, rvh T4SSs encode two VirB8-like channel proteins, RvhB8a and RvhB8b, with both proteins full length across all systems. In comparison with published VirB8 structures (Terradot et al., 2005; Bailey et al., 2006), as well as sequence analysis of hundreds of VirB8-like proteins, we observed that RvhB8a proteins possess mutations...
Figure 8. Two specialized type IV secretion systems are known from *Rickettsia* spp. (a) Relationship between the P-T4SS and F-T4SS encoded within *Rickettsia* genomes. Models for the *Rickettsiales* vir homolog (rvh) P-T4SS (left) and *Rickettsiales*-amplified genetic element (RAGE) F-T4SS (right) are based on the vir (*Agrobacterium* tumefaciens plasmid pTi) and tra/trb (*Escherichia coli* plasmid F) paradigms, respectively. Components analogous across both systems are colored similarly. Gray, components specific to P-T4SSs (VirB11) and F-T4SSs (TraN, TraH, TrbC, TraW, TraF, TraU, TraE). White, components not present in the rvh T4SS (VirB5) and RAGE T4SS (TraT, TraS, TraX, TraQ, Orf169) relative to *vir* (pTi) and *tra/trb* (pF), respectively. Asterisks in the F-T4SS model depict proteins encoded outside of the typical array of F conjugation genes in *E. coli*. Inset: Subcellular localization of *R. typhi* pilin (RvhB2) determined by immunogold (IG) electron microscopy of RvhB2 expression on intact, negative-stained rickettsia. IG-labeled RvhB2 predominantly traverses the PP:OM, white arrowheads; IM, black arrowheads. (b) Comparison of gene structures for P-T4SSs (*A. tumefaciens* plasmid pTi and *R. prowazekii* chromosome) and F-T4SSs (*Rickettsia endosymbiont of Ixodes scapularis* (REIS) chromosome and *E. coli* plasmid F). For each comparison, genes homologous across both systems are indicated by tan shading. Gene sizes and coding strands are shown for plasmids pTi and F, with the *Rickettsia* chromosomal sequences shown as models. Color schema as in (a), with light gray depicting genes in the plasmids pTi and F that are not represented in the models (and not present in *Rickettsia* genomes). For P-T4SSs, TraG is shown with the N- and CT domains analogous to VirB6 and VirB8 proteins, respectively, from P-T4SSs (Arutyunov and Frost, 2013), although the bitopic channel protein TraE is positioned in the model as a possible VirB8 analog. Duplicate genes within the rvh T4SS are illustrated by stars. For the RAGE T4SS: dashed box, relaxase (TraAI) and conjugative TraD protein (pfam06412); black, non-T4SS genes.

in the most conserved motif, NPxG, located in the CT region (Gillespie et al., 2009, 2010). As mutations in this region are predicted to alter the sharp turn confirmation between helix α4 and strand β4 (Terradot et al., 2005), RvhB8α proteins may function differently within rvh channels. Collectively, the redundancy associated with RvhB4, RvhB9, and RvhB8 proteins is unique to the rvh T4SS, and its significance to rvh assembly and function remains unknown.

rvh genes encoding homologs to the other OM pore components (RvhB7, RvhB10), the VirB3 IM protein (RvhB3), the
IM/cytosolic ATPase VirB11 (RvhB11), and the T4CP (RvhD4) are all single copy and highly conserved across Rickettsia (Gillespie et al., 2009) and Rickettsiales (Gillespie et al., 2010; Al-Khedery et al., 2012) genomes. Another distinguishing characteristic of the rvh T4SS (Gillespie et al., 2010) is the scattered chromosomal distribution of rvh gene islets relative to the usual arrangement of vir genes within P-T4SSs (Fig. 8b). Minimal variation in islet composition and chromosomal position exists within genera and species (Rances et al., 2008; Gillespie et al., 2009; Pichon et al., 2009; Al-Khedery et al., 2012), indicating that tight regulation of disparate rvh loci is a conserved mechanism in rickettsial genomes. Supporting this, a protein of the helix-turn-helix/xenobiotic response element family, EcXR, has been shown to coordinately express several of the rvh islets in E. chaffeensis (Cheng et al., 2008). Such rvh regulators are unknown from Rickettsia and Orientia genomes, although we previously noted EcXR homologs are present in all genomes of the Anaplasmataceae (Gillespie et al., 2010). Subsequently, EcXR-like proteins from the Wolbachia endosymbiont of a filarial nematode (Brugia malayi) were identified and shown to co-regulate rvh gene islets as well as some genes for the biosynthesis of riboflavin, which may be supplied to B. malayi (Li and Carlow, 2012). Given that host conditions impact T4SS operon regulation in other intracellular bacteria (Schmiederer et al., 2001; Boschi et al., 2002), it is probable that these cues work in concert with regulatory elements to coordinate the assembly and function of components encoded across rvh islets.

To date, no substrates of the Rickettsia rvh T4SS have been identified. However, rvh effectors have been characterized for A. phagocytophilum (Lin et al., 2007; Niu et al., 2010; Rikihisa and Lin, 2010; Rikihisa et al., 2010), A. marginale (Lockwood et al., 2011), and E. chaffeensis (Rikihisa and Lin, 2010; Liu et al., 2012). To our knowledge, only two of these effectors have been shown to directly interact with their cognate RvhD4 proteins: Atp-1 of A. phagocytophilum (Niu et al., 2010) and ECHO825 of E. chaffeensis (Liu et al., 2012). All other rvh substrates, including the well-characterized AnkA effector, were shown to be translocated by heterologous T4SSs from either A. tumefaciens (vir) (Lin et al., 2007) or L. pneumophila (dot/icm) (Lockwood et al., 2011). Remarkably, RvhD4 sequences of Anaplasmataceae species contain highly variable CT extensions (Rikihisa and Lin, 2010), which range from 50 (Neorickettsia spp.) to 250 (A. centrale) aa and are extremely enriched with acidic residues (data not shown). This is reminiscent of the extended CTD of TraD (T4CP of E. coli F plasmid) that interacts with TraM of the relaxosome (Beranek et al., 2004; Lu and Frost, 2005; Lu et al., 2008; Wong et al., 2011), forming the only known specific contact mediating a T4CP-substrate interaction. The significance of these extended CT sequences of Anaplasmataceae T4CPs, particularly their role in substrate recognition, as well as their absence in other rvh T4CPs, remains to be determined. Importantly, the future characterization of the rvh T4CP-substrate interaction, especially the nature of the substrate SS, will be paramount for facilitating the discovery of the first Rickettsia rvh substrates.

