1 **Title:** Description of a widespread bacterial secretion system with chemically diverse substrates

2 **Running title:** The Type Eleven Secretion System

3 **Byline:** Alex S. Grossman[a¶], Terra J. Mauer[b* ¶], Katrina T. Forest[b], and Heidi Goodrich-Blair[a,b#]

4

5 **Affiliations:**

6 [a]University of Tennessee-Knoxville, Department of Microbiology, Knoxville, TN

7 [b]University of Wisconsin-Madison, Department of Bacteriology, Madison, WI

8 #Address correspondence to Heidi Goodrich-Blair, hgblair@utk.edu

9 *Present affiliation: Terra Mauer, Washtenaw Community College, Ann Arbor, MI

10 ¶ These authors contributed equally to this work.

Abstract:

In host-associated bacteria, surface and secreted proteins mediate acquisition of nutrients, interactions with host cells, and specificity of tissue-localization. In Gram-negative bacteria, the mechanism by which many proteins cross or become tethered to the outer membrane remains unclear. The domain of unknown function (DUF)560 occurs in outer membrane proteins throughout Proteobacteria and has been implicated in host-bacteria interactions and lipoprotein surface exposure. We used sequence similarity networking to reveal three subfamilies of DUF560 homologs. One subfamily includes those DUF560 proteins experimentally characterized to date: NilB, a host-range determinant of the nematode-mutualist *Xenorhabdus nematophila*, and the surface lipoprotein assembly modulators Slam1 and Slam2, which facilitate surface exposure of lipoproteins in *Neisseria meningitidis* (1, 2). We show that DUF560 proteins from a second subfamily facilitate secretion of soluble, non-lipidated proteins across the outer membrane. Using *in silico* analysis, we demonstrate that DUF560 gene complement correlates with bacterial environment at a macro level and host association at a species level. The DUF560 protein superfamily represents a newly characterized Gram-negative secretion system capable of lipoprotein surface exposure and soluble protein secretion with conserved roles in facilitating symbiosis. In light of these data, we propose that it be titled the type eleven secretion system (TXISS).

Importance:

The microbial constituents of a host associated microbiome are decided by a complex interplay of microbial colonization factors, host surface conditions, and host immunological responses. Filling such niches requires bacteria to encode an arsenal of surface and secreted proteins to effectively interact with the host and co-occurring microbes. Bioinformatic predictions of the localization and function of putative bacterial colonization factors are essential

35 for assessing the potential of bacteria to engage in pathogenic, mutualistic, or commensal
36 activities. This study uses publicly available genome sequence data, alongside experimental
37 results from representative gene products from *Xenorhabdus nematophila*, to demonstrate a role
38 for DUF560 family proteins in the secretion of bacterial effectors of host interactions. Our
39 research delineates a broadly distributed family of proteins and enables more accurate
40 predictions of the localization of colonization factors throughout Proteobacteria.

41 **Introduction:**

42 All plants and animals exist in association with bacterial symbionts that can contribute to
43 nutrition, protection, development, and reproduction. These symbionts express surface and
44 secreted proteins that facilitate host interactions through a variety of functions, including
45 acquisition of nutrients (e.g. TonB-dependent transporters in Gram-negative host-associated
46 bacteria (3, 4)), interaction with host cells (e.g. *Acinetobacter baumanii* OmpA binding host
47 desmoplakin (5)), and specificity in host-range and tissue-localization (e.g. *Vibrio fischeri* pilin
48 variability mediating host specificity (6)).

49 Recently, a mechanism of lipoprotein surface tethering was identified in the human
50 pathogen *Neisseria meningitidis* and termed the Slam (Surface lipoprotein assembly modulator)
51 machinery (1, 2). Slam proteins containing the β-barrel domain DUF560 are required for surface
52 presentation of certain lipoproteins, including those that capture metal-carrying compounds used
53 by hosts to sequester nutrients from bacteria (1). Two *N. meningitidis* Slam proteins have been
54 characterized, each with distinct lipoprotein substrates. Slam activity also has been demonstrated
55 for DUF560 representatives from pathogens *Pasteurella multocida, Moraxella catarrhalis,* and
56 *Haemophilus influenzae* (1, 2). However, most Slam homologs have no bioinformatically
57 predicted substrate to date, and one study has found that *N. meningitidis* Slam1 can surface

58 expose unlipidated fHbp variants (7). Thus, the full functional potential of DUF560 proteins is
59 not yet known.

60 The DUF560 homolog NilB is a host-association and species-specificity factor in the
61 nematode symbiont *Xenorhabdus nematophila*, a proteobacterium in the family *Morganellaceae*
62 (8-10). A screen for *X. nematophila* mutants defective in colonizing *Steinernema carpocapsae*
63 intestines revealed the Nematode intestinal localization locus (Nil)(10, 11). The Nil locus
64 contains the genes *nilB* and *nilC*, each of which is independently necessary for colonization of
65 nematodes. Biochemical and bioinformatic analyses have established that NilC is an outer-
66 membrane-associated lipoprotein and NilB is an outer-membrane β–barrel protein in the
67 DUF560 protein family with a ~140 amino acid periplasmic N-terminal domain that contains
68 tetratricopeptide repeats (11-14).

69 To begin to understand the range of functions of DUF560 proteins, we assessed their
70 ecological distribution, genomic context, and relatedness. We experimentally examined the *X.*
71 *nematophila* DUF560 homolog HrpB, which is not predicted to interact with a lipoprotein
72 substrate. Finally, to better understand the potential role of DUF560 proteins in host-symbiont
73 interactions we analyzed DUF560 distribution among symbiotic *Xenorhabdus*. Our data
74 demonstrate that the activities of the DUF560 family extend beyond lipoprotein surface
75 presentation and constitute a new type XI bacterial secretion system (TXISS) which, like the
76 type II secretion system, is capable of acting on either membrane anchored or soluble proteins
77 (15).

78 **Results:**

79 **TXISS cluster according to environment**

80 Using homology to NilB or Slam proteins, previous work identified a wide distribution of
81 DUF560 proteins within mucosal associated bacteria (2, 10, 11). To quantifiably delineate
82 subfamilies within the TXISS we generated a sequence similarity network (SSN) using the
83 Enzyme Function Initiative toolset (EFI) (16-18) and annotated it to highlight environmental
84 source or taxonomic grouping of the isolates containing DUF560 homologs (Fig 1; Supplemental
85 Table 1). In this network analysis, protein sequences with a high identity (40% or greater) were

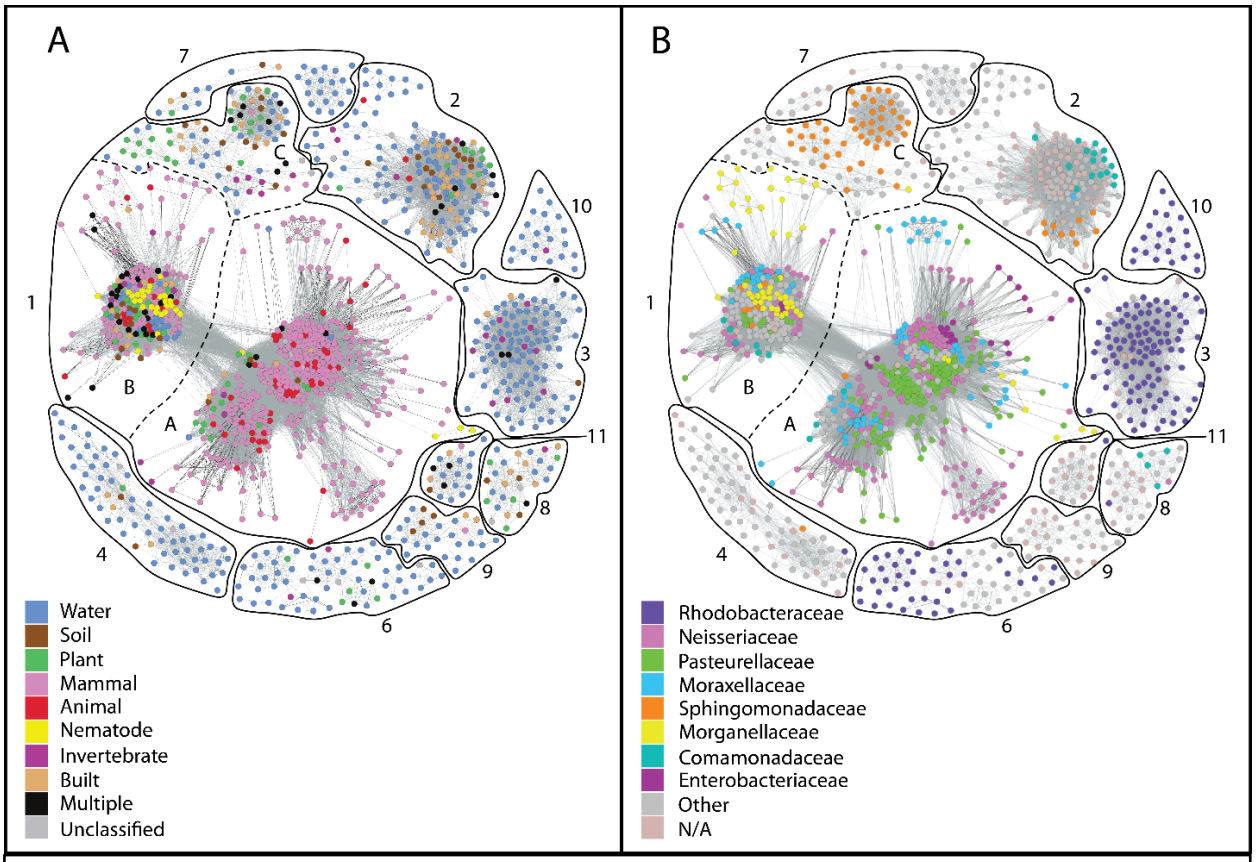


Figure 1. Sequence similarity network of all DUF560 proteins. SSN of DUF560 homologs generated by EFI-EST as accessed on 4.24.19 (16-18). Edges were drawn using an alignment score of 38, and any sequences which shared ≥ 40% identity were placed in a single node to allow the separation of clusters. Each node represents a group of highly similar sequences, with edge darkness demonstrating similarity, and the distance between nodes being determined via the Fruchterman-Reingold algorithm to optimize edge lengths (53). Each node was color coded to show the isolates' environmental origin(s) (A) and taxonomic class (B).

86 gathered into data points called nodes and connected together with edges based on sequence
87 similarity. Using all homologs in Interpro 73 and Uniprot 2019-02, we identified 10 major
88 clusters of TXISS proteins. Cluster 1 was chosen for in-depth analysis since it contained the
89 majority of nodes in the network (62.4%) and could be visually divided into three subclusters
90 (1A, 1B, and 1C) using force directed node placement (Fig. 2). The remaining clusters displayed
91 a preponderance of water and soil associated organisms and contained no characterized proteins.
92 Consistent with our previous observations (11), the majority of cluster 1 nodes (75%) comprise
93 sequences from various animal-associated isolates, while another 20% contain sequences from
94 marine, freshwater, soil, or built-environment isolates.

