Identification of the Type Eleven Secretion System (T11SS) and Characterization of T11SS-dependent Effector Proteins

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Alex S. Grossman

December 2022

Copyright © 2022 by Alex Grossman
All rights reserved.

Acknowledgements

Sir Isaac Newton's famous quote "If I have seen further, it is by standing on the shoulders of Giants" applies not just to the incremental development of science, but also to the incremental development of the people who make a life in science. The efforts of so many people have contributed to my life that it would be impossible to include everyone who deserves to be included in this list. That being said, I'm going to try my best. First in impact and chronology, I would like to thank my parents Susan Grossman and Allen Grossman for their ceaseless efforts to nurture and provide for me, even after they no longer had any idea what I did for a living. I would like to thank my early friends, Nicholas Pollicelli and Sheyn Denisov who taught me that there is nothing wrong with embracing nerdiness in all its glory. I would like to thank several schoolteachers who were instrumental in instilling a love of learning from an early age; Ilka Hanzelman (expert code-cracker and fashion icon), Timothy Snelling (the quiz master), and Rob Marsteller (who let us do all sorts of dissections). During my undergraduate education at Case Western Reserve University, I met many friends, from many backgrounds, with diverse expertise who were instrumental in turning an adrift nerd from Appalachia into a reasonable facsimile of a well-rounded scientist of the world. In alphabetical order, I would like to thank Dr. Alexander Strang, Andrew Schmidt, Brandon Carreño, Dr. Daniel Brandt, Jerry Huang, Jesse Ofsowits, Joe Fennimore, Joseph Lerchbacker, Kayla Kindig, Kevin Ye, Luciana Copeland, Stephan Brennen, and Dr. Yumi Matsuyama. *Vivat cultus*. I would like to thank three professors from undergrad specifically; Dr. Michael Benard who was the first person to give me the opportunity to participate directly in research and contributed much to my development, Dr. Susan Burden-Gulley who served as my academic advisor, gave me the opportunity to TA a fantastic course, and did a lot to dispel my fear of academia, and finally Dr. Amy Absher who gets a special shout out for being one of the most kind, thoughtful, and blindingly brilliant humans I have ever had the opportunity to study under. My first exposure to the University of Tennessee was the Microbiology REU program in 2015 where I worked with Dr. Jonelle Basso who was, and is, an inspiration for her kindness, dedication to education, and scientific curiosity. I would like to thank all the members of my dissertation committee and my PI, Dr. Heidi Goodrich-Blair. Heidi has always been available and excited to discuss science with me. Her dedication and ceaseless effort have given me the opportunity to learn so many new things about the natural world, the culture of academia, and the development of my career. I would like to thank the many friends and mentors I have worked with in the Goodrich-Blair lab, particularly Sarah Kauffman and Dr. Terra Mauer who were instrumental in my education. Finally, I would like to thank my partner Rose Pickrell who helped carry the weight of the world when it seemed too heavy for human shoulders and who brought so much happiness into my life.

"Tiger got to hunt, bird got to fly;
Man got to sit and wonder 'why, why, why?'
Tiger got to sleep, bird got to land;
Man got to tell himself he understand."

~Kurt Vonnegut, Cat's Cradle~

Abstract

Host-associated microbes live in dangerous environments as a result of host immune killing, nutrient limitation, and physiological conditions. Bacteria have evolved a toolkit of surface and secreted proteins to help interact with this host environment and overcome nutrient limitation. The studies included within this dissertation describe the identification of a novel bacterial secretion system which has evolved to transport these symbiosis-mediating proteins. This system, termed the type eleven secretion system (T11SS), is present throughout the Gram negative phylum Proteobacteria, including many human pathogens such as *Neisseria meningitidis*, *Acinetobacter baumanii*, *Haemophilus haemolyticus*, and *Proteus vulgaris*. Furthermore, these studies describe how novel cargo proteins of this secretion system were identified and characterized using molecular biology and physicochemical techniques. Chapter 1 establishes the importance of nematode model systems in researching symbiosis, highlighting how research in entomopathogenic nematodes identified the first T11SS. Chapters 2 and 3 use a T11SS-dependent hemophore named hemophilin and its transporter protein to demonstrate T11SS secretion and its mechanisms of cargo specificity. Chapter 3 also explores the role of hemophilin within the nematode symbiont *X. nematophila* in surviving heme starvation and facilitating nematode fitness. Chapter 4 demonstrates that the lipidated symbiosis factor NilC is surface exposed by the T11SS NilB and uses a combination of metabolomics, proteomics, and lectin library analysis to describe the role of NilC in colonization. Chapter 5 describes a protocol for bioinformatically controlling genome neighborhood co-occurrence analyses and utilizes this technique to demonstrate significant co-occurrence of T11SS with metal uptake pathways, single carbon metabolism, and mobile genetic elements. Additionally, this protocol allowed prediction of 141 T11SS-dependent cargo falling into 10 distinct architectures, including never before seen T11SS-dependent surface proteins. Finally, Chapter 6 summarizes our findings and contextualizes how the T11SS plays essential roles in host-microbe association in mutualistic bacteria and pathogenic bacteria alike.

Table of Contents

Chapter 1: Leveraging nematode microbiomes as model systems to identify and characterize molecular determinants of symbiosis 1
- **Introduction:** 2
- ***C. elegans* as a Model for Symbiosis:** 9
- **Entomopathogenic Nematodes as a Model for Symbiosis:** 18
- **Goals of this Dissertation:** 33

Chapter 2: A Widespread Bacterial Secretion System with Diverse Substrates 37
- **Introduction:** 39
- **Results:** 42
- **Discussion:** 61
- **Materials and Methods** 67

Chapter 3: Bacterial hemophilin homologs have conserved roles in host-microbe heme capture that are dependent on highly specific type eleven secretor proteins 74
- **Introduction:** 76
- **Results:** 80
- **Discussion:** 118
- **Materials and Methods:** 126

Chapter 4: A Surface Exposed, Two-Domain Lipoprotein Cargo of a Type 11 Secretion System Promotes Colonization of Host Intestinal Epithelia Expressing Glycans 144
- **Introduction:** 147
- **Results:** 149
- **Discussion:** 190
- **Materials and Methods:** 197

Chapter 5: Bioinformatic investigation of effectors and substrates for the novel type 11 secretion system (T11SS) 214
- **Introduction:** 216
- **Results:** 219
- **Discussion:** 241
- **Materials and methods:** 250

Chapter 6: Conclusions and Avenues for Expansion 254

Bibliography: 268

Vita: 296

List of Figures

Figure 1.1……...……...3
Different types of symbioses and examples discussed in this chapter

Figure 1.2……..…..…...8
Illustrated representation of the well-studied types of nematode-bacterium symbiosis

Figure 2.1……..…..…..…..43
Conceptual model of the type 11 secretion system

Figure 2.2 ……..…..…….44
Sequence similarity network of all DUF560 proteins

Figure 2.3……...…..…….46
Cluster 1 of the T11SS sequence similarity network (SSN) forms subclusters according to environment of isolation and signal sequence of predicted cargo

Figure 2.4……48
Increased stringency separates subcluster 1A and 1B into functional groups

Figure 2.5 ……..……...…50
Examples of TbpBBDlip and TbpBBDsol signal peptides and domain architectures

Figure 2.6……51
Phyre2 models of select TbpBBDsol proteins

Figure 2.7……53
HrpB increases secretion of HrpC

Figure 2.8……55
Supernatant samples have a less complex protein profile than whole cells

Figure 2.9……57
Cladograms of *Xenorhabdus* and *Steinernema* color coded according to *Xenorhabdus* DUF560 Class

Figure 2.10……...……58
Representative genomes of *Xenorhabdus* DUF560 Classes A-F

Figure 2.11……..……..…59
Bayesian posterior probability phylogenies

Figure 2.12……...……62
The Nil locus is necessary for colonization of *S. anatoliense*, *S. carpocapsae*, and *S. websteri* infective juveniles

Figure 3.1……79
Conceptual model of hemophilin mediated heme acquisition

Figure 3.2……81
Distribution and relatedness of soluble T11SS-dependent cargo proteins

Figure 3.3……83
Distribution and relatedness of hemophilin family proteins

Figure 3.4……86
Alignment of the N-terminal region of *Xenorhabdus* hemophilin homologs

Figure 3.5……...…….87
Hemophilin homolog N-terminal domain consensus sequence logo based on alignments

Figure 3.6……88
Single locus cladogram of T11SS genes from the genus *Xenorhabdus*

Figure 3.7……92
Secretion of hemophilin proteins by their cognate and non-cognate T11SS proteins

Figure 3.8……...……..93
Secretion of domain swapped chimeric hemophilin proteins by $HrpB_{X.nem}$ and $HrpB_{H.haem}$

Figure 3.9……...……..96
Heme binding properties of hemophilin homologs

Figure 3.10……..…….....……..99
UV-visible spectra of porphyrin in the presence of bovine serum albumin (BSA)

Figure 3.11……..…….……..101
Protoporphyrin IX binding properties of hemophilin homologs

Figure 3.12……..…….……..102
Propensity of Hpl and CrpC to bind to selected linear and cyclic tetrapyrroles

Figure 3.13……..…….……..104
Structural modeling of HrpC and CrpC

Figure 3.14……..…….……..107
Transcriptional expression from P_{hrpB} and P_{hrpC} in *X. nematophila*

Figure 3.15……..…….…..111
Growth curve of *X. nematophila* wildtype, *hrpB*::*kanR*, and a *tonBhrpAB* complement strain with and without metal starvation

Figure 3.16……..…..……..113
HrpB facilitates the growth of conspecific *Xenorhabdus* strains under metal starvation and does not impact the growth of heterospecific *E. faecalis*

Figure 3.17……..…..……..114
General phenotypic characterization of the *X. nematophila hrpB*::*kanR* mutant

Figure 3.18……..…..……...116
HrpB is essential for nematode fitness *in vitro*, but does not detectably impact nematode fitness *in vivo*

Figure 3.19……..…...……..117
X. nematophila hrpB*::*kanR* lawns generate mucoid regions when cultured with *S. carpocapsae

Figure 3.20……..……...……..119
Colonization of *S. carpocapsae* nematodes by *X. nematophila hrpB*::*kanR*

Figure 4.1……..……..150
NilB positively influences the amount of NilC present on the surface of *E. coli* when co-expressed

Figure 4.2……..……..151
Immuno-dot blots reveal impact of NilB on surface exposure of NilC

Figure 4.3……...153
NilC surface exposure is not driven by cell lysis

Figure 4.4……..……..154
Histidine tag does not negatively affect NilC function in colonization whereas loss of lipidation site does

Figure 4.5……..……..156
NilC secondary structure and structural models

Figure 4.6……..……..158
Schematic diagram of the genomic context of *nilB* and *nilC* homologs

Figure 4.7……..……..160
NilC solution NMR

Figure 4.8……..……..161
NilC limited proteolysis

Figure 4.9……..……..164
***Xenorhabdus nematophila* NilC is surface exposed by NilB when its expression is de-repressed**

Figure 4.10……..……166
Immuno-dot blots of *X. nematophila* cells differing in nil or lrp alleles

Figure 4.11……..……168
NilB variants influence *X. nematophila* NilC surface exposure

Figure 4.12……..……170
Glycan content differs between nematode species at bacterial colonization sites

Figure 4.13……..……172
Summary of predicted identity and localization of glycan sugar residues in the gastrointestinal tract of *S. carpocapsae, S. scapterisci,* and *S. feltiae*

Figure 4.14……..……175
Addition of soluble, purified NilC increases colonization of the anterior intestinal cecum

Figure 4.15……..……177
Lipidated NilC is surface exposed and soluble NilC is secreted

Figure 4.16……..……181
Growth curves of strains used for sampling to conduct metabolome and proteome analyses

Figure 4.17……..……183
Differentially translated proteins between Δ*SR1* and WT show significant differences in metal-related activities (binding, homeostasis, and metabolic pathways).

Figure 4.18……..……184
Significant metabolome differences between WT and Δ*SR1* indicate T11SS role of amino acid and amino sugar metabolism

Figure 4.19……..……186
Predicted pathways for amino sugar metabolism leading to peptidoglycan (PG), exopolysaccharide [Poly-N-acetylglucosamine (PNAG)], lipopolysaccharide Lipid A biosynthesis, and glycolysis pathways

Figure 4.20……189
NilC binds peptidoglycan

Figure 4.21……192
The model of potential molecular interactions occurring during *X. nematophila* colonization of *S. carpocapsae* anterior intestinal cecum epithelial cell surfaces.

Figure 5.1……221
T11SS (DUF560) protein size distribution (in amino acids)

Figure 5.2……222
Design for a novel control technique for genomic neighborhood co-occurrence analysis

Figure 5.3……224
Significantly co-occurring genes with animal associated T11SS reveal conserved association with iron/heme uptake, protein export, and single carbon metabolism

Figure 5.4……226
BlastKOALA supports T11SS association with iron/heme uptake, protein export, and single carbon metabolism

Figure 5.5……229
PsiPred annotation of predicted T11SS-dependent cargo reveals diverse N-terminal domains featuring proline rich repeats, intrinsically disordered regions, α helical repeats, and ligand binding handles

Figure 5.6……232
RoseTTAFold annotation of predicted T11SS-dependent cargo reveals conservation of C-terminal or centrally located hydrophilic β-barrel domains

Figure 5.7……234
Localization of FLAG tagged Pls protein in the absence and presence of its cognate T11SS secretor

Figure 5.8……237
Significantly co-occurring genes with marine associated T11SS reveal conserved association with single carbon metabolism, amino acid metabolism, and the glyoxalase detoxification pathway

Figure 5.9……238
BlastKOALA supports T11SS association with single carbon metabolism, amino acid metabolism, and the glyoxalase detoxification pathway

Figure 5.10……242
Representative structures of DUF1194 proteins located within a T11SS genomic neighborhood in *Ascidiaceihabitans donghaensis*

List of Attachments

DUF560 sequence similarity network (TableS1_Clusters1-4,6-11WithEnvTax.xlsx)

DUF560 TbpBBD protein Co-occurrence (TableS2_DUF560TbpBBDCo-occurrence.xlsx)

Xenorhabdus DUF560 classes (TableS3_XenorhabdusDUF560Classes.xlsx)

Loci tags used in phylogenetic analysis of Xenorhabdus and Steinernema species (TableS4_CombinedLociTable.xlsx)

Hemophilin Homologs by subcluster (Chap3SubclusterRodeoResultsAndNetworkNodeTable.xlsx)

Immunoblot images of T11SS activity (Chapter3WesternImages.xlsx)

Strain list and primer list from chapter 3 (Chapter3StrainAndPrimerTable.xlsx)

Lectin Binding specificity data (Chapter4LectinInfo.xlsx)

NilBC mutant metabolomic data (Chapter4MassSpectrometryInfo.xlsx)

Strain list and proteomic data tables from chapter 4 (Chapter4SupplementalTables.pdf)

DUF560 co-occurrence analysis and cargo prediction (Chapter5CooccurenceDatasets.xlsx)

Functional analysis of co-occurring genes (Chapter5FunctionalAnalysis.xlsx)

DUF1194-DUF560 co-occurrence dataset (Chapter5DUF1194-DUF560Coocurrence.xlsx)

Chapter 1: Leveraging nematode microbiomes as model systems to identify and characterize molecular determinants of symbiosis

Authors: Alex S. Grossman, Elizabeth M. Ransone, Jennifer K. Heppert, Terra J. Mauer, and Heidi Goodrich-Blair

Publication disclosure:

Some portions of this chapter were previously published in the book Nematodes As Model Organisms as "Chapter 5 Nematodes As Models for Symbiosis" (Heppert et al. 2022) (https://cabidigitallibrary.org/doi/10.1079/9781789248814.0000). The initial draft of the Introduction was written by Jennifer K. Heppert. The initial drafts of: What makes a good symbiosis model system?:, *C. elegans* as a Model for Symbiosis, Nematode–microbiome models of insect and gastropod pathogenesis, and Dissertation Goals, were written by Alex S. Grossman. The initial drafts of the *Steinernema* and *Heterorhabditis* sections were written by Elizabeth M. Ransone. Figures 1.1 and 1.2A were generated by Alex S. Grossman, Figure 1.2B was generated by Jennifer K. Heppert. Alex S. Grossman, Elizabeth M. Ransone, Jennifer K. Heppert, Terra J. Mauer, and Heidi Goodrich-Blair all contributed to proofing of the original publication. Alex S. Grossman and Heidi Goodrich-Blair collaborated to adapt the manuscript for inclusion in this dissertation.

Introduction:

Microbes are ubiquitous, being present throughout earth's oceans, in the soil, on and in plants and animals, in ice present at both poles, and even in the atmosphere (Giovannoni 2017; Stolz 2017). All multicellular eukaryotes therefore live in association with microbes. Some of these interactions are symbioses, defined as unlike organisms living closely together in long-term relationships, often with intertwined life cycles and evolutionarily derived structures and signals that promote specific associations (Stubbendieck, Li, and Currie 2019). Symbioses can span a range of interaction outcomes, from mutualism, in which all parties receive a net benefit from the interaction, to parasitism, in which at least one party receives a net benefit while another incurs a net cost (Fig. 1.1). Common types of microbial mutualisms are nutritional, in which a partner provides the other

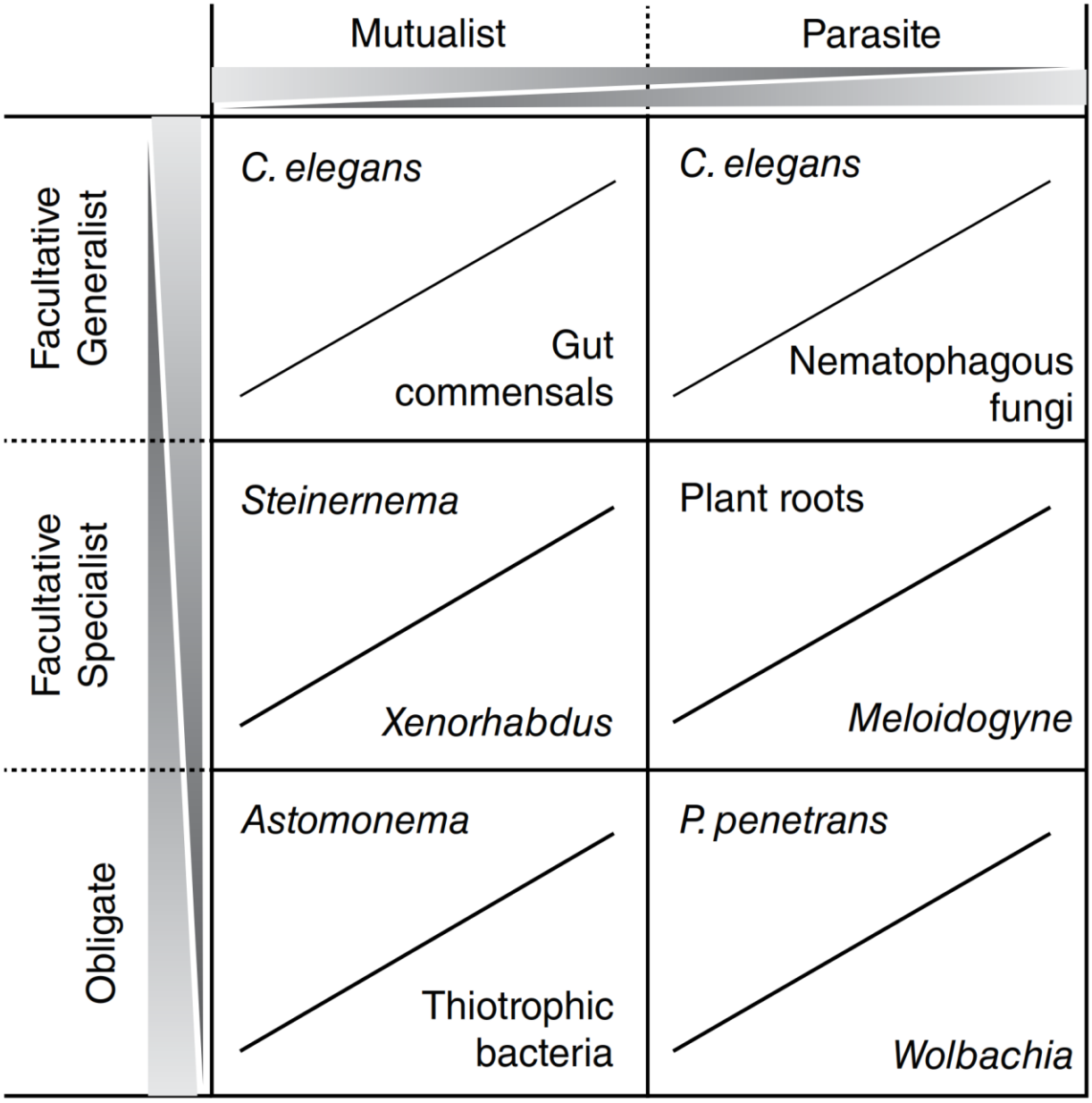


Figure 1.1 **Different types of symbioses and examples discussed in this chapter**. Symbiotic interactions can be classified from mutualist to parasitic and from facultative to obligate. Symbioses can move along the depicted gradients based on changes in environmental or evolutionary pressures. Several of the interactions discussed in this chapter are categorized as examples of the symbiosis types depicted.

with essential metabolites not present in the diet, and defensive, in which a partner provides the other with protection against competitors, predators, and parasites (Douglas 2010). Microbial symbionts can be passed directly from parent to offspring in vertical transmission events, or horizontally acquired from the environment, and symbioses can exhibit mixed modalities that combine these two transmission modalities to different degrees (Bright and Bulgheresi 2010; Ebert 2013; Stoy et al. 2020). Symbioses can be further categorized by the level of interdependence of the participants and the specificity of engagement with potential partners (Fig. 1.1). In obligate associations, one or more of the partners cannot exist without the other(s); while in facultative symbioses, interdependence is context dependent (Chomicki, Kiers, and Renner 2020). In generalist associations, symbionts can associate with a broad range of partners; while in specialist associations the partner number is limited to one or two, though the taxonomic level of the partners (for example, a type of organism, one species, or one subspecies) can vary across symbioses (Chomicki, Kiers, and Renner 2020). Whether antagonistic or mutualistic, obligatory or facultative, or generalist or specialist, bacterial interactions with hosts occur through conserved mechanisms. This conservation means that the study of experimentally tractable organisms can yield insights into processes of host–microbe associations all along the continuum. However, because such diverse types of host–microbe symbioses exist, it is important to study many different examples of such associations to understand the full range of parameters that modulate the specificity and outcomes of these associations.

Nematoda is a diverse phylum whose members participate in many types of symbiosis with different bacterial taxa, making them ideal models for understanding host–microbe interactions. Symbioses can result in the participants gaining complex traits through their partnership that would take millennia to evolve *de novo*. This means that symbioses are a particularly potent biological force, because they allow organisms to break the evolutionary "rule" of descent with modification

(Douglas 2010). This review focuses primarily on bacterial symbionts, which vary greatly in their genomic potential, even among members of a single genus or species. Such variation among mutualistic symbiotic partners can result in variation in symbiotic traits and the ability to contribute to host fitness. This raises the question of what "goods and services" the symbionts provide to each other and at what costs. The corollary to this question is if, and how, hosts evaluate potential partners for their symbiotic potential, recruit and provide their own inducements, and transmit to new generations those symbionts that are most likely to provide optimal fitness to offspring. Potential bacterial symbionts also may compete with each other for colonization of specific host tissues and likely have evolved mechanisms to gain advantage over both distant and closely related competitors. The molecular foundations of symbiont partner recognition and selection are not well understood, particularly in complex systems, in which many players influence the outcomes. Finally, how do microbes that are facultatively associated with a host, or that occupy more than one host over their life cycle, adapt to these changing environments for optimal fitness? How do hosts defend themselves from harmful associations while recruiting beneficial partners? In this chapter, I outline how nematodes are useful as biological models of animal–microbe symbioses, and how they have the potential to expand our understanding of fundamental principles of host–microbe interactions, including community recruitment and formation, transmission between individuals and generations, nutrient sharing and metabolic cooperation, and the molecular and cellular adaptations by both host and microbe that facilitate these interactions.

Diverse Nematode Models of Symbiosis

On the surface, the members of the phylum Nematoda look very similar. They are thread-like roundworms with a relatively simple, cylindrical body and digestive system surrounded by a protective cuticle. Many, but not all, are small and transparent, facilitating the observation of symbionts by light microscopy. However, that is largely where the commonalities among members

of this phylum end. Although difficult to quantify, nematodes are one of the most abundant multicellular organisms on earth, and they have been found in nearly every ecological niche that has been explored (Shah and Mahamood 2017). These animals thrive in environments as different as soil, the human microbiome, insects, the sea floor, and plant roots, and they have the trophic strategies to match. Their diets include other invertebrates, bacteria, blood, and plant fluids and tissue. Based on what we know about this diversity of nematode habitats and diets, it is perhaps unsurprising that there is an equivalent level of diversity in their interactions with microbes.

Even within bacterial species that form symbioses with closely related nematodes, there can be a remarkable amount of subspecies- or strain-level genomic diversity and differences in genetic capacity (Denver et al. 2016; Maher et al. 2021; Murfin, Whooley, et al. 2015; Scott et al. 2012). Further, there is variability in nematode dependence on symbiosis: even considering two closely related nematode species, one of them can require a bacterial symbiont to live and reproduce, while a sister species has no association with a symbiont (Bouchery et al. 2013). It is clear that symbionts can be lost, gained, and changed even within closely related nematodes that occupy similar ecological niches (Bouchery et al. 2013; Murfin, Lee, et al. 2015a; Wasala et al. 2019). There are specialist species (such as entomopathogenic nematodes, filarial nematodes, and marine nematodes with ectosymbionts) that have a dominant interaction with a single major symbiont, and generalists (such as *Caenorhabditis elegans* and other free-living nematodes) harboring a community of associated symbionts, many of which are dispensable or interchangeable with other taxa, more similar to the human gut microbiome. The location and nature of the host association can vary dramatically, from endosymbionts that are within the body, both within (intracellular) and between (extracellular) nematode cells, to ectosymbionts that are associated on the surface of the nematode.

Overall, this rich diversity in the types and features of symbioses makes nematode–bacterium associations useful for modeling a similarly diverse set of scientific questions. Nematode–

bacterium symbioses provide important opportunities for comparative study of community assembly, the impact of environmental factors such as stress and diet on the association, how reproductive mode relates to symbiont transmission, and host–microbe coevolution (Morran et al. 2016). Further, studying a broad range of nematode symbioses will reveal universal commonalities such as genomic features and metabolites that facilitate interactions, and patterns of adaptation in both host and microbe. In the following sections I discuss these advances, focusing on the best-studied examples of nematode–bacterium interactions and how they have contributed to our understanding of animal–microbe symbioses (Fig. 1.2)

What makes a good symbiosis model system?

When choosing a model system to explore symbiosis, it is important to remember that what makes a good genetic model system is not necessarily or entirely what makes a good symbiosis model system. Model systems have historically been chosen for their ubiquity, ease of propagation, and ease of genetic manipulation. The same strategies that made common household pests such as mice, vinegar flies, and compost nematodes globally ubiquitous also made them ideal laboratory tools, with high fecundity and uncomplicated husbandry. Once a model system is established, it grows more powerful as researchers develop, borrow, and adapt the tools developed to answer a series of biological questions. This culminates in community-curated genomic and transcriptomic datasets, techniques, and strain collections which can represent, in some cases, genetic control of entire genomes (Kamath and Ahringer 2003). As genetic engineering techniques and next-generation sequencing have improved, scientists have brought more and more species into the fold as model systems. The process is even speeding up over time; the progression from transgenesis to binary gene expression systems in *Caenorhabditis elegans* took 45 years (Brenner 2009; Wei et al. 2012), while that same progression took only 13 years in the more recent model system of *Aedes aegypti* (Matthews and Vosshall 2020). Establishing a symbiosis model is even more complex, since it requires

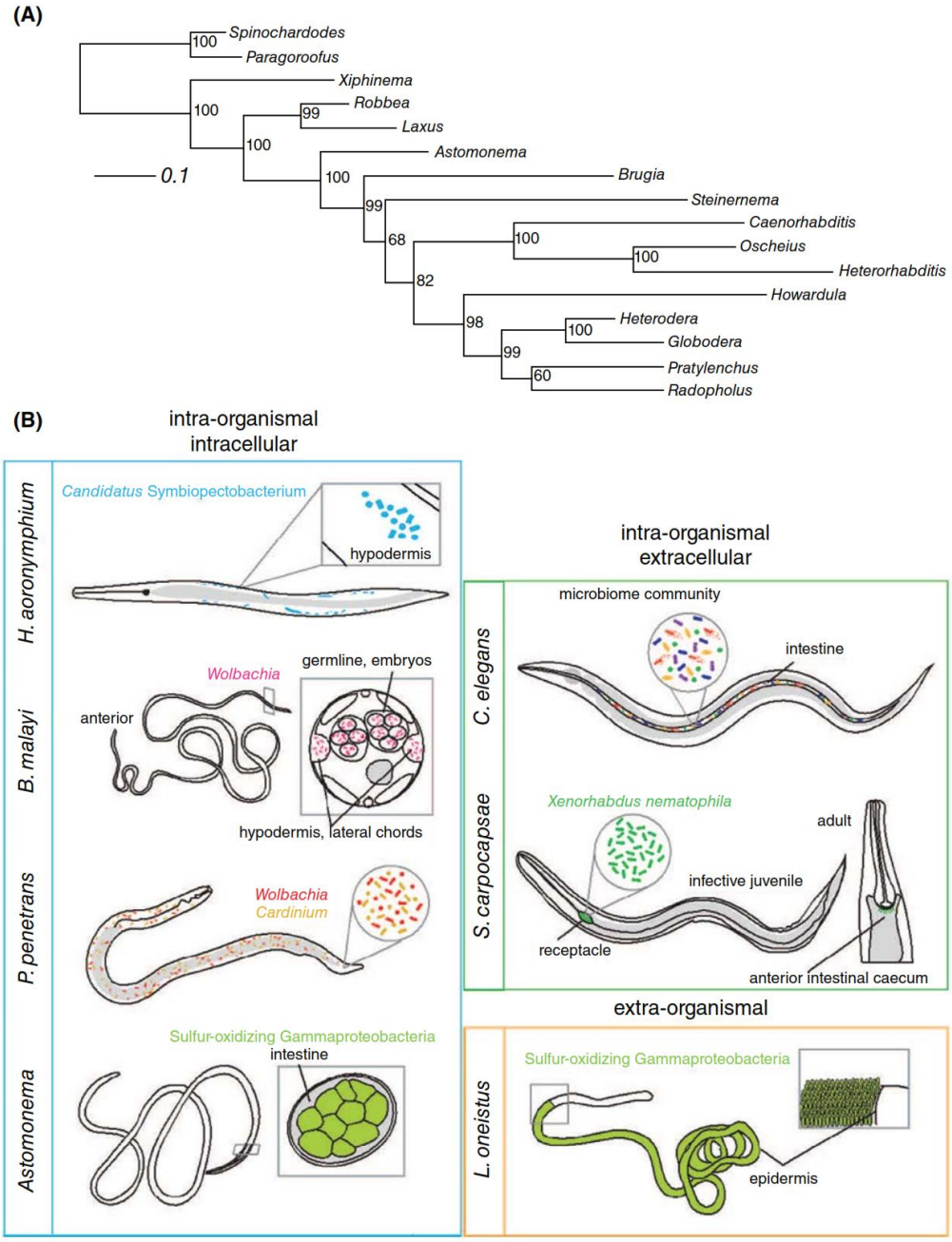


Figure 1.2 **Illustrated representation of the well-studied types of nematode-bacterium symbiosis**. **(A)** Phylogeny of select nematode genera with well-studied bacterial symbionts based on 3 loci (ITS, 18S, COI). Evolutionary relationships were estimated using a GTR + γ + I substitution model and bootstrapped 1000 times. Percent support is indicated on the phylogeny. Within the phylum Nematoda diverse lineages display convergent evolution of intimate host-symbiont relationships. **(B)** Representation of nematodes and bacterial symbionts. Symbiont niche tissues are specified when known. Bacterial symbionts are depicted and named in color. Insets show magnified views of symbiont colonized tissues. Nematode names are listed horizontally beside each organism, and intestine is gray where pictured.

finding two (or more) organisms that live in close association with each other but that can be propagated readily and, ideally, separately in the laboratory. It also requires that both can be made genetically tractable. What is more, not every pair of symbionts is going to provide the same opportunities for learning; some symbioses are obligate and lifelong, while others are facultative and transient. Some symbioses are mutualistic, while others are pathogenic. Fortunately, we do not need to find a single pair of organisms to explore all of these interactions. The phylum Nematoda is diverse, possessing multiple convergent evolutions of each type of symbiotic relationship discussed here (generalist, specialist, obligate). This breadth of possibility is precisely what makes nematodes such an exciting system.

C. elegans **as a Model for Symbiosis:**

The microbial community of the *Caenorhabditis elegans* intestine was characterized for the first time relatively recently. *C. elegans* is a useful model of facultative generalist symbiotic interactions where microbes from the environment colonize host tissue, and therefore the symbiotic community is diverse and variable. This raises interesting and important questions about host mechanisms of selection or restriction, if and how benefits are gained from symbionts, and community interactions between symbionts, that can be addressed using *C. elegans* as a model of symbiosis. These questions are also fundamental to understanding human microbial communities that have a parallel symbiotic structure. The established genetic tools, behavioral assays, and developmental progression of *C. elegans* position it particularly well for studies of host biology.

History of microbial community research in *C. elegans*

C. elegans has a long and storied history as a genetically tractable model organism. In the early 1960s, the field of molecular biology coalesced around the discoveries of mRNA (Brenner, Jacob, and Meselson 1961) and the triplet genetic code (Crick et al. 1961). The discoveries of this time period revealed for the first time the basic unit of biological information: the gene. However, in

order to determine how minute genes could build complex organisms of flesh and blood required incorporating an animal into the molecular biologist's toolset (Goldstein 2016). In a 1963 letter, Sydney Brenner, a scientist whose name would become synonymous with *C. elegans* genetics, wrote that he "would like to tame a small metazoan organism to study development directly". *C. elegans* turned out to be an ideal tool, propagating readily and autogamously in the lab (Brenner 1974). Over the decades these tiny worms facilitated the discovery of programmed cell death (Ellis and Horvitz 1986) and RNA interference (Fire et al. 1998) alongside advancing almost every aspect of developmental biology (Packer et al. 2019). They even bear the honor of being the first multicellular organism to have its whole genome sequenced (Genome sequence of the nematode *C. elegans*: a platform for investigating biology. 1998). Despite the pivotal role played by *C. elegans* and its experimental value as a genetically tractable organism, this nematode has only recently been explored as a model system for symbiosis.

C. elegans as a model for host–microbe interaction

Researchers spent years overlooking the ways in which microbes and *C. elegans* interact, choosing instead to sterilize nematodes and raise them axenically with the allopatric bacterium *E. coli* (Brenner 1974; Stiernagle 2006). The earliest forays into adapting *C. elegans* into a model for microbial symbiosis did not begin until 1999 and focused on pathogenesis. Human pathogens such as *Pseudomonas aeruginosa* (M.-W. Tan et al. 1999) and *Salmonella typhimurium* (Aballay, Yorgey, and Ausubel 2000) were introduced to nematodes in order to discover cellular mechanisms of infection and resistance. While profoundly valuable in developing our understanding of innate immune signaling (Irazoqui, Urbach, and Ausubel 2010), these studies failed to recreate the microbial communities with which *C. elegans* would interact in their natural habitat amongst decomposing plant matter. Mammalian pathogens frequently display adaptations specific for growth in a mammal (such as thermal tolerance, halotolerance, immune evasion strategies, etc.), which are unlikely to

reconstruct the patterns of infection natively encountered by wild worms. Observing the impact of *C. elegans* on soil communities has suggested a potential application of *C. elegans* in soil remediation. *Bacillus anthracis* is a potent human pathogen that can be difficult to deplete from natural reservoirs or contaminated soil, due to spore formation, but application of nematodes alongside spore-germinating compounds significantly depletes spores (Schelkle et al. 2018). While more recent studies have looked at ecologically relevant pathogens such as the microsporidia *Nematocida parisii* (Troemel et al. 2008), the bacterium *Microbacterium nematophilum* (Hodgkin, Kuwabara, and Corneliussen 2000), the fungus *Drechmeria coniospora* (L. Zhang et al. 2016), and the virus OrV (Frézal et al. 2019), pathogenic interactions also have revealed heritable mechanisms of sensing and avoiding specific bacteria (Moore, Kaletsky, and Murphy 2019). However, it remains to be seen if similar mechanisms are involved in sensing and recruiting commensal community members, and fewer studies in general have looked beyond pathogenesis into the commensal and mutualistic communities associated with *C. elegans*.

Early attempts to study commensal communities in *C. elegans* followed in the footsteps of pathogen research by focusing on commensals native to humans. Worms that were fed hypothetical human probiotic species of *Lactobacillus* and *Bifidobacterium* displayed extended lifespans and increased resistance to infection by *Salmonella* (Ikeda et al. 2007). Strains of *Lactobacillus acidophilus* have protective effects against infection by Gram-positive bacteria, *Enterococcus faecalis* and *Staphylococcus aureus* (Y. Kim and Mylonakis 2012). The common interbacterial signaling molecule indole endows worms with relative resistance to colonization by the opportunistic fungal pathogen *Candida albicans* (Oh et al. 2012). Additionally, recent work revealed that the probiotic strain *Bacillus subtilis* PXN21, or metabolites derived from it, can prevent the aggregation of α-synuclein in a DAF-16-dependent manner. Such aggregation is associated with disease severity in Parkinson's disease (Goya et al. 2020).

To provide a more natural context, later investigations focused on the impacts of nematode-derived microbes on *C. elegans* biology. *C. elegans* transcriptomics, enabled by the nematode's well-annotated genome, revealed that exposure to soil bacteria in general altered expression of genes involved in metabolism, innate immunity, and cuticle biosynthesis (Coolon et al. 2009). More specifically, nematodes that were exposed to an artificial community of *C. elegans* microbiome isolates displayed upregulation of immune genes and hydrolases known to be expressed in the intestine (M. Berg et al. 2019). In another study, culture-based methods isolated two soil bacteria, *Bacillus megaterium* and *Pseudomonas mendocina*, which were protective against *Pseudomonas aeruginosa* infection. The protective effects of *B. megaterium* were mediated by a behavioral change in nematodes that resulted in diversion of energy resources away from egg-laying and toward immunity. The protective effects of *P. mendocina* were dependent on stimulating the innate immune response via the p38 MAPK pathway (Montalvo-Katz et al. 2013). Collectively, strains isolated from nematodes suggest that bacterial exposure is important for establishing gene expression, priming the innate immune system against pathogen exposure, and potentially modulating nematode behavior.

Characterization of native microbiome of *Caenorhabditis*

In 2016, three independent studies (M. Berg et al. 2016; Dirksen et al. 2016; Samuel et al. 2016) utilized amplicon sequencing of the V4 region of the 16S ribosomal subunit gene to characterize the microbiome of *C. elegans*. Samuel et al. (2016) holistically examined bacterial communities from rotting vegetative matter inhabited by *C. elegans* and revealed 564 culturable bacteria and about 2400 operational taxonomic units (OTUs). Of the culturable isolates, 80% enhanced nematode growth compared with *E. coli*. The remaining 20% impaired growth and/or upregulated expression of one or more stress genes. Berg et al. (2016) allowed aposymbiotic laboratory-strain nematodes to recruit bacteria from environmental soil samples prior to amplicon sequencing. Of the approximately 2400 OTUs found within the worm intestine, members of nine

bacterial families were actively recruited and were proposed to constitute a "core microbiome". Additionally, spatial distribution and soil temperature impacted the composition of the *C. elegans* gut community. Finally, Dirksen et al. (2016) sequenced amplicons from wild isolates of *C. elegans* as well as the nematodes *Caenorhabditis briggsae* and *Caenorhabditis remanei*. This analysis demonstrated that gut microbiomes differ according to nematode species, genotype, and developmental stage. Using individual isolates alongside an artificial community of 14 bacteria, the authors demonstrated that commensal isolates enhanced nematode growth compared with *E. coli* and increased resistance to pathogenic fungal infection by *D. coniospora*. A meta-analysis of all three datasets used principal components analysis to show that while bacterial communities vary a great deal by soil microenvironments (for example compost, fruit, microcosms), nematode gut communities across all studies cluster closely together and are dominated by bacteria from the phyla Proteobacteria, Bacteroidetes, and Firmicutes (F. Zhang et al. 2017). Furthermore, certain bacterial isolates, such as *Ochrobactrum* MYb71 and *Stenotrophomonas* MYb57, are enriched in worms compared with the environment. This enrichment supports the hypothesis that these isolates are colonizing the intestines and actively reproducing (Dirksen et al. 2016). Like the human gut microbiome, the *C. elegans* microbiome is selected from environmental reservoirs for specific representatives that support increased growth, prime the immune system, and increase resistance to colonization by pathogens.

Tools for microbiome research

The above meta-analysis resulted in the generation of the *Caenorhabditis elegans* Microbiome Resource, henceforth CeMbio (Dirksen et al. 2020). CeMbio is a collection of 12 bacterial isolates that represent the most common OTUs found in *C. elegans* collected from the wild. These 12 isolates represent all nine of the bacterial families identified as being constituents of the *C. elegans* core microbiome. Each of these representatives has a fully sequenced genome and accompanying

metabolic network to facilitate microbiome studies. These strains can be identified and enumerated via quantitative polymerase chain reaction (qPCR) amplification with diagnostic primers, enabling the construction of defined artificial communities (Dirksen et al. 2020). This toolset complements the already robust strain collections available for *C. elegans* and paves the way for future studies to determine mechanisms of host–microbe nutrient exchange, immune activation, and protection from pathogens. Resources on the nematode side of the symbiosis include the *Caenorhabditis* Genetics Center (CGC), which offers an extensive collection of *C. elegans* mutants, and the *Caenorhabditis elegans* Natural Diversity Resource (CeNDR), which offers fully sequenced wild isolates of *C. elegans* with detailed geographic metadata (Cook et al. 2017). These resources make *C. elegans* a uniquely valuable tool for a broad range of investigators to examine genetic determinants for microbiome acquisition, maintenance, and specificity.

Responding to bacteria: *C. elegans* innate immunity

C. elegans is uniquely qualified for exploring the role of host innate immunity in establishing and maintaining a resident microbiome. *C. elegans* entirely depends on humoral immune signaling to respond to microbial challenge, since nematodes have no adaptive immune system. Genetic immune determinants can be examined without the confounding variables of acquired immunity (Shapira 2009). Also, the array of genetic tools available for *C. elegans* enables measuring the impact of immune pathway disruption on microbiome and nematode health/fitness/behavior. Forward genetic screens can be achieved via mutagenesis, while targeted reverse genetic manipulations can be accomplished by RNAi exposure (Shapira and Tan 2008) or CRISPR-Cas9 editing (Dickinson and Goldstein 2016). Finally, the *C. elegans* microbiome can be flexibly populated with a range of members from a single monoclonal isolate to thousands of OTUs from the wild. The core microbiome, as represented by CeMbio, facilitates well-controlled studies on the role played by

innate immunity in community recruitment, while human pathogens that also infect *C. elegans* are used to discover molecular mechanisms by which innate immunity combats pathogenic infection.

C. elegans has already been essential to a number of discoveries in innate immunity. The process of RNAi, now used experimentally to knock down gene expression in a wide range of organisms, was originally discovered in *C. elegans* where it naturally serves as a protective measure against RNA viruses (Fire et al. 1998). This process is conserved in many eukaryotes, including plants and animals, where, in addition to defending against viral infection, it acts as a regulatory element of endogenous genes. Investigation of the *C. elegans* RNAi system yielded knowledge of viral defense against its native pathogens, such as the OrV virus (Frézal et al. 2019), as well as diverse mammalian viruses such as Indiana vesiculovirus and human immunodeficiency virus-1 (Diogo and Bratanich 2014).

C. elegans does not appear to have the canonical immune deficiency (IMD) pathway found in insects, or its mammalian equivalent, the tumor necrosis factor (TNF) receptor pathway. However, *C. elegans* does have a single Toll-like receptor (TLR) protein whose role in development expands the functional range of the TLR family (Irazoqui, Urbach, and Ausubel 2010). The Toll protein was first discovered in *Drosophila* where, in addition to its involvement in dorso-ventral patterning, it has roles in detecting microbe-associated molecular patterns (MAMPs) and initiating cellular development. TLRs in humans and mice serve similar roles. The *C. elegans* TLR is essential for pathogen avoidance behaviors but this activity does not involve detection of MAMPs. Instead it serves a strictly developmental role in maturing the BAG neurons, which are responsible for CO_2 detection (Brandt and Ringstad 2015). As such, reducing TLR function leaves worms susceptible to infection by some microbes (*Salmonella marcescens* and *Salmonella enterica*) but has no effect on others (*Salmonella aureus*, *P. aeruginosa*, and *D. coniospora*). Potentially, this disparity could be caused by pathogenic microbes

having different CO_2 signatures in the environment. This finding highlights that, the roles of TLR in development and immune signaling may be more tightly integrated than currently appreciated.

As in humans, the p38 immune signaling pathway is integral to innate immunity in *C. elegans*. *C. elegans* has been used to demonstrate that neuronal tissues are responsible for coordinating responses to infection via stimulation of the p38 MAPK pathway (Styer et al. 2008). As such, neurons need to be able to differentiate commensal bacterial signals from pathogenic bacterial signals in order to respond appropriately to microbiome recruitment. As mentioned previously, the p38 pathway is essential for *C. elegans* to benefit from the protective effects of priming by the gut-dwelling commensal *P. mendocina*. *P. mendocina* activates the p38 pathway in a manner similar to *P. aeruginosa*, causing the worm to pre-emptively adapt for potential infection. The MAPK cascade initiated within *C. elegans* is homologous to that seen in humans; however, the transcription factors activated by this cascade have yet to be identified (Montalvo-Katz et al. 2013). Despite the knowledge gaps in this immune pathway, it is hypothesized to drive production of antimicrobial peptides and programmed cell death based on its responses to challenge with *P. aeruginosa* (Irazoqui, Urbach, and Ausubel 2010). The similarities and differences in p38 pathways between *C. elegans* and mammals help to bridge the gap between vertebrate and invertebrate immunology.

C. elegans possesses a transforming growth factor-β (TGFβ) signaling pathway which is activated by the neuronally derived growth factor DBL1 (Irazoqui, Urbach, and Ausubel 2010). This pathway has been implicated in defense against diverse microbes and across multiple tissue types. The TGFβ pathway facilitates resistance against epithelial infection by the fungus *D. coniospora* by detecting paracrine signals from neuronal tissue and inducing expression of the antimicrobial peptide caenacin (Zugasti and Ewbank 2009). Meanwhile in the gut, TGFβ pathway mutants lacking the gene encoding DBL-1 show an altered microbiome, in which total bacterial abundance is only modestly increased, but *Enterobacteriaceae* abundance doubled. This community disruption occurred

in a tissue-specific manner centered on the anterior gut and enabled a normally commensal *Enterobacter* to reach a pathogenic abundance (M. Berg et al. 2019). The homologous TGFβ signaling seen in humans also shapes microbial communities in a number of ways, including regulation of inflammatory response and priming lymphocytes to be tolerogenic of commensal bacteria (Zeuthen, Fink, and Frokiaer 2008). These observations demonstrate the vital role played by neuroendocrine signaling in innate immunity and establishes *C. elegans* as a valuable system for exploring neuroendocrine signaling effects on microbial communities, and vice versa.

The insulin-like signaling pathway, represented by the insulin receptor DAF-2 and the transcription factor DAF-16, is also capable of mediating microbiome response in *C. elegans* (Irazoqui, Urbach, and Ausubel 2010). Insulin-like signaling contributes independently to nematode longevity and bacterial resistance (Miyata et al. 2008). DAF-16 is responsible for transcription of many genes implicated in bacterial defense, but only specific infections result in activation of this pathway. For example, *P. aeruginosa* infection repressed expression of DAF-16 and enteropathogenic *E. coli* (EPEC) infection induced DAF-16 expression resulting in resistance to subsequent infection (Anyanful et al. 2009). Encoding multiple signaling pathways, like *C. elegans* does, is essential for organisms to fine-tune immune responses against distinct microbes. It should be noted that some bacteria, such as *S. aureus*, induce a response in nematodes that is independent of the previously mentioned pathways (Irazoqui, Urbach, and Ausubel 2010). This suggests that *C. elegans* has novel signaling pathways yet to be characterized, leaving the field open for growth. Additionally, future research could utilize more commensal communities instead of allopatric pathogens to determine how the innate immune system avoids unnecessary activation against its own microbiome.

Advantages of *C. elegans* as a model system for symbiosis research

C. elegans has been a staple model organism for decades and its value has not yet been exhausted, especially as it relates to host–microbe interactions. *C. elegans* offers an experimental

sandbox to explore facultative generalist interactions in a simplified microbiome. *C. elegans* can reproduce by self-fertilization, rapidly, and in the laboratory. It has one of the most extensive genetic toolkits of any metazoan and the developmental history of every cell in its body has been carefully mapped (Sulston et al. 1983). Protocols and automation pipelines have already been developed to describe life stages and behavioral outputs, such as egg-laying, activity, and microbial prey seeking (Collins et al. 2016; Pitt et al. 2019). Thanks to the extensive mapping of the *C. elegans* nervous system, exposure to the commensal *Providencia rettgeri* has already been demonstrated to impact the worm's olfactory responses and feeding preferences via the production of the neurotransmitter tyramine (O'Donnell et al. 2020). Finally, isolated microbiome constituents and host nematodes have been collected from around the globe and can be reared together or in isolation to parse apart genotypic determinants of symbiosis. The combination of host genetic tools, fate mapping, neurobiology, and established behavioral assays makes *C. elegans* an especially powerful model for understanding how animal hosts assemble beneficial microbial communities. The genetics of the recently identified microbial community members are just beginning to come online, with full genome sequences adding power to possibility (Dirksen et al. 2020). The lives of *C. elegans* nematodes are spent seeking out bacteria in their own terrain, where being able to separate commensal from pathogen can be the difference between dinner and death. In a sense, they are some of the original microbiologists and, if we listen closely enough, they might just tell us their secrets.

Entomopathogenic Nematodes as a Model for Symbiosis:

Entomopathogenic nematodes (EPNs) infect and kill insects in cooperation with intergenerationally transmitted bacterial symbionts (Dillman et al. 2012). This tripartite relationship between the insect, the nematode, and the bacterium has allowed researchers to explore the range of symbiosis outcomes, as the partners in the system exhibit activities from mutualism to antagonism at

various points in the life cycle. EPNs can be reared with their symbionts outside of insects, and *Steinernema*, though not *Heterorhabditis*, can be cultivated on specialized media without bacteria, enabling the establishment of a range of tools for bacterial manipulation and an ever-expanding toolset for nematodes. These tools are used to answer questions about the molecular basis of colonization and host recognition, the necessity of metabolites and nutrient environments for symbiosis, and the evolution and maintenance of precise, strain-level host–microbe interactions relevant to a range of fields, from the human microbiome to agricultural pest control.

Discovery of entomopathogenic nematodes

EPNs were first discovered in 1923 (Poinar and Grewal 2012). Because up to one-fourth of crops are destroyed by insects worldwide (Clarke 2020; Kergunteuil et al. 2016), early research focused on utilizing EPNs for agricultural pest control. Although difficulties in producing EPNs on an industrial scale initially limited enthusiasm about their use in biocontrol of insect pests (Ehlers 2001), EPN research was revived in the mid-20th century when, in Europe and North America, scientists independently isolated EPN populations from codling moth larvae (Dutky and Hough 1955; Shishiniova, Budurova, and Gradinarov 1998). These nematodes became a cornerstone of the EPN research field and were later classified as *Steinernema carpocapsae* (Poinar and Leutenegger 1968; Poinar and Thomas 1967). Today, two EPN families, *Heterorhabditidae* and *Steinernematidae*, are commercially produced for use against insect pests in agricultural products such as NemAttack™, which consists of *Steinernema feltiae* nematodes (Lacey et al. 2015; Shapiro-Ilan et al. 2013). Most research on EPN–microbe symbiosis to date focuses on *Heterorhabditis* and *Steinernema* nematodes because of their development for use as entomopathogens throughout the 20th century.

Although researchers had observed the presence of bacteria inside the nematodes as early as 1938 (Bovian 1938), the roles of the symbionts in the life cycles of EPNs would not be discovered for nearly 30 years (Poinar and Thomas 1966, 1967). Symbiont laboratory isolation and cultivation

required the development of media enabling propagation of the nematode–bacterium pairs outside of insects, as well as separately from each other (House, Welch, and Cleugh 1965; Stoll 1953). Poinar and Thomas's isolation of the *S. carpocapsae* bacterial symbiont *Xenorhabdus nematophila* in 1965 paved the way for the development of EPNs as a symbiosis model system (Poinar and Thomas 1965). We now know that all members of both EPN families are associated with and dependent on a microbial symbiont: *Heterorhabditidae* with *Photorhabdus* bacteria and *Steinernematidae* with *Xenorhabdus* bacteria (Poinar 1990).

Advancement of EPNs as a model for host–microbe interaction

The value of EPN nematodes as a model system for symbiosis lies partly in their life cycle, which includes measurable and reproducible stages of reproduction and development as well as discrete and observable interactions with the bacterium (Clarke 2020; Goodrich-Blair 2007; Richards and Goodrich-Blair 2009; Stock 2005, 2019). Early research showed that the non-feeding, soil-dwelling infective juvenile stage of an EPN nematode is intestinally colonized by the bacterial symbiont (Poinar 1966; Poinar and Leutenegger 1968; Poinar, Thomas, and Hess 1977). Poinar discovered that *Xenorhabdus* symbionts colonize the anterior intestinal lumen of infective juveniles in a specialized compartment referred to as the receptacle (previously termed the vesicle) (S. K. Kim, Flores-Lara, and Stock 2012; Poinar 1966), and that *Photorhabdus* symbionts colonize the lumen of the pharynx and intestine of infective juveniles (Poinar, Thomas, and Hess 1977). After finding prey, the nematode–bacterium pair infects the insect through the mouth, anus, spiracles, or cuticle (Barbercheck 2008; A. Peters and Ehlers 1994). The nematode releases its symbiont into the insect hemolymph, where the bacteria replicate, kill the insect, and degrade the cadaver for nutrients. The nematode farms the bacteria for its own food until nutrients are depleted (Clarke 2020; Mucci et al. 2022; Stock 2019). At that point, a new generation of infective juveniles associates with its bacterial symbiont and exits the cadaver in search of a new insect host to repeat the cycle.

The relationship between EPN hosts and their bacterial symbionts is most likely to be obligate in nature, as to date neither has been isolated without the other in field samples. This statement is supported by the observations that the bacteria depend on the nematode for transport between insect prey (Poinar 1966) and, in turn, serve as the major EPN food source (Clarke 2020; Mucci et al. 2022; Stock 2019). Further, *Xenorhabdus* bacteria are auxotrophic for nicotinic acid dinucleotide (NAD), an essential coenzyme for cellular metabolism, and must derive this molecule from host environments (Chaston et al. 2011). However, in the laboratory when nutrients are supplemented abiotically, the symbiosis can be made facultative as the bacteria can be cultured independently, and, at least for *Steinernema*, the nematodes can be raised axenically on specialized media. The technical advances in culturing the *Photorhabdus* and *Xenorhabdus* symbionts independently in the laboratory (Poinar and Thomas 1965; Stoll 1953) and in genetically manipulating them (J Xu et al. 1989, 1991) paved the way for breakthroughs like the expression of the green fluorescent protein (GFP) in the bacterial symbionts. This enabled investigation of processes such as symbiont intergenerational transmission and laid the foundation for EPNs as a model system.

The bacterial symbionts of EPN are transmitted to progeny infective juveniles (and therefore to new insect hosts) through nematode tissue specific localization and symbiont bottlenecks that select for one or a few individual cells (Ciche et al. 2008; Martens, Heungens, and Goodrich-Blair 2003). In both *Heterorhabditis bacteriophora* and *S. carpocapsae*, as the nematodes develop into infective juveniles a few bacterial cells localize to the pharyngeal intestinal valve (Chaston et al. 2013; Ciche et al. 2008). During this stage, the *S. carpocapsae* intestine fully constricts before the anterior intestinal region expands so that the lumen forms the receptacle, which is colonized by 1–2 *X. nematophila* cells that are presumed to derive from those that colonized the pharyngeal intestinal valve. These few receptacle-colonizing *X. nematophila* then replicate to fill the receptacle (Bird and Akhurst 1983;

Chaston et al. 2013; Martens, Heungens, and Goodrich-Blair 2003; Snyder et al. 2007). In *H. bacteriophora*, the transmission process by which *Photorhabdus luminescens* cells colonize the infective juvenile pharyngeal-intestinal valve has been elucidated. In adult female *H. bacteriophora* nematodes, *P. luminescens* adheres to and invades posterior intestinal cells, and it is these maternally colonizing cells that are transferred to progeny infective juveniles that develop inside the mother. As in the *X. nematophila–S. carpocapsae* symbiosis, these *P. luminescens* cells colonize the pharyngeal-intestinal valve of the developing infective juveniles (Ciche et al. 2008). Curiously, in *S. carpocapsae* adult nematodes, *X. nematophila* bacteria localize to the anterior intestinal caecum, and posterior colonization has not been observed (Chaston et al. 2013). The role, if any, of anterior intestinal colonization in transmission has not been experimentally established. Overall, these findings suggest that while details of the transmission process vary, there is a bottleneck in both systems that limits the diversity of symbionts that are transmitted to the next generation. These observations added to a growing number of examples across diverse symbiotic systems demonstrating that symbiont populations become genetically bottlenecked at specific life-cycle stages (Chaston et al. 2013). The concept that symbiont bottlenecks are common among horizontally transmitted systems contradicts the prevailing theory that such bottlenecks will not occur because they are expected to result in accumulation of deleterious mutations in the symbiont that will negatively impact symbiotic fitness. They also lend support for the evolutionary concept that high population densities of host and symbiont can balance the negative effects of bottlenecks (Chaston and Goodrich-Blair 2010; Muller 1964; Pettersson and Berg 2007).

The findings summarized above have implications for how hosts select and transmit specific symbiotic partners. Colonization and transmission by only one or a few bacterial cells necessarily indicate that other cells were restricted access to these niches. These observations further bolstered the idea that selection and partner recognition mechanisms play a role in EPN symbiosis. Such

selective processes appear to include physical (for example, the colonization site, such as the pharyngeal-intestinal valve, is spatially limited and only accommodates a few cells) and molecular (for example, interactions between surface molecules present on both the nematode and bacterial cell surfaces). For instance, within the *S. carpocapsae* infective juvenile receptacle, *X. nematophila* associates with a mucosal layer surrounding a cluster of non-cellular (i.e., lacking a nucleus) spherical bodies, termed the intravesicular structure (IVS). Similarly, both *P. luminescens* and *X. nematophila* appear to be embedded in a polysaccharide material within their respective infective juvenile host's intestine (Ffrench-Constant et al. 2003; Martens, Russell, and Goodrich-Blair 2005). Bacterial attachment to host mucus could indicate a lectin–glycan mechanism of attachment and specificity, with a bacterial protein acting as the lectin (Chaston et al. 2013; Hooper and Gordon 2001; Martens, Russell, and Goodrich-Blair 2005).

Advantages of EPNs as a model system for symbiosis research

As noted in previous sections, EPNs present several distinct advantages as nematode model systems for understanding symbiosis. Unlike other models of beneficial animal–microbe interactions, where complexity necessitates artificial simplification of the microbiota, this is a naturally occurring, highly specific animal–microbe association. Both the nematode host and the bacterial symbiont can be cultured and manipulated independently or together in the laboratory and are amenable to experimental evolution studies (Morran et al. 2016). High-quality genomes and multiple transcriptomes exist for multiple nematode–microbe species pairs (to date, most notably: *S. carpocapsae*–*X. nematophila*, *S. feltiae*–*X. bovienii*, and *H. bacteriophora*–*P. luminescens*), providing a strong foundation for future evolutionary comparisons (Bai et al. 2013; Chang et al. 2019; Chaston et al. 2011; Dillman et al. 2015; Duchaud et al. 2003; Fu et al. 2020; Lu et al. 2017; Murfin, Whooley, et al. 2015; Rougon-Cardoso et al. 2016; Serra et al. 2019). Evidence is accumulating that pheromone signaling between EPNs, both within and amongst species, via conserved signaling molecules called

ascarosides, is likely an important regulator of nematode behavior (Choe et al. 2012; Hartley et al. 2019; Kaplan et al. 2012). Future experiments will reveal if and how these signaling molecules are involved in aspects of symbiosis and whether symbionts can sense and respond to these important host-generated cues. A full suite of genetic tools has been developed for modifying the *Xenorhabdus* and *Photorhabdus* bacteria. The development of the nematodes as genetic model systems has lagged behind their bacterial counterparts, with relatively scarce reports of successful gene knockdown by RNAi and transgenic gene expression (Ciche and Sternberg 2007; Hashmi, Hashmi, and Gaugler 1995; Morris et al. 2017; Ratnappan et al. 2016). New advances in genome manipulation, including CRISPR-Cas9 gene editing, have been successfully applied in other "non-model" nematodes and are likely to work in EPNs as well (Castelletto, Gang, and Hallem 2020; Ward 2018). These efforts are well worth the investment of time and resources, as EPNs have a wealth of interesting biology to offer, including diverse host-seeking behaviors, reproductive patterns such as mating and egg-laying, a nutrient responsive life-cycle transition, insect virulence, and questions surrounding aging, all of which may be influenced by their symbiotic partnership.

Role of non-obligate symbionts in EPN life cycles

Advances in next-generation sequencing have expanded our knowledge of the nematode associated microbial communities, including in EPNs. Early investigations of EPN colonizing bacteria focused primarily on the cognate bacterial symbiont, though there were reports using culture-dependent techniques which suggested that there may be additional members of the *Steinernema* microbiome (Bonifassi et al. 1999; Gouge and Snyder 2006). Recently, a study using metabarcoding of the V3V4 region of the 16S rRNA gene and housekeeping *rpoB* markers showed that both laboratory and wild populations of Steinernema species frequently are associated with a dozen additional Proteobacteria species in addition to their obligate *Xenorhabdus* symbiont (Ogier et al. 2020). Several species of *Pseudomonas* were predicted to play a role in the insect-killing

"pathobiome" of *Steinernema* due to their entomopathogenic nature, but, in general, the function of the recently discovered "frequently associated microbiome" is unknown (Ogier et al. 2020). *Xenorhabdus* is well equipped to compete with other bacterial species (Ciezki et al. 2017; Dreyer, Malan, and Dicks 2018; Thaler, Baghdiguian, and Boemare 1995; Thappeta et al. 2020), so it will be interesting to explore in the future how these additional members of the EPN microbiome survive within the insect cadaver and how they are transmitted by nematodes between insects. With the increasing availability of sequencing tools, our understanding of the traditional EPN tripartite insect–nematode– symbiont life cycle will expand to include these additional non-obligate symbionts.

Nematode–microbiome models of insect and gastropod pathogenesis

While *Steinernema* and *Heterorhabditis* are accepted as the "true" entomopathogenic nematodes, the family *Rhabditidae* contains other organisms that have adapted this lifestyle as a function of their microbiomes. Most notably, the nematode genera *Heterorhabditidoides* (not to be confused with *Heterorhabditis*) and *Oscheius* include members that are facultatively entomopathogenic in insects of multiple orders (Dillman et al. 2012; Torres-Barragan et al. 2011; K. Y. Zhang et al. 2012). These nematodes can exist as free-living bacterivores but gain insect-killing potential when their external cuticle is colonized by specific strains of *Serratia marcescens* or *Serratia nematodiphila* (C.-X. Zhang et al. 2009). This flexible microbiome is facultatively generalist and could represent an intermediate step towards the evolution of a more intimate host–microbe association, like that seen in *Steinernema*. Studying the mechanisms of microbiome assembly in these facultative EPNs and comparing them with the obligate EPNs would provide insight into the evolution of parasitism as well as demonstrating the evolutionary trade-offs of carrying a surface associated symbiont compared with a gut associated symbiont.

The genus *Phasmarhabditis* has been extensively studied as a biological control for slugs and snails that act as agricultural pests (El-Danasoury and Iglesias-Piñeiro 2017; M. J. Wilson and Rae 2015). In fact, *Phasmarhabditis hermaphrodita* is commercially available under the trade name Nemaslug® throughout Europe. Like the aforementioned *Heterorhabditidoides*, *Phasmarhabditis* spp. can exist as free-living bacterivores or as facultative parasites. *Phasmarhabditis* spp. were initially assumed to utilize highly specific microbial associations to promote gastropod pathogenicity, supported by evidence that specific symbionts such as *Moraxella osloensis* displayed toxicity to slugs (L. Tan and Grewal 2001). However, later discoveries showed that *Phasmarhabditis* can associate with variable communities and that no specific bacterium is uniquely necessary for pathogenesis (Rae, Tourna, and Wilson 2010). Sequencing-based studies could be utilized in the future to identify a core microbiome for this economically relevant nematode. Nematodes have independently evolved to become invertebrate parasites on many occasions and these groups each provide unique snapshots of host–microbe specificity, symbiont acquisition, and inter-kingdom communication which can be reconstructed readily in the laboratory and contextualized across evolutionary space. Few biological systems provide so many examples of convergent evolution that can be phylogenetically linked.

Molecular mechanisms of host–microbe interaction

One of the major contributions of the EPN– bacterium model to the symbiosis field is the expansion of our understanding of the bacterial proteins and molecular mechanisms of animal host colonization. The development of new experimental tools for genetic manipulation of the symbiont (J Xu et al. 1989, 1991) enabled identification of loci necessary to establish and maintain symbiosis (Ciche et al. 2001; Heungens, Cowles, and Goodrich-Blair 2002). These included both site-directed and transposon mutagenesis followed by assessment of host colonization phenotypes (Ciche et al. 2001; Easom, Joyce, and Clarke 2010; Heungens, Cowles, and Goodrich-Blair 2002; Martens et al. 2003; Martens, Heungens, and Goodrich-Blair 2003; Somvanshi et al. 2010; Vivas and Goodrich-

Blair 2001). The development of a signature-tagged mutagenesis approach, the precursor to currently used techniques such as RbTnSeq (Shields and Jensen 2019), allowed higher-throughput pooled screening of *X. nematophila* mutants for a colonization defective phenotype (Heungens, Cowles, and Goodrich-Blair 2002). Further, the availability of sequenced bacterial genomes enabled the rapid identification of transposon-disrupted loci (Chaston et al. 2011; Wilkinson et al. 2009).

The discovery of colonization-deficient bacterial mutants in both *Xenorhabdus* and *Photorhabdus* allowed researchers to delve into the molecular mechanisms that enable bacterial colonization of animal hosts, using EPNs as a model system. In particular, unbiased screens have revealed novel factors not previously implicated in symbioses. For instance, the screen for colonization-defective *X. nematophila* signature tagged mutants revealed a locus dubbed "Symbiosis Region 2", or SR2, in which the transposon had inserted into a sequence (*nilD*) with palindromic repeats that were also present at other locations in the genome and with sequence identity to the mysterious "iap" repeats found in *Escherichia coli* (Heungens, Cowles, and Goodrich-Blair 2002; Nakata, Amemura, and Makino 1989). In the same year that the repeat sequence *nilD* of SR2 was reported as being necessary for colonization, the term CRISPR was coined to describe "clustered regularly interspaced palindromic repeats", including the *E. coli* iap repeats (Jansen et al. 2002). Subsequent analyses confirmed that *nilD* is a CRISPR RNA, the first, and perhaps only, known example of a CRISPR RNA necessary for mutualistic colonization of a host, broadening the scope of CRISPR biological function beyond bacterial defense (Veesenmeyer et al. 2014).

Nutrition and secondary metabolism

Lrp is a member of the feast-or-famine regulator family (Yokoyama et al. 2006). In *X. nematophila*, this regulator of population heterogeneity and symbiotic behaviors is likely sensing and responding to the prevailing metabolic conditions which may indicate its symbiotic state. Similarly, specific metabolites control the DNA-binding activity of LysR-type regulators of which the *P.*

luminescens transcription factor HexA is a member (Joyce and Clarke 2003). HexA controls numerous symbiotic activities in *P. luminescens*, and, like Lrp, may be important for controlling gene expression temporally over the symbiotic life cycle (Clarke 2020; Joyce and Clarke 2003). An important aspect of metabolism controlled by each of these regulators in their respective bacteria is the production of complex small molecules that fall outside primary conserved metabolism and are therefore known as secondary metabolites (Engel et al. 2017). The importance of bacterial secondary metabolites in *Steinernema* and *Heterorhabditis* symbiosis is suggested by the fact that in both *X. nematophila* and *P. luminescens*, *ngrA* mutants lacking 4'-phosphopantetheine (Ppant) transferase, a key enzyme in the nonribosomal-peptide-synthetase (NRPS)-dependent biosynthesis of secondary metabolites, do not support the reproduction and development of their respective nematode partners (Ciche et al. 2001; Swati et al. 2015).

Secondary metabolites produced by *Photorhabdus* may have garnered curiosity over a century ago in the form of the bioluminescence it produces. In fact, *Photorhabdus* is the only terrestrial bacterium known to date with bioluminescent properties. Insects infected with *Heterorhabditis* nematodes associated with their symbiont will glow in the dark (Poinar and Grewal 2012). Legends from the battlefields of the American Civil War describe soldiers with glowing wounds (Nealson and Hastings 1979). Although the cause of the luminescence was not discovered at the time and there is no way to investigate the reports, modern microbiologists have speculated that the luminescence was due to opportunistic *Photorhabdus* infections. Physicians in the Civil War era believed that the presence of the bacterium would lead to good patient outcomes (Nealson and Hastings 1979). The discovery of antibiotic properties of *Xenorhabdus* and *Photorhabdus* means that this belief could have some credibility (Dutky 1959; Webster et al. 2002), although there is at least one bioluminescent species of *Photorhabdus* that can be rarely pathogenic to humans (Gerrard et al. 2003, 2004; Hapeshi and Waterfield 2017). To date, little is known about the functionality of the luminescence to the

nematode or the symbiont, but it has been suggested to help protect an insect cadaver from predation (Maher et al. 2021).

Prior work on the nutritional demands of *X. nematophila* elucidated some of the components of the host environment needed for colonization and symbiosis, including that *X. nematophila* is a nicotinamide adenine dinucleotide (NAD) auxotroph. Using *X. nematophila* mutants that were found to be auxotrophic for various nutrients in culture, Martens et al. (2005) examined the colonization levels of *S. carpocapsae* to probe the nutrient environment of the nematode. *X. nematophila* mutants unable to produce select amino acids, vitamins, and nutrients were able to colonize *S. carpocapsae* infective juveniles to normal wild-type levels, suggesting that the colonization environment of the nematode receptacle is able to provide some nutrients to the symbiont (Martens, Russell, and Goodrich-Blair 2005). However, colonization levels were decreased among *X. nematophila* with mutations in the para-aminobenzoate, pyridoxine, and L-threonine biosynthesis pathways (Martens, Russell, and Goodrich-Blair 2005), leading to the conclusion that the nematode environment lacks sufficient quantities of one or more essential metabolites related to these pathways, with one-carbon metabolism representing a common thread. Therefore, at least at the infective juvenile stage, the symbiont appears to be dependent on some host-provisioned nutrients.

EPN–symbiont pairs as models of symbiont specificity and host switching

Strain variation can play a part in the fitness outcomes of one or more of the partners in a symbiosis, and the *Steinernema–Xenorhabdus* model represented an excellent opportunity to investigate this concept. A *Steinernema* nematode exhibits the greatest fitness when associated with its native symbiont strain (Chapuis et al. 2009; Murfin, Lee, et al. 2015b; Murfin, Whooley, et al. 2015), and cannot be well colonized by non-native symbionts (Bird and Akhurst 1983; Cowles and Goodrich-Blair 2008; Sicard et al. 2004). Further, *Xenorhabdus bovienii* provides an opportunity to assess strain-level specificity, because the bacterial strains have a 96% average nucleotide identity with each other

yet associate with *Steinernema* nematodes that segregate phylogenetically into different clades within the *Steinernema* genus (McMullen et al. 2017; Murfin, Lee, et al. 2015b; Murfin, Whooley, et al. 2015). Within the *X. bovienii* species, there is variation in genomic coding regions between strains, especially in genes predicted to be involved in host interactions, indicating the potential for distinctive symbiotic traits expressed within the species (Murfin, Lee, et al. 2015b).

The effect of strain variation on host fitness was tested using *X. bovienii* strains and the *Steinernema* nematodes that they colonize. When pairs of *X. bovienii* strains and *Steinernema* nematodes were combined and examined for fitness using several different parameters (including percentage of productive infections, number of progeny produced, and progeny infective potential), the strains that were genetically less related to the native symbiont of the nematode species resulted in lowered fitness for the host (McMullen et al. 2017; Murfin, Whooley, et al. 2015). In addition, *X. bovienii* strains isolated from different strains of *Steinernema feltiae* nematodes produce bacteriocins (tailocins) that are active against other *X. bovienii* strains capable of colonizing the same *S. feltiae*, indicating the potential for interspecies competition during colonization (Ciezki et al. 2017). In some cases, an association with the *X. bovienii* bacterial symbiont can even lead to death of the non-native nematode host but is not harmful to the native nematode host (McMullen et al. 2017; Murfin et al. 2018; Murfin, Whooley, et al. 2015). This indicates that nematodes and bacterial symbionts might compete for the resources of an insect cadaver during co-infection with closely related nematode–bacterium partners.