RAGE T4SS

Originally described from the O. tsutsugamushi str. Boryong genome, the RAGE is comprised of F-T4SS genes, as well as genes encoding histidine kinases (HK), ANK repeat-containing proteins (Anks), and tetratricopeptide repeat (TPR) domain-containing proteins (among others) (Cho et al., 2007). While extensively proliferated and highly pseudogenized, sequencing of a second O. tsutsugamushi genome allowed for reconstruction of a hypothetical complete element (Nakayama et al., 2008), in which the F-T4SS was shown to be similar to F-T4SSs encoded within the R. bellii (Ogata et al., 2006) and R. massiliae (Blanc et al., 2007) genomes. The RAGE tra/trb genes encode 15 proteins that form the scaffold of typical F-T4SSs (Fig. 8a). The tra/trb gene order within the RAGE is highly similar to the structure of the E. coli plasmid F, except that the components primarily involved in the regulation of conjugation are absent (Fig. 8b). Furthermore, the TraAl relaxase is more typical of integrative conjugal elements than F-T4SSs (Gillespie et al., 2012b). Our previous report on the genome sequencing of the Rickettsia endosymbiont of Ixodes scapularis (REIS) revealed nine complete (or nearly complete) RAGEs, as opposed to the single RAGEs present in the R. bellii and R. massiliae genomes (Gillespie et al., 2012b). Moreover, two of the four plasmids carried by REIS (pREIS1 and pREIS3) were shown to encode RAGEs, with phylogeny estimation indicating multiple invasions of RAGEs within Rickettsia genomes throughout evolution. Remarkably, one REIS RAGE is highly similar to the RAGE encoded in R. bellii genomes, despite the distant relatedness between these rickettsial lineages, suggesting that RAGEs are actively mobile and exchangeable between diverse species.

In addition to syntenic F-T4SS genes, RAGEs across Rickettsia and Orientia share common genes encoding DNA methyltransferases (D12 class), PolC-like DNA helicases, stringent response hydrolases and synthetases, and proteins containing HK and TPR domains. Thus, these factors are likely essential for regulation and mobility of RAGEs. However, compared with Orientia RAGEs, Rickettsia RAGEs lack Anks and the HK genes are especially the nature of the substrate SS, will be paramount for facilitating the discovery of the first Rickettsia rvh substrates.

CONSERVATION OF RICKETTSIA SECRETION MACHINES

Aside from the T5SSs and F-T4SS, in which Scas (Blanc et al., 2005b; Sears et al., 2012) and RAGEs (Gillespie et al., 2012b) are variably encoded across Rickettsia genomes, there is strong conservation of protein secretion systems for nearly all Rickettsia species and strains. A notable exception is seen for the rvh T4SS of REIS, for which we identified several components as probable pseudogenes (rvhB1, rvhB6a, rvhB6b, rvhB6c, rvhB6e, rvhB9b, rvhB10, and rvhD4) (Gillespie et al., 2012b). As hundreds of pseudogenes were part of the REIS genome assembly, which included 109 regions of low sequence coverage, the validity of these pseudogenes remains to be determined. We previously noted a lack of rvhB6e from R. massiliae str. MTUS and a split rvhB6d from R. bellii str. OSU 85 389 (Gillespie et al., 2009); however, other strains sequenced for these species contain full-length genes. Thus, there may
be flexibility for the number of encoded paralogs within the rnu T4SS, as seen by a truncated rnuB60a gene from the recently sequenced strain GvF12 of R. prowazekii. However, many of the more recently sequenced Rickettsia genomes have not been assembled or manually curated; thus, missing genes or putative pseudogenes may be a product of low sequencing coverage. Collectively, there is a strict conservation in Rickettsia genomes for genes encoding protein secretion systems. Given the varying degrees of pathogenicity (or lack thereof) reported across these species and strains, it is likely that many underlying determinants of pathogenicity are encoded by the variable substrates of these secretion systems.

SECRETORY PROTEINS AT THE HOST INTERFACE

In addition to those proteins discussed above (Sca passenger domains, RARP-1), other Rickettsia secretory proteins have been studied primarily for their roles in host infection, yet the secretion pathways for nearly all of these proteins are unknown (Fig. 9).

Adhesins

Rickettsial OM proteins of a poorly characterized family, COG3637 (opacity proteins and related surface antigens), have previously been investigated for possible roles in host cell invasion. Implementing a proteomics-based screen to capture rickettsial ligands interacting with host epithelial cells, Renesto et al. (2006) identified a 30-kDa protein of R. conorii and R. prowazekii, named adhesion of Rickettsiae (Adr) 1, belonging to COG3637. This protein family is conserved across many proteobacterial genomes and nearly all genomes of Alphaproteobacteria (data not shown), and members are predicted to form β-barrels in the bacterial OM. In all Rickettsia genomes, two Adr paralogs are encoded adjacent to one another and share c. 45% aa identity. Like most members of COG3637, rickettsial Adr1 and Adr2 proteins contain strongly predicted NT Sec SSs and no predicted TMS regions, implying their likely Sec-mediated translocation to the PP and subsequent insertion into the OM. Consistent with a role in host cell entry, both Adr1 and Adr2 of R. prowazekii were detected on the surface of bacterial cells, with anti-Adr2 antibodies inhibiting host cell invasion (Vellaswamy et al., 2011). However, a recent study showed that Adr1 of R. conorii is involved in the recruitment of human vitronectin, a mammalian glycoprotein that inhibits the membrane-damaging effect of the terminal cytolytic complement pathway (among other cellular functions) (Riley et al., 2014). When expressed in serum sensitive E. coli, Adr1 of R. conorii, as well as that of R. rickettsii, R. prowazekii, and R. typhi, mediated resistance to serum killing, suggesting that evasion of host complement-mediated killing is an evolutionarily conserved process across Rickettsia species (Riley et al., 2014).