95 The division of nodes among the three subclusters more strongly reflected environmental
96 origin than bacterial taxonomy (Fig. 2A). Subcluster 1A almost exclusively comprises animal-
97 associated bacteria and contains all previously characterized TXISS, which separate according to
98 predicted substrate when analyzed with higher stringency (Supplemental Fig. 1). Subclusters 1B
99 and 1C have no previously characterized representatives. Subcluster 1B contains a mixture of
100 sequences from host-associated and free-living bacteria, while subcluster 1C contains sequences
101 largely from environmentally isolated *Sphingomonadaceae*. Subclusters 1A and 1B do not
102 correlate well with taxonomy of the isolates (Fig. 2B). For example, cluster 1 contains 81 nodes
103 with *Moraxellaceae* sequences. Of these, 79% are predominantly animal-associated and 12.3%
104 are predominantly environmental isolates. The animal-associated isolates are enriched in
105 subcluster 1A and environmental-isolates are enriched in subcluster 1B. These data demonstrate
106 a correlation with lifestyle (e.g. free-living vs host associated) as opposed to taxonomy and
107 suggest that subclusters indicate divergent molecular functions.

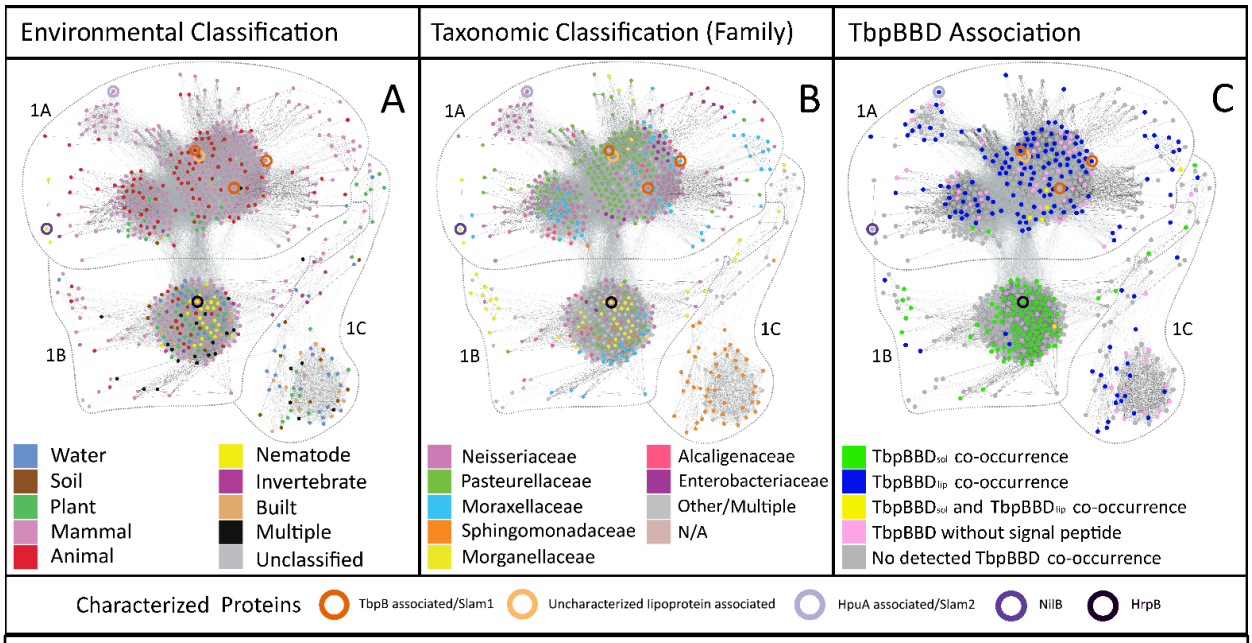


Figure 2. Cluster 1 of the TXISS sequence similarity network (SSN) forms subclusters according to environment of isolation and signal sequence of predicted cargo. All nodes from cluster 1 of the TXISS SSN were positioned using the Fruchterman-Reingold algorithm. The resulting graph was annotated according to sequences A) environmental origin(s) or B) taxonomic class of the isolates, or C) whether the node homolog(s) co-occur with a TbpBBD-domain encoding gene. Nodes containing experimentally characterized proteins are highlighted using colored circles as indicated at the bottom of the figure.

TXISS cluster according to substrate

The Slam acronym was defined on the basis that DUF560 homologs from *N. meningitidis* facilitate surface exposure of lipoproteins, such as TbpB, LbpB, HpuA, and fHbp which are frequently encoded nearby (1, 2). The lipid tail is not essential for Slam-dependent surface exposure of a target (1, 7, 19). This result prompted us to consider whether co-occurrence with lipoproteins is a constant throughout cluster 1. We used the EFI Genome-Neighborhood Tool (16, 18, 20) to assay the genomic context of each subcluster. This analysis corroborated previous work demonstrating genomic association of DUF560 proteins with TonB, TonB-dependent receptors and proteins that have a Pfam TbpB_B_D domain, which will be referred to hereafter as TbpBBD (2, 21).

119 Given the prevalence of TbpBBD domains in the genome neighborhoods of DUF560
120 genes, and their known occurrence in lipoproteins surface exposed by Slams we examined their
121 gene structures to detect potential patterns. Using a combination of genome-neighborhood-
122 network data (18), Rapid ORF Description & Evaluation Online (RODEO) data (22), and
123 manual annotation we analyzed all TbpBBD domain proteins co-inherited with sequences
124 present in cluster 1 (Supplemental Table 2). The majority of TbpBBD-bearing proteins
125 associated with subcluster 1A are predicted to be lipidated and have two TbpBBD domains
126 similar to TbpB in *N. meningitidis* (2), therefore they will be referred to as $TbpBBD_{lip}$. In
127 contrast, the TbpBBD-bearing proteins associated with subcluster 1B are almost exclusively

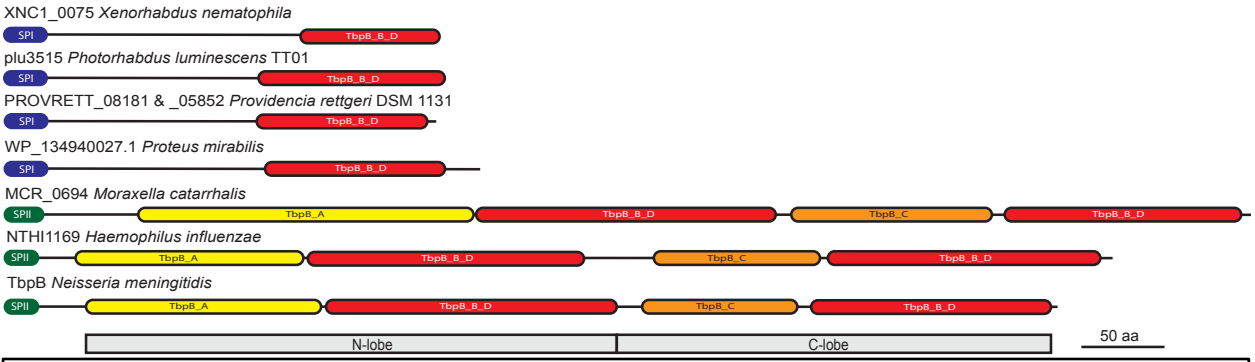


Figure 3. Examples of $TbpBBD_{sol}$ and $TbpBBD_{lip}$ domain architectures. The schematic diagram shows features found in select TbpBBD-domain containing proteins predicted to be exported by TXISS mechanisms. TXISS-1B associated proteins from *X. nematophila, P. luminescens, P. rettgeri, P. mirabilis,* and *H. haemolyticus* have signal-peptidase 1 (SPI) signal sequences, lack annotated "handle" domains, and have a single TbpBBD domain. TXISS-1A associated proteins from *M. catarrhalis, H. influenzae, and N. meningitidis* have signal-peptidase 2 (SPII) type signal sequences, N-lobe and C-lobe handle domains (TbpB_A and TbpB_C, respectively), and two TbpBBD domains. Predicted signal peptidase sites and domains are color-coded.

128 predicted to be soluble proteins and only have a single TbpBBD domain similar to hemophilin in
129 *Haemophilus haemolyticus*, therefore they will be referred to as $TbpBBD_{sol}$ (Fig. 2C; Fig. 3).
130 Biochemical and structural evidence support the conclusion that hemophilin is a soluble
131 secreted protein that binds free heme and facilitates heme uptake into the cell (23). Three-
132 dimensional homology modeling (Phyre[2]) was used to visualize potential structural similarities

133 between hemophilin and several $TbpBBD_{sol}$ proteins (24, 25) (Supplemental Fig. 2). In light of
134 sequence and structural level similarities we hypothesized that subcluster 1B proteins transport
135 $TbpBBD_{sol}$ proteins across the outer membrane via a mechanism analogous to subcluster
136 1A proteins transporting $TbpBBD_{lip}$ substrates. Our prediction separates the TXISS into two
137 subfamilies, TXISS-1A which has members that flip $TbpBBD_{lip}$ substrates across the outer
138 membrane and TXISS-1B which we predict has members that secrete $TbpBBD_{sol}$ substrates into
139 the extracellular milieu.
140
141 **TXISS-1B activity reconstructed *in vivo***
142 To test our prediction that the TXISS-1B can secrete $TbpBBD_{sol}$ substrates, we
143 investigated the heme receptor protein (Hrp) locus of *X. nematophila.* This locus, which is
144 conserved across the *Xenorhabdus* genus, consists of genes predicted to encode TonB, a TonB-
145 dependent heme receptor named HrpA (XNC1_0073), the TXISS-1B homolog HrpB
146 (XNC1_0074), and its predicted $TbpBBD_{sol}$ substrate HrpC (XNC1_0075) (11), a homolog of
147 the heme-binding protein hemophilin (55% similar 39% identical) (23). Specifically, we sought
148 to test whether the TXISS-1B homolog HrpB mediates secretion of the putative heme-binding
149 protein HrpC. Using the expression vector pETDuet-1, HrpC-FLAG was expressed with and
150 without HrpB-FLAG and whole cell and supernatant fractions were separated and analyzed by
151 immunoblotting with anti-FLAG antibodies (26). We found that in the presence of HrpB-FLAG
152 the levels of HrpC-FLAG detected in the supernatant increased 9.9-fold at 1-h post-induction and
153 17.0-fold at 2.5-h post-induction (Fig. 4A). Whole cell lysates demonstrated equivalent
154 expression of HrpC-FLAG in both treatments (Fig. 4B). Coomassie staining of supernatants
155 show that expression of HrpB-FLAG did not induce cell lysis or non-specific protein secretion

156 (Fig. 4C, Supplemental Fig. 3). Consistent with the fact that in the absence of Slam1, several
157 strains of *E. coli* surface expose a fraction of the substrate fHbp (27, 28), some HrpC-FLAG
158 reached the supernatant in the absence of HrpB-FLAG.