A similar study in *Heterorhabditids* found differential fitness in *Photorhabdus* symbiont species that colonize a common host nematode, *Heterorhabditis downesi*, and prey on a common insect population. Traits important for both virulence and mutualism were compared in *Photorhabdus cinerea* and *Photorhabdus temperata* (Maher et al. 2021). *P. cinerea* had increased antibiotic activities against non-*Photorhabdus* bacteria and deterred insect cadaver invasion by other nematodes at a higher rate than *P.*

temperata (Maher et al. 2021). In plate choice experiments, *P. cinerea* was preferred by *H. downesi* nematodes, and in local head-to-head competition between the two strains in insects, *P. cinerea* displayed a significant colonization advantage in infective juveniles and colonized to a greater level in infective juveniles co-colonized by both species (Maher et al. 2021). Based on these studies, it is unclear what fitness advantages might be conveyed by *P. temperata* that allow it to persist in competition with *P. cinerea*, but it may be linked to the role of bioluminescence, as *P. temperata* was shown to be brighter than *P. cinerea* (Maher et al. 2021). The evolution of bacterial symbiont traits that impact their virulence and mutualism phenotypes, and how those traits might be selected for by hosts, is still an area of active exploration. However, it appears that selection for the maintenance of mutualism is stronger than other selection pressures (Morran et al. 2016). The number of culturable nematode and bacterial symbiont pairs, existing genomes, and high-confidence predictions of their phylogenetic relationships put EPNs and their bacterial symbionts in an extremely favorable position for studying the evolution of symbiont specificity down to the strain level.

Bacterial surface proteins mediating host interactions

Symbiotic partners often physically interact at a cellular level where surface-exposed molecules of one species come in contact with surface exposed molecules from the other in order to elicit some response. In both *X. nematophila* and *P. luminescens*, transposon mutagenesis screens revealed surface-localized proteins necessary for interactions with host nematodes (Heungens, Cowles, and Goodrich-Blair 2002; Somvanshi et al. 2010). In a screen for *X. nematophila* mutants defective in colonizing the infective juvenile intestine, five independent mutants each had the transposon insertion located in one region, termed Symbiosis Region 1 (SR1) (Heungens, Cowles, and Goodrich-Blair 2002). Each of the transposons were within one of three genes in the SR1 locus, and these were named *nilA*, *nilB*, and *nilC* to denote their role in nematode intestinal localization. The absence of SR1 from other *Xenorhabdus* genomes prompted investigations into its role in

determining *X. nematophila* specificity for associating with *S. carpocapsae*. Nonnative *Xenorhabdus* species that lacked the SR1 locus were unable to colonize *S. carpocapsae* infective juveniles. The SR1 locus was sufficient to confer upon these non-native *Xenorhabdus* species the ability to colonize *S. carpocapsae* infective juveniles, demonstrating that the SR1 locus is a specificity determinant that determines the nematode host range of the symbiont (Chaston et al. 2013; Cowles and Goodrich-Blair 2008). Furthermore, the SR1 locus also is necessary for species-specific colonization of the anterior intestinal caecum of developing juvenile and adult nematodes (Chaston et al. 2013). SR1 was the first discovered example of a single locus capable of expanding host range in an animal–bacterium mutualism. Each SR1 *nil* gene is individually important for achieving robust colonization of infective juvenile nematodes (Cowles and Goodrich-Blair 2008; Heungens, Cowles, and Goodrich-Blair 2002), but at the time they were discovered, none had homologs of known function in public databases (Heungens, Cowles, and Goodrich-Blair 2002). As such, elucidating their roles in promoting bacterial symbiont colonization requires extensive subsequent studies..

Surface proteins of the bacteria or host can act as lectins or adhesins that adhere to specific polysaccharides on the surface of the other organism. NilB-mediated colonization of the anterior intestinal caecum (AIC) is a potential site of adhesion and glycan involvement, as it involves an intimate association of symbiont with a nematode tissue (Chaston et al. 2013). In addition, within the infective juvenile receptacle, bacteria associate with a wheat-germ agglutinin-reactive substance, indicating the presence of N-acetyl glucosamine and/or N-acetyl neuraminic acid (Martens and Goodrich-Blair 2005). Therefore, it is tempting to speculate that the Nil proteins are involved in some sort of host recognition process or adhesion to molecules presented at these colonization sites. Future work on further understanding the roles of these molecular players, in particular how they function in specificity between the nematode and bacterium partner pairs and with which, if any, host molecules they interact, is an exciting direction of EPN symbiosis research.

Goals of this Dissertation:

Bioinformatic investigation of NilB homologs revealed that *Xenorhabdus* species universally encoded between 1 and 3 uncharacterized DUF560 family proteins, and the complement of DUF560 genes encoded by a *Xenorhabdus* species correlated with phylogenetically detected host switching events (Grossman, Mauer, et al. 2022). Since DUF560 proteins had already been demonstrated to facilitate host-microbe symbioses in *X. nematophila* and *N. meningitidis* (Heungens, Cowles, and Goodrich-Blair 2002; Hooda et al. 2016), this collection of diverse homologs presented a powerful tool to characterize how DUF560 proteins facilitated symbiosis. However, many of the DUF560 proteins found in *Xenorhabdus* differed from those previously studied in that they were not associated with surface lipoproteins and were instead predicted to transport soluble proteins.

The studies described in chapter 2 of this dissertation will describe how I utilized one of these *Xenorhabdus* DUF560 proteins, named Heme receptor protein B (HrpB), to demonstrate that DUF560 proteins could secrete soluble proteins in addition to lipoproteins (Grossman, Mauer, et al. 2022). In light of this discovery, I proposed that the DUF560 protein family is a novel bacterial secretion system and named it the type eleven secretion system (henceforth T11SS). Chapter 2 also details how I also leveraged sequence similarity networking and genomic co-occurrence to predict 851 T11SS/cargo pairs throughout Proteobacteria and to separate them into sub-clusters indicative of cargo type, lipoprotein or soluble protein. This chapter establishes the type eleven secretion system, illustrates how well conserved it is in Proteobacteria that exist in host mucosa, and builds a groundwork for the work presented in the subsequent chapters.

The studies in chapter 3 focus on the T11SS transporter, HrpB and its secreted cargo protein, hemophilin. Hemophilin is a soluble bacterial hemaphore that acts as a virulence factor in *Neisseria* (Hooda et al. 2016), a probiotic factor in *Haemophilus* (Atto et al. 2020), and a potential symbiosis factor in *Xenorhabdus*. Within chapter 3 I demonstrate that HrpB homologs in *Xenorhabdus*,

Haemophilus, and *Acinetobacter* also act to secrete their respective hemophilin homologs. I also demonstrate that this secretion is highly specific between a given T11SS and its cognate cargo protein. Furthermore, this specificity seems to be conveyed by a 6 β-strand hydrophilic β-barrel domain present in all known T11SS cargo proteins, and specificity can be partially transferred by exchanging these domains between cargo. Examining purified hemophilin's heme binding behavior *in vitro* revealed that these hemophores share a ligand binding handle domain, but differ in their affinities for metallated and unmetallated porphyrins. Examining hemophilin and its cognate T11SS within *X. nematophila* demonstrated the T11SS is required for this bacterium to cope with heme starvation, is transcriptionally upregulated by metal starvation, and is required for host nematode fitness when grown *in vitro*.

In chapter 4, I demonstrate that NilC is flipped from the periplasm to the cell surface by NilB, utilize a lectin library to characterize glycan interaction within the nematode receptacle, and use metabolomics and proteomics to explore the function of NilC (Grossman, Escobar, et al. 2022). My findings indicate that NilB-dependent translocation is relatively tightly regulated compared to other known T11SS and could only be observed when *nilB* expression was de-repressed by deletion of the gene which encodes its repressor protein, NilR. Additionally, I demonstrated that *Steinernema* species differ significantly in glycan content within the receptacle. The lectin wheat germ agglutinin (WGA) bound strongly within the receptacle of *Steinernema carpocapsae* and was capable of excluding colonization of that nematode when present, demonstrating a role for WGA reactive glycans in bacterial colonization. Addition of purified NilC had the opposite effect and increased bacterial colonization, suggesting that NilC is unlikely to be competing with WGA for glycan binding, but may instead be involved in liberating host-derived glycans. The multi-omics analysis of *nil* mutants revealed disruption of amino sugar metabolism leading to peptidoglycan, exopolysaccharide, and lipopolysaccharide biosynthesis. Finally, I demonstrate that purified NilC can bind peptidoglycan,

potentially indicating that periplasmically-oriented NilC participates in cell wall interactions when not surface exposed.

In chapter 5, I develop a controlled method for performing genome neighborhood co-occurrence analysis and utilize this to search for novel T11SS-dependent cargo proteins. To control for non-specific co-occurrences, I generated a database of random proteins that were biophysically similar to the proteins of interest and used this database to assign false discovery rates to all co-occurring motifs. This analysis demonstrated conserved co-occurrence of T11SS proteins within iron/metal homeostasis loci, single carbon metabolism loci, and mobile genetic regions. Additionally, this analysis resulted in 141 predicted cargo families including 10 distinct domain architectures, several of which had never been identified previously as a T11SS-dependent cargo. One of these families had large α helical repeat regions with sequence similarity to surface glycoproteins such as Plasmin sensitive protein (Pls) from the Gram-positive organism *Staphylococcus aureus*. To provide validation of the T11SS-dependent cargo predictions, I co-expressed one of these predicted T11SS/Pls pairs from *H. parahaemolyticus* and demonstrated that this homolog of Pls was secreted by its cognate T11SS. This expands the known functional range of T11SS proteins and, if this *H. parahaemolyticus* Pls homolog functions similarly to Pls from *S. aureus*, it may be another example of a T11SS dependent symbiosis factor.

In chapter 6, I compile and synthesize discoveries regarding the type eleven secretion system, its diverse subfamilies, the cargo proteins they have evolved secrete, and their conserved roles in maintaining symbioses. The availability of publicly available sequencing data and the diversity of T11SSs present within *Xenorhabdus* made it possible to discover this large and vital membrane transport system often encoded within human pathogens and used for the surface exposure of surface associated virulence factors. I have begun to systematically identify novel cargo proteins within host-associated and environmental isolates and as we learn more about the

molecular determinant of secretion specificity and localization of symbiosis factor, I hypothesize that there are many more that await discovery.

Chapter 2: A Widespread Bacterial Secretion System with Diverse Substrates

Authors: Alex S. Grossman, Terra J. Mauer, Katrina T. Forest, and Heidi Goodrich-Blair

Publication disclosure:

Previous versions of this chapter were published in mBio (Grossman, Mauer, et al. 2022) (https://doi.org/10.1128/mBio.01956-21) and BioRχiv (Mauer et al. 2020) (https://doi.org/10.1101/2020.01.20.912956). The authors have no competing interests that might be perceived to influence the results and/or discussion reported in this chapter. Alex Grossman, Terra Mauer, and Heidi Goodrich-Blair wrote the text of this chapter and composed the figures. Phylogenetic analysis and sequence similarity networking was performed by Alex Grossman. Annotation of environmental association and T11SS loci was performed by Terra Mauer. Cloning and dot blots were performed by Alex Grossman and Terra Mauer. Alex Grossman performed all secretion experiments in *E. coli*. Terra Mauer performed the flow cytometry experiment and colonization assays in *Steinernema* nematodes. Katrina Forest provided sustained intellectual contributions to the research design and edits to the text and figures.

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(Hooda et al. 2016; Hooda, Lai, and Moraes 2017). 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*. 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 (T11SS).

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

Introduction:

All plants and animals exist in association with bacterial symbionts that can contribute to nutrition, protection, development, and reproduction. These symbionts express surface and secreted

proteins that facilitate host interactions through a variety of functions, including acquisition of nutrients (Chu et al. 2010; Noinaj et al. 2010), interaction with host cells (Schweppe et al. 2015), and specificity in host-range and tissue-localization (Chavez-Dozal et al. 2014). Possibly due to the complexity of bacterial membranes and the breadth of biophysical characteristics of secreted proteins, there is no singular export pathway.

Bacteria have two broadly distributed export systems, including the inner membrane spanning Sec and twin arginine translocation (Tat) systems which are shared between Gram-positive and Gram-negative bacteria. Diderms require additional secretion systems for transport into and across the outer membrane. The type 1, 2, 5, 9, and 10 secretion systems work to transport substrates across that outer membrane, while types 3, 4, and 6 go one step further, moving effector proteins across the outer membrane and directly into another organism (Green and Mecsas 2016; Lasica et al. 2017; Palmer et al. 2021). Despite the many secretion systems described thus far, there are still a number of secreted proteins which are known to contribute to symbiosis but for which no transport system is known. Here we describe a machinery present throughout the phylum Proteobacteria that is responsible for secreting soluble proteins and lipoproteins from the periplasm across the outer membrane.

Recently, a mechanism of lipoprotein surface tethering was identified in the human pathogen *Neisseria meningitidis* and termed the Slam (Surface lipoprotein assembly modulator) machinery (Hooda et al. 2016; Hooda, Lai, and Moraes 2017). Slam proteins containing the β-barrel domain DUF560 (also termed SlipAM domain) (Mistry et al. 2021) are required for surface presentation of certain lipoproteins, including those that capture metal-carrying compounds used by hosts to sequester nutrients (Hooda et al. 2016). Two *N. meningitidis* Slam proteins have been characterized, each with distinct lipoprotein substrates. Slam activity also has been demonstrated for DUF560

representatives from pathogens *Pasteurella multocida, Moraxella catarrhalis,* and *Haemophilus influenzae* (Hooda et al. 2016; Hooda, Lai, and Moraes 2017). However, most Slam homologs have no bioinformatically predicted substrate thus far, and one study has found that *N. meningitidis* Slam1 can surface expose unprocessed factor H binding protein (fHbp) single nucleotide polymorphism (SNP) variants (da Silva et al. 2019). Thus, the full functional potential of DUF560 proteins is not yet known.

The DUF560 homolog NilB is a host-association and species-specificity factor in the nematode symbiont *Xenorhabdus nematophila*, a proteobacterium in the family *Morganellaceae* (Akhurst 1983; Chaston et al. 2011; Heungens, Cowles, and Goodrich-Blair 2002). A screen for *X. nematophila* mutants defective in colonizing *Steinernema carpocapsae* intestines revealed the Nematode intestinal localization locus (*nil*) (Bhasin, Chaston, and Goodrich-Blair 2012; Heungens, Cowles, and Goodrich-Blair 2002). The Nil locus contains the genes *nilB* and *nilC*, each of which is independently necessary for colonization of nematodes. Biochemical and bioinformatic analyses have established that NilC is an outer-membrane-associated lipoprotein and NilB is an outer-membrane β-barrel protein in the DUF560 protein family with a ~140 amino acid periplasmic N-terminal domain that contains tetratricopeptide repeats (Bhasin, Chaston, and Goodrich-Blair 2012; Blatch and Lässle 1999; Cowles and Goodrich-Blair 2004, 2008).

To begin to understand the range of functions of DUF560 proteins, we assessed their ecological distribution, genomic context, and relatedness. We experimentally examined the *X. nematophila* DUF560 homolog HrpB, which is not predicted to interact with a lipoprotein substrate. Finally, to better understand the potential role of DUF560 proteins in host-symbiont interactions we analyzed DUF560 distribution among symbiotic *Xenorhabdus*. Our data demonstrate that the activities of the DUF560 family extend beyond lipoprotein surface presentation and constitute a new

type XI bacterial secretion system (T11SS) that is capable of acting on either membrane anchored or soluble protein (Figure 2.1) (D'Enfert, Chapon, and Pugsley 1987).

Results:

T11SS cluster according to environment

Using homology to NilB or Slam proteins, previous work identified a wide distribution of DUF560 proteins within mucosal associated bacteria (Bhasin, Chaston, and Goodrich-Blair 2012; Heungens, Cowles, and Goodrich-Blair 2002; Hooda, Lai, and Moraes 2017). To quantifiably delineate subfamilies within the T11SS we generated a sequence similarity network (SSN) using the Enzyme Function Initiative toolset (EFI) (Gerlt 2017; Zallot, Oberg, and Gerlt 2018, 2019) and annotated it to highlight environmental source or taxonomic grouping of the isolates containing DUF560 homologs (Fig 2.2; TableS1_Clusters1-4,6-11WithEnvTax.xlsx). In this network analysis, protein sequences with a high identity (40% or greater) were gathered into data points called nodes and connected together with edges based on sequence similarity. Using all homologs in Interpro 73 and UniProt 2019-02, we identified 10 major clusters of T11SS proteins. Cluster 1 was chosen for in-depth analysis since it contained the majority of nodes in the network (62.4%) and could be visually divided into three subclusters (1A, 1B, and 1C) using force directed node placement (Fig. 2.3). The remaining clusters displayed a preponderance of water and soil associated organisms and contained no characterized proteins. Consistent with our previous observations (Bhasin, Chaston, and Goodrich-Blair 2012), the majority of cluster 1 nodes (75%) comprise sequences from various animal-associated isolates, while another 20% contain sequences from marine, freshwater, soil, or built-environment isolates.

The division of nodes among the three subclusters more strongly reflected environmental origin than bacterial taxonomy (Fig. 2.3A). Subcluster 1A almost exclusively comprises animal-

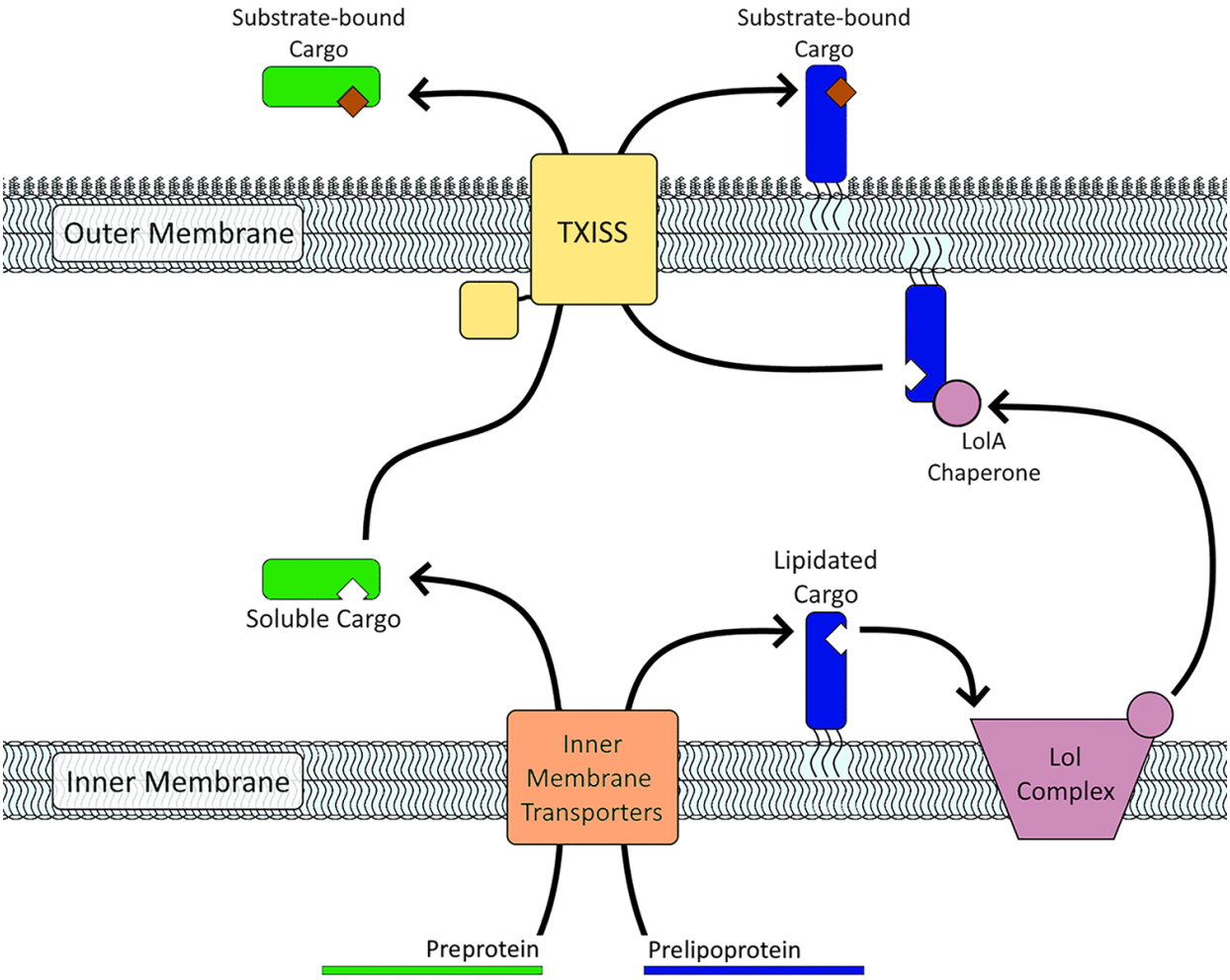


Figure 2.1. **Conceptual model of the type 11 secretion system**. T11SS outer membrane proteins (yellow) of the DUF560 family are necessary and sufficient to secrete lipidated (blue) and soluble (green) cargo proteins across the outer membrane to the extracellular milieu. T11SS-dependent cargo proteins are exported into the periplasm via inner membrane transporters (orange). Once in the periplasm, lipidated T11SS cargo proteins, such as TbpB, are expected to be chaperoned across the hydrophilic periplasm by the localization of lipoproteins (Lol) complex. It is currently unknown whether periplasmic soluble T11SS cargo proteins have dedicated molecular chaperones to reach the outer membrane. Once secreted by T11SS, cargo proteins can bind to their specific substrates.

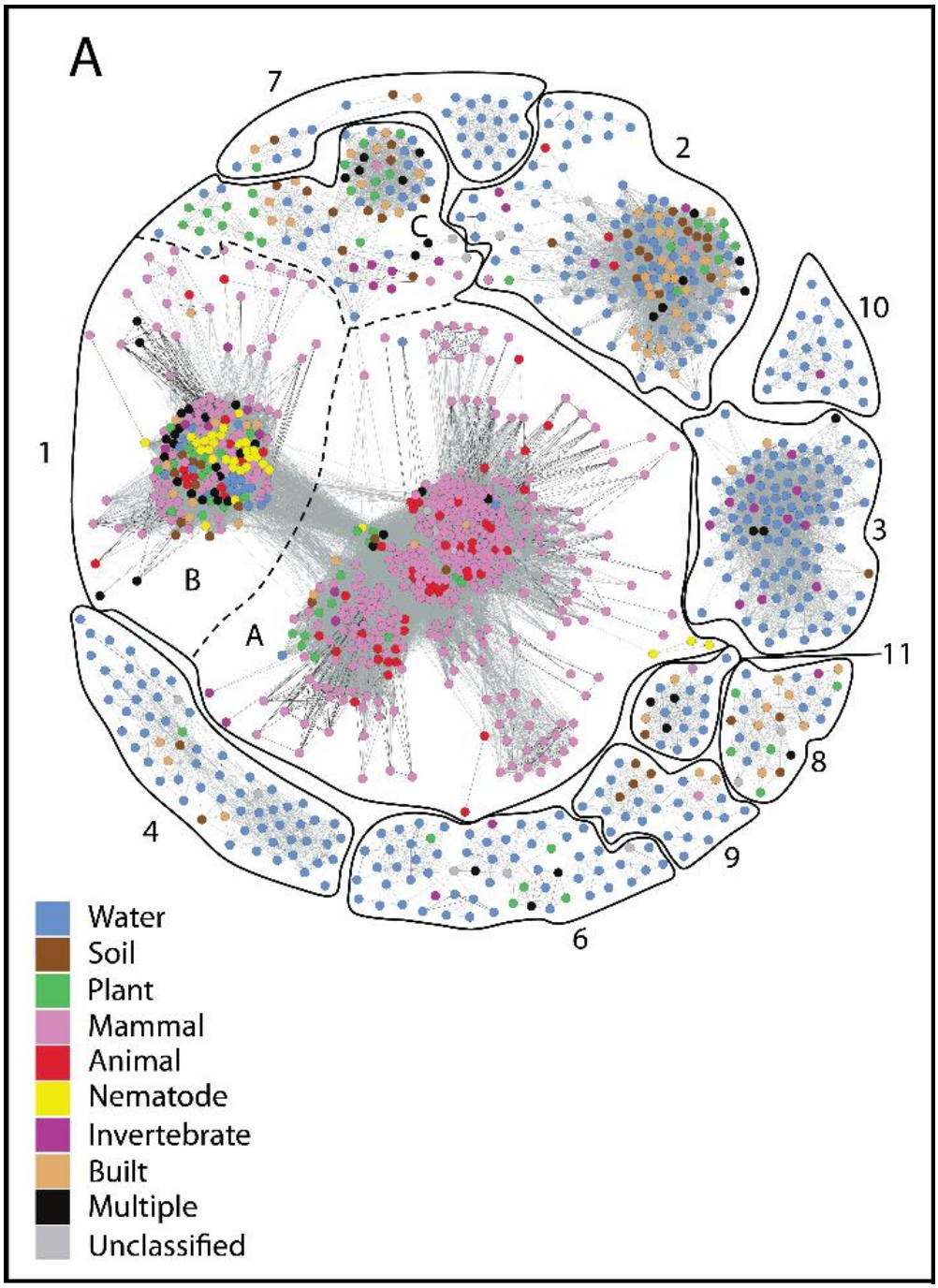


Figure 2.2. **Sequence similarity network of all DUF560 proteins.** SSN of DUF560 homologs generated by EFI-EST as accessed on 4.24.19. Edges were drawn using an alignment score of 38, and any sequences which shared ≥ 40% identity were placed in a single node to allow the separation of clusters. Each 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. Each node was color coded to show the isolates' environmental origin(s) **(A)** and taxonomic class **(B)**.

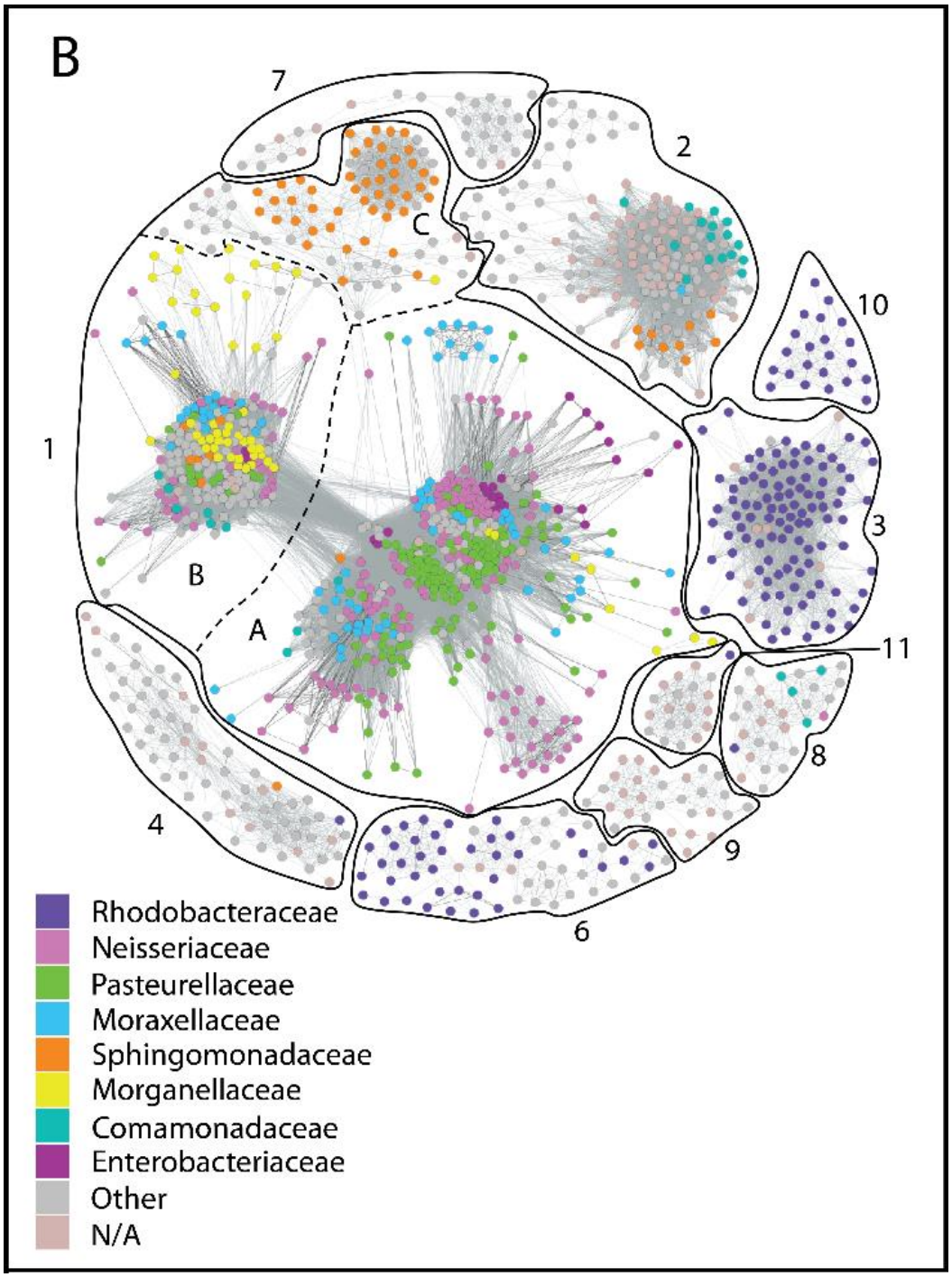


Figure 2.2. (continued)

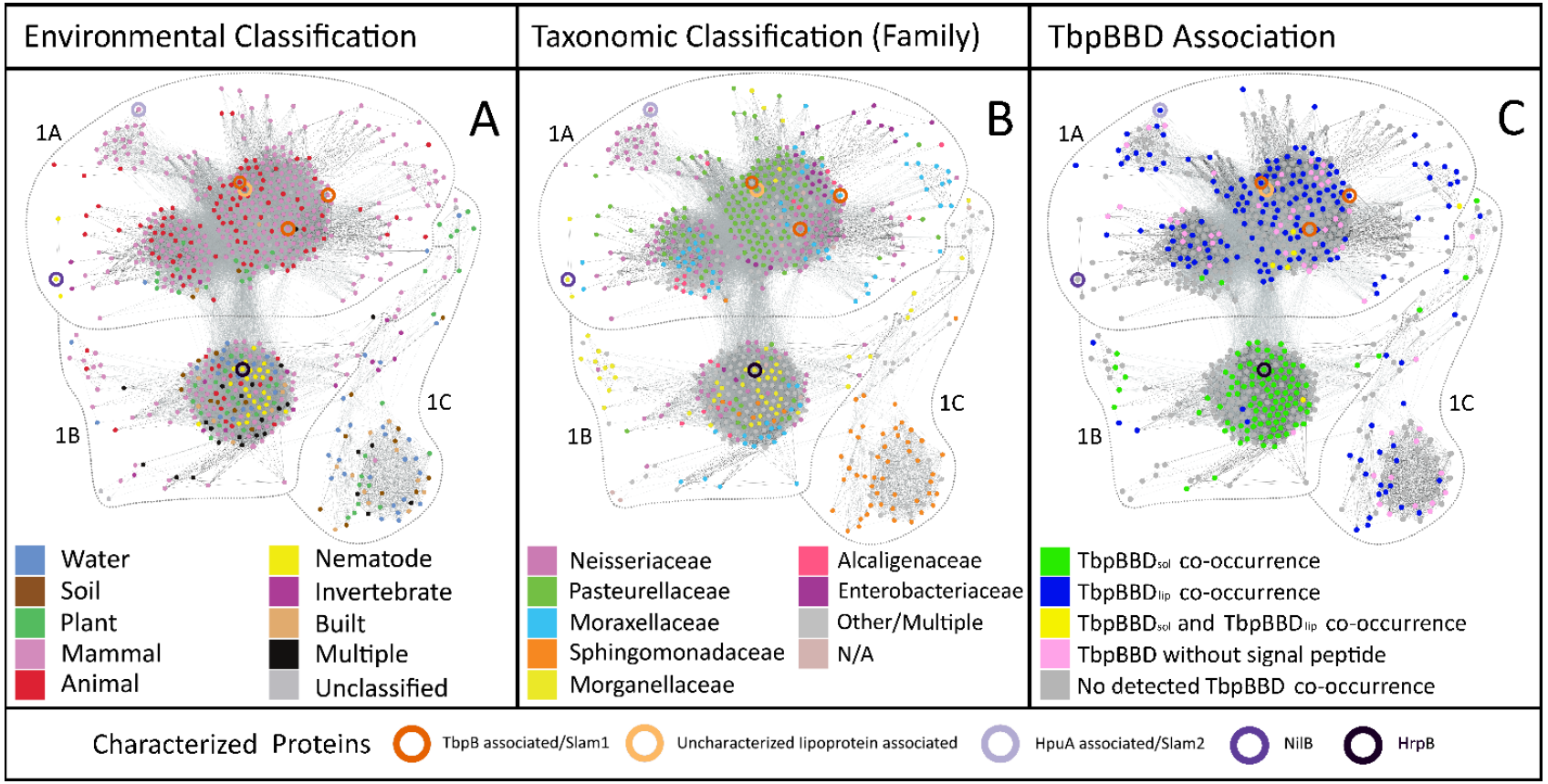


Figure 2.3. **Cluster 1 of the TXISS sequence similarity network (SSN) forms subclusters according to environment of isolation and signal sequence of predicted cargo**. All nodes from cluster 1 of the TXISS SSN were positioned using the Fruchterman-Reingold algorithm. The resulting graph was annotated 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.

associated bacteria and contains all previously characterized T11SS, which separate according to predicted substrate when analyzed with higher stringency (Fig. 2.4). Subclusters 1B and 1C have no previously characterized representatives. Subcluster 1B contains a mixture of sequences from host-associated and free-living bacteria, while subcluster 1C contains sequences largely from environmentally isolated *Sphingomonadaceae*. Subclusters 1A and 1B do not correlate well with taxonomy of the isolates (Fig. 2.3B). For example, cluster 1 contains 81 nodes with *Moraxellaceae* sequences. Of these, 79% are predominantly animal-associated and 12.3% are predominantly environmental isolates. The animal-associated isolates are enriched in subcluster 1A and environmental-isolates are enriched in subcluster 1B. These data demonstrate a correlation with lifestyle (e.g., free-living vs host associated) as opposed to taxonomy and suggest that subclusters indicate divergent molecular functions.

T11SS 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 transferrin binding protein B (TbpB), lactoferrin binding protein B (LbpB), hemoglobin/haptoglobin binding protein A (HpuA), and fHbp, which are frequently encoded nearby (Hooda et al. 2016; Hooda, Lai, and Moraes 2017). Yet, the lipid tail is not essential for Slam-dependent surface exposure of a target (Hooda et al. 2016; Ostberg et al. 2013; da Silva et al. 2019). This result prompted us to consider whether co-occurrence with lipoproteins is a constant throughout cluster 1. We used the EFI Genome-Neighborhood Tool to assay the genomic context of each subcluster (Gerlt 2017; Gerlt et al. 2015). 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 (El-Gebali et al. 2019; Hooda, Lai, and Moraes 2017).

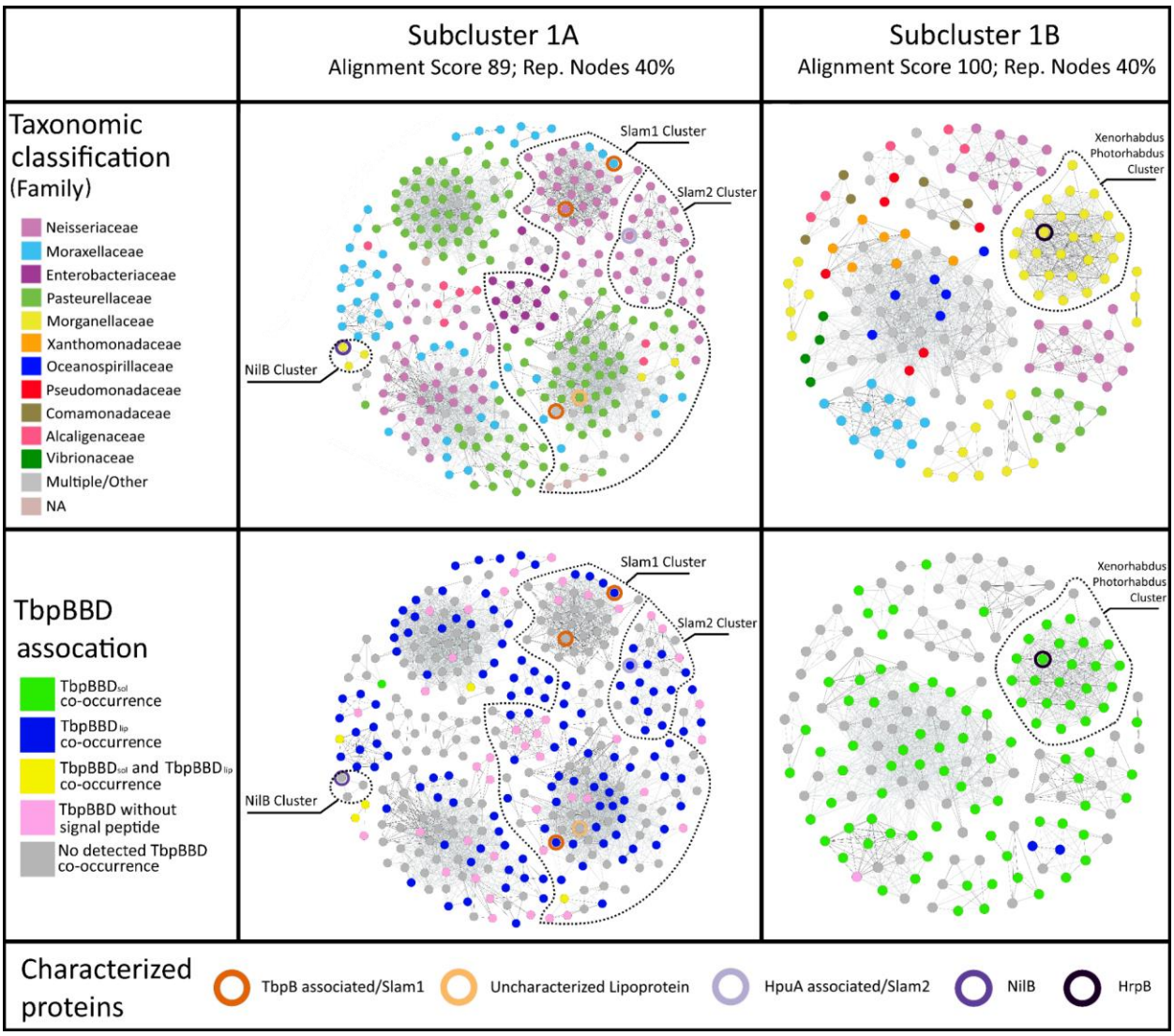


Figure 2.4. **Increased stringency separates subcluster 1A and 1B into functional groups.** A series of stringent EFI-EST sequence similarity networks highlights detail in cluster 1 of the DUF560 homologs. Edge darkness demonstrates similarity. Node positioning was optimized using the Fruchterman-Reingold algorithm. 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.

Given the prevalence of TbpBBD domains in the genome neighborhoods of DUF560 genes, and their known occurrence in lipoproteins surface exposed by Slams we examined their gene structures to detect potential patterns. Using a combination of genome-neighborhood-network data (Gerlt 2017), Rapid ORF Description & Evaluation Online (RODEO) data (Tietz et al. 2017), and manual annotation we analyzed all TbpBBD domain proteins co-inherited with sequences present in cluster 1 (TableS2_DUF560TbpBBDCo-occurrence.xlsx). The majority (75.1%) of TbpBBD-bearing proteins associated with subcluster 1A have a signal peptidase 2 (SPII) signal peptide, indicating lipidation, and two TbpBBD domains, similar to TbpB in *N. meningitidis* (Hooda, Lai, and Moraes 2017). These are referred to here as $TbpBBD_{lip}$ (lipidated TbpBBD). In contrast, the TbpBBD-bearing proteins associated with subcluster 1B are almost exclusively predicted to be soluble proteins (97.8%) with signal peptidase 1 (SPI) signal peptides and a single TbpBBD domain, similar to hemophilin in *Haemophilus haemolyticus*. Hereafter we refer to this class of proteins as $TbpBBD_{sol}$ (for soluble TbpBBD) (Fig. 2.3C; Fig. 2.5B). $TbpBBD_{lip}$ and $TbpBBD_{sol}$ proteins are predicted to be translocated across the inner membrane through the Sec secretion pathway. However, at least some $TbpBBD_{sol}$ proteins have similarities to *Escherichia coli* OmpA and DsbA, which can be secreted through either the Sec or Tat secretion pathways (Fig. 2.5A), leaving open the possibility that some T11SS-1B $TbpBBD_{sol}$ proteins are conditionally Tat secreted (Qi and Tracy 2017). Biochemical and structural evidence support the conclusion that hemophilin is a soluble secreted protein that binds free heme and facilitates heme uptake into the cell (Latham et al. 2020). Three-dimensional homology modeling (Phyre^2) was used to visualize potential structural similarities between hemophilin and several $TbpBBD_{sol}$ proteins : a previously described homolog X. nematophila XNC1_0075 and two of its most closely related homologs, *Providencia rettgeri* PROVRETT_08181 and *Proteus mirabilis* WP_134940027.1 (Altschul et al. 1997; Bhasin, Chaston, and Goodrich-Blair 2012; Kelley et al. 2015; Kelley and Sternberg 2009) (Fig. 2.6). No high

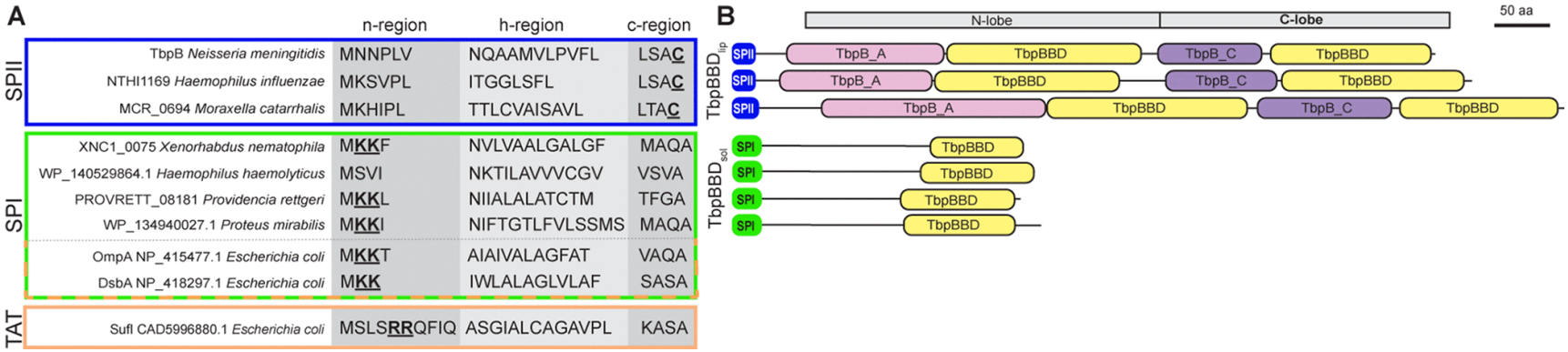


Figure 2.5. **Examples of TbpBBDlip and TbpBBDsol signal peptides and domain architectures**. T11SS-1A-associated proteins from *N. meningitidis*, *H. influenzae*, and *M. catarrhalis* have signal peptidase 2 (SPII; blue) signal sequences, while T11SS-1B associated proteins from *X. nematophila*, *H. hemolyticus*, *P. rettgeri*, and *P. mirabilis* have signal peptidase 1 (SPI; green) signal sequences. (A) SPII and SPI signal sequences, comprising n-, h-, and c-regions of representative T11SS-1A (blue) and T11SS-1B (green, above dashed line) cargo proteins and two *E. coli* proteins (OmpA and DsbA) that can be secreted through either the Sec or Tat machinery (below dashed line, green-orange border), are compared to a canonical Tat secretion signal (orange) found in the *E. coli* protein SufI. Underlined amino acids highlight conserved features of signal sequences, including the acylated cysteine of lipoproteins, the twin arginine motif of Tat signal peptides, and n-region twin lysines that are permissive for Tat secretion. (B) The schematic diagram shows features found in select TbpBBD domain-containing proteins predicted to be exported by T11SS mechanisms. T11SS-1A-associated cargo proteins shown have SPII (blue), N-lobe and C-lobe handle domains TbpB_A (pink) and TbpB_C (purple), respectively, and two TbpBBD domains (yellow). T11SS-1B-associated cargo proteins shown have SPI (green), lack annotated handle domains, and have a single TbpBBD domain (yellow).

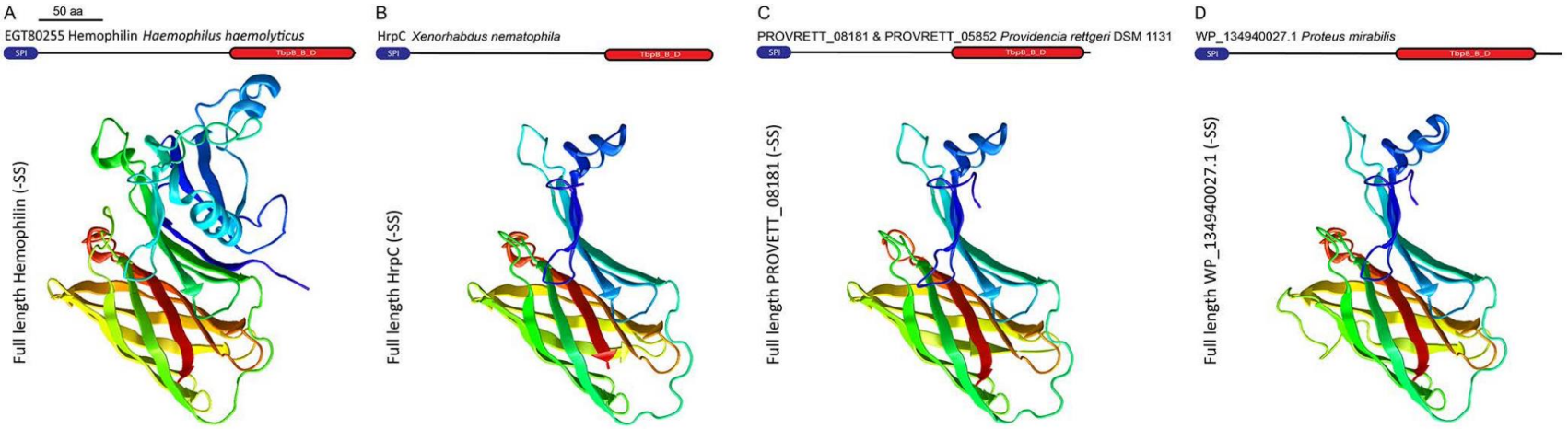


Figure 2.6. **Phyre[2] models of select TbpBBD$_{sol}$ proteins.** TbpBBD$_{sol}$ proteins, lacking the signal sequence (-SS), from *X. nematophila* (HrpC), *P. rettgeri* (PROVRETT_08181 and 05852), and *P. mirabilis* (WP_134940027.1) were queried through the Phyre[2] Protein Homology/analogy Recognition Engine v. 2.0 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). 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

confidence models were found for the first ~50 residues, which correspond to the position of the variable N-terminal handle domains of $TbpBBD_{lip}$ proteins. However, the structures of the central regions and the C-terminal TbpBBD β-barrel domains were predicted with high confidence (>99%) based on hemophilin. In light of sequence and structural level similarities, we predict two functional T11SS subfamilies; T11SS-1A members flip $TbpBBD_{lip}$ substrates across the outer membrane, and T11SS-1B members secrete $TbpBBD_{sol}$ substrates into the extracellular milieu.

T11SS-1B activity reconstructed *in vivo*

To experimentally evaluate our prediction that the T11SS-1B can secrete $TbpBBD_{sol}$ substrates, we investigated the heme receptor protein (Hrp) locus of *X. nematophila*. This locus, which is conserved across the *Xenorhabdus* genus, consists of genes predicted to encode TonB, a TonB-dependent heme receptor named HrpA (XNC1_0073), the T11SS-1B homolog HrpB (XNC1_0074), and its predicted $TbpBBD_{sol}$ substrate HrpC (XNC1_0075) (Bhasin, Chaston, and Goodrich-Blair 2012), a homolog of the heme-binding protein hemophilin (55% similar 39% identical) (Latham et al. 2020). Specifically, we sought to test whether the T11SS-1B homolog HrpB mediates secretion of the putative heme-binding protein HrpC. Immunotagged HrpC was expressed with or without immunotagged HrpB in *E. coli*. Whole-cell and culture supernatant fractions were separated analyzed by immunoblotting with anti-FLAG antibody (Einhauer and Jungbauer 2001). We found that in the presence of HrpB-FLAG the levels of HrpC-FLAG detected in the supernatant increased 9.9-fold at 1-h post-induction and 17.0-fold at 2.5-h post-induction (Fig. 2.7A). Whole cell lysates demonstrated equivalent expression of HrpC-FLAG in both treatments (Fig. 2.7B). A trivial explanation for these data could be that HrpB expression leads to cell lysis. The supernatant protein profile revealed by Coomassie blue staining of sodium dodecyl sulfate-polyacrylamide gels (SDS-PAGs) did not indicate cell lysis or nonspecific protein secretion

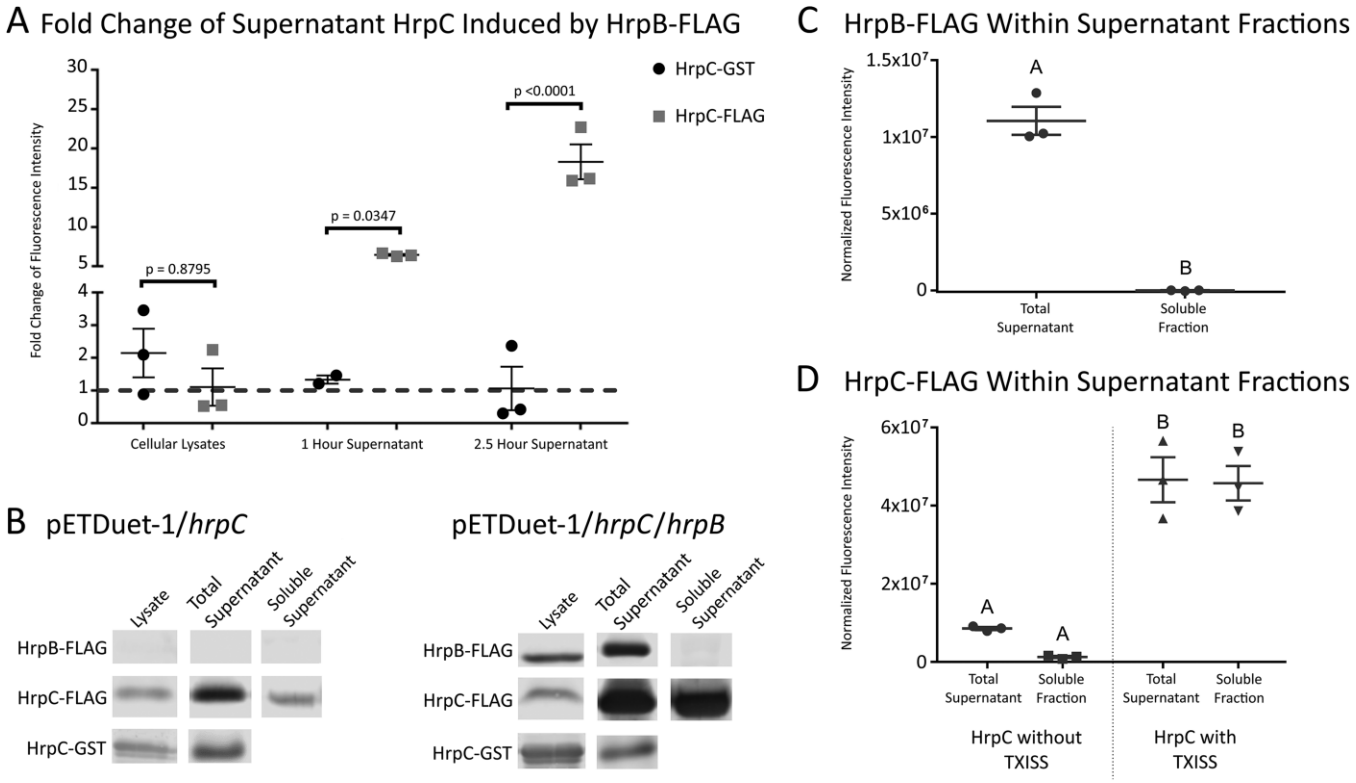


Figure 2.7. **HrpB increases secretion of HrpC**. (A) Demonstration of HrpB-dependent secretion. HrpC was detected and quantified as fluorescence intensity observed in immunoblots with anti-tag (FLAG or GST) antibodies. Fold change values shown are the fluorescence intensity of HrpC co-expressed with HrpB, divided by the intensity of HrpC expressed alone as observed from immunoblots. Each dot represents fold change derived from a distinct biological replicate pair and standard error is shown. At both assayed time points, the total supernatant concentration of HrpC-FLAG was increased by co-expression with FLAG-HrpB compared to the HrpC-GST treatment which was seemingly unaffected by the presence of FLAG-HrpB (unpaired *t* test). (B) Representative immunoblots comparing FLAG-HrpB, HrpC-FLAG, and HrpC-GST in cellular and supernatant fractions. (C) Fluorescence intensity of HrpB in supernatant fractions normalized to the Coomassie blue loading control. Each dot represents a distinct technical replicate. HrpB is not present in the soluble fraction, suggesting that it is likely in OMVs (unpaired *t* test). (D) Fluorescence intensity of HrpC-FLAG in supernatant fractions normalized to the Coomassie blue loading control. Each dot represents a distinct technical replicate pair. The T11SS-secreted HrpC is mostly located in the soluble fraction (Tukey's HSD test).

(Fig. 2.7C, 2.8). To further rule out this possibility, we created an HrpC−glutathione S-transferase (GST) fusion protein and co-expressed it with FLAG-HrpB. Previous observations had demonstrated that GST-fused TbpB is not surface exposed by Slam1(Hooda et al. 2016). Consistent with this finding, the levels of HrpC-GST in a culture supernatant were unaffected by co-expression with FLAG-HrpB, supporting the conclusion that expression of the outer membrane protein HrpB does not cause bacterial lysis (Fig. 2.7) and moreover demonstrating that HrpB cannot secrete its substrate when HrpC is fused to a 26-kDa protein. Consistent with the fact that in the absence of Slam1, several strains of *E. coli* surface expose a fraction of the substrate fHbp, some HrpC-FLAG reached the supernatant in the absence of HrpB-FLAG.

In the absence of Slam1, a fraction of the cargo protein fHbp is surface exposed when expressed in *E. coli* BL21(DE3) (Fantappiè et al. 2017; Konar et al. 2015). Similarly, we found that some HrpC-FLAG and HrpC-GST reached the supernatant in the absence of FLAG-HrpB. Furthermore, FLAG-HrpB was unexpectedly detected in the supernatant fraction (Fig. 2.7B). To distinguish soluble supernatant proteins from those that may be associated with insoluble membrane components (e.g., outer membrane vesicles [OMVs]), we depleted insoluble components of the supernatant via ultracentrifugation and tested the clarified soluble fraction for Hrp proteins. Effectively all HrpB in the supernatant was removed with ultracentrifugation, suggesting it is localized in OMVs or other insoluble membrane fragments (Fig. 2.7C) (Kohl et al. 2018). Also, in the absence of HrpB co-expression, the levels of HrpC detected in the supernatant were reduced upon removal of insoluble material, while supernatant levels of HrpC detected after co-expression with HrpB were not affected by removal of insoluble material (Fig. 2.7D). We conclude that HrpB is membrane bound and that it secretes soluble HrpC protein into the extracellular milieu.

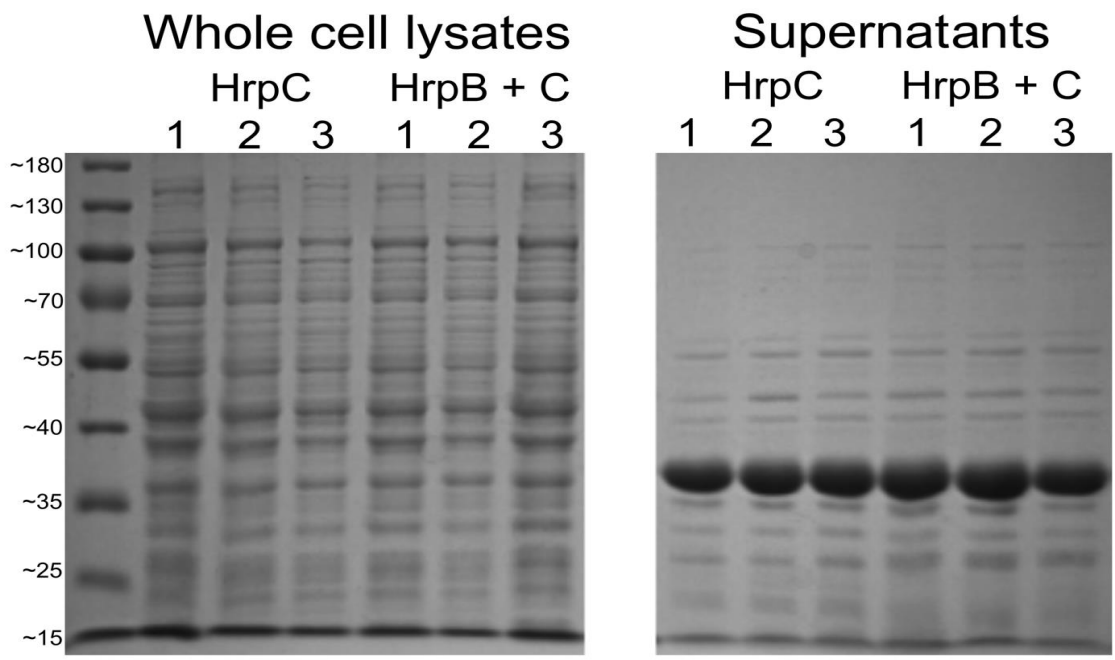


Figure 2.8. **Supernatant samples have a less complex protein profile than whole cells.** Coomassie stained SDS-PAGE gels showing side by side comparison of protein banding in whole cell lysates and supernatant samples of *E. coli* expressing the indicated proteins. Units on the left side are in kilodaltons. Expression of HrpB does not cause cell lysis or nonspecific protein export; if it did, the supernatant and lysate banding patterns would be similar to each other for samples from HrpB-expressing cells. 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.

Host environment drives T11SS 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 T11SS 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 (Chaston et al. 2013; Kämpfer et al. 2017; M.-M. Lee and Stock 2010; Spiridonov et al. 2004) (Fig. 2.9). All 46 *Xenorhabdus* genomes on the Magnifying Genome (MaGe) platform (Vallenet et al. 2017) encode between one and three T11SS, with five distinct homologs represented across the genus (one T11SS-1A and four T11SS-1B) (McGinnis and Madden 2004). Each unique combination of T11SS homologs was assigned one of six classes (A-F) as depicted in Fig. 2.10 and TableS3_XenorhabdusDUF560Classes.xlsx. To visualize correlations between T11SS class and host identity, we constructed Maximum Likelihood and Bayesian phylogenetic trees for *Xenorhabdus* and *Steinernema*. *Xenorhabdus* trees were generated with whole genome data while *Steinernema* trees used five loci as available (Fig. 2.11; TableS4_CombinedLociTable). Aligning the Maximum likelihood phylogenies reveals that the T11SS complement of a *Xenorhabdus* species is more predictive of nematode host than phylogenetic position. This alignment of the phylogenetic placement of a given *Steinernema* host with the T11SS complement of the symbiont provides insights into the nematode internal environment experienced by the symbiont. For example, most class B/C *Xenorhabdus* with two HrpB paralogs at the Hrp locus are symbionts of nematodes within the phylogenetic Clade IV, suggesting that these nematodes present a distinctive environment in which an additional Hrp locus is beneficial. Class D *Xenorhabdus* with an HrpB paralog encoded adjacent to genes predicted to encode a cobalt ABC transporter, are solely symbionts of Clade V nematodes. *X. innexi* and *X. stockiae* have seemingly diverged from this

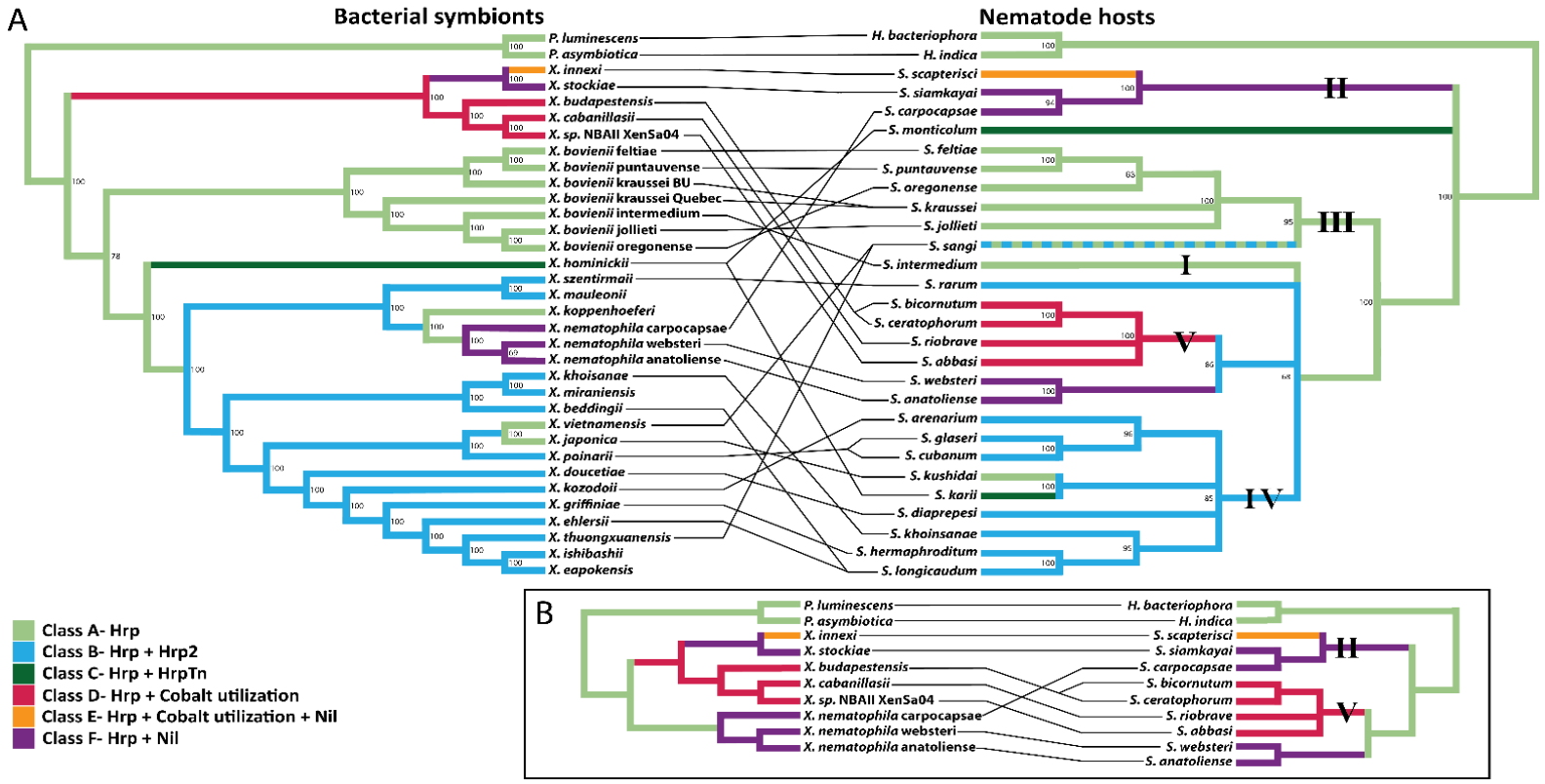


Figure 2.9. **Cladograms of *Xenorhabdus* and *Steinernema* color coded according to *Xenorhabdus* DUF560 Class.** Co-phylogeny of nematode species and their colonizing bacteria A) across the Steinernema genus or B) with a focus on specific clades. Numbers on branches indicate bootstrap support values. 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).

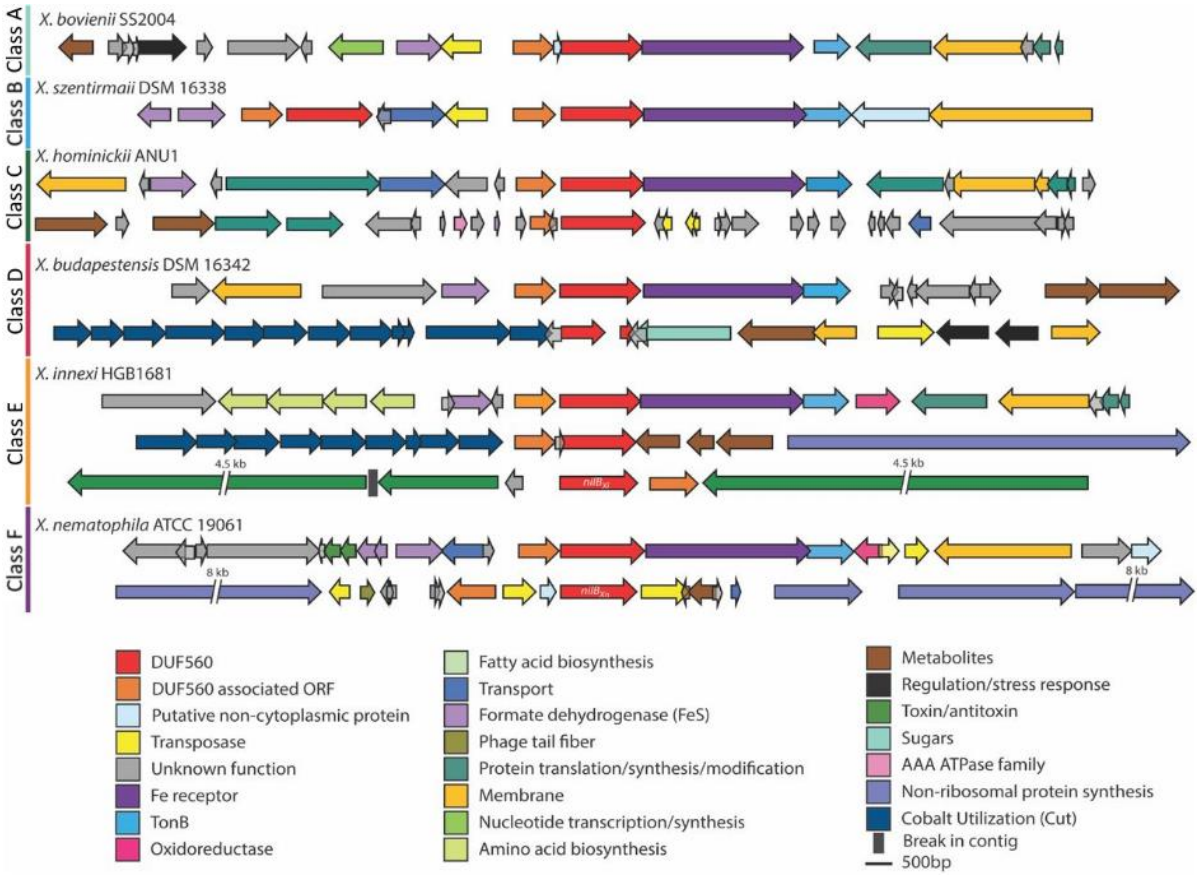


Figure 2.10. **Representative genomes of *Xenorhabdus* DUF560 Classes A-F.** Schematic diagrams of *Xenorhabdus* T11SS loci representing each of the six classes defined in the text. One species from each of the classes was selected for presentation. Box arrows represent open reading frames (ORFs), which are color coded according to predicted annotated function as indicated by the legend. The DUF560 homolog is shown in red and the predicted T11SS 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.

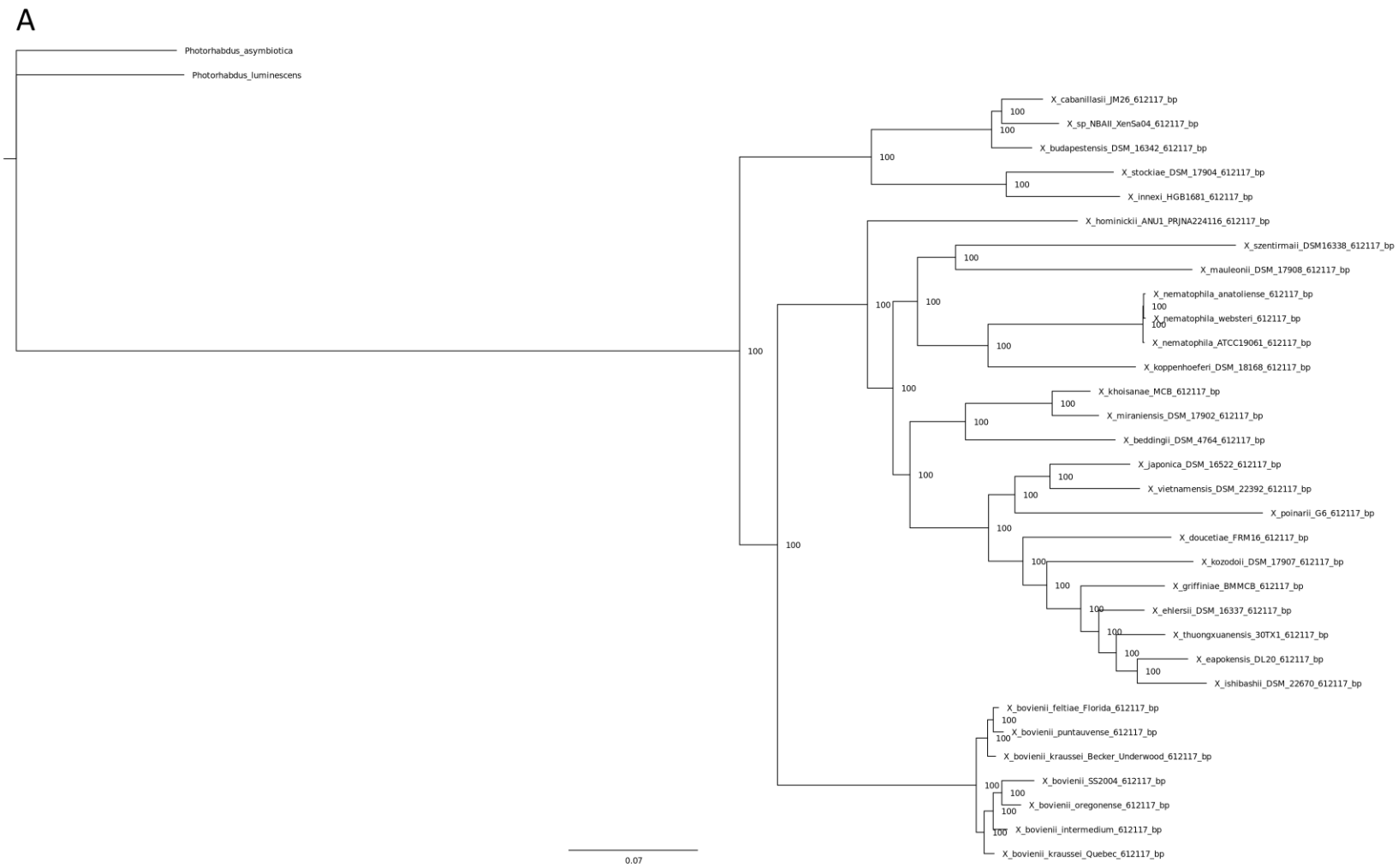


Figure 2.11. **Bayesian posterior probability phylogenies**. A) Phylogram of select *Xenorhabdus* bacteria, based on concatenations of 665 conserved core genes. Numbers indicate posterior probability values. Distances indicate substitutions per base pair. B) Bayesian phylogeny of select entomopathogenic 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 TableS4_CombinedLociTable.