Host cytoskeleton targeting proteins

Aside from interacting with vinculin (Sca4), and the direct (Sca2) and indirect (Sca5) interactions of Scas with host cell actin, two additional Rickettsia proteins also target the host cytoskeleton, although their functions are less understood. The RickA protein of some Rickettsia spp. activates host Arp2/3 complexes, resulting in actin nucleation (Gouin et al., 2004; Jeng et al., 2004). RickA is localized to the bacterial surface (Gouin et al., 1999) and has been considered a likely substrate of the rnu T4SS (Gouin et al., 2004). While the lack of NT Sec SSs and TMS regions in all RickA proteins supports this hypothesis, no association with the rnu T4SS has been characterized. RickA-mediated actin nucleation has been implicated in the intercellular spread of some species (Heinzen et al., 1993; Gouin et al., 1999; Heinzen et al., 1999; Van Kirk et al., 2000), while RickA pseudogenes correlate with a lack of HAP in others. For instance, we originally identified a transposon (ISRpe1) interrupting rickA of the nonpathogen R. peacockii, which correlates with a lack of observed HAP by this species (Simser et al., 2005). We also identified a similar transposase-interrupted RickA gene within the genome of REIS, another nonpathogenic species that lacks HAP (Gillespie et al., 2012). However, other Rickettsia species, namely R. canadensis and R. felis, contain genes encoding full-length RickA proteins yet have not been observed spreading intercellularly via HAP. Furthermore, R. raoultii, which was determined to express RickA in levels comparable to SFG rickettsiae actin nucleators, displays atypical ABM patterns (Balraj et al., 2008). These data indicate that RickA may provide a redundant role to Sca2 formlike proteins for Arp2/3 complex-independent actin nucleation, but alone is not essential for ABM for some rickettsial species. However, very recent research on R. parkeri demonstrated that RickA and Sca2 each direct independent modes of ABM, with RickA generating short, curved actin tails early during infection and Sca2 generating long, straight actin tails later in infection (Reed et al., 2014). This is partially consistent with a previous suggestion that RickA targets actin for the facilitation of host cell entry (Haglund et al., 2010), but reveals that RickA-induced HAP, like that of Sca2, can facilitate intercellular spread in some species. Collectively, the variable genotypic profiles (sca2, rickA) of Rickettsia spp. are beginning to shed light on the substantial diversity in motility-associated phenotypes observed across species.

Another actin-associated protein, RalF, is also differentially encoded across Rickettsia genomes (Fig. 9). RalF, known in prokaryotes only from some Rickettsia and Legionella spp. (Cox et al., 2004), is remarkable in that it contains a Sec7 domain. This domain is conserved across eukaryotic guanine nucleotide exchange factors (GEF), which activate ADP-ribosylation factors (Arfs), proteins involved in vesicle trafficking and actin remodeling. Legionella RalF (RalFα) is an I-T4SS effector that activates and recruits host Arfs to the Legionella-containing vacuole (modified phagosome) (Nagai et al., 2002). The structure of RalFα revealed an NT Sec7 domain, as well as a central Sec7-capping domain (SCD) that regulates active site access to Arfs (Amor et al., 2005). Additionally, the RalFα CT mediates secretion through the I-T4SS (Nagai et al., 2005). The Sec7 domain and SCD of Rickettsia RalF (RalFβ) share c. 45% aa identity with RalFα, although a PRR flanking the CT of the SCD distinguishes RalFα from RalFβ (Gillespie et al., 2008). Furthermore, a CT rnu T4SS translocation signal for RalFα has not been identified, although the absence of an NT Sec SS and TMS regions suggests it is a substrate of the rnu T4SS or the Rickettsia T1SS.

Comparative studies of RalFα and RalFβ determined similar GEF activity in the Sec7 domains of both proteins, yet divergent subcellular localization patterns that are mediated by the SCDs (Alix et al., 2012). The SCD of RalFα affords the protein access to membrane transport in the host secretory pathway, while the RalFβ SCD mediates the protein’s interaction with actin filaments near the host IM. Regarding RalFβ, it was further determined that the SCD localizes the protein to the plasma membrane, but the PRR is responsible for interacting with components of the actin cytoskeleton (Alix et al., 2012). Membrane sensor regions were recently identified within the SCDs of RalFα and RalFβ, with differential enrichments in aromatic and positively charged residues determining divergent lipid

...
Figure 9. Conservation and distribution across 55 *Rickettsia* genomes of genes encoding 18 secretory proteins. Phylogeny at left estimated as previously described (Driscoll et al., 2013), with additional genomes annotated using RAST (Aziz et al., 2008). Classification scheme for *Rickettsia* spp. follows previous studies: red, ancestral group; blue, transitional group; aquamarine, typhus group; and brown, spotted fever group (Gillespie et al., 2007, 2008). *Rickettsia helvetica* is unclassified following recent recommendations (Driscoll et al., 2013). Protein names are listed at top and further described in the text (NOTE: Atypical surface-exposed proteins are not shown, see Moonlighting Proteins). Secretion pathway information is above protein names. Large red circles, full-length proteins; medium-sized blue circles, truncated proteins that may have additional fragments encoded by separate genes; and small black circles, probable pseudogenes with one or more fragments that do not span the complete protein. A lack of circles denotes missing genes, with zero significant matches using blastp (in some cases, small fragments are detectable with tblastn). For Sca2, D denotes divergent passenger domains fused to the conserved Sca2 β domain. Numbers on other circles depict proteins encoded by multiple genes. Information for each protein is provided in Table S1.
substrates that regulate Arf-GEF activities (Folly-Klan et al., 2013). Interestingly, these studies were performed with R. prowazekii Ralf, which together with R. typhi Ralf have a PRR c. 205 aa less than the PRRs of RalF proteins from R. bellii and TRG rickettsiae. This may indicate divergent functions for RalF across the limited rickettsial species that contain the protein. Curiously, small ORFs encoding the PRR of RalF are present in most SFG rickettsiae genomes and indicate that the larger PRR was likely the ancestral domain, with TG rickettsiae RalF undergoing a minimization (data not shown). Whether or not these RalF-like PRR-containing proteins have a function in SFG rickettsiae, particularly in targeting actin, remains to be determined.