159 The Hrp locus is conserved in many proteobacteria including *Proteus mirabilis*,
160 *Providencia rettgeri, Acinetobacter baumannii, N. gonorrhoeae*, and *H. haemolyticus.* For
161 example, the $TbpBBD_{sol}$ protein present in *N. gonorrhoeae* (NGO0554) is flanked by a gene
162 predicted to encode a TonB-dependent receptor (NGO0553) and a gene predicted to encode a
163 DUF560 family protein (NGO0555). Consistent with the predicted role of these $TbpBBD_{sol}$
164 proteins in metal homeostasis, NGO0554 is repressed by iron, upregulated in response to
165 oxidative stress, and contributes to resistance to peroxide challenge (29-31). These TXISS-
166 dependent HrpC homologs likely contribute to metal homeostasis in a variety of human
167 microbiome constituents by acting as hemophores akin to hemophilin and HasA (23, 32, 33).

168

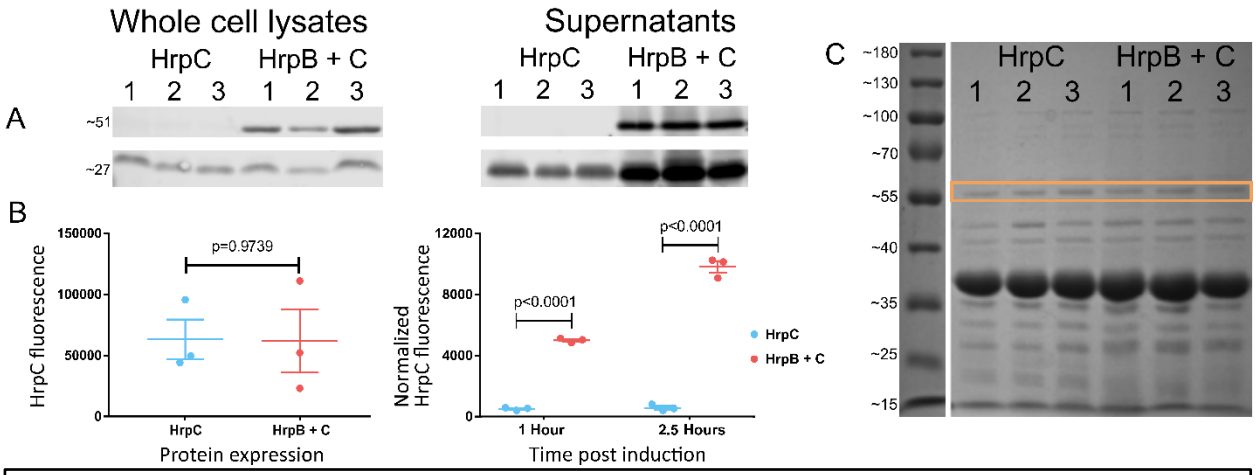


Figure 4. HrpB increases secretion of HrpC. A) Representative immunoblots of 2.5 h post-induction whole cell lysates and supernatant samples. HrpC-FLAG and HrpB-FLAG were distinguished by molecular weights (~27 and ~51 kDa, respectively). B) Fluorescence intensity of HrpC levels was not significantly impacted by HrpB in cell lysates (left graph, unpaired t-test) but was in supernatants (right graph, Sidak's multiple comparisons test). C) Representative Coomassie stained SDS-PAGE gel of supernatant samples indicating HrpB causes neither cell lysis nor nonspecific protein export. A distinct band (in orange) was used as a loading control to normalize supernatant samples. The increase in supernatant HrpC is faintly visible at ~27 kilodaltons in the last three supernatant lanes. Relicates labeled 1-3.

169

170

Host environment drives TXISS class

Having established that DUF560 homologs represent a bona fide secretion system, we next used the *Xenorhabdus* system to expand on our observation that the presence and type of TXISS corresponds to bacterial environmental niche. *Xenorhabdus* are species-specific obligate mutualists of *Steinernema* nematodes, and NilB is a known host range determinant. Therefore, we considered host species as an environmental niche and bioinformatically examined whether the complement of DUF560 genes in a microbe corresponds with host phylogeny (34-38). All 46 *Xenorhabdus* genomes on the Magnifying Genome (MaGe) platform (39) encode between one and three TXISS, with five distinct homologs represented across the genus (one TXISS-1A and four TXISS-1B) (40). Each unique combination of TXISS homologs was assigned one of six classes (A-F) as depicted in Supplemental Data Fig. 4 and Supplemental Table 3. To visualize

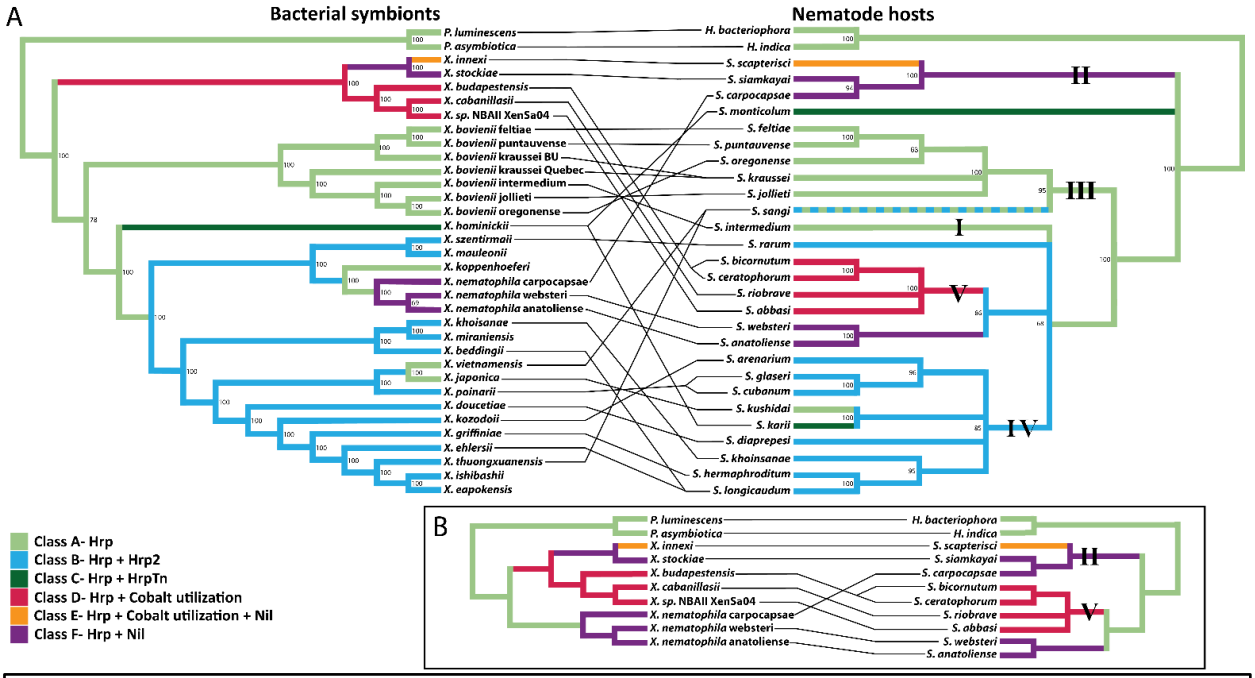


Figure 5: Cladograms of *Xenorhabdus* and *Steinernema* color coded according to *Xenorhabdus* DUF560 Class. Co-phylogeny of nematode species and their colonizing bacteria A) across the *Steinernema* genus or B) with a focus on specific clades. Numbers on branches indicate bootstrap support values. Bootstrap values below 60% were contracted. Lines connecting the phylogenies indicate mutualist pairs. Roman numerals highlight the 5 Steinernema clades described by Spridinov et al. 2004. Colored overlays indicate the DUF560 class of a given bacterium or a given nematode's symbiont. Class A (light green); class B (light blue); class C (dark green); class D (red); class E (orange); class F (purple).

182 correlations between TXISS class and host identity, we constructed Maximum Likelihood and
183 Bayesian phylogenetic trees for *Xenorhabdus* and *Steinernema*. *Xenorhabdus* trees were
184 generated with whole genome data while *Steinernema* trees used five loci as available
185 (Supplemental Fig. 5; Supplemental Table 4). Aligning the Maximum likelihood phylogenies
186 reveals that the TXISS complement of a *Xenorhabdus* species is more predictive of nematode
187 host than phylogenetic position (Fig. 5).

188 This alignment of the phylogenetic placement of a given *Steinernema* host with the
189 TXISS complement of the symbiont provides insights into the nematode internal environment
190 experienced by the symbiont. For example, most class B/C *Xenorhabdus* with two HrpB
191 paralogs at the Hrp locus are symbionts of nematodes within the phylogenetic Clade IV,
192 suggesting that these nematodes present a distinctive environment in which an additional Hrp
193 locus is beneficial. Class D *Xenorhabdus* with an HrpB paralog encoded adjacent to genes
194 predicted to encode a cobalt ABC transporter, are solely symbionts of Clade V nematodes. *X.*
195 *innexi* and *X. stockiae* have seemingly diverged from this lineage through acquisition of a *nilB*
196 homolog and switching to hosts within Clade II. Similarly, *X. nematophila* independently gained
197 *nilB* and switched into a Clade II host (Fig. 5B). These acquisitions, alongside the varied
198 genomic contexts of *nilB*/*nilC* pairs, are consistent with previous suggestions that the Nil locus
199 was horizontally acquired among *Xenorhabdus* (14).