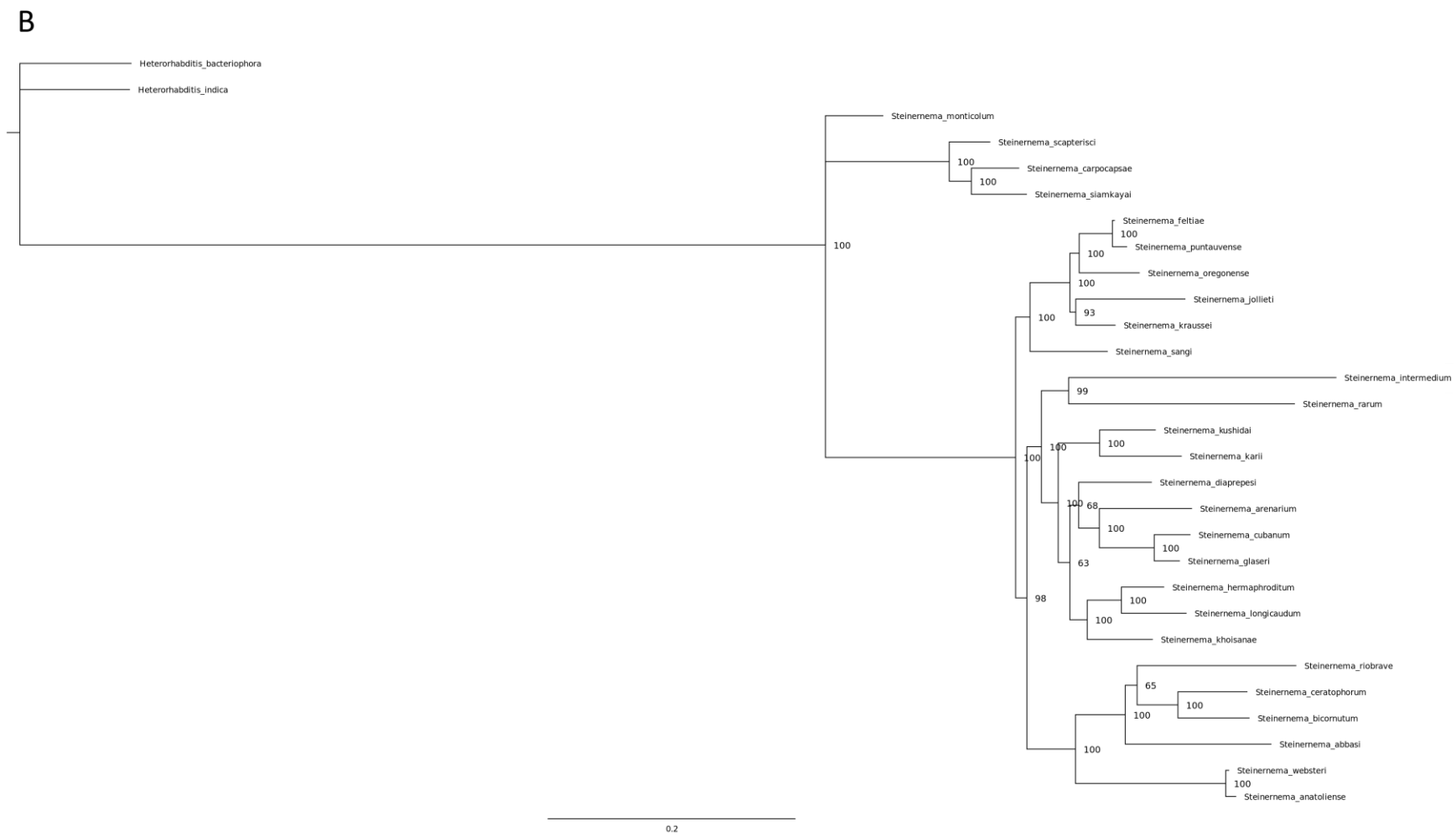


Figure 2.11. (continued)

lineage through acquisition of a *nilB* homolog and switching to hosts within Clade II. Similarly, *X. nematophila* independently gained *nilB* and switched into a Clade II host (Fig. 2.9B). These acquisitions, alongside the varied genomic contexts of *nilB*/*nilC* pairs, are consistent with previous suggestions that the Nil locus was horizontally acquired among *Xenorhabdus* (Cowles and Goodrich-Blair 2008).

The T11SS NilB and the lipoprotein NilC are necessary for *X. nematophila* to colonize the Clade II nematode *S. carpocapsae* (Cowles and Goodrich-Blair 2008; Heungens, Cowles, and Goodrich-Blair 2002). However, *X. nematophila* also colonizes two nematodes that are phylogenetically separate from Clade II, *S. anatoliense* and *S. websteri* (Fig. 2.9). Our hypothesis that T11SS are involved in host-environment adaptations leads to the prediction that *X. nematophila* will require the Nil locus to colonize these nematodes. To test this hypothesis, bacteria-free *S. anatoliense*, *S. websteri*, and *S. carpocapsae* eggs (Murfin, Chaston, and Goodrich-Blair 2012) were exposed to an *X. nematophila* ATCC19061 ⊿*nil* mutant and a *nil* complemented strain (Bhasin, Chaston, and Goodrich-Blair 2012). Consistent with our prediction, the ⊿*nil* mutant was deficient in infective juvenile colonization in all three nematode species (Fig. 2.12), demonstrating that *nil* genes are necessary for infective juvenile colonization of *S. anatoliense* and *S. websteri* and supporting our hypothesis that T11SS promote adaptation to host environments.

Discussion:

Bacteria have evolved specialized secretion systems for delivery of effectors that facilitate the host-associated lifestyle. Knowledge of cargo protein identities and sorting processes facilitates predictions from genomic information of bacterial secretome composition, regulation, and localization. Despite the diverse secretion systems now recognized, more remain to be identified based on the fact that some proteins predicted to be secreted lack known secretion

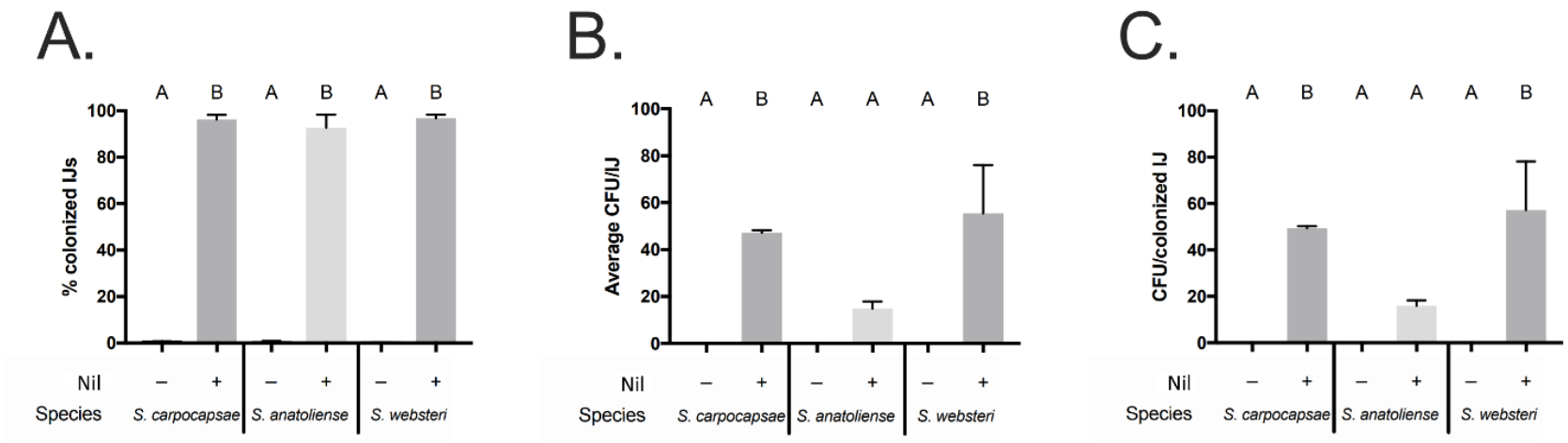


Figure 2.12. **The Nil locus is necessary for colonization of *S. anatoliense*, *S. carpocapsae*, and *S. websteri* infective juveniles.** Bacteria free *S. carpocapsae*, *S. anatoliense*, and *S. websteri* were exposed to GFP expressing *X. nematophila* lacking or bearing the *nil* locus. The resulting 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.

pathways (Bendtsen et al. 2005). The type 10 secretory pathway was described in 2020 shedding light on long-standing mysteries surrounding the dependence on a muramidase for secretion of typhoid toxin across the cell wall (Palmer et al. 2021). Enabled by availability of genomic data from myriad environments and ever improving bioinformatic visualization tools, we have presented data that the T11SS is a broadly distributed molecular vehicle for moving proteins across the Gram-negative outer membrane. Our network analysis has revealed functionally relevant T11SS clusters, with cluster 1 members having a conserved role in host-microbe interactions.

The DUF560 (domain of unknown function) family presence in animal-associated bacteria was first recognized when it was noted that the host colonization factor NilB has homologs in several human pathogens (Heungens, Cowles, and Goodrich-Blair 2002). This observation was strengthened by subsequent demonstration that Slam1 and Slam2 from these bacteria facilitate surface presentation of host metal acquisition proteins (Hooda et al. 2016; Hooda, Lai, and Moraes 2017). Using the *Steinernema-Xenorhabdus* symbiosis, here we provide evidence that the composition of T11SS in a bacterial symbiont genome correlates with host organism. During the evolutionary history of *Xenorhabdus*, the gain or loss of T11SS loci correlated with host switching events (Fig. 2.9). These data indicate that T11SS activity contributes to bacterial adaptation to new host environments, which is particularly relevant given the varied distribution of T11SS homologs among human pathogens. For instance, our work enables categorization of T11SS among animal-associated *Neisseria* species, including the human pathogens *Neisseria meningitidis* and *Neisseria gonorrhoeae* (Cohen, Powderly, and Opal 2017). *Neisseria* strains can encode up to 6 T11SS paralogs. *N. meningitidis* MC58 has two functionally characterized T11SS, Slam1 and Slam2. Our network analysis indicates *N. meningitidis* encodes a third Slam, NMB1466/NP_274965, that also falls within cluster 1A. *N. gonorrhoeae* T11SS are represented in more nodes than *N. meningitidis*, notably occupying 19.3% of nodes in subcluster 1B (Fig. 2.3). The cluster framework we present here indicates that *Neisseria* use

T11SS to secrete both lipoproteins and soluble cargo and that variation in T11SS composition among *Neisseria* strains may be predictive of host or tissue association phenotypes.

While DUF560 proteins were originally thought to represent a mechanism for lipoprotein surface exposure (Hooda et al. 2016), recent studies have expanded that functional range to include peripheral membrane proteins (da Silva et al. 2019) and now soluble secreted proteins, establishing the function of DUF560 outer membrane proteins in secretion of varied substrates. We predict that further study will uncover even more chemically diverse cargo for the distinct classes of T11SS, revealed through our network analysis. Here we focused on a single cluster of a DUF560 sequence similarity network (SSN). The remaining nine clusters likely represent diverse subfamilies responsible for transporting as-yet-unknown cargo. The lipoprotein cargo proteins for which structural conformations are known, share a C-terminal eight-stranded β-barrel (Hooda, Lai, and Moraes 2017). Our discovery that HrpC is a cargo protein for T11SS-1B HrpB strengthens the concept that this barrel is an important characteristic of T11SS cargo; HrpC is a homolog of the *H. haemolyticus* hemophilin, the structure of which likewise adopts a C-terminal eight-stranded β-barrel (Latham et al. 2020). These data support the concept that T11SS cargo have bifunctional structures in which the N terminus is the host effector domain, while the C terminus targets secretion. This framework will facilitate identification of as-yet-unknown T11SS cargo among the genes that co-occur with T11SS outer membrane proteins. It is our hope that the DUF560-TbpBBD co-occurrence table used for our network annotation will become a resource for other microbiologists studying secreted proteins. This database identifies 851 potential T11SS/cargo pairs spread across 463 bacterial isolates, only 4 of which have been experimentally investigated thus far. Furthermore, our analyses suggest that SSN clusters have predictive power for other characteristics of T11SS cargo, including whether they are surface attached or secreted. The network-enabled classification presented here will facilitate the investigation of both T11SS outer membrane proteins (OMPs) and their cargo in diverse bacteria.

Slam1 and Slam2 T11SS-dependent cargo with known molecular function include those that serve as co-receptors for TonB-dependent metal uptake systems (Noinaj et al. 2010). Similarly, HrpC, which we show here is a T11SS-dependent cargo protein, and its homologs also likely function in metal acquisition by acting as hemophores akin to hemophilin and HasA (Dent et al. 2019; Latham et al. 2020; Yukl et al. 2010). This idea is supported by the fact that an *N. gonorrhoeae hrpC* homolog (NGO0554) is repressed by iron, upregulated in response to oxidative stress, and contributes to resistance to peroxide challenge (Jackson et al. 2010; Quillin et al. 2018; Stohl, Criss, and Seifert 2005). Further, the *hrp* locus, which is conserved across proteobacteria, is predicted to encode TonB and a TonB-dependent receptor (Fig. 2.10). Our working model is that the metal bound by secreted HrpC is passed to its respective TonB-dependent receptor and imported into the cell through TonB energization (Noinaj et al. 2010). Given the conservation of the T11SS *hrp* locus among all *Xenorhabdus* and throughout human microbiome constituents, it will be important in the future to examine the regulation of HrpC T11SS-dependent secretion and the roles of *hrp* machinery in binding and acquiring host metals in a mucosal environment.

In addition to their long-recognized roles in import, TonB and TonB-dependent receptors may also function in protein export. The *Myxococcus xanthus* TonB-dependent transporter Oar can export the protease PopC using the proton motive force to energize membrane translocation (Gómez-Santos et al. 2019). Despite this expansion of the known functional range of TonB-dependent transporters, in the case of the T11SS clusters, TonB is likely responsible for energizing uptake systems, but not T11SS-mediated export. This suggestion is supported by the fact that both the T11SS-1A function of Slam-1 (Hooda et al. 2016) and the T11SS-1B function of HrpB shown here could be reconstructed in *E. coli* without co-expression of native TonB. However, it will be interesting for future studies to investigate possible functional interactions between TonB-dependent transporter-mediated secretion and T11SS-mediated secretion.

Our initial network generated from the entire PF04575 (DUF560) data set included cluster 5, comprising 14 nodes of *Klebsiella* homologs and 2 nodes with *Klebsiella* and *Escherichia* homologs annotated as PgaA/HmsH. This cluster was removed based on its limited number of nodes, predicted topology differences relative to the rest of the network (16-stranded versus 14-stranded barrel), and the fact that the majority of known PgaA homologs were not represented within the cluster. PgaA is a component of one of several machineries for secretion of exopolysaccharides (EPSs) that comprise biofilm matrices of Gram-negative bacteria (Low and Howell 2018; Y. Wang et al. 2016; Whitney and Howell 2013). Despite the topological and substrate (polysaccharide versus protein) differences, the PF04575 assignment of cluster 5 PgaA homologs hints that there could be evolutionary or structural parallels between exopolysaccharide synthase-dependent secretion systems and T11SS. For instance, both T11SS and EPS secretion machineries either have, or associate with proteins that have, tetratricopeptide repeat (TPR) domains (Low and Howell 2018). In EPS secretion systems, the TPR are necessary for secretion or for cargo modification (Marmont et al. 2017; Y. Wang et al. 2016). T11SS TPR domains may similarly modulate activities of other proteins that influence secretion of T11SS substrates. Support for a role of TPR domains in T11SS activity comes from the fact that *X. nematophila* expressing versions of the NilB TPR domain with small deletions display colonization defects that are ameliorated by deletion of the entire N-terminal periplasmic domain (Bhasin, Chaston, and Goodrich-Blair 2012). Our establishment of T11SS as a bona fide secretion system opens new avenues for exploration of its integration with other export and secretion machineries and its coordinated contributions to host-associated phenotypes, including metal homeostasis, aggregation, and biofilm formation.

Materials and Methods

DUF560 sequence similarity network analysis

Enzyme Function Initiative's Enzyme Similarity Tool (EFI-EST) was used to collect all predicted DUF560-domain-containing protein sequences from Interpro 73 and UniProt 2019-02 (accessed 4.24.19) and BLAST for similarity (Gerlt 2017; Gerlt et al. 2015; Zallot, Oberg, and Gerlt 2018). Representative networks collapsed nodes which shared ≥ 40% identity. On an EFI-EST network, edges are drawn according to a database-independent value called alignment score. A greater alignment score requirement means draws fewer edges. For separation of DUF560-domain-containing proteins an alignment score of 38 was chosen (Fig. 2.2). For subcluster 1A 89 was chosen and for subcluster 1B 100 was chosen (Fig. 2.4). The EFI-EST Color SSN tool was used to assign cluster numbers. Networks were visualized and interpreted using Cytoscape v3.7.1 (Shannon 2003) and Gephi v0.9.2 (Bastian, Heymann, and Jacomy 2009). Nodes were arranged with the Fruchterman-Reingold force-directed layout algorithm (Fruchterman and Reingold 1991).

The contents of each cluster were compared to Pfam DUF560 (PF04575) to ensure that clusters were legitimate DUF560 proteins (El-Gebali et al. 2019). Any clusters for which fewer than 18% of sequences were present in Pfam, or which included fewer than 20 sequences were excluded from downstream analysis. This filtering removed cluster 5, which was composed mostly of *Klebsiella pneumoniae* PgaA. This generated a sequence similarity network that contains 10 clusters, 1222 nodes, and 52190 edges with 1589 TaxIDs represented (Fig. 2.4). Using NCBI Taxonomy Browser, each node was examined and manually curated for the isolates' environmental origin(s) among the following categories: water, soil, plant, mammal, animal, invertebrate, nematode, built (environments such as sewage, bioreactors, etc.), multiple environments, and unclassified. If no citation was available, the isolation source was searched for in the biosamples or bioprojects records. If neither

resource was available, strain source was searched for through other resources (NCBI linkout, Google search). A node was assigned an environment if the majority of strains within the node fell into that environment. If a node had no majority environment, the category "multiple environments" was assigned. Any node with animal associated microbes that did qualify as mammal, insect, or nematode associated was designated as generic "animal associated". For fine scale interpretation, analysis focused on cluster 1 and its subclusters (Fig. 2.3; Fig. 2.4)

Three different techniques were used to determine if DUF560 proteins present within our network were genomically associated with TbpB_B_D-domain-containing proteins, hereafter TbpBBD-domain-containing proteins. First, using EFI-GNN, genome-neighborhood-networks were generated for subcluster 1A (Alignment Score 89; 20 ORFs around), subcluster 1B (Alignment Score 38; 10 ORFs around), and subcluster 1C (Alignment Score 38; 10 ORFs around) resulting in 352 DUF560-TbpBBD pairs (Zallot, Oberg, and Gerlt 2018). Next, the RODEO web tool was used to analyze co-occurrence using profile Hidden Markov Models to assign domains to local ORFs resulting in 712 DUF560-TbpBBD pairs (Tietz et al. 2017). Seven additional DUF560-TbpBBD pairs in *Xenorhabdus* were manually annotated. All three datasets were combined for a total of 851 non-redundant protein pairs. SignalP-5 was used to predict the signal peptides of all TbpBBD-domain-containing proteins (Almagro Armenteros et al. 2019; Nielsen et al. 1997). These predictions were used to annotate each node (Fig. 2.3C). Any node that was associated with both signal peptide bearing proteins and those with no predicted signal peptide were annotated according to the signal peptide bearing proteins.

DUF560 genome neighborhood analysis

Subclusters 1A-C were separated and analyzed in EFI-EST with an alignment score of 38 as described above. Each network was then analyzed through EFI-GNN and visualized in Cytoscape v3.7.1 (Shannon 2003) (Alignment Score 38; 10 ORFs up and downstream).

Phyre^2 analysis

TbpBBD$_{sol}$ protein sequences from *Xenorhabdus nematophila* (HrpC), *Providencia rettgeri* (PROVETT_08181/PROVETT_05852), and *Proteus mirabilis* (WP_134940027.1) were collected and the first 22 amino acids were trimmed to remove the signal sequences. These sequences were entered into the Phyre^2 Protein Homology/analogy Recognition Engine v.2.0 to predict potential 3-dimensional structures (Kelley et al. 2015). The top predicted structural model output for all three proteins, hemophilin (6OM5), was used to generate structural models (Fig. 2.6). Subsequent PDB files were visualized with Protean 3D v15. (Protean 3D®. Version 15.0. DNASTAR. Madison, WI).

Co-expression of HrpB and HrpC

The gene encoding HrpB25_26insDYKDDDDK (FLAG-HrpB) was synthesized and cloned into the second multiple cloning site (MCS) of pETDuet-1 by Genscript. The genomic region containing HrpC was amplified from the *X. nematophila* ATCC 19061 (HGB800) genome using primers 1 and 2, digested with restriction enzymes SacI and SalI, and ligated into the pUC19 MCS. Site-directed mutagenesis was used to add a C-terminal FLAG tag onto HrpC using primers 3 and 4. The gene encoding HrpC1_2insV246_247insDYKDDDDK (henceforth HrpC-FLAG) was amplified from pUC19 using primers 5 and 6. The product was ligated into MCS1 of both pETDuet-1 and pETDuet-1/FLAG-HrpB, resulting in expression plasmids pETDuet-1/HrpC-FLAG and pETDuet-1/HrpC-FLAG/FLAG-HrpB. HrpC-FLAG was used to create HrpC1_2insV246_247ins3x(GGGGS)-GST domain (HrpC-GST) using Hi-Fi assembly and primer

pair 7 and 8 and primer pair 9 and 10 to integrate a 15-amino-acid linker and GST into the previous FLAG locus. This process yielded pETDuet-1/HrpC-GST and pETDuet-1/HrpC-GST/FLAG-HrpB. All clones were confirmed by Sanger sequencing using primers 11 and 14 at the University of Tennessee (UT) Genomics Core.

Expression plasmids were transformed into *E. coli* B21(DE3) and *E. coli* C43(DE3) via electroporation. Results were similar for the two strains. Strains were grown in defined medium with 150 μg/ml ampicillin (Orchard and Goodrich-Blair 2004). Bacteria were sub-cultured into 100 ml of broth at an initial optical density at 600 nm (OD_{600}) of 0.028, grown for 18 h at 37°C to reach late logarithmic growth, and induced with 500 μM isopropyl-β-d-thiogalactopyranoside (IPTG). After 1 h, 700 μl of each culture was filtered for subsequent use. At 2.5 h, whole cells were collected by centrifugation and lysed using a bead beater. Remaining supernatants were filtered. Protein concentration of whole-cell lysates was measured via Bradford assay. For supernatant samples, 700 μl of supernatant was precipitated via 10% trichloroacetic acid (TCA) (Koontz 2014a).

Samples were analyzed by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting. For lysates, wells were loaded with 9.5 μg of protein. For supernatants, wells were loaded with half the TCA precipitate collected media. Western blots were probed with either rat anti-FLAG or rabbit anti-GST primary antibody anti-IgG secondary antibody that fluoresces at 680 nm. Intensities were recorded for FLAG or GST reactive bands which were the correct size for mature HrpB and HrpC. For every supernatant sample, a distinct band from the Coomassie blue-stained gel was used as a loading control to normalize intensities of supernatant samples prior to analysis. Unpaired *t* tests were used to compare HrpC-FLAG to HrpC-GST secretion and supernatant fraction HrpB. One replicate of the HrpC-GST 1-h supernatant was

excluded from analysis due to cellular lysis. A Tukey's honestly significant difference (HSD) test was used for comparing supernatant fraction HrpC.

For additional analysis of OMVs, bacteria producing each expression plasmid were sub-cultured in triplicate into 40 ml of broth at an initial OD_{600} of 0.04. After 5 h at 37°C to reach mid-logarithmic growth, cultures were induced with 500 μM IPTG. After an additional 2.5 h, each culture was clarified via centrifugation, then the supernatant was filtered and centrifuged at 150,000 relative centrifugal force (RCF) for 3 h. The entire OMV pellet was solubilized in SDS sample buffer. Six hundred microliters of the total supernatant or the clarified (post-ultracentrifugation) supernatant were precipitated via 10% TCA and suspended in SDS sample buffer (Koontz 2014a). Samples were analyzed by SDS-PAGE and Western blotting as described above.

Phylogenetic tree generation

Phylogenetic analysis was performed as described previously. Briefly, select *Xenorhabdus* and *Photorhabdus* species were analyzed using MicroScope MaGe's Gene Phyloprofile tool (Chaston et al. 2013; Vallenet et al. 2017) to identify homologous protein sets which were conserved across all assayed genomes. Putative paralogs were excluded from downstream analysis to ensure homolog relatedness, resulting in 665 homologous sets. Homolog sets were retrieved via locus tag indexing using BioPython (Cock et al. 2009), individually aligned using Muscle v3.8.31 (Edgar 2004), concatenated using Sequence Matrix v1.8 (Vaidya, Lohman, and Meier 2011), and trimmed of nucleotide gaps using TrimAL v1.3 (Capella-Gutiérrez, Silla-Martínez, and Gabaldón 2009). JmodelTest v2.1.10 (Darriba et al. 2012) was used to choose the GTR+γ substitution model for maximum likelihood and Bayesian analysis.

For nematode phylogenetic analysis select *Steinernema* and *Heterorhabditis* species were analyzed. The Internal Transcribed Spacer, 18S rRNA, 28S rRNA, Cytochrome Oxidase I, and 12S

rRNA loci were collected from Genbank as available and used as homologous sets (TableS4_CombinedLociTable). Nematode species which had fewer than 3 of the 5 loci sequenced were excluded from downstream analysis. Homologous sets were individually aligned, concatenated, and trimmed using the same methods as the *Xenorhabdus* sequences. JmodelTest v2.1.10 was used to choose the GTR+γ+I substitution model (Darriba et al. 2012).

Maximum likelihood analyses were performed via RAxML v8.2.10 (Stamatakis 2014) using rapid bootstrapping and 1000 replicates, and were visualized via Dendroscope v3.6.2 (Huson and Scornavacca 2012). Nodes with less than 60% bootstrap support were collapsed. Bayesian analyses were performed via MrBayes v3.2.6 with BEAGLE (Altekar et al. 2004; Ayres et al. 2012; Ronquist et al. 2012) on the Cipres Science Gateway platform (M. A. Miller, Pfeiffer, and Schwartz 2010). 500,000 MCMC replicates were performed for the bacterial tree, 4,000,000 were performed for the nematode tree. 25% of replicates were discarded as burn-in, and posterior probabilities were sampled every 500 replicates. Two runs were performed with 3 heated and one cold chain. The final standard deviation of split frequencies was 0.000000 for the bacterial tree, and 0.002557 for the nematode tree. Bayesian trees were visualized with FigTree v1.4.4 (Rambaut 2018). Bayesian and maximum likelihood methods generated phylogenies with consistent topologies (Fig. 2.9; Fig. 2.11).

WT vs. *Δ nil* colonization of *S. anatoliense*, *S. carpocapsae*, and *S. websteri*

Strains are described in Table 2. Bacteria-free eggs of *S. anatoliense*, *S. carpocapsae*, and *S. websteri* were generated and exposed to a *nil* locus mutant (HGB1495), and a *nil* locus complemented strain (HGB1496) grown on lipid agar plates for 2 days at 25°C. Lipid agar plates were placed into White traps 1 week after plating eggs to collect infective juvenile (IJ) nematodes. Nematode colonization was visualized using fluorescence microscopy on the Keyence BZX-700 to observe GFP expressing bacteria in the receptacle. This was done for each species in biological triplicate and

technical duplicate. To determine the number of CFU per IJ, nematodes were surface sterilized, and ground for 2 min using a Fisher brand motorized tissue grinder (CAT# 12-1413-61) to homogenize the nematodes and release colonizing bacteria. Serial dilutions in PBS were performed and plated on LB agar, which were incubated at 30°C for 1 day before enumerating CFUs (Fig. 2.12). To calculate the CFU per colonized IJ, the percent colonized nematodes was divided by the CFU/IJ for each biological replicate. The data were analyzed using a one-way ANOVA with Tukey's multiple comparison's test to compare the mean of each treatment.

Acknowledgements:

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

Chapter 3: Bacterial hemophilin homologs have conserved roles in host-microbe heme capture that are dependent on highly specific type eleven secretor proteins

Authors: Alex S. Grossman, David A. Gell, Derek G. Wu, Jacob Elrod, Heidi Goodrich-Blair

Author contributions:

Alex S. Grossman performed, and generated figures for all network analyses, co-expression experiments, transcriptional activity experiments, and *in vivo* study of *X. nematophila* and its nematode host *S. carpocapsae*. Alex S. Grossman and Jacob Elrod collaborated to generate a mutant library of transcriptional reporters. Alex S. Grossman and Derek G. Wu collaborated on one replicate of the chimeric hemophilin secretion experiment, the *S. carpocapsae* emergence/fitness assay, and the cell free conditioned medium experiment. Heidi Goodrich-Blair performed sequence alignment of HrpC homologs in *Xenorhabdus* and generated Figure 3.4. David A. Gell performed all physicochemical characterization experiments on the identified hemophilin homologs, including ligand binding and structural modelling. Alex S. Grossman wrote the text of this chapter, David A. Gell, Heidi Goodrich-Blair, and Derek G. Wu assisted in proofreading

Abstract:

The biosynthesis of metabolites essential to cellular life is energetically costly and is impossible in the case of the metal cofactors required for enzymatic processes. As such, bacteria utilize surface and secreted proteins to acquire valuable nutrients from their environment. The cargo proteins of the recently described type eleven secretion system (T11SS) have been connected to host specificity (Grossman, Mauer, et al. 2022), metal homeostasis (Grossman, Escobar, et al. 2022), and nutritional immunity evasion (Hooda et al. 2016). This Sec-dependent, Gram-negative secretion system consists of an outer membrane protein that is encoded by organisms throughout the phylum Proteobacteria, including human pathogens such as *Neisseria meningitidis, Proteus mirabilis, Acinetobacter baumannii,* and *Haemophilus influenzae*. Experimentally verified cargo proteins of this secretion system

include the host metal acquisition proteins transferrin binding protein B (TbpB), Lactoferrin binding protein B (LbpB), and Hemoglobin/Haptoglobin binding protein A (HpuA), as well as the hemophilin homologs Heme receptor protein C (HrpC) and Hemophilin A (HphA), the complement immune evasion protein factor-H binding protein (fHbp), and the host symbiosis factor Nematode intestinal localization protein C (NilC). In this study, we sought to define the specificity of T11SS systems for their cognate cargo proteins using the taxonomically distributed T11SS-hemophilin cargo pairs. We conducted bioinformatic and physicochemical characterization of four select hemophilin family proteins and discovered previously unknown ligand binding diversity among members of this protein family. *In vivo* expression of hemophilin homologs revealed that each was secreted in a specific manner by its cognate T11SS protein. Furthermore, secretion assays of chimeric hemophilin proteins revealed that this specificity is partially dictated by the C-terminal domain of the cognate cargo. Finally, *in vivo* examination of hemophilin and its cognate T11SS protein within *Xenorhabdus nematophila* demonstrated that this is a heme-binding system that is essential for growth under metal starvation and contributes to host nematode fitness.

Introduction:

Many enzymes have evolved expanded catalytic potential through the incorporation of metallic cofactors and prosthetic groups. Iron cofactors are essential to most living organisms due to their functional contributions to enzymes required for DNA synthesis, photosynthesis, respiration, and nitrogen metabolism (Bertini, I., Gray, H.B., Stiefel, E.I. and Selverstone Valentine 2007). Without the biochemical flexibility provided by metallic cofactors, life as we know it would be impossible. Because of this, and the limited bioavailability of essential metals, competition among organisms for these ions can be fierce. In some cases, including many marine environments, competition for bioavailable iron is actually the major limiting factor of microbial growth (H.-R. Zhang et al. 2021). This phenomenon is exploited by animals through nutritional immunity, in which

valuable ions, such as iron, are sequestered to slow or deter the pathogenic growth of microbes. Within animal hosts iron is sequestered by proteins such as hemopexin, transferrin, lactoferrin, and ferritin (Hood and Skaar 2012). Medical conditions, such as hemochromatosis, that increase the serum iron concentration or prevent effective storage of iron increase a patient's risk of infection from many bacteria, including *Escherichia coli*, *Listeria monocytogenes*, and *Yersinia enterocolitica* (Khan, Fisher, and Khakoo 2007). In turn, to overcome iron limitation within an animal host, bacteria have evolved means of countering nutritional immunity, including adaptations to use alternative catalytic metals (Posey and Gherardini 2000), production of high-affinity siderophores (Holden and Bachman 2015), and or membrane bound uptake receptors (Hood and Skaar 2012) that facilitate acquisition of iron from host metalloproteins..

Several recent studies have linked the type eleven secretion system (T11SS) outer membrane proteins (OMPs) and their cargo proteins to iron uptake strategies in Gram-negative bacteria. In *Neisseria*, the T11SS proteins Slam1 and Slam2 surface expose cargo proteins that are responsible for binding host-metal carriers: Transferrin binding protein B (TbpB) and Lactoferrin binding protein B (LbpB) are surface exposed by Slam1, and Hemoglobin/Haptoglobin binding protein A (HpuA) is surface exposed by Slam2 (Calmettes et al. 2012; C. J. Chen et al. 1996; Hooda et al. 2016; Ostan et al. 2017). These surface exposed outer membrane lipoproteins facilitate bacterial colonization by capturing their respective host factors (transferrin, lactoferrin, or hemoglobin/haptoglobin) and complexing with a TonB-dependent uptake channel capable of importing the iron cofactor. Since surface exposure is essential for the function of these lipoproteins, genetic inactivation of the T11SS OMP, Slam1 prevents effective colonization and pathogenesis by *Neisseria* (Hooda, Lai, and Moraes 2017). While *Neisseria* Slam1 and Slam2 have specificity for their respective cargo (Hooda et al. 2016), no mechanism has yet been proposed to explain this level of specificity and it is unknown if all T11SS have specificity for their cognate cargo. Bioinformatic analyses revealed a large number of

potential T11SS-dependent cargo, lipid anchored and unanchored, which frequently exist in cognate pairs/groupings according to genomic co-occurrence analyses (Grossman, Mauer, et al. 2022; Hooda, Lai, and Moraes 2017). To date, all verified or predicted T11SS-dependent cargo have two distinct domains: an N-terminal domain that varies in predicted structure and function across cargo, and a C-terminal, 8-stranded β-barrel domain from the TbpBBD or lipoprotein C family (Grossman, Mauer, et al. 2022; Hooda et al. 2016; Hooda, Lai, and Moraes 2017).

T11SS are capable of secreting unlipidated cargo, such as soluble hemophores like the heme receptor protein C described within this manuscript (HrpC) and hemophilin A from *A. baumanii* (HphA) (Bateman et al. 2021; Grossman, Mauer, et al. 2022). HphA can bind hemoglobin directly to pirate heme from this prevalent host metalloprotein, and contributes to the virulence of *Acinetobacter baumannii* in a murine infection model through its role as a co-receptor to the TonB-dependent heme receptor HphR (Bateman et al. 2021). Thus, hemophilin proteins represent a high-affinity heme acquisition system comparable to HasA from *Serratia marcescens* (Arnoux et al. 2000) or IsdB from *Staphylococcus aureus* (Gianquinto et al. 2019). Known members of the hemophilin protein family function to import heme from a host environment as depicted in Figure 3.1. Compiling the results of published data from multiple organisms into a single model suggests that that hemophilin crosses the inner membrane through the Sec translocon to reach the periplasm (Grossman, Mauer, et al. 2022), may interact with chaperones such as Skp to traverse the periplasm (Huynh et al. 2022), and then crosses the outer membrane in a T11SS-dependent manner to reach the extracellular milieu (Bateman et al. 2021; Fantappiè et al. 2017; Grossman, Mauer, et al. 2022; Hooda et al. 2016). From here, hemophilin captures heme that is released from hemoglobin or other host derived hemoproteins using a high affinity binding domain (Bateman et al. 2021). This domain seems to bind porphyrins with very little conformational change (Latham et al. 2020). Holo-hemophilin interacts with a TonB-dependent outer membrane co-receptor which imports the heme molecule

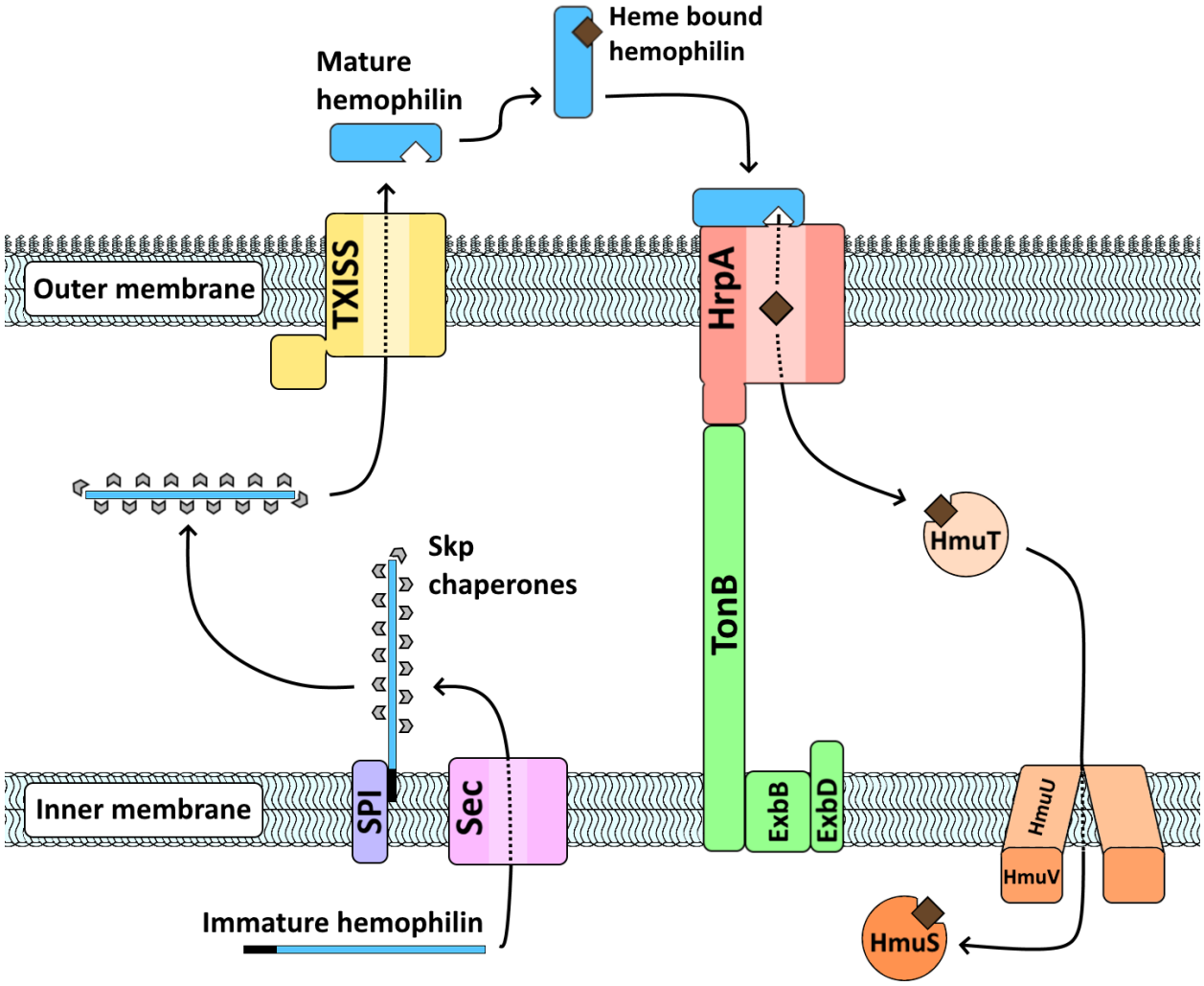


Figure 3.1. **Conceptual model of hemophilin mediated heme acquisition.** Described hemophilin homologs have a signal peptide directing them to the Sec translocon and signal peptidase I. Based on analogy to the T11SS-dependent lipoprotein TbpB, hemophilin may depend on Skp for periplasmic transport. Each hemophilin protein has a cognate T11SS which translocates it through the outer membrane where it acts to bring heme to a TonB-dependent receptor for uptake. Based on analogy to the *Yersinia pestis* Has hemophore system, heme is transported through the periplasm and across the inner membrane by the hemin utilization system (Hmu). Once brought into the cytoplasm, heme can be degraded by heme oxygenase (HmuS), an enzyme which is frequently encoded adjacent to hemophilin loci.

into the periplasm and releases apo-hemophilin. Finally, the hemin utilization system chaperones the periplasmic heme into the cytoplasm for incorporation into the cell or digestion by heme oxygenase to free the iron cation (Schwiesow et al. 2018).

Within this overall framework, hemophilin can have diverse functional roles within different organisms and environments. For example, in *Haemophilus haemolyticus* hemophilin can act as a probiotic factor by making bioavailable iron inaccessible to *H. influenzae* (Atto et al. 2020), while in *A. baumannii* hemophilin can act as a virulence factor important for systemic infection in a murine model (Bateman et al. 2021). Furthermore, hemophilin homologs found in sequence databases display sequence variation within the heme-binding handle domain, suggesting possible variability in ligand binding. Since previous research on hemophilin-family-proteins has focused primarily on human associated pathogens, we sought to expand our understanding of hemophilin activities by characterizing homologs from nematode mutualistic bacteria in the genus *Xenorhabdus*. Further, we took advantage of the availability of orthologous T11SS-hemophilin cargo pairs to further examine the mechanisms underlying T11SS cargo recognition and specificity.

Results:

Sequence similarity networking

To explore the relatedness of hemophilin family proteins and to identify subcluster divisions that may reflect divergent function, a sequence similarity network generated through Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) was overlaid with a taxonomic framework and used (Zallot, Oberg, and Gerlt 2019)
(Chap3SubclusterRodeoResultsAndNetworkNodeTable.xlsx and Fig. 3.2). The network was populated with the previously published dataset of T11SS-associated cargo that were predicted to be soluble (Grossman, Mauer, et al. 2022), revealing a single major cluster containing all previously

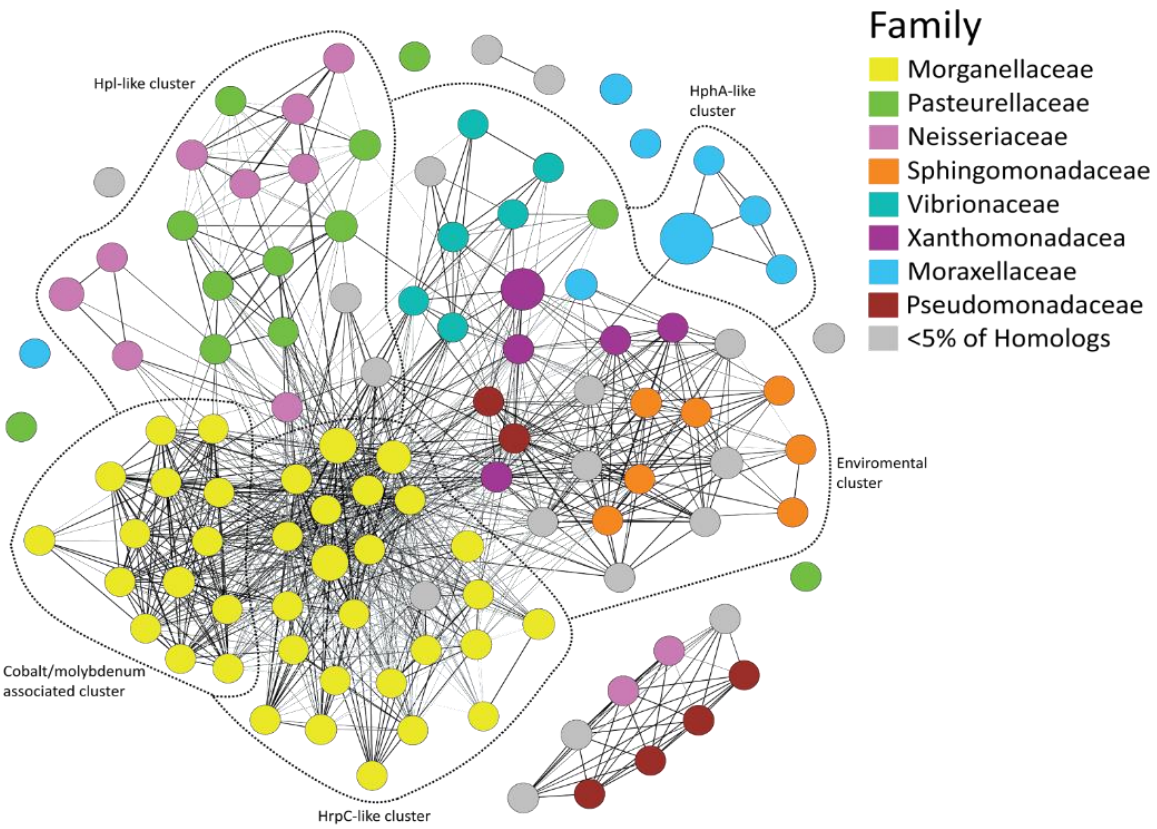


Figure 3.2. **Distribution and relatedness of soluble T11SS-dependent cargo proteins.** A sequence similarity network of all soluble TbpBBD-domain-containing proteins identified in Grossman and Mauer et all (2022). All of these sequences are encoded alongside a T11SS gene (± 6 genes) Sequences were submitted to EFI-EST to generate the network. Each node represents one or more protein sequences with 80% or greater identity, the larger the node the more sequences it contains. Edges indicate an alignment score of 35 or greater. Edge darkness indicates shared sequence identity, with the darkest edges being the most identical. Dotted lines indicate proposed subclusters as defined using the Fruchterman-Reingold algorithm and the distribution of characterized proteins. Sequences which did not connect to any characterized hemophilin family protein were excluded from further analysis (Figure 3.3). Despite having similar domain architectures to hemophilin, none of these divergent sequences contained experimentally characterized proteins.

described hemophilin proteins (88/107 nodes), one smaller cluster containing uncharacterized proteins from predominantly *Pseudomonas* and *Neisseria* species (9/107 nodes), and a few unassociated doublets and singletons (10/107 nodes). To focus this study specifically on hemophilin and its direct evolutionary relatives all nodes not within the central cluster were removed. The remaining nodes were labeled according to taxonomic family and separated using a force-directed separation algorithm. Regions of higher interconnectivity were separated visually into subclusters revealing five subclusters predominantly populated by seven families from three different classes of *Proteobacteria* (Fig. 3.3AC). The subclusters that contained Hpl, HrpC, and HphA were named Hpl-like, HrpC-like, and HphA-like after their respective characterized member. One novel subcluster was termed cobalt/molybdenum associated due to its genomic co-occurrence with genes predicted to encode cobalt- or molybdenum-dependent enzymes. Another novel subcluster was termed "environmental" due to the dominant presence of homologs encoded by microbes found in soil, water, and plant-associated environments. The subclusters did not fall exclusively along taxonomic lines. For instance, the Hpl-like subcluster includes *Neisseria* and *Pasteurella* hemophilin homologs that clustered closely together, possibly indicating that these genes may have been horizontally exchanged. Cross-taxonomic-class sequence similarity also was observed in the environmental cluster of hemophilin homologs.

To identify potential distinguishing features among the five subclusters, we examined the identities of genes that commonly occur within genomic neighborhoods (± 6 genes) with *hrpB*. All sequences from the network were submitted to the Rapid ORF Detection & Evaluation Online (RODEO) web tool that uses profile hidden Markov models to identify co-occurring protein domains (Chap3SubclusterRodeoResultsAndNetworkNodeTable.xlsx) (Tietz et al. 2017). Consistent with previous observations, 92.3% (240/260 sequences) of hemophilin homologs, regardless of subcluster, were encoded in association with genes predicted to encode TonB and TonB-dependent

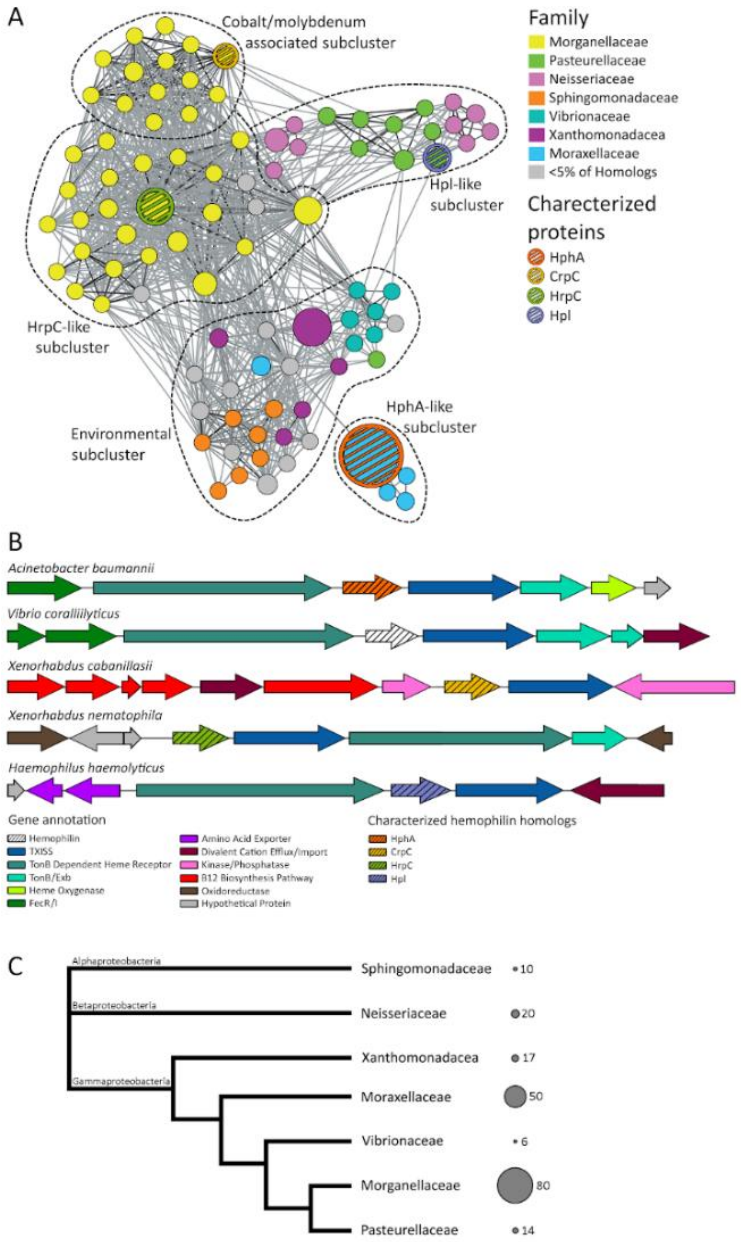


Figure 3.3. **Distribution and relatedness of hemophilin family proteins. A)** A sequence similarity network of hemophilin homologs generated with EFI-EST. Each node represents one or more protein sequences with 80% or greater identity, the larger the node the more sequences it contains. Edges indicate an alignment score of 35 or greater. Edge darkness indicates shared sequence identity, with the darkest edges being the most identical. Dotted lines indicate proposed subclusters as defined using a force-directed algorithm and the distribution of characterized proteins. **B)** Representative genomic neighborhoods from the subclusters identified within the sequence similarity network demonstrating common co-occurring genes. **C)** A cladogram of the seven genera which encode the most hemophilin family proteins, split between Alpha-, Beta-, and Gammaproteobacteria. The circles and numbers to the right of the cladogram indicate the relative abundance of known hemophilin homologs encoded by each genus.

receptors, suggesting a strong and consistent link between T11SS cargo and TonB-dependent uptake across the outer membrane (Grossman, Mauer, et al. 2022; Hooda, Lai, and Moraes 2017). However, some subclusters had informative co-occurrences. For example, hemophilin homologs in the HphA-like subcluster showed nearly universal co-occurrence with genes predicted to encode heme oxygenase (51/53) and the iron-sensing regulator FecR (48/53). Similarly, homologs in the "environmental" subcluster typically were encoded near genes predicted to encode metal responsive regulatory proteins (FecR or Fur) (64/72) and heme oxygenase (19/72). Additionally, 69/72 "environmental" subcluster loci encoded additional regulatory genes such as RpoE, IscR-family regulators, and LysR-family regulators. RpoE is a sigma factor that responds to extra-cytoplasmic stress and is essential for metal resistance (Egler et al. 2005). IscR regulates iron-sulfur cluster biosynthesis according to cellular demand (Schwartz et al. 2001), and LysR-family regulators drive diverse pathways by binding DNA directly in response to co-inducing/co-repressing ligands (Maddocks and Oyston 2008). Homologs in the cobalt/molybdenum-associated subcluster occurred alongside other predicted T11SS-dependent cargo (14/18) and were either located adjacent to a B12 biosynthetic locus (3/18) or near a formate dehydrogenase locus (13/18). Formate dehydrogenase activity relies on molybdenum metal cofactors, and their co-occurrence with T11SS OMPs may hint at a role for the latter in molybdenum acquisition. The HrpC-like subcluster includes genes predicted to encode redox enzymes, such as formate dehydrogenase (48/81) and NADPH:quinone oxidoreductase (12/81), tRNA synthases and modification systems (30/81), and regulatory proteins including TetR family regulators, FaeA family regulators, FecR, and FecI (32/81). Homologs in the Hpl-like subcluster had few unifying co-occurrences, however many co-occurred with tRNA synthases/modification systems (14/36), specifically selenocysteine tRNA synthases (6/36) (Fig. 3.3B).

Sequence and phylogenetic analysis of hemophilin and T11SS proteins within *Xenorhabdus*

Genes predicted to encode the hemophilin homolog HrpC are conserved throughout the *Xenorhabdus* genus, as well as the closely related *Photorhabdus* genus (see the Morganellaceae nodes in Fig. 3.3A). However, this is not the only hemophilin homolog encoded by these genera. Some *Xenorhabdus* encode an additional paralog, HrpC2, in the same neighborhood as HrpC, *X. hominickii* encodes a transposon shifted version of HrpC2 termed $HrpC_{tn}$, and some species encode CrpC adjacent to a B12 biosynthetic locus (Grossman, Mauer, et al. 2022). Amino acid alignments of HrpC homologs from all *Xenorhabdus* species revealed a striking delineation into at least three subtypes based on divergent sequence in the N-terminal region, spanning from ~aa41-107 of the *X. nematophila* HrpC (XNC1_0075) sequence (Fig. 3.4). All HrpC homologs encoded a 10-aa motif with a conserved histidine (H49), and this domain was absent from all HrpC2/CprC homologs. A subset of homologs, termed HrpC3, lack the N-terminal region, including an otherwise universally conserved DXNG motif (Fig. 3.5).

Given our observation that hemophilin homologs are diversifying among *Xenorhabdus* species, we next examined whether parallel diversification is present among the cognate, co-occurring T11SS OMPs. To examine the diversity of *Xenorhabdus* hemophilin-associated T11SS OMPs, we estimated their phylogenetic relatedness using nucleotide sequences and maximum likelihood estimation with 1000 bootstraps. The T11SS-encoding *hrpB* genes from *Photorhabdus asymbiotica* and *Photorhabdus luminescens* were included in the analysis as an outgroup. For an additional comparison, we included the *Xenorhabdus* T11SS OMP *nilB*, which is found in only three species of *Xenorhabdus*. The resulting phylogeny (Fig. 3.6) revealed that *nilB* genes are more divergent from other *Xenorhabdus* T11SS genes than are the *hrpB* homologs from *Photorhabdus*. In addition, correlating with the diversification of the *hrpC* hemophilin homologs, the *Xenorhabdus hrpB* homologs separated into two clades of *hrpB* and *hrpB2* (note that the *hrpC3* homologs noted above

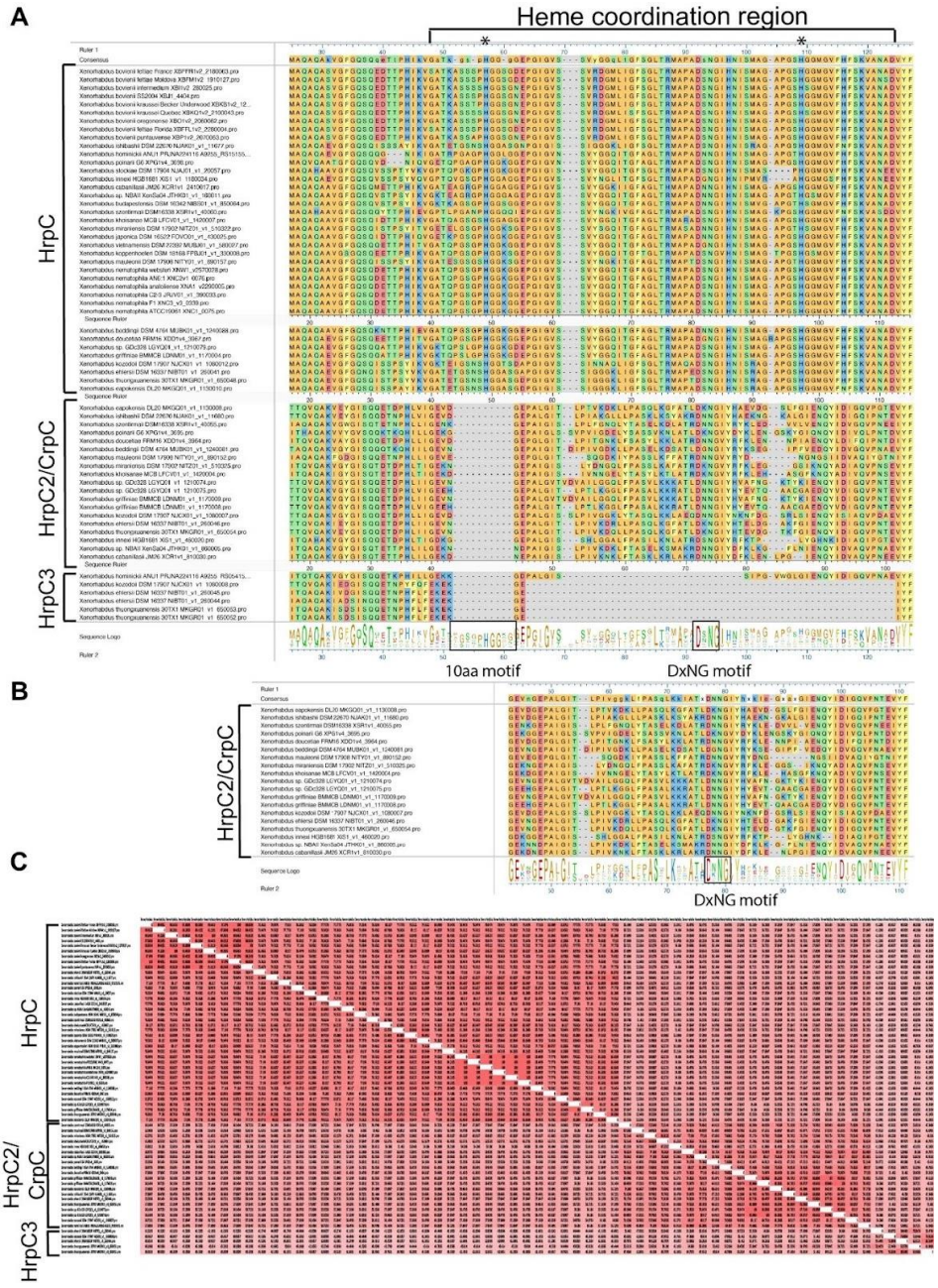


Figure 3.4. **Alignment of the N-terminal region of *Xenorhabdus* hemophilin homologs**. Shown is multiple sequence alignment and sequence consensus logo (generated using DNA MegAlign Pro MUSCLE) of 63 (**A**) or 19 (**B**) *Xenorhabdus* hemophilin family homologs and the resulting distance matrix with increasing percent pairwise identity represented by increasing intensity of red highlighting (**C**). Based on the 63 sequence alignment shown in (**A**), at least three distinct groups of homologs could be discerned: HrpC (**top**), HrpC2/CrpC (**middle**), and HrpC3 (**bottom**). All sequences found in the second and third groups have a truncated version of the N-terminal handle domain. All representatives from this truncated clade have lost a 10 aa motif centered on a histidine residue with 100% conservation amongst the ancestral group that is predicted to be essential for heme binding. The HrpC3 group is also missing a ~60 aa region centered on a DXNG motif with 100% conservation amongst other homologs.

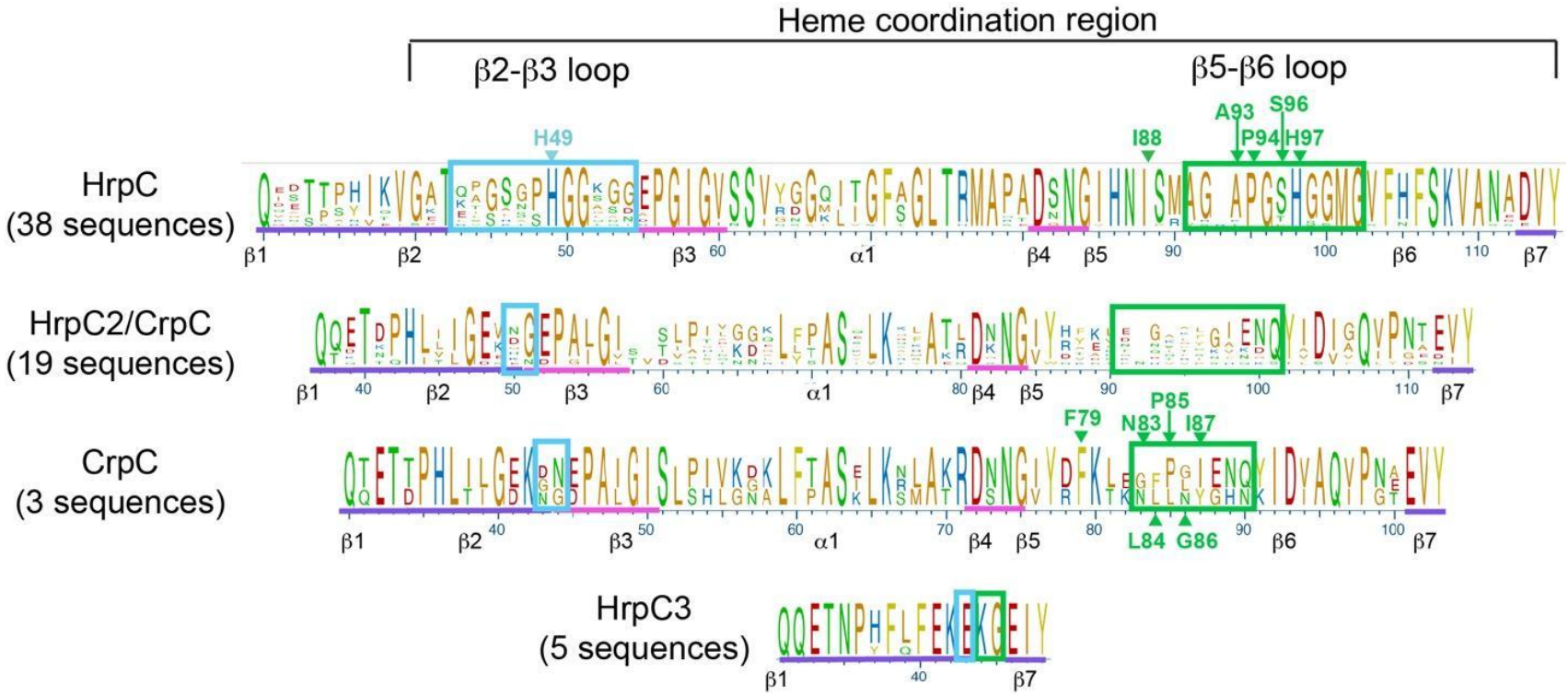


Figure 3.5. **Hemophilin homolog N-terminal domain consensus sequence logo based on alignments**. Multiple sequence alignment (generated using DNA MegAlign Pro MUSCLE) of the HrpC group (38 sequences), the HrpC2/CrpC group (19 sequences), CrpC only (3 sequences), and HrpC3 only (5 sequences). Sequences underlined in purple (representing β1-β2 and β7 of the Hpl and HpuA structures), and pink (representing β3, β4, and β5 of the Hpl and HpuA structures) are conserved regions present in all *Xenorhabdus* HrpC homologs (purple), or present in all HrpC/HrpC2/CrpC homologs, but not in HrpC3 homologs (pink). The region shown includes the heme coordination region of hemophilin, composed of a β2-β3 loop (turquoise box) with a conserved histidine (H49 of *X. nematophila* HrpC) and a β5-β6 loop (green box, amino acid numbering according to *X. nematophila* HrpC) as described in the text. The β2-β3 loop is absent from HrpC2/CrpC and HrpC3 homologs. The β5-β6 loop shows strong conservation among the HrpC homologs, including residues known to coordinate heme in Hpl and HpuA, but variability among the HrpC2/CrpC homologs. The CrpC homologs, including *X. cabanillasii*, have a conserved phenylalanine-lysine pair in the β5-β6 loop sequence (amino acids F79 and K80 in *X. cabanillasii* CrpC). The HrpC3 homologs lack the heme coordination region and contain only a lysine-glycine amino acid pair in that region.

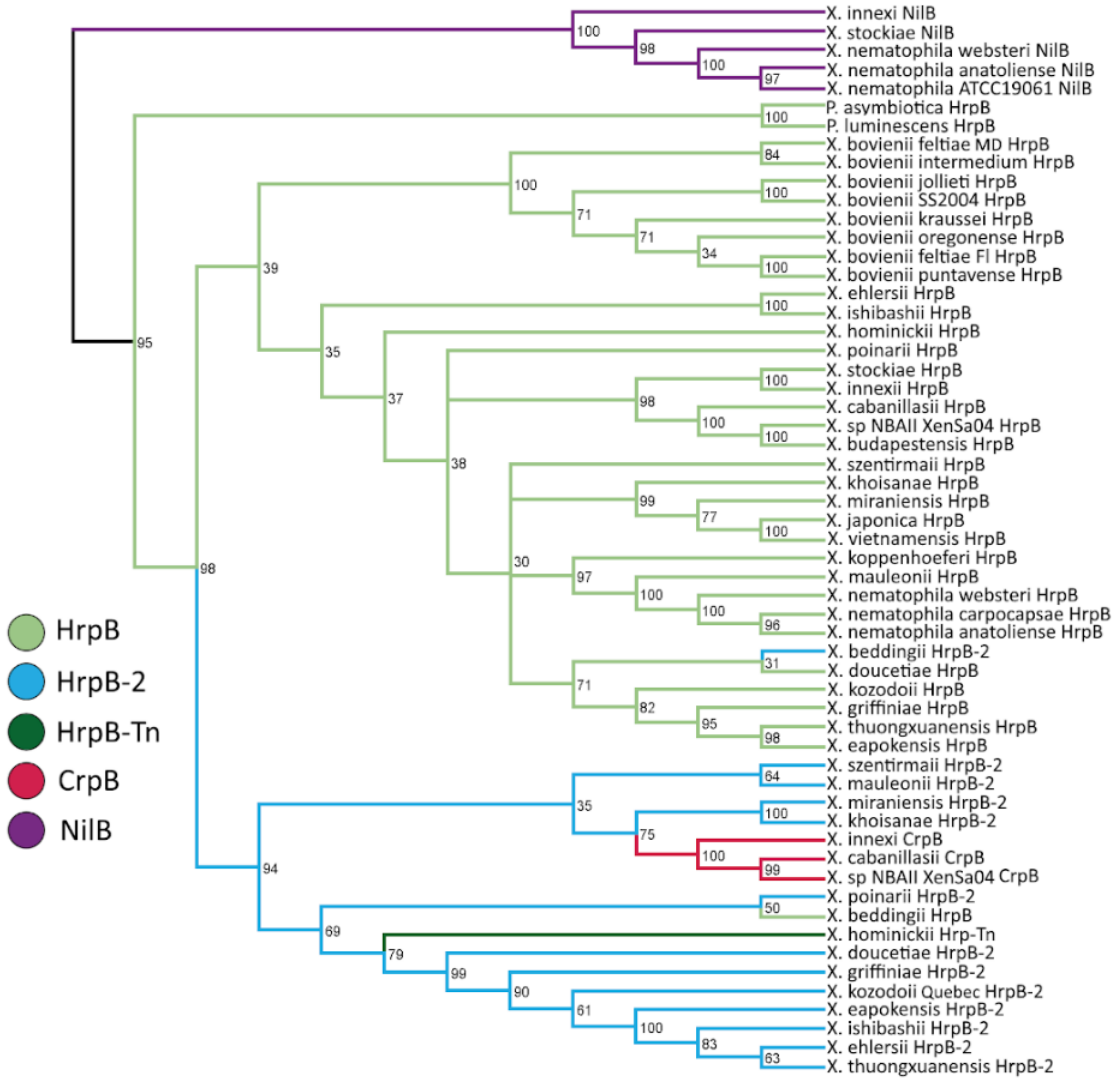


Figure 3.6. **Single locus cladogram of T11SS genes from the genus *Xenorhabdus*.** T11SS sequences from *Photorhabdus luminescens* and *P. asymbiotica* were used as outgroup samples. Node labels indicate percent bootstrap support amongst 1000 bootstraps. $HrpB_{X.nem}$ and HrpB2 cluster separately (with the exception of *Xenorhabdus beddingii* in which the *hrpB* and *hrpB2* homologs have switched genomic locations). $HrpB_{Tn}$ seems to have originated from the HrpB2 clade prior to its horizontal acquisition by *X. hominickii*. CrpB seems to have evolved from a paralog of HrpB2. NilB is exceptionally divergent, supporting the hypothesis that it was horizontally acquired.

do not have a corresponding additional *hrpB* partner). These data indicate that the most recent common ancestor of *Photorhabdus* and *Xenorhabdus* encoded *hrpB* and that in some *Xenorhabdus* species, this gene was later duplicated (presumably along with a *hrpC* cargo partner). In addition, these data support our previous suggestion that the *nilB* T11SS homolog was horizontally acquired by a subset of *Xenorhabdus* species (Cowles and Goodrich-Blair 2008). We noted that the *X. beddingii hrp* locus may have undergone a genomic inversion, since the paralog we designated as *hrpB2* based on its location actually clustered with the *hrpB* genes.

No experimental data exists for the specific function of HrpB2 and its respective hemophilin homolog, but its broad distribution, consistent retention, and apparent correlation with the cladistics of *Xenorhabdus*' nematode hosts (Grossman, Mauer, et al. 2022) suggests it provides some unique function or advantage. The existence of $hrpB_{tn}$ and a corresponding $hrpC_{tn}$ supports this since it appears to be the result of horizontal acquisition of the $hrpB_2$ gene by *X. hominickii*, potentially coinciding with a host-switching event where *X. hominickii* became capable of colonizing *Steinernema karii*. Finally, the phylogeny shows that the cycle of T11SS diversification via duplication and subsequent paralogous evolution may be repeating itself within the *hrpB2* gene cluster, resulting in *crpB* and its respective cargo encoding gene *crpC*. Despite the fact that no organism which encodes *crpB* also encodes *hrpB2*, *crpB* falls in the *hrpB2* cluster, indicating another horizontal acquisition and subsequent divergence. As previously hypothesized, this acquisition may coincide with another host switching event in *Xenorhabdus* evolutionary history, which shifted a lineage of bacteria from colonizing one group of nematodes (clade II) to another group (clade V) (Grossman, Mauer, et al. 2022). Overall, our sequence and phylogenetic analyses of Xenorhabdus hemophilin homologs and their TXISS OMP partners indicates that these cargo-secretion pairs are diversifying in response to selective pressures imposed by the host environments in which these bacteria live.

Co-expression of diverse T11SS/hemophilin pairs

Our data described above indicate that TXISS OMPs and their cargo have been horizontally transferred or duplicated as pairs during the evolution of *Xenorhabdus* genomes. This supports the concept that TXISS OMPs and their cargo are co-diversifying and may be specific for each other. To test the extent of specificity between hemophilin cargo proteins and their paired TXISS OMP secretor, and to further establish the role of TXISS OMPs in secretion of hemophilin cargo, we conducted secretion assays on different cognate or non-cognate pairs, chosen from the different subclusters of the hemophilin family network. Four T11SS-cargo pairs were selected for expression within *E. coli* Bl21 C43 for secretion assays. $HrpB_{X.nem}$/HrpC from *Xenorhabdus nematophila*, $HrpB_{H.haem}$/Hpl from *Haemophilus haemolyticus*, HsmA/HphA from *A. baumannii*, and CrpB/CrpC from *Xenorhabdus cabanillasii* were cloned into pETDuet-1 based expression vectors to perform co-expression and secretion experiments. Additionally, plasmids were constructed which co-expressed $HrpB_{X.nem}$ alongside the non-cognate hemophilin homologs Hpl, HphA, and CrpC. Western blotting of supernatant and cellular lysates was performed to monitor cargo localization in the extracellular milieu. Extracellular HrpC and HphA both ran as single protein bands somewhat larger than would be predicted from mature protein sequence alone, with HrpC appearing at ~30kD (~24kD expected) and HrpA appearing at 29kD (~25kD expected) (Science Gateway n.d.). In both cases this increase in apparent size is too large to be explained by the presence of the signal peptide, though this larger apparent size is consistent with our previous studies of HrpC (Grossman, Mauer, et al. 2022). Extracellular Hpl and CrpC were both observed running as two different bands. Hpl had a predominant band at ~32kD and a minor one at the expected size of ~26kD. CrpC had a predominant band at ~19kD and a minor band at the predicted size of ~27kD. Interestingly, the ~19kD CrpC band was absent or reduced when the protein was expressed alone or with the non-cognate $HrpB_{X.nem}$, suggesting potential protein modification (e.g., cleavage) during or after secretion

by its cognate T11SS. We predict that if the protein is being cleaved, the potential cleavage site is likely between residues 74 and 75 (-KRDN/NGIY-) based on the difference between the expected and observed size of CrpC-FLAG. For enumeration of Hpl and CrpC localization we opted to sum both protein bands to fairly represent all of the secreted protein.

Co-expression of hemophilin homologs with a cognate T11SS protein always significantly increased the concentration of cargo protein found in the supernatant, though the level of secretion varied greatly among T11SS proteins. At one hour after induction, the presence relative to the absence of the cognate T11SS OMP increased the average level of extracellular hemophilin homolog protein ($HrpB_{X.nem}$/HrpC: 34.2-fold; $HrpB_{H.haem}$/Hpl: 59.3-fold, HsmA/HphA: 4.6-fold; CrpB/CrpC: 56.9-fold) (Fig. 3.7 and Chapter3WesternImages.xlsx). $HrpB_{X.nem}$ significantly increased (4.8-fold) the average extracellular levels of the non-cognate cargo Hpl, though it was significantly less effective at doing so than the cognate T11SS $HrpB_{H.haem}$. $HrpB_{X.nem}$ did not significantly impact average extracellular levels of HphA, (1.1 fold), or CrpC (1.7 fold).