A previous study on R. parkeri has defined a core set of proteins that are involved in the nucleation process that leads to ABM: profilin, fimbrin/T-plastin, capping protein, and coflin (Serio et al., 2010). The role of Sca2 in this Arp2/3 complex-independent process is clear (Haglund et al., 2010; Kleba et al., 2010), and most rickettsial species known to nucleate actin for intracellular spread encode a Sca2 protein with requisite FH2 and WH2 domains (except R. bellii) (Fig. 9). Combined with studies demonstrating that inhibition of the host Arp2/3 complex fails to abolish rickettsial ABM (Harlander et al., 2003; Heinzen, 2003; Serio et al., 2010), it is likely that for some species, rickettsial proteins targeting the Arp2/3 complex are functioning outside of induction of ABM (Ireton, 2013). Indeed, the Arp2/3 complex has recently been shown to be critical for R. parkeri invasion of various cell types (Reed et al., 2012). Remarkably, this study demonstrated that upstream pathways for actin nucleation during the host cell entry process are divergent between arthropod and mammalian cells, before collectively converging on the Arp2/3 complex. Given the multitude of eukaryotic actin polymerization pathways, as well as the wide range of metazoan cells colonized by the collective species of Rickettsia, it is not surprising that some cytoskeletal interacting proteins are either differentially encoded across rickettsial genomes (Sca2, RickA, RalF) or variable in their regions that directly engage host cytoskeletal proteins (Sca4). Such a fundamental process as host cell infection is likely governed by additional rickettsial factors (Gong et al., 2013), predominantly secretory proteins, which will undoubtedly be discovered in the near future.

**Hemolysins and phospholipases**

Efforts to understand another fundamental process in the rickettsial lifestyle, phagosomal escape, have resulted in the characterization of several membranolytic enzymes potentially involved in lysing the endocytic vacuole. Hemolysins are considered candidates for this process, given their ability to lyse the membranes of a variety of different cell types (Braun and Focareta, 1991). All sequenced Rickettsia genomes encode two enzymes annotated as hemolysins: TlyA and TlyC (Fig. 9). Sequence and structural analysis (data not shown) suggest that both of these proteins are related to CorC of a conserved bacterial magnesium and cobalt transporter (Gibson et al., 1991). Indeed, annotations for a wide range of proteins related to TlyA and TlyC are diverse and possibly misleading (Arenas et al., 2011), with some studies yielding divergent functions not always related to membrane lysis. For example, TlyA of Mycobacterium tuberculosis has been shown to possess membranolytic activity in vitro, but also functions as an rRNA methyltransferase (Rahman et al., 2010). Additionally, surface-exposed leptospiral TlyC was demonstrated to lack membranolytic activity, yet binds to host laminin, collagen IV, and fibronectin to facilitate infection (Carvalho et al., 2009). Rickettsia TlyA, which has not been functionally characterized, differs from TlyC primarily in the presence of an NT Sec SS and adjacent TMS region. For TlyC, we previously demonstrated its membranolytic activity by transforming a hemolysin-negative mutant strain of Proteus mirabilis with recombinant R. typhi TlyC, which restored the hemolytic phenotype of P. mirabilis (Radulovic et al., 1999). Recombinant R. typhi TlyC also conferred a hemolytic phenotype to E. coli str. K-12, yet the mechanism of secretion in this surrogate host, as well as P. mirabilis, remains unknown.

Phospholipases have also been investigated for their possible roles in phagosomal escape. pld, which encodes phospholipase D (PLD), is a highly conserved gene present in all sequenced rickettsial genomes (Fig. 9). Functioning as a dimer, PLD was shown to exhibit phospholipase activity in vitro, and rickettsiae pretreated with anti-PLD antibodies had a less cytotoxic effect on Vero cells relative to untreated bacteria (Renesto et al., 2003). In a study designed to identify factors directly mediating rickettsial phagosomal escape, Salmonella enterica serovar typhimurium, which does not lyse its endocytic vacuole during host infection, was transformed with R. prowazekii genes encoding TlyC or PLD (Whitworth et al., 2005). A small portion of salmonellae transformed with tlyC were observed free in the host cytoplasm, whereas 100% of pld-transformed salmonellae were cytosolic, strongly implicating PLD in endosomal lysis. Despite this, a pld mutant generated in a subsequent study, which used homologous recombination to replace R. prowazekii pld with a truncated form of the gene, showed no discernible phenotype in tissue culture compared with the wild-type strain (Madrid Evir) (Driskell et al., 2009). However, this pld mutant did not show attenuated virulence in a guinea pig infection model. Collectively, several lines of evidence support PLD as a secreted protein involved in host infection. The presence of functional NT Sec SS in R. typhi PLD (Ammerman et al., 2008) suggests it may localize to the rickettsial OM from the PPI, or possibly secreted in a manner similar to RARP-1 outlined above.

Enzymes eliciting phospholipase A2 (PLA2) activity have long been suspected to mediate phagosomal escape, as well as entry and exit of host cells (Winkler and Miller, 1982; Winkler and Daugherty, 1989; Silverman et al., 1992; Ojcius et al., 1995; Walker et al., 2001). PLA2 activity derived from rickettsiae (and not the host) was demonstrated to be involved in rickettsial uptake prior to identification of rickettsial PLA2 phospholipases (Silverman et al., 1992; Walker et al., 2001). As Rickettsia genome sequences accumulated, a conserved patatin (Pat)-like PLA2 encoding gene, pat1, was identified in silico and shown to be conserved in all Rickettsia genomes, with the pRF plasmid of R. felis carrying an additional Pat-like gene (Blanc et al., 2005a). A previous report from our laboratory identified a second Pat encoding gene, pat2, which is present in all TG rickettsiae genomes, yet sporadically encoded in other rickettsial genomes (Rahman et al., 2010). We demonstrated that R. typhi Pat2 possesses PLA2 activity and is cytotoxic to yeast cells and that the protein is secreted into the host cell cytoplasm during infection and requires a host cofactor for enzymatic activity. A subsequent report confirmed the PLA2 activity of R. prowazekii Pat2 (Housley et al., 2011). Recently, we determined that Pat1 of R. typhi also possesses PLA2 activity and is cytotoxic to yeast cells and, like Pat2, is also secreted extracellularly during infection and requires host cofactor(s) for activation (Rahman et al., 2013). Pretreatment of rickettsiae with anti-Pat1 and anti-Pat2 antibodies significantly decreased bacterial survival, underscoring the roles for both patatins during the early stage of R. typhi host cell infection. These data, combined with phylogenomics analyses,
suggest that R. typhi (and probably R. prowazekii) utilizes two evolutionary divergent phospholipases during host cell infection, presenting a factor that distinguishes TG rickettsiae from most other Rickettsia species (Rahman et al., 2013) (Fig. 9). As both of these secretory proteins lack NT Sec SSs, Pat1 and Pat2 are promising candidates for effectors of either the T1SS or T4SS.