200 The TXISS NilB and the lipoprotein NilC are necessary for *X. nematophila* to colonize
201 the Clade II nematode *S. carpocapsae* (10, 14). However, *X. nematophila* also colonizes two
202 nematodes that are phylogenetically separate from Clade II, *S. anatoliense* and *S. websteri* (Fig.
203 5). Our hypothesis that TXISS are involved in host-environment adaptations leads to the
204 prediction that *X. nematophila* will require the Nil locus to colonize these nematodes. To test this

205 hypothesis, bacteria-free *S. anatoliense*, *S. websteri*, and *S. carpocapsae* eggs (41) were exposed
206 to an *X. nematophila* ATCC19061 *Δnil* mutant and a Nil complemented strain (11). Consistent

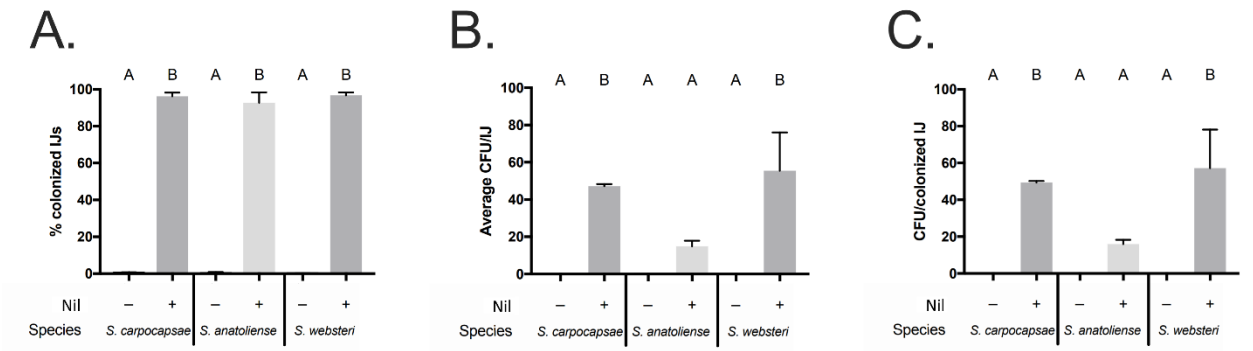


Figure 6. The Nil locus is necessary for colonization of *S. anatoliense*, *S. carpocapsae*, and *S. websteri* infective juveniles. Bacteria free *S. carpocapsae, S. anatoliense,* and *S. websteri* were exposed to GFP expressing *X. nematophila* lacking or bearing the Nil locus. The resulting progeny infective juveniles (IJs) were monitored for colonization either by A) microscopy or by B) plating lysates for average CFU/IJ. C) The average CFU per colonized IJ combines both of these values to show bacterial load per organism. Treatments were analyzed via one-way ANOVA and Tukey's post-hoc test.

207 with our prediction, the *Δnil* mutant was deficient in infective juvenile colonization in all three
208 nematode species (Fig. 6), demonstrating that *nil* genes are necessary for infective juvenile
209 colonization of *S. anatoliense* and *S. websteri* and supporting our hypothesis that TXISS promote
210 adaptation to host environments.

211 **Discussion:**

212 Bacteria that live inside animal hosts face many challenges as a result of their
213 environment. Gram-negative bacteria that occupy host environments produce diverse factors that
214 manipulate host cell and tissue physiology for their own benefit. Delivery of these effectors to
215 their site of action requires the crossing of two bacterial membranes and possibly additional host
216 cell membranes depending on the effector. Bacteria have evolved specialized secretion systems
217 for delivery of effectors that facilitate the host-associated lifestyle. For instance, the type II
218 secretion system can secrete hydrolyzing enzymes essential for commensals to grow, but it can
219 also secrete clinically important toxins such as Cholera toxin and Diphtheria toxin (42, 43).
220 Other secretion systems such as type III and type VI are specialized for the delivery of effectors

221 across both bacterial and host cell membranes. Knowledge of cargo protein identities and sorting
222 processes facilitates predictions from genomic information of bacterial secretome composition,
223 regulation, and localization. Despite the diverse secretion systems now recognized, there are
224 more to be identified based on the fact that some proteins predicted to be secreted lack known
225 secretion pathways (44). The type X secretory pathway was described in 2020 and shed light on
226 long standing mysteries surrounding the dependence on a muramidase for secretion of Typhoid
227 toxin across the cell wall (45). Secretion pathways are the tools with which bacteria interact with
228 their world, and as such each one discovered provides a slightly more complete picture of the
229 toolbox.

230 The DUF560 (domain of unknown function) family presence in animal-associated
231 bacteria was first recognized when it was noted that the host-colonization factor NilB has
232 homologs in *N. meningitidis, M. catarrhalis, P. multocida,* and *H. influenzae* (10). This
233 observation was strengthened by subsequent demonstration that Slam1 from human pathogenic
234 bacteria facilitates the surface presentation of host-metal acquisition proteins (1, 2, 10). Enabled
235 by the availability of extensive genomic data from many environments and ever-improving
236 bioinformatic visualization tools, we have presented data demonstrating that the type XI
237 secretion system is a broadly distributed molecular vehicle for moving proteins across the Gram-
238 negative outer membrane. While DUF560 proteins were originally thought to represent a
239 mechanism for lipoprotein surface exposure (1), recent studies have expanded that functional
240 range to include peripheral membrane proteins (7) and now soluble secreted proteins,
241 establishing the function of DUF560 OM proteins in secretion of varied substrates. We predict
242 that further study will uncover a great diversity of cargo for the distinct classes of TXISS
243 revealed through our network analysis. Here we focused on one single cluster of a DUF560

244 sequence similarity network and highlighted two subfamilies of TXISS (TXISS-1A and 1B). The
245 remaining 9 clusters likely represent diverse subfamilies responsible for transporting as-yet-
246 unknown cargo. The lipoprotein cargo proteins for which structural data are known revealed a
247 common C-terminal 8-stranded β-barrel motif (2). Our discovery that HrpC is a cargo protein for
248 TXISS-1B HrpB strengthens the concept that this barrel is an important characteristic of TXISS
249 cargo; HrpC is a homolog of the *H. haemolyticus* hemophilin, the structure of which likewise
250 adopts a C-terminal 8-stranded β-barrel (23) (Supplemental Fig. 2). These data support the
251 concept that TXISS cargo have bifunctional structures in which the N-terminus is the host
252 effector domain while the C-terminus directs the protein for secretion. This framework will
253 facilitate identification of as-yet-unknown TXISS cargo among the genes that co-occur with
254 TXISS OMPs. Further, our analyses suggest that sequence similarity network clusters have
255 predictive power for other characteristics of TXISS cargo, including whether they are surface-
256 attached or secreted. The network-enabled classification presented here will facilitate the
257 investigation of both TXISS OMPs and their cargo in diverse bacteria.

258 Our work enables categorization of TXISS within *Neisseria* species that are found in
259 many animal hosts, including the human pathogens *N. meningitidis* and *N. gonorrhoeae* (46).
260 *Neisseria* strains can encode up to 6 TXISS paralogs. *N. meningitidis* MC58 has two functionally
261 characterized TXISS: Slam1 and Slam2. Our network analysis indicates *N. meningitidis* encodes
262 a third SPAM, NMB1466/ NP_274965, that also falls within Cluster 1A. *N. gonorrhoeae* TXISS
263 are represented in more nodes than *N. meningitidis*, notably occupying 19.3% of nodes in
264 subcluster 1B (Fig. 2). Taken together, these data indicate that *Neisseria* may be using TXISS to
265 secrete both lipoproteins and soluble products.

266 There is a broadly conserved role for the TXISS in host-microbe interactions. Using the
267 *Steinernema-Xenorhabdus* symbiosis we have demonstrated that the composition of TXISS in a
268 bacterial symbiont genome correlates with host organism. During the evolutionary history of
269 *Xenorhabdus* the gain or loss of TXISS loci correlated with host switching events (Fig. 5). Given
270 the conservation of the TXISS Hrp locus, amongst all *Xenorhabdus* and throughout human
271 microbiome constituents, it will be important in the future to examine the regulation of Hrp
272 TXISS-dependent secretion and the roles of Hrp machinery in binding and acquiring host metals
273 in a mucosal environment.

274 Finally, the whole family network first generated from the PF04575 (DUF560) dataset
275 included cluster 5, comprising 14 nodes of *Klebsiella* homologs and 2 nodes with *Klebsiella* and
276 *Escherichia* homologs. However, based on its limited number of nodes and predicted topology
277 differences relative the rest of the network (16-stranded vs. 14-stranded barrel), cluster 5 was
278 filtered out prior to our in-depth analysis. Cluster 5 homologs are annotated as PgaA/HmsH
279 OMPs responsible for the synthase-dependent secretion of the exopolysaccharide poly-beta-1,6
280 N-acetyl-D-glucosamine (PNAG), which is necessary for biofilm formation (47, 48). The
281 structure of the C-terminus of *E. coli* PgaA has been solved and forms a 16-stranded porin
282 through which PNAG is exported (49). Although PgaA homologs exist in many bacteria,
283 including *X. nematophila*, the only PgaA homologs represented in any cluster of the original
284 network output were those from *Klebsiella* spp., further strengthening exclusion of cluster 5 from
285 the TXISS network presented here (48). However, despite the topological and substrate
286 (polysaccharide vs. protein) differences, the PF04575 assignment of cluster 5 members hints that
287 there could be evolutionary or structural parallels between the PNAG exopolysaccharide (EPS)
288 synthase-dependent secretion system and TXISS. PgaA is a component of one of several

289 machineries responsible for export of EPS polymers including PNAG, colonic acid, alginate,
290 cellulose, Psl, and Pel that comprise biofilm matrices of Gram-negative bacteria (47). In each
291 case, like the TXISS system described here, the beta-barrel porin either has or associates with
292 another protein that has TPR domains (47). The TPR region of *E. coli* PgaA is necessary for
293 PNAG secretion and biofilm formation (49) and the TPR region of PelB modulates the activity
294 of the polymer-modifying enzyme PelA, with impacts the chemistry of the resulting polymer
295 (50). TXISS TPR domains may similarly modulate the activities of other proteins that influence
296 secretion of the TXISS protein substrates. Indeed, *X. nematophila* expressing versions of the
297 NilB TPR domain with small deletions display colonization defects that are ameliorated by
298 deletion of the entire N-terminal periplasmic domain (11). It will be interesting for future studies
299 to examine the commonalities between the TXISS and EPS synthase-dependent secretion
300 systems and if there is any coordination in secretion of their respective substrates and/or host-
301 associated phenotypes including aggregation and biofilm formation.
302

303 **Materials and Methods**

304 **DUF560 sequence similarity network analysis**

305 Enzyme Function Initiative's Enzyme Similarity Tool (EFI-EST) was used to collect all
306 predicted DUF560-domain-containing protein sequences from Interpro 73 and Uniprot 2019-02
307 (accessed 4.24.19) and BLAST for similarity (16, 17, 20). Representative networks collapsed
308 nodes which shared ≥ 40% identity. On an EFI-EST network, edges are drawn according to a
309 database-independent value called alignment score. A greater alignment score requirement
310 means draws fewer edges. For separation of DUF560-domain-containing proteins an alignment
311 score of 38 was chosen (Fig. 1). For subcluster 1A 89 was chosen and for subcluster 1B 100 was