Relative secretion of chimeric hemophilin by HrpB homologs

The data described above indicates that individual hemophilin family cargo proteins and their cognate T11SS OMPs have specificity for each other. Published literature has implicated the C-terminal β-barrel domain of cargo proteins in directing secretion by T11SS OMPs (Hooda, Lai, and Moraes 2017). To assess the role of the hemophilin C-terminal domain on T11SS specificity, two chimeric hemophilin cargo were engineered. The first chimeric cargo had the N-terminal handle domain from HrpC and the C-terminal β-barrel domain from Hpl (henceforth HrpC-Hpl), while the second had the N-terminal handle from Hpl and the C-terminal β-barrel domain from HrpC (henceforth Hpl-HrpC). pETDuet-1 constructs were assembled to independently co-express $HrpB_{H.haem}$ and $HrpB_{X.nem}$ alongside both chimeric cargo proteins (Fig. 3.8A) and Western blotting of supernatant and cellular lysates was again performed to monitor cargo secretion. The HrpC-Hpl

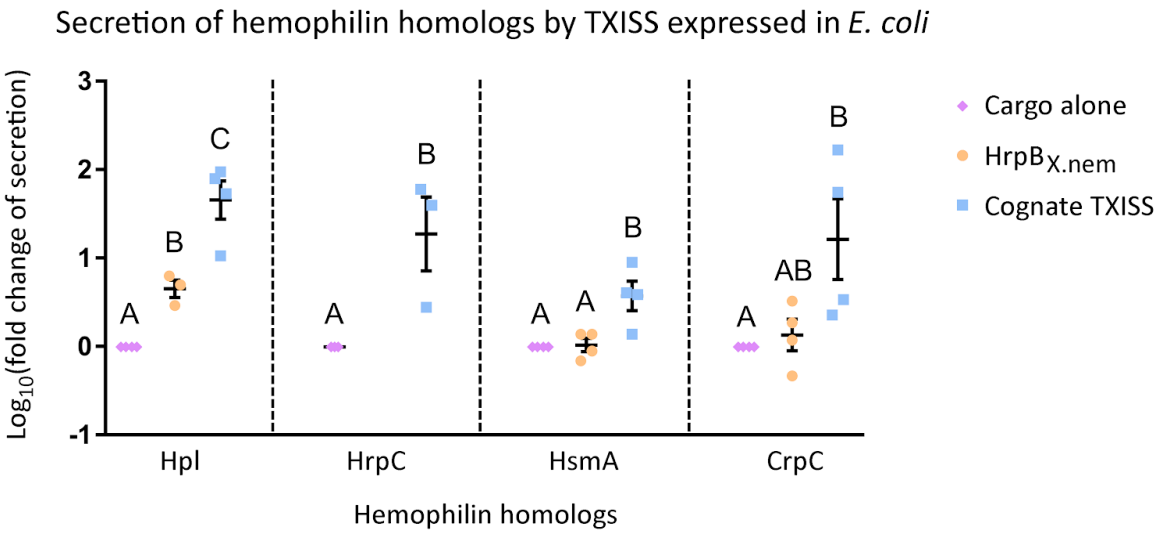


Figure 3.7. **Secretion of hemophilin proteins by their cognate and non-cognate T11SS proteins.** Each cargo protein was co-expressed alongside their cognate T11SS, co-expressed alongside the non-cognate $HrpB_{X.nem}$, and expressed in isolation. The supernatant proteins were precipitated and quantified via Immuno-blotting. Fold change of secretion was determined by dividing the amount of extracellular cargo in co-expression treatments by the amount seen in the respective cargo alone treatment. Data were transformed with a $\log_{10}$ function prior to performing a Tukey's HSD test for each hemophilin homolog. Letters indicate significance groups. All four T11SS proteins assayed significantly increased the concentration of their cognate cargo in the extracellular milieu. The non-cognate T11SS significantly increased secretion for Hpl but not HsmA or CrpC, and when it did so secretion was much less effectively than when co-expressed with the cognate T11SS.

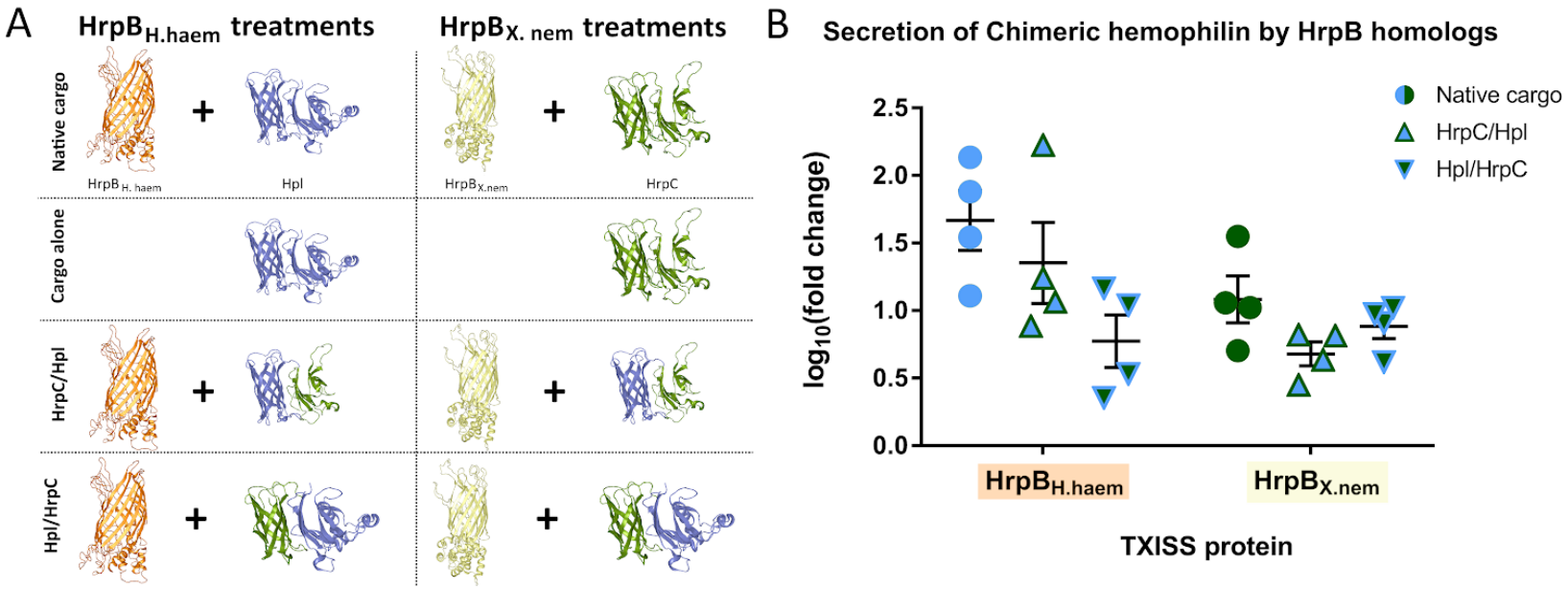


Figure 3.8. **Secretion of domain swapped chimeric hemophilin proteins by HrpB$_{X.nem}$ and HrpB$_{H.haem}$. A)** A pictorial depiction of the experimental treatments used to assess secretion of chimeric hemophilin by HrpB homologs from *H. haemolyticus* and *X. nematophila*. Each T11SS protein was co-expressed alongside its cognate cargo protein, and two chimeric cargo proteins generated by swapping the two domains of Hpl and HrpC. Fold change of secretion was determined by dividing the amount of extracellular cargo in co-expression treatments by the amount seen in the respective cargo alone treatment. **B)** Secretion data were made normal with a $\log_{10}$ transformation. Both chimeric hemophilin proteins were preferentially secreted by the T11SS which was cognate to their C-terminal domain, however neither chimera was as effectively secreted as the completely cognate hemophilin protein.

chimera ran as a single band at ~27 kD, which is larger than the predicted size of ~25 kD. The Hpl-HrpC chimera ran as two bands with apparent sizes of ~33 kD and ~24 kD, which straddle the expected size of ~27 kD. Since Hpl-HrpC ran as two bands, similarly to Hpl, we opted to sum both bands for the purpose of enumeration. Co-expression with $HrpB_{H.haem}$ increased on average, the extracellular levels of its cognate cargo Hpl by 65.0-fold, the HrpC-Hpl chimera by 51.0-fold, and the Hpl-HrpC chimera by 7.8-fold, indicating that the presence of the cognate C-terminal domain of Hpl was optimal for secretion by $HrpB_{H.haem}$. Co-expression with $HrpB_{X.nem}$ increased on average, the extracellular levels of the cognate HrpC by 15.6-fold, the HrpC-Hpl chimera by 5.1-fold, and the Hpl-HrpC chimera by 8.1-fold on average (Fig. 3.8B; Chapter3WesternImages.xlsx). Overall, both the presence of the C-terminal domain of the cognate cargo, relative to the non-cognate cargo, resulted in higher levels of secreted chimeric hemophilin homolog protein. This finding reinforces the concept that T11SS cargo are two-domain proteins: An N-terminal effector domain and a C-terminal β-barrel domain that directs cargo for T11SS-mediated secretion (Grossman, Escobar, et al. 2022). Our findings additionally implicate the C-terminal β-barrel domain in mediating specificity of cargo for particular T11SS OMPs.

Physicochemical characterization of purified hemophilin homologs

The N-terminal effector domains of known T11SS cargo proteins vary widely in their structure and function (Grossman, Escobar, et al. 2022; Grossman, Mauer, et al. 2022). As described above, the N-terminal region of *Xenorhabdus* hemophilin homologs has conservation interrupted by regions of variability that may suggest diversification of ligand binding among members of this family. To begin to investigate the ligand binding properties of hemophilin homologs, Hpl from *H. haemolyticus*, HrpC from *X. nematophila*, HphA from *A. baumannii*, and CrpC from *X. cabanillasii* were expressed without signal peptides in the cytoplasm of *E. coli* and purified. Hpl from *H. haemolyticus* was recovered from *E. coli* cytoplasm as an approximately 50:50 mix of heme-bound and heme-free

protein, with the level of heme saturation likely reflecting competition for heme binding and limitations of heme biosynthesis *in vivo* (Latham et al. 2020). Preparations of HrpC (*X. nematophila*) and HphA (*A. baumannii*) had a brownish appearance and an absorbance peak at ~413 nm, consistent with the presence of sub-saturating levels of a porphyrin ligand, whereas CrpC (*X. cabanillasii)* was colorless with no peaks in the visible absorption spectrum, indicating the lack of a porphyrin. Heme-free (apo-protein) preparations of Hpl, HphA, and HrpC were produced by acid-acetone extract and reversed-phase HPLC, and porphyrin-binding affinities were determined by titration (Fig. 3.9; Table 1). Similar large changes in the UV-visible spectrum of hemin occurred upon titration with *H. haemolyticus* Hpl, *A. baumannii* HphA or *X. nematophila* HrpC (Fig. 3.9AC). In particular, similarities in the Soret (412–414 nm) and Q-band regions (500–600 nm) between HrpC and HphA suggest that the heme coordination structure of these two proteins is similar. The binding curves for Hpl, HphA, and HrpC (Fig. 3.9AC) were close to linear, indicating that binding was too strong to reliably extract binding constants. On the other hand, spectral changes upon addition of CrpC to heme were more gradual (Fig. 3.9D) and were fit with a one-to-one binding model (apparent K_d ~4.7 μM). These values are similar to the binding affinity we determined for BSA (apparent K_d ~1.5 μM; Fig. 3.10), and are substantially weaker than, for example, human serum albumin affinity for heme (Kd ~11 nM) (Adams and Berman 1980).

To quantify the porphyrin binding affinities of Hpl, HphA, and HrpC, we assayed binding to Zn(II)-PPIX (a fluorescent heme analog) in the presence of competition from excess BSA (Fig. 3.9F-J; Fig. 3.10). These experiments yielded apparent dissociation constants shown in Table 1, assuming the affinity of BSA for Zn(II)-PPIX to be the same as for Fe(III)-PPIX (1.5 μM; Fig. 3.10). The affinities of Hpl (0.8 nM) and HphA (1.8 nM) for Zn(II)-PPIX were not significantly different within error of the measurements (Table 1). The affinity of *X. nematophila* HrpC for Zn(II)-PPIX was approximately 10-fold weaker (K_d ~11 nM). To investigate the contribution of the

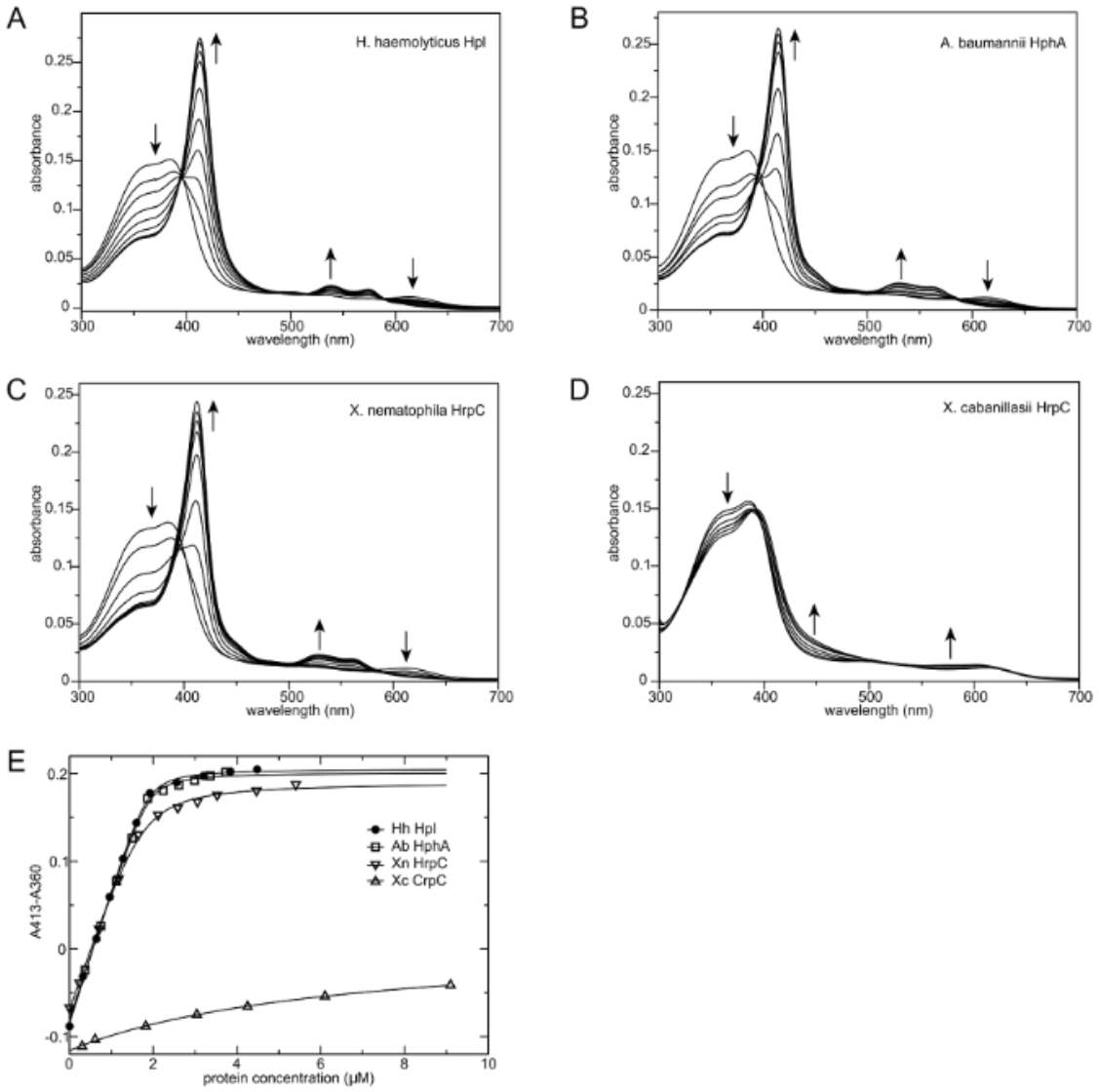


Figure 3.9 **Heme binding properties of hemophilin homologs.** (**A–D**) Titrations of hemophilin proteins into Fe-PPIX (hemin) solution (2.5 μM porphyrin in 20 mM Tris.HCl, pH 8.0 at 21°C) were monitored by UV-visible absorption spectroscopy. Arrows indicate the direction of spectral changes. (**E**) Spectral change, derived from A–D, is plotted as a function of protein concentration. Data points (symbols) were fit with a 1:1 binding model accounting for ligand depletion (solid lines). (**F–H**) Titrations of protein into Zn-PPIX solution (0.5 mM porphyrin in 20 mM Tris.HCl, pH 8.0 at 21°C supplemented with 1 mg/mL BSA) were monitored by fluorescence spectroscopy. (**J**) Titration of CrpC into Zn-PPIX (1.0 mM porphyrin). Data points (+) were fit with a 1:1 binding model accounting for ligand depletion to yield equilibrium dissociation constants shown in table 1.

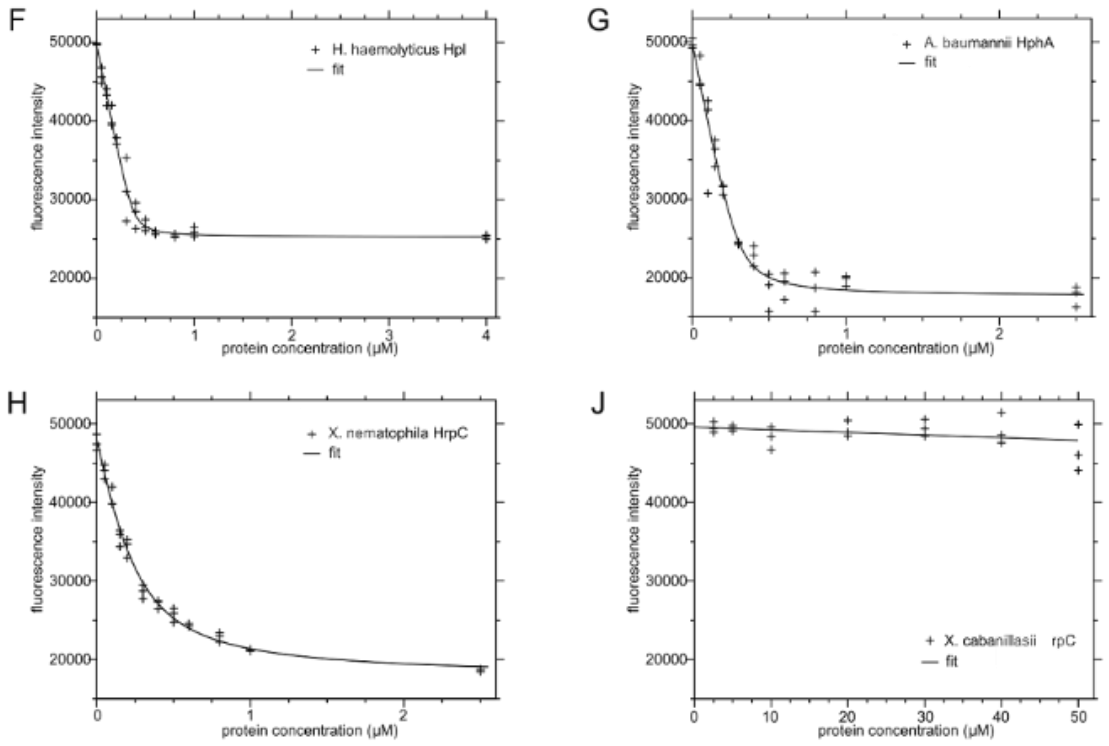


Figure 3.9. (continued)

Table 1. **Relative porphyrin binding affinity of hemophilin homologs**

Protein	Kd Heme	Zn(II)-PPIX $K_{d,app}$ (nM) (95% CI)	PPIX $K_{d,app}$ (nM) (95% CI)
HrpC	N/A	11 (6.5 – 17.6) *	18 (10 – 26) †
Hpl	N/A	0.8 (0.1 – 2.3) *	163 (90 – 250) †
HphA	N/A	1.8 (0.2 – 5.3) *	88 (32 – 202) †
CrpC	7,100 †	N/A	4,700 †

* Determined by Fluorescence spectroscopy in competition with BSA (in 30-fold molar excess of heme) assuming apparent *K*d = 1.5 μM for BSA binding to heme (Fig. 3.9)

† Determined by UV-visible spectroscopy

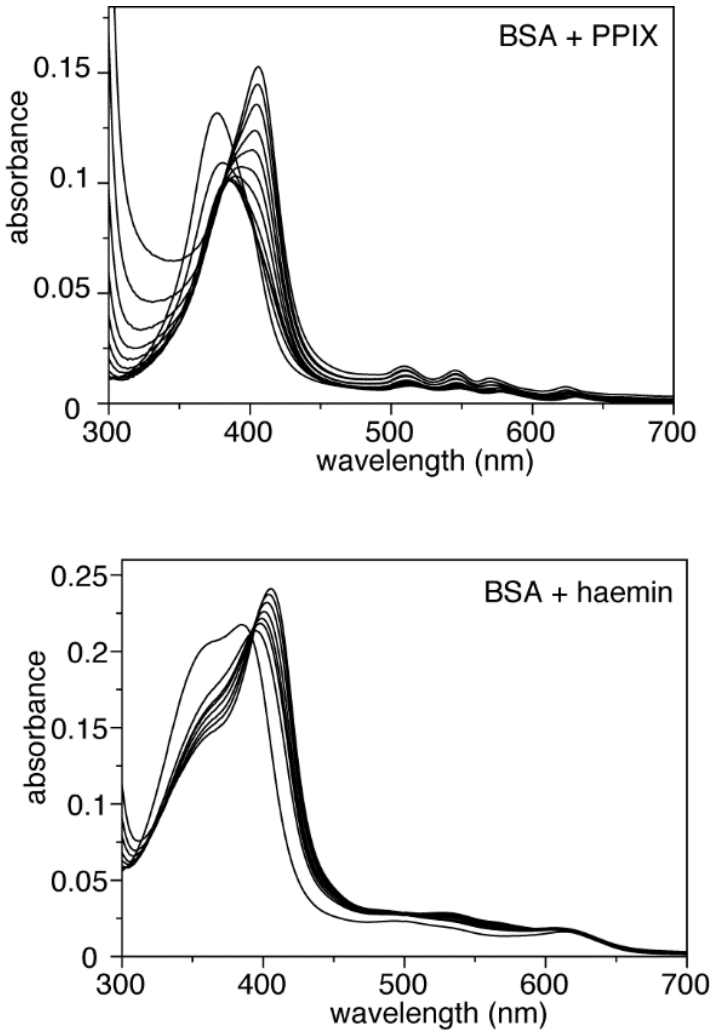


Figure 3.10. **UV-visible spectra of porphyrin in the presence of bovine serum albumin (BSA)**. BSA was titrated into protoporphyrin IX (PPIX) or Fe(III)-PPIX (hemin). Curves are shown for PPIX (1.5 μM) with BSA at the following concentrations: 0, 0.16, 0.48, 0.97, 1.6, 3.2, 6.0, 9.6, 16, 29, 55, 92 μM. Curves are shown for Fe(III)-PPIX (1.5 μM) with BSA at the following concentrations: 0, 1, 2, 3, 5, 9, 13, 21 μM.

porphyrin metal to binding, we performed titrations with unmetallated PPIX, monitored by UV-visible spectroscopy (Fig. 3.11). Hpl and HphA bound to unmetallated PPIX with similar affinity (*K*d ~100 nM; Table 1), which was ~2 orders of magnitude weaker than binding to metallated Zn(II)-PPIX. *X. nematophila* HrpC appeared to bind metallated and unmetallated porphyrin with similar affinities (10–20 nM). CrpC showed weak binding to PPIX, similar to Fe(III)-PPIX/hemin (K_d values ~5 μM; Table 1). These affinities suggesting that Hpl, HphA, and HrpC might effectively scavenge metallated or unmetallated porphyrins from the environment, whereas CrpC has only weakly detectible heme binding, similar to that of BSA.

To investigate if CrpC might bind selectively to porphyrins with different physicochemical properties we screened for interactions with coproporphyrin III (copro III), uroporphyrin I ethyl ester, and the linear tetrapyrrole, BV, a natural breakdown product of hemin, by UV-visible spectroscopy. For reference, spectra of Hpl or CrpC bound to PPIX, Fe(III)-PPIX or Zn(II)-PPIX are shown under comparable experimental conditions (Fig. 3.12). Addition of Hpl to PPIX, Zn(II)-PPIX, Fe(III)-PPIX caused large spectral peak shifts consistent with stoichiometric binding. By comparison, minor changes in absorption peak wavelengths accompanied addition of CrpC to PPIX, Zn(II)-PPIX and Fe(III)-PPIX reflecting the weak binding characterized in titration studies. Copro III contains four propionate groups and is therefore larger and more polar than PPIX. Weak binding of Hpl to copro III was indicated by small spectroscopic shifts, whereas no binding was evident with CrpC. Similarly, small spectral changes indicated some propensity for Hpl, but not CrpC, to bind to the linear tetrapyrrole, BV. Neither Hpl nor CrpC showed any indication of binding to uroporphyrin I ethyl ester. Compared to PPIX, this is a larger amphipathic porphyrin with four alkylated propionates. Given that bioinformatics placed CrpC in a cobalt/molybdenum associated cluster, we also tested spectroscopically for binding to cobalamin, Co(II) chloride, and

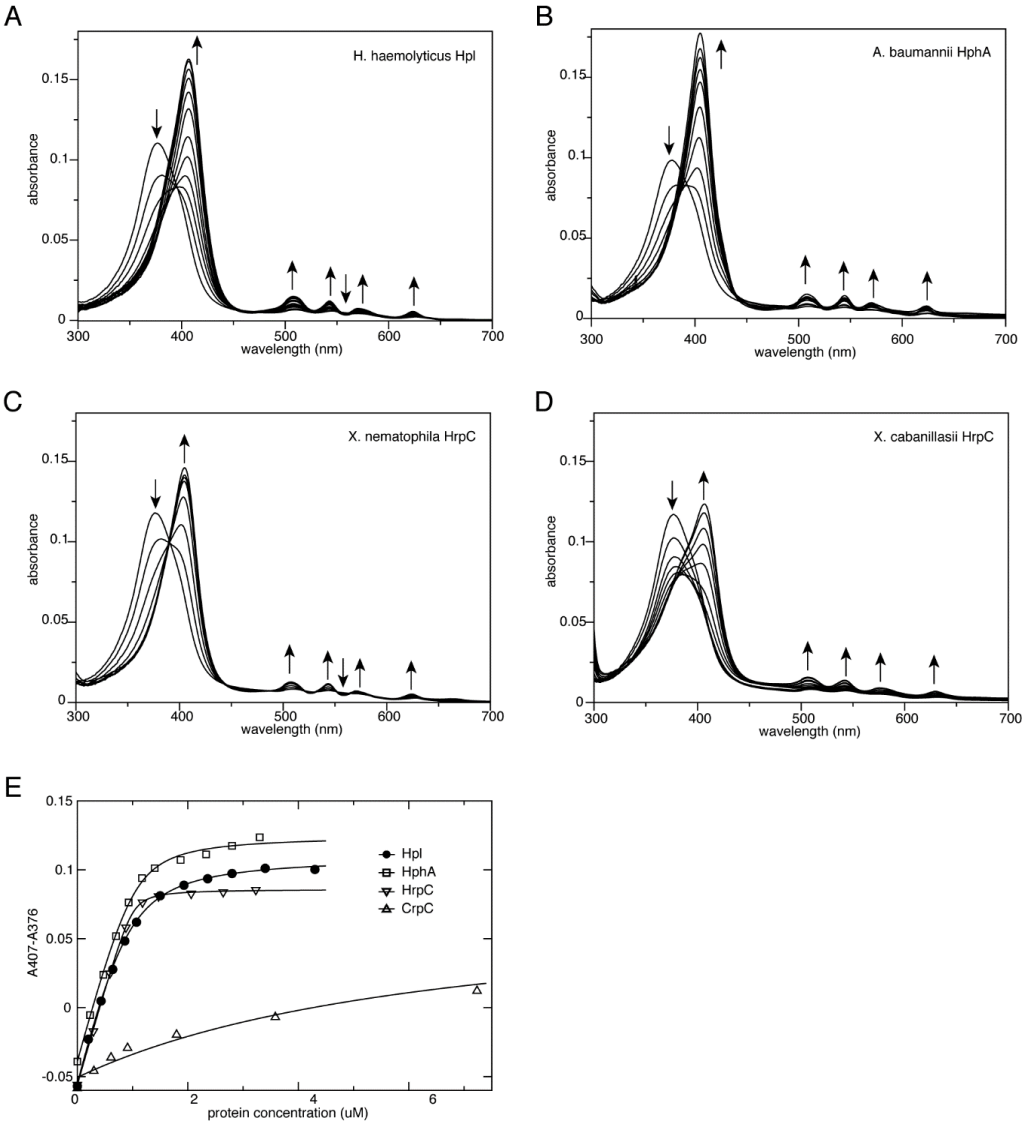


Figure 3.11. **Protoporphyrin IX binding properties of hemophilin homologs.** (A–D) Titrations of Hpl/HrpC proteins into protoporphyrin IX (PPIX) solution (1.5 μM porphyrin in 20 mM Tris-HCl, pH 8.0 at 21°C) were monitored by UV-visible absorption spectroscopy. Arrows indicate the direction of spectral changes. (E) Spectral change, derived from A–D, is plotted as a function of protein concentration. Data points (symbols) were fit with a 1:1 binding model accounting for ligand depletion (solid lines) to yield equilibrium dissociation constants shown in table 1.

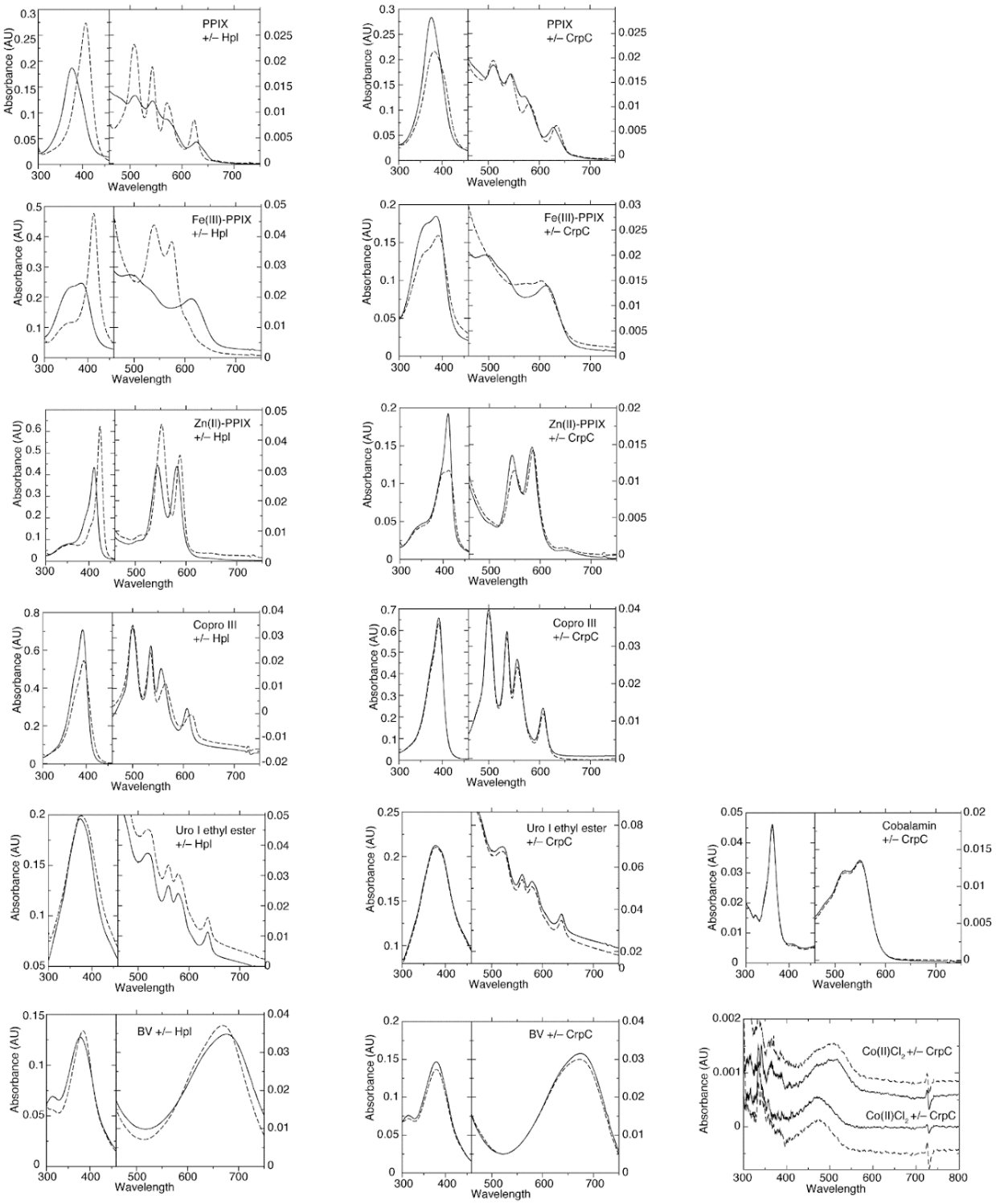


Figure 3.12. **Propensity of Hpl and CrpC to bind to selected linear and cyclic tetrapyrroles.** UV-visible spectra of selected tetrapyrroles (4 μM) in the absence (solid lines) or presence of Hpl (left panels, broken lines) or CrpC (right panels, broken lines) at a concentration of 6 μM. The buffer condition was 20 mM Tris.HCl, pH 8.0 (21°C).

hexamine Co(III) chloride. No spectroscopic changes were induced upon mixing CrpC with these reagents, thus no specific high-affinity ligand for CrpC was identified.

To investigate a structural basis for ligand binding in HrpC from *X. nematophila* and CrpC from *X. cabanillasii* we generated structural models using the program MODELLER (Eswar et al. 2006). Models were based on the crystal structures of Hpl and HphA (Bateman et al. 2021; Latham et al. 2020) and alignment to protein sequences with similarity, detectable by BLASTP (Altschul et al. 1990, 1997), specifically to the N-terminal α/β sub-domains of HphA and Hpl responsible for ligand binding.

Before describing conclusions from protein modeling, we briefly identify some salient features of Hpl and HphA. Comparison of Hpl and HphA crystal structures with other T11SS cargo proteins identifies a structurally conserved region that comprises β6-β7-β8 strands of Hpl and HphA together with their C-terminal β-barrels. Strands β6-8 are contiguous with the C-terminal β-barrel domain and form a β-sheet that packs against the β-barrel. The C-terminal 8-stranded β-barrel has meander topology with a shear value 8 and a tightly packed hydrophobic core. The combination of these features is unique to T11SS cargo proteins including Hpl and HphA, which bind heme, and HpuA, TbpB, and fHbp, which bind to protein targets (Latham et al. 2020). This sheet-barrel structure presumably forms an ancestral scaffold onto which different structural features have been molded to achieve binding to a diverse range of ligands (proteins and small molecules).

Hemophilin family proteins show a high level of sequence conservation within the sheet-barrel region described above, giving a high level of confidence in MODELLER predictions across this region. Further, modelling suggests that the complete N-terminal domains of HrpC and CrpC adopt the same β1-β2-β3-α-β4-β5-β6-β7-β8 topology as seen in Hpl and HphA (Fig. 3.13). Both Hpl and HphA coordinate heme on the β-face of the porphyrin through a histidine residue in a conserved position, 5-residues N-terminal to the sheet-barrel scaffold. However, the mode of heme

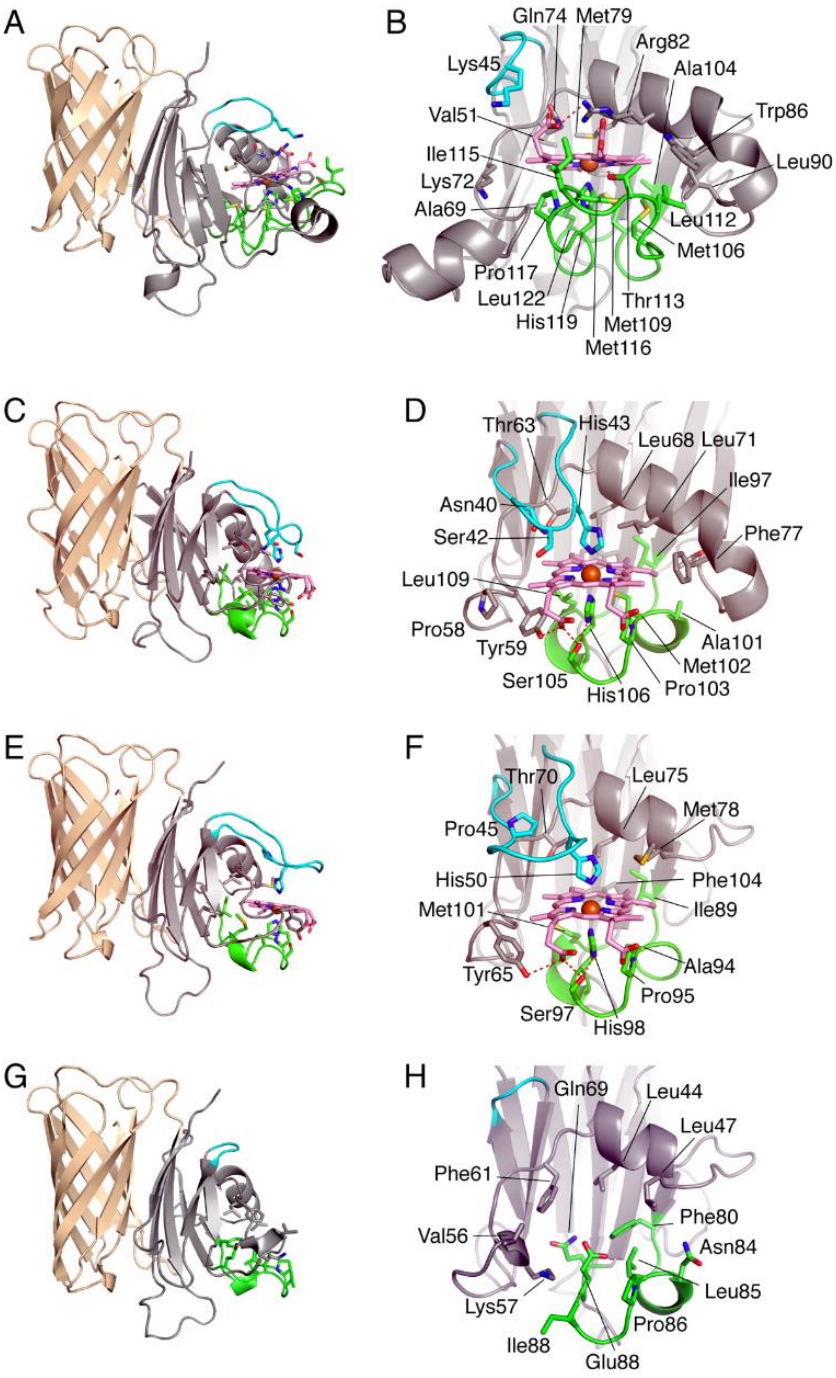


Figure 3.13. **Structural modeling of HrpC and CrpC**. Comparison of the heme-binding sites of Hpl (A, B; pdb 6om5; ref 19) and HphA (C, D; pdb 7red; ref 12) with structural models of X. nematophila HrpC (E, F) and X. cabanillasii CrpC (G, H) produced using the program, MODELLER. Panels A, C, E, G illustrate the overall structure with N-terminal α/β domain (grey/cyan/green) and C-terminal 8-strand β-barrel (brown). Panels B, D, F, H show detail of the known or predicted porphyrin binding site. The β2-β3 loop (cyan) and β5-β6 loop (green) are described in the main text. The heme porphyrin and central iron atom are shown in pink and orange, respectively.

binding between Hpl and HphA differs in several important respects. First, the β5-β6 loop carrying the β-facial His is substantially longer in Hpl compared to HphA (20 vs 14 residues), and more completely covers the porphyrin face in Hpl. Second, ferric heme in Hpl is 5-coordinate (although a weak 6th ligand in the form of a water molecule or Cl^- ion may be present), whereas HphA has a six-coordinate, bis-histidyl, ferric heme center with the second axial His ligand donated by the β2-β3 loop. Third, the β2-β3 loop in HphA forms a 'lid' over the porphyrin α-face, and a structure of the apo-protein suggest that the β2-3 loop unfolds to accommodate binding and dissociation of the heme. In Hpl the β2-β3 loop is shorter (7 vs 11 residues) with limited heme contacts. Instead, residues from the α-helix in the β3-β4 loop make more extensive contacts with heme in Hpl. A consequence of these differences is that the heme iron sits between the β3-β4 α-helix and β5-β6 loop in Hpl, but between the β2-β3 and β5-6 loops in HphA (Fig. 3.11, compare A, B with C, D). Hpl also features a cavity between the β3-β4 α-helix and the heme iron that can accommodate a range of diatomic heme ligands in the Fe(II) and Fe(III) oxidation sates.

Modeling suggests that the heme environment in HrpC (*X. nematophila*) is similar to that in HphA (*A. baumannii*), including bis-histidyl heme ligands (His49 and His97 in HrpC), hydrogen bond network (comprising Tyr64, the porphyrin 17-propionate, Ser96 and His97) and non-polar contacts with the porphyrin ring contributed by the β3-4 α-helix, β5-6 loop and underlying β6-7-8 sheet (Fig. 3.13, compare C, D with E, F). Notable differences include a shorter β3-4 α-helix than either HphA or Hpl and a β4-5 hairpin that is more similar to Hpl. The pyrrole rings B and C appear to be more exposed to solvent in HrpC than in Hpl or HphA, which may partly explain a lower affinity for hemin and Zn(II)-PPIX because hydration increases the rate of scission of the Fe-His bond. Possible reasons for higher binding affinity to unmetallated PPIX are not evident.

Overall CrpC is predicted to have a structure similar to HrpC, but the heme-ligating His residues from HrpC are missing in CrpC. The β-facial His (His97 in HrpC) is replaced by glutamate

in CrpC. Only one heme binding protein, the terminal oxidase cytochrome bd, is known to position a glutamate side chain within coordinate bonding distance of heme iron (Safarian et al. 2016), likely connected to a unique structural solution to oxygen reduction, and in this case glutamate does not appear to contribute substantially to heme affinity (Murali and Gennis 2018). The β2-3 loop, which carries the second heme-ligating His in HrpC is reduced to a three-residue hairpin and is too small to contribute to a porphyrin pocket of the kind seen in other hemophilin homologs. The β3-4 α-helix also lacks bulky side chains that can contribute pi-stacking interactions with the porphyrin as seen in Hpl (Arg82 and Trp86). In summary, models of CrpC suggest that two important heme-binding features, coordination of the heme iron and extensive contact area on both sides of the porphyrin plane, are absent.

Transcriptional reporters for HrpB and HrpC in *X. nematophila*

Having established that *X. nematophila* HrpC is capable of binding heme, we predicted that the expression of the *X. nematophila hprC* gene and its cognate T11SS-encoding gene $hrpB_{X.nem.}$, may be regulated by metal availability or other nutritional conditions. To test this, transcriptional reporters were constructed in *X. nematophila* by incorporating a β-galactosidase-encoding gene into the genomic positions of *hrpC* or $hrpB_{X.nem.}$ These reporter strains were exposed to varying nutrient conditions (glucose minimal media vs LB) or conditions representing low metal (250μM 2,2'-bipyridyl) or high heme (23μM hemin chloride). β-galactosidase assays revealed that P_{hrpC}-*lacZ* and P_{hrpB}-*lacZ* are undetectable when cells are grown in glucose minimal medium, and neither hemin supplementation nor metal starvation induced expression under these conditions (Fig. 3.14). In nutrient rich, LB medium, P_{hrpC}-*lacZ* and P_{hrpB}-*lacZ* were similarly undetectable under neutral conditions and in the presence of hemin. However, a significant increase in P_{hrpC} activity (59.5-fold) and P_{hrpB} activity (1.2-fold) occurred when cells were starved of divalent cations by 2,2'-bipyridyl treatment in LB medium (Fig. 3.14).

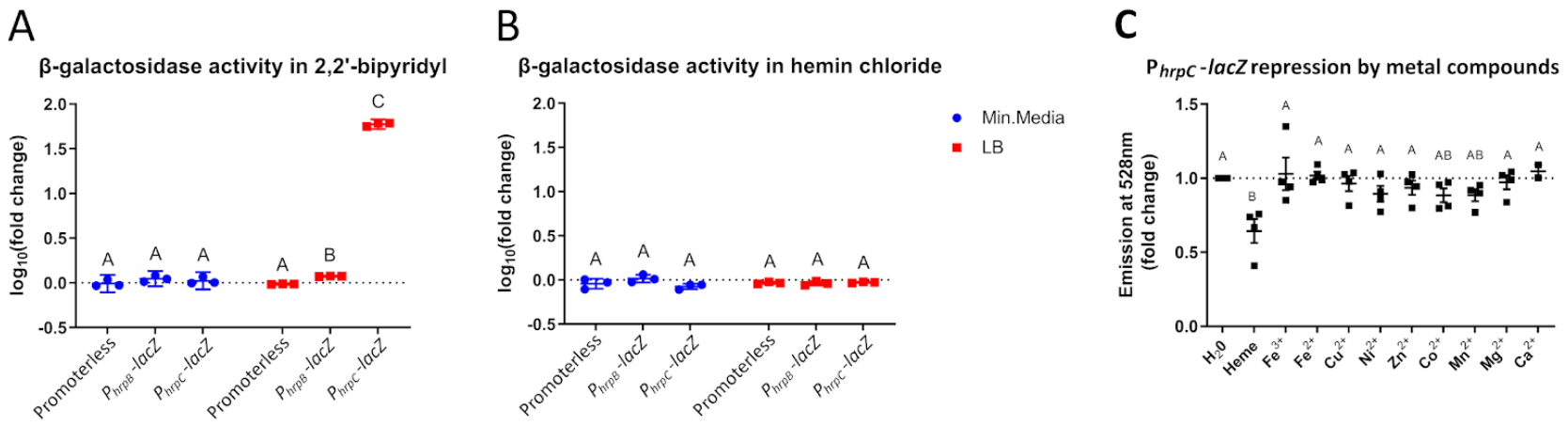


Figure 3.14. **Transcriptional expression from P_{hrpB} and P_{hrpC} in *X. nematophila*.** The *lacZ* gene was integrated into the *X. nematophila* genome in the positions of *hrpB* and *hrpC* to use expression of β-galactosidase as a proxy for promoter transcriptional activity. Letters above the data points indicate significantly different groups according to one-way ANOVA with Tukey's HSD Test. **A-B)** Reporter experiments were performed in both glucose minimal medium and nutrient rich LB medium. β-galactosidase activity in the presence of 2,2'-bypyridyl and hemin chloride was divided by activity in a vehicle control to find fold change of activity. Expression from P_{hrpB} and P_{hrpC} was only significantly impacted in nutrient rich LB media supplemented with 2,2'-bypyridyl. **C)** The P_{hrpC} reporter strain was grown under metal starvation and then exposed to a series of metallic compounds to determine if any restored normal expression levels. Of the compounds tested, only hemin chloride significantly reduced P_{hrpC} activity.

Since 2,2'-bipyridyl is a non-specific chelator of all divalent cations a follow-up experiment was performed to determine which metal compounds most effectively repressed the chelation-induced increase in expression of P_{hrpC}-*lacZ*. The transcriptional reporter strain was starved using 2,2'-bypyridyl in dark LB, then rinsed and exposed to dark LB containing 100μM concentrations of each divalent cation. Subsequent β-galactosidase activity indicated that only hemin significantly prevented chelation-induced P_{hrpC}-*lacZ* expression (0.64-fold), while neither ferric (1.03-fold) nor ferrous (1.02-fold) iron suppressed P_{hrpC}-*lacZ* detectably. These data indicate that hemophilin expression in *X. nematophila* responds to metal starvation generally, and that one aspect of this regulation involves transcriptional repression in response to high concentrations (100 μM) of environmental heme.

Tn5 mutagenesis reveals potential regulators of P_{hrpB} and P_{hrpC} activity

To reveal potential regulatory proteins contributing to the expression of hemophilin and its cognate T11SS within *X. nematophila*, we generated random mutagenesis libraries of the *hrpB* and *hrpC* reporter strains using a mini-Tn*5* transposon. The mini-Tn*5* transposon present in pBSL118 was conjugated into each strain and exconjugants were screened for changes in reporter expression using blue/white screening on X-gal (Alexeyev, Shokolenko, and Croughan 1995). Colonies that were optically more or less blue than unmutated reporter controls were selected for further analysis. Of the 190 exconjugants from each reporter strain that were assayed, no exconjugants isolated from P_{hrpB}-*lacZ* detectably altered expression levels. In the P_{hrpC}-*lacZ* mutant library we identified 6 isolates with drastically altered expression levels (0.29-fold minimum; 2.11-fold maximum). The Mini-Tn5 cassettes were then subcloned into pUC19 and sequenced to identify where they had integrated into the genome.

Mapping of insertion sites revealed that a mutant with 2.11-fold increased P_{hrpB}-*lacZ* activity relative to wild type had integrated into the NADPH-dependent oxidoreductase *ygfF* (XNC1_4576). Intriguingly, this enzyme family frequently co-occurs with hemophilin homologs in the HrpC-like

subcluster and may be responsive to metal concentration within the cell. Three mutants with decreased expression of P_{hrpB}-*lacZ*, (0.29-fold, 0.32-fold, and 0.39-fold) relative to wild type had insertions in *pgaC* (XNC1_2839), *btuF* (XNC1_0087), and a hypothetical gene (XNC1_0295) respectively. In *E. coli, pgaC* encodes a glycosyl transferase essential for the generation of extracellular poly *N*-acetyl glucosamine (Butland et al. 2005). This gene is also essential for pathogenesis in *Y. pestis* where it is somewhat misleadingly annotated as hemin storage protein R (*hmsR*). In *Y. pestis*, the *hms* genes are essential to block the proventriculus region of the flea gut via the synthesis of exopolysaccharide biofilms, which in turn forces their flea vectors to bite mammalian host more often and to regurgitate bacteria during blood meals (Hinnebusch, Perry, and Schwan 1996; Lillard et al. 1999). Vitamin B12 utilization F (*btuF*) encodes a periplasmic cyanocobalamin binding protein essential for import and is a homolog of the periplasmic heme transporter *hmuT* (53% similarity; 30% identity in *E. coli*) (Van Bibber et al. 1999; Borths et al. 2002). Nothing is currently known about the 73 amino acid hypothetical protein XNC1_0295, however using BLAST and PSI-BLAST reveals that it is entirely specific to *X. nematophila* strains (Altschul et al. 1990, 1997). The two remaining Mini-Tn5 cassettes did not carry enough genomic context with them after subcloning to identify the genomic region they had inserted into (15bp each).

Impact of *hrpB* deletion in *X. nematophila*

We have established that the *X. nematophila* HrpC protein can bind PPIX, that its expression is influenced by hemin availability, and that its secretion into the extracellular milieu is facilitated by the T11SS OMP, HrpB. To examine the function of this system in *X. nematophila* biology, we created, by allelic exchange, an insertion-deletion null *hrpB* mutant (*hrpB*::*kan*R). Since the *hrpB* gene lies upstream of genes predicted to encode a TonB-dependent receptor and a TonB homolog, we accounted for possible polar effects by creating a complementation strain in which all three genes were integrated at the *att*Tn*7* site of the *X. nematophila* chromosome. GFP-expressing versions of

both the mutant and the complemented strain were constructed to facilitate identification of bacteria in fluorescent microscopy imaging.

Growth curves were performed to determine if *hrpB* is essential to growth under nutrient replete conditions and under metal starved conditions caused by 2,2'-bypyridyl supplementation (Constable and Housecroft 2019). In nutrient-replete LB, both the mutant and complemented strains were indistinguishable from wild type (Fig. 3.15A). However, in metal-starved conditions, the *hrpB* mutant displayed a sizable growth defect that was completely rescued in the complemented strain (Fig. 3.15B). To see if this growth defect was specific to hemin starvation the growth curve was performed in LB medium supplemented with both 2,2'-bypyridyl and hemin chloride. The presence of excess hemin completely rescued the growth defect of the *hrpB* mutant, demonstrating that *hrpB* is essential for growth under nutrient rich, but heme-starved conditions *in vitro* (Fig. 3.15C).

The secretion of a heme-binding protein in the extracellular milieu is expected to sequester heme away from other organisms that do not encode an appropriate uptake system (Atto et al. 2020; Latham et al. 2020). To determine if HrpB-dependent secretion of hemophilin impacts the growth of competing microbes we tested the impact of *X. nematophila* cell free conditioned medium (henceforth CFCM) on the growth of three strains of bacteria likely to be encountered by *X. nematophila* in its natural environment. To generate the conditioned media, stationary phase *X. nematophila* wild type*, hrpB* mutant, and complement strain bacteria were incubated in fresh 2,2'-bypyridyl supplemented LB to maximize expression of HrpC without completely depleting the nutrients in the medium. We tested two different strains of *X. nematophila*, ATCC19061 isolated from the nematode *S. carpocapsae* and *X. nematophila* Websteri isolated from *S. websteri*, both of which have an *hrp* locus, including the presumed TonB-dependent heme uptake system, and a strain of *E. faecalis* isolated from an *X. nematophila*-infected insect (Singh et al. 2014), which, as a Gram-positive bacterium, does not encode any TonB-dependent systems. The three bacterial strains were cultured

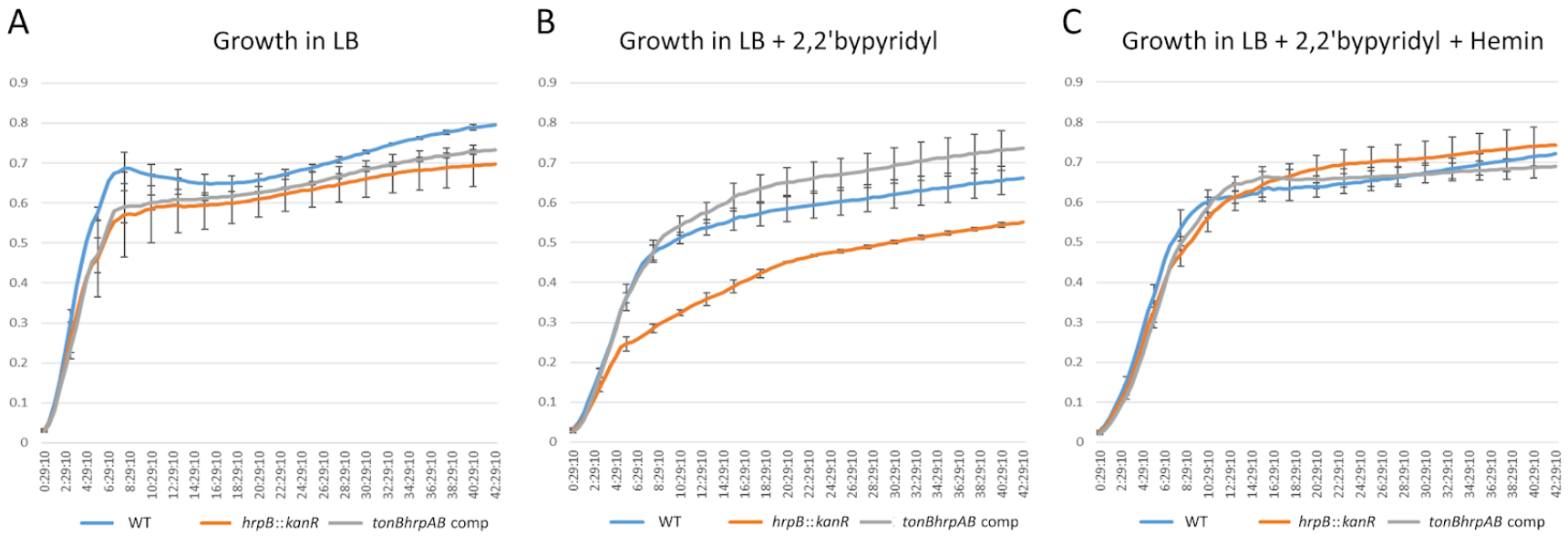


Figure 3.15. **Growth curve of *X. nematophila* wildtype, *hrpB*::*kanR*, and a *tonBhrpAB* complement strain with and without metal starvation**. Each strain was grown for 42.5 hours in LB media supplemented with either vehicle control A), 2,2'-bypyridyl B), or 2,2'-bypyridyl and hemin chloride C). Optical density at 600nm is reported on the y-axis and time is indicated on the x-axis. The *hrpB* mutant displayed a significant growth defect under metal starved conditions which could be rescued by complementation or with excess hemin in solution.

in the three types of CFCM and monitored for growth. When the resulting data were fit with a logarithmic growth curve, we found that the two *X. nematophila* strains behaved similarly; those grown in CFCM from wild type bacteria and the *hrpB* complement grew at similar rates and to similar densities, while the bacteria grown in CFCM from the *hrpB* mutant reached a lower final carrying capacity as indicated by OD_{600} (Fig. 3.16A-E). However, the difference in carrying capacity was only significant for *X. nematophila* Websteri. Conversely, the growth rate and carrying capacity of *E. faecalis* did not visibly or statistically differ between the three CFCM used (Fig. 3.16CF).

The *hrpB* mutant was subjected to a range of general phenotypic assays including protease activity, lipase activity, motility, and biofilm formation. Deletion of *hrpB* had no discernable effect on any of these general phenotypes (Fig. 3.17A-D). To determine if *hrpB* is essential for *X. nematophila* to function as an insect pathogen, a *Manduca sexta* (tobacco hornworm moth) injection model was used (Hussa and Goodrich-Blair 2012). Approximately 300 CFUs of wild type and *hrpB* mutant bacteria were injected into insects and mortality was monitored over several days. The *hrpB* mutant displayed no defect in insect virulence compared to wild type (Fig. 3.17E). These data suggest that hemophilin secretion is not essential for *X. nematophila* virulence, which is not unexpected given that *X. nematophila* encodes genes for heme biosynthesis, unlike *Haemophilus influenzae* and *H. haemolyticus* (Hariadi et al. 2015).

To determine if T11SS mediated hemophilin secretion impacted the fitness *S. carpocapsae*, the mutualistic nematode host of *X. nematophila*, we raised axenic nematodes alongside wildtype *X. nematophila*, the *hrpB* mutant, or the complemented strain under two separate conditions, on bacterial lawns and within an insect carcass. In both conditions nematodes grew and reproduced until they depleted their food sources, at which time they transitioned to a dauer-like state known as the infective juvenile (IJ) stage and emerged into a water trap. Nematodes were collected from water traps at 6, 12, and 18 days post-trapping. When tested in biological sextuplicate, *S. carpocapsae*

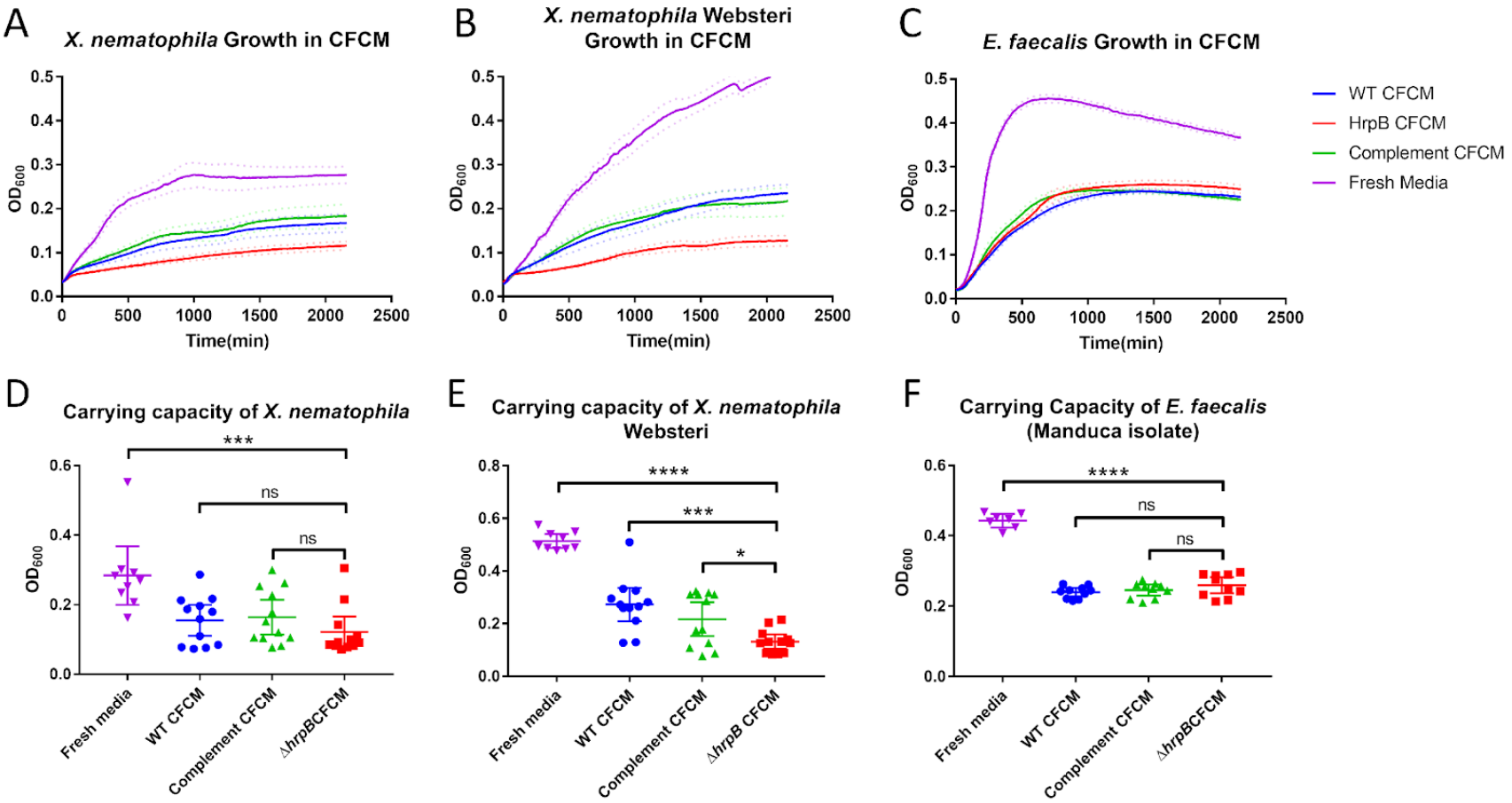


Figure 3.16. **HrpB facilitates the growth of conspecific *Xenorhabdus* strains under metal starvation and does not impact the growth of heterospecific *E. faecalis*.** A-C) Growth of *X. nematophila* ATCC19061, *X. nematophila* subspecies Websteri, and *E. faecalis* (Manduca isolate) respectively when cultured in metal starved CFCM over 36 hours in biological dodecaplicate. Dotted lines above and below the average indicate the standard error of the means depicted. D-F) Carrying capacities for the above growth curves as derived from the R package GrowthCurver. Asterisks indicate significant differences according to a Dunnett's test.

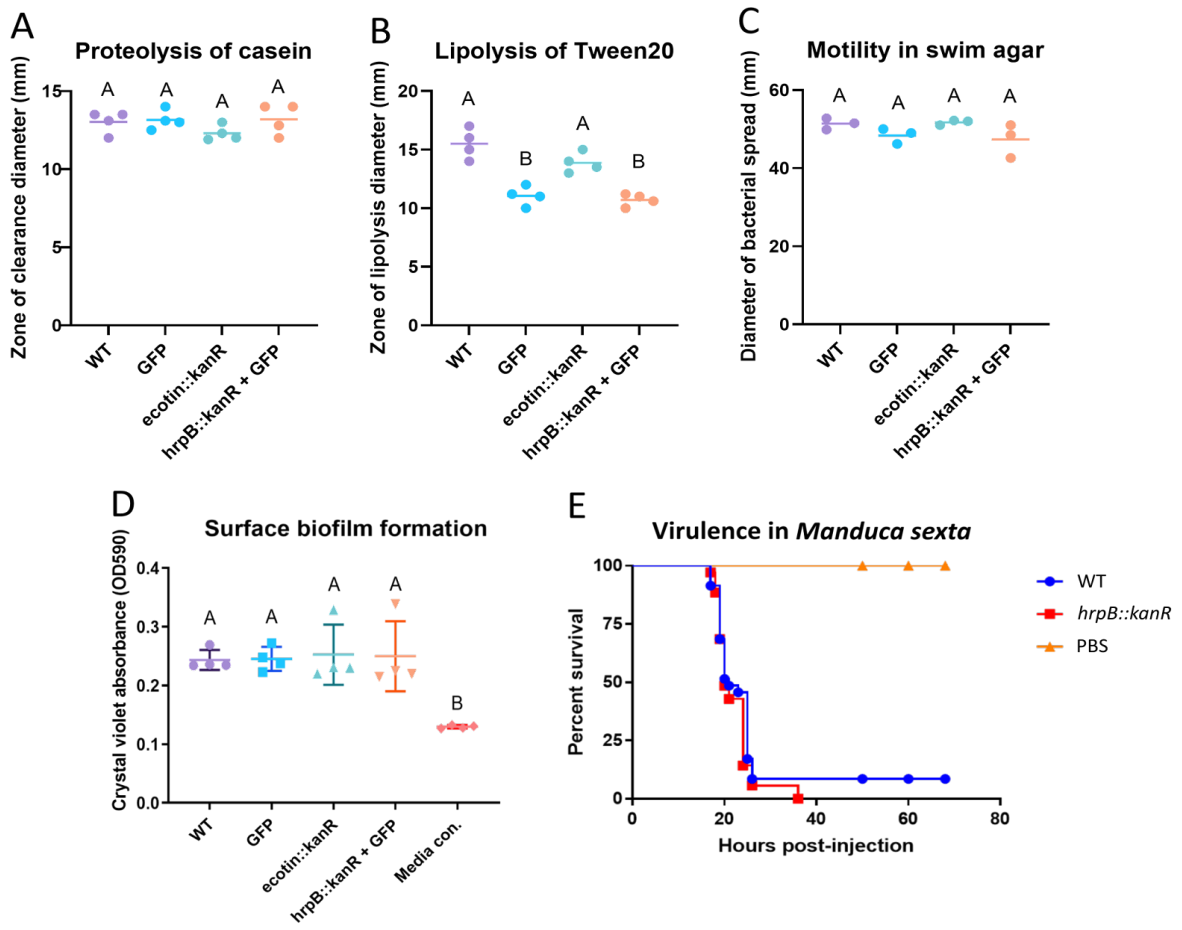


Figure 3.17. **General phenotypic characterization of the *X. nematophila hrpB*::*kanR* mutant**. For all general phenotypic assays a wildtype treatment, a GFP expressing treatment (*kefA*::GFP), and a T11SS-unrelated allelic exchange treatment (*ecotin*::*kanR*) acted as controls for genetic manipulation. **A)** Bacteria plated on powdered milk media in biological quadruplicate were monitored to determine if loss of *hrpB* impacted *X. nematophila*'s ability to degrade casein protein. **B)** Bacteria plated on Tween20 Lipase media in biological quadruplicate were monitored to determine if loss of *hrpB* impacted lipid degradation. Both GFP expressing strains showed decreased lipolytic activity, suggesting that GFP insertion or expression is responsible for the defect, rather than deletion of *hrpB*. **C)** Bacteria plated on low density swim agar in biological triplicate were monitored to determine if loss of *hrpB* impacted flagellar motility. **D)** Bacterial biofilms grown on polypropylene 96 well plates in biological quadruplicate were quantified via crystal violet staining and optical detection at 590nm. **E)** Bacterial strains were injected into 4th instar M. sexta insects in biological triplicate alongside a PBS control injection (giving a total of 35 insects per treatment). Then insect mortality was observed over approximately 60 hours post injection. Parts **A-D** were analyzed via one-way ANOVA with a Tukey's HSD Test. Part **E** was analyzed using a series of Mantel-cox tests. *X. nematophila hrpB*::*kanR* displayed no significant difference from appropriate controls in the above assays.

nematodes raised on the *hrpB* mutant produced significantly fewer IJs than those raised on wildtype bacteria (21.0%) or the complemented strain (20.1%) (Fig. 3.18A). Observation of the nematode growth plates revealed that the *hrpB* mutant lawn slowly accumulated a "mountainous topology" of mucoid mass absent from the wildtype and complemented lawns (Fig. 3.19A). Mucoid mass was quantified by crystal violet staining, showing that the *hrpB* mutant lawn produced more adherent material than *hrpB* encoding strains (Fig.3.19B). To test if the fitness defect was present *in insecta*, *S. carpocapsae* IJs were axenically colonized by one of the three bacterial genotypes, surface sterilized, and injected into *Galleria melonella* larvae (200 nematodes/insect) to track insect mortality and subsequent IJ emergence. Within the insect these nematodes release their bacterial symbionts and use them as their central food source (Mucci et al. 2022). Three insects were injected with each nematode biological replicate for a total of nine insects per bacterial treatment. All insects died within 20 hours without any significant difference in timing. No significant difference in nematode emergence was detected between treatments *in insecta*, suggesting that the insect carcass environment ameliorates the negative effects of not expressing *hrpB* seen on agar plates either as a function of its nutritional content or physical conditions (Fig.3.18B).

To determine if *hrpB* is essential for *X. nematophila* to colonize or persist within the intestines of its nematode host *S. carpocapsae*, axenic nematodes were raised on lawns of GFP expressing *X. nematophila*. Wildtype bacteria and an *ecotin*::*kan*R allelic exchange strain were used as controls for genetic manipulation. Infective juvenile nematodes were collected as they emerged from the lawns of bacteria. Frequency of colonization was assessed by counting the percentage of nematodes possessing GFP expressing bacteria within their intestines and density of colonization was assessed by homogenizing nematodes and plating for CFUs. To make sure that the bacteria had no defect for

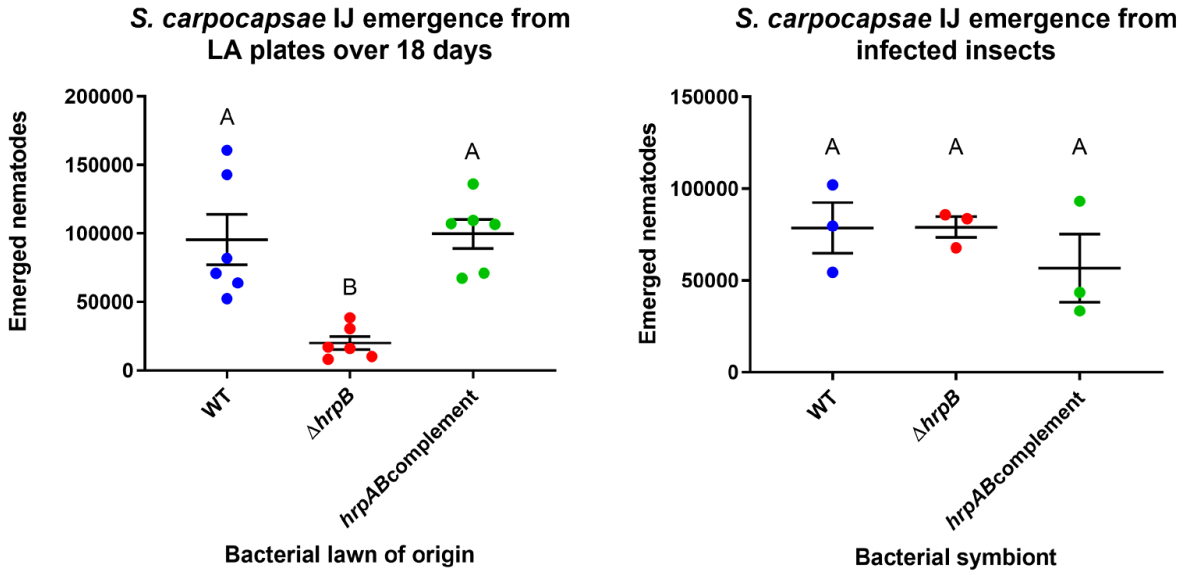


Figure 3.18. **HrpB is essential for nematode fitness *in vitro*, but does not detectably impact nematode fitness *in vivo*. A)** Axenic nematodes were reared on LA plates containing bacterial lawns differing in their ability to express HrpB, their progeny were collected over an 18-day period. The *hrpB*::*kan*R mutant supported significantly fewer progeny than either *hrpB* expressing strain. **B)** Nematodes carrying the same bacterial symbionts were injected into *Galleria mellonella* larvae and after the insects died the nematode progeny which emerged from the carcasses were collected over an 18-day period. The deletion of *hrpB* did not seem to have an impact on nematode fitness within the insect environment.

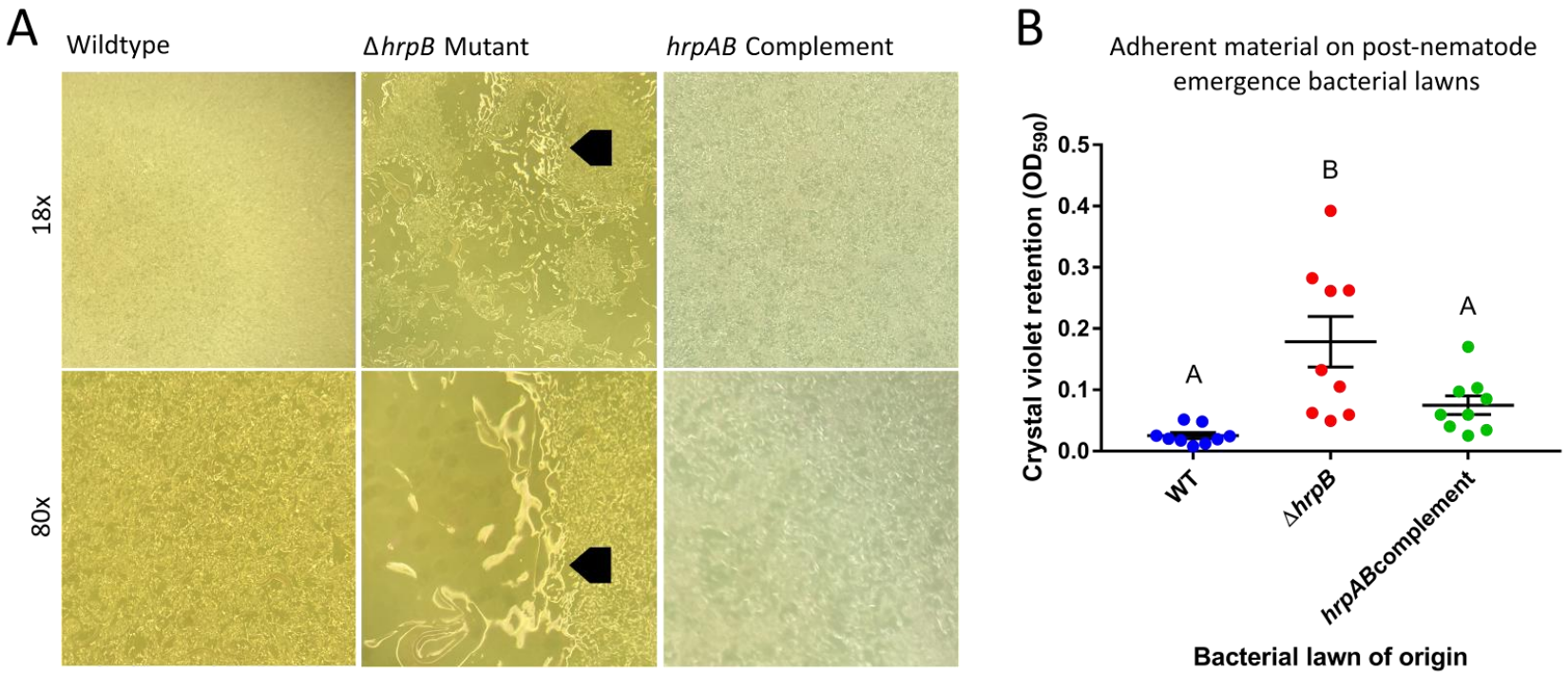


Figure 3.19. ***X. nematophila hrpB*::*kanR* lawns generate mucoid regions when cultured with *S. carpocapsae*. A)** 18x and 80x magnification images of bacterial lawns after 26 days of nematode grazing and emergence. Wildtype bacterial lawns, alongside the *hrpAB* complemented strain, have a very consistent density and texture, while the *hrpB* mutant has accumulated highly mucoid regions that resemble exopolysaccharide derived biofilms. These mucoid regions are indicated with a black arrow. **B)** The adherent material on these plates was collected on glass cover slips and quantified using crystal violet staining of the adherent material. The observed difference in mucoid mass was supported by an increase in adherent mass from the *hrpB* mutant lawns. Significance was determined using a Tukey's HSD test.

persistence with the host, frequency was assessed 3 days post-trapping and 3 weeks post-trapping. The *hrpB* mutant colonized nematodes at the same frequency and density as wild type *X. nematophila* and showed no defect for persistence over time (Fig. 3.20).