**Toxin-antitoxin modules**

Toxin-antitoxin (TA) modules are a diverse family of proteins that were originally characterized as regulators of programmed cell death (Jensen and Gerdes, 1995; Yarmolinsky, 1995). Specifically, plasmid-encoded modules express (toxin) and neutralize (antitoxin) factors that maintain perfect plasmid segregation after cell division, but result in cell death as a consequence of imperfect plasmid segregation (Hayes, 2003). More recent studies have shown that toxins of chromosomal-encoded TA modules directly cleave mRNA in response to nutritional stress, suggesting that the majority of TA modules probably function as control mechanisms in response to nutrient stress (Gerdes et al., 2005). While abundantly encoded in the genomes of free-living bacterial species, TA modules are scarcely encoded within the genomes of host-associated species (Pandey and Gerdes, 2005). Our previous phylogenomics analyses determined that one to five TA modules are encoded in most Rickettsia genomes (except TG rickettsiae, which contain none), yet the number of single toxin and antitoxin genes is quite variable across genomes (Gillespie et al., 2008). Subsequent reports have identified more putative toxin and antitoxin genes (Audoly et al., 2011; Socolovschi et al., 2013), although their function as TA modules has been largely unexplored. Comprehensive genome analyses suggest TA modules are elevated in pathogenic species vs. their closest nonpathogenic relatives Georgiades and Raoult, 2011, b); however, for Rickettsia spp., the absence of TA modules in TG rickettsiae and presence in nonpathogenic species refute this claim. Furthermore, TA modules are lacking in Orientia and Anaplasmataceae genomes. Thus, it is likely that Rickettsia TA modules have more species-specific functions without correlations between major groups and pathogenicity.

Despite 23 toxin and 15 antitoxin genes encoded in the R. felis genome (Ogata et al., 2005), only one protein (antitoxin VapB) was previously identified in a proteome analysis (Ogawa et al., 2007). Subsequent structural analysis of the VapBC module revealed that VapB can block VapC, and thus regulate the level of toxin expression (Mate et al., 2012). Remarkably, VapC was later shown to be secreted into host cells during R. felis infection, with the toxin exhibiting RNase activity (Audoly et al., 2011). VapC homologs (VapC-1) are present in most rickettsial genomes, but other VapC proteins are encoded by divergent genes we collectively assigned to VapC-d (Fig. 9). Importantly, most genomes encode one or both proteins, except TG rickettsiae, R. australis and R. akari, which encode none. Whether VapC and other toxins are secreted by other Rickettsia spp. remains to be determined, as does the functional significance of toxin secretion to pathogenesis.

**Anks**

Anks are encoded within the genomes of many intracellular bacterial species (Pan et al., 2008; Voth, 2011; Kaur et al., 2012), with amoeba-associated species tending to be enriched with Ank genes (Schmitz-Esser et al., 2010). Anks are variably encoded across Rickettsia genomes (Gillespie et al., 2008; Merhej and Raoult, 2011), with larger genomes indicative of high rates of lateral gene transfer, namely R. felis (Ogata et al., 2005), R. bellii (Ogata et al., 2006), R. massiliae (Blanc et al., 2007), and REIS (Gillespie et al., 2012b), enriched with Ank genes relative to the smaller Rickettsia genomes. Despite this, aside from the ubiquitous RARP-1 characterized by us, no Ank genes have been studied thus far for their roles in pathogenicity. We have very recently begun investigating another Rickettsia Ank gene (R. typhi gene RT0600), encoding the protein RARP-2, which is secreted extracellularly by R. typhi during host cell infection (unpublished data). RARP-2 proteins contain an NT domain of unknown function, coupled with a CTD comprised of variable ANK repeats (R. typhi and R. felis RARP-2 contain 11 and 5 repeats, respectively). Comparative analysis of 55 Rickettsia genomes indicates (1) five RARP-2 pseudogenization events, mostly in nonpathogenic species (Fig. 9); (2) extraordinary variation in ANK repeats across RARP-2 homologs; and (3) a highly conserved NT domain with no recognizable homology to any sequences in GenBank. Unlike RARP-1, RARP-2 does not encode an NT Sec SS, suggesting it is a substrate of either the T1SS or the ruf T4SS. Our efforts are currently aimed at determining the secretion pathway of RARP-2 and its target within host cells.

**Moonlighting proteins**

There is increasing evidence for the surface localization of some rickettsial proteins that typically function in the bacterial cytoplasm or PP. Dozens of these candidate moonlighting proteins (MLPs) (Henderson and Martin, 2011), which are encoded by single-copy genes, were identified within the surface proteomes for R. typhi (Sears et al., 2012), R. parkeri (Pornwiroon et al., 2009), R. conorii (Renesto et al., 2005, 2006), and R. helongiangensis (Qi et al., 2013), yet have a wide range of ascribed cellular and metabolic functions. Four of these proteins, chaperone GroEL, elongation factor Tu 1 (TuF), ATP synthase subunit beta (AtPD), and foldase PrsA, were identified on the surface of all four species, providing robust evidence for their extracellular localization. Indeed, across a variety of Rickettsia species, several candidate MLPs have demonstrated immunoreactivity, for example, cell division protein FtsZ, translation initiation factor 2, cysteinyll tRNA synthetase, TuF, ribosomal proteins RplY and RpsB, cytosol aminopeptidase, uncharacterized protein YbgF, AtPD, and chaperones GroES, GroEL, PrsA, SurA, and DnaK (Renesto et al., 2005; Hajem et al., 2009; Pornwiroon et al., 2009; Qi et al., 2013). Furthermore, some of these candidate MLPs have been identified on the cell surfaces of Anaplasmataceae species (Ge and Rikihisa, 2007a, b; Noh et al., 2008; Gibson et al., 2010), as well as other bacteria (Rhomberg et al., 2004; Boonjakuakul et al., 2007; Desvaux et al., 2010; Sommer et al., 2010). Collectively, these data suggest that highly conserved proteins with usual cytoplasmic or periplasmic functions are moonlighting on the cell surface of divergent rickettsial species.