312 chosen (Supplemental Fig. 1). The EFI-EST Color SSN tool was used to assign cluster numbers.
313 Networks were visualized and interpreted using Cytoscape v3.7.1 (51) and Gephi v0.9.2 (52).
314 Nodes were arranged with the Fruchterman-Reingold force-directed layout algorithm (53).
315 The contents of each cluster were compared to Pfam DUF560 (PF04575) to ensure that
316 clusters were legitimate DUF560 proteins (21). Any clusters for which fewer than 18% of
317 sequences were present in Pfam, or which included fewer than 20 sequences were excluded from
318 downstream analysis. This filtering removed cluster 5, which was composed mostly of *Klebsiella*
319 *pneumoniae* PgaA. This generated a sequence similarity network that contains 10 clusters, 1222
320 nodes, and 52190 edges with 1589 TaxIDs represented (Fig. 1). Using NCBI Taxonomy
321 Browser, each node was examined and manually curated for the isolates' environmental origin(s)
322 among the following categories: water, soil, plant, mammal, animal, invertebrate, nematode,
323 built (environments such as sewage, bioreactors, etc.), multiple environments, and unclassified
324 (Supplemental Table 1). If no citation was available, the isolation source was searched for in the
325 biosamples or bioprojects records. If neither resource was available, strain source was searched
326 for through other resources (NCBI linkout, Google search). A node was assigned an environment
327 if the majority of strains within the node fell into that environment. If a node had no majority
328 environment, the category "multiple environments" was assigned. Any node with animal
329 associated microbes that did qualify as mammal, insect, or nematode associated was designated
330 as generic "animal associated". For fine scale interpretation, analysis focused on cluster 1 and its
331 subclusters (Fig. 2; Supplemental Fig. 1)
332 Three different techniques were used to determine if DUF560 proteins present within our
333 network were genomically associated with TbpB_B_D-domain-containing proteins, hereafter
334 TbpBBD-domain-containing proteins. First, using EFI-GNN, genome-neighborhood-networks

335 were generated for subcluster 1A (Alignment Score 89; 20 ORFs around), subcluster 1B
336 (Alignment Score 38; 10 ORFs around), and subcluster 1C (Alignment Score 38; 10 ORFs
337 around) resulting in 352 DUF560-TbpBBD pairs (17). Next, the RODEO web tool was used to
338 analyze co-occurrence using profile Hidden Markov Models to assign domains to local ORFs
339 resulting in 712 DUF560-TbpBBD pairs (22). 7 additional DUF560-TbpBBD pairs in
340 *Xenorhabdus* were manually annotated. All three datasets were combined for a total of 851 non-
341 redundant protein pairs (Supplemental Table 2). SignalP-5 (54, 55) was used to predict the signal
342 peptides of all TbpBBD-domain-containing proteins. These predictions were used to annotate
343 each node (Fig. 2C). Any node that was associated with both signal peptide bearing proteins and
344 those with no predicted signal peptide were annotated according to the signal peptide bearing
345 proteins.
346
347 **DUF560 genome neighborhood analysis**
348 Subclusters 1A-C were separated and analyzed in EFI-EST with an alignment score of 38
349 as described above. Each network was then analyzed through EFI-GNN and visualized in
350 Cytoscape v3.7.1 (51) (Alignment Score 38; 10 ORFs up and downstream).
351
352 **Phyre^2 analysis**
353 TbpBBD$_{sol}$ protein sequences from *Xenorhabdus nematophila* (HrpC), *Providencia*
354 *rettgeri* (PROVETT_08181/PROVETT_05852), and *Proteus mirabilis* (WP_134940027.1) were
355 collected and the first 22 amino acids were trimmed to remove the signal sequences. These
356 sequences were entered into the Phyre^2 Protein Homology/analogy Recognition Engine v.2.0 to
357 predict potential 3-dimensional structures (25). The top predicted structural model output for all

358 three proteins, hemophilin (6OM5), was used to generate structural models (Supplemental Fig.
359 2). Subsequent PDB files were visualized with Protean 3D v15. (Protean 3D®. Version 15.0.
360 DNASTAR. Madison, WI).
361
362 **Co-expression of HrpB and HrpC**
363 Primers and strains in Supplemental Table 5. Gene synthesis was performed by
364 Genscript® to synthesize the gene encoding HrpB25_26insDYKDDDDK (henceforth HrpB-
365 FLAG). Genscript® subsequently cloned HrpB-FLAG into multiple cloning site 2 (MCS2) of
366 pETDuet-1. The genomic region containing HrpC was amplified from the *X. nematophila* ATCC
367 19061 (HGB800) genome using primers #11-12, digested with restriction enzymes SacI and SalI,
368 and ligated into the multicloning site of pUC19. Site directed mutagenesis was used to add a C-
369 terminal FLAG-tag onto HrpC using primers #13-14. The gene encoding
370 HrpC1_2insV246_247insDYKDDDDK (henceforth HrpC-FLAG) was amplified from pUC19
371 using the primers #15-16. The gene product was ligated into multiple cloning site 1 (MCS1) of
372 both pETDuet-1 and pETDuet-1/HrpB-FLAG, resulting in the expression plasmids pETDuet-
373 1/HrpC-FLAG and pETDuet-1/HrpC-FLAG/HrpB-FLAG. Correct insertion of all genes was
374 confirmed using primers #17-20 and Sanger sequencing at the University of Tennessee
375 Genomics Core.
376 Expression plasmids were transformed into *E. coli* B21DE3 via electroporation
377 (HGB2402 and HGB2459). All strains were grown solely in defined medium with 150 ug/ml
378 ampicillin as previously described (56). Bacteria producing each expression plasmid were
379 subcultured into 100 ml of broth at an initial OD600 of 0.028. Cultures were grown for 18 hours
380 at 37°C to reach late logarithmic growth and then induced with 500 μM IPTG. One hour after

381 induction a 700 µl sub-sample of each culture was taken and filter sterilized for subsequent
382 protein precipitation. At 2.5 hours the whole cells were separated from the remaining supernatant
383 and lysed open via a bead beater. Remaining supernatants were filter sterilized. Whole cell
384 lysates were equalized via Bradford assay. For 1 and 2.5 hour supernatant samples 700 µl of
385 supernatant was precipitated via 10% Trichloroacetic acid incubation (57).

386 Samples were run on 1% SDS-PAGE gels. For lysates, wells were loaded with 9.5ug of
387 protein. For supernatants, wells were loaded with the precipitate of 350µl of media. Western
388 blots were probed with a rat derived α-FLAG tag primary antibody and an α-IgG secondary
389 antibody that fluoresces at 680 nm. Intensities for FLAG reactive bands were recorded. A
390 distinct band from the Coomassie stained gel was used as a loading control to normalize
391 intensities of supernatant samples prior to analysis (Fig. 4, Supplemental Fig. 3). For whole cell
392 samples the data were analyzed with an unpaired t-test. For supernatant samples a two-way
393 ANOVA was performed alongside Sidak's multiple comparisons test.
394

395 **Phylogenetic tree generation**

396 Phylogenetic analysis was performed as described previously (58). Briefly, select
397 *Xenorhabdus* and *Photorhabdus* species were analyzed using MicroScope MaGe's Gene
398 Phyloprofile tool (34, 39) to identify homologous protein sets which were conserved across all
399 assayed genomes. Putative paralogs were excluded from downstream analysis to ensure homolog
400 relatedness, resulting in 665 homologous sets (Supplemental Table 5). Homolog sets were
401 retrieved via locus tag indexing using BioPython (59), individually aligned using Muscle
402 v3.8.31(60), concatenated using Sequence Matrix v1.8 (61), and trimmed of nucleotide gaps

403 using TrimAL v1.3 (62). JmodelTest v2.1.10 (63) was used to choose the GTR+γ substitution
404 model for maximum likelihood and Bayesian analysis.

405 For nematode phylogenetic analysis select *Steinernema* and *Heterorhabditis* species were
406 analyzed. The Internal Transcribed Spacer, 18S rRNA, 28S rRNA, Cytochrome Oxidase I, and
407 12S rRNA loci were collected from Genbank's as available and used as homologous sets
408 (Supplemental Table 4). Nematode species which had fewer than 3 of the 5 loci sequenced were
409 excluded from downstream analysis. Homologous sets were individually aligned, concatenated,
410 and trimmed using the same methods as the *Xenorhabdus* sequences. JmodelTest v2.1.10 (63)
411 was used to choose the GTR+γ+I substitution model.

412 Maximum likelihood analyses were performed via RAxML v8.2.10 (64) using rapid
413 bootstrapping and 1000 replicates, and were visualized via Dendroscope v3.6.2 (65). Nodes with
414 less than 60% bootstrap support were collapsed. Bayesian analyses were performed via MrBayes
415 v3.2.6 with BEAGLE (66-68) on the Cipres Science Gateway platform (69). 500,000 MCMC
416 replicates were performed for the bacterial tree, 4,000,000 were performed for the nematode tree.
417 25% of replicates were discarded as burn-in, and posterior probabilities were sampled every 500
418 replicates. Two runs were performed with 3 heated and one cold chain. The final standard
419 deviation of split frequencies was 0.000000 for the bacterial tree, and 0.002557 for the nematode
420 tree. Bayesian trees were visualized with FigTree v1.4.4 (70). Bayesian and maximum likelihood
421 methods generated phylogenies with consistent topologies (Fig. 5; Supplemental Fig. 5).
422

423 **WT vs. *Δnil* colonization of *S. anatoliense*, *S. carpocapsae*, and *S. websteri***

424 Strains are described in Supplemental Table 5. Bacteria free eggs of *S. anatoliense*, *S.*
425 *carpocapsae*, and *S. websteri* were generated (41) and exposed to a Nil locus mutant

426 (HGB1495), and a Nil locus complemented strain (HGB1496) grown on lipid agar plates for 2
427 days at 25°C (11). Lipid agar plates were placed into White traps 1 week after plating eggs to
428 collect infective juvenile (IJ) nematodes. Nematode colonization was visualized using
429 fluorescence microscopy on the Keyence BZX-700 to observe GFP expressing bacteria in the
430 receptacle. This was done for each species in biological triplicate and technical duplicate. To
431 determine the number of CFU per IJ, nematodes were surface sterilized, and ground for 2 min
432 using a Fisher brand motorized tissue grinder (CAT# 12-1413-61) to homogenize the nematodes
433 and release colonizing bacteria. Serial dilutions in PBS were performed and plated on LB agar,
434 which were incubated at 30°C for 1 day before enumerating CFUs (Fig. 6). To calculate the CFU
435 per colonized IJ, the percent colonized nematodes was divided by the CFU/IJ for each biological
436 replicate. The data were analyzed using a one-way ANOVA with Tukey's multiple comparison's
437 test to compare the mean of each treatment.
438

439 **Acknowledgements:**

440 This work was supported by a grant from the National Science Foundation (IOS-
441 1353674) to HGB and KTF and the University of Tennessee-Knoxville to HGB. TJM was
442 supported by an NIH National Research Service Award T32-GM07215.
443

444 **Author Contributions:**

445 TJM, ASG, and HGB wrote the text of this article and composed the figures.
446 Bioinformatic analysis was performed by ASG and TJM. Phylogenetic analysis was performed
447 by ASG. Cloning and experiments were performed by ASG and TJM. KTF provided sustained
448 intellectual contributions to the research design and edits to the text and figures.