Discussion:

The hemophilin protein family presents a powerful tool for characterizing T11SS-dependent secretion because it is well conserved across diverse bacterial genera with distinct lifestyles. This includes the entomopathogenic bacteria of the genus *Xenorhabdus*, where hemophilin homologs are universally present, though the complement of homologs varies between species. Enough *Xenorhabdus* species have been sequenced that we can relatively confidently reconstruct the evolutionary trajectory of these hemophilin homologs, revealing that within this genus the ancestral T11SS (HrpB) was responsible for secreting hemophilin (HrpC). However, by examining the paralogues that evolved from this ancestral system we can also see that the family is diversifying its N-terminal domain to potentially gain distinct ligand binding activity. Conversely, by using the diversity present within the hemophilin protein family we were also able to demonstrate that T11SS specifically secrete their cognate cargo proteins and that this specificity is predominantly driven by the C-terminal domain. Additionally, while three of the hemophilin homologs we examined were capable of binding heme via similar domains, the function of these proteins varies according to the needs and lifestyle of the bacteria that encodes them. Hemophilin can act as a pathogenesis factor in *A. baumanii* and a tool for interbacterial competition in *H. haemolyticus*. Our data expand this functional range into mutualist interactions as well by demonstrating that the hemophilin expressed by *X. nematophila* can influence some aspect of nematode fitness.

These results support several previously hypothesized features of type eleven secretions systems by using hemophilin family proteins to demonstrate that T11SS secretors and their cargo

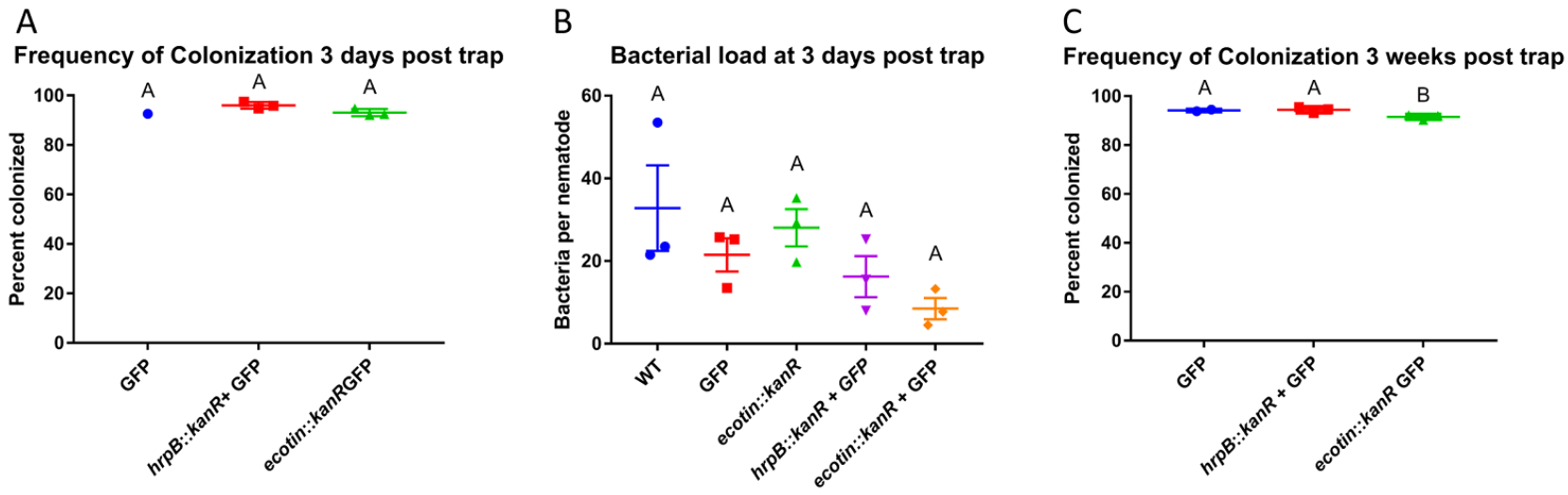


Figure 3.20. **Colonization of *S. carpocapsae* nematodes by *X. nematophila hrpB*::*kanR*.** Axenic *S. carpocapsae* eggs were plated on lawns of each bacterial treatment in biological triplicate and allowed to reproduce freely. Infective juvenile nematodes which emerged from these plates were assayed for frequency of colonization at 3 days post-trapping **(A)** and 3 weeks post-trapping **(C)** using fluorescent microscopy of green fluorescent protein. The percentage of nematodes colonized by bacteria was unaffected by *hrpB* deletion for both time points. Additionally, bacterial load was assessed at 3 days post-trapping by homogenizing a set number of nematodes and plating the homogenate to determine CFUs/nematode **(B)**. All three colonization assays were analyzed using one-way ANOVA with a Tukey's HSD Test. Deletion of *hrpB* had no impact on colonization efficacy, bacterial density per nematode, or persistence within the nematode intestines.

proteins exist as cognate pairs, T11SS-dependent cargo proteins have conserved roles in metal and metalloprotein binding, and T11SS loci seem to be subject to horizontal transmission between diverse proteobacteria species (Bhasin, Chaston, and Goodrich-Blair 2012; Cowles and Goodrich-Blair 2008; Grossman, Mauer, et al. 2022). These results also lend new nuance to our understanding by establishing that 1) T11SS specificity is not shared between cargo, even between homologs 2) the efficacy or rate of secretion varies between T11SS/cargo pairings, 3) the affinity of heme binding varies between hemophilin proteins, 4) effective secretion of hemophilin by *X. nematophila* contributes to host animal fitness, and 5) some members of the hemophilin family may be binding non-porphyrin substrates.

The high level of cognate specificity displayed by T11SS and their cargo is fascinating for both molecular and evolutionary reasons. All T11SS-dependent proteins characterized to date share a similar C-terminal 8-stranded hydrophilic β-barrel domain (Grossman, Mauer, et al. 2022; Hooda, Lai, and Moraes 2017). Our observations support the hypothesis that this C-terminal domain contributes to cargo for their T11SS-dependent localization. $HrpB_{H.haem}$ and $HrpB_{X.nem}$ both secreted their cognate hemophilin protein more effectively than either of the chimeric hemophilin proteins generated by swapping their domains. However, both T11SS HrpB homologs were more effective at secreting the chimeric hemophilin that possessed the C-terminal domain from their cognate protein relative to the one with a heterologous C-terminal domain, even though the latter chimera had a cognate N-terminal domain. This partial reconstruction of specificity could indicate that the C-terminal domain predominantly dictates T11SS targeting while the N-terminal domain contributes in a smaller way. Alternatively, it may indicate that the chimeric proteins generated have compromised stability relative to the native cargo sequences that slows transposition or speeds up protein degradation. Regardless, the ability of these T11SS proteins to distinguish their cognate C-terminal domain is striking given that the two cargo proteins are closely related in sequence (39% identity)

and function. Further, our data suggest that the efficacy of T11SS secretion varied in magnitude among the T11SS OMP homologs that were tested in parallel in an *E. coli* expression strain. This information is important to consider for potential bioengineering applications of T11SS. For instance, our finding that the C-terminus of cargo can alter specificity for closely-related but different T11SS systems, raises the possibility that novel T11SS-dependent cargo could be engineered for specific secretion by fusing novel N-terminal functional domains to the C-terminal domain from a characterized cargo protein. Similarly, Slam proteins, the subset of T11SS responsible for surface exposure of lipoproteins, have been proposed as a novel mechanism for surface presentation of immunogenic antigens as a potential vaccination strategy (Moraes et al. 2021), and our discovery that different T11SS may have different levels of secretion efficacy could be exploited to fine-tune the levels of surface exposure or secretion. An important next step towards achieving this goal will be to identify the individual sequence and structural motifs responsible for the specificity between cargo and T11SS, to enable automatic annotation of T11SS cognate pairs and informed engineering of novel pairings.

From an evolutionary perspective, our observations of specificity between and horizontal genomic acquisition of T11SS/cargo pairs, as well as the distinctive co-occurrence patterns revealed by our network analysis, suggests that these pairs have been co-evolving together for long periods of time, often co-inherited as a single locus. This hypothesis would also explain the consistency of co-occurrence patterns seen amongst T11SS loci, often carrying their own *tonB* and TonB-dependent receptor genes (Grossman, Mauer, et al. 2022). During evolution of the *Xenorhabdus* genus, the Hrp locus has been duplicated on at least two separate occasions, once by genomic duplication (Hrp2) and once by putative transposition (Hrp_{Tn}), based on association with a transposon. Single protein phylogenetic analysis suggests that CrpB evolved from an ancient duplication of the Hrp2 locus (Fig. 3.5). Additionally, our network analysis revealed that 82.4% of the hemophilin homologs within the

"cobalt/molybdenum associated" cluster co-occurred with at least one additional T11SS-dependent cargo, suggesting that this cluster contains representatives derived from duplication of cargo proteins and evolution of paralogs over time, all within the same locus and likely sharing a single T11SS transporter. Based on the sequence differences between HrpC, HrpC2, and HrpC3 within the ligand binding domain this paralogous adaptation may drive the evolution of novel ligand binding affinities (Fig. 3.5).

Purified hemophilin homologs have variable affinity for hemin in titration curves, suggesting potential diversity in their ecological roles (Fig. 3.8; Table 1). *H. haemolyticus* Hpl is a heme chelating protein that can prevent heterospecific organisms from accessing environmental heme (Latham et al. 2020). This suggests that *H. haemolyticus* Hpl has a dual function: promoting heme uptake by the producing strain and as a tool for competing against other organisms by sequestering a limiting nutrient away from any bacterium that lacks the appropriate TonB-dependent uptake system (a form of nutritional immunity). This role in competition may be facilitated by the relatively high hemin affinity of Hpl, the strongest observed amongst the three heme-binding hemophilin homologs tested here. Conversely, we found that, of the three hemin-binding hemophilin homologs tested here, HrpC had the lowest binding affinity. This may be a reflection of the distinctive niches occupied by *X. nematophila* relative to the other two bacterial sources of hemophilin homologs tested (*H. haemolyticus* and *A. baumannii*). *Xenorhabdus* are capable of infecting diverse insects, which are notoriously heme poor (Burmester and Hankeln 2007) and its mutualistic nematode host is a heme auxotroph (Buecher and Popiel 1989; Hieb, Stokstad, and Rothstein 1970). *Drosophila melanogaster* transferrin-1 binds heme with lower affinity than mammalian transferrin and is susceptible to low pH conditions (Weber, Kanost, and Gorman 2020). Therefore, to access heme in an insect environment, *X. nematophila* HrpC may not need high affinity to overcome host chelation. In turn, lower affinity may offer a selective advantage by enabling resource sharing with its mutualistic

nematode host, *S. carpocapsae,* with the resulting improved fitness benefitting both mutualistic partners. Dissociation of heme from HrpC would be essential for heme to be transferred to nematode iron chelating proteins. Entomopathogenic nematodes can live for months at a time without feeding while in their free living infective stage. During this time they have a sealed intestine and no access to exogenous nutrients (Goodrich-Blair 2007). While deletion of *hrpB* had no impact on *X. nematophila* colonization of, or persistence within, the infective stage nematode, future experiments could determine if valuable bacterially derived cofactors like heme and cobalamin are being provisioned to the nematode during this prolonged starvation period.

CrpC, a member of the HrpC2 group of hemophilin homologs, lacks the β2- β3 and β5- β6 loop residues that are conserved among HrpC homologs, including two histidines that are found in HrpB. Consistent with this distinctive sequence at the heme coordination sites, CrpC did not display strong affinity for hemin or any other porphyrin or tetrapyrrole tested. CrpC and its T11SS partner, CrpB are encoded adjacent to an anaerobic B12 biosynthesis locus (Fig. 3.3B) in three strains of *Xenorhabdus* (HrpBC_Classes_Table_Xenorhabdus.xlsx). At this locus in another strain, *X. budapestensis*, the *crpB* gene is apparent, but appears to have been fragmented, and no partner *crpC* homolog is present. The genomic context of CrpC and CrpB initially led us to hypothesize that this cargo/secretion pair functions in cobalt or cobalamin binding. In most tetrapods, including humans, B12 is chelated by haptocorrins and gastric intrinsic factor (Gif) (Lopes-Marques et al. 2014), however very little is known about the B12 chelating compounds in insects and nematodes. *D. melanogaster* is known to encode a homolog of the cubulin receptor (Kozyraki, Verroust, and Cases 2022) which uptakes B12 from *gif* in humans, however in *D. melanogaster* it is instead associated with dorsal/ventral differentiation during embryonic development (Shimell et al. 1991). *C. elegans* is not known to encode homologs of haptocorrin or Gif, despite its dependence on B12 for fertility and longevity (Bito et al. 2013). While seemingly absent or repurposed in arthropods and nematodes, the

hemichordate acorn worm species *Saccoglossus kowalevskii* does encode homologs of two haptocorrins and *gif*, suggesting that these genes are at least basal to the superphylum Deuterostomia (Lopes-Marques et al. 2014). None of the ligands we tested were capable of high-affinity binding with CrpC, consistent with our structural predictions that indicate that CrpC is unlikely to bind porphyrins directly. It is possible that CrpC has evolved to bind some as-yet-unknown ligand that may be present in either the insect or nematode environment. One way to identify such a ligand could be by extracting it from insect and nematode homogenates by virtue of its ability to bind purified CrpC. Overall, our findings, suggest that in *Xenorhabdus*, the hemophilin family comprises bona-fide hemophilins, such as HrpC, as well as paralogs, including HrpC2 and CrpC homologs, that have evolved new, as yet unknown, ligand binding activities.

The crystal structures of Hpl and HphA reveal a heme binding site in which the heme group is positioned between a conserved α helix and a loop extended from the β5-6 hairpin (Fig. 3.12A-D). A threading model of HrpC reveals a similar heme binding site (Fig. 3.12E-F). Previous studies based on circular dichroism have demonstrated that the base of the active site experiences little conformational change between holo- and apo- forms (Latham et al. 2020), and that the low complexity β5-6 loop shifts with heme binding to lock the ligand into place until release (Bateman et al. 2021). In support of this model, our results show that the affinity of Hpl, HphA, and HrpC correlates with the length of the flexible loop in structural models (Fig. 3.12). Given the novel nature of this heme binding motif, its conservation throughout the hemophilin family of proteins, and the presence of two solved structures we propose that the domain be canonized as the Hemophilin ligand binding domain (Hlb).

The ligand binding affinities of hemophilin proteins appears to vary among homologs and species. Structural variations within the hemophilin ligand binding domain may reflect subtle optimizations for these varied functions. In *H. haemolyticus*, hemophilin plays a role in inter-bacterial

competition (Atto et al. 2020), while in *A. baumannii,* hemophilin is a virulence determinant (Bateman et al. 2021). In contrast, our findings indicate that *X. nematophila* hemophilin does not inhibit the growth of a competitor bacterium found within its insect host environment, nor is it necessary for virulence in *Manduca sexta* insects. While investigation of other bacterial competitors or insect hosts may reveal such competition and virulence functions, our data instead point to the possibility that *X. nematophila* hemophilin serves as a heme acquisition factor, possibly for both the bacterium itself and for its mutualistic nematode host, *S. carpocapsae*. This idea is based on several observations: first the *X. nematophila hrp* locus is induced by, and facilitates survival during, periods of heme starvation experienced by the bacterium. Further, the production of hemophilin as a "communal good" is supported by our finding that 'recipient' *X. nematophila* cells have improved growth under metal starvation conditions when grown in media conditioned by HrpC secretion. Second, *S. carpocapsae* nematodes grown on lawns of *X. nematophila* that lack the HrpB T11SS produced significantly fewer offspring than wildtype or complemented bacterial lawns, indicating that the *hrp* locus, which is present in all symbionts of entomopathogenic nematodes, may be contributing directly to host success. Future studies will explore this interaction in detail to determine if this fitness effect is driven by the observed mucoidy of the *hrpB* mutant, the nutritional state of the bacteria, or direct interactions between hemophilin and *Steinernema* nematodes. Like all previously studied nematodes, entomopathogenic nematodes are heme autotrophs (Luck et al. 2016; Rao et al. 2005), and their insect prey are typically poor in heme as a result of using alternative oxygen carrying molecules (Burmester and Hankeln 2007), making their bacterial symbionts the most likely source of heme (Mucci et al. 2022). In this system, a secreted, relatively low affinity hemophilin capable of transmitting heme to the nematode, without lysing the producing bacterial symbiont cell, could be invaluable during periods of starvation.

In conclusion our data show that hemophilin homologs are secreted by cognate partner T11SS OMPs and that partner pairs are specific for each other. Specificity is at least partially conferred by the hemophilin C-terminal β-barrel domain, which can direct efficient secretion by the partner T11SS when fused to a different hemophilin N-terminal domain. Sequence similarity network-based analysis continues to be a useful tool for grouping T11SS systems by function and has proved instrumental in linking T11SS proteins to potential cargo proteins for subsequent experimental characterization. While this study reconstructed hemophilin secretion in *E. coli*, much work remains to discover how transcriptional regulation, post-transcriptional regulation, modulation of secretion rate, and variable binding affinity all come together to facilitate mutualist, commensal, and pathogenic infection of animal hosts by Proteobacteria encoding hemophilin loci.

Materials and Methods:

Sequence similarity network analysis

Protein sequence similarity networks were generated using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) (Zallot, Oberg, and Gerlt 2019). As an input we used the previously reported database of soluble TbpBBD cargo encoding genes which co-occurred with T11SS encoding genes (Grossman, Mauer, et al. 2022). EFI-EST performs an all-by-all BLAST of query sequences to assess relatedness, and then generates a network where each node represents a protein sequence, and the color of the edges indicates relatedness between nodes. A minimum alignment score of 35 was chosen to reduce total network edges enough to visualize protein subclusters. To simplify visualization proteins sharing ≥ 80% identity were compressed into representative nodes. Networks were visualized and interpreted using Cytoscape v3.7.1 (Shannon 2003) and Gephi v0.9.5 (Bastian, Heymann, and Jacomy 2009). For the complete network of soluble TbpBBD domain proteins, nodes were organized with the Fruchterman-Reingold force-directed

algorithm (Fruchterman and Reingold 1991). For the network containing only hemophilin family proteins, nodes were organized with the ForceAtlas2 algorithm for continuous force-directed arrangement (Jacomy et al. 2014). Subclusters were identified by eye (Fig. 3.3A) and representative loci were chosen to provide syntenic context (Fig. 3.3B).

Bacterial culture conditions

All strains and plasmids utilized in this study are described in table 2. All primers utilized in this study are described in table 3. All cultures were grown in either glucose minimal media (Bhasin, Chaston, and Goodrich-Blair 2012), LB stored in the dark to prevent formation of oxidative radicals (henceforth dark LB), Case amino acid media with defined metals (Orchard and Goodrich-Blair 2004), or glucose minimal media supplemented with 1% dark LB. Plate based cultures were grown on either LB supplemented with pyruvate to prevent the formation of reactive oxygen radicals (henceforth LBP), lipid agar optimized for nematode growth (McMullen II and Stock 2014), or glucose minimal plates (Bhasin, Chaston, and Goodrich-Blair 2012). For plasmid-based expression, chemically competent *Escherichia coli* strain BL21-DE3 (C43) were chosen for ease of transformation and their ability to non-toxically express membrane proteins (Dumon-Seignovert, Cariot, and Vuillard 2004; Miroux and Walker 1996). Strains of *Escherichia coli* were grown at 37°C. Strains of *Xenorhabdus nematophila* and *Enterococcus faecalis* were grown at 30°C. Where appropriate media was supplemented with ampicillin at a concentration of 150 μg/ml, chloramphenicol at 15 μg/ml or kanamycin at 50 μg/ml unless another concentration is stated. Protein expression was induced at the midlog point of bacterial growth via addition of isopropyl β-d-1-thiogalactopyranoside (henceforth IPTG) at a concentration of 0.5mM.

Construction of T11SS and T11SS-dependent cargo expression plasmids

Expression plasmids for HrpC alone (HGB2531) and HrpBC co-expression (HGB2530) were previously generated and reported on (Grossman, Mauer, et al. 2022). FLAG-*hrpB* was

Table 2. **Strains used in chapter 2**.

HGB #	Genotype	Plasmid	Antibiotic markers
238	E. coli BL21 DE3	None	Amp150
800	X. nematophila; ATCC 19061	Native plasmid	Amp150
1495	X. nematophila (HGB007) delta Nil locus, delta nilR, kefA::GFP, attTn7::Nil locus	Native plasmid	Clor15 Amp150 Kan50 Erm200 Strep12.5
1496	X. nematophila (HGB007) delta Nil locus delta nilR, kefA::GFP, attTn7::emptyTn7	Native plasmid	Clor15 Amp150 kan50 Erm200 Strep12.5
2402	E. coli BL21 DE3	pETDuet/MCS1:0075CtermFlag	Amp150
2459	E. coli BL21 DE3	pETDuet/MCS1:0075CtermFlag/MCS2:0074NtermFlag	Amp150
2485	E. coli BL21 DE3	pETDuet/MCS1:0075Cterm3x(GGGGS)+GSTfusion/MCS2:0074NtermFlag	Amp150
2486	E. coli BL21 DE3	pETDuet/MCS1:0075Cterm3x(GGGGS)+GSTfusion	Amp150
2530	E. coli BL21 DE3 C43	pETDuet/MCS1:0075CtermFlag/MCS2:0074NtermFlag	Amp150
2531	E. coli BL21 DE3 C43	pETDuet/MCS2:0075NtermFlag	Amp150

Table 3. **Primers used in chapter 3**.

Primer #	Name	Annealing temp	Product length	Sequence
1	XnemSLAM_Substrates_F_SacI	57.5	2995bp	nnnnngagctcttatccctgcctcgatttttg
2	XnemSLAM_Substrates_R_SalI			nnnnngtcgacatatgtctcccactgagctaaatga
3	0075_Cterm_FlagTag_F_3	59	~750bp	gatgatgataaataatcactaatttttcagagaacag
4	0075_FlagTruncated_R			tctttataatcctgtttagaaccaccgaatg
5	0075CFlag_NcoISite_F	52.9	786bp	nnnnccatggtaaaaaagttcaacgtatt
6	0075CFlag_NotISite_R			nnnngcggccgcttatttatcatcatcatctttat
7	GST_link_F	61.6	760	ggaggaggaggatctggaggaggaggatctatgtcccctatacta ggtattggaaaattaaggccttgtgc
8	GST_link_R			cggccgcttatcagtcacgatgcggccgctcgagt
9	Duet74NFlag-75_link_F	56.7	7578	tcgtgactgataagcggccgcataatgcttaagtc
10	Duet74NFlag-75_link_R			agatcctcctcctccagatcctcctcctccctgtttagaacca ccgaatgcggtatc
11	DuetDOWN1_R			gattatgcggccgtgtacaa
12	DuetUP1_F			atgcgtccggcgtaga
13	DuetUP2_F			ttgtacacggccgcataatc
14	0074_25aa_FlagTruncated_R			gatgatgataaagatgaggatacttcccgtaatatc

amplified from pETDuet-1/*hrpBC*$_{X.nem}$ using primers 1-2. *hpl*-FLAG and its adjacently encoded T11SS neighbor, FLAG-*hrpB*$_{H.baem,}$ were amplified from the purified genome of *Haemophilus haemolyticus* BW1 using primers 3-6. *crpC*-FLAG and its adjacently encoded T11SS neighbor, FLAG-*crpB*, were amplified from the purified genome of *Xenorhabdus cabanillasii* (HGB2490) using primers 7-10. *hphA* (ACJ40780.1) and *hsmA* (ACJ40781.1) were generated via gene synthesis by Genscript. To make cargo-only expression plasmids, pETDuet-1, *hpl*-FLAG, *crpC*-FLAG, and *hphA*-FLAG were digested with NcoI and NotI. Each cargo protein was independently ligated into MCS1 of pETDuet-1 via T4 DNA ligase, resulting in pETDuet-1/*hpl* (HGB2526), pETDuet-1/*crpC* (HGB2525), and pETDuet-1/*hphA* (HGB2532). Integration of each T11SS-dependent cargo was confirmed via digestion with NcoI and NotI as well as Sanger sequencing using primers 11-12 at the University of Tennessee (UT) Genomics Core.

To make T11SS/cargo co-expression plasmids, each of the above cargo-only expression plasmids were digested with KpnI and NdeI, alongside the PCR products for FLAG-*hrpB*$_{H.baem,}$ FLAG-*crpB*, and FLAG-*hsmA*. Each T11SS protein was then independently ligated into MCS2 of the plasmid containing its cognate cargo via T4 DNA ligase, resulting in pETDuet-1/*hpl*/*hrpB*$_{H.baem}$ (HGB2523), pETDuet-1/*crpC*/*crpB* (HGB2524), and pETDuet-1/*hphA*/*hsmA* (HGB2533). Additionally, the PCR product for FLAG-*hrpB* from *X. nematophila* was digested with KpnI and NdeI and ligated into MCS2 of all the cargo-only expression plasmids, resulting in pETDuet-1/*hpl*/*hrpB*$_{X.nem}$ (HGB2529), pETDuet-1/*crpC*/*hrpB*$_{X.nem}$ (HGB2528), and pETDuet-1/*hphA*/*hrpB*$_{X.nem}$ (HGB2527). Integration of each T11SS protein was confirmed via digestion with KpnI and NdeI as well as Sanger sequencing using primer 13 the at the University of Tennessee (UT) Genomics Core.

To construct chimeric hemophilin homologs, *hrpC* and *hpl* were split into two domains based on multiple sequence alignment and the NCBI conserved domain database. *hrpC* was split between position 402 and 403, while *hpl* was split between position 474 and 475. Primers 14-15 were used to

amplify the hemophilin handle domain from pETDuet-1/*hrpC*/*hrpB*$_{X.nem}$ (HGB2530). Primers 16-17 were used to amplify *hrpB* and the hemophilin β-barrel domain from pETDuet-1/*hrpB*$_{X.nem}$/*hpl* (HGB2529). These two products were assembled into pETDuet-1/Chimeric hemophilin(*hrpC-hpl*)/*hrpB*$_{X.nem}$ (HGB2595). Primers 18-19 were used to amplify the hemophilin handle domain from pETDuet-1/*hrpB*$_{X.nem}$/*hpl* (HGB2529). Primers 20-21 were used to amplify *hrpB*$_{X.nem}$ and the hemophilin β-barrel domain from pETDuet-1/*hrpC*/*hrpB*$_{X.nem}$ (HGB2530). These two products were assembled into pETDuet-1/Chimeric hemophilin(*hpl-hrpC*)/*hrpB*$_{X.nem}$ (HGB2596). Primers 14-15 were used to amplify the hemophilin handle domain from pETDuet-1/*hrpC*/*hrpB*$_{X.nem}$ (HGB2530). Primers 16-17 were used to amplify *hrpB*$_{H.haem}$ and the hemophilin β-barrel domain from pETDuet-1/*hpl*/*hrpB*$_{H.haem}$ (HGB2523). These two products were assembled into pETDuet-1/ Chimeric hemophilin (*hrpC-hpl*)/*hrpB*$_{H.haem}$ (HGB2597). Finally, pETDuet-1/Chimeric hemophilin (*hpl-hrpC*)/*hrpB*$_{X.nem}$ (HGB2596) and pETDuet-1/*hpl*/*hrpB*$_{H.haem}$ (HGB2523) were digested with NotI and NcoI, liberating the hemophilin homolog from each vector. The chimeric hemophilin from HGB2596 was isolated via gel electrophoresis and then ligated into MCS1 of the vector isolated from HGB2523, resulting in pETDuet-1/Chimeric hemophilin (*hpl-hrpC*)/*hrpB*$_{H.haem}$ (HGB2598). Integration of each T11SS protein and chimeric cargo was confirmed via Sanger sequencing using primers 12-13 at the University of Tennessee (UT) Genomics Core.

Protein expression and Immunoblotting

E. coli strains used for expression experiments were taken fresh from storage at -80°C for each experiment. Strains were cultured on glucose minimal media plates + ampicillin overnight. For each biological replicate, 10 colonies were pooled and inoculated into 5ml of fresh minimal media glucose + ampicillin broth and incubated rotating overnight. Each replicate of each strain was rinsed 2x in PBS and normalized to an OD_{600} of 0.05 in 60ml of glucose minimal media + 1% LB + ampicillin. These were grown shaking at 225rpm until they reached mid log growth ($OD_{600} \approx 1$),

typically between 5 and 8.5 hours. Upon reaching midlog growth, 25mls of each culture was removed and used as an uninduced T0 control. The remaining 35mls were supplemented with IPTG to a concentration of 0.5mM. One milliliter supernatant samples were taken at 1 and 2.5 hour(s) post induction. Supernatant samples were clarified via centrifugation and filter sterilized. At 2.5 hours post induction the remaining cultures were concentrated via centrifugation, rinsed 2x in PBS, and lysed via sonication (30 s at ~500-rms volts). Supernatant samples and cellular lysate samples were supplemented with PMSF (1.7 μg/ml), Leupeptin (4.75 μg/ml), and Pepstatin A (0.69 μg/ml) to inhibit proteinase activity. The no plasmid control was performed identically except without the presence of ampicillin in the media.

The protein concentration of cellular lysates was normalized via the Pierce 660nm Protein Assay (REF22660). For supernatant samples 600μl of each filtered sample was precipitated via 10% Trichloroacetic acid precipitation as previously described (Grossman, Mauer, et al. 2022; Koontz 2014a). Samples were boiled for 10-25 min. prior to performing SDS-PAGE to ensure complete unfolding in the protein sample buffer. SDS-PAGE was performed in duplicate using 10% polyacrylamide gels. The first gel was used to perform Coomassie staining for total protein content while the second gel was transferred to a PVDF membrane for Western immunoblotting. Immunoblots were incubated in 50% Ly-cor blocking buffer:50% Tris buffered saline (TBS) for 1 hour to block. Immunoblots were then incubated in 50% Ly-cor blocking buffer:50% TBS supplemented 0.1% Tween20 and 1:5000 rat α-FLAG antibody for 1 hour. Subsequently the blots were incubated in 50% Ly-cor blocking buffer:50% TBS supplemented 0.1% Tween20 and 1:5000 goat α-rat antibody bound to a 680CW fluorophore for 1 hour. Finally, immunoblots were visualized using a Li-cor odyssey imaging the 700nm wavelength. The intensity of supernatant samples was normalized to a clearly visible, non-target protein band in the Coomassie stain to control for protein concentration. Efficacy of secretion was measured as the fold change of cargo protein present in the

supernatant when co-expressed with a T11SS protein relative to cargo protein present in the supernatant when expressed alone. Fold changes were not normally distributed initially, so they were $\log_{10}$ transformed prior to analysis. The resulting data were analyzed via a one-way ANOVA and a Tukey's HSD Test (Tukey 1949).

Purification of hemophilin homologs

Hemophilin from *H. haemolyticus* was expressed and purified as previously described to yield low and high heme-content fractions after anion exchange chromatography (Latham et al. 2020). Heme was removed by cold acid acetone treatment to yield an apo hemophilin fraction, as previously described (Ascoli, Rossi Fanelli, and Antonini 1981). Residual Fe(III) heme was estimated at 1.8% of sites, based on extinction coefficients of met-hemophilin being 96,100 M^{-1} cm^{-1} and 38,600 M^{-1} cm^{-1} at 414 nm and 280 nm, respectively, and extinction coefficient of the apo-protein being 25,900 M^{-1} cm^{-1} at 280 nm.

Expression constructs encoding the hemophilin homologs from *X. nematophila* (residues 23–247), *X. cabanillasii* (residues 23–238) and *A. baumannii* (residues 21–264) were constructed in pET28a. In each case, the native N-terminal signal peptide was omitted and replaced with a hexa-histidine tag and engineered tobacco etch virus (TEV) protease cleavage site. Clones were transformed into *E. coli* strain Rossetta-2 (Novagen), grown in LB containing 34 μg/ml chloramphenicol and 25 μg/ml kanamycin; expression was induced with 1 mM IPTG for 3 hours shaking at 37°C. Cells were suspended in lysis buffer (0.5 M NaCl, 0.05 M sodium phosphate, 0.02 M imidazole, 100 μM phenylmethylsulfonyl fluoride, pH 7.2) and lysed by sonication (Branson), then hemophilin homologs were captured by Ni-affinity chromatography. TEV protease was expressed and purified as described (S. van den Berg et al. 2006). The His-tag was cleaved from hemophilin homologs by TEV protease treatment overnight at room temperature, to liberate hemophilin proteins with an additional N-terminal Gly-His-Met tripeptide residual from the TEV

cleavage site. TEV protease and His-tag peptides were removed over a second Ni-affinity column. Hemophilin preparations from *X. nematophila* and *A. baumannii* had a brownish appearance and an absorbance peak at ~413 nm characteristic of a porphyrin ligand, as well as less intense absorption peaks at 533 and 659 nm. Ligand was estimated to occupy ~25% of sites based on comparison with spectra of hemophilin from *H. haemolyticus*. In contrast, CrpC from *X. cabanillasii* was colorless. Acid acetone or methyl ethyl ketone extraction was not effective to remove colored contaminants from HrpC of *X. nematophila* or *A. baumannii*. Apo-protein fractions of these proteins were prepared by reversed-phase HPLC over a C4 stationary phase (Waters Symmetry) developed with a CH_3CN:water mobile phase gradient containing 0.1% trifluoroacetic acid. Solvent was removed by lyophilisation. Apo-CrpC from *X. cabanillasii* was applied to a strong anion exchange resin (Q sepharose, Pharmacia) in 25 mM Tris-HCl buffer (pH 8.25 at 21°C) and collected in the flow-through. All apo-proteins were dialyzed into 20 mM Tris-HCl buffer (pH 7.9 at 21°C) prior to storage at –80°C. Apo-protein concentrations were determined by absorption extinction coefficient at 280 nm calculated from amino acid composition.

UV-visible absorption and fluorescence spectroscopy

UV-visible spectra were recorded on a Jasco V-630 spectrophotometer fitted with a temperature-controlled sample holder (Jasco) and spectrosil quartz cuvettes with a path length of 1.0 cm (Starna, Baulkham Hills, Australia). Porphyrin concentrations were determined according to the following molar extinction coefficients. Fluorescence measurements were made on a Tecan Spark M20 plate reader in 96-well format. To prepare Zn-PPIX, 0.5 g PPIX (Frontier Scientific) was dissolved in boiling chloroform (100 mL), to which a saturated solution of Zn acetate in MeOH (1 mL) was added. The mixture was refluxed for 20 min and then a small amount of MeOH was added and, after cooling, the dark red solid was filtered off (50:50 Zn-PPIX:PPIX by HPLC). Zn-PPIX was purified by RP-HPLC over a C18 solid phase (Phenomenex) with isocratic

acetone:MeOH:water:formic acid (280:120:100:1) mobile phase, which achieved baseline separation of the Zn-PPIX fraction.

Hemin binding measurements in absorbance mode were made by successive additions of apo-protein (0.4 mM stock) into 2.5 mM Zn-PPIX in 20 mM Tris.HCl, pH 8 supplemented with BSA (1 mg/mL) as a blocking agent. Fluorescence measurements were made in 96-well format with 36 (Hpl, HphA, HrpC) or 21 (CrpC) individually prepared samples covering an appropriate concentration range. Data were fitted to a 1:1 binding model accounting for ligand depletion.

$F_{obs} = F_0 - (F_{sat}/2M)\{(L+K_d+M) - sqrt[(L+K_d+M)^2 - 4ML]\}$, where F_{obs} is fluorescence signal, F_0 is the starting fluorescence, F_{sat} is the fluorescence at saturation, L and M are the ligand and macromolecule concentrations, respectively, and K_d is the equilibrium dissociation constant. Data were fitted using GNUPLOT (v. 4.6). Data were also fitted with a second simpler linear model that assumed F_{obs} as proportional to protein concentration (positing that K_d is, in fact, too small to measure with the experimental setup). The second order (corrected) Akaike's Information Criterion (AIC) was used to compare the two models. Comparing AIC for the two model-fits (ΔAIC) indicates which model is the more probable explanation of the data evidence. We quote the evidence ratio ($1/e^{-0.5 \times \Delta AIC}$) in favor of the more probable model. A 95% confidence interval for the K_d parameter was obtained by determining, from the F-distribution, a threshold sum-of-squares for which a fit with fixed K_d would not be significantly different from the best-fit model at a significance level of P=0.05. The K_d value was then fixed at values above and below the best-fit value until the threshold sum-of-squares was reached.

Xenorhabdus nematophila P_{hrpB} and P_{hrpC} transcriptional reporters

Transcriptional reporter strains of *Xenorhabdus nematophila* were generated via homologous recombination of a β-galactosidase cassette immediately after the promoter region of *hrpB* and *hrpC* in the genome. This β-galactosidase cassette includes a promoterless *lacZ* gene, a ribosome binding

site, and a complete chloramphenicol resistance cassette. One kilobase fragments preceding *hrpB* and *hrpC* were amplified from *X. nematophila* ATCC19061 (HGB0800) using primers 22-23 and 24-25 respectively. These fragments were introduced to the β-galactosidase cassette of pKV124/Δtnp using digestion-ligation between the SacII and XbaI sites. The resulting plasmids were transformed into *E. coli* S17-1 λ pir (HGB0066), generating pKV124/Δtnp/P_{hrpB} (HGB2322) and pKV124/Δtnp/P_{hrpC} (HGB2323). Plasmids were introduced to *X. nematophila* ATCC19061 via conjugation at a 2:3 *E. coli* to *X. nematophila* ratio, resulting in *X. nematophila* P_{hrpB}::*lacZ* (HGB2325) and *X. nematophila* P_{hrpC}::*lacZ* (HGB2326). Successful integration was confirmed via chloramphenicol resistance and PCR performed with primers 26-27. Strains were grown under constant chloramphenicol selection to maintain the insertion. A previously generated P_{rpoS} transcriptional reporter (HGB2012) was used as a positive control for β-galactosidase since this promoter produces strong, consistent expression. Meanwhile a promoterless control (HGB1828) served as negative control for experiments.

All transcriptional reporter strains were taken fresh from storage at -80°C for each experiment. Strains were cultured on LBP supplemented with chloramphenicol and ampicillin overnight. For each biological replicate ~10 colonies were pooled and inoculated into 5ml of fresh dark LB supplemented with chloramphenicol and ampicillin and incubated overnight. Each replicate of each strain was rinsed 2x in PBS and normalized to an OD_{600} of 0.1 in 5ml of dark LB. These were grown rotating at 30°C for 3 hours to reach mid log growth. All cultures were then normalized to an OD_{600} of 1.5 in PBS. Concentrated cultures were supplemented with either 250μM 2,2'-bipyridyl, 23μM hemin chloride, or a vehicle control (PBS with 1mM NaOH). All cultures were incubated shaking at 30°C for 90 min. to allow the effector compounds to induce expression. After incubation, 90μl each culture was aliquoted into a 96 well plate alongside 20μl of fluorescent reporter solution (150mM PIPES pH 7.2 + 0.25% Triton X-100 + 20μM Fluorescein Di-β-D-

Galactopyranoside). Reactions were incubated at 37°C for 90 min. in the dark. β-galactosidase activity was measured by exciting the samples at 485nm and reading the subsequence emission at 528nm. Intensity values were converted into fold change relative to the vehicle control treatment. Fold changes were not normally distributed, so they were $\log_{10}$ transformed prior to analysis. Fluorescence data were subjected to a one-way ANOVA with a Tukey's HSD Test (Tukey 1949).

For the metal compound repression analysis, *X. nematophila* P_{hrpC}::*lacZ* (HGB2326) was raised and prepared exactly as described above until the addition of effector. All cultures were supplemented with 250μM 2,2'-bipyridyl and incubated at 30°C for 90 min. while shaking to induce high levels of expression from P_{hrpC}. 300μl of each culture was aliquot into a 96 well plate at an OD600 of 1.5 and supplemented with hemin chloride, ferric chloride, ferrous sulfate, cobalt sulfate, copper sulfate, magnesium sulfate, manganese sulfate, nickel sulfate, zinc chloride, or vehicle control at a final concentration of 100μM. Cells were then incubated again at 30°C for 90 min. while shaking to allow the effector compounds to drive *lacZ* expression. Intensity values were collected as described above and were subjected to a one-way ANOVA with a Tukey's HSD Test (Tukey 1949).

Tn*5* mutagenesis of P_{hrpC} transcriptional reporter

To perform a Tn5 random mutagenesis of *X. nematophila* P_{hrpB}::*lacZ* (HGB2325) and P_{hrpC}::*lacZ* (HGB2326) overnight cultures of both recipient strains as well as *E. coli* carrying the Mini-Tn5 donor plasmid pBSL118 (HGB0069) were raised in dark LB at 30°C (Alexeyev, Shokolenko, and Croughan 1995). The reporter cultures were supplemented with chloramphenicol and the donor strain culture was supplemented with kanamycin. Overnight cultures were sub-cultured 1:10 into unsupplemented dark LB and allowed to grow for 3 hours to reach logarithmic growth. To facilitate conjugation, each reporter strain was independently combined with the donor strain at a 3:2 ratio (900μl:600μl) and spun for 1 minute at 16,000xG to pellet the cells. The conjugation mixture was then suspended in 250μl of dark LB and spotted onto permissive LBP plates which were incubated

for 24 hours at 30°C. Streaks from the resulting conjugation spots were suspended in 1 ml of PBS, diluted 1:10, and then spread plated in triplicate onto LBP supplemented with chloramphenicol, kanamycin, and 80 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (henceforth X-gal) to select for exconjugants and simultaneously screen for mutants with modified expression from P_{hrpB} and P_{hrpC}. Selection plates were incubated for 72 hours at 30°C since Hrp expression is strongest under stationary growth conditions. 190 individual exconjugants were picked from the plates and inoculated into dark LB in two 96-well plates. Initially colonies with atypically high or low β-galactosidase activity (atypically blue or white colonies) were intentionally selected, however there were not atypical colonies enough to fill both plates so the second 96-well plate was populated haphazardly instead.

Isolated exconjugants were incubated 24 hours shaking at to reach stationary phase. Alongside unmutated versions of each reporter, each culture was then sub-cultured into fresh 96-well plates containing X-gal at a final concentration of 1mg/ml and incubated 16 hours shaking at 30°C to react with the X-gal. Absorbance at 610nm was used to detect β-galactosidase activity and compare it to the level of activity seen in the unmutated control. No exconjugants altered the expression of P_{hrpB}::*lacZ* detectably. Six exconjugants were chosen as interesting based on increased (2 of 6) or decreased (4 of 6) expression of P_{hrpC}::*lacZ*. The genomes of these six strains were purified and digested with HincII. The cloning plasmid pUC19 (HGB0101) was also digested with HincII. The genomic DNA from the exconjugants of interest was then ligated into the pUC19 plasmid using T4 DNA ligase. The resulting plasmids were transformed into electrocompetent DH5α for plating on LBP supplemented with ampicillin and kanamycin. Plasmids were purified from these transformants and sequenced by SNPsaurus.

In vivo analysis of *hrpB* in *Xenorhabdus nematophila*

To assess the role of the HrpB T11SS protein within *X. nematophila,* a deletion mutant was constructed. 1000bp regions from upstream and downstream of *hrpB* were amplified from *X. nematophila* ATCC19061 (HGB0800) using primers 28-29 and 30-31 respectively. A kanamycin resistance cassette was amplified using primers 32-33 and a pUC19 backbone was amplified using primers 34-35. These four fragments were ligated via Hi-Fi assembly. The *hrpB* allelic exchange cassette was digested out of pUC19 using SacI and SalI and ligated into the suicide vector pKR100 to generate pKR100/*hrpB*::*kan*R (HGB2300). This plasmid was conjugated into *X. nematophila* ATCC19061 (HGB0800) and exconjugants were selected for using kanamycin resistance, resulting in *X. nematophila hrpB*::*kan*R (HGB2301). Recombination was confirmed via elimination of chloramphenicol resistance and Sanger sequencing using primers 36-37. To assess the ability of the bacteria to colonize nematodes, a GFP expressing strain was constructed by conjugating pJMC001 (HGB1783) into *X. nematophila hrpB*::*kan*R (HGB2301) to generate *X. nematophila hrpB*::*kan*R *kefA*::GFP (HGB2302). A GFP expressing control was also generated in this manner resulting in *X. nematophila kefA*::GFP (HGB2106). To construct a complemented strain, the entire predicted operon *hrpB* is encoded in was amplified using primers 38-39, digested with KpnI, and ligated into the Tn*7* transposon site present on the suicide vector pEVS107 (HGB0280) to form pEVS107/*tonBhrpAB* (HGB2375). This transposon was then integrated into the *hrpB* deletion mutant via triparental λ pir conjugation alongside the helper plasmid pUX-BF13 (HGB0282), resulting in *X. nematophila hrpB*::*kan*R *att*Tn7::*tonBhrpAB* (HGB2390).

Growth curves were performed using *X. nematophila* wild type (HGB0800), *hrpB*::*kan*R (HGB2301), and the *hrpB*::*kan*R *att*Tn7:: *tonBhrpAB* complemented strain (HGB2390). Cultures were raised overnight on minimal media glucose agar at 30°C to isolate three biological replicates. Each culture was then raised for 48 hours in case amino acid media with defined metals to deplete the

bacteria's metal ion stores. Cultures were then rinsed twice in PBS, normalized to an OD600 of 0.05, suspended into three different bacterial medium treatments, and aliquoted into a 96 well plate in technical quadruplicate. The treatments consisted of dark LB + NaOH vehicle, dark LB + 250μM 2,2'-bypyridyl + NaOH vehicle, and dark LB + 250μM bypyridyl + 11.5μM hemin chloride in NaOH. Hemin chloride was suspended immediately before addition to cultures to reduce degradation and the final concentration of NaOH in each treatment was 100μM. Cultures were incubated at 30°C in a Biotek Synergy H1 shaking microplate reader while taking OD_{600} readings every 15 minutes for 42.5 hours. Growth curves were analyzed via the Growthcurver R package (Sprouffske and Wagner 2016).

Cell free conditioned medium assays were used to determine if HrpB impacted *X. nematophila*'s ability to compete with other microbes. 15 ml dark LB overnight cultures of *X. nematophila* wild type (HGB0800), *hrpB*::*kan*R (HGB2301), and the *hrpB*::*kan*R *att*Tn7:: *tonBhrpAB* complemented strain (HGB2390) were raised in biological triplicate. These cultures were rinsed thrice in dark LB supplemented with 250 μM bypyridyl and suspended in that same medium. Cultures were incubated for 3 hours in dark LB supplemented with 2,2'-bypyridyl in order to maximize expression of HrpC. The cultures were then spun down at ~3200xG for 10 minutes and the spent medium was filter sterilized with a 0.22 μm filter to generate cell free conditioned medium (henceforth CFCM). CFCM was stored at -20°C until use. For growth curves, *X. nematophila* wild type (HGB0800), *X. nematophila* Websteri (HGB1419), and a strain of *E. fecali*s isolated from a *Manduca sexta* infected with *X. nematophila* (HGB2115) were grown in the CFCM. 3-6 biological replicates of each strain were grown in each previously generated batch of CFCM for a total of 12 biological replicates. Cultures were grown shaking for 36 hours taking OD_{600} readings every 15 minutes. Growth curves were analyzed via the Growthcurver R package (Sprouffske and Wagner 2016).

Proteolytic activity of the mutant was measured using LB agar supplemented with 3% skim powdered milk. Lipolytic activity of the mutant was measured using Tween20 agar at 43 hours (Sierra 1957). Motility of the *hrpB* mutant was measured on 0.125% LB agar plates in a humidity chamber, taking measurements at 16 and 43 hours. For biofilm formation assays, 240µl cultures were grown aerobically, without light, in technical quadruplicate in a 96 well plate for 48 hours. Then the wells were decanted, rinsed 3 times, and stained with 0.1% crystal violet solution for 15 min. After drying the crystal violet was suspended in 95% ethanol for 1 hour and optical density was measured at 590nm to enumerate the amount of crystal violet retained by biofilms. Motility, lipolysis, proteolysis, and biofilm data were subjected to one-way ANOVA analysis with a Tukey's HSD Test. Virulence of *X. nematophila* was tested using lab reared *Manduca sexta* (tobacco hornworm moth) as previously reported (Hussa and Goodrich-Blair 2012). Briefly, *Manduca sexta* eggs were purchased from Carolina Scientific, surface sterilized in a 0.6% bleach solution, and raised at 26°C with a 16hr light: 8hr dark photoperiod until they reached 4th instar. Cultures of WT (HGB0800), an *ecotin*::*kan*R control (HGB2275), and *hrpB*::*kan*R (HGB2301) were diluted to ~30 CFUs/µl and then 10 µl of each was injected into *M. sexta*. Insects were observed regularly for death over a 60-hour period. Injections were performed in biological triplicate with each biological rep. consisting of 10 or 15 technical replicates for a total of 35 insects per treatment. Survival curves were analyzed using a series of Mantel-cox tests (Mantel 1966).

To test the impact of T11SS dependent hemophilin secretion on the fitness of *X. nematophila*'s mutualistic nematode host *S. carpocapsae*, nematodes were raised on LA plates containing lawns of *X. nematophila* wild type (HGB0800), *hrpB*::*kan*R (HGB2301), and the *hrpB*::*kan*R *att*Tn7::*tonBhrpAB* complemented strain (HGB2390) in biological sextuplicate. Prior to addition, axenic nematodes were generated by collecting gravid adult *S. carpocapsae* from plates of *X. nematophila* and surface sterilizing them as previously reported (Yadav et al. 2015). 5000 nematodes

were added to the surface of each experimental lawn and incubated at 25°C for 8 days to allow growth and reproduction. After this incubation the lawns containing bacteria and nematodes are transferred into White traps in order to collect emerging infective juveniles (IJs). IJs were collected at 6, 12, and 18 days post trapping and enumerated via microscopy. On day 18 images of the plate surfaces were taken using an Olympus SZ macroscope and adherent material was sampled from each plate in triplicate. 18x18 mm glass cover slips were arrayed to avoid overlap and firmly pressed into the medium. Cover slips were then collected, dried completely, stained with 0.5% crystal violet in a solution of methanol:water (1:4) for 60 seconds, rinsed 2x in sterile H_2O, and boiled in 300 µl of 95% ethanol to solubilize the adherent crystal violet. Samples were diluted 1:10 to keep optical densities within the optimal reading range of our instrument (BioTek Synergy H1). To quantify crystal violet, absorbance was measured at OD590 and analyzed with a Tukey's HSD test (Tukey 1949).

For insect experiments the nematodes collected from the first three biological replicates of the LA plate emergence experiment were then surface sterilized in 1% bleach for 3 minutes, concentrated to 10 nematodes/µl, and injected into *Galleria mellonella* larvae purchased from Grubco (200 IJs in 20 µl). Each replicate was injected into 3 insects for a total of 9 per treatment. Additionally, 10 *G. mellonella* were injected with PBS to act as a negative control. Mortality was tracked over the next 8 days and then all insect cadavers were placed into White traps in order to collect emerging infective juveniles (IJs). IJs were collected at 6, 12, and 18 days post trapping and enumerated via microscopy.

To determine if HrpB is essential for nematode colonization, axenic nematodes were generated as described above and then added in biological triplicate to the surface of LA plates containing lawns of WT (HGB0800), GFP control (HGB2106), an *ecotin*::*kan*R + GFP control (HGB2305), and *hrpB*::*kan*R + GFP (HGB2302). Some replicates of 2106 were lost due to adult

nematodes contaminating the IJ stocks. Lawns were incubated at 25°C for 9 days and then transferred to white traps to collect emerging infective juveniles. Nematodes were collected and stored in sterile water at 25°C in tissue culture flasks. At 72 hours post-emergence, frequency of colonization was determined by counting colonized nematodes using fluorescence microscopy on a Keyence BZX-700, and bacterial abundance was determined via surface sterilization in 1% bleach solution and homogenization of nematodes using a motorized mortar and pestle and subsequent plating on minimal media glucose agar. These agar plates were incubated for 30 hours before CFUs were counted. At 3-weeks post-emergence colonization frequency was enumerated a second time to ensure that bacteria did not have a persistence defect. Colonization data were subjected to a one-way ANOVA with a Tukey's HSD Test (Tukey 1949).

Phylogenetic analysis

Phylogenetic analysis was performed as described previously (Grossman, Mauer, et al. 2022). Briefly, T11SS proteins from *Xenorhabdus* and *Photorhabdus* were extracted using MicroScope MaGe (Vallenet et al. 2017), individually aligned using Muscle v3.8.31 (Edgar 2004), concatenated using Sequence Matrix v1.8 (Vaidya, Lohman, and Meier 2011), and trimmed of nucleotide gaps using TrimAL v1.3 (Capella-Gutiérrez, Silla-Martínez, and Gabaldón 2009). JmodelTest v2.1.10 (Darriba et al. 2012) was used to choose the GTR + γ + I substitution model for maximum likelihood analysis. Maximum likelihood analyses were performed via RAxML v8.2.10 (Stamatakis 2014) using rapid bootstrapping and 1,000 replicates and were visualized via Dendroscope v3.6.2 (Huson and Scornavacca 2012).

Acknowledgements:

This work was funded in part by the University of Tennessee-Knoxville. This work was funded in part by a grant from the Clifford Craig Foundation. We thank Dr. Brianna Atto for useful discussions and Sarah Kauffman for technical assistance.

Chapter 4: A Surface Exposed, Two-Domain Lipoprotein Cargo of a Type 11 Secretion System Promotes Colonization of Host Intestinal Epithelia Expressing Glycans

Authors: Alex S. Grossman, Cristian A. Escobar, Erin J. Mans, Nicholas C. Mucci, Terra J. Mauer, Katarina A. Jones, Cameron C. Moore, Paul E. Abraham, Robert L. Hettich, Liesel Schneider, Shawn R. Campagna, Katrina T. Forest, and Heidi Goodrich-Blair

Publication disclosure:

A previous version of this chapter was published in Frontiers in Microbiology as part of their research topics E-book Bacterial Secretion Systems, Volume II (Grossman, Escobar, et al. 2022) (https://doi.org/10.3389/fmicb.2022.800366). Alex S. Grossman, Cristian A. Escobar, and Erin J. Mans share primary authorship on the original manuscript. Alex S. Grossman and Heidi Goodrich-Blair wrote the text of this chapter. Alex S. Grossman performed the secretion experiments in *E. coli* and *X. nematophila* and generated all associated figures, with the exception of the flow cytometry data, which was generated by Terra J. Mauer, and the Lrp mutant dot blots, which were performed collaboratively by Alex S. Grossman and Cameron C. Moore. Alex S. Grossman performed the chitin binding experiment and generated the associated figure. Cristian A. Escobar and Katrina Forest performed the biophysical characterization of NilC, circular dichroism, NMR, and partial proteinase assays and generated the associated figure. Erin J. Mans performed the adult nematode/AIC colonization assays, as well as the differential lectin analysis and differential NilC analysis of those experiments. Statistical analysis of the colonization assays was performed by Liesel Schneider, the associated figure was generated collaboratively by Erin Mans, Liesel Schneider, and Heidi Goodrich-Blair. Sample preparation for proteomics and metabolomics was performed by Terra J. Mauer, mass spectrometry was performed by Katarina A. Jones, Paul E. Abraham, Robert L. Hettich, and Shawn R. Campagna. Analysis of proteomics and metabolomics was performed by Katarina A. Jones and Nicholas C. Mucci. All authors contributed to manuscript revision, read, and approved the published version.

Abstract:

The only known required component of the newly described type 11 secretion system (T11SS) is an outer membrane protein (OMP) of the DUF560 family. $T11SS_{OMPs}$ are broadly distributed across proteobacteria, but properties of the cargo proteins they secrete are largely unexplored. We report biophysical, histochemical, and phenotypic evidence that *Xenorhabdus nematophila* NilC is surface exposed. Biophysical data and structure predictions indicate that NilC is a two-domain protein with a C-terminal, 8-stranded β-barrel. This structure has been noted as a common feature of T11SS effectors and may be important for interactions with the $T11SS_{OMP}$. The NilC N-terminal domain is more enigmatic, but our results indicate it is ordered and forms a β-sheet structure, and bioinformatics suggest structural similarities to carbohydrate-binding proteins. *X. nematophila* NilC and its presumptive $T11SS_{OMP}$ partner NilB are required for colonizing the anterior intestine of *Steinernema carpocapsae* nematodes: the receptacle of free-living, infective juveniles and the anterior intestinal cecum (AIC) in juveniles and adults. We show that, in adult nematodes, the AIC expresses a Wheat Germ Agglutinin (WGA)-reactive material, indicating the presence of *N*-acetylglucosamine or *N*-acetylneuraminic acid sugars on the AIC surface. A role for this material in colonization is supported by the fact that exogenous addition of WGA can inhibit AIC colonization by *X. nematophila*. Conversely, the addition of exogenous purified NilC increases the frequency with which *X. nematophila* is observed at the AIC, demonstrating that abundant extracellular NilC can enhance colonization. NilC may facilitate *X. nematophila* adherence to the nematode intestinal surface by binding to host glycans, it might support *X. nematophila* nutrition by cleaving sugars from the host surface, or it might help protect *X. nematophila* from nematode host immunity. Proteomic and metabolomic analyses of wild type *X. nematophila* compared to those lacking *nilB* and *nilC* revealed differences in cell wall and secreted polysaccharide metabolic pathways. Additionally, purified NilC is capable of binding peptidoglycan, suggesting that periplasmic NilC may interact with the bacterial

cell wall. Overall, these findings support a model that NilB-regulated surface exposure of NilC mediates interactions between *X. nematophila* and host surface glycans during colonization. This is a previously unknown function for a T11SS.

Introduction:

Bacteria rely on secretion systems to convey proteins across membranes to the cell surface and extracellular environment. In host-associated bacteria, such effector proteins, which can include surface-exposed lipoproteins, mediate acquisition of nutrients, signaling interactions with host cells, mechanical interactions with host surfaces, and specificity in host range. These processes make effector proteins potential targets for pharmaceutical treatment and vaccine development (Kinkead et al. 2018; Konovalova and Silhavy 2015; M. M. Wilson and Bernstein 2016). Bacterial lipoproteins are classified by N-terminal lipidation at a cysteine residue but, otherwise, are diverse in sequence, function, and subcellular localization. The mechanisms by which certain classes of proteins, including lipoproteins, are targeted to and oriented within the outer-membrane are still largely unknown. The newly described type 11 secretion system (T11SS), comprising an outer membrane protein (OMP) containing a DUF560 (a domain of unknown function 560), is broadly distributed among proteobacteria and mediates translocation of lipoprotein and a soluble protein cargo across the outer membrane (Bhasin, Chaston, and Goodrich-Blair 2012; Grossman, Mauer, et al. 2022; Heungens, Cowles, and Goodrich-Blair 2002; Hooda, Lai, and Moraes 2017).

A sequence-similarity-based network analysis provided functionally relevant categorization of these T11SS OMPs, hereafter referred to as $T11SS_{OMPs}$, into 10 clusters. Cluster 1, which contains the largest number of $T11SS_{OMPs}$, was further refined into three subclusters: A, B, and C. Clusters 1A and 1B contain the few characterized T11SS OMPs, and its members predominantly are encoded by microbes isolated from animals (Grossman, Mauer, et al. 2022). *Neisseria meningitidis* Slam1 and Slam2

are T11SS$_{OMPs}$ responsible for secretion of transferrin-, lactoferrin-, factor H- or hemoglobin/haptoglobin-binding lipoproteins. *X. nematophila* HrpB and *Acinetobacter baumanii* HsmA are T11SS$_{OMPs}$ that secrete the soluble (non-lipidated) hemophores HrpC and HrpA, respectively. *Xenorhabdus nematophila* NilB is a T11SS$_{OMP}$ that, along with the associated lipoprotein NilC, is necessary for mutualistic colonization of the nematode *Steinernema carpocapsae* (Bateman et al. 2021; Cowles and Goodrich-Blair 2008; Fantappiè et al. 2017; Grossman, Mauer, et al. 2022; Hooda et al. 2016; Hooda, Lai, and Moraes 2017; da Silva et al. 2019).

Current evidence indicates T11SS$_{OMPs}$ have specificity for their cargo. When expressed in *Escherichia coli* BL21-C43, *Neisseria* T11SS-1A homologs Slam1 and Slam2 do not translocate each other's cargo proteins, nor can Slam1 translocate the *E. coli* periplasmic lipoprotein PgaB (Hooda et al. 2016). T11SS cargo proteins for which there are known structures (TbpB, LbpB, fHbp, HupA) have limited sequence similarity but a common organization of an N-terminal effector domain and a C-terminal 8-stranded β-barrel that may direct a cargo for secretion (Hooda, Lai, and Moraes 2017). The TbpB C-terminal β-barrel is the first part of the cargo protein to be surface exposed during secretion (Hooda et al. 2016), which may indicate this domain initiates interactions with T11SS$_{OMPs}$. This suggests a general conceptual framework in which the cargo C-terminus is necessary for its interaction with a T11SS$_{OMP}$, while the N-terminus encodes the host interaction (or other) effector domain.

An *X. nematophila* T11SS$_{OMP}$, NilB, is encoded near an outer membrane lipoprotein NilC on a locus known as Symbiosis Region 1 (SR1) (Bhasin, Chaston, and Goodrich-Blair 2012; Cowles and Goodrich-Blair 2004; Heungens, Cowles, and Goodrich-Blair 2002). In *X. nematophila*, the SR1 locus, which encodes both *nilB* and *nilC*, is necessary and sufficient for normal levels of colonization of *S. carpocapsae* intestines (Chaston et al. 2013; Cowles and Goodrich-Blair 2008). *nilB* and *nilC* are not

organized in an operon and have two different promoter sequences, but they are coordinately downregulated at the transcriptional level by the transcription factors NilR and Lrp in a synergistic manner (Cowles and Goodrich-Blair 2006). Although NilB and NilC function in *S. carpocapsae* colonization is well-established, their cellular and molecular functions remain unclear. Given that NilB is a member of the $T11SS_{OMP}$ family that facilitates the surface exposure of target lipoproteins, we considered the model that it functions to facilitate NilC surface exposure, and that NilC is a host-interaction effector. We describe here experiments to determine if NilB and NilC form a $T11SS_{OMP}$–host interaction effector pair by determining if NilC lipoprotein can be surface exposed, if NilB facilitates NilC lipoprotein surface exposure, and what effector activity one or both of them might have with respect to host interactions.

Results:

NilB Facilitates NilC Surface Exposure During Heterologous Expression in *Escherichia coli*

To determine if NilC is a cargo protein for NilB, we monitored NilC surface exposure during heterologous expression in *Escherichia coli*, with and without co-expression with NilB. We constructed plasmids in which expression of *nilC* and, when present, *nilB* are under control of an IPTG-inducible T7 RNA polymerase (see section "Materials and Methods"). NilC surface exposure experiments were conducted by immuno-dot blotting with an anti-NilC antibody. As expected, *E. coli* without an expression plasmid did not react with the antibody. For NilC-expressing strains, we found that NilC was present on the surface of *E. coli* at significantly higher levels in the presence of NilB than in its absence after cultivation in either LB (Fig. 4.1B and 4.2) or a minimal glucose medium supplemented with 1% LB (MM:LB) (Fig. 4.1A). When considering all treatments together, we noted a positive correlation between the total amount of NilC expressed and the amount on the

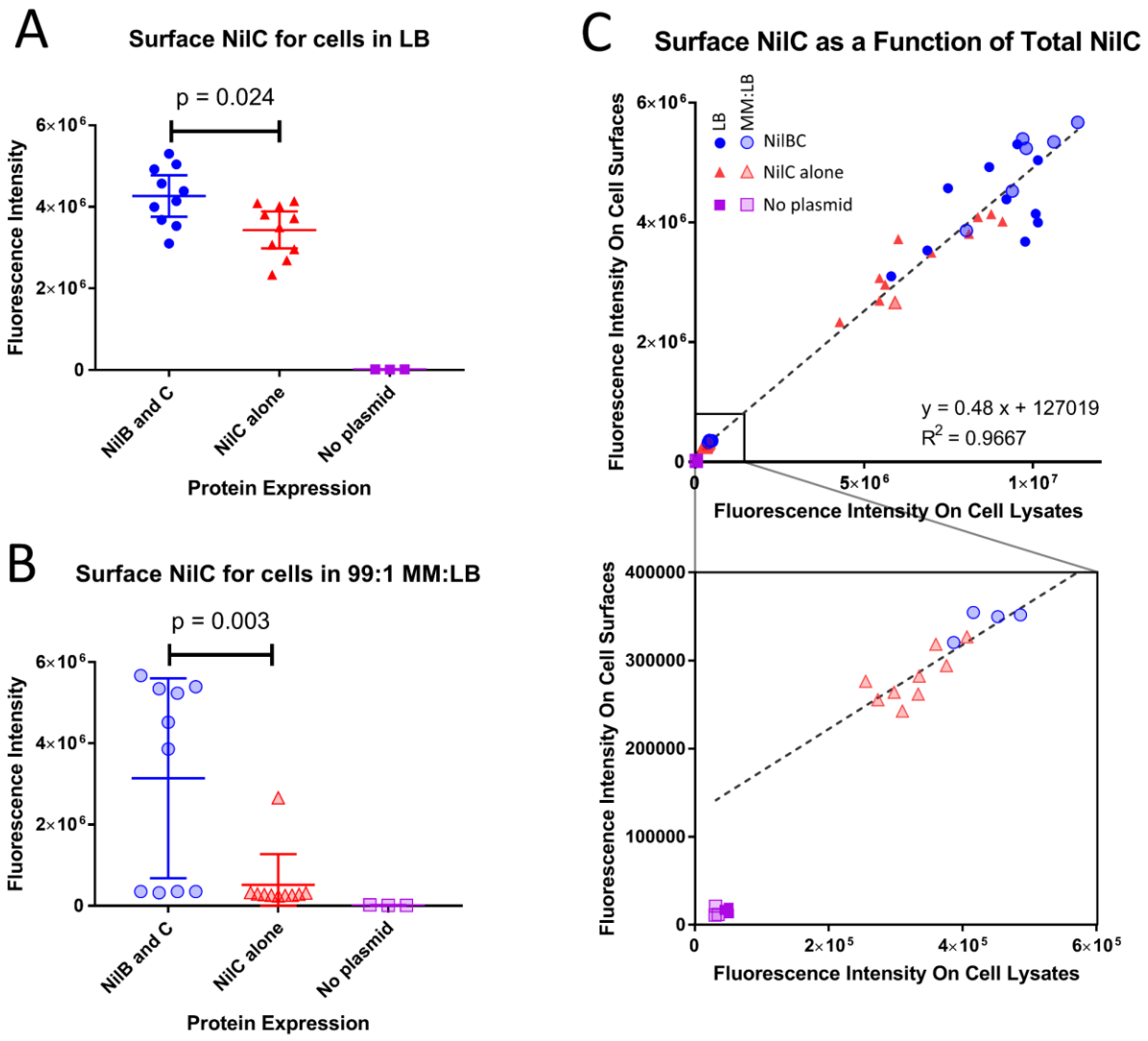


Figure 4.1. **NilB positively influences the amount of NilC present on the surface of *E. coli* when co-expressed.** *E. coli* BL21 DE3 C43 was transformed with pETDuet-1 plasmids carrying *nilC* alone (red triangles) or both *nilC* and *nilB* (blue circles) under control of the inducible T7 polymerase promoter. A no-plasmid control (purple squares) was included to account for a potential cross-reacting signal. Surface exposure of NilC was quantified by immunofluorescence with anti-NilC antibody dot blotting of cells grown in LB [**(B,C)** solid symbols], or MM:LB [**(A,C)** transparent symbols]. p-values were determined via one-way ANOVA with Tukey's multiple comparison. Total cellular levels of NilC were quantified by immunoblotting whole cell lysates from the same samples and surface NilC as a function of total cellular NilC is shown in **(C).** A linear regression of data from both media conditions revealed that total NilC expression was a strong predictor of surface exposure of NilC (R-squared = 0.9667). The bottom graph provides an expanded view of the region of the top graph indicated with a square.

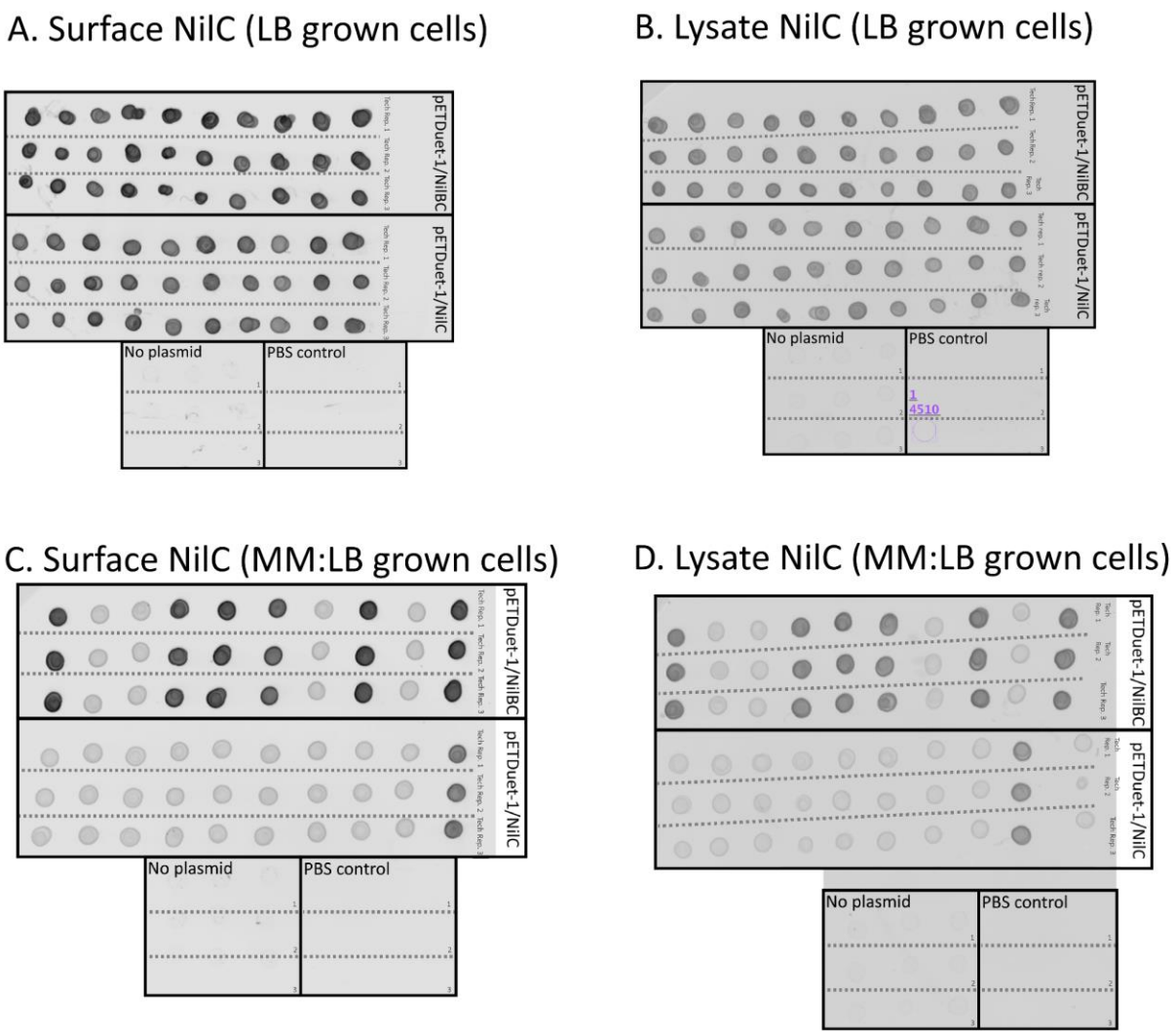


Figure 4.2. **Immuno-dot blots reveal impact of NilB on surface exposure of NilC.** Whole cells (**A,C**) or lysates (**B,D)** of *E. coli* carrying pET-Duet containing P_{T7}-*nilC* and P_{T7}-*nilB* (top three rows), P_{T7}-*nilC* only (middle three rows), or no plasmid (bottom three rows, left) after growth in LB. (**A,B**) or minimal medium glucose (MM) supplemented with 1% LB (MM:LB) and 2 h of induction with IPTG.1x PBS was used as a control for background fluorescence (bottom three rows, right). Samples were spotted in technical triplicate onto nitrocellulose membranes and probed with rabbit anti-NilC primary antibody and goat anti-rabbit secondary antibody bound to a IRDye 680RD fluorophore. Emission intensity was quantitated using an Odyssey Infrared Imaging System and displayed in Figure 1. For the MM:LB treatments, after subculturing, 4/10 and 9/10 replicates of *E. coli* pET-DUET *nilB*/*nilC* and pET-DUET *nilC* alone, respectively, did not leave lag phase growth (measured via OD_{600}) and did not express high levels of NilC (either in lysates or on the surface).

cell surface. This indicates that NilC can efficiently reach the *E. coli* cell surface even in the absence of NilB (Fig. 4.1C). However, when NilC was expressed alone, its levels on the cell surface were 16.5% or 80% of those observed when co-expressed with NilB, during incubation in MM:LB or LB, respectively, indicating that NilB enhances the surface exposure of NilC. Furthermore, supernatants of induced *E. coli* cells expressing *nilC* with or without *nilB* did not have NilC levels detectable by the immuno-dot-blotting assay, indicating that cell lysis does not explain the surface NilC levels detected (Fig. 4.3). These data indicate that the presence of NilB supports greater overall levels of NilC expression in *E. coli* and that, like its $T11SS_{OMP}$ family relatives, NilB facilitates surface exposure of its cargo: NilC lipoprotein. NilC surface exposure enhancement by, rather than complete dependence on, NilB $T11SS_{OMP}$ is similar to the relationship between *N. meningitidis* fHbp and its $T11SS_{OMP}$, Slam1. In either *E. coli* or *N. meningitides*, the fHbp lipoprotein can be surface exposed even in the absence of Slam1, particularly when expressed at high levels, but fHbp surface levels are elevated in the presence of Slam1 (Fantappiè et al. 2017; Hooda et al. 2016; da Silva et al. 2019).