One particular candidate MLP, the OM-anchored lipoprotein PrsA, was originally characterized as a surface protein of R. prowazekii, potentially catalyzing conformational changes in surface-localized proteins (Emelyanov and Demyanova, 1999). A moonlighting function as an adhesin was reported for R. felis PrsA, which was shown to interact with tick histone H2B, a protein known to localize to the surface of eukaryotic cells (Thepparit et al., 2010). Despite identical gene sequences, electrophoretic PrsA variants are known across R. prowazekii strains of variable pathogenicity, illustrating different capacities for post-translational modification of proteins across rickettsial strains (Emelyanov and Loukianov, 2004). Interestingly, different lysine methyltransferase profiles were reported across several R. prowazekii strains (Abeykoon et al., 2012), a finding that
possibly explains the observed SDS-PAGE mobility differences of PrsA proteins across these strains. Remarkably, homologs to one of these lysine methyltransferases (RP789) were found on the surfaces of both R. typhi (Sears et al., 2012) and R. helongiiangensis (Qi et al., 2013), suggesting that surface-exposed proteins may undergo methylation of Lys residues, a process that would help neutralize the highly acidic rickettsial proteins exposed to the host cytoplasm, such as Sca5 (Sahni et al., 2013). It is tempting to speculate that extracellular chaperones, such as PrsA, may primarily function in folding other secretory proteins, but also may themselves undergo post-translational modification to provide rickettsial with additional surface variability and/or ligands for mediating interactions with host receptors.

Collectively, our observations suggest that secreted MLPs may contribute to the secretory proteins that operate at the host/microorganism interface, with some also functioning in the direct modification of other surface-exposed proteins. The characterization of secreted MLPs is an exciting area of future research for Rickettsiologists. Of primary importance is understanding the mechanism(s) for OM translocation of substrates that either completely lack Sec SSs or are Sec-processed but primarily known for their activities in the P. Despite their highly conserved nature, characterized MLPs from other bacterial species have remarkably divergent functions, with moonlighting activities often directly involved in pathogenesis (Henderson and Martin, 2013). A potential abundance of MLPs on rickettsial cell surfaces is consistent with reductive genome evolution, as functional redundancy via moonlighting would be an effective strategy for survival in multiple diverse eukaryotic hosts despite limited genetic resources.

CONSERVATION OF SECRETORY PROTEINS

In contrast to the outlined protein secretion systems and pathways, a substantial lack of conservation is seen within most families of secretory proteins (Fig. 9). All but seven families (Sca5, Sca4, Adr1, Adr2, TlyC, PLD, RARP-1) lack genes in one or more Rickettsia genomes. Most of these variably present genes encode proteins rich in repeat motifs and other hyper-variable regions, suggestive of rapid sequence evolution likely driven by host selective pressures. Despite being encoded across all Rickettsia genomes, Sca5, Sca4, and RARP-1 are also hyper-variable in sequence length and composition, with proteins from many species containing repeat regions. Furthermore, Adr1 and Adr2 have minimal sequence variation present within surface-exposed loops of their beta-barrel structures, while the NT region of PLD proteins is hypervariable in composition across Rickettsia spp. This leaves TlyC as the lone highly conserved secretory protein (77% identity across R. typhi and R. bellii), being comparable with evolutionary constraints characteristic of many surface candidate MLPs (data not shown). Thus, TlyC and surface candidate MLPs aside, the characterized secretory proteins of Rickettsia spp. are highly variable across genomes, with nearly every species containing a different repertoire (Fig. 9). These variable repertoires, in conjunction with yet undescribed secretory proteins, likely account for the differences in phenotype and host ranges observed across species, and are undoubtedly critical factors contributing to pathogenesis.

CONCLUSION AND FUTURE DIRECTIONS

Through the combination of previous research with current comparative genome analyses, we provide our perspective on the Rickettsia secretome. This work sheds light on the architecture of protein secretion systems for highly reductive genomes, revealing common themes shared with other bacteria, but also illustrating several innovative systems and pathways unique to Rickettsia species. Surprisingly, many of the well-characterized secretory proteins of Ricketttsia spp. (e.g. Sca4, RkA, RaIF, TlyC, PLD, Pat1, Pat2) have yet to be linked to their cognate secretion pathways. Once achieved, an understanding of substrate characteristics unique to each pathway will help identify novel secretory proteins and their potential contributions to rickettssioses. This information will also pave the way for understanding how components of the secretome can be utilized as therapeutic targets and for the generation of vaccines. Challenges will continue to come from the AT-rich nature of these genomes, which complicates informatics tools and annotation methods, as well as hampers many experimental techniques. Nonetheless, informatics approaches that accommodate these characteristics, coupled with recent advances in molecular characterization tailored to host-dependent bacteria, will pave the way for a better understanding of protein secretion in these fascinating intracellular microorganisms.

ACKNOWLEDGEMENTS

We thank members of the Azad laboratory (U. Maryland, School of Medicine), Dr. Timothy Driscoll (Virginia Bioinformatics Institute at Virginia Tech), and Dr. Joao Pedra (U. Maryland, School of Medicine) for invaluable feedback and discussion. We are grateful to Dr. Ted Whitworth and Dr. Marco Quevedo Diaz for providing the immungold electron microscopy images of R. typhi pilin. The research presented in this review has been supported with funds from the National Institute of Health/National Institute of Allergy and Infectious Diseases grants (R01AI071828, R01AI043006, and R01AI59118). 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.

Conflict of interest statement. None declared.

SUPPLEMENTARY DATA

Supplementary material is available at FEMS Microbiology Reviews online.

Fig. S1. Conservation of Vinculin Binding Sites (VBS) across rickettsial Sca4 sequences.

Fig. S2. The NT sequences of rickettsial AprE proteins are highly conserved within rickettsial genera, but extraordinarily divergent across genera.

Table S1. Information for proteins illustrated in the secretion pathway models.

Table S2. Select T5SS proteins sharing characteristics within the autocatalytic region.

Table S3. Characterized and predicted TatB proteins from Rickettsiales and other select bacterial species.

Table S4. Twin-arginine translocation (Tat) system proteins from select Rickettsiales species.

Table S5. Ubiquinol-cytochrome c reductase iron-sulfur subunit (PetA) proteins from select Rickettsiales species.

Table S6. T1SS MFP and ABC transporter proteins from select proteobacterial species.


Bogsch E, Brink S, Robinson CG Pathway specificity for a delta pH-dependent precursor thylakoid lumen protein is governed by a ‘Sec-avoidance’ motif in the transfer peptide and a ‘Sec-incompatible’ mature protein. EMBO J 1997;16:3851–9.