449

450 **Additional Information:**

451 The authors have no competing interests that might be perceived to influence the results
452 and/or discussion reported in this paper. A previous version of this article was published on
453 BioRχiv (https://doi.org/10.1101/2020.01.20.912956). Correspondence and requests for materials
454 should be addressed to H. Goodrich-Blair at hgblair@utk.edu

455

456 **References:**

457

458 1. Hooda Y, Lai CC, Judd A, Buckwalter CM, Shin HE, Gray-Owen SD, Moraes TF. 2016.
459 Slam is an outer membrane protein that is required for the surface display of lipidated virulence
460 factors in Neisseria. Nat Microbiol 1:16009.

461 2. Hooda Y, Lai CCL, Moraes TF. 2017. Identification of a Large Family of Slam-
462 Dependent Surface Lipoproteins in Gram-Negative Bacteria. Front Cell Infect Microbiol 7:207.

463 3. Chu BC, Garcia-Herrero A, Johanson TH, Krewulak KD, Lau CK, Peacock RS,
464 Slavinskaya Z, Vogel HJ. 2010. Siderophore uptake in bacteria and the battle for iron with the
465 host; a bird's eye view. Biometals 23:601-11.

466 4. Noinaj N, Guillier M, Barnard TJ, Buchanan SK. 2010. TonB-dependent transporters:
467 regulation, structure, and function. Annual Review of Microbiology 64:43-60.

468 5. Schweppe DK, Harding C, Chavez JD, Wu X, Ramage E, Singh PK, Manoil C, Bruce JE.
469 2015. Host-Microbe Protein Interactions during Bacterial Infection. Chem Biol 22:1521-1530.

470 6. Chavez-Dozal AA, Gorman C, Lostroh CP, Nishiguchi MK. 2014. Gene-swapping
471 mediates host specificity among symbiotic bacteria in a beneficial symbiosis. PLoS One
472 9:e101691.

473 7. da Silva RAG, Karlyshev AV, Oldfield NJ, Wooldridge KG, Bayliss CD, Ryan A,
474 Griffin R. 2019. Variant Signal Peptides of Vaccine Antigen, FHbp, Impair Processing Affecting
475 Surface Localization and Antibody-Mediated Killing in Most Meningococcal Isolates. Front
476 Microbiol 10:2847.

477 8. Akhurst RJ. 1983. Neoplectana species: specificity of association with bacteria of the
478 genus Xenorhabdus. Exp Parasitol 55:258-263.

479 9. Chaston JM, Suen G, Tucker SL, Andersen AW, Bhasin A, Bode E, Bode HB,
480 Brachmann AO, Cowles CE, Cowles KN, Darby C, de Leon L, Drace K, Du Z, Givaudan A,

481 Herbert Tran EE, Jewell KA, Knack JJ, Krasomil-Osterfeld KC, Kukor R, Lanois A, Latreille P,
482 Leimgruber NK, Lipke CM, Liu R, Lu X, Martens EC, Marri PR, Medigue C, Menard ML,
483 Miller NM, Morales-Soto N, Norton S, Ogier JC, Orchard SS, Park D, Park Y, Qurollo BA,
484 Sugar DR, Richards GR, Rouy Z, Slominski B, Slominski K, Snyder H, Tjaden BC, van der
485 Hoeven R, Welch RD, Wheeler C, Xiang B, Barbazuk B, et al. 2011. The entomopathogenic
486 bacterial endosymbionts Xenorhabdus and Photorhabdus: convergent lifestyles from divergent
487 genomes. PLoS One 6:e27909.

488 10. Heungens K, Cowles CE, Goodrich-Blair H. 2002. Identification of Xenorhabdus
489 nematophila genes required for mutualistic colonization of Steinernema carpocapsae nematodes.
490 Mol Microbiol 45:1337-53.

491 11. Bhasin A, Chaston JM, Goodrich-Blair H. 2012. Mutational analyses reveal overall
492 topology and functional regions of NilB, a bacterial outer membrane protein required for host
493 association in a model of animal-microbe mutualism. J Bacteriol 194:1763-76.

494 12. Blatch GL, Lassle M. 1999. The tetratricopeptide repeat: a structural motif mediating
495 protein-protein interactions. Bioessays 21:932-9.

496 13. Cowles CE, Goodrich-Blair H. 2004. Characterization of a lipoprotein, NilC, required by
497 Xenorhabdus nematophila for mutualism with its nematode host. Mol Microbiol 54:464-77.

498 14. Cowles CE, Goodrich-Blair H. 2008. The *Xenorhabdus nematophila nilABC* genes
499 confer the ability of *Xenorhabdus* spp. to colonize *Steinernema carpocapsae* nematodes. J
500 Bacteriol 190:4121-8.

501 15. d'Enfert C, Chapon C, Pugsley AP. 1987. Export and secretion of the lipoprotein
502 pullulanase by Klebsiella pneumoniae. Mol Microbiol 1:107-16.

503 16. Gerlt JA, Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., Whalen,
504 K.L. 2015. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for
505 generating protein sequence similarity networks. Biochemica et Biophysica Acta 1854:1019-
506 1037.

507 17. Zallot R, Oberg, N.O., Gerlt, J.A. 2018.
508 'Democratized'genomicenzymologywebtoolsforfunctionalassignment. Current Opinion in
509 Chemical Biology 47:77-85.

510 18. Zallot R, Oberg N, Gerlt JA. 2019. The EFI Web Resource for Genomic Enzymology
511 Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes
512 and Metabolic Pathways. Biochemistry 58:4169-4182.

513 19. Ostberg KL, DeRocco, A.J., Mistry, S.D., Dickinson, M.K., Cornelissen, C.N. 2013.
514 Conserved Regions of Gonococcal TbpB Are Critical for Surface Exposure and Transferrin Iron
515 Utilization. Infection and Immunity 81:3442-3450.

516 20. Gerlt JA. 2017. Genomic Enzymology: Web Tools for Leveraging Protein Family
517 Sequence−Function Space and Genome Context to Discover Novel Functions. Biochemistry
518 56:4293-4308.

519 21. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M,
520 Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D,
521 Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids
522 Research 47:D427–D432.

523 22. Choe K, Mitchell Laboratory. 2018. RODEO (Rapid ORF Description & Evaluation
524 Online) is an algorithm to help biosynthetic gene cluster (BGC) analysis, with an emphasis on
525 ribosomal natural product (RiPP) discovery., p http://ripp.rodeo/advanced.html, University of
526 Illinois at Urbana-Champaign.

527 23. Latham RD, Torrado M, Atto B, Walshe JL, Wilson R, Guss JM, Mackay JP, Tristram S,
528 Gell DA. 2020. A heme-binding protein produced by Haemophilus haemolyticus inhibits non-
529 typeable Haemophilus influenzae. Mol Microbiol 113:381-398.

530 24. Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study
531 using the Phyre server. Nat Protoc 4:363-71.

532 25. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal
533 for protein modeling, prediction and analysis. Nat Protoc 10:845-58.

534 26. Einhauer A, Jungbauer A. 2001. The FLAG peptide, a versatile fusion tag for the
535 purification of recombinant proteins. J Biochem Biophys Methods 49:455-65.

536 27. Fantappie L, Irene C, De Santis M, Armini A, Gagliardi A, Tomasi M, Parri M, Cafardi
537 V, Bonomi S, Ganfini L, Zerbini F, Zanella I, Carnemolla C, Bini L, Grandi A, Grandi G. 2017.
538 Some Gram-negative lipoproteins keep their surface topology when transplanted from one
539 species to another and deliver foreign polypeptides to the bacterial surface. Mol Cell Proteomics
540 16:1348-1364.

541 28. Konar M, Rossi R, Walter H, Pajon R, Beernink PT. 2015. A mutant library approach to
542 identify improved meningococcal factor H binding protein vaccine antigens. PLoS One
543 10:e0128185.

544 29. Jackson LA, Ducey TF, Day MW, Zaitshik JB, Orvis J, Dyer DW. 2010. Transcriptional
545 and functional analysis of the Neisseria gonorrhoeae Fur regulon. J Bacteriol 192:77-85.

546 30. Quillin SJ, Hockenberry AJ, Jewett MC, Seiferta HS. 2018. Neisseria gonorrhoeae
547 exposed to sublethal levels of hydrogen peroxide mounts a complex transcriptional tesponse.
548 mSystems 3:e00156-18.

549 31. Stohl EA, Criss AK, Seifert HS. 2005. The transcriptome response of Neisseria
550 gonorrhoeae to hydrogen peroxide reveals genes with previously uncharacterized roles in
551 oxidative damage protection. Mol Microbiol 58:520-532.

552 32. Dent AT, Mouriño S, Huang W, Wilks A. 2018. Post-transcriptional regulation of the
553 Pseudomonas aeruginosa heme assimilation system (Has) fine-tunes extracellular heme sensing.
554 J Biol Chem 294:2771-2785.

555 33. Yukl ET, Jepkorir G, Alontaga AY, Pautsch L, Rodriguez JC, Rivera M, Moënne-Loccoz
556 P. 2010. Kinetic and spectroscopic studies of hemin acquisition in the hemophore HasAp from
557 Pseudomonas aeruginosa. Biochemistry 49:6646-6654.

558 34. Chaston JM, Murfin KE, Heath-Heckman EA, Goodrich-Blair H. 2013. Previously
559 unrecognized stages of species-specific colonization in the mutualism between *Xenorhabdus*
560 bacteria and *Steinernema* nematodes. Cellular Microbiology 15:1545-1559.

561 35. Kampfer P, Tobias NJ, Ke LP, Bode HB, Glaeser SP. 2017. Xenorhabdus
562 thuongxuanensis sp. nov. and Xenorhabdus eapokensis sp. nov., isolated from Steinernema
563 species. International Journal of Systematic and Evolutionary Microbiology 67:1107-1114.

564 36. Lee MM, Stock SP. 2010. A multigene approach for assessing evolutionary relationships
565 of *Xenorhabdus* spp. (gamma-Proteobacteria), the bacterial symbionts of entomopathogenic
566 *Steinernema* nematodes. J Invertebr Pathol 104:67-74.

567 37. Snyder HA, Stock SP, Kim SK, Flores-Lara Y, Forst S. 2007. New insights into the
568 colonization and release process of *Xenorhabdus nematophila* and the morphology and
569 ultrastructure of the bacterial receptacle of its nematode host, *Steinernema carpocapsae*. Applied
570 and Environmental Microbiology 73:5338-5346.

571 38. Spiridonov SE, Reid AP, Podrucka K, Subbotin SA, Moens M. 2004. Phylogenetic
572 relationships within the genus *Steinernema* (Nematoda: Rhabditida) as inferred from analyses of
573 sequences of the ITS1-5.8S-ITS2 region of rDNA and morphological features. Nematology
574 6:547-566.

575 39. Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A, Josso A, Mercier J, Renaux A,
576 Rollin J, Rouy Z, Roche D, Scarpelli C, Medigue C. 2017. MicroScope in 2017: an expanding
577 and evolving integrated resource for community expertise of microbial genomes. Nucleic Acids
578 Research 45:D517–D528.

579 40. McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of
580 sequence analysis tools. Nucleic Acids Res 32:W20-5.