NilC Is a Two-Domain Protein with a Predicted C-Terminal, 8-Stranded β-Barrel

Given these data that indicate NilC is a surface exposed cargo protein for the NilB $T11SS_{OMP}$, we sought to understand its molecular properties and biological function. Our investigation of the structural and biophysical characteristics of NilC required purified protein. A recombinant NilC soluble domain ($NilC_{22–282}$) with a C-terminal 6X-His tag was expressed in *E. coli* and purified by nickel affinity and size exclusion chromatographies. The C-terminal His-tag did not impede the ability of NilC to function in colonization when it was expressed instead of the native sequence in *X. nematophila* (Fig. 4.4). The purified protein is remarkably highly soluble, remaining in solution until at least 80 mg/ml. Secondary structure prediction based on the NilC amino acid sequence indicates a protein that is largely random coil (~71–74%) with only ~24% of

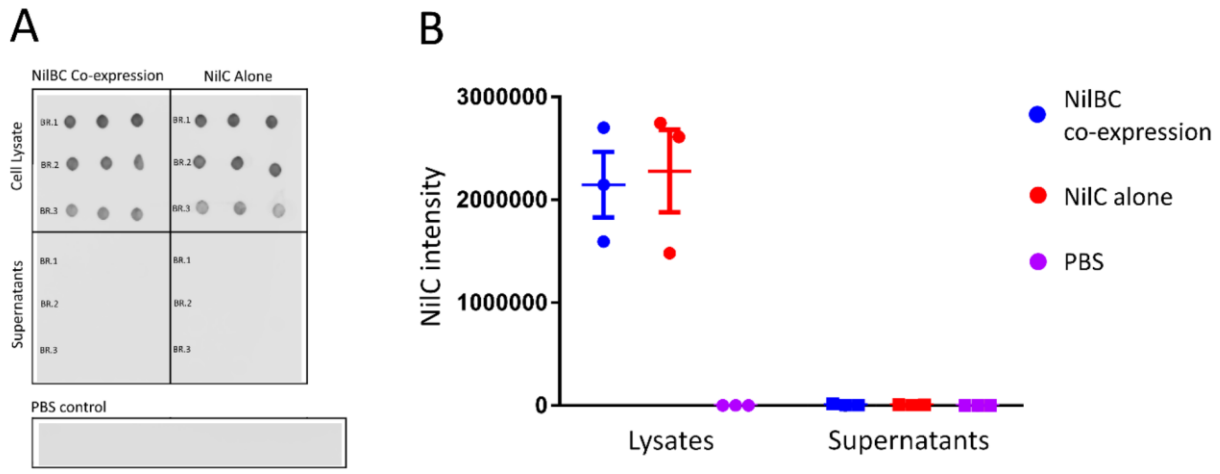


Figure 4.3. **NilC surface exposure is not driven by cell lysis.** (A) Immuno-dot blots of *E. coli* BL21 C43 co-expressing NilB and NilC or expressing NilC alone. Sterile PBS was used as a control for background fluorescent signal. Horizontal dots represent technical replication; vertical dots represent biological replication. The immunoblot was probed with anti-NilC antibody and a secondary antibody bound to a IRDye 680RD fluorophore. Emission intensity was quantitated using an Odyssey Infrared Imaging System. (B) Cellular lysate fractions for both strains demonstrate comparable expression of NilC in both strains. Error bars display the standard error of the mean. A Tukey's honestly significant difference test was performed and revealed no significant difference between NilC present in the spent media of either strain and NilC present in the sterile PBS control.

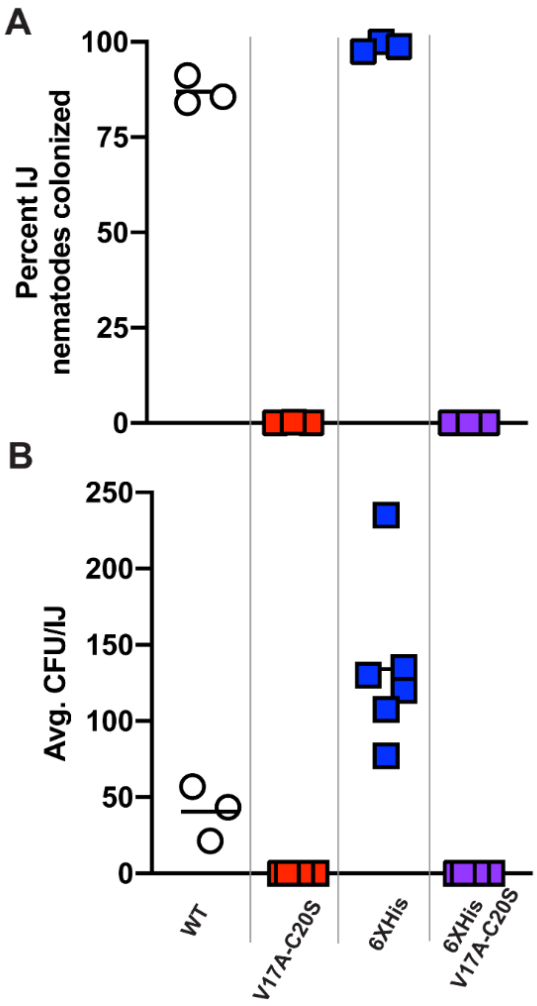


Figure 4.4. **Histidine tag does not negatively affect NilC function in colonization whereas loss of lipidation site does.** *X. nematophila* colonization of *S. carpocapsae* nematodes was measured as **A)** percent of infective juvenile (IJ) nematodes with visible green-fluorescent protein expressing *X. nematophila* within the intestinal receptacle, observed by fluorescence microscopy or **B)** the average colony forming units (CFU) per IJ as determined by surface sterilization, grinding and plating. Wild type (WT) (open circles) and *X. nematophila* Δ*SR1* carrying ectopic SR1 with *nilC* modifications (filled squares): two signal sequence amino acid changes (V17A-C20S) that eliminate the lipobox (red and purple squares), and a C-terminal 6X-His tag (6XHis) (blue and purple squares) insertion, either independently (red and blue squares) or combined (purple squares). Strains used to monitor colonization frequency expressed the green fluorescent protein from the *kefA* locus.

the protein predicted to form β-strand and little to no α-helix predicted (~0–4%) (Fig. 4.5A). This prediction was corroborated by circular dichroism (CD). The $NilC_{22–282}$ CD spectrum is noteworthy for the fact that it has apparently no characteristic negative signal of α-helix (208 and 222 nm) or β-strand (218 nm) secondary structure elements (Figure 4.5B). On the other hand, the NilC spectrum has a positive band centered at 200 nm, which indicates it is not completely unfolded, as disordered proteins have a negative band at 195 nm. We suspected that a negative signal at 218 nm for β-strand might be masked in the experimental spectra by the high concentration of random coil with a positive signal at a nearby wavelength. CD spectrum deconvolution using CONTIN and BeStSel algorithms supported this interpretation and indicates a β-strand content between 43 and 49%, with no significant presence of α-helix (~2%).

Tertiary structure predictions of NilC – both *via* well-established remote homology modeling in Phyre2 (Kelley et al. 2015) or *via* neural net analysis as implemented in RoseTTAFold (Baek et al. 2021), AlphaFold2 (Jumper et al. 2021), and their combined implementation in ColabFold (Mirdita et al. 2022) predict with high statistical significance the 8-stranded C-terminal β barrel of the TX1SS targeting domain. The structure of this domain is robustly modeled using *Haemophilus haemolyticus* hemophilin (PDB: 6OM5) as a template (Latham et al. 2020). The N-terminal effector domain, on the other hand, is poorly defined even by these recent powerful structure prediction methods, although all models of the N-terminal domain have a large fraction of random coil, ~7 β-strands or extended segments without well-defined secondary structure, and, essentially, no α-helical content (Fig. 4.5C). These predictions are an excellent match to our experimental measurements. Intriguingly, a surface electrostatic calculation of the model with the highest amount of secondary structure shows a large non-polar patch, an unexpected feature for a protein so readily soluble (Fig. 4.5D). Indeed, because the first 40 amino acids of the mature NilC sequence contain a high fraction

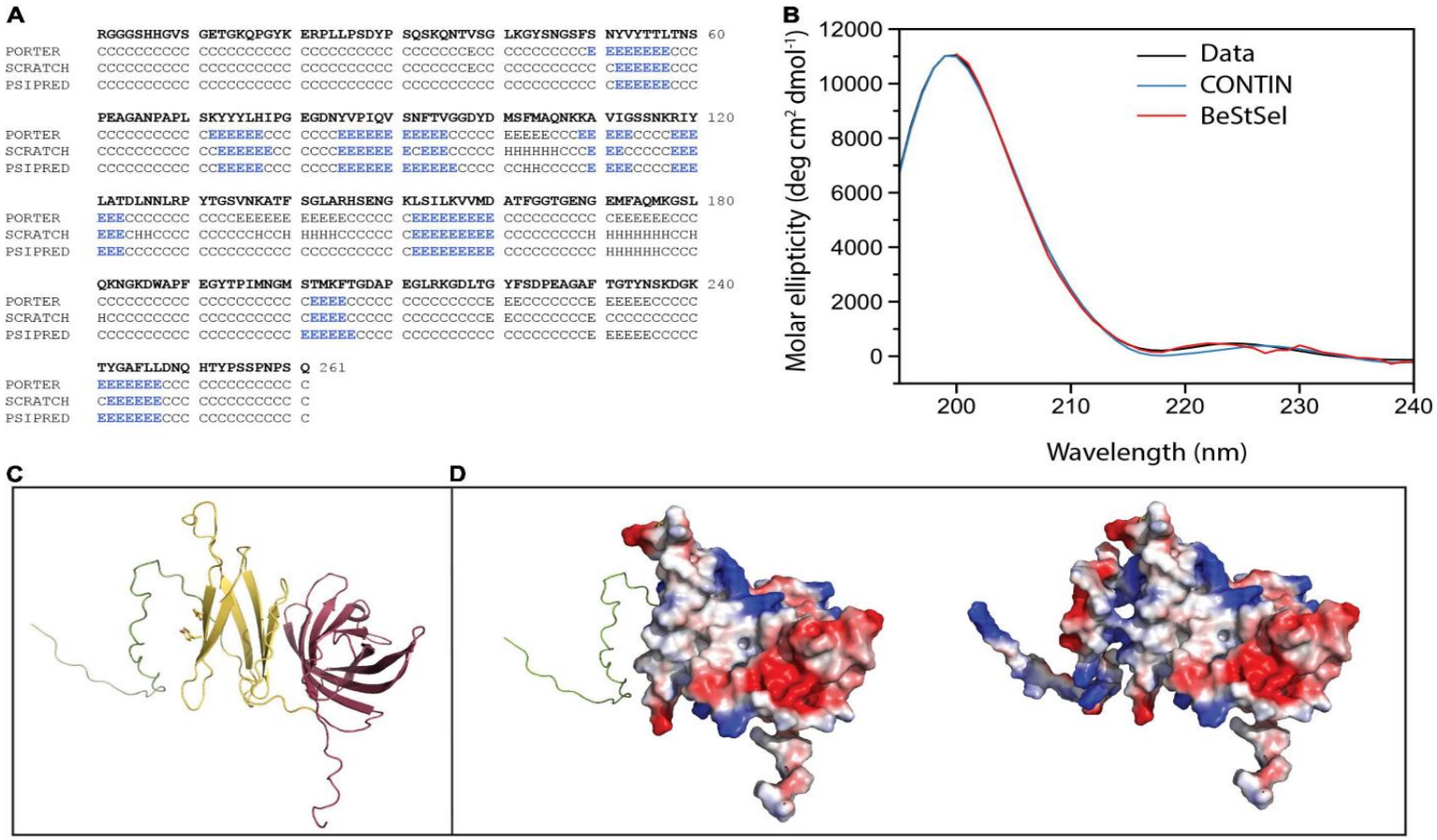
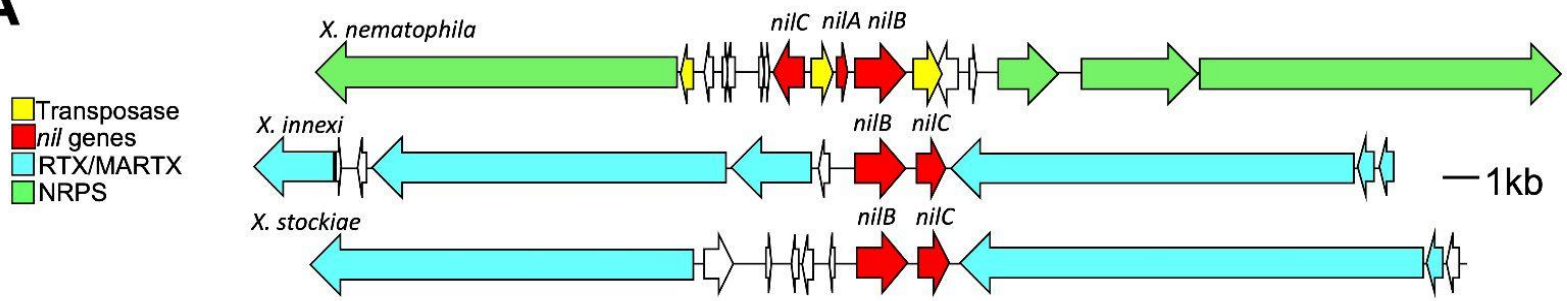


Figure 4.5. **NilC secondary structure and structural models. (A)** Secondary structure prediction for $NilC_{22–282}$ was performed using Porter, Scratch, and PsiPred algorithms (E β-strand, H α-helix, C random coil) with blue highlighting indicating the β-strand elements predicted by each. **(B)** A $NilC_{22–282}$ CD spectrum was collected at 25°C in an AVIV model 420 CD spectrometer. Deconvolution of the spectra using two algorithms leads to excellent agreement with experimental measurement. **(C)** The NilC structural model showing disordered N-term (green), a relatively flexible N-terminal effector domain (yellow), and an 8-stranded barrel T11SS-targeting domain (purple). The two tyrosines conserved among *Xenorhabdus* NilC homologs are shown. **(D)** On the left, the electrostatic surface of effector and barrel domains (red, negative; blue, positive; and white, neutral charge densities) is shown, and, on the right, the electrostatic surface of entire protein, suggesting the charged N-terminal extension may help solubilize the protein despite the hydrophobic patch.

of glycine and proline, we surmised they might be highly disordered. We thus created a second NilC construct, and expressed and purified $NilC_{62–282}$. This protein is significantly less soluble than $NilC_{22–282}$, bolstering the hypothesis that this N-terminal extension may protect the hydrophobic patch of purified $NilC_{22–282}$ in solution. We also noticed an unusually high concentration of tyrosine in the primary sequence, with 12 Tyr in the first 133 amino acids of the secreted protein (9%) (Fig. 4.6C). Tyrosine is an important element in sugar-binding sites for its aromatic stacking and, to a lesser extent, -OH hydrogen bond-donating abilities (Banno et al. 2017; Hudson et al. 2015; Quiocho 1986; Weis and Drickamer 1996). The only homologs of *X. nematophila nilC* identified to date occur in two other species (of ~23 with sequenced genomes) of nematode-associated *Xenorhabdus: X. innexi* and *X. stockiae*, both of which also contain the associated *nilB* gene (Grossman, Mauer, et al. 2022) (Fig. 4.6A,B). In each case, the NilC polypeptide is predicted to be a two-domain protein with a C-terminal barrel and conserved tyrosines at positions 75 and 77 of the mature sequence in the N-terminal presumed effector domain (Fig. 4.6C). Both of these tyrosines occur within a region predicted to be a β-strand (Fig. 4.5A,C).

To further our experimental characterization of $NilC_{22–282}$, we collected a 2D ^{1}H-^{15}N HSQC spectrum of ^{15}N labeled protein. The nuclear magnetic resonance (NMR) spectrum shows good signal dispersion, covering a range of more than 4 ppm on the proton axis, indicating NilC is well-folded under the conditions used (Felli and Pierattelli 2015) (Fig. 4.7). The most intense peaks, which are collapsed into the center of the spectrum, imply that some parts of the protein are highly flexible. Finally, we carried out limited protease digestion of $NilC_{22–282}$ in order to experimentally investigate the potential for stable subdomains. Several fragments of sizes between 12 and 27 kD are relatively stable intermediate breakdown products of NilC (Fig. 4.8A). Mass spectrometry of the peptide mixture after partial digestion identified the prominent fragment as having a mass of approximately 12,358 Da (Fig. 4.8B). Mapping of this mass onto the possible array of all Proteinase

A



B

	XnNilC	XnC23NilC	XnF1NilC	XnwebsteriNilC	XnanatolienseNilC	XiDSMNilC	XiATCCNilC	XsDSMNilC	XsKK7NilC	XsKJ12NilC
XnNilC	0.00	0.00	0.00	0.00	0.07	0.69	0.69	0.70	0.71	0.71
XnC23NilC	0.00	0.00	0.00	0.00	0.06	0.69	0.69	0.70	0.71	0.71
XnF1NilC	0.00	0.00	0.00	0.00	0.07	0.69	0.69	0.70	0.71	0.71
XnwebsteriNilC	0.00	0.00	0.00	0.00	0.06	0.69	0.69	0.70	0.71	0.71
XnanatolienseNilC	0.07	0.06	0.07	0.06	0.00	0.69	0.69	0.71	0.72	0.71
XiDSMNilC	0.69	0.69	0.69	0.69	0.69	0.00	0.00	0.69	0.75	0.70
XiATCCNilC	0.69	0.69	0.69	0.69	0.69	0.00	0.00	0.69	0.75	0.70
XsDSMNilC	0.70	0.70	0.70	0.70	0.71	0.69	0.69	0.00	0.54	0.06
XsKK7NilC	0.71	0.71	0.71	0.71	0.72	0.75	0.75	0.54	0.00	0.54
XsKJ12NilC	0.71	0.71	0.71	0.71	0.71	0.70	0.70	0.06	0.54	0.00

Figure 4.6. **Schematic diagram of the genomic context of *nilB* and *nilC* homologs**. **B)** CLUSTAL Omega uncorrected pairwise distance of and **C)** amino acid sequence alignment of NilC from strains of *X. nematophila* (ATCC19061; F1; C2-3; anatoliense, and websteri), *X. innexi* (DSM 16336 and HGB1681) and *X. stockiae* (DSM 17904, KK7.4, and KJ12). The amino acid alignment is divided according to the signal sequence and N- and C-terminal domains identified in this study. N-terminal domain tyrosines conserved among all *Xenorhabdus* NilC homologs are indicated with arrows, and those conserved among the *X. nematophila* NilC homologs are indicated by asterisks. **D)** Sequence homology alignment of segments (amino acids noted at the start and end of each segment) of NilC and d1gua, a galactose-binding domain-like, CBM4/9 family, and NilC and c4qpwa, a carbohydrate binding 1, glycosyl hydrolase domain, as detected using Phyre 2.0. The blue block arrows indicate regions of β-sheet secondary structure, either predicted (NilC) or known (d1gua and c4qpwA).

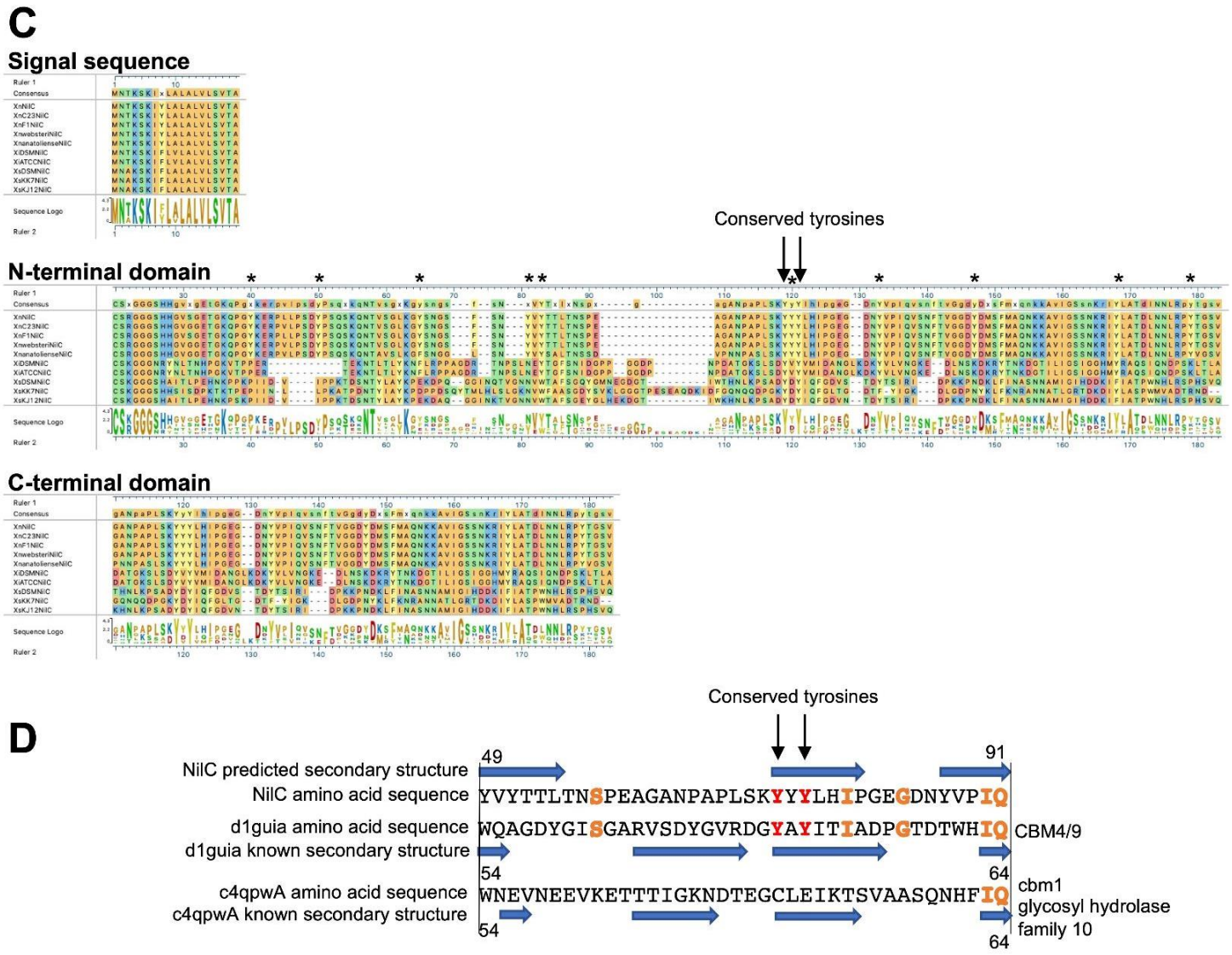


Figure 4.6. (continued)

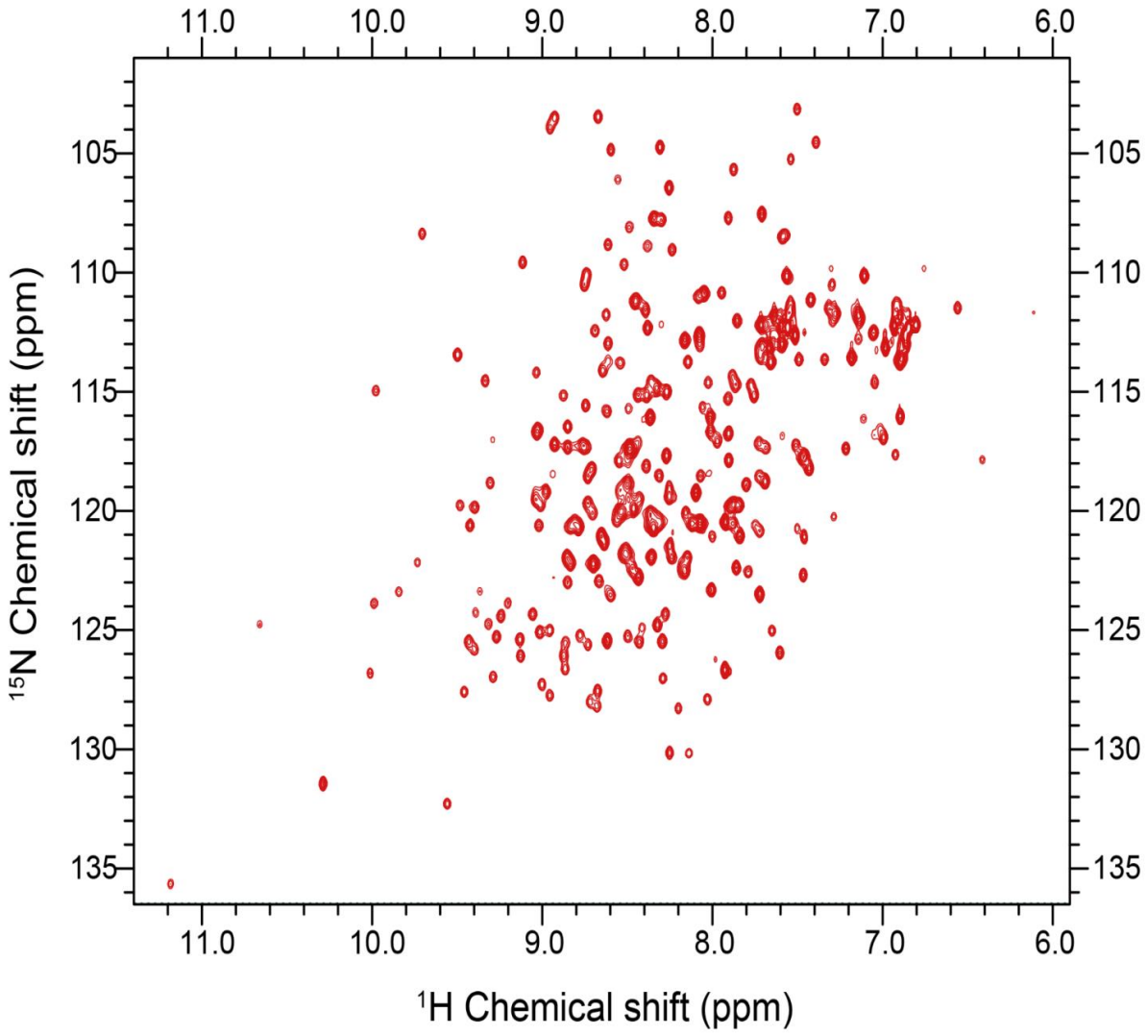


Figure 4.7. **NilC solution NMR**. 2D ^{1}H-^{15}N HSQC of ^{15}N uniformly labeled NilC was collected at 27 °C in a 750 MHz NMR spectrometer.

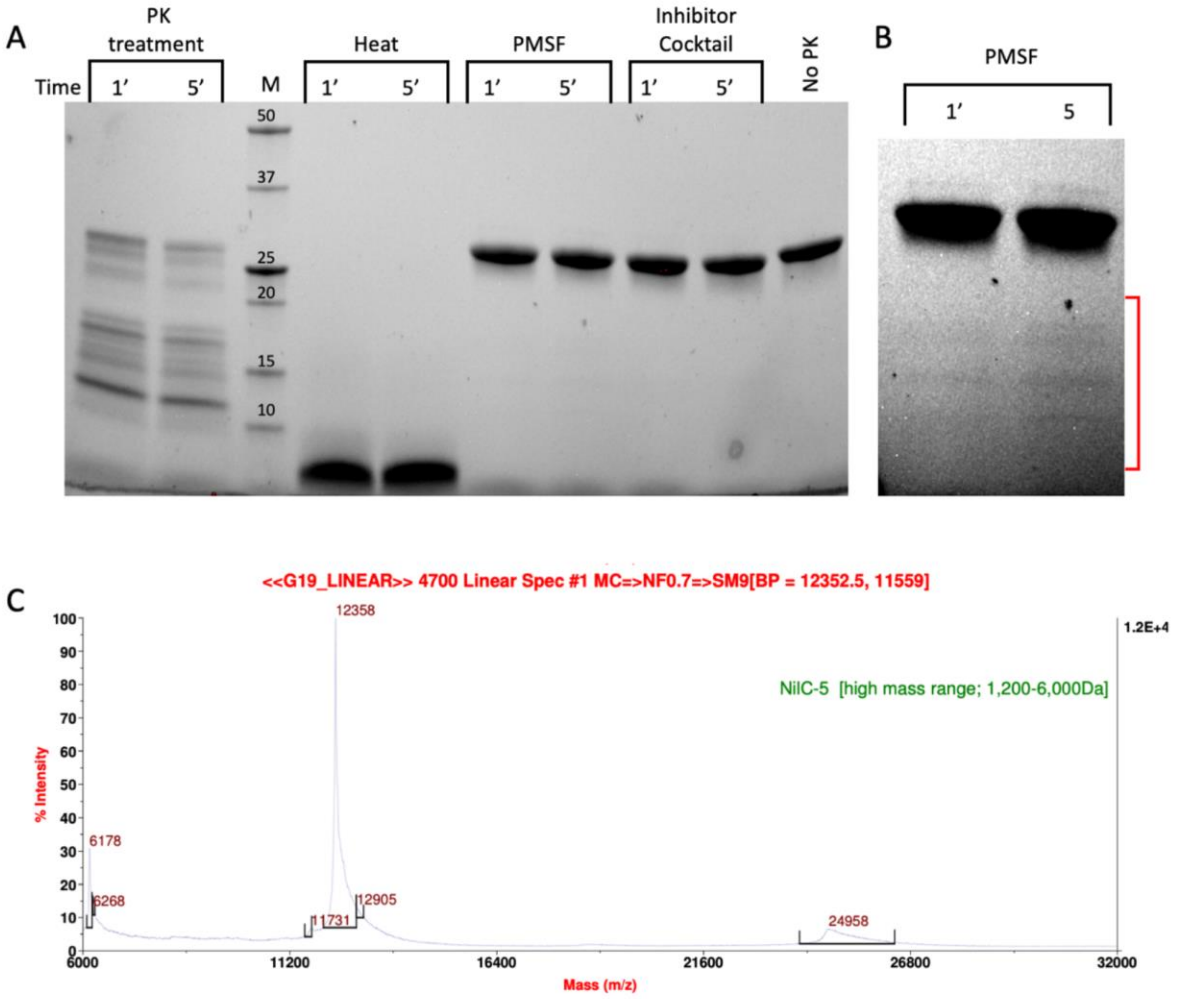


Figure 4.8. **NilC limited proteolysis**. (A) Proteinase K was used to digest purified NilC for 1 or 5 minutes, at which time samples were further subjected to no treatment, heat, or one of two protease inhibitor regimens. All samples were left at 4°C for ~1 week before running SDSPAGE. The heat treatment accelerated Proteinase K activity and led to complete digestion, whereas PMSF or protease inhibitor cocktail prevented any further digestion. (B) Indeed, very little proteolysis had occurred after 5 minutes, as seen in this higher contrast enlarged section of the same gel. Fortuitously, however, the extended time at 4°C created the opportunity for Proteinase K to act on the SDS-treated NilC, leading to the limited digestion. (C) The mass of the major breakdown product was determined by Electrospray Ionization Mass Spectrometry run in the positive mode.

K partial digestion fragments provides insight into the folded core of NilC. This mass corresponds to the C-terminal barrel without its last strand, implying once again that the two domains of NilC are structurally distinct, the first more open and flexible, the second stably folded (Fig. 4.5C).

Taken together, these data present a consistent picture of non-acylated NilC as a highly soluble and folded two-domain protein formed of sections of β-structure and a large fraction of random coil. The C-terminal domain is an 8-stranded barrel, adding support to the hypothesis that this structural motif is the common feature of T11SS cargo proteins. On the other hand, all of our experimental results indicate the N-terminal domain does not adopt a very well-defined structure, and that, under the conditions tested and in the absence of any potential ligands, NilC includes flexible regions and protease-accessible sites.

NilC Is Surface Exposed in *Xenorhabdus nematophila* Cells When Expressed at High Levels

Given the evidence that NilC can be surface exposed in *E. coli* and that its C-terminal domain is predicted to form an 8-stranded barrel T11SS-targeting domain, we revisited the question of whether NilC is also surface exposed in the native context of an *X. nematophila* cell. Previous whole-cell protease-digestion data demonstrated periplasmic orientation of the lipoprotein NilC, based on the observation of protease resistance of NilC in the whole cell but not lysate samples of wild type *X. nematophila* (Cowles and Goodrich-Blair 2004). Later, another protease digestion experiment to detect surface NilC was performed on an *X. nematophila* Δ*nilR* mutant, in which the absence of the transcription factor NilR causes *nilB* and *nilC* expression to be de-repressed (Bhasin, Chaston, and Goodrich-Blair 2012; Cowles and Goodrich-Blair 2006). In this analysis, slight shaving of NilC was detected in whole cells, indicating some surface exposure (Bhasin, Chaston, and Goodrich-Blair 2012). To further examine *X. nematophila* NilC cellular localization, we used the same

immuno-dot-blotting approach we had used for *E. coli* (Fig. 4.9A,B). We assessed NilC levels in *X. nematophila* wild type compared to an isogenic Δ*nilR* mutant. In addition, we included an isogenic pair of Δ*nilR* strains in which the SR1 locus has been deleted from its native locus, and reintroduced, either as a wild type sequence, or including a *nilCM1Z* start-to-stop codon mutation, at the *att*Tn*7* site downstream of the conserved gene *glmS*, which is involved in peptidoglycan biosynthesis (Choi et al. 2005; Craig 1996; J. E. Peters and Craig 2001). The latter strain served as a negative control for NilC detection by an antibody.

In cell lysates (indicative of overall expression), we detected antibody reactivity in all strains tested, including the *nilCM1Z* negative control, indicating some cross-reactivity by the polyclonal antibody (Fig. 4.9A). However, NilC-expressing strains had significantly higher antibody reactivity for whole cells than the *nilCM1Z* control, demonstrating the effective detection of NilC surface exposure using this method. When assessing levels of surface NilC in whole cell preparations of *X. nematophila*, we observed that wild type (*nilR*+) cells had little detectable NilC, while the Δ*nilR* strain had significantly higher signal intensity, indicative of surface NilC being present when *nilC* is de-repressed by deletion of *nilR*. Curiously, in strains in which SR1 had been introduced into the *att*Tn*7* site downstream of *glmS*, we observed a significant reduction in the amount of surface- exposed NilC detected relative to the isogenic parent Δ*nilR* strain, although total NilC levels were not significantly different (Fig. 4.9B).

NilR synergistically represses *nilB* and *nilC* with another transcription factor, Lrp (Cowles and Goodrich-Blair 2004, 2006). *X. nematophila* Lrp controls both mutualistic (with nematodes) and pathogenic (with insects) phenotypes. *X. nematophila* cells with fixed high levels of Lrp display higher levels of biofilm formation, nematode reproduction, and intestinal colonization, while *X. nematophila* cells with fixed low levels of Lrp display greater virulence in insects, compared to each Lrp

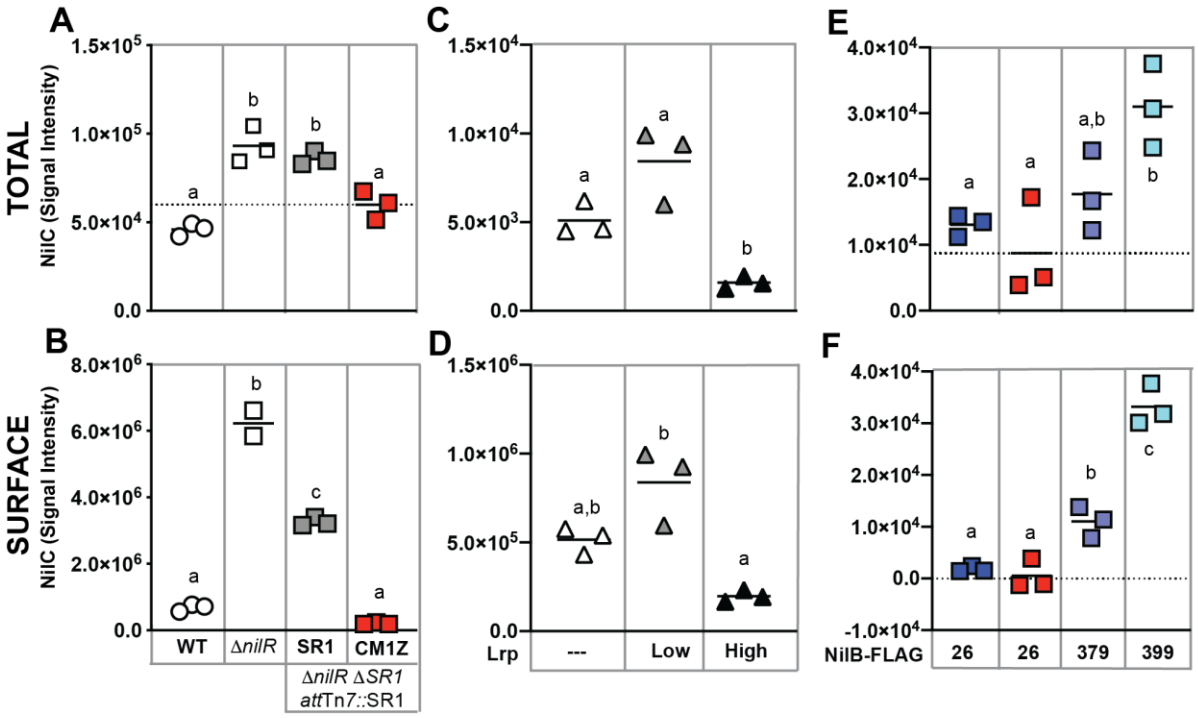


Figure 4.9. ***Xenorhabdus nematophila*** **NilC is surface exposed by NilB when its expression is de-repressed.** Cell lysate **(A,C,E)** or whole cell **(B,D,F)** preparations of LB-grown *X. nematophila* wild type (WT, circles), Δ*nil*R (squares), or Δ*lrp* (triangles) strains. At the *att*Tn*7* site of Δ*nil*RΔ*SR1* strains, wild-type SR1 (gray squares with a black outline) or SR1 with a *nilCM1Z* mutation (a red outline) and/or *nilB*-FLAG insertions (blue-shaded squares, sites of insertion noted beneath and on the NilB schematic bottom) were introduced **(A,B,E,F)**. Δ*lrp* strains were transformed with a vector-only control (white triangles) or plasmids expressing low or high levels of Lrp (gray and black triangles, respectively) **(C,D)**. Treatments were spotted onto nitrocellulose membranes. NilC was detected by immunoblotting with anti-NilC antibodies. Significantly different groups within each panel are indicated with different letters (tested using one-way ANOVA with Tukey's *post hoc* multiple comparisons analysis). Dashed axis lines indicate the mean of the *nilCM1Z* data points **(A,E,F)**.

expression on NilC surface exposure and found that, as expected, NilC levels and surface exposure were inversely correlated with Lrp levels (Fig. 4.9CD); *X. nematophila* cells expressing high levels of Lrp had significantly lower levels of total and surface-exposed NilC relative to cells lacking or expressing low levels of Lrp.

Having demonstrated above that NilB facilitates NilC surface exposure during heterologous expression in *E. coli*, we endeavored to determine if this is also the case in the native *X. nematophila* context. We used immuno-dot blotting to examine NilC surface exposure in cells expressing select NilB FLAG-tag variants encoded by the SR1 locus integrated at the *att*Tn*7* site in a Δ*nil*RΔ*SR1* strain background (Fig. 4.9E,F and 4.10). Our previous work using these and other FLAG-tag insertions across the length of the $T11SS_{OMP}$ NilB revealed a topology consisting of a ~138-amino acid N-terminal domain and 7 surface loops (Bhasin, Chaston, and Goodrich-Blair 2012) (Fig. 4.9). A variant with a FLAG-tag insertion at the mature, periplasmically oriented N-terminus of NilB (FLAG-26), expresses detectable levels of NilB and functions as well as wild type in colonization. Another variant, with an insertion in a transmembrane helix (FLAG-379), does not express detectable levels of NilB and is insufficient for colonization (Bhasin, Chaston, and Goodrich-Blair 2012). These two variants were used to compare surface levels of NilC surface exposure in the relative presence (FLAG-26) or absence (FLAG-379) of detectable NilB. Again, as was noted in other experiments, the total levels of NilC detected within cells correlated with the amount of NilC on the cell surface. Surprisingly, significantly less surface NilC was detected in the FLAG-26 strain (that expresses detectable levels of NilB) relative to the FLAG-379 strain (which does not express detectable levels of NilB) during growth in LB. This indicates that a basal level of NilC can be surface exposed in the absence of NilB (Fig. 4.9F), and that NilB can limit surface exposure of NilC. We next assessed the impact on NilC surface exposure of a FLAG-399 insertion in surface loop 6 of NilB. This variant expressed detectable NilB but is not functional in colonization.

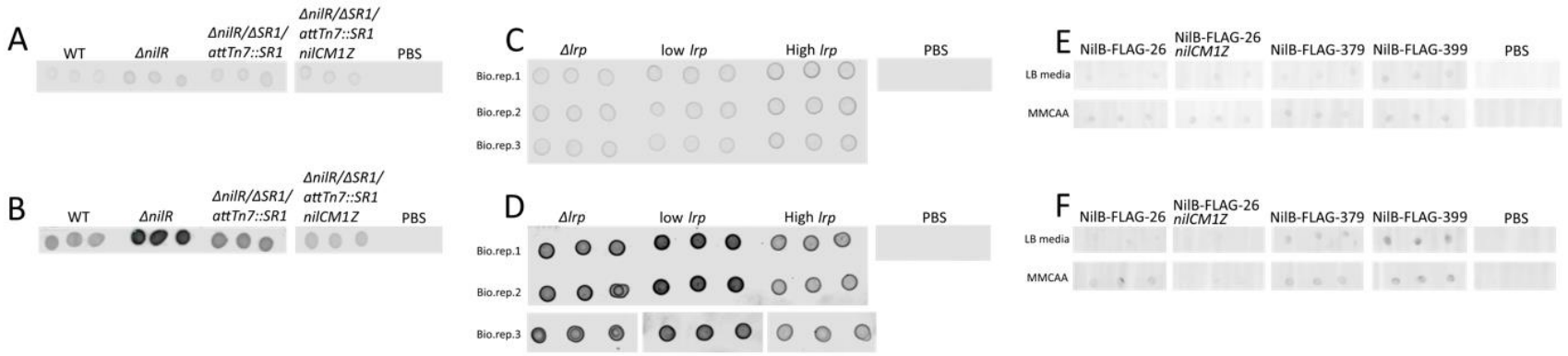


Figure 4.10. **Immuno-dot blots of *X. nematophila* cells differing in *nil* or *lrp* alleles**. lysates (**A,C,E**) or whole cells (**B,D,F**) of *X. nematophila*preparations of *X. nematophila* wild type (WT), ⊿*nilR*, ⊿*nilR* ⊿*SR1* with wild type SR1, SR1 with a *nilCM1Z* mutation (*nilCM1Z*) or *nilB*-FLAG insertions (FLAG-26, FLAG-379, FLAG-399) at the *att*Tn*7* site downstream of *glmS*, and ⊿*lrp* with vector only (⊿*lrp*) or plasmids expressing low or high levels of Lrp were spotted onto nitrocellulose membranes. All cells were grown in LB medium except for samples shown in E and F that were grown in minimal medium with casamino acids (MMCAA; quantified data shown in Fig. S8). NilC was detected by immunoblotting with anti-NilC antibodies. 1x PBS was used as a control for background fluorescence. Samples were spotted in technical triplicate onto nitrocellulose membranes and probed with Rabbit anti-NilC primary antibody and Goat anti-rabbit secondary antibody bound to a IRDye 680RD fluorophore. Emission intensity was quantified using an Odyssey Infrared Imaging System and displayed in Fig. 3 and Fig. S8.

The FLAG-399 variant had significantly more surface NilC than either of the other two FLAG variants, suggesting that mutating the sixth extracellular NilB loop results in increased secretion of NilC cargo protein. A similar trend was noted when these same strains were grown in a minimal medium supplemented with casamino acids (Fig. 4.11AB), although, in this case, the FLAG-26 expressing strain displayed relatively higher and more variable surface levels of NilC. As an alternate method of detecting surface NilC, *X. nematophila* whole cells grown in defined media supplemented with casamino acids and incubated with anti-NilC antibody (Cowles and Goodrich-Blair 2004) and a fluorescent secondary antibody was observed using flow cytometry (Fig. 4.11C). Overall, these data using different growth conditions and immunodetection methods support the conclusion that, in *X. nematophila*, $T11SS_{OMP}$ NilB can either inhibit or promote NilC surface presentation, relative to cells without NilB, which may suggest that $T11SS_{OMP}$ activity is modulated depending on environmental conditions.

Taken together, the data described above are consistent with the model that NilB facilitates surface exposure of NilC, and that, when *nilB* and *nilC* expression is de-repressed by deletion of the transcription factors NilR or Lrp, NilC is surface exposed in *X. nematophila*. Our biophysical data indicate that NilC is a two-domain protein, with a C-terminal barrel domain proposed to be the T11SS targeting motif, and an N-terminal domain proposed to be a host-interaction effector. To further explore the possible effector role of *X. nematophila* NilC, we took a two-pronged approach. In the first approach, we used histochemistry to examine the molecules that are specifically presented on the host nematode intestinal surface, where *X. nematophila* adheres in a manner positively influenced by NilB and NilC (Chaston et al. 2013). In the second, we compared the proteomes and metabolomes of *X. nematophila* strains with and without SR1 to identify metabolic pathways and activities that may be impacted by the presence or absence of NilB and NilC.

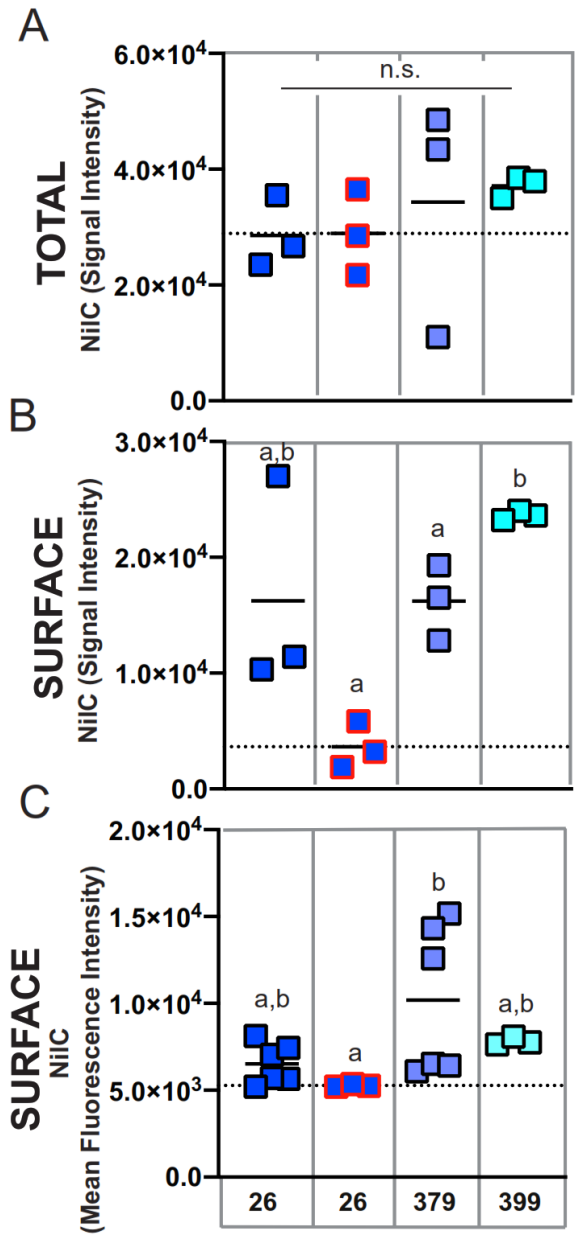


Figure 4.11. **NilB variants influence *X. nematophila* NilC surface exposure.** Cell lysate (**A**) or whole cell (**B, C**) preparations of *X. nematophila* ⊿*nilR* ⊿*SR1* strains grown in minimal medium supplemented with casamino acids. ⊿*nilR* ⊿*SR1* with SR1 with wild type *nilC* (black outline) or a *nilCM1Z* mutation (red outline) and *nilB*-FLAG insertions (blue shaded squares, amino acid site of insertion noted beneath) at the *att*Tn*7* site were spotted onto nitrocellulose membranes (**A, B**) or analyzed by flow cytometry (**C**). NilC protein was detected by immunoblotting with anti-NilC antibodies. Each symbol represents an individual replicate, and lines indicate mean. Significantly different groups within each panel are indicated with different letters (tested using Welch's ANOVA test with Dunnett's T3 post-hoc multiple comparisons analysis). Dashed axis lines indicate the mean of the *nilCM1Z* data points.

The *Steinernema carpocapsae* Anterior Intestinal Cecum Expresses a Wheat-Germ Agglutinin-Reactive Material That Contributes to SR1-Mediated *Xenorhabdus nematophila* Colonization

Xenorhabdus bacteria colonize the anterior intestinal cecum (AIC) of their *Steinernema* nematode hosts during the reproductive juvenile and adult stages (Fig. 4.12AD). The AIC is immediately posterior to the basal bulb, a pumping organ that drives ingestion from the mouth, through the esophagus, and into the intestine (Fig. 4.12A). To begin investigating the surface chemistry of the AIC bacterial colonization site, we selected three nematodes: *S. carpocapsae*, *S. scapterisci*, and *S. feltiae*. The symbionts of the first two nematodes, *X. nematophila* and *X. innexi*, respectively, both encode NilB and NilC homologs (Fig. 4.6). The symbiont of *S. feltiae, X. bovienii*, does not. Adult nematodes that had been raised in the presence or absence of their bacterial symbiont were treated with lectin fluorescent conjugates and observed by fluorescence microscopy for lectin binding to host tissues, monitoring frequency of binding to the mouth, the esophagus, the basal bulb, the AIC, and the intestine (Fig. 4.12B). We found that all lectins tested had some binding to various tissues in all three nematodes tested. Binding varied according to the media type (lipid agar or liver kidney agar) as well as the presence/absence of bacterial symbiont (Chapter4LectinInfo.xlsx and Fig. 4.13).

We focused our attention on fluorescent conjugate wheat germ agglutinin (F-WGA, which reacts with *N*-acetyl glucosamine or *N*-acetyl neuraminic acid), because this lectin had consistent reactivity with *S. carpocapsae* nematode tissues, and because it previously had been observed to react with material extruded from the intestinal receptacle colonization site of the infective juvenile stage *S. carpocapsae* nematode (Martens, Russell, and Goodrich-Blair 2005). As a control, we included fluorescent conjugate *Ulex Europaeus* Agglutinin (F-UEA), which reacts with alpha-linked fucose.

Figure 4.12. **Glycan content differs between nematode species at bacterial colonization sites**. **(A)** *S. carpocapsae* nematode colonized at the anterior intestinal cecum (AIC) by *X. nematophila* expressing green fluorescent protein. Nematode tissue is stained with rhodamine phalloidin. **(B)** False color overlay (a bright field and fluorescence) of uncolonized *S. carpocapsae* grown on liver kidney agar without symbiont, and then incubated with F-WGA (a blue color) for 24 h before imaging. **(C)** At the top is shown a DIC image of *S. carpocapsae* with assayed body parts (mouth, esophagus, basal bulb, AIC, and intestine), roughly defined by dashed lines. *S. carpocapsae, S. scapterisci*, and *S. feltiae* were grown on lipid agar or liver kidney agar (LA or LK) with their respective symbionts (+S) or on liver kidney agar without their symbionts. Nematodes were incubated for 24 h with either F-WGA (blue, top section) or F-UEA (orange, bottom section) before observation for lectin binding to individual body parts. The percentage of all observed nematodes that had lectin binding at that body part is indicated within each cell, and is also represented by the intensity of shading (darker shading = higher percentage). WGA, but not UEA, reproducibly binds to the AIC, with the highest frequencies observed in *S. carpocapsae* and *S. scapterisci* grown on lipid agar with their respective symbionts. **(D,E)** Competition experiments were conducted in which nematodes cultivated on liver kidney agar without their symbionts were exposed to WGA, UEA, or a PBS control and with their green fluorescent protein-expressing *Xenorhabdus* symbiont. After 24 h, individual nematodes were imaged and assessed for colonization at the AIC. **(D)** Shows *X. nematophila* AIC colonization within two *S. carpocapsae* nematodes incubated without lectin (PBS control). **(E)** Probability of AIC colonization by the symbiont of each nematode (*X. nematophila* colonizing *S. carpocapsae, X. innexi* colonizing *S. scapterisci*, and *X. bovienii* colonizing *S. feltiae*) after incubation without (–; white bars) or with (+) WGA (blue bars) or UEA (orange bars). The probability of AIC colonization in *S. carpocapsae* and *S. scapterisci*, but not *S. feltiae*, was significantly affected by treatment with WGA (*S. carpocapsae*: $F_{3,59} = 3.91$, $p = 0.01$; *S. scapterisci*: $F_{1,72} = 5.37$, $p = 0.02$; *S. feltiae*: $F_{1,32} = 1.5$, $p = 0.23$), but not UEA (*S. carpocapsae*: $F_{1, 2684} = 3.70$, $p = 0.05$; *S. scapterisci*: $F_{1,2696} = 0$, $p = 0.9833$).

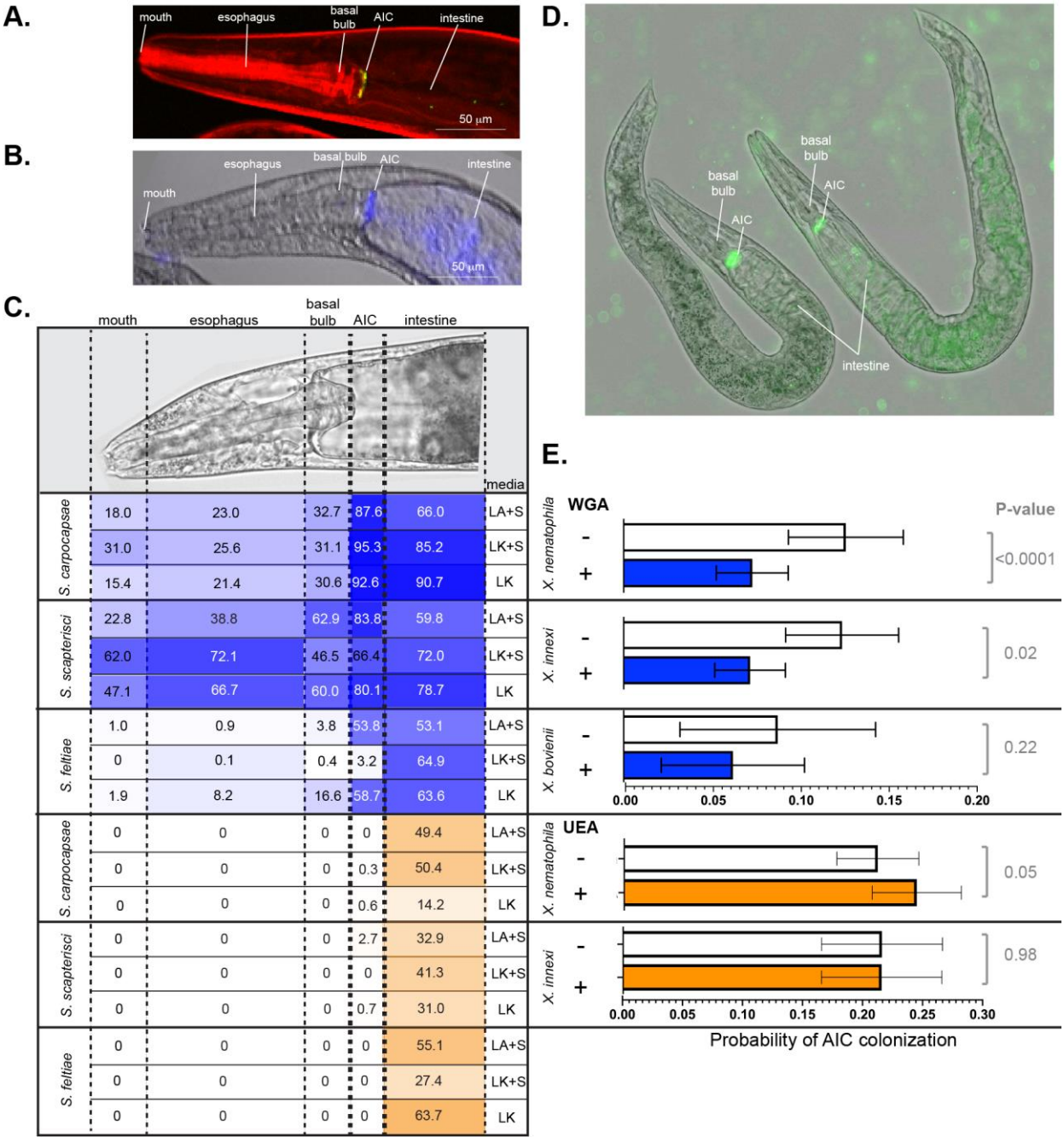


Figure 4.12. (continued)

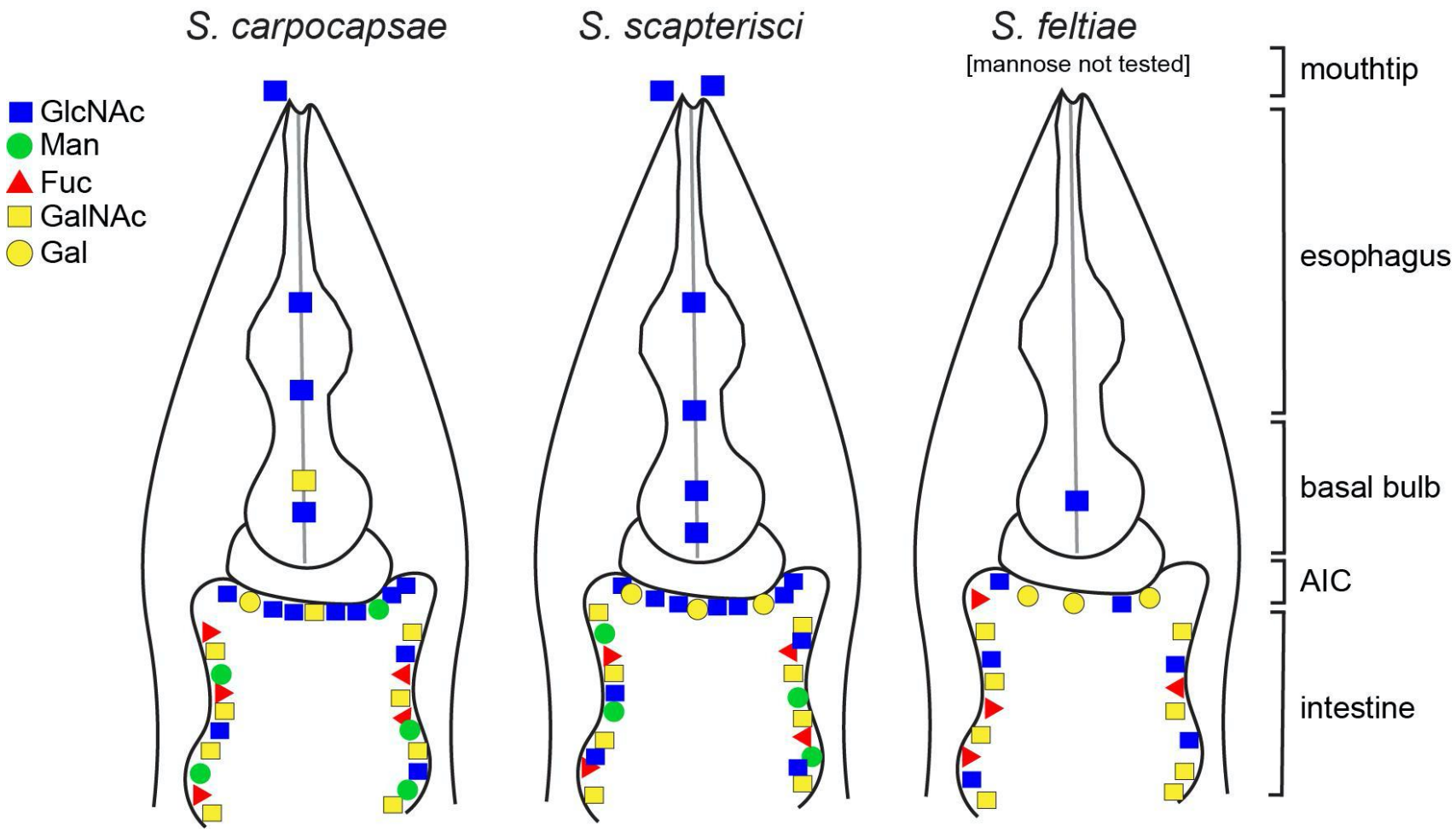


Figure 4.13. **Summary of predicted identity and localization of glycan sugar residues in the gastrointestinal tract of *S. carpocapsae*, *S. scapterisci*, and *S. feltiae*.** The number at each tissue location of symbols representing an individual sugar at each tissue site represents the general frequency at which the lectin corresponding to that sugar was observed as detailed in Chapter4LectinInfo.xlsx.

Observed individuals from all three nematode species from all three cultivation conditions displayed both F-WGA and F-UEA localization to the intestine (Fig. 4.12C). In contrast, F-WGA but not F-UEA localized to the AIC. F-WGA localization to the AIC occurred in *S. carpocapsae* and *S. scapterisci*, but not *S. feltiae* consistently at high frequency (88–95%, 66–84%, and 3–59%, respectively), indicating that the two nematode hosts of SR1-encoding symbionts consistently express F-WGA-reactive material at the symbiont colonization site, while the host of the non-SR1-encoding symbiont does not (Fig. 4.12C).

Given that both F-WGA and the respective *Xenorhabdus* bacterial symbionts are able to bind to the AIC of both *S. carpocapsae* and *S. scapterisci* in the majority of the nematode population, we hypothesized that the WGA-reactive material might be important for colonization of their bacterial symbionts as a binding ligand, a nutrient source, or both. We reasoned that if the WGA-reactive material is involved in colonization, that the addition of soluble WGA to adult nematodes would block bacterial interaction with the WGA-reactive material on the AIC surface. To test this idea, we exposed adult nematodes simultaneously to unconjugated WGA and GFP-expressing bacterial symbionts for 24 h before monitoring by fluorescence microscopy the presence of bacterial colonization at the AIC (Fig. 4.12D). The presence of WGA significantly reduced the presence at the AIC of GFP-expressing *Xenorhabdus* symbionts in both *S. carpocapsae* ($p < 0.0001$; note that composite data including both wild type and Δ*SR1* treatments were included in this analysis) and *S. scapterisci* ($p = 0.02$) but not in *S. feltiae* ($p = 0.23$) when compared to a control that was exposed only to the GFP-expressing bacterial symbionts (Fig. 4.12E). In contrast, the presence of UEA did not decrease the probability for either *X. nematophila* or *X. innexi* to colonize their nematode hosts (*S. feltiae* was not examined) (Fig. 4.12E).

We next considered the possibility that the Nil proteins are responsible for interaction with the AIC WGA-reactive material. If so, then we predict that any Nil-independent adherence to the AIC would not be inhibited by WGA. To test this idea, we analyzed the data to discern the influence of the presence or absence of the SR1 locus on *X. nematophila* colonization of *S. carpocapsae* nematodes exposed or not to WGA. The probability of *X. nematophila* colonization of the *S. carpocapsae* AIC was significantly higher ($p < 0.05$) for SR1 + treatments (0.114 ± 0.022 and 0.146 ± 0.027 with and without WGA, respectively), relative to Δ*SR1* treatments (0.095 ± 0.019 and 0.083 ± 0.017, with and without WGA, respectively). Also, consistent with the role for SR1 in mediating interactions with the WGA-reactive material, the probability of AIC colonization by Δ*SR1 X. nematophila* was not significantly altered by the presence of WGA ($p = 0.36$).

We hypothesized that *X. nematophila* surface-exposed NilC interacts with molecules on the AIC surface. If so, then we predicted that, as with WGA, the addition of free, soluble NilC would inhibit bacterial interaction with the AIC. We exposed nematodes (cultivated in the absence of their symbiont) to purified NilC protein and GFP-expressing *X. nematophila* cells with or without the SR1 locus. In this experiment, the isogenic Δ*SR1 X. nematophila* strains with and without the SR1 locus at the *att*Tn*7* site colonized to similar levels, indicating that the conditions of the experiment were insufficient to distinguish the two strains. However, for both strains, we observed that nematodes exposed to NilC exhibited significantly (*S. carpocapsae* WT $p = 0.0042$, Δ*SR1* $p = 0.0088$) higher rates of colonization compared to nematodes exposed to just the GFP-expressing symbionts (Fig. 4.14). This finding was contrary to our prediction and may indicate that free NilC can facilitate colonization, perhaps by signaling for, or directly catalyzing, release of a nutrient from the host-cell surface. To pursue this hypothesis, we created an *X. nematophila* strain, expressing a non-lipidated version of NilC (generated by the introduction of two lipobox mutations: V17A and C20S). This strain was unable to colonize the infective juvenile stage of *S. carpocapsae* nematodes (Fig. 4.4).

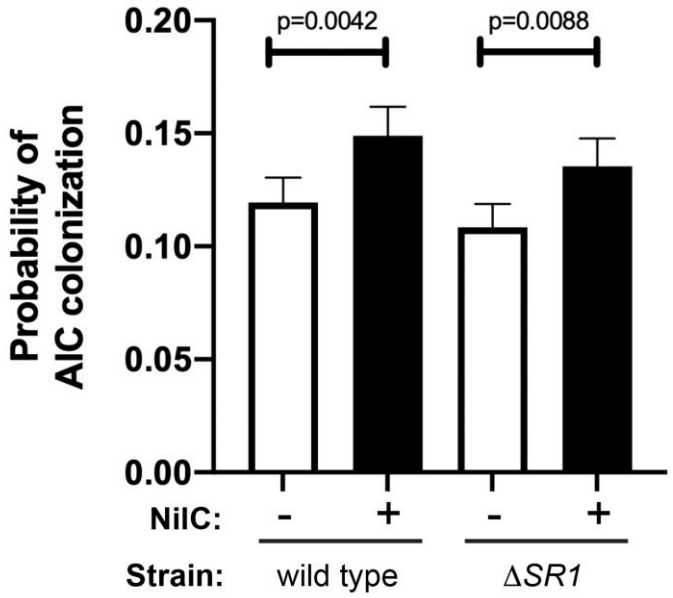


Figure 4.14. **Addition of soluble, purified NilC increases colonization of the anterior intestinal cecum.** Mean with standard error of the probability of AIC colonization is shown. Green-fluorescent protein expressing *X. nematophila* wild type or Δ*SR1* mutant was exposed for ~24 h to *S. carpocapsae* nematodes in a buffer with (+) or without (–) purified, soluble NilC protein. Nematodes were observed by microscopy for the presence or absence of *X. nematophila* at the anterior intestinal cecum (AIC). A total of 8,384 observations were made and analyzed using PROC GLIMMIX of SAS to determine if treatment combination (NilC and strain) differed in probability for colonization. There were differences in probability for colonization ($p = 0.0004$), and there was significantly higher AIC colonization in the presence of NilC for both wild type and Δ*SR1*.

Indeed, unlike wild type NilC, the V17A-C20S NilC protein was not detected on the *X. nematophila* cell surface, and was only detected in the supernatant in 2/6 replicate samples (Fig. 4.15). These data suggest that non-lipidated NilC is predominantly retained within the periplasm.

Deletion of *nilB* and *nilC* Causes Global Metabolome and Proteome Changes Including Impacts on a Peptidoglycan Precursor and Exopolysaccharide Biosynthesis

Based on the data described above, our working model is that NilB is a $T11SS_{OMP}$ that conditionally facilitates surface exposure of the lipoprotein NilC, which is a host-interaction effector. Since external treatment with WGA inhibits *X. nematophila* interaction at the nematode AIC surface, while soluble NilC enhances this interaction, we infer that NilC either interferes with a host defense pathway, as for the T11SS cargo factor H-binding protein, or helps to acquire a host-derived nutrient, similar to the function of T11SS lipoprotein effectors transferrin and lactoferrin-binding proteins. To examine downstream physiological effects of NilC action, we compared the proteomes and metabolomes of *X. nematophila*Δ*nil*RΔ*SR1* emptyTn*7* relative to the Δ*nil*RΔ*SR1* Tn*7*:*SR1* (for this comparison, referred to as Δ*SR1* and WT, respectively) (Table 4). We grew the *X. nematophila* cells in glucose minimal medium (Fig. 4.16) and harvested cells and supernatant at OD_{600} ~0.6 for processing and analysis. Whole cell and supernatant samples were analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS) for proteomic analysis and ultra-high-performance liquid chromatography high-resolution mass spectrometry (UPLC-HRMS) for metabolomic analysis.

In the proteomics comparison of WT and Δ*SR1*, 3,336 proteins were detected in total: 1,742 proteins in whole cell samples and 1,594 proteins in the supernatant samples. Of the 3,336 proteins detected, 61 were considered to be significantly different in abundance between Δ*SR1* and WT based on a Student's *T*-test filter ($p < 0.05$) and a fold change filter ($|FC| > 1$) (See sections 2 and 3

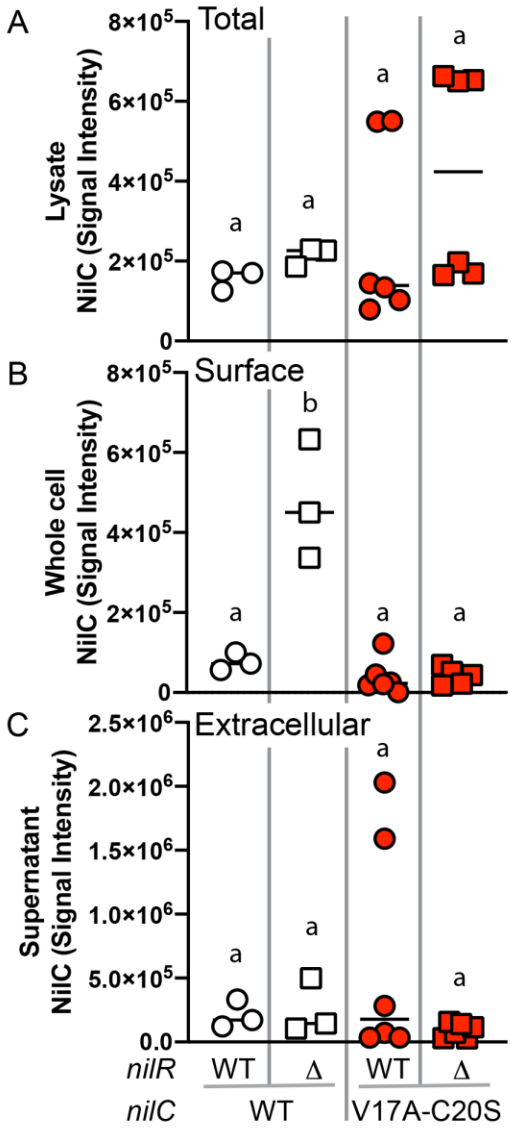


Figure 4.15. **Lipidated NilC is surface exposed and soluble NilC is secreted.** NilC levels in total lysed cells **(A)**, on whole cell surfaces **(B)**, and in cell-free supernatants **(C)** of *X. nematophila* carrying wild type *nilC* (white symbols) or a *nilC* lipobox mutant allele (V17A-C20S; red symbols) and grown in LB medium. The *nilC* alleles were expressed in *nil*R+ (circles) or ⊿*nil*R (squares) genetic backgrounds (see methods). Significantly different groups within each panel are indicated with different letters above the bar graph (tested using one-way ANOVA with Tukey's post-hoc multiple comparisons analysis). As expected, significantly more WT NilC was detected on the cell surface of *X. nematophila* cells lacking *nil*R relative to those with *nil*R. Also as expected, *X. nematophila* expressing the *nilC* lipobox mutant allele did not display NilC on the cell surface, in either the ⊿*nil*R or *nil*R+ strain backgrounds. Instead, in two *nil*R+ background samples, NilCV17A-C20S was detected in the supernatant, suggesting that in these two samples, NilC was either secreted, or cell lysis occurred.