Duong F, Wickner W The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. EMBO J 1997;16:4871–9.


Georgiades K, Raoult D Genomes of the most dangerous epi- demic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. PLoS ONE 2011b;6:e17962.


Gibson MM, Bagga DA, Miller CG, Maguire ME Magnesium transport in Salmonella typhimurium: the influence of new


Hofreuter D, Odenbreit S, Haas R Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system. Mol Microbiol 2001;41:379–91.


Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden


Lee M, Jun SY, Yoon BY, Song S, Lee K, Ha NC Membrane fusion proteins of type I secretion system and tripartite efflux pumps share a binding motif for TolC in gram-negative bacteria. PLoS ONE 2012b;7:e40460.


Li Z, Carlow CK Characterization of transcription factors that regulate the type IV secretion system and riboflavin biosynthesis in Wolbachia of Brugia malayi. PLoS ONE 2012;7:e51597.

Li H, Walker DH rOmpA is a critical protein for the adhesion of Rickettsia rickettsii to host cells. Microb Pathog 1998;24:289–98.


Madasu Y, Suarez C, Kast DJ, Kovar DR, Dominguez R Rickettsia Sca2 has evolved formin-like activity through a different molecular mechanism. Proc Natl Acad Sci USA 2013;110:E2677–86.


Mate MJ, Vincentelli R, Foos N, Raoult D, Cambillau C, Ortiz-Lombardia M Crystal structure of the DNA-bound VapBC2

Matsuyama Y, Fujita Y, Mizushima S SecD is involved in the re-
lease of translocated secretory proteins from the cytoplasmic

McBride MJ, Zhu Y Gliding motility and Por secretion system
genes are widespread among members of the phylum bac-

McDonough JA, Hacker KE, Flores AR, Pavelka MS, Jr, Braunstein
M The twin-arginine translocation pathway of Mycobacterium
smegmatis is functional and required for the export of my-

Merhej V, Raoul D Rickettsial evolution in the light of compar-

Molik S, Karnauchov I, Weidlich C, Herrmann RG, Klosgen RB
The Rieske Fe/S protein of the cytochrome b6/f complex in
chloroplasts: missing link in the evolution of protein trans-

Moncalian G, Cabezón E, Alkorta I, Valle M, Moro F, Valpuesta
JM, Goni FM, de La Cruz F Characterization of ATP and DNA
binding activities of TrwB, the coupling protein essential in

Morse K, Norimine J, Palmer GH, Sutten EL, Baszler TV, Brown
WC Association and evidence for linked recognition of type
IV secretion system proteins VirB9-1, VirB9-2, and VirB10 in

Nagai H, Cambronne ED, Kagan JC, Amor JC, Kahn RA, Roy
CR A bacterial gua-
nine nucleotide exchange factor activates ARF on Legionella

Nagai H, Cambronne ED, Kagan JC, Amor JC, Kahn RA, Roy
CR A C-terminal translocation signal required for Dot/Icm-
dependent delivery of the Legionella RalF protein to host cells.

Nakayama K, Yamashita A, Kurokawa K, et al. The Whole-
genome sequencing of the obligate intracellular bacterium
Orientia tsutsugamushi revealed massive gene amplification

Narita S ABC transporters involved in the biogenesis of the
outer membrane in gram-negative bacteria. Biosci Biotechnol

Narita S, Tanaka K, Matsuyama S, Tokuda H Disruption of IolCDE,
encoding an ATP-binding cassette transporter, is lethal for
Escherichia coli and prevents release of lipoproteins from the

Nelson CM, Herron DC, Felsheim RF, Schloder BR, Grindle SM,
Chavez AO, Kurtti TJ, Munderloh UG Whole genome tran-
scription profiling of Anaplasma phagocytophilum in human
and tick host cells by tiling array analysis. BMC Genomics
2008;9:364.

Neumann-Haefelin C, Schafer U, Muller M, Koch HG SRP-
dependent co-translational targeting and SecA-dependent
translocation analyzed as individual steps in the export of

Nicosia A, Perugini M, Franzini C, Casagli MC, Borri MG, Antoni G,
Almoni M, Neri P, Ratti G, Rappuoli R Cloning and sequenc-
ing of the pertussis toxin genes: operon structure and gene

Nishi J, Sheikh J, Mizuguchi K, Luisi B, Burland V, Boutin A, Rose
DJ, Blatterner FR, Nataro JP The export of coat protein from en-
teroaggregative Escherichia coli by a specific ATP-binding cas-

Nishie M, Shioya K, Nagao J, Jikuya H, Sonomoto K ATP-
dependent leader peptide cleavage by NukT, a bifunctional
ABC transporter, during lantibiotic biosynthesis. J Biosci Bio-

Niu H, Kozjak-Pavlovic V, Rudel T, Rikhisa Y Anaplasma
phagocytophilum Als-1 is imported into host cell mitochond-
dria and interferes with apoptosis induction. PLoS Pathog
2010;6:e1000774.

Noh SM, Brayton KA, Brown WC, Norimine J, Munske GR, Davitt
CM, Palmer GH Composition of the surface proteome of
Anaplasma marginale and its role in protective immunity
induced by outer membrane immunization. Infect Immun

Nouwen N, Piwowarek M, Berrelkamp G, Driessen AJ The large
first periplasmic loop of SecD and SecF plays an impor-

Nunez PA, Soria M, Farber MD The twin-arginine translocation
pathway in alpha-proteobacteria is functionally preserved ir-
respective of genomic and regulatory divergence. PLoS ONE
2012;7:e33605.

Oberhettinger P, Schutz M, Leo JC, Heinz N, Berger J, Aut-
enrieth IB, Linke D Intimin and invasin export their C-
terminus to the bacterial cell surface using an inverse mech-
anism compared to classical autotransport. PLoS ONE 2012;7:
e47069.

Ochsner UA, Snyder A, Vasil AI, Vasil ML Effects of the twin-
arginine translocase on secretion of virulence factors, stress

Ogata H, Renesto P, Audic S, Robert C, Blanc G, Fournier PE,
Parinello H, Claverie JM, Raoul D The genome sequence of
Rickettsia felis identifies the first putative conjugative plas-
mid in an obligate intracellular parasite. PLoS Biol 2005;3:
e248.