581 41. Murfin KE, Chaston J, Goodrich-Blair H. 2012. Visualizing bacteria in nematodes using
582 fluorescence microscopy. J Vis Exp 68:e4298.

583 42. Costa TR, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, Waksman
584 G. 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat
585 Rev Microbiol 13:343-59.

586 43. Zielke RA, Simmons RS, Park BR, Nonogaki M, Emerson S, Sikora AE. 2014. The type
587 II secretion pathway in Vibrio cholerae is characterized by growth phase-dependent expression
588 of exoprotein genes and is positively regulated by sigmaE. Infect Immun 82:2788-801.

589 44. Bendtsen JD, Kiemer L, Fausboll A, Brunak S. 2005. Non-classical protein secretion in
590 bacteria. BMC Microbiol 5:58.

591 45. Palmer T, Finney AJ, Saha CK, Atkinson GC, Sargent F. 2020. A holin/peptidoglycan
592 hydrolase-dependent protein secretion system. Mol Microbiol doi:10.1111/mmi.14599.

593 46. Cohen J, Powderly W, Opal S. 2017. Preface to the Fourth Edition, Infectious Diseases,
594 4th edition. Elsevier.

595 47. Low KE, Howell PL. 2018. Gram-negative synthase-dependent exopolysaccharide
596 biosynthetic machines. Curr Opin Struct Biol 53:32-44.

597 48. Whitney JC, Howell PL. 2013. Synthase-dependent exopolysaccharide secretion in
598 Gram-negative bacteria. Trends Microbiol 21:63-72.

599 49. Wang Y, Andole Pannuri A, Ni D, Zhou H, Cao X, Lu X, Romeo T, Huang Y. 2016.
600 Structural Basis for Translocation of a Biofilm-supporting Exopolysaccharide across the
601 Bacterial Outer Membrane. J Biol Chem 291:10046-57.

602 50. Marmont LS, Whitfield GB, Rich JD, Yip P, Giesbrecht LB, Stremick CA, Whitney JC,
603 Parsek MR, Harrison JJ, Howell PL. 2017. PelA and PelB proteins form a modification and
604 secretion complex essential for Pel polysaccharide-dependent biofilm formation in Pseudomonas
605 aeruginosa. J Biol Chem 292:19411-19422.

606 51. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski
607 B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular
608 interaction networks. Genome Res 13:2498-504.

609 52. Bastian M, Heymann S, Jacomy M. 2009. Gephi : An Open Source Software for
610 Exploring and Manipulating Networks. International AAAI Conference on Weblogs and Social
611 Media.

612 53. Fruchterman TMJ, Reingold EM. 1991. Graph Drawing by Force-directed Placement.
613 Software - Practice and Experience 21:1129-1164.

614 54. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S,
615 von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep
616 neural networks. Nat Biotechnol 37:420-423.

617 55. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997. Identification of prokaryotic
618 and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10:1-6.

619 56. Orchard SS, Goodrich-Blair H. 2004. Identification and functional characterization of a
620 *Xenorhabdus nematophila* oligopeptide permease. Appl Environ Microbiol 70:5621-7.

621 57. Koontz L. 2014. TCA Precipitation, p 3-10. *In* Lorsch J (ed), Laboratory Methods in
622 Enzymology: Protein Part C, vol 536. Academic Press Elsevier.

623 58. Murfin KE, Lee MM, Klassen JL, McDonald BR, Larget B, Forst S, Stock SP, Currie
624 CR, Goodrich-Blair H. 2015. *Xenorhabdus bovienii* strain diversity impacts coevolution and
625 symbiotic maintenance with *Steinernema* spp. nematode hosts. MBio 6:e00076.

626 59. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T,
627 Kauff F, Wilczynski B, de Hoon MJ. 2009. Biopython: freely available Python tools for
628 computational molecular biology and bioinformatics. Bioinformatics 25:1422-3.

629 60. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
630 throughput. Nucleic Acids Res 32:1792-7.

631 61. Vaidya G, Lohman, D. J., Meier, R. 2010. SequenceMatrix: Concatenation software for
632 the fast assembly of multi-gene datasets with character set and codon information. Cladistics
633 27:171-180.

634 62. Capella-Gutiérrez S, Silla-Martínez, J. M., Gabaldón, T. 2009. trimAl: a tool for
635 automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-
636 1973.

637 63. Darriba D, Taboada, G. L., Doallo, R., Posada, D. 2012. jModelTest 2: more models, new
638 heuristics and highperformance computing. Nat Methods 9:772.

639 64. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis
640 of large phylogenies. Bioinformatics 30:1312-1313.

641 65. Huson DH, Scornavacca, C. 2012. Dendroscope 3: An Interactive Tool for Rooted
642 Phylogenetic Trees and Networks. Syst Biol 61:1061-1067.

643 66. Altekar G, Dwarkadas, S., Huelsenbeck, J. P., Ronquist, F. 2004. Parallel Metropolis
644 coupled Markov chain Monte Carlo for Bayesian phylogenetic inference Bioinformatics 20:407-
645 415.

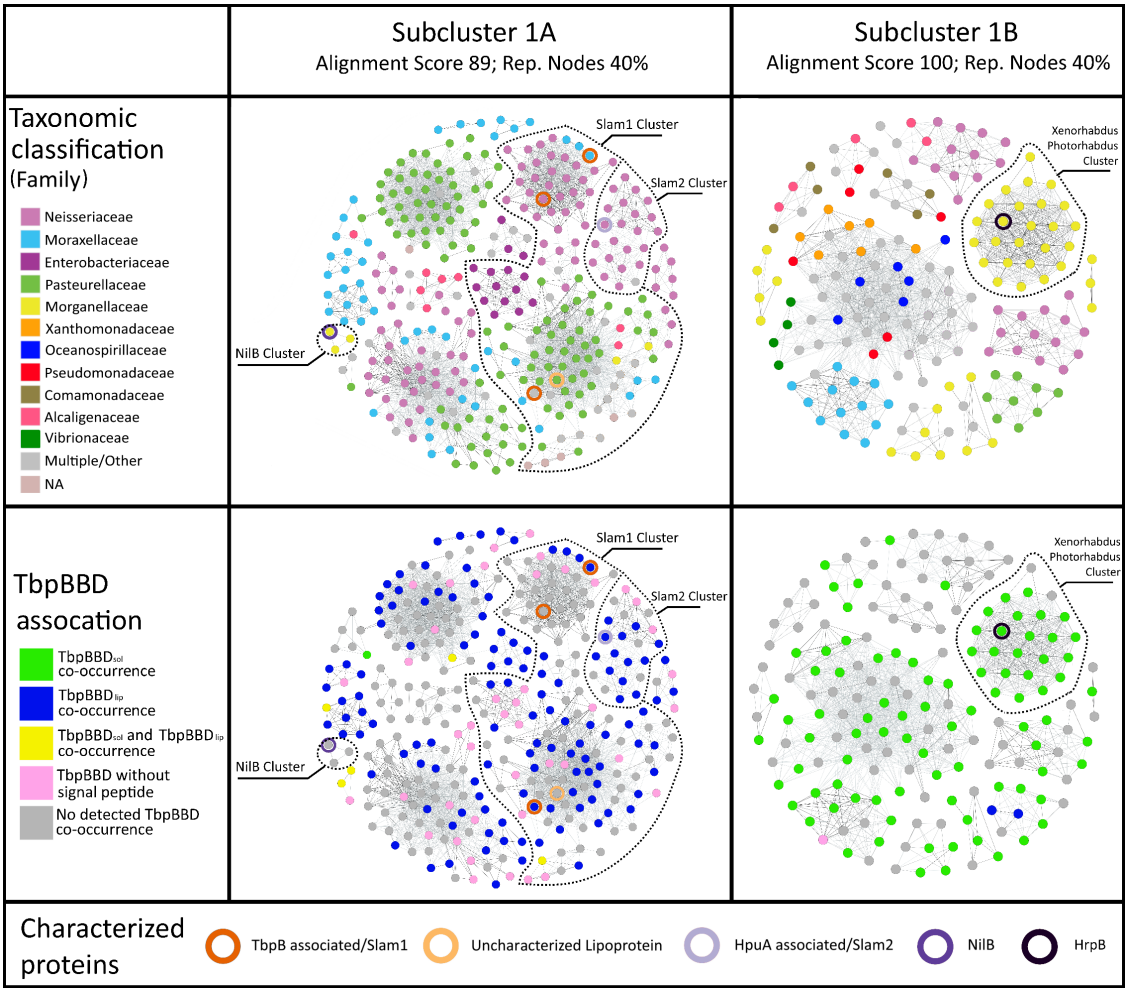
646 67. Ayres DL, Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O.,
647 Huelsenbeck, J. O., Ronquist, F., Swofford, D. L., Cummings, M. P., Rambaut, A., Suchard, M.
648 A. 2011. BEAGLE: An Application Programming Interface and High-Performance Computing
649 Library for Statistical Phylogenetics. Syst Biol 61:170-173.

650 68. Ronquist F, Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., Larget,
651 B., Liu, L., Suchard, M. A., Huelsenbeck, J. P. 2012. MrBayes 3.2: Efficient Bayesian
652 Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol 61:539-542.

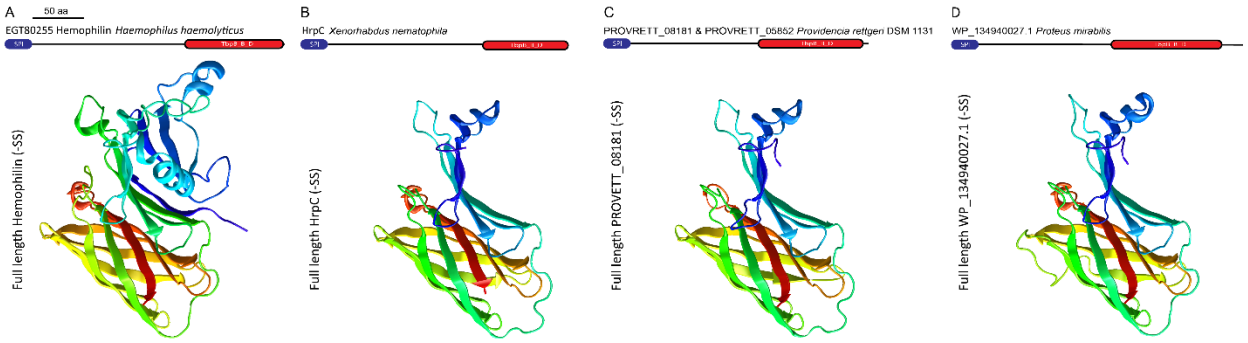
653 69. Miller MA, Pfeiffer, W., Schwartz, T. Creating the CIPRES Science Gateway for
654 inference of large phylogenetic trees, p 1-8. *In* (ed), Proceedings of the Gateway Computing
655 Environments Workshop (GCE),