Table 4. **Strains and plasmids used within chapter 4**.

Strain	Genotype	Plasmid	Antibiotic Resistance	Reference/Source
HGB2534	*Escherichia coli* BL21 DE3 C43	pETDuet/MCS2:NilB-FLAG-26/MCS1:NilC_Cterm6xHis	Amp	This Study
HGB2535	*Escherichia coli* BL21 DE3 C43	pETDuet/MCS1:NilC Cterm6xHis	Amp	This Study
HGB2536	*Escherichia coli* BL21 DE3 C43	NA	NA	(Kwon et al., 2015)
NA	*Escherichia coli* BL21 DE3	pET-28a(+)-TEV-*nilC*$_{22-282}$	Kan	This Study; Genscript
HGB1783	*Escherichia coli* S17-1 lpir	pJMC001	Cm	(Bhasin et al., 2012)
HGB0283	*Escherichia coli* S17-1 lpir	pUX-BF13 Tn*7* transposition helper	Amp	(Bao et al., 1991)
HGB2308	*Escherichia coli* BW29427 (DAP-requiring)	pTn*7*/SR1- *nilC57*-V17A-C20S	Kan, Strep, Erm	This Study
HGB2309	*Escherichia coli* BW29427 (DAP-requiring)	pTn*7*/SR1-*nilB26*-FLAG26 *nilC58*-6XHis	Kan, Strep, Erm	This Study
HGB2321	*Escherichia coli* BW29427 (DAP-requiring)	pTn*7*/SR1- *nilC59*-V17A-C20S-6XHis	Kan, Strep, Erm	This Study
HGB2018	*Xenorhabdus nematophila* HGB800 *att*Tn*7*::Tn*7*/GFP	NA	Amp, Cm	Lab strain; (Sugar et al., 2012)
HGB2106	*Xenorhabdus nematophila* HGB800 *kefA*::pJMC001-GFP (from HGB1783)	NA	Amp, Cm	This Study
HGB2330	*Xenorhabdus nematophila* Δ*SR1-7*::*kan att*Tn*7*::Tn*7*/SR1-*nilC57*-V17A-C20S	NA	Amp, Kan, Erm	This Study
HGB2331	*Xenorhabdus nematophila* Δ*SR1-7*::*kan att*Tn*7*::Tn*7*/SR1-*nilB26*-FLAG26 *nilC58*-6XHis	NA	Amp, Kan, Erm	This Study
HGB2332	*Xenorhabdus nematophila* Δ*SR1-7*::*kan att*Tn*7*::Tn*7*/SR1- *nilC59*- V17A-C20S-6XHis	NA	Amp, Kan, Erm	This Study
HGB2368	*Xenorhabdus nematophila* Δ*SR1-7*::*kan att*Tn*7*::Tn*7*/SR1-*nilC57*-V17A-C20S *kefA*::pJMC001-GFP (from HGB1783)	NA	Amp, Kan, Erm, Cm	This Study

Table 4. (continued)

Strain	Genotype	Plasmid	Antibiotic Resistance	Reference/Source
HGB2369	*Xenorhabdus nematophila* Δ*SR1-7::kan att*Tn*7*::Tn*7*/SR1-*nilB26*-FLAG26 *nilC58*-6XHis *kefA*::pJMC001-GFP (from HGB1783)	NA	Amp, Kan, Erm, Cm	This Study
HGB2370	*Xenorhabdus nematophila* Δ*SR1-7::kan att*Tn*7*::Tn*7*/SR1-*nilC59*- V17A-C20S-6XHis *kefA*::pJMC001-GFP (from HGB1783)	NA	Amp, Kan, Erm, Cm	This Study
HGB2371	*Xenorhabdus nematophila* Δ*nilR16*::*Sm* Δ*SR1-7::kan att*Tn*7*::Tn*7*/SR1-*nilC57*-V17A-C20S	NA	Amp, Kan, Erm, Str	This Study
HGB2372	*Xenorhabdus nematophila* Δ*nilR16*::*Sm* Δ*SR1-7::kan att*Tn*7*::Tn*7*/SR1-*nilB26*-FLAG26 *nilC58*-6XHis	NA	Amp, Kan, Erm, Str	This Study
HGB2373	*Xenorhabdus nematophila* Δ*nilR16*::*Sm* Δ*SR1-7::kan att*Tn*7*::Tn*7*/SR1-*nilC59*-V17A-C20S-6XHis	NA	Amp, Kan, Erm, Str	This Study
HGB800	*Xenorhabdus nematophila* WT isolated from *Steinernema carpocapsae* nematodes	NA	Amp	ATCC 19061 (Chaston et al., 2011)
HGB1102	*Xenorhabdus nematophila* HGB800 Δ*nilR16*::*Strep*	NA	Str	(Cowles and Goodrich-Blair, 2006)
HGB1103	*Xenorhabdus nematophila* Δ*nilR16*::*Sm*	NA	Amp, Str	(Cowles and Goodrich-Blair, 2006)
HGB1255	*Xenorhabdus nematophila* Δ*nilR16*::*Sm* Δ*SR1-7::kan att*Tn*7*::Tn*7*-SR1	NA	Amp, Kan, Erm, Str	(Bhasin et al., 2012)
HGB1103	*Xenorhabdus nematophila* Δ*nilR16*::*Sm*	NA	Amp, Str	(Cowles and Goodrich-Blair, 2006)
HGB1966	*X. nematophila lrp-2::kan*	pKV69 (vector)	Cm	(Hussa et al., 2015)
HGB1967	*X. nematophila lrp-2::kan*	pEH54 (low-Lrp plasmid)	Cm	(Hussa et al., 2015)
HGB1968	*X. nematophila lrp-2::kan*	pEH56 (high-Lrp plasmid)	Cm	(Hussa et al., 2015)
HGB1521	*Escherichia coli* S17-1 lpir	pEVS107::SR1/*nilB*-FLAG26; *nilC*(M1Z)	Kan, Erm, Str, Spec	This Study
HGB1200	*Xenorhabdus nematophila* Δ*nilR16*::*Str* Δ*SR1-7::kan att*Tn*7*::Tn*7*-SR1/*nilB26*-FLAG-26	NA	Kan, Erm, Str	(Bhasin et al., 2012)

Table 4. (continued)

Strain	Genotype	Plasmid	Antibiotic Resistance	Reference/Source
HGB1211	*Xenorhabdus nematophila* Δ*nilR16*::*Str* Δ*SR1-7*::*kan att*Tn*7*::Tn*7*-SR1/*nilB*-FLAG26 *nilC19-M1Z* (from HGB1521)	NA	Amp, Kan, Erm, Str	This Study
HGB1808	*Xenorhabdus nematophila* Δ*nilR16*::*Str* Δ*SR1-7*::*kan att*Tn*7*::Tn*7*-SR1/*nilB37*-FLAG-379	NA	Amp, Kan, Erm, Str	(Bhasin et al., 2012)
HGB1207	*Xenorhabdus nematophila* Δ*nilR16*::*Str* Δ*SR1-7*::*kan att*Tn*7*::Tn*7*-SR1/*nilB38*-FLAG-399	NA	Amp, Kan, Erm, Str	(Bhasin et al., 2012)
HGB1681	*Xenorhabdus innexi* WT isolated from *Steinernema scapterisci* nematodes	NA	NA	(Kim et al., 2017)
HGB1699	*Xenorhabdus bovienii*-Sf-FL WT isolated from *Steinernema feltiae* Florida nematodes	NA	NA	(Murfin et al., 2015)
HGB1262	*Escherichia coli* BW29427 (CGSC#: 14194)	pURR25	Strep, Kan	(Teal et al., 2006;Sugar et al., 2012)
HGB2018	*X. nematophila* (HGB800) *att*Tn*7*::Tn*7*-GFP (from HGB1262)	NA	Cm, Amp	This Study
HGB2171	*X. innexi* (HGB 1681) *att*Tn*7*::Tn*7*-GFP (from HGB1262)	NA	Amp, Kan	(Kim et al., 2017)
HGB1865	*X. bovienii* (HGB1699) *att*Tn*7*::Tn*7*-GFP (from HGB1262)	NA	Kan	(Murfin et al., 2018)
HGB1430	*X. nematophila* (HGB007) Δ*SR1 kefA*::pJMC001-GFP (from HGB1783)	NA	Amp, Kan, Cm	(Bhasin et al., 2012)
HGB1431	*X. nematophila* (HGB007) *kefA*::pJMC001-GFP (from HGB1783)	NA	Amp, Kan, Cm	This Study
HGB1508	*X. nematophila* (HGB007) Δ*SR1 kefA*::pJMC001-GFP *att*Tn*7*::eTn*7* (from HGB1783)	NA	Amp, Kan, Cm, Erm	(Chaston et al., 2013)
HGB1509	*X. nematophila* (HGB007) Δ*SR1*, *kefA*::pJMC001-GFP *att*Tn*7*::Tn*7*-SR1 (from HGB1783)	NA	Amp, Kan, Cm, Erm	(Chaston et al., 2013)
HGB1495	Δ*nilR16*::*Str* Δ*SR1-7*::*kan att*Tn*7*::empty Tn*7* *kefA*::pJMC001-GFP (from HGB1783)	NA	Amp, Kan, Strep, Cm	(Bhasin et al., 2012)
HGB1496	Δ*nilR16*::*Str* Δ*SR1-7*::*kan att*Tn*7*::Tn*7*-SR1 *kefA*::pJMC001-GFP (from HGB1783)	NA	Amp, Kan, Strep, Cm	(Bhasin et al., 2012)

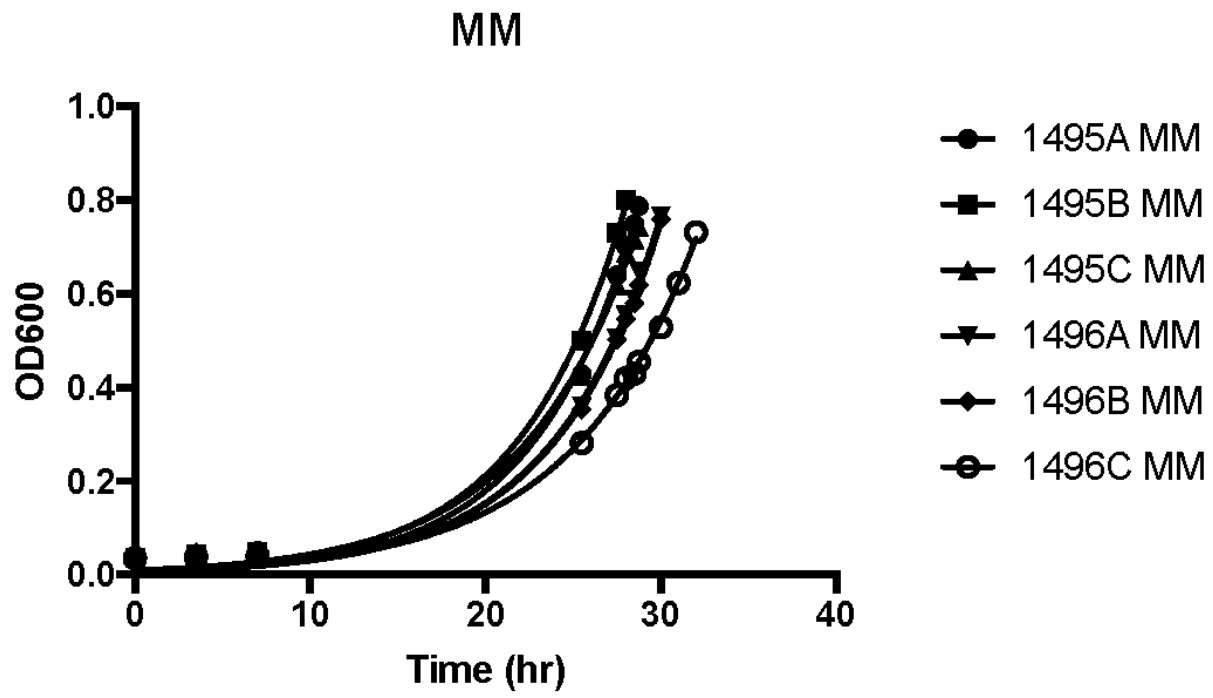


Figure 4.16. **Growth curves of strains used for sampling to conduct metabolome and proteome analyses**. Triplicate cultures (A, B, and C) of each strain (1495: ⊿*lSR1* *att*Tn*7*::eTn*7*; 1496: ⊿*lSR1* attTn*7*::Tn*7*/SR1) were grown in defined medium with glucose (no casamino acids) in 500 ml flasks with 100 ml media for >24 h until OD600 ≅0.6 when they were harvested for proteomics and metabolomics analyses.

in Chapter4SupplementalTables.pdf as well as Fig. 4.17AB). In the metabolomics dataset, 85 metabolites were identified in total, and comparisons were made between the metabolomes using partial least square discriminant analysis (PLS-DA) to determine the separation of the strain metabolic profiles (Fig. 4.18). For both the supernatant and whole cell fractions, there is clear separation of the metabolic profiles between Δ*SR1* and WT. For the whole cell fraction, 18 metabolites were significant ($p < 0.1$), for the supernatant fraction 10 metabolites were significant ($p < 0.1$) (Chapter4MassSpectrometryInfo.xlsx and Fig. 4.18C).

Among the metabolites with elevated abundance in the WT strain were the amino sugars *N*-acetylglucosamine 1/6-phosphate, glucosamine phosphate, and the nucleotide sugar UDP-glucoronate, and, from the PLS-DA analysis (VIP > 1), the nucleotide amino sugar UDP-*N*-acetylglucosamine, the precursor to the exopolysaccharide poly-β - 1,6-*N*-acetylglucosamine (PNAG) (Fig 4.18 and 4.19). UDP-glucoronate is a central intermediate in the synthesis of precursors for exopolysaccharide and cell wall polysaccharide biosynthesis. The proteome revealed that a putative UDP-glucoronate epimerase (XNC1_2486) (Borg et al. 2021) has lower abundance in WT relative to Δ*SR1* (See table 2 in Chapter4SupplementalTables.pdf), consistent with the elevated UDP-glucoronate abundances detected in the former. Amino sugars are involved in peptidoglycan and exopolysaccharide biosynthesis and can be used by bacteria as sources of carbon and nitrogen by catabolism through glycolysis (Dobrogosz 1968). Consistent with this metabolic connection, according to the PLS-DA analysis, relative to Δ*SR1*, whole cell samples of WT had higher abundances of the glycolytic intermediates 3-phosphoglycerate and fructose 1,6-bisphosphate, as well as UDP-glucoronate and UDP-glucose, lipopolysaccharide precursors. The proteomic data also indicated differences in polysaccharide and glycolytic processes.

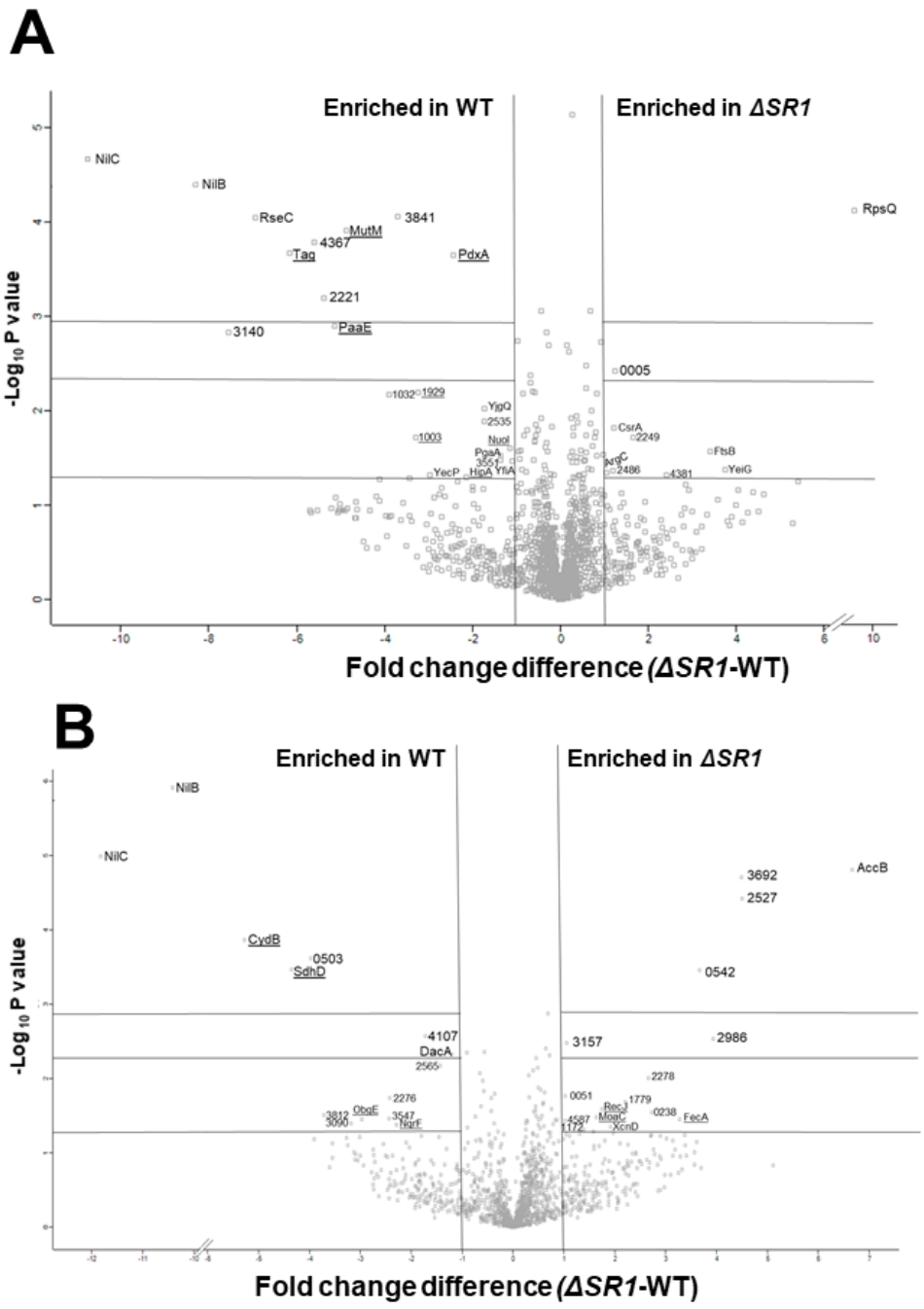


Figure 4.17. **Differentially translated proteins between Δ*SR1* and WT show significant differences in metal-related activities (binding, homeostasis, and metabolic pathways).** Volcano plot representations of proteins that are differentially present in *X. nematophila* wild type and Δ*SR1* strains for **(A)** whole cell fraction and **(B)** supernatant fraction. Proteins indicated on the left of the vertical lines were detected at significantly higher levels (| FC| > 1) in wild type, while those listed to the right were detected at significantly higher levels (| FC| > 1) in the Δ*SR1* mutant. Underlined proteins represent proteins with expected metal-binding activity. Significance values (from bottom to top) are indicated by the horizontal lines at $p < 0.05$, 0.005, and 0.001.

Figure 4.18. **Significant metabolome differences between WT and Δ*SR1* indicate T11SS role of amino acid and amino sugar metabolism**. A) Whole cell metabolomics partial least squares-discriminant analysis (PLS-DA) between WT and Δ*SR1* and the associated variable of importance in projection (VIP) scores for the metabolites that contribute to the separation of metabolic profiles. B) Supernatant metabolomics PLS-DA between WT and Δ*SR1* and the associated VIP scores for the metabolites that contribute to the separation of metabolic profiles. C) Heat map of all metabolites detected in the screen with fold change and Student's t-test p-values listed between strains for both the whole cell and supernatant analyses.

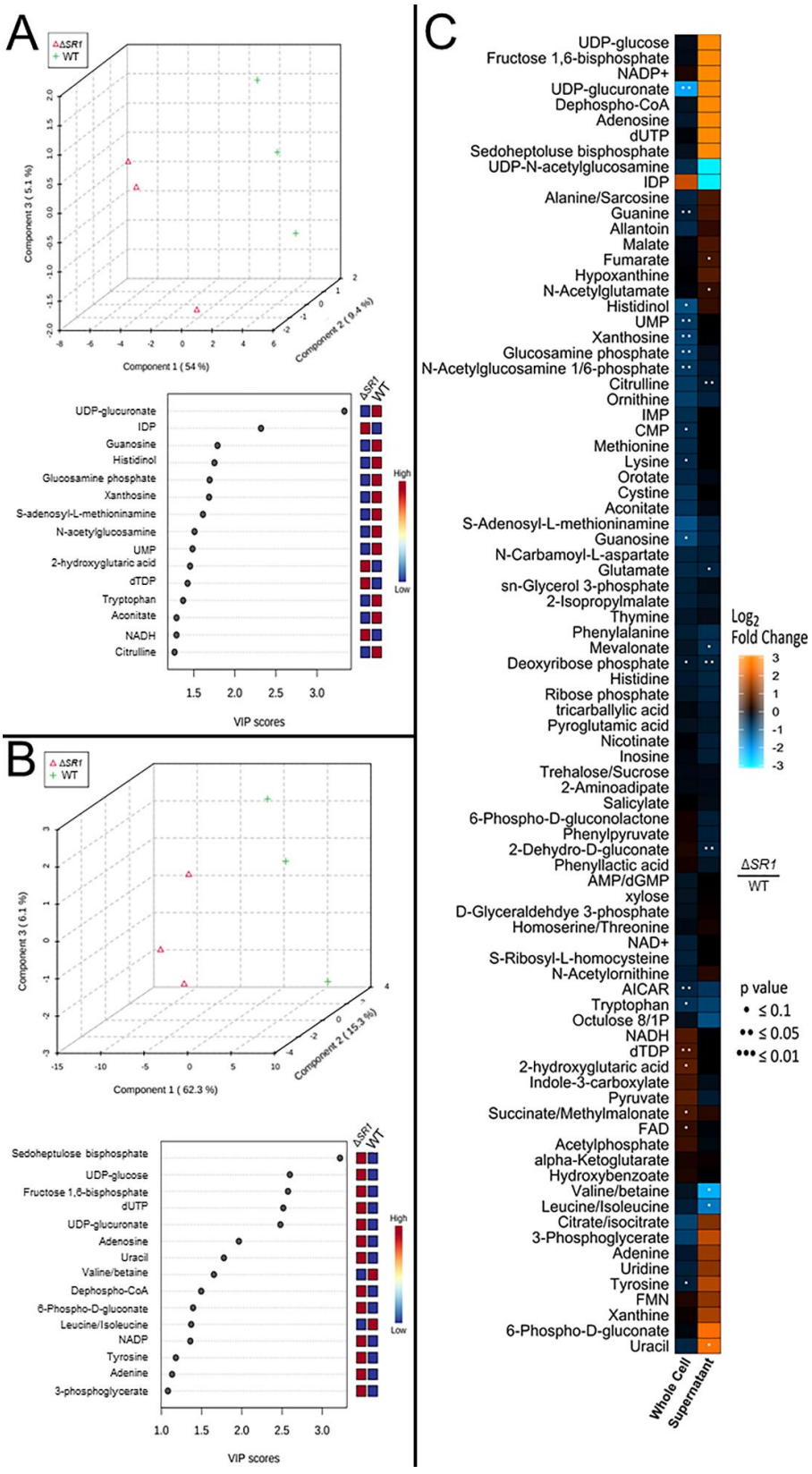


Figure 4.18. (continued)

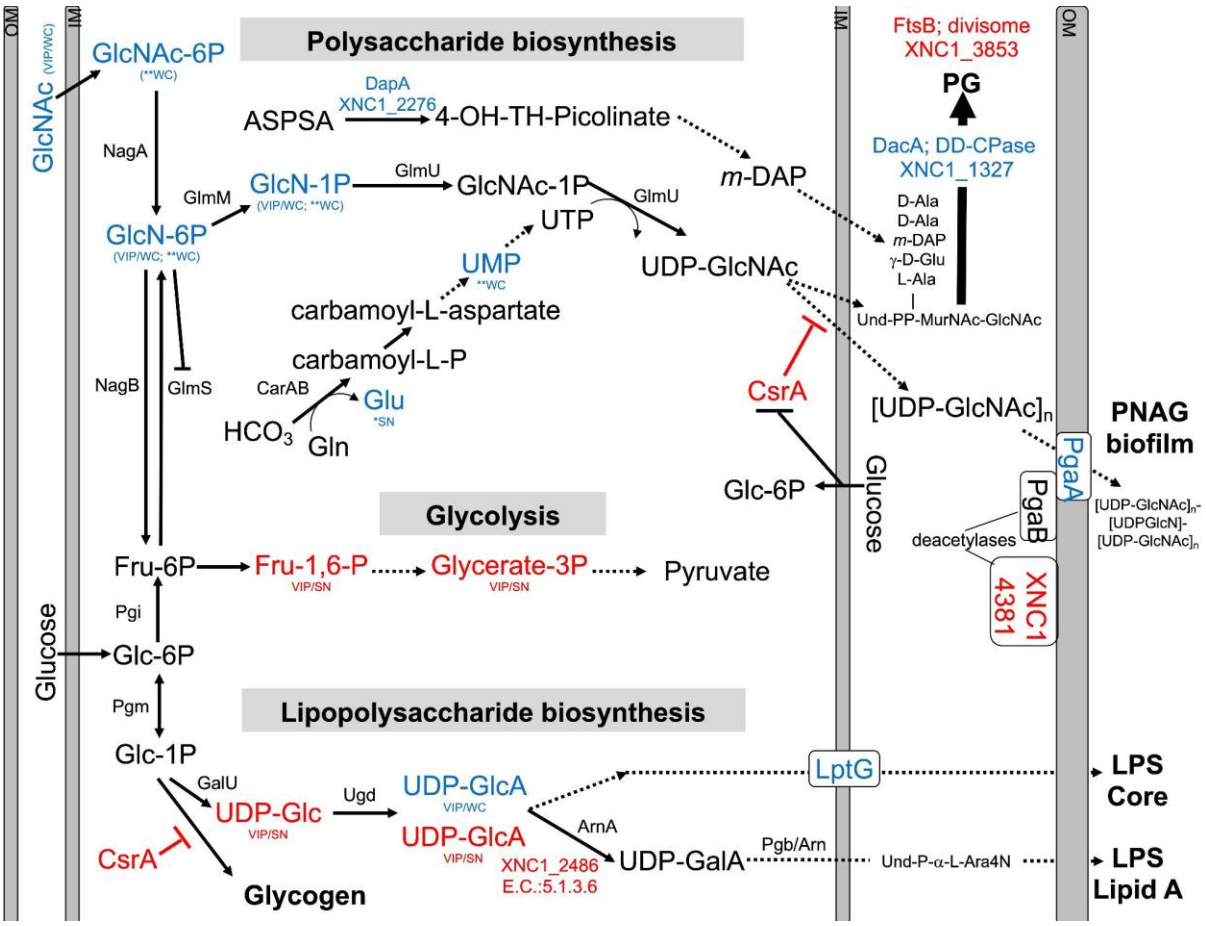


Figure 4.19. **Predicted pathways for amino sugar metabolism leading to peptidoglycan (PG), exopolysaccharide [poly-N-acetylglucosamine (PNAG)], lipopolysaccharide LipidA biosynthesis, and glycolysis pathways**. Blue and red font texts indicate higher and lower abundance of the indicated metabolites and proteins in WT relative to ΔSR1. Asterisks (*) indicate significant differences based on Student's T-test, while VIP indicates VIP > 1 in PLS-DA plots, with WC and SN indicating whether the observed difference was in a whole cell or supernatant, respectively. XNC1_4381 shows similarity to PgaB deacetylase, and XNC1_2486 is predicted to convert UDP-glucuronate (UDP-GlcA) to UDP-galacturonate (UDP-GalA), similar to the enzyme ArnA. GlcNAc-6P, N-Acetyl-Glucosamine; GlcN-6P, Glucosamine-6-phosphate; GlcN-1P, Glucosamine-1-phosphate; GlcNAc-1P, N-Acetyl-Glucosamine-1-phosphate; UDP-GlcNAc, Uridine diphosphate N-acetyl glucosamine; Und, Undecaprenyl; MurNac, N-Acetyl-Muramic Acid; Fru-6P, Fructose-6-phosphate; Fru-1,6-P, Fructose 1,6 bisphosphate; Glycerate-3P, 3-Phosphoglycerate; Glc-6-P, Glucose-6-phosphate; Glc-1P, Glucose-1-phosphate; UDP-Glc, Uridine diphosphate glucose; UDP-GlcA, Uridine diphosphate glucuronate; UDP-GalA, Uridine diphosphate galacturonate; L-Ara4N, L-4-aminoarabinose. Dashed lines indicate multiple steps in the indicated pathway.

The differentially abundant proteome included XNC1_2986, predicted to encode a sugar-phosphate-binding transcriptional regulator of the RpiR family, which was more abundant in the Δ*SR1* mutant relative to WT. In addition, LptG (XNC1_4255), a component of the ABC transporter that exports lipopolysaccharide across the inner membrane (Ruiz et al. 2008), was at higher abundance in wild type relative to Δ*SR1*. DacA, a predicted D-alanyl-D-alanine carboxypeptidase that removes the C-terminal D-alanyl from sugar-peptide cell wall precursors during peptidoglycan biosynthesis, was at higher abundance in the WT relative to Δ*SR1*. FtsB, a cell division protein that regulates peptidoglycan biosynthesis, was lower in WT relative to Δ*SR1* (Boes et al. 2019). WT had higher abundance of a PgaA (a.k.a. HmsH) homolog, the secretin for the exopolysaccharide PNAG (Drace and Darby 2008; Low and Howell 2018). The mutant had elevated abundance of CsrA, which regulates PNAG biofilm formation negatively and positively in *E. coli* and *Yersinia pestis*, respectively (Berndt et al. 2019; Parker et al. 2017; Silva-Rohwer et al. 2021), and XNC1_4381, predicted to encode a lipoprotein with poly-β-1,6-*N*-acetylglucosamine de-*N*-acetylase activity, similar to PgaB/HmsF (Calder et al. 2020). The *pga* locus is necessary for *X. nematophila* (and *Yersinia pestis*) biofilm formation on the external surfaces of *Caenorhabditis elegans* nematodes, but is not necessary for colonization of *S. carpocapsae* infective juvenile nematode receptacles (Couillault and Ewbank 2002; Drace and Darby 2008). Overall, the combined proteomics data indicate that the flux of amino sugar intermediates toward peptidoglycan, exopolysaccharide, and lipopolysaccharide structures is altered in the Δ*SR1* mutant relative to wild type (Fig. 4.19).

Biofilm formation can be regulated by c-di-GMP, levels of which are affected by the pool of purine nucleotides (J. K. Kim et al. 2014). Among the metabolites that were differentially abundant between the WT and Δ*SR1* were those involved in purine metabolism. Xanthosine, AICAR, guanine, and guanosine were detected at significantly higher abundances in the WT background whole cell fraction than in the corresponding Δ*SR1* fraction (Fig. 4.18). Metal binding proteins

involved in regulating DNA repair, transcription, and translation are differentially abundant between strains, coinciding with the significant purine and pyrimidine metabolites found in the metabolomics analysis. Tag (XNC1_4499), MutM (XNC1_0165), and RecJ (XNC1_1136) are proteins involved in the base-excision repair (BER) pathway. RecJ binds Mg, Mn, and Co and has a significantly higher abundance in the Δ*SR1* mutant in the supernatant fraction. The BER pathway repairs DNA damage caused by oxidation, alkylation, and deamination, and their differential abundance may indicate differences in DNA damage occurring in these strains (Krokan and Bjørås 2013).

The metabolomics and proteomics analysis described above suggests that the presence of NilB and NilC influences polysaccharide homeostasis. This, combined with the ability of NilC to increase *X. nematophila* colonization at a glycan-rich tissue and our observations of slight structural similarity between NilC and CBM domains (Fig. 4.6), led us to hypothesize that NilC might have binding and/or cleavage activity for polysaccharides. We tested this in preliminary assays by assessing the ability of purified NilC to bind to chitin-coated beads or to cleave various disaccharide substrates or polysaccharides using fluorogenic substrates with negative results in both cases. We next tested if NilC can bind peptidoglycan (PG), which it would encounter when periplasmically oriented. We chose to use commercially available PG purified from the Gram-positive bacterium *Bacillus subtilis* to avoid potential complications of contaminating lipopolysaccharide from Gram-negative bacterial sources. BSA or NilC was incubated with PG or chitin as a negative control, and the levels of protein that pelleted with these insoluble substrates were quantified (Fig. 4.20). We found that, at the highest concentration of polysaccharide tested (500 mg/ml), NilC associated with PG significantly more than with chitin ($p = 0.0454$). Furthermore, NilC associated with PG significantly more than did BSA ($p < 0.0001$). These data suggest that, when in the periplasm, NilC may bind to peptidoglycan.

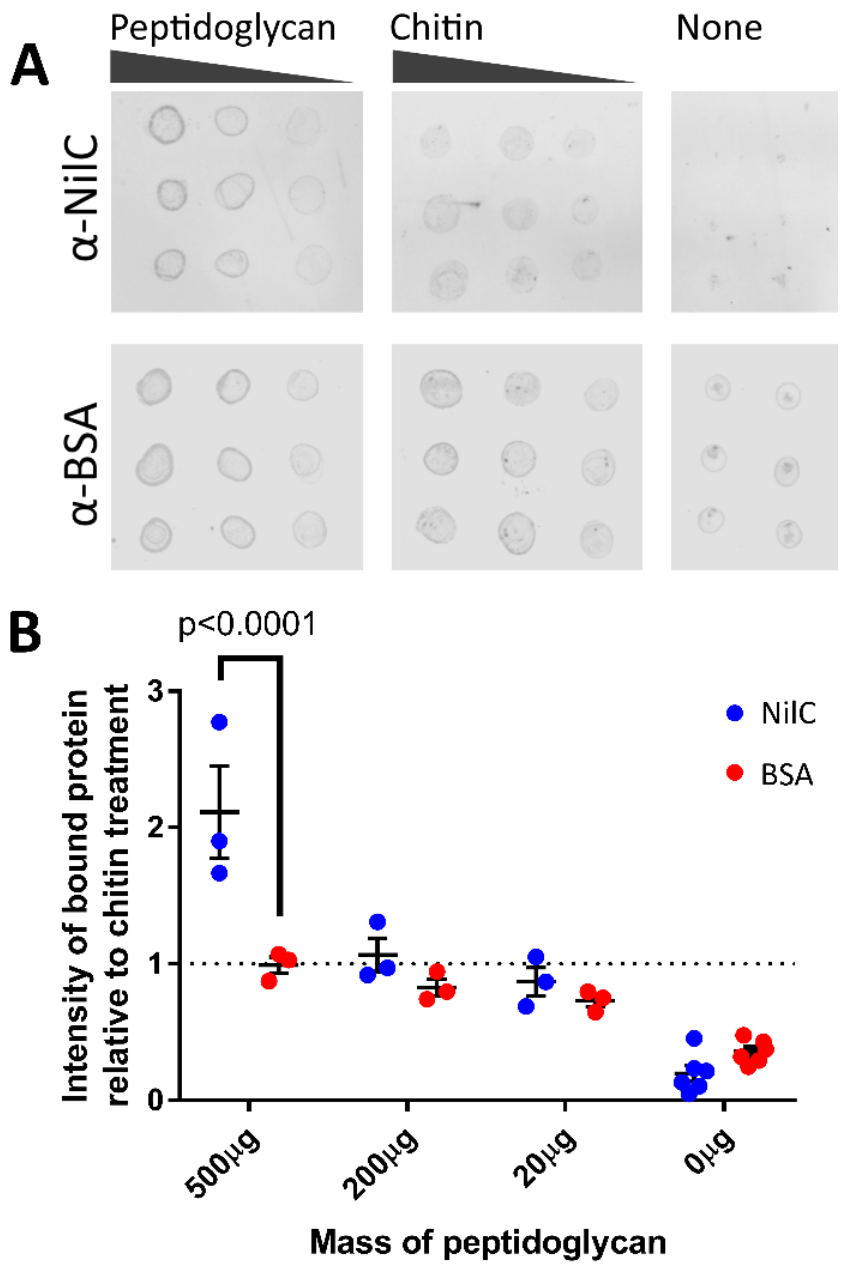


Figure 4.20. **NilC binds peptidoglycan. (A)** Immuno-dot blots depicting the amount of NilC or bovine serum albumin (BSA) bound to *B. subtilis* peptidoglycan or chitin after an hour of incubation. Three different concentrations of peptidoglycan and chitin were used to provide a gradient (horizontal dots), and 2 μl of each reaction were spotted in triplicate (vertical dots). Immunoblots were probed with the indicated primary antibodies and then labeled with an α-rabbit IgG antibody conjugated to a 680-nm fluorophore for imaging. **(B)** To visualize peptidoglycan binding, protein intensity values were normalized to the average intensity of protein in the chitin samples and plotted. Error bars depict the standard error of each mean. A Tukey's honest significance test was used to compare NilC and BSA intensities; a Dunnett's test was used to compare peptidoglycan-associated protein intensities to chitin-associated protein intensities.

Discussion:

Recent discoveries have revealed that members of the DUF560 family of OMPs are type 11 secretion systems (T11SS) that can move proteins across the outer membrane (Bateman et al. 2021; Fantappiè et al. 2017; Grossman, Mauer, et al. 2022; Hooda et al. 2016; Hooda, Lai, and Moraes 2017; da Silva et al. 2019). T11SS-dependent effectors include binding proteins for transferrin, lactoferrin, factor H, and heme. However, the full range, characteristics, and functions of T11SS cargo remain to be explored. Here, we demonstrate that the lipoprotein NilC, a mutualism species-specificity factor, is a cargo protein for the $T11SS_{OMP}$, NilB. Our biophysical data confirm the emerging concept that the T11SS cargo is characterized by an N-terminal functional domain that mediates host interactions and a C-terminal domain that targets cargo proteins for secretion by T11SS. Hooda et al. (2017) noted divergent functional N-terminal domains, and a common C-terminal barrel domain of the known T11SS cargo proteins TbpB, HpuA, and fHbp for which structural data were available at that time (Hooda, Lai, and Moraes 2017). We now add experimental and bioinformatics data that show this structural framework extends to a T11SS cargo protein with a distinctive N-terminal domain and is thus more generalizable than previously realized. We found that the N-terminal 40 amino acids of NilC are disordered, supporting the emerging concept that many lipoproteins destined for the outer membrane have a long, disordered linker at their N-terminus (El Rayes et al. 2021). In *E. coli*, this linker is important for trafficking by the Lol (localization of lipoproteins) system and for appropriate surface localization of the four extracellular lipoproteins tested, including RcsF (El Rayes et al. 2021). Recent structural data have indicated that RcsF associates with BamA before being transferred to a nascent OMP that is folded by the Bam complex. In this model, the RcsF lipid anchor is embedded in the inner leaflet of the outer membrane, with the long, disordered region threading through the OMP to the surface exposed

remainder of the protein (Rodríguez-Alonso et al. 2020). A similar assembly process and resulting topography may also occur for NilC and T11SS (Fig. 4.21).

According to the emerging two-domain model for the T11SS cargo, the NilC N-terminal protein sequence between the disordered linker and the C-terminal β-barrel likely comprises the functional effector domain. The molecular function of this effector presumably facilitates mutualistic colonization of the nematode host intestine, for which NilC is required. Some T11SS cargo proteins (e.g., TbpB) are co-receptors for nutrient uptake systems (Eisenbeis et al. 2008; Konopka 2012; Pogoutse and Moraes 2017), while another (fHbp) protects from host immunity (Schneider et al. 2006). Our data do not yet distinguish between these two possible functions for NilC. Indeed, our data raise a third possibility that the T11SS cargo can have both periplasmic and extracellular functions. Our demonstration that NilC can bind purified bacterial peptidoglycan suggests that periplasmic NilC binds to the cell wall (Fig. 4.21), and is the first experimental indication that NilC may be a carbohydrate binding protein. In this context, we noted an intriguing abundance of tyrosine residues in the N-terminal effector domain. Carbohydrate-aromatic amino acid contacts are a defining characteristic of non-covalent protein carbohydrate-binding pockets due in particular to the favorable CH-π interactions, with tyrosine being the second most frequently carbohydrate-contacting aromatic amino acid (Hudson et al. 2015; Quiocho 1986; Weis and Drickamer 1996). A 62-residue stretch of NilC aligns (albeit with low confidence) with the carbohydrate-binding module (CBM) 4/9 domain d1guia (29% identity over 62 residues) (Fig. 4.6D). While different CBM4 domains recognize diverse glycans using tryptophan or tyrosine residues, in at least one example, the CH-π - binding pocket interactions are donated entirely by tyrosines (Boraston et al. 2002). Conserved tyrosine residues are involved in peptidoglycan peptide recognition in both bacterial lysostaphin (Mitkowski et al. 2019), and eukaryotic peptidoglycan recognition protein-Iα (Guan et al. 2004).

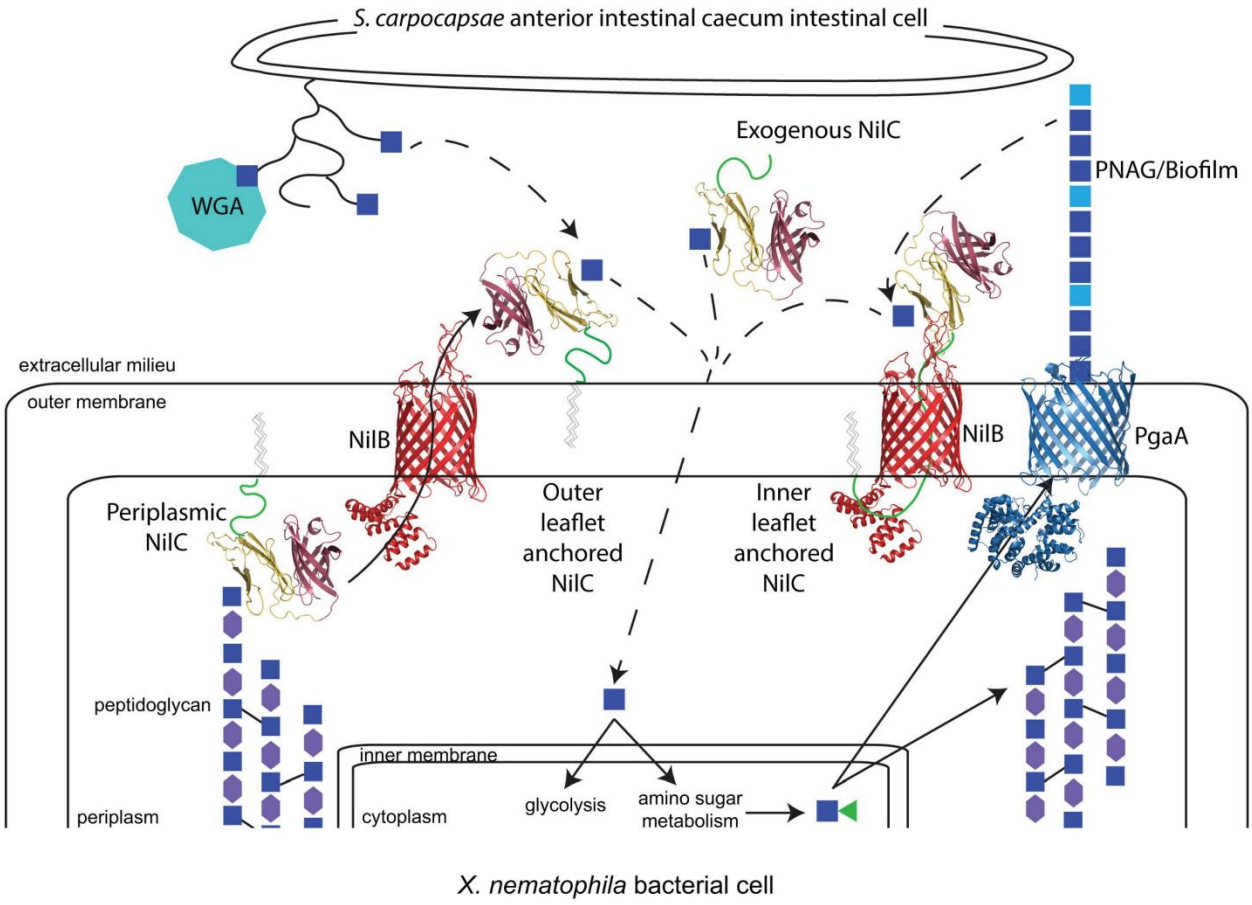


Figure 4.21. **The model of potential molecular interactions occurring during *X. nematophila* colonization of *S. carpocapsae* anterior intestinal cecum epithelial cell surfaces.** *N*-acetyl-glucosamine (GlcNAc; blue squares) is present on the *S. carpocapsae* epithelial cell surface, possibly as part of mucin proteins (curved lines), as demonstrated by the binding of wheat germ agglutinin (WGA; teal octagon) to this tissue. GlcNAc is also a component of bacterial peptidoglycan, alternating in β-1,4 linkage with *N*-acetyl Muramic acid (MurNac; purple hexagons) and cross-linked to other polysaccharide chains through peptide linkages (black solid lines). *X. nematophila* encodes homologs of the Pga synthase-dependent exopolysaccharide secretion system for the production of poly-β-1,6-*N*-acetyl-glucosamine (PNAG), composed of GlcNAc and de-acetylated GlcN (light blue squares). A working model based on evidence presented in this study is that periplasm-oriented NilC binds through its N-terminal effector domain (yellow) to GlcNAc residues within peptidoglycan, possibly modulating turnover. NilC can be extracellular, and its transport across the outer membrane by NilB (red) is directed by virtue of its C-terminal β-barrel T11SS-targeting domain (purple). The lipid anchor of NilC may be embedded either in the inner or outer leaflet of the outer membrane. In the former possibility, the long, disordered N-terminal region of NilC (a green cartoon line) may thread through the barrel of NilB, analogous to the structural interactions between the lipoprotein RcsF and associated OMPs. Extracellular NilC may bind to nematode or bacterial glycan-derived GlcNAc and facilitate its uptake (dashed arrows). NilB (red) and NilC models are predictions from Phyre2.0 and AlphaFold, with the N-terminal random coil of NilC represented as a green cartoon line. The structure of PgaA (blue) is PDB 4Y25.

What role, if any, periplasmic NilC plays in nematode host colonization remains to be investigated. *X. nematophila* expressing a V17A-C20S NilC mutant variant is colonization deficient and was not detected consistently in the supernatant, suggesting that, without its covalently attached lipid, this variant NilC is retained in the periplasm. This may be due to sequestration by binding to peptidoglycan, or due to lower efficiency of secretion through T11SS. Regardless, these data argue against a function for periplasmic NilC in facilitating nematode colonization, but do not preclude the possibility that it plays some other cellular role. Periplasmic NilC may modulate peptidoglycan remodeling or stability, activities that could explain the influence of SR1 on polysaccharide and glycolytic metabolism revealed by this study. Similar effects have been observed for the major outer membrane lipoprotein, Lpp, which is predominantly located in the bacterial periplasm where it binds peptidoglycan, and is situationally surface exposed. (Narita and Tokuda 2011). The predominant function of periplasmic Lpp is in cell envelope structure maintenance. The route by which Lpp is surface exposed and its function, if any, at that location are poorly understood, but are postulated to occur as part of a stress response (Bahadur, Chodisetti, and Reddy 2021; Konovalova and Silhavy 2015; Winkle et al. 2021). Regardless, our findings here suggest that periplasm-surface duality of lipoproteins may be more common than currently appreciated.

The possibility that extracellular NilC serves as the nematode-host interaction effector is supported by our finding that the exogenous addition of soluble, purified NilC can increase *X. nematophila* colonization of the nematode anterior intestinal cecum. If extracellular NilC binds carbohydrates, as suggested above, its substrate may be derived from this surface, which we show here is coated in glycans, including either GlcNAc (a component of peptidoglycan) or *N*-acetylneuraminic acid (Guérardel et al. 2001). Although the glycomes of *Steinernema* nematodes have yet to be elucidated, other nematodes lack *N*-acetylneuraminic (sialic) acid (Bacic, Kahane, and Zuckerman 1990), and the intestinal mucus layer is composed, in part, of O-linked serine/threonine

glycosylated proteins (mucins) with terminal GlcNAc residues. This mucus layer is part of the nematode host-pathogen interface, and is considered part of the immune defense system. In *C. elegans*, mucins are upregulated in response to *Pseudomonas aeruginosa* infection and elicit downregulation of *P. aeruginosa* metal-binding siderophores and biofilm formation (Head, Olaitan, and Aballay 2017; Troemel et al. 2006). In turn, *P. aeruginosa* can counteract these effects and exploit mucins on *C. elegans*' intestinal surfaces to increase virulence (Hoffman, Lalsiamthara, and Aballay 2020; Wheeler et al. 2019). Specifically, *P. aeruginosa* colonization and biofilm formation are increased by the presence of *N*-acetyl-galactosamine monosaccharides derived from the Mul-1 mucin (Hoffman, Lalsiamthara, and Aballay 2020). Similarly, *X. nematophila* NilC may be involved in sensing host-derived monosaccharides to promote colonization and biofilm formation, although, in this case, to facilitate a mutualistic association.

The interface between a host and bacterium is not exclusively composed of host-derived glycans, and extracellular NilC may promote host interactions through binding of a bacterial-derived carbohydrate, such as an exopolysaccharide component of a surface-adherent biofilm. We identified several links between NilB and NilC and biofilm exopolysaccharides. Our analysis of the global transcription factor Lrp, the regulon of which includes *nilC*, revealed an inverse relationship between NilC surface levels and glass surface biofilm formation, for which the exopolysaccharide is as yet unknown (compare Figure 4.12 in this study to Figure 2 of (M. Cao et al. 2017). Furthermore, the abundance of the PgaA secretin of PNAG biofilm exopolysaccharide is reduced in *X. nematophila* cells-lacking *nilB* and *nilC*, suggesting that the Δ*SR1* strain may be defective in PNAG exopolysaccharide biofilm formation. This change in PgaA abundance could be achieved through the RNA-binding protein CsrA (a carbon storage regulator), which is more abundant in Δ*SR1* relative to the SR1+ strain. In *E. coli*, CsrA is a global post-transcriptional regulator that coordinates diverse physiological processes, including iron storage and biofilm formation (Berndt et al. 2019;

Parker et al. 2017; Pourciau et al. 2019; Willias et al. 2015) based on its negative regulation of PgaA translation (Figueroa-Bossi et al. 2014; X. Wang et al. 2005). In *Y. pestis*, biofilm formation is positively regulated by CsrA during growth on alternative carbon sources (Silva-Rohwer et al. 2021). Furthermore, in *Aggregatibacter actinomycetemcomitans*, CsrA-mediated carbon (glycogen) storage and peptidoglycan recycling are both modulated by the presence or absence of PgaA, indicating a complex system of feedback-signaling controlling flux of carbon through energy-deriving, storage, cell wall, and biofilm exopolysaccharide pathways (Shanmugam, El Abbar, and Ramasubbu 2015). Taken together, our data indicate that, as part of their function in nematode colonization, NilB and NilC are integrated with a complex system that balances growth, biofilm formation, carbon uptake and storage, and stress resistance pathways. Future work to identify the *X. nematophila* biofilm exopolysaccharides expressed at the nematode intestinal surface and the role, if any, of NilC in modulating their abundance should shed light on these questions.

A dual function for NilC in both the periplasm and extracellular space raises a further intriguing possibility that the $T11SS_{OMP}$ NilB can control the abundance of NilC that exists in either orientation, perhaps modulated in response to different environmental conditions. Consistent with this idea, we found that, in *X. nematophila*, detectable surface exposure of NilC was observed only for strains in which *nilC* and *nilB* transcription is de-repressed by deletion of either of two transcription factors, NilR and Lrp. While the presence of NilB facilitates surface exposure of NilC during heterologous expression in *E. coli*, we noted a strong and direct correlation between total NilC levels and surface exposure. This phenomenon is similar to that observed for the *N. meningitidis* T11SS cargo protein fHbp. Low-level expression of some fHbp variants is tolerated, but overexpression induces surface exposure (da Silva et al. 2019). In these strains, the presence of the $T11SS_{OMP}$ (Slam 1) is not necessary for, but does enhance the surface localization of fHbp (Fantappiè et al. 2017). da Silva et al. (2019) found that the majority of *N. meningitidis* isolates express variant fHbp that is not

secreted, suggesting that limiting surface exposure of fHbp likely confers a fitness benefit to the bacterium in a clinical setting (da Silva et al. 2019). Similarly, our finding that wild type *X. nematophila* has little detectable surface exposure of NilC indicates that there is selective pressure to limit surface exposure except under certain environmental conditions. Our finding that *X. nematophila*-expressing NilB variants with FLAG-tag insertions at the periplasmic N-terminus and in a surface-exposed loop display lower and higher NilC surface levels, respectively, relative to a cell-lacking detectable NilB (FLAG-379) indicates that this $T11SS_{OMP}$ may regulate NilC surface exposure in response to periplasmic and extracellular signals.

Our overall working model (Fig. 4.21) is that NilC secretion is triggered by the presence of extracellular glycan residues, potentially sensed by the $T11SS_{OMP}$ NilB through one of its surface loop motifs (Bhasin, Chaston, and Goodrich-Blair 2012). In the absence of such a glycan (which presumably indicates a non-nematode-host environment), NilC remains periplasmic and bound to peptidoglycan. In the presence of the activating extracellular glycan, T11SS-mediated secretion of NilC, either lipid-anchored in the outer leaflet, or threaded through $T11SS_{OMP}$ NilB and lipid anchored in the inner membrane, facilitates binding or import of amino sugar molecules derived from the host cell surface. This model provides a framework for future studies, elucidating the identities and species specificity of nematode glycans and mucins, testing the ability of the $T11SS_{OMP}$ NilB to recognize and respond to these host molecules by modulating secretion activity, establishing the structural orientation of NilC within the outer membrane, identifying the ligands of NilC N-terminal effector domains and determining the role, if any, of *Xenorhabdus*-derived PNAG or other exopolysaccharides on colonization of the *Steinernema* nematode anterior intestinal cecum.

Materials and Methods:

Bacterial Strains and Media

Bacterial strains and plasmids used in this study are listed in table 4 and are described in more detail in the following sections. Bacteria were grown in lysogeny broth (LB) culture media (J. H. Miller 1972) or minimal glucose media (Bhasin, Chaston, and Goodrich-Blair 2012), supplemented with 1% LB at either 30°C (for *Xenorhabdus* strains) or 37°C (for *Escherichia coli* strains). For growth of *Xenorhabdus* strains, the LB medium was either stored in the dark (dark LB) or supplemented with 0.1% sodium pyruvate (Jimin Xu and Hurlbert 1990). Strains of *X. nematophila* include those in which the SR1 region has been inserted into the *att*Tn*7* site, downstream of *glmS*. The Tn*7* transposon insertion introduces a new transcriptional terminator that results in 127-bp increase in total *glmS* transcript length (Gay, Tybulewicz, and Walker 1986). Previous studies in the closely related *E. coli* and *P. mirabilis* have not detected any detrimental effects from *att*Tn*7* insertion (Choi and Schweizer 2006; Craig 1996). Antibiotics were used as indicated at the following concentrations: ampicillin (Amp), 150 μg/ml; chloramphenicol (Cm), 15 μg/ml (*X. nematophila*) and 30 μg/ml (*E. coli)*; kanamycin (Kan), 50 μg/ml; erythromycin (Erm), 200 μg/ml; and streptomycin (7.5 μg/ml). All transformations were performed *via* electroporation at 2.5 kV for 6 milliseconds, followed by an hour of outgrowth at 37°C in SOC media.

Generation of pETDuet-1 Expression Vectors

pETDuet-1 (Novagen ®) served as the plasmid backbone for cloning and expression. A 6xHis tag was added to the C-terminus of NilC *via* site-directed mutagenesis (using forward primer 5'-A GTCCATGGTAAATACAAAATCTAAAATCTATTTAGCTC-3' and reverse primer 5'-GCTGCGGCCGCTTAGTGGTGGTG ATGATGATG-3') prior to restriction cloning into MCS1 using the *Nco*I and *Not*I restriction sites to create pETDuet/MCS1:NilC Cterm6xHis. NilB-Flag26

was amplified from HGB1200 (Bhasin, Chaston, and Goodrich-Blair 2012) and cloned into MCS2 *via* Gibson assembly (using two pairs of primers: NilB forward 5'-ATTAGTTAAGTATAAGAAGGAGATATACATATGAAAAA AATCAAATCCATCGTTATAAC-3' and NilB reverse 5'-GCTC TCGAGTTAAAAATTACGCTTGAAGTCCAG-3' plus pETDuet forward 5'-GACTTCAAGCGTAATTTTTAACTCGAGAGCTAATTAACCTAGGCTGCTGCCAC-3' and pETDuet reverse 5'-A TGTATATCTCCTTCTTATACTTAACTAATATACTAAGATG G-3') to create the plasmid pETDuet/MCS2:NilB-FLAG-26/MCS1:NilC_Cterm6xHis. Plasmids were transformed into *E. coli* DE3 C43 (Lucigen ®) *via* electroporation prior to expression. *nilB* and *nilC* sequences were confirmed by Sanger sequencing at the University of Tennessee (UT) Genomics Core. Expression of NilB and NilC was confirmed *via* immunoblotting of induced lysates using α-FLAG and α-NilC antibodies, respectively.

Measuring NilC Surface Exposure in *Escherichia coli* Co-Expression Strains by Immuno-Dot Blotting

Escherichia coli BL21 DE3 C43 strains containing pETDuet-1/MCS2:NilB_26aaflag/MCS1:NilC_Cterm6xHis, pETDuet-1/MCS1:NilC_Cterm6xHis, and no plasmid (HGB 2534, 2535, and 2536, respectively) were raised on defined glucose medium (Orchard and Goodrich-Blair 2004) plates at 37°C for ~16 h. HGB 2534 and 2535 plates were supplemented with ampicillin at 150 μg/ml. Biological replicates (10 each) of HGB 2534 and 2535 were cultured into 5 ml of a defined glucose medium supplemented with 1% LB (Amp 150 μg/ml) and incubated at 37°C for ~16 h. HGB 2536 was cultured in biological triplicate into 5 ml of a defined glucose medium supplemented with 1% LB. Overnight cultures were rinsed 2 x in sterile-defined media, and then half was aliquoted into 5-ml fresh defined glucose media supplemented with 1% LB, and half

was aliquoted into 5-ml fresh dark LB. These cultures were incubated at 37°C for 3 h prior to induction with 0.5-mM IPTG. Cultures were induced at 37°C for 2 h. Cultures were equalized to an OD_{600} of 0.6 and rinsed 3 x in Phosphate Buffered Saline (PBS). Note that, even for those cultures that did not grow, sufficient cells were present to allow spotting of the normalized OD amount. For each whole cell sample, 2 µl was spotted onto a nitrocellulose membrane and allowed to dry. Cells were lysed by sonication (30 s at ~500-rms volts), and 2 µl of each lysate sample was spotted onto a nitrocellulose membrane and allowed to dry. Membranes were immunoblotted and enumerated as described under "Measuring *X. nematophila* NilC surface exposure" below.

Construction of *Xenorhabdus nematophila* Mutants-Expressing Modified NilC

Site-directed mutagenesis of pTn7/SR1 (Cowles and Goodrich-Blair 2004) was used to generate *nilC*57-V17A-C20S wherein the lipobox was removed by adding a signal peptidase I site and by removing the conserved lipobox cysteine residue (using forward primer 5'-GCCTCTTCTAGAGGAGGGGGTTCT-3' and reverse primer 5'-TGTAGCAGAAAGTACCAGTGCGAG-3'). Site-directed mutagenesis of pAB001 (Bhasin, Chaston, Goodrich-Blair) and pTn7/SR1-*nilC*57-V17A-C20S was used to generate *nilC*58-6xHis and *nilC*59-V17A-C20S-6xHis wherein 6 histidine residues were encoded onto the C-terminal end of *nilC* (using forward primer 5'-AGTCC ATGGTAAATACAAAATCTAAAATCTATTTAGCTC- 3' and reverse primer 5'-GCTGCGGCCGCTTAGTGGTGGTGATGAT GATG-3'. The three resulting plasmids were transformed into *E. coli* BW29427, a DAP-dependent, conjugation proficient strain, resulting in HGB2308, HGB2309, and HGB2321. Triparental conjugation was performed using an *X. nematophila* background (either HGB0777 or HGB1251), the appropriate pTn*7* plasmid, and the helper plasmid pUX-BF13 (HGB0283) at a ratio of 3:1:1. Exconjugants were selected for using kanamycin resistance, and insertion into the *att*Tn7 site was confirmed *via* Sanger sequencing (using

forward primer 5'-TGTTGGTTTCACATCC-3' and reverse primer 5-TACTTATGAGCAAGTATTGTC-3'), resulting in HGB2330, HGB2331, HGB2332, HGB2371, HGB2372, and HGB2373. GFP-expressing strains possessing the modified *nilC* alleles were constructed by conjugating *X. nematophila* stains HGB2330, HGB2331, and HGB2332 and *E. coli* S17-1 λ pir-containing pJMC001 (HGB1783) at a ratio of 3:2. Exconjugants were selected for using chloramphenicol resistance and screened for GFP expression using fluorescence microscopy, resulting in HGB2368, HGB2369, and HGB 2370.

Infective Juvenile Colonization Assays

To prepare nematodes for colonization assays, nematode eggs were collected from gravid adult females, developing from infective juvenile nematodes inoculated onto wild type lawns of symbiont and surface sterilized as previously described (Murfin, Chaston, and Goodrich-Blair 2012). They were then applied in biological triplicate to LA plates with lawns of treatment bacteria (HGB800, HGB2330, HGB2331, HGB2332, HGB2106, HGB2368, HGB2369, and HGB2370. Lawns were incubated at 27°C for 9 days and then transferred to white traps to collect emerging infective juveniles. Nematodes were collected and stored in sterile water at 27°C in tissue culture flasks. Frequency of colonization was determined *via* fluorescence microscopy on a Keyence BZX-700. Briefly, nematodes, which had been raised on the GFP-expressing bacterial strains, were observed for intestinal GFP presence at 4 and 20 days post emergence. Both time points gave similar results, suggesting that no mutant suffered from a persistence defect. Direct counts of average CFU/nematode were determined *via* a grinding assay wherein nematodes that had been raised on non-fluorescent bacterial strains were collected 6 days post emergence, equalized by density, homogenized *via* an electric mortar and pestle, serially diluted, and plated on LB plates to

determine CFU. Dilutions of homogenate equivalent to 4 and 0.4 nematodes were used for quantification.

Expression and Purification of the NilC Soluble Domain

NilC lacking its predicted signal peptide (NilC amino acids 22-282) was cloned by GenScript into the vector pET-28a(+)-TEV using restriction enzymes Nde1 and Not1. The resulting ORF includes an N-terminal hexahistidine tag, followed by a Tobacco Etch Virus protease (TEV) cleavage site. Three amino acids from the cloning (GHM) remain prepended to the NilC sequence. $NilC_{22–282}$ was expressed in *E. coli* BL21(DE3) cells. The $NilC_{62–282}$ construct was prepared using pET-28a(+)-TEV-NilC22-282 as a template where the sequence for residues 22–61 was deleted by inverse PCR with the following primers: forward: 5'-CTTAAGGGATATTCCAACG-3', reverse: 5'-CATATGGCCCTGAAAATAAAG-3'. The PCR product was treated with a Dpn1 restriction enzyme for 3 h at 37°C, and then phosphorylated with T4 DNA kinase, followed by ligation with T4 DNA ligase at 16°C overnight. Since this construct is based on pET-28A(+) TEV NilC22-282 construct, it has no signal peptide, so it is expressed in the cytoplasm.

For expression, cells were grown at 37°C in LB media until OD_{600} reached ~0.7, at which point expression was induced by adding IPTG to a final concentration of 1 mm. Protein expression was performed at 25°C overnight. For isotopically labeled protein for NMR experiments, cells were again grown in LB but then transferred to M9 minimal media, supplemented with ^{15}N ammonium chloride (Cambridge Isotope Laboratories) (1g/L of culture). Protein induction and expression were performed as above.

For purification, the cell pellet was suspended in Buffer A (50 mM Tris-HCl pH 8.0, containing 300-mM NaCl). Cell lysis was preceded by French press. Cell lysate was clarified by centrifugation at 25,000 × g for 30 min at 8°C before being loaded into a 5-ml Ni-NTA column

(Qiagen, Germantown, MD, United States). After washing with 100-ml of Buffer A containing 50-mM imidazole, protein was eluted using Buffer A with 500-mM Imidazole. Fractions with target protein were pooled and mixed with TEV protease (final concentration of 1 mg/ml), removing the hexahistidine tag. The sample was dialyzed against Buffer A overnight before removal of the TEV protease using a 1-ml Ni-NTA column. Fractions containing NilC were concentrated before loading into a Sephacryl S-100 HiPrep 16/60 column (Cytiva) for size exclusion chromatography. This step was performed in 50-mM Tris-HCl pH 8.0, containing 150-mM NaCl.

Solution NMR

2D ^{1}H-^{15}N HSQC was performed in a 750-MHz NMR spectrometer equipped with a cryoprobe at the National Magnetic Resonance Facility at Madison. A NilC sample was prepared in 25-mM sodium phosphate buffer pH 6.5, containing 25-mM NaCl, 0.01% sodium azide, 50-µM 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), and 8% D_2O. Experiments were performed at 27°C using a 180-µM NilC concentration. Spectra were processed using Bruker Topspin and referenced to a DSS standard.

Circular Dichroism Spectroscopy

Circular dichroism spectra were collected in an AVIV model 420 CD spectrometer. NilC was prepared at 0.092 mg/ml in a 25-mM sodium phosphate buffer, pH 6.5. Data acquisition was performed at 25°C in a 1-mm path length cuvette using a 5-s averaging time. The CD spectrum of a buffer was measured and subtracted from the NilC spectrum. Secondary structure calculation was performed using the programs Bestsel (Micsonai et al. 2015) and CONTIN/LL (Provencher and Glöckner 1981; van Stokkum et al. 1990).

Limited Proteolysis

Purified NilC at 1 mg/ml (35 mm) was mixed with proteinase K (PK) in a 100:1 molar ratio in 100 ml. The reaction was treated after 1 or 5 min using the following methods: (1) no treatment (2) heating at 95°C for 10 min (3) addition of PMSF to 1-mM final concentration (4) addition of a protease inhibitor cocktail (Roche) to 1 X final concentration. Each sample was then diluted with an SDS-PAGE sample buffer and heated at 95°C for 10 min. All samples were left at 4°C for ~1 week before running SDS-PAGE. The mass of the major breakdown product was determined by an Electrospray Ionization Mass Spectrometry run in the positive mode. To determine which structural element of NilC this corresponded to, we used both Peptide Cutter and Mass Peptide (which consider ion charges) to predict all of the possible Proteinase K digestion peptides and their masses, and a simple spread sheet to sum the masses of all possible contiguous peptides to identify those which could be the observed stable fragment.

HGB1211 (*X. nematophila* with *nilR* and SR1 deleted and the SR1 locus carrying a *nilB*-FLAG26 insertion and a *nilC* (M1Z) start-to-stop codon mutation introduced at the *att*Tn*7* locus) were created as described previously (Bhasin, Chaston, and Goodrich-Blair 2012; Cowles and Goodrich-Blair 2004). Briefly, the plasmid pAB001 (Bhasin, Chaston, and Goodrich-Blair 2012) (pTn*7*-SR1 *nilB*-FLAG26) was used as a template for site-directed mutagenesis with primers NilCMtoZfor 5'-CAAATTGGAATCATTATTAGAATACAAAATCTAAAATC-3' and NilCMtoZrev 5'-TTTAGATTTTGTATTCTAATAATGATT CCAATTTGTTT-3' (Cowles and Goodrich-Blair 2004). The resulting plasmid, pEVS107:SR1/*nilB*-FLAG26; *nilC* (M1Z), was conjugated into *X. nematophila*Δ*nilR16:Str*Δ*SR1-7:kan.*

Measuring *Xenorhabdus nematophila* NilC Surface Exposure by Immuno-Dot Blotting and Flow Cytometry

NilC surface exposure in *X. nematophila* cells was monitored by immuno-dot blotting of whole bacterial cells. For Fig. 4.9AB, these were HGB800, HGB1102, HGB778, and HGB779; for Fig. 4.9CD, these were HGB1966, 1967, and 1968; for Fig. 3EF, these were HGB1200, HGB1211, HGB1207, and HGB1808 (Table 4). For Fig. 4.15, strains tested were HGB800, 1103, 2330, 2371. Test strains were struck on LB pyruvate plates with appropriate antibiotics and incubated at 30°C overnight. For each strain, 5 ml of LB that was kept in the dark to prevent generation of reactive oxygen species (dark LB) was inoculated and grown shaking overnight at 30°C. Strains were then sub-cultured 1:1,000 in dark LB and grown overnight (~16 h) at 30°C with shaking. After growth, cells were spun down from the media and rinsed two times in 1xPBS. Where appropriate, the supernatants were collected and sterilized using a 0.22-mm filter to generate cell-free supernatants. Cell concentrations were normalized to an OD_{600} of 6 and serially diluted 1:3 two times, giving OD_{600} values of 6, 2, and 0.67. Each dilution of whole cells was then spotted in technical triplicate onto a PVDF or a nitrocellulose membrane in 2-µl aliquots and allowed to dry. Lysates were obtained by adding glass beads and shaking vigorously using a FastPrep homogenizer in three cycles of 20 s, shaking, 5 min waiting at 4°C. Lysed cells were centrifuged at 15,871 × *g* for 5 min to remove cellular debris. Lysates and cell-free supernatants were spotted onto nitrocellulose as above allowed to dry. Membranes were blocked with a 50:50 mixture of a Li-cor Intercept Protein-Free Blocking Buffer (P/N: 927-80001) and Tris buffered saline (TBS: 50-mM Tris-Cl, pH 7.6; 150-mM NaCl) for 1 h. Membranes were incubated in a 50:50 Li-cor blocking buffer: TBS supplemented with 0.1% Tween 20 and a 1:1,500 Rabbit Anti-NilC antibody (Pocono labs, received July 23, 2019) for 1 h. Membranes were rinsed 4 times in TBS supplemented with 0.1% Tween 20 prior to incubation with a 1:5,000 Goat anti-rabbit secondary antibody bound to an IRDye 680RD fluorophore.

Emission intensity at 700 nm was quantified using an Odyssey Infrared Imaging System and statistically analyzed *via* one-way ANOVA with Tukey's *post hoc* multiple comparisons analysis.

For flow cytometry, test strains were struck on LB pyruvate plates with appropriate antibiotics and grown 30°C for 1–2 days. Cultures of 5 ml of a minimal medium (MM) (Bhasin, Chaston, and Goodrich-Blair 2012), supplemented with 0.1% casamino acids with no antibiotics, were inoculated with three biological replicates for each test strain and grown 30°C overnight shaking to a stationary phase. Cells were pelleted by centrifugation at 3,220 × *g* for 5 min and suspended in 1 ml of phosphate buffered saline (1X PBS:0.137-M NaCl, 0.0027-M KCl, 0.01-M Na_2HPO_4, 0.0018-M KH_2PO_4). Each biological replicate was divided into two treatment groups, one with permeabilization (0.1% TritonX100 in 1X PBS final concentration), and one without permeabilization (PBS) and incubated at 25°C shaking for 10 min. All remaining centrifugation was performed at 6,010 × *g* for 1 min. Samples were washed in 500-µl 1X PBS, and the primary α-NilC IgG antibody (Cowles and Goodrich-Blair 2004) was added at 1:200 concentration, and incubated at 25°C shaking for 1 h. The samples were washed again in PBS and the secondary goat α-rabbit IgG antibody was added at 1:200 and incubated at 25°C shaking for 1 h. The samples were washed a final time in PBS, and then measured for fluorescence at 488 nm using a BD Biosciences LSR flow cytometer. HGB2018 expressing GFP was used as a 488nm positive control. The gate for positive signal at 488nm was set at 97% of the HGB2018 cells.

Nematode Strains and Preparation

Steinernema carpocapsae ALL strain was obtained from Dr. Harry Kaya (UC-Davis). *S. feltiae* was obtained from the laboratories of Dr. S. Patricia Stock (U Arizona) and was originally isolated in FL, United States. Each of these nematodes is propagated approximately every 3 months through *Galleria mellonella* insects (Vivas and Goodrich-Blair 2001). *S. scapterisci* was obtained from Becker

Underwood Inc. and is maintained by propagation approximately every 3 months through *Acheta domesticus* crickets obtained from local pet stores (I.-H. Kim et al. 2017). Infective juvenile stage nematodes are stored in water in 50-ml tissue culture flasks in between insect infections.

Lectin Binding Assays

To prepare nematodes for lectin binding and colonization assays, nematode eggs were collected from gravid adult females, developing from infective juvenile nematodes inoculated onto wild type lawns of symbiont and surface sterilized as previously described (Murfin, Chaston, and Goodrich-Blair 2012). They were then applied to treatment plates (e.g., bacterial symbiont or liver kidney agar, which enables nematode development without a bacterial symbiont; (Martens, Russell, and Goodrich-Blair 2005) as detailed for each experiment. If nematode eggs were not immediately seeded onto treatment plates, they were stored in a sealed dish in dark LB for up to 4 days. After the eggs had been placed on the plates, they were stored at 27°C in the dark for 9 days. Every 3 days, a fresh plate was selected from each treatment and the nematodes collected off the plates by washing and pipetting with 1x PBS. Collected nematodes were placed in a 1.5-ml centrifuge tube and rinsed three additional times with 1x PBS. After rinsing, as much liquid was removed from the tube and 200 µl of 1x PBS added back to equalize the volume within each tube. For treatment with lectin, a final concentration of 27 mM (~2 µl) of a green fluorescein conjugate lectin was added to the nematode preparation. Tubes were wrapped in foil to inhibit light bleaching and shaken gently overnight on an end-over-end mixer. Nematodes were then rinsed three times with 1X PBS to remove excess lectin before observation *via* fluorescence microscopy on a Keyence BZX-700 microscope. For all data shown in Fig. 4.12 and 4.14, experimental observations of lectin or bacterial localization were blinded, such that the observer was not aware of the treatment. For Fig. 4.12A, *S. carpocapsae* nematodes were stained with 6.6-mM rhodamine phalloidin (Sigma) as described in

Chaston et al. (2013) and visualized on a Zeiss LSM 510 confocal microscope (Zeiss, Thornwood, NY, United States) (Sugar et al. 2012). Images for Fig. 4.12BD were taken on a Nikon Eclipse TE300 epifluorescence-inverted microscope. Brightfield and FITC filter images were overlaid and false colored using Image-J.