Ogata H, La Scola B, Audic S, Renesto P, Blanc G, Robert C,
Fournier PE, Claverie JM, Raoul D Genome sequence of Rik-
ettsia bellii illuminates the role of amoebae in gene exchanges

Ogawa M, Renesto P, Azza S, Moinier D, Fourquet P, Gorvel
JP, Raoul D Proteome analysis of Rickettsia felis highlights
the expression profile of intracellular bacteria. Proteomics

Ojcius DM, Thibon M, Mounier C, Dautry-Varsat A pH and cal-
cium dependence of hemolysis due to Rickettsia prowazekii:
comparison with phospholipase activity. Infect Immun

Okuda S, Tokuda H Lipoprotein sorting in bacteria. Annu Rev Mi-

Oliver DC, Huang G, Fernandez RC Identification of secretion de-
terminants of the Bordetella pertussis BrkA autotransporter.

Oliver JD, Burkhardt NY, Felsheim RF, Kurtti TJ, Munderloh UG
Motility characteristics are altered for Rickettsia bellii trans-
formed to overexpress a heterologous rickA gene. Appl Envi-

Oomen CJ, van Ulsen P, van Gelder P, Feijen M, Tommassen J,
Gross P Structure of the translocator domain of a bacterial au-

Paetzel M, Karla A, Strynadka NC, Dalbey RE Signal peptidases.

Palmer T, Berks BC The twin-arginine translocation (Tat) protein

Ankyrin repeat proteins comprise a diverse family of bacte-
Mycobacterium tuberculosis
Infect Immun
autotransporter protein Sca1
J Bacteriol
2010;286:35096–103.

Pavlova O, Peterson JH, Ieva R, Bernstein HD
Mechanistic link between beta barrel assembly and the initiation of autotransporter secretion.

Unexpected diversity of signal peptides in prokaryotes.
MBio 2012;3 pii e00339–12.

Perlman SJ, Hunter MS, Zchori-Fein E
The emerging diversity of Rickettsia.

Petersen TN, Brunak S, von Heijne G, Nielsen H
SignaP 4.0: discriminating signal peptides from transmembrane regions.
Nat Methods 2011;8:785–86.

Peterson JH, Tian P, Ieva R, Dautin N, Bernstein HD
Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment.

Pichon S, Bouchon D, Cordaux R, Chen L, Garrett RA, Greve P
Conservation of the Type IV secretion system throughout Wolbachia evolution.

Pohlner J, Halter R, Beyreuther K, Meyer TF
Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA protease.

Pornsriwong W, Bourchookarn A, Paddock CD, Macaluso KR
Proteomic analysis of Rickettsia parkeri strain portsmouth.

Contribution of the twin arginine translocation system to the virulence of enterohemorrhagic Escherichia coli O157:H7.

Prakash A, Yogeeshwari S, Sircar S, Agrawal S
Protein domain of unknown function 3233 is a translocation domain of autotransporter secretory mechanism in gamma proteobacteria.
PloS ONE 2011;6:e25570.

Price CE, Driessen AJ
YidC is involved in the biogenesis of anaerobic respiratory complexes in the inner membrane of Escherichia coli.

Price CE, Driessen AJ
Conserved negative charges in the transmembrane segments of subunit K of the NADH:ubiquinone oxidoreductase determine its dependence on YidC for membrane insertion.

Cysteine scanning mutagenesis and topological mapping of the Escherichia coli twin-arginine translocase TatC Component.

Qi HY, Bernstein HD
SecA is required for the insertion of inner membrane proteins targeted by the Escherichia coli signal recognition particle.

Proteome analysis and serological characterization of surface-exposed proteins of Rickettsia heilongjiangensis.
PloS ONE 2013;8:e70440.

Radulovic S, Troyer JM, Beier MS, Lau AO, Azad AF
Identification and molecular analysis of the gene encoding Rickettsia typhi hemolysin.

Rahman MS, Simser JA, Macaluso KR, Azad AF
Molecular and functional analysis of the lwp gene, encoding a type I signal peptidase from Rickettsia rickettsii and Rickettsia typhi.

Rahman MS, Simser JA, Macaluso KR, Azad AF
Functional analysis of secA homologues from rickettsiae.

Rahman MS, Ceral SM, Drehner-Lesnick SM, Beier MS, Azad AF
The lwpA gene, encoding the type II signal peptidase of Rickettsia typhi: transcriptional and functional analysis.

Rahman A, Srivastava SS, Sneh A, Ahmed N, Krishnasasrty
MV Molecular characterization of YlaA gene product, Rv1694 of Mycobacterium tuberculosis: a non-conventional hemolysin and a ribosomal RNA methyl transferase.
M Biochem 2010a;11:35.

Rahman MS, Ammerman NC, Sears KT, Ceral SM, Azad AF
Functional characterization of a phospholipase A(2) homolog from Rickettsia typhi.

Rahman MS, Gillespie JJ, Kaur SJ, Sears KT, Ceral SM, Beier-Sexton M, Azad AF
Rickettsia typhi possesses phospholipase A2 enzymes that are involved in infection of host cells.

Rances E, Voronin D, Tran-Van V, Mavingui P
Genetic and functional characterization of the type IV secretion system in Wolbachia.

Reed SC, Serio AW, Welch MD
Rickettsia parkeri invasion of diverse host cells involves an Arp2/3 complex, WAVE complex and Rh-family GTPase-dependent pathway.

Reed SC, Lamason RL, Ricsa VI, Abernathy E, Welch MD
Rickettsia actin-based motility occurs in distinct phases mediated by different actin nucleators.

Rego AT, Chandran V, Waksman G
Two-step and one-step secretion mechanisms in Gram-negative bacteria: contrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis.

Fiber formation across the bacterial outer membrane by the chaperone/usher pathway.

Renesto P, Dehoux P, Gouin E, Touqui L, Cossart P, Raoult D
Identification and characterization of a phospholipase D-superfamily gene in rickettsiae.

Proteome analysis of Rickettsia conorii by two-dimensional gel electrophoresis coupled with mass spectrometry.

Identification of two putative rickettsial adhesins by proteomic analysis.

Proteomic analysis of the sarcosine-insoluble outer membrane fraction of the bacterial pathogen Bartonella henselae.

Rikihisa Y, Lin M
Anaplasma phagocytophilum and Ehrlichia chaffeensis type IV secretion and Ank proteins.

Rikihisa Y, Lin M, Niu H
Type IV secretion in the obligatory intracellular bacterium Anaplasma phagocytophilum.