656 70. Rambaut A. 2018. FigTree v1.4.4 phttp://tree.bio.ed.ac.uk/software/figtree/.
657

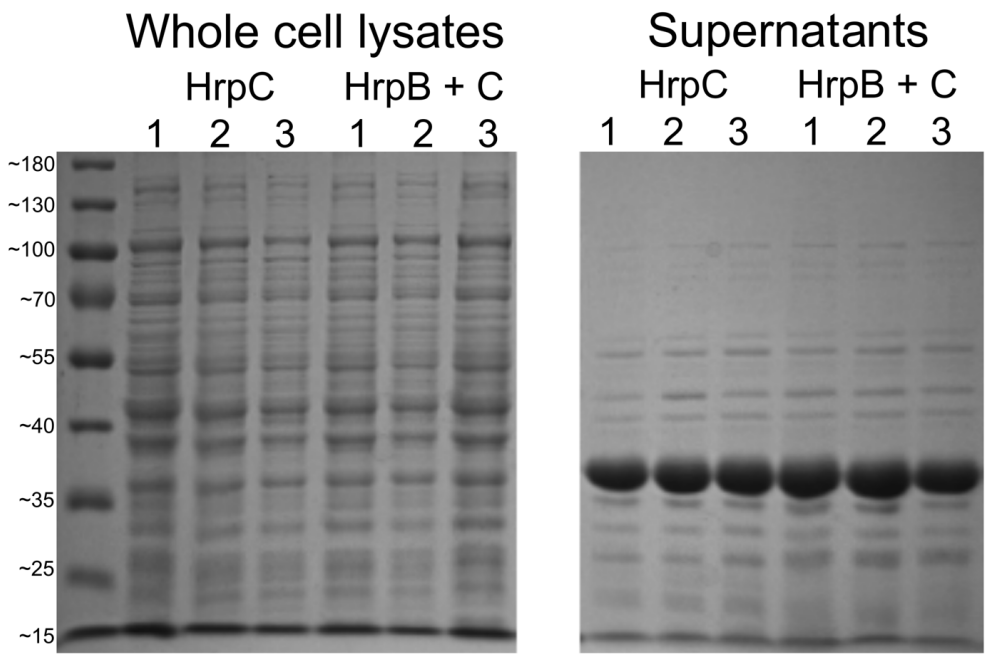
Supplemental Data



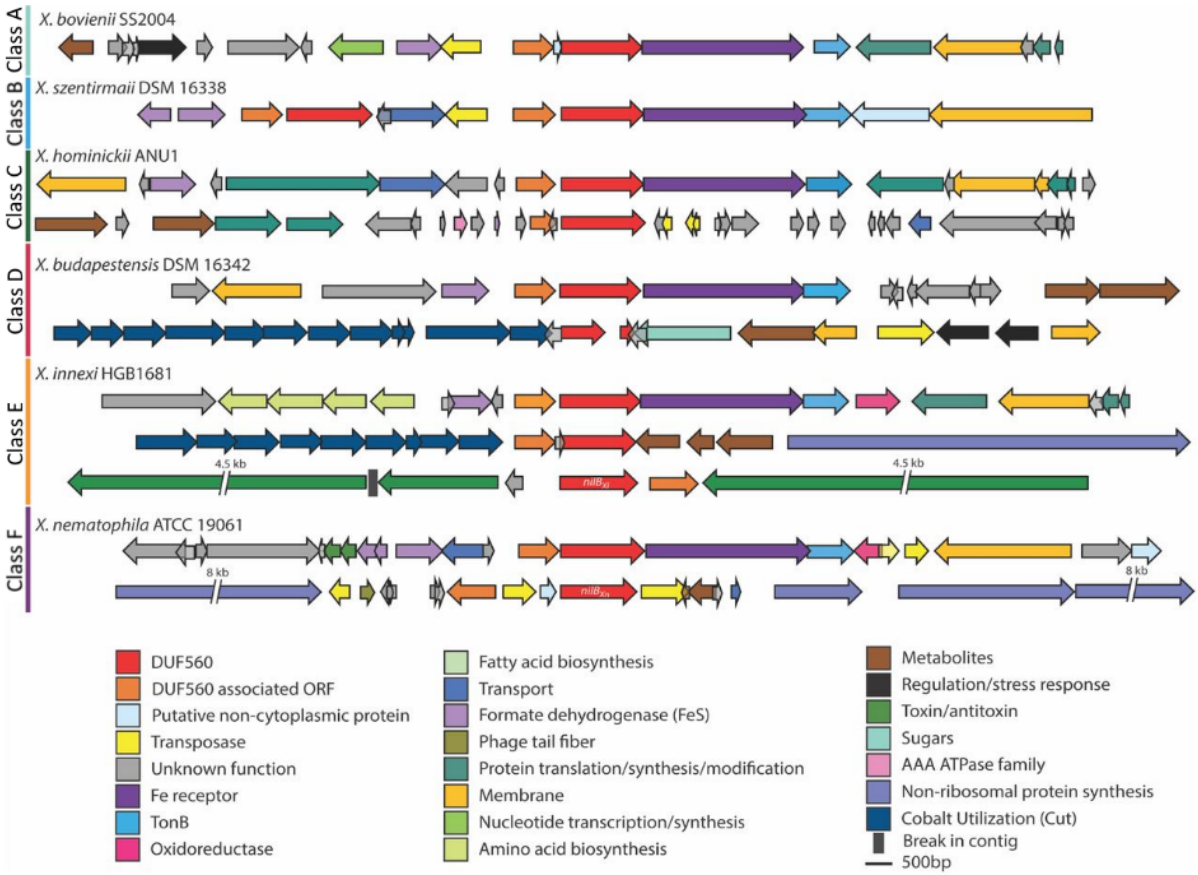
Supplemental Figure 1. Increased stringency separates subcluster 1A and 1B into functional groups. A series of stringent EFI-EST sequence similarity networks highlights detail in cluster 1 of the DUF560 homologs. Edge darkness demonstrates similarity. Node positioning was optimized using the Fruchterman-Reingold algorithm (53). Dotted lines indicate hypothetical functional clusters based on previous molecular data. Circled nodes indicate proteins which have been molecularly characterized. Subclusters 1A and 1B were analyzed separately to allow fine tuning of alignment score (89 and 100 respectively). Networks were color coded to display either taxonomic categories or co-inheritance with TbpBBD-domain containing proteins.



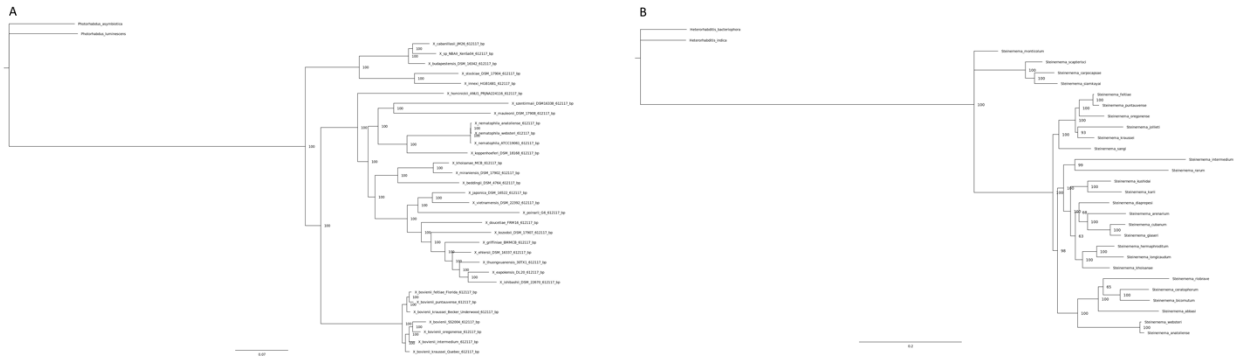
Supplemental Figure 2. Phyre^2 models of select TbpBBD$_{sol}$ proteins. TbpBBD$_{sol}$ proteins, lacking the signal sequence (-SS), from *X. nematophila* (HrpC), *P. rettgeri* (PROVRETT_08181 and 05852), and *P. mirabilis* (WP_134940027.1) were queried through the Phyre^2 Protein Homology/analogy Recognition Engine v. 2.0 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (25). The top predicted structural model output for each is shown alongside the solved crystal structure of hemophilin from *H. haemolyticus* (protein data bank file 6OM5) which the algorithm selected as the template for all queries (23).



Supplemental Figure 3. Supernatant samples have distinct protein profile. Coomassie stained SDS-PAGE gels showing side by side comparison of protein banding in whole cell lysates and supernatant samples of *E. coli* expressing the indicated proteins. Units on the left side are in kilodaltons. Expression of HrpB does not cause cell lysis or nonspecific protein export, if it did the banding pattern of the supernatant would look more like that of the lysates. The increase in supernatant HrpC is faintly visible at ~27 kilodaltons in the last three supernatant sample lanes. All experiments were repeated in biological triplicate as labeled above the lanes.



Supplemental Figure 4. Representative genomes of *Xenorhabdus* DUF560 Classes A-F. Schematic diagrams of *Xenorhabdus* TXISS loci representing each of the six classes defined in the text. One species from each of the classes was selected for presentation. Box arrows represent open reading frames (ORFs), which are color coded according to predicted annotated function as indicated by the legend. The DUF560 homolog is shown in red and the predicted TXISS cargo is shown in orange. Large ORFs were not presented in their entirety and the length of the gap is indicated above the break line shown within such ORFs.



Supplemental Figure 5. Bayesian posterior probability phylogenies. A) Phylogram of select *Xenorhabdus* bacteria, based on concatenations of 665 conserved core genes. Numbers indicate posterior probability values. Distances indicate substitutions per base pair. B) Bayesian phylogeny of select entomopathogenic nematodes, based on concatenations of the ITS, 18S rRNA, 28S rRNA, COI, and 12S rRNA loci. Two members of the sister taxon *Photorhabdus* were chosen as an outgroup. Loci are recorded in Supplemental table 4.

705 **Supplemental Table 1. DUF560 sequence similarity network clusters 1-4 and 6-11 node**

706 **table, with subcluster, environmental annotations, and sources (when possible)**

707 *Associated file: TableS1_Clusters1-4,6-11WithEnvTax.xlsx*

708 **Supplemental Table 2. DUF560 TbpBBD protein Co-occurrence**

709 *Associated file: TableS2_DUF560TbpBBDCo-occurrence.xlsx*

710 **Supplemental Table 3. *Xenorhabdus* DUF560 classes**

711 *Associated file: TableS3_XenorhabdusDUF560Classes.xlsx*

712 **Supplemental Table 4. Loci tags used in phylogenetic analysis of *Xenorhabdus* and**

713 ***Steinernema* species**

714 *Associated file: TableS4_CombinedLociTable*

715 **Supplemental Table 5. Primer and strain list**

716 *Associated file: TableS5_PrimersAndStrains.xlsx*

717

718

719