NilC Protein and Unconjugated *Ulex Europaeus Agglutinin* and Wheat Germ Agglutinin Lectin Colonization Competition Assays

Xenorhabdus nematophila HGB2018, *X. innexi* HGB2171, and *X. bovienii* HGB1865 bacteria expressing the green fluorescent protein from the *att*Tn*7* locus were used to visualize bacteria at the anterior intestinal cecum colonization site of *S. carpocapsae, S. scapterisci*, and *S. feltiae* FL nematodes, respectively, in the presence or absence of lectin or NilC protein. In addition, to determine the effects of WGA on the colonization of *X. nematophila* with and without SR1, isogenic Δ*SR1* strains with either empty Tn*7* or the Tn*7*-SR1 locus were compared. The green-fluorescent protein was introduced into these strains by integrating the plasmid pJMC001 at the *kefA* site (Bhasin, Chaston, and Goodrich-Blair 2012). Nematodes isolated as described above were exposed to GFP-expressing bacterial symbionts alone or with unconjugated WGA, unconjugated UEA, or purified NilC (see above) at 27-mM final concentration, and processed as described above before observing by fluorescence microscopy for the presence or absence of bacteria at the AIC. As above, these experiments were blinded.

Statistical Analysis

Separate one-way analyses of variance were performed in GraphPad Prism to test if treatments differed with respect to mean fluorescence intensity. Tukey's *post hoc* multiple comparisons analysis was used to compare between all levels of treatment. A simple linear regression was used to test if total NilC expression predicted surface exposure of NilC.

For WGA, UEA, and soluble NilC protein experiments related to individual nematode colonization, the binary outcome of colonization was analyzed within generalized linear mixed models using the GLIMMIX procedure (SAS v 9.4, Cary, NC, United States) (Schabenberger 2005) with a binary distribution and a logit link. The fixed effect of treatment, as well as other confounding variables, such as days observed (6 days total) or the life stage of the nematode (male, female, juvenile), were included in the model. Additionally, random effects to account for biological replicate were included. The *inverse link* option was used within the LS Means statement to obtain model-adjusted probability for colonization by bacterial symbionts. Additional variables, including addition of unconjugated WGA and bacteria, or bacteria alone, the life stage of the nematode, and the day observed that can contribute to colonization status in the assay, were included in models as appropriate. Statistical significance was set at $\alpha = 0.05$.

Proteomic and Metabolomic Sample Preparation

HGB 1495 (*X. nematophila* ATCC19061 Δ*SR1*Δ*nil*R GFP with integrated empty Tn*7*) and HGB1496 (*X. nematophila* ATCC19061 Δ*SR1*Δ*nil*RGFP with integrated Tn*7* containing SR1) were struck from frozen stocks onto LB pyruvate plates and incubated at 30°C overnight. Three biological replicates for each strain were inoculated in 50 ml of a minimal medium (Bhasin, Chaston, and Goodrich-Blair 2012) in a sterile 500-ml flask, and incubated at 30°C, 200 rpm to OD_{600} of 0.6. Based on the findings of Bhasin et al. (2012), growth in this medium to OD_{600} of 0.6 for ~55 h was expected to maximize *nilB* expression in the HGB1496 strain. Cell lysis was monitored at 260–280 Abs. The 50-ml culture was divided for the metabolomics and proteomics analyses. For proteomic analysis, cell pellets were suspended in an SDS lysis buffer (2% in 100 mM of NH4HCO3, 10-mM DTT). Samples were physically disrupted by bead beating (0.15 mm) at 8,000 rpm for 5 min. Crude lysates were boiled for 5 min at 90°C. Cysteines were blocked by adjusting each sample to 30-mM

iodoacetamide and incubating the reaction in the dark for 15 min at room temperature. Proteins were precipitated using a chloroform/methanol/water extraction. Dried protein pellets were suspended in 2% sodium deoxycholate (SDC) (100-mM NH_4HCO_3), and protein amounts were estimated by performing a bicinchoninic acid (BCA) assay. For each sample, an aliquot of ~500 µg of protein was digested *via* two aliquots of sequencing-grade trypsin [Promega, 1:75 (w:w)] at two different sample dilutions, (overnight) and subsequent incubation for 3 h at 37°C. The peptide mixture was adjusted to 0.5% formic acid to precipitate SDC. Hydrated ethyl acetate was added to each sample at a 1:1 [v:v] ratio three times to effectively remove SDC. The samples were then placed in a SpeedVac Concentrator (Thermo Fisher Scientific, Waltham, MA, United States) to remove ethyl acetate and further concentrate the sample. The peptide-enriched flow through was quantified by a BCA assay, desalted on RPC18 stage tips (Pierce Biotechnology, Waltham, MA, United States) and then stored at −80°C. For metabolomic analysis, frozen samples were thawed at 4°C prior to extraction. Extractions were performed using 1.5 ml of 0.1-M formic acid in 4:4:2 acetonitrile:water:methanol according to the procedure described previously (Dearth et al. 2018).

LC-MS/MS and UPLC-HRMS Analysis

For proteomics analyses, all samples were analyzed on a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States), coupled with a Proxeon EASY-nLC 1200 liquid chromatography (LC) pump (Thermo Fisher Scientific, Waltham, MA, United States). Peptides were separated on a 75-µm inner diameter microcapillary column packed with 25 cm of Kinetex C18 resin (1.7 µm, 100 Å, Phenomenex). For each sample, a 2 µg aliquot was loaded in Buffer A (0.1% formic acid, 2% acetonitrile) and eluted with a linear 150-min gradient of 2–20% of Buffer B (0.1% formic acid, 80% acetonitrile), followed by an increase in Buffer B to 30% for 10 min, another increase to a 50% buffer for 10 min and concluding with a 10-min wash at 98% Buffer

A. The flow rate was kept at 200 nl/min. MS data were acquired with the Thermo Xcalibur software version 4.27.19, a topN method where N could be up to 15. Target values for the full scan MS spectra were 1 × 106 charges in the 300–1,500 m/z range with a maximum injection time of 25 ms. Transient times corresponding to a resolution of 70,000 at *m/z* 200 were chosen. A 1.6 *m/z* isolation window and fragmentation of precursor ions were performed by higher-energy C-trap dissociation (HCD) with a normalized collision energy of 30 eV. MS/MS scans were performed at a resolution of 17,500 at m/z 200 with an ion target value of 1 × 106 and a maximum injection time of 50 ms. Dynamic exclusion was set to 45 s to avoid repeated sequencing of peptides.

An established untargeted metabolomics method utilizing ultra-high-performance liquid chromatography, coupled to high-resolution mass spectrometry (UHPLC-HRMS) (Thermo Scientific, San Jose, CA, United States), was used to analyze water-soluble metabolites (Metabolomic Analysis *via* Reversed-Phase Ion-Pairing Liquid Chromatography Coupled to a Stand-alone Orbitrap Mass Spectrometer). Synergi 2.6-µm Hydro RP column 100 Å, 100 mm × 2.1 mm (Phenomenex, Torrance, CA, United States) and an UltiMate 3000 pump (Thermo Fisher Scientific, Waltham, MA, United States) were used to carry out the chromatographic separations prior to full scan mass analysis by an Exactive Plus Orbitrap MS (Thermo Fisher Scientific, Waltham, MA, United States). HPLC grade solvents (Thermo Fisher Scientific, Waltham, MA, United States) were used. Chromatographic peak areas for each detected metabolite were integrated using an open-source software package, Metabolomic Analysis and Visualization Engine (MAVEN). Area under the curve (AUC) was used for further analyses. The raw metabolomics data have been submitted to the MetaboLights data repository under study ID MTBLS3857.

Proteome Database Search Analysis

MS raw data files were searched against the predicted proteins of the *X. nematophila* ATCC 19061 genome (accession FN667742; downloaded 12/20/2017) (Chaston et al. 2013) to which common contaminant proteins had been added. A decoy database, consisting of the reversed sequences of the target database, was appended in order to discern the false-discovery rate (FDR) at the spectral level. For standard database searching, the peptide fragmentation spectra (MS/MS) were analyzed by the Crux pipeline v3.0. The MS/MS was searched using the Tide algorithm and was configured to derive fully tryptic peptides using default settings, except for the following parameters: allowed clip nterm-methionine, precursor mass tolerance of 10 parts per million (ppm), a static modification on cysteines (iodoacetamide; + 57.0214 Da), and dynamic modifications on methionine (oxidation; 15.9949). The results were processed by Percolator to estimate q values. Peptide spectrum matches (PSMs) and peptides were considered to be identified at a q value < 0.01. Across the entire experimental dataset, proteins were required to have at least 2 distinct peptide sequences and 2 minimum spectra per protein. For label-free quantification, MS1-level precursor intensities were derived from MOFF using the following parameters: 10 ppm mass tolerance, a retention time window for extracted ion chromatogram was 3 min, a time window to get the apex for the MS/MS precursor was 30 s. Protein intensity-based values, which were calculated by summing together quantified peptides, normalized by dividing by protein length and then LOESS and median central tendency procedures, were performed on $\log_2$-transformed. Using the freely available software Perseus[1], missing values were replaced by random numbers drawn from a normal distribution (width = 0.3 and downshift = 2.8). This platform was also used to generate the volcano plots.

Peptidoglycan-Binding Assays

NilC was purified as described above in "Expression and purification of the NilC soluble domain." Bovine serum albumin (BSA) was purchased from Fisher Scientific (BP9700100) and used as a negative control for peptidoglycan binding. Peptidoglycan (PG) from *Bacillus subtills* (SMB00288) and chitin (C7170-100G) were purchased from Sigma Aldrich. All reactions were performed in a protein storage buffer (50-mM Tris HCL, 500-mM NaCl pH8). In 50-µl binding reactions, 1 µg of either pure NilC or BSA was combined with 500, 200, 20, or 0 µg of PG or chitin. All reactions were incubated, rotating for 1 h at 37°C. After incubation, reactions were centrifuged at 21,000 G for 3 min to pellet the insoluble fraction and collect the supernatant. This process was repeated two times to ensure complete removal of insoluble components. The remaining pellet was rinsed three times with a protein storage buffer. About 2 µl of each supernatant sample and a pellet sample was dot blotted in technical triplicate onto nitrocellulose membranes. Membranes were heat fixed at 50°C for 10 min to adhere the PG. Membranes were blocked with a 50:50 mixture of Li-cor Intercept Protein-Free Blocking Buffer (P/N: 927-80001) and Tris buffered saline (TBS: 50-mM Tris-Cl, pH 7.6; 150-mM NaCl) for 1 h. NilC membranes were immunoblotted and enumerated as described under "Measuring *X. nematophila* NilC surface exposure" above. BSA membranes were incubated in a 50:50 Li-cor blocking buffer, TBS supplemented with 0.1% Tween 20 and 1:5,000 rabbit anti-BSA IgG [Thermo Fisher Scientific (Waltham, MA, United States), A11133] for 1 h. The membranes were rinsed four times in TBS supplemented with 0.1% Tween 20 prior to incubation with 1:5,000 a Goat anti-rabbit secondary antibody bound to an IRDye 680RD fluorophore. For all membranes, emission intensity at 700 nm was quantified using an Odyssey Infrared Imaging System. A Tukey's honest significance test was used to assess differences between NilC and BSA intensities; a Dunnett's test was used to assess differences between PG binding and chitin binding for each concentration.

Data Availability Statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: MetaboLights data repository under study ID MTBLS3857.

Acknowledgments

This work was supported by grants from the National Science Foundation (IOS- 1353674) to HG-B and KF and the University of Tennessee, Knoxville to HG-B. TM was supported by NIH National Research Service Award T32-GM07215. Funding (or partial funding) for open access to this research was provided by the University of Tennessee's Open Publishing Support Fund. The authors thank Mengyi Cao, John Chaston, Archna Bhasin, and Xiaojun Lu for generating strains used in this study, Joseph Jackson for advice on flow cytometry analyses, and Madeleine Roberts and Brian Weaver for assistance with ligand-binding assays.

Chapter 5: Bioinformatic investigation of effectors and substrates for the novel type 11 secretion system (T11SS)

Authors: Alex S. Grossman, Nicholas C. Mucci, Heidi Goodrich-Blair

Author contributions:

Alex S. Grossman performed the co-occurrence analysis, functional analysis, T11SS cargo prediction, Pls/TepS co-expression analysis, and generated the accompanying figures. Nicholas C. Mucci generated the database of non-specific control sequences and generated figure 5.1 and 5.2. Alex S. Grossman wrote the manuscript and Heidi Goodrich-Blair assisted with proofreading.

Abstract:

The recently identified type 11 secretion system (T11SS) is broadly distributed among proteobacteria, in which it can serve as a symbiosis factor. In the bacterium *Xenorhabdus nematophila* this system a T11SS mediates mutualistic colonization of *Steinernema carpocapsae* nematodes, while in human pathogens such as *Neisseria gonorrhoeae* T11SS are required for pathogenic colonization. Publicly available databases suggest that we have yet to understand the full functional range of T11SS proteins and their cargo. Of the over 3,000 T11SS family proteins currently annotated in the database, only 7 have experimentally demonstrated cargo, highlighting that there is still much unknown about the cellular functions of T11SS and their potential cargo. To begin to identify novel cargo of diverse T11SS, we capitalized on the fact that gene products in shared functional pathways often cluster together within a genome, and used genome neighborhood analyses to identify gene families that consistently co-occur with T11SS ORFs. As part of our analysis, we developed bioinformatic controls to ensure perceived co-occurrence is specific to T11SS, and not an artifact of some other general characteristic common among OMPs. The novel techniques developed in this manuscript enabled removal of ubiquitous, non-specific co-occurring gene families. The controlled co-occurrence analysis supports previously predicted roles for T11SS in protein export, metal uptake, and heme catabolism, while identifying new associations with single carbon metabolism,

nucleotide synthesis, and mobile genetic islands. By comparing significantly co-occurring genes to experimentally confirmed T11SS-dependent cargo we identified 2556 predicted cargo proteins which fall into 141 predicted UniRef50 clusters and approximately 10 different architectures. These architectures include TbpB-like protein, Hemophilin- like proteins, fHbp-like proteins, LbpB-like proteins, and 6 uncharacterized types, all unified by the presence of a C-terminal β-barrel domain. To test the reliability of our predictions we demonstrated secretion of a representative predicted *Escherichia coli* T11SS-dependent cargo, a homolog of the virulence modulating glycoprotein Plasmin sensitive protein (Pls). This finding not only expands the list of experimentally confirmed T11SS-secreted cargo, but also revealed a novel architecture of the Pls domain as an N-terminal domain fused to the C-terminal β-barrel domain typical of T11SS cargo. We next utilized our controlled co-occurrence technique to examine an unexplored cluster of T11SS homologs encoded by marine microorganisms. Unlike the T11SS genes of the main cluster, this marine cluster showed no sign of association with iron uptake pathways. However, this marine cluster did share the association with single carbon metabolism and nucleotide synthesis. Additionally, while the marine cluster did not co-occur with the C-terminal β-barrel domain ubiquitous to known T11SS-dependent cargo, they instead frequently co-occur with DUF1194 containing protein, suggesting that this domain of unknown function may be directly interacting with T11SS. These results show that provided enough genomic data, careful bioinformatic analysis can accurately predict T11SS-dependent cargo proteins, and could reveal novel T11SS targeting domains.

Introduction:

Identifying protein function remains a difficult task in the postgenomic age. As of the Pfam 2021 update v33.1, there are 4244 domains of unknown function (DUF) families, which constitutes around 23% of the database (Mistry et al. 2021). This means roughly one quarter of the Pfam database remains uncharacterized, and these domains make up proteins of unknown function. When

DUFs have their functions defined by molecular biology techniques they can be annotated and taken off the list, improving our datasets permanently. Molecular biology techniques are necessary to properly assign protein function, yet they are time consuming and labor intensive and so cannot typically be performed in a high-throughput manner. *In silico* approaches can be applied to protein domains to focus attention on the likely functions of a particular protein family (Mudgal et al. 2015). Computational research on protein family function allows for efficient and directed hypothesis development so that the proverbial baton can be passed to molecular biologists for experimental confirmation.

The domain of unknown function 560 (DUF560) protein family was recently defined as the type 11 bacterial secretion system (T11SS) and split into distinct clusters using sequence level diversity (Grossman, Mauer, et al. 2022). This outer membrane protein (OMP) family has homologs in a diverse array of Proteobacteria, including human pathogens in the *Neisseria*, *Haemophilus*, and *Moraxella* genera (Hooda et al. 2016; Hooda, Lai, and Moraes 2017). In the gamma-proteobacterium *Xenorhabdus nematophila*, a T11SS system encoded within the nematode intestinal localization locus (*nil*) is necessary for colonization of the mutualistic nematode host *Steinernema carpocapsae* (Cowles and Goodrich-Blair 2008). In this system the T11SS, NilB, regulates the surface exposure of a surface lipoprotein, NilC, both of which are necessary for colonization. Other known lipidated T11SS-dependent effectors include co-receptors for metal-containing compounds heme, transferrin, and lactoferrin (Bateman et al. 2021; Hooda et al. 2016; Hooda, Lai, and Moraes 2017; Latham et al. 2020). These proteins are composed of two types of domains, N-terminal β-sheet ligand binding domains and 8-stranded C-terminal β-barrel domains (such as TbpBBD and Lipoprotein C). T11SS is capable of transporting unlipidated, soluble cargo proteins across the outer membrane. For example, T11SS-dependent hemophilin homologs have been characterized from *X. nematophila* (HrpC), *Acinetobacter baumannii* (HphA), and *Haemophilus haemolyticus* (Hpl). However, the majority of

T11SS OMPs do not have a known or predicted cognate cargo. A sequence similarity network analysis suggested that previously characterized T11SS only capture a small portion of the functional diversity present in this family (Grossman, Mauer, et al. 2022).

The availability and accessibility of sequencing data have skyrocketed since the first bacterial genome was sequenced in 1995, with over 200,000 bacterial and archaeal genomes currently deposited on public databases (Z. Zhang et al. 2020). This tremendous repository of genomic sequence data allows computational biologists to probe these genomes for patterns in genetic co-occurrence, as measured through proximity to a gene of interest. Genes within a common functional pathway often cluster together within a genome. Such patterns can be useful when forming hypotheses about the function of uncharacterized proteins. Gene neighborhood studies analyze the genomic regions upstream and downstream of any given gene of interest with the assumption that nearby genetic partners could illuminate the role(s) of the unknown gene (Rogozin et al. 2002).

However, one flaw in the current methodology used to derive co-occurrence lists is the lack of bioinformatic controls capable of removing non-specific co-occurring genes. For example, commonly inherited, deeply ancestral domains, such as transcription and translation machinery, often appear as the top result of a co-occurrence analysis, but do not yield useful information about the specific query gene since their appearance is a coincidence of ubiquity. Also, meaningful results can be obscured by co-occurrence based on structural or other common features present within the query protein, such as those that target the gene product to specific cellular locations, which are expected to interact commonly with the relevant protein trafficking machinery. Herein, we use computational approaches to elucidate potential functional roles played by T11SS-dependent cargo and predict novel cargo. Using these results, we then generated a co-expression system in *Escherichia coli* to experimentally demonstrate the secretion of one novel, bioinformatically predicted cargo

family using the plasmin sensitive protein (Pls) from *Haemophilus parahaemolyticus* and the type eleven Pls Secretor (TepS).

Results:

Establishing a control for domain of unknown function (DUF) gene neighborhood co-occurrence analysis

Characterized T11SS-cargo pairs encoded by Proteobacteria have conserved roles in host-metal acquisition and immune evasion (Bateman et al. 2021; Grossman, Mauer, et al. 2022; Hooda et al. 2016). Only a small portion of predicted T11SS-cargo pairs have been experimentally characterized, leaving open the possibility that as-yet uncharacterized systems may be participating in diverse symbiotic and metabolic pathways. To gain insights into the potential breadth of biological processes in which T11SS-cargo pairs may function, we collected all T11SS sequences identified as part of the largest T11SS sequence similarity network cluster described in Grossman *et al.* 2021 (cluster 1, containing 1111 genes) and utilized the Rapid ORF Description & Evaluation Online (RODEO) software to identify all protein domains encoded within six open reading frames upstream or downstream of the query genes, resulting in a data set of 1504 domains, co-occurring between 2 and 928 times (see tab 1 of Chapter5CooccurenceDatasets.xlsx) (Tietz et al. 2017). To avoid spurious correlations, we chose to exclude any domains that co-occurred fewer than 25 times (~2.7% of the maximum frequency co-occurrence). This number was arrived at by generating a histogram of co-occurrence frequency and identifying the inflection point between the rare co-occurring sequences which made up the bulk of the dataset (1331 or 88.5%) and those domains whose co-occurrence frequency was greater than expected randomly (173 or 11.5%).

To ensure that the observed co-occurrences were specific to the T11SS protein family and not due to common features of genomic loci encoding outer membrane proteins, we developed and

implemented a control process. Briefly, RODEO co-occurrence neighborhoods were generated using randomly selected proteins with biophysical characteristics similar to our DUF560 query proteins such as protein length and outer membrane localization (GO term: outer membrane GO0019867), and then used as a point of comparison for gene neighborhoods from T11SS genes. The control sequences obtained were limited to the Proteobacteria phylum, where T11SS predominantly occurs (Grossman, Mauer, et al. 2022). Due to a left-skewed distribution, the median length of all DUF560 domain containing proteins in Pfam was used to reduce bias from pseudogenes and gene fragments (Fig. 5.1) (Mistry et al. 2021). The median size of all predicted T11SS proteins was 481 amino acids, so we extracted all outer membrane ORFs between 480 and 482 amino acids in size. Sequences that appeared in the T11SS database were removed from the control database in order to ensure unbiased selection. A random subsample of this control database was taken equal to the number of sequences in the original query database (see tab 11 of Chapter5CooccurenceDatasets.xlsx) (Fig. 5.2).

OMP sequences from the subsampled control dataset were then submitted as a query to the genome neighborhood network function of RODEO using the same parameters as the T11SS query list. The output of this control co-occurrence was used to generate a false discovery rate (FDR) of any given domain by dividing frequency of co-occurrence in with random OMPs by frequency of co-occurrences with T11SS genes. We then gated the dataset for domains with an FDR less than 0.1 (10%) to exclude non-specific correlations, resulting in a final set of 51 significantly co-occurring domains. (see tab 2 of Chapter5CooccurenceDatasets.xlsx). The most common significant co-occurring domains were other DUF560 proteins (928/1111), the TbpBBD β barrel domain (503/1111), the TonB C-terminal domains (364/1111), heme oxygenase (137/1111), and DUF454 (118/1111). Next, all genes that co-occurred with a T11SS gene and possessed one of the identified domains were extracted resulting in 3405 significantly co-occurring genes (see tab 3 of

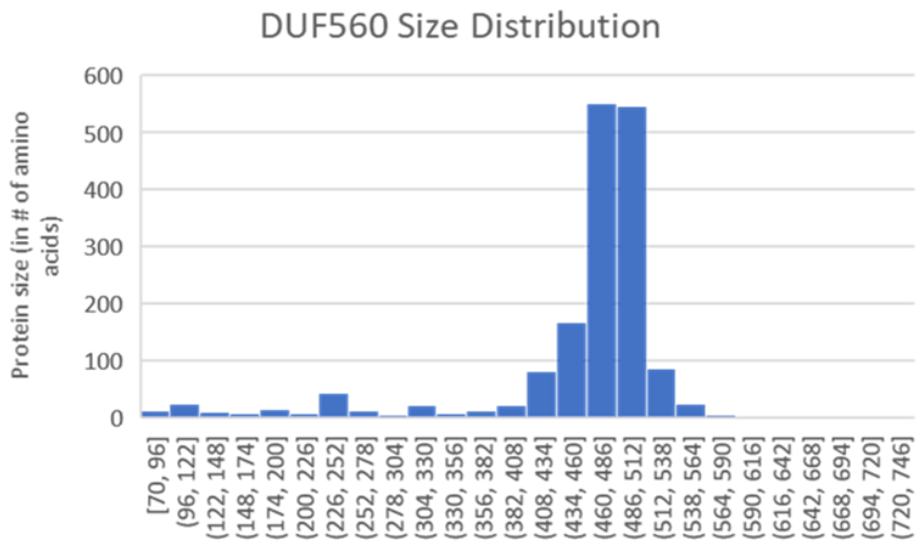


Figure 5.1. **T11SS (DUF560) protein size distribution (in amino acids).** After examination of the size distribution, median was chosen as an appropriate measure of representative protein size over mean due to left-skewed distribution.

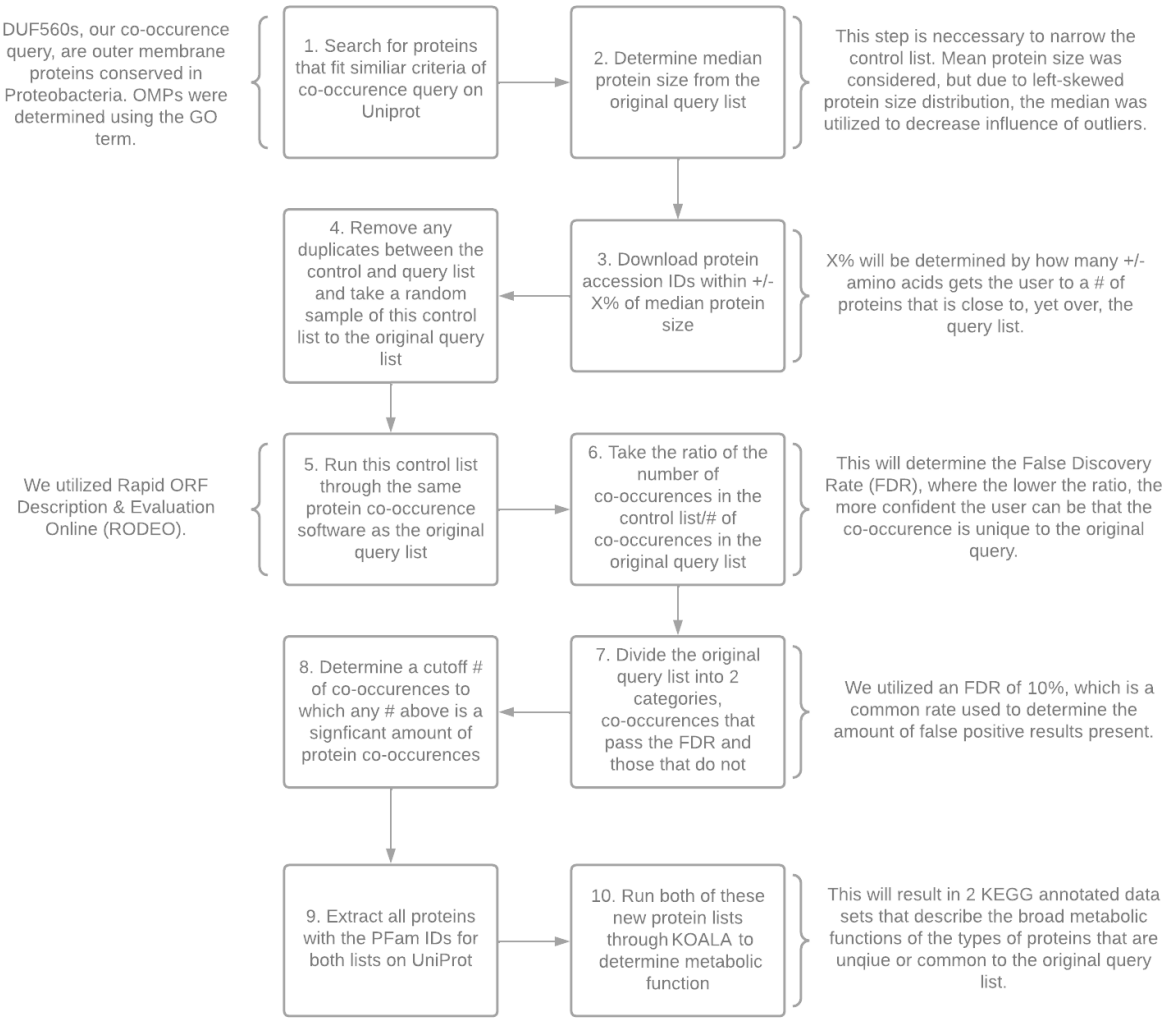


Figure 5.2. **Design for a novel control technique for genomic neighborhood co-occurrence analysis.** Schematic diagram outlining the process we used to remove nonspecific co-occurrences from a genomic neighborhood analysis, with commentary highlighting how we used this outline to refine our T11SS co-occurrence database. Created in Lucidchart.

Chapter5CooccurenceDatasets.xlsx). These genes were submitted to KofamKOALA and BlastKOALA in parallel for annotation into molecular pathways and KEGG specific functional categories called "brite hierarchies" (Aramaki et al. 2020; Kanehisa, Sato, and Morishima 2016).

Functional assessment of T11SS co-occurring proteins identifies a role of iron uptake, protein export, and single carbon metabolism via folate

We examined the co-occurrence gene set for those with annotations in the curated Kyoto Encyclopedia of Genes and Genomes (KEGG) ortholog (KO) database (Aramaki et al. 2020). A hidden Markov model search using KofamKOALA (HMMER/HMMSEARCH) was used to annotate the genes according to the KO database. The software found matches for 1488 of the 3405 significant co-occurring genes (see tab 1 of Chapter5FunctionalAnalysis.xlsx). When matched to cellular functions (brite hierarchies) the most common categories were transporters (periplasmic TonB, ExbD, Signal peptidase II, TolA, etc.), enzymes (heme oxygenase, formate dehydrogenase, exopolyphosphatase, dihydrofolate reductase, etc.), and tRNA biogenesis (tRNA modifying GTPase, tryptophanyl-tRNA synthase, Ribonuclease P, aminoacyl tRNA synthase) (Fig. 5.3A). Additionally, T11SS genes were found to significantly co-occur with transposase genes. Of the 1488 matched sequences 259 had known pathway association, the most common pathways being biosynthesis of secondary metabolites (heme oxygenase, cobalamin-dependent methionine synthase, vitamin K biosynthesis protein MenH, etc), porphyrin metabolism (heme oxygenase), and protein export (signal peptidase II and YidC intertase) (Fig. 5.3B) (see tab 2 of Chapter5FunctionalAnalysis.xlsx). YidC/Oxa1 insertases suggest that a chaperone may be required for T11SS membrane insertion (Kumazaki et al. 2014). The list of pathways also included ribosome components (RpmB, RpmG, RpmH) and aminoacyl tRNA biosynthesis (TyrS, TrpS). A high frequency of co-occurrence with transposases, ribosome components, and tRNA synthesis genes is characteristic of pathogenicity

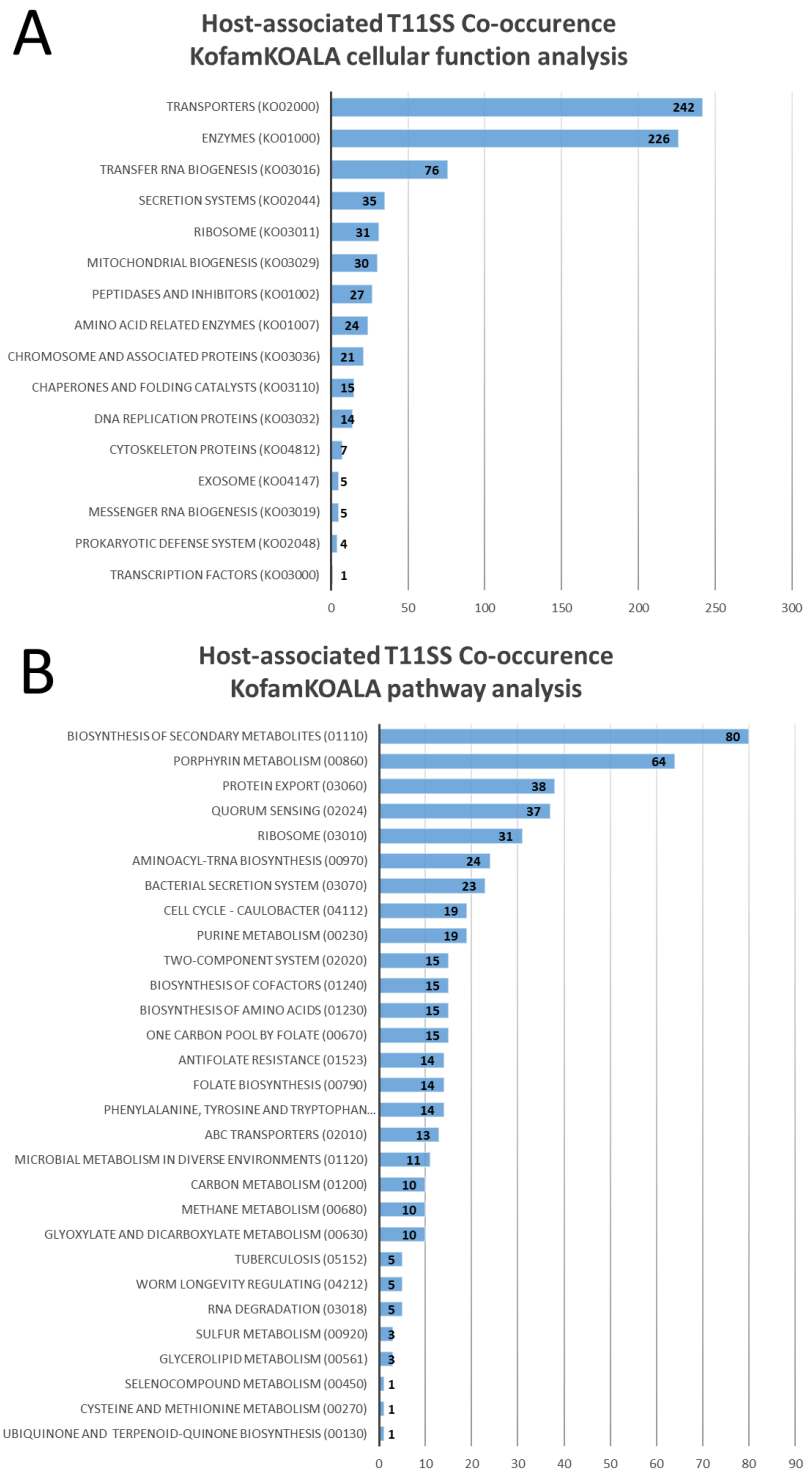


Figure 5.3. **Significantly co-occurring genes with animal associated T11SS reveal conserved association with iron/heme uptake, protein export, and single carbon metabolism**. KofamKOALA uses hidden Markov models to assign functions to query sequences and reveal shared pathways. Cellular functions were estimated using brite hierarchies and assigned to known pathways.

islands, phage regions, and other mobile genetic elements, since many mobile genetic elements use universally conserved ribosome and tRNA genes as anchors for genomic insertion (Dobrindt and Reidl 2000). The frequency at which T11SS genes are encoded within mobile genetic element islands suggests that T11SS may frequently be horizontally acquired. Finally, the list included several carbohydrate biosynthesis pathways (one carbon pool by folate, purine metabolism).

To support our observed co-occurrences, we used BLAST alignment, rather than hidden Markov searching, to establish sequence similarity using BlastKOALA, which compares query sequences to the non-redundant KEGG dataset (Kanehisa, Sato, and Morishima 2016). The software found matches for 1387 of the 3405 significantly co-occurring genes. Categorization into cellular functions (brite hierarchies) again revealed the largest categories to be enzymes, transporters, and tRNA biogenesis (Fig. 5.4A) (see tab 3 of Chapter5FunctionalAnalysis.xlsx). BlastKOALA identified more co-occurring amino acid importers (arginine, lysine, tyrosine, tryptophan, and serine transporters) in the dataset than KofamKOALA. Of the 1387 identified genes 260 had known pathways. This analysis identified similar pathways to those identified by KofamKOALA, including secondary metabolite biosynthesis, protein export, porphyrin metabolism, and carbohydrate biosynthesis (Fig. 5.4B) (see tab 4 of Chapter5FunctionalAnalysis.xlsx).

Bioinformatic investigation of additional T11SS cargo and their structural prediction

To bioinformatically identify potential T11SS-dependent cargo within our dataset of co-occurring genes we collected the sequences of co-occurring proteins that contain one or more of the significantly co-occurring domains and searched them for homology to known cargo. To consolidate potential homologs in the identified sequences we used UniProt ID mapper to find the UniRef50 clusters each of our significantly co-occurring genes belonged to. UniRef50 is a database of genes in which all genes with 50% or greater identity at an amino acid level are clustered into a single

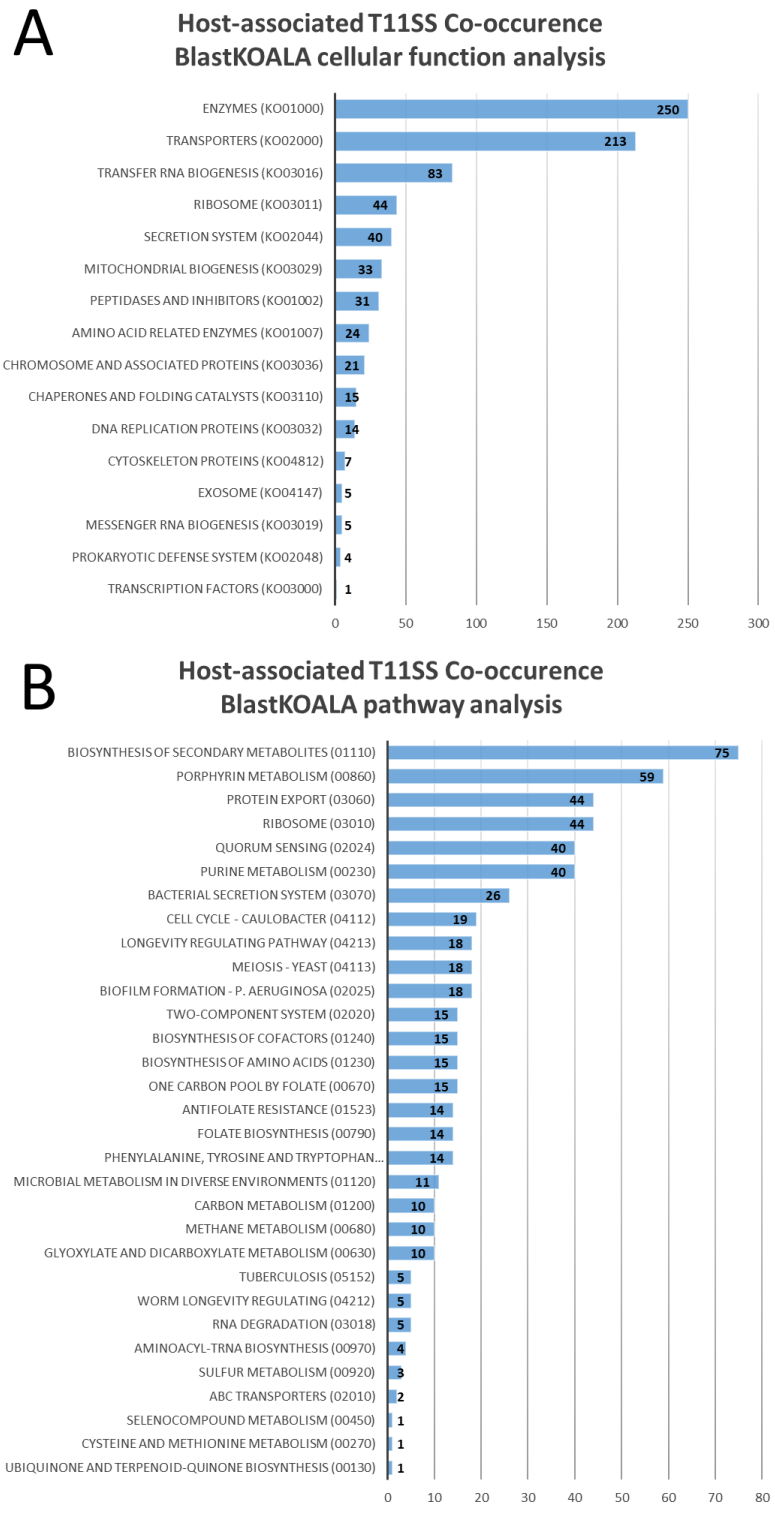


Figure 5.4. **BlastKOALA supports T11SS association with iron/heme uptake, protein export, and single carbon metabolism**. BlastKOALA uses the BLAST alignment algorithm to assign functions to query sequences and reveal shared pathways. Cellular functions were estimated using brite hierarchies and assigned to known pathways.

homology group (Suzek et al. 2007). The 3405 significantly co-occurring sequences fell into 666 UniRef50 clusters (see tab 4 of Chapter5CooccurenceDatasets.xlsx). The reference sequences of these UniRef50 were BLAST searched using the N-terminal handle domains and the C-terminal β-barrel domains from experimentally verified T11SS cargo proteins: NilC, TbpB, LbpB, HrpC, CrpC, Hpl, and Factor H binding protein with a low stringency E-value cutoff of 0.05 (Grossman, Escobar, et al. 2022; Grossman, Mauer, et al. 2022; Hooda et al. 2016; Hooda, Lai, and Moraes 2017). No matches were found for NilC or the N-terminal ligand binding domain of fHbp, but all other queries matched at least one cluster. This analysis revealed 141 clusters of potential T11SS-dependent cargo proteins representing 2656 ORFs across 1048 species/taxons (see tab 5 of Chapter5CooccurenceDatasets.xlsx). Each sequence was also annotated with SignalP 6 (Almagro Armenteros et al. 2019) revealing 81 Sec secreted lipoproteins, 33 Sec secreted soluble proteins, and 27 proteins with no predicted signal peptide. Several of the proteins with no signal peptide were annotated as fragments or pseudogenes, supporting the previous observations of T11SS transporting Sec dependent lipoproteins and soluble proteins.

We utilized multiple sequence alignment to separate the predicted cargo protein clusters into distinct architectures and annotate them (see tab 5 of Chapter5CooccurenceDatasets.xlsx). Predicted cargo clusters with homology to hemophilin proteins (Hpl, HrpC, HsmA) were termed Hemophilin-like proteins 1 (96 lipoproteins, 912 soluble). A separate group of homologs were found with similar N-terminal ligand binding domains and C-terminal β-barrel domains that we termed Hemophilin-like proteins 2 (65 lipoproteins, 7 soluble). Predicted cargo clusters with homology to TbpB proteins were termed TbpB-like proteins (638 lipoproteins, 7 soluble). Predicted cargo clusters with homology to fHbp protein were termed fHbp-like proteins (54 lipoproteins, 285 soluble). One cluster appeared to be lactoferrin binding protein (105 lipoproteins). Several homologous clusters encoded lipoproteins with an elongated N-terminal ligand binding domain with a disordered region,

these were termed disordered N-terminus 1 (196 lipoproteins, 8 soluble) and 2 (1 lipoprotein). Another group had elongated N-terminal domains with repeat rich (EEARKA motif) α-helical regions with homology to the Pls surface glycoprotein in *Staphylococcus aureus*, termed Pls-like proteins (173 lipoproteins, 20 soluble). Two phylogenetically restricted architectures were discovered that had multiple TbpBBD domains, *Sphingomonas* surface proteins which had an N-terminal sequence of proline-threonine repeats (18 lipoproteins, 12 soluble) and *Psychrobacter* surface proteins which had an N-terminal sequence of proline-threonine-aspartic acid repeats (55 lipoproteins, 4 soluble).

Using PsiPred 4.0 (Buchan and Jones 2019) we predicted the secondary structures of representative proteins from the largest UniRef50 cluster from each architecture with more than one representative (Fig 5.5). All of the architectures were predicted to have at least one domain, generally at the C-terminus, composed of β-strands linked by disordered loops. Using AlphaFold2 (Jumper et al. 2021) we predicted the tertiary structures of these representatives of the novel architectures, (Fig 5.6). The β -strands present in the N-terminal domains of most cargo were predicted to form handle domains similar to those seen in hemophilin or TbpB (Grossman, Mauer, et al. 2022; Hooda, Lai, and Moraes 2017). When present, the disordered regions and long α-helical repeat regions would precede this handle domain. Each handle domain was accompanied by a C-terminal β-barrel domain similar to the TbpBBD domain or the Lipoprotein C domain.

Demonstrating T11SS-dependent secretion of Plasmin sensitive surface protein

To determine if our bioinformatic predictions of T11SS cargo were accurate, we chose one of the novel cargo types and experimentally expressed a representative protein alongside its cognate T11SS protein within *E. coli*. Since no previously characterized T11SS cargo possessed the large **α**-helix repeat region we observed in several of our predicted cargo we chose to use the 525 amino acid Plasmin sensitive protein (Pls) from the "colicin transporter" cluster and its genomically



Figure 5.5. **PsiPred annotation of predicted T11SS-dependent cargo reveals diverse N-terminal domains featuring proline rich repeats, intrinsically disordered regions, α helical repeats, and ligand binding handles**. PsiPred 4 and DisoPred 3 use amino acid sequence to annotate secondary structures. For each of the detected cargo architectures the representative sequence from the largest UniRef50 cluster was trimmed of it signal peptide to reflect a mature protein and submitted.



Figure 5.5. (continued)

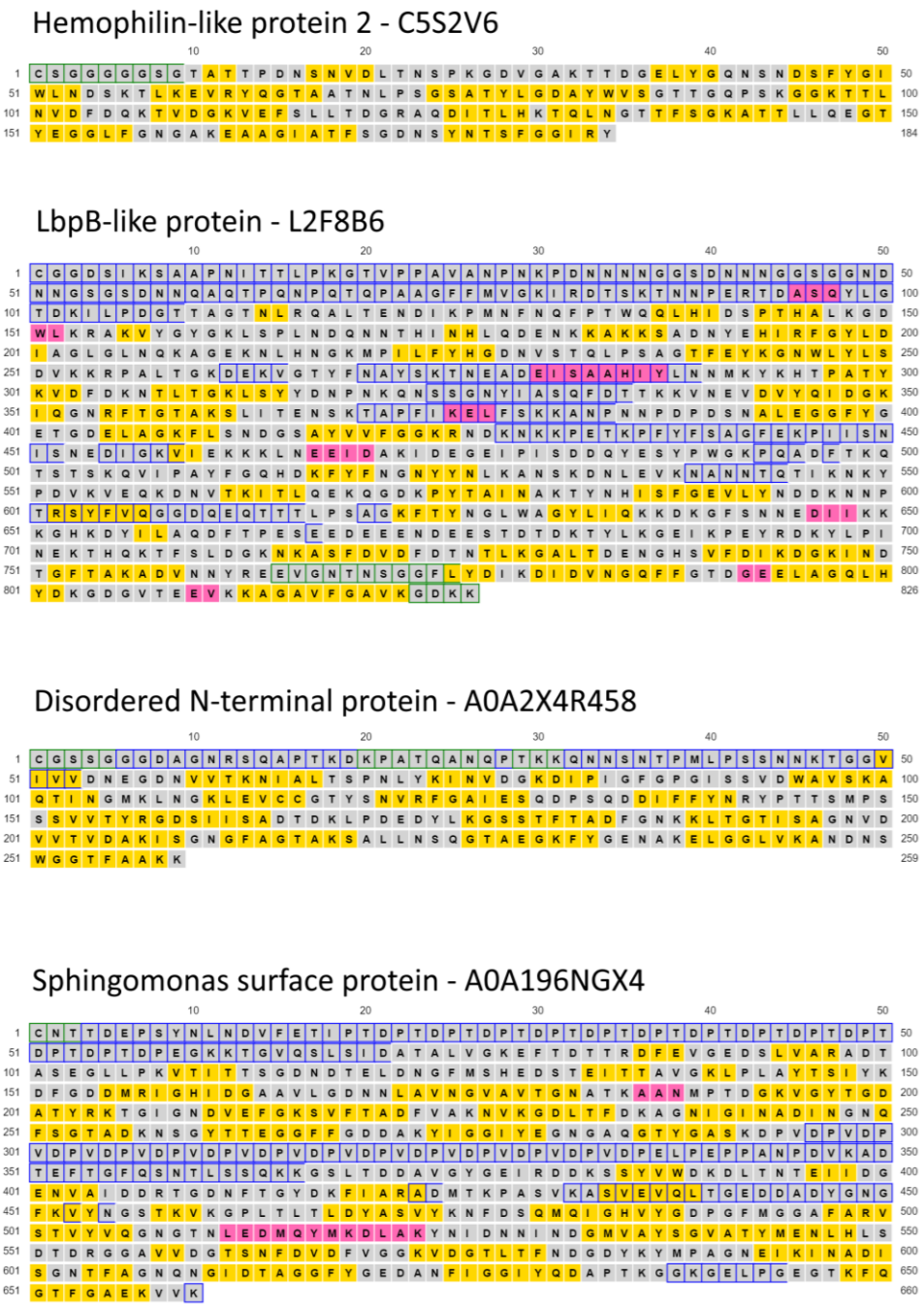


Figure 5.5. (continued)

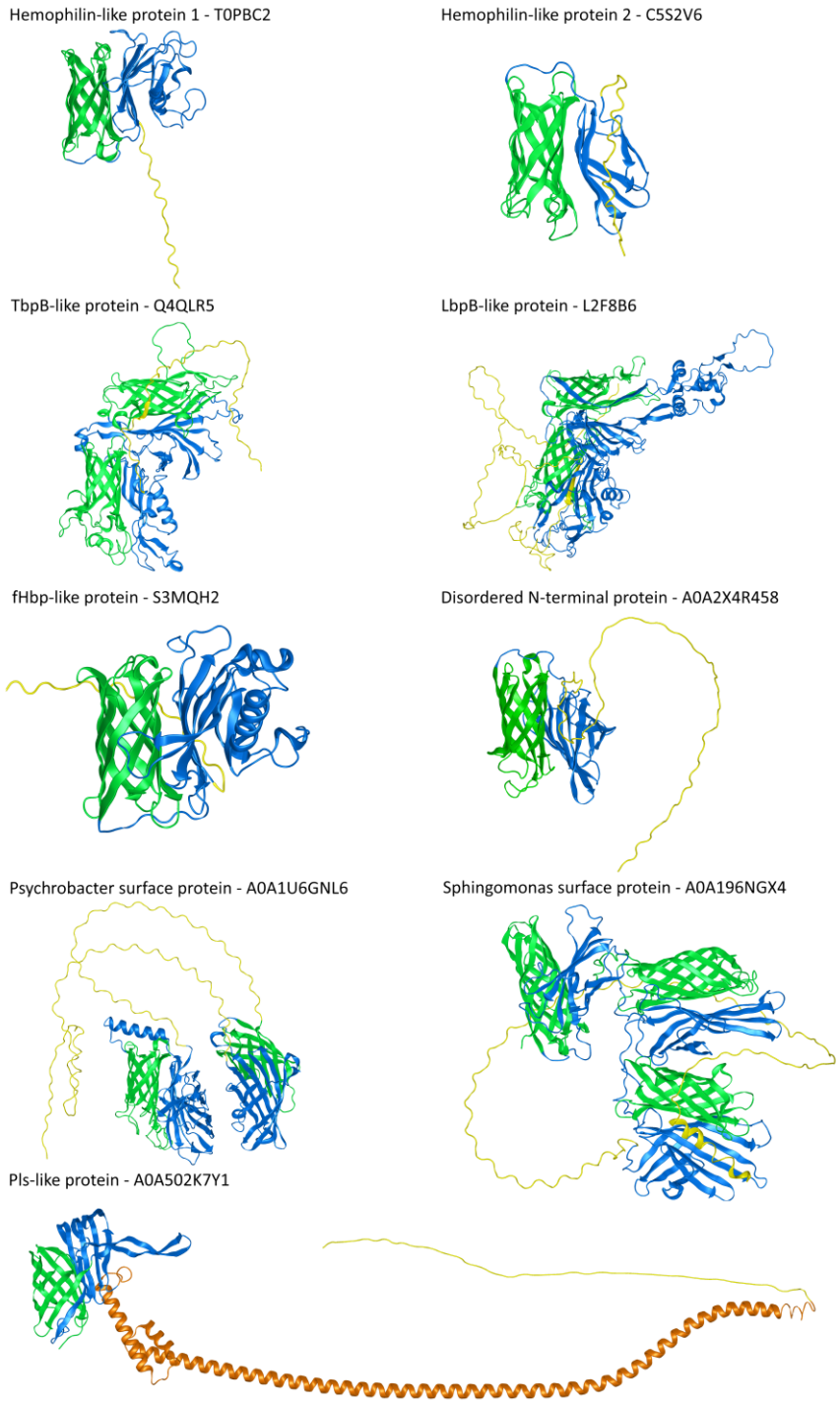


Figure 5.6. **AlphaFold2 annotation of predicted T11SS-dependent cargo reveals conservation of C-terminal or centrally located β-barrel domains**. For each of the detected cargo architectures the representative sequence from the largest UniRef50 cluster was used to predict a representative structure. Green residues indicate C-terminal β-barrel domains. Blue residues indicate putative ligand binding handle domains. Orange residues indicate EEARKA motif repeats within the Pls-like protein architecture. Yellow residues indicate regions predicted to be disordered.

associated putative type eleven Pls secretor (TepS) from *H. parahaemolyticus*. In *Staphylococcus aureus* Pls is a surface glycoprotein coated in N-acetylhexosaminyl residues that contributes to the virulence of MRSA strains by stimulating biofilm formation (Bleiziffer et al. 2017). Pls sequences from *Staphylococcus* species differ from the homolog seen in *H. parahaemolyticus* since they are larger (818-1540 residues in *S. aureus*) and because they do not have the C-terminal β-barrel domain common to T11SS-dependent proteins. Despite these architectural differences, the α-helix repeat region of Pls from *H. parahaemolyticus* shares between 49 and 62% identity with the repeat regions of Pls proteins from *S. aureus* (WP_256928426, WP_256927786, WP_258808513, WP_256933640, and WP_257570445).

To experimentally test for TepS-mediated secretion of Pls, pETDuet-1 based plasmids were constructed to either express Pls-FLAG alone, or to co-express Pls-FLAG alongside its predicted cognate T11SS, TepS. The empty vector was included as a negative control. These plasmids were transformed into *E. coli* BL21 DE3 C43 and induced with IPTG. Immunodotblots probing for the FLAG tag in intact and lysed cells were used to compare among treatments the surface exposure of FLAG tagged Pls. This analysis revealed that co-expression with its predicted cognate T11SS did not significantly impact surface exposure of Pls or the intracellular concentration of Pls (Fig. 5.7). This result was initially surprising, since Pls was predicted to be a surface anchored lipoprotein. However, since *S. aureus* Pls can be cleaved from the cell surface spontaneously and by proteinases such as plasmin (Bleiziffer et al. 2017; Savolainen et al. 2001) we monitored for the presence of secreted FLAG-tagged Pls in the supernatant fraction. SDS-PAGE and Western immunoblotting was used to compare protein present in clarified supernatants and within extracellular vesicles. These data revealed that Pls was present in the supernatant, but only when co-expressed with the T11SS protein TepS. The data also revealed that some Pls was located in extracellular vesicles independent of TepS expression, and that TepS expression increased the amount of Pls contained within those vesicles

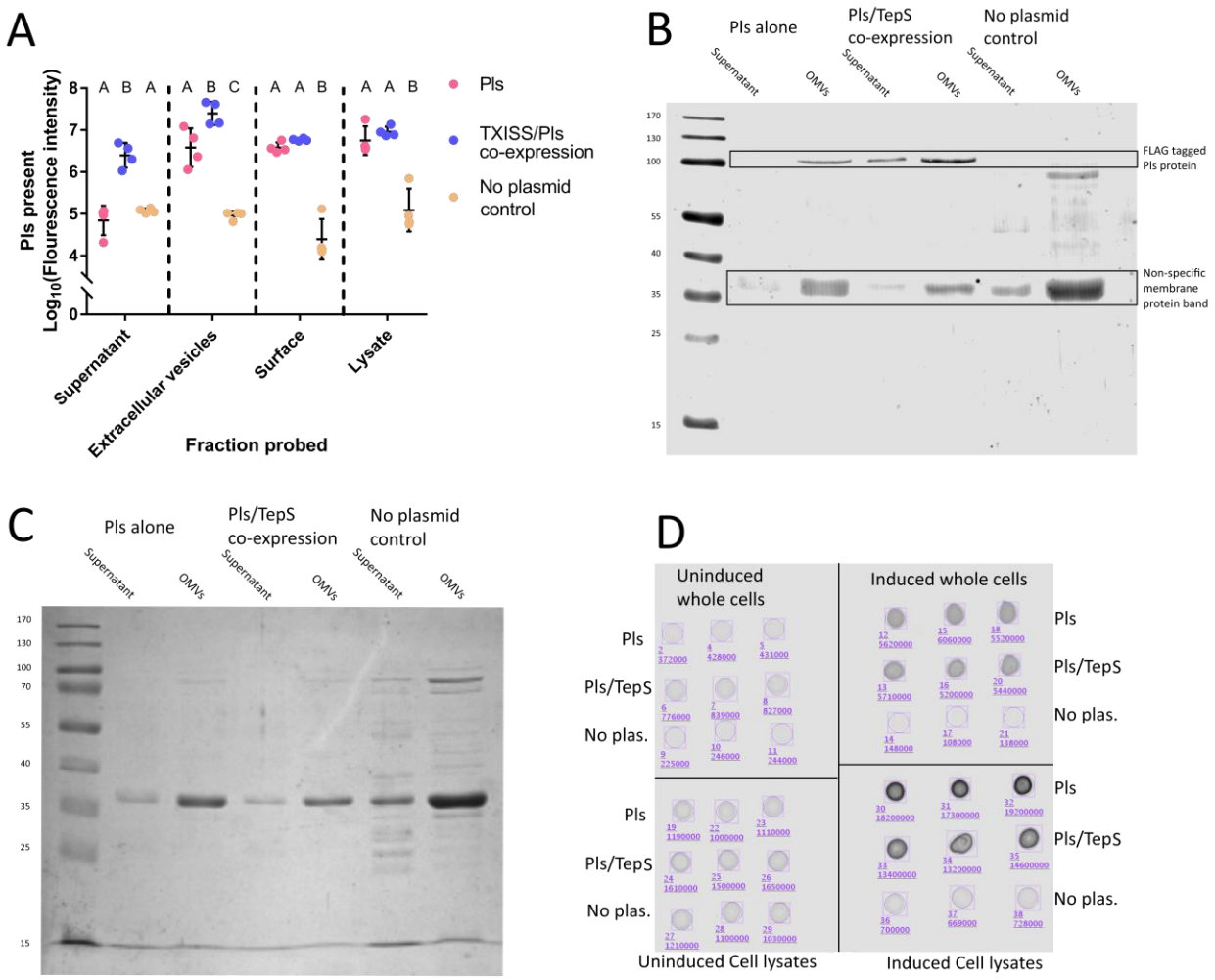


Figure 5.7. **Localization of FLAG tagged Pls protein in the absence and presence of its cognate T11SS secretor.** *E. coli* cells expressing Pls-FLAG alone or Pls-FLAG alongside its cognate T11SS (TepS) were probed using dot blots (cell surfaces and lysates) or Western blotting (supernatant and extracellular vesicles) to see how expression of the T11SS protein impacted protein localization. **B)** a representative Western blot demonstrating TepS dependent secretion of Pls. **C)** A representative Coomassie stain showing total supernatant protein content. **D)** Immuno-dot blots demonstrating that TepS was not required for surface exposure of Pls when expressed in *E. coli*. Supernatant secretion of Pls was only detectable in the presence of the T11SS TepS. The extracellular vesicle fraction collected via ultracentrifugation showed some Pls localization in the absence of TepS, however that localization was significantly increased by TepS. Pls was detected on the surface of both treatments equally.

(Fig. 5.7). In the immunoblot, FLAG-tagged Pls is visible as a single clear band in individual lanes, however the band varied in apparent size between biological replicates from 82.3 to 121.3 kD with an average of 103kD, which is almost twice the value predicted from sequence alone (56.8kD). This large and variable size may suggest that Pls is either being significantly slowed by glycan moieties as it travels through polyacrylamide, or that it exists as a dimer.

Bioinformatic analysis of T11SS proteins from marine environments suggests distinct roles and co-occurrence patterns

To test the robustness of our analysis we then utilized our co-occurrence technique to explore another cluster of potential T11SS proteins identified by Grossman et al. 2021 which predominantly contained sequences from marine microbes (cluster 3). This cluster almost exclusively contains sequences from the family Rhodobacteraceae, and in addition to containing pelagic microbes also includes algal symbionts (*Silicimonas algicola*), mollusc symbionts (*Aliiroseovarius crassostreae*), tunicate symbionts (*Ascidiaceihabitans donghaensis*), echinoderms symbionts (*Sulfitobacter delicatus*), and coral symbionts (*Roseivivax isoporae*) (M.-H. Chen et al. 2012; Crenn et al. 2016; Ivanova et al. 2004; Kessner et al. 2016; Y.-O. Kim et al. 2014). As above, we submitted all genes from the cluster (145 sequences) as queries to the genome neighborhood network function of RODEO, using the same parameters as the original query list, resulting in 203 co-occurring domains with frequencies between 2 and 102 (see tab 6 of Chapter5CooccurenceDatasets.xlsx). Domains which co-occurred fewer than 5 times were excluded from analysis (122 or ~60.0%). Co-occurring domains were compared to another random subset of the non-specific control co-occurrence dataset to assign FDR values. All domains with an FDR greater than 0.1 were excluded from analysis, resulting in 42 significantly co-occurring domains (see tab 7 of Chapter5CooccurenceDatasets.xlsx). The most common co-occurring domains were DUF560

(102/145 loci), DUF1194 (102/145 loci), glyoxalase domains (89/145 loci), and thymidylate synthase (77/145 loci). All genes that co-occurred with a marine cluster DUF560 and that contained any of the significant co-occurring domains were extracted resulting in 860 significant co-occurring genes which we submitted to KofamKOALA and BlastKOALA for functional analysis (Aramaki et al. 2020; Kanehisa, Sato, and Morishima 2016).

KofamKOALA found matches for 595 of the 860 significant co-occurring genes (see tab 5 of Chapter5FunctionalAnalysis.xlsx). When matched to cellular functions (brite hierarchies) the most common categories were enzymes (lactoylglutathione lyase, thymidylate synthase, aspartate aminotransferase, etc), prokaryotic defense systems (antitoxins CptB and HigA-1, topoisomerase IV B), and transcription factors (Cell division repressor DicA, cold shock protein) (Fig. 5.8A). Of the 595 matched sequences, 270 had known pathway association, the most common pathways being biosynthesis of one carbon pool by folate (dihydrofolate reductase, thymidylate synthase), pyrimidine/nucleotide metabolism (thymidylate synthase), pyruvate metabolism (lactoylglutathione lyase), and biosynthesis of amino acids (aspartate aminotransferase) (see tab 6 of Chapter5FunctionalAnalysis.xlsx). Unlike cluster 1, transposases were not among the significantly co-occurring genes of cluster 3. However, aminoacyl-tRNA synthesis functions, which can be associated themselves with mobile genetic elements (Alamos et al. 2018; Dobrindt and Reidl 2000), do appear on the co-occurrence list, including alanyl-tRNA, prolyl-tRNA, and threonyl-tRNA synthases (Fig. 5.8B). Parallel analysis with BlastKOALA identified 514 of the 860 significant co-occurring genes, of which 267 had known pathway association (see tabs 7 and 8 of Chapter5FunctionalAnalysis.xlsx). BlastKOALA did not identify any additional cellular functions or pathways not already revealed by KofamKOALA (Fig. 5.9).

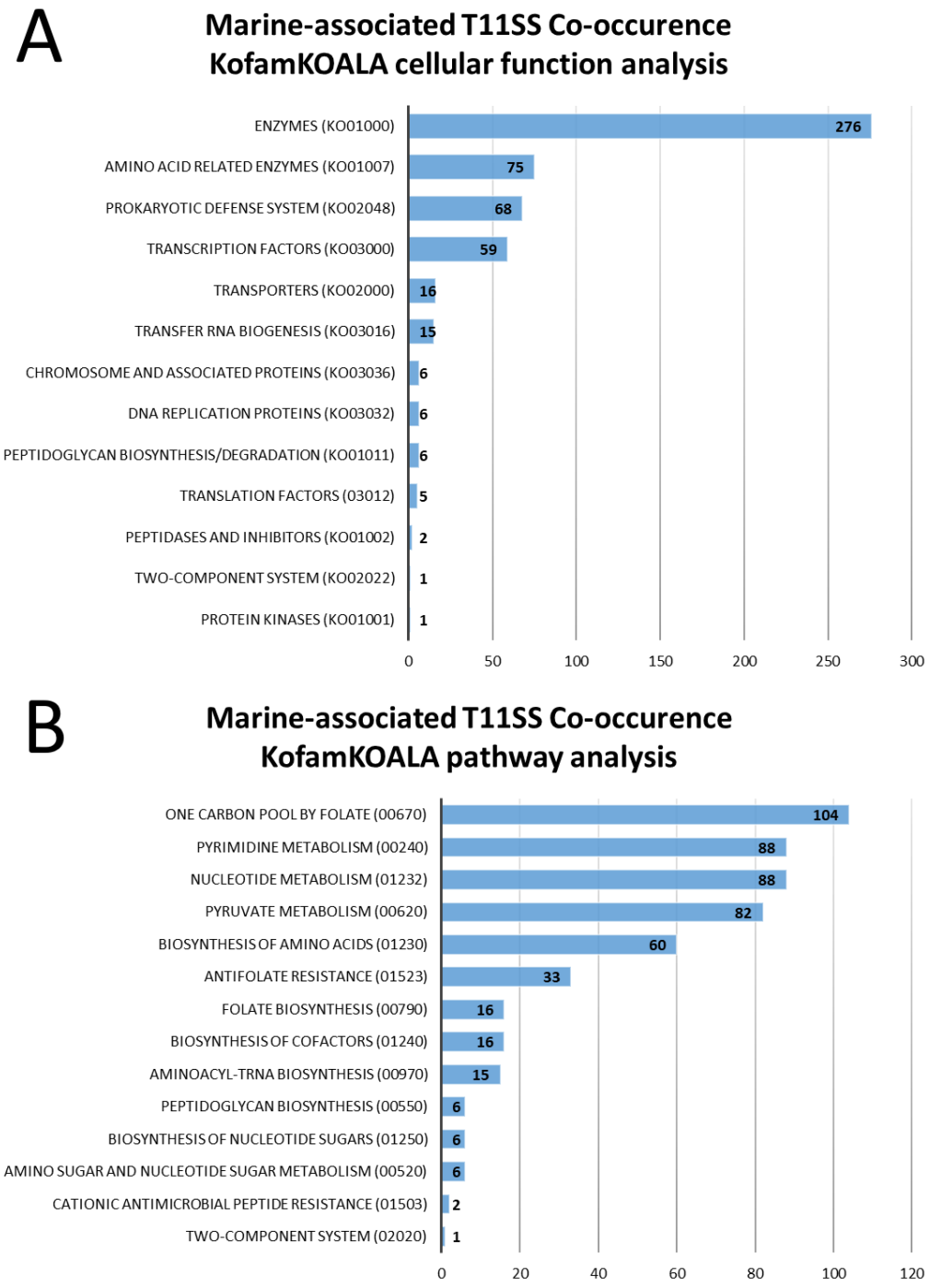


Figure 5.8 **Significantly co-occurring genes with marine associated T11SS reveal conserved association with single carbon metabolism, amino acid metabolism, and the glyoxalase detoxification pathway**. KofamKOALA uses hidden Markov models to assign functions to query sequences and reveal shared pathways. Cellular functions were estimated using brite hierarchies and assigned to known pathways.

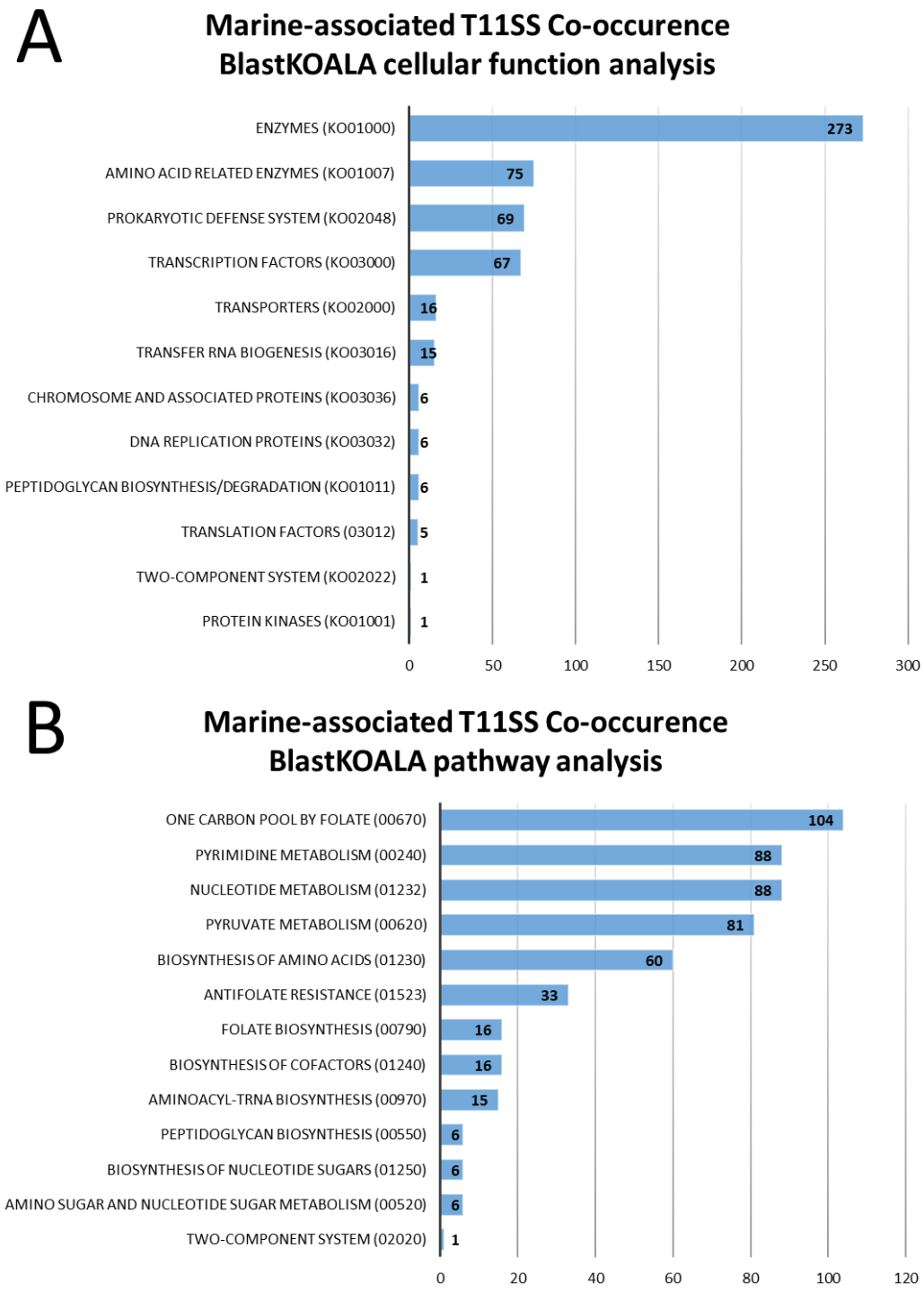


Figure 5.9 **BlastKOALA supports T11SS association with single carbon metabolism, amino acid metabolism, and the glyoxalase detoxification pathway**. BlastKOALA uses the BLAST alignment algorithm to assign functions to query sequences and reveal shared pathways. Cellular functions were estimated using brite hierarchies and assigned to known pathways.

Bioinformatic investigation of potential cluster 3 T11SS cargo

Initial attempts to identify new cargo proteins within the co-occurrence datasets from cluster 3 used the same methods used for cluster 1. Sequences were clustered into UniRef50 clusters with ≥50% identity, resulting in 218 homologous groups (see tab 9 of Chapter5CooccurenceDatasets.xlsx). These groups were again searched, using BLASTP, for sequence level homology to N-terminal handle domains and C-terminal β-barrel domains from experimentally characterized T11SS-dependent cargo. Only groups displayed similarity to these domains. Two are unlikely to be T11SS cargo since they are predicted to encode integral outer membrane proteins with weak homology to the ligand binding domains of previous cargo. The third candidate resembled the disordered N-terminus 1 architecture identified in the previous cargo prediction. However, this cluster only contained a single sequence (see tab 10 of Chapter5CooccurenceDatasets.xlsx). The near total absence of homologs of known T11SS cargo co-occurring with T11SS from the marine cluster suggests that T11SS proteins from the marine cluster either seldom co-occurred with their own cargo proteins, or that the cargo proteins have different domains from or below-threshold sequence similarity than those found in cluster 1.

To identify potential cargo proteins in another way, we compared the dominant co-occurrence patterns of the marine cluster 3 to those seen in the primary cluster 1. T11SS genes from cluster 1 co-occur with genes predicted to encode putative cargo proteins containing a TbpBBD or lipoprotein C domain, TonB-dependent receptors, and TonB, proteins that function together to export co-receptors and outer membrane uptake systems. These three domains are very rare within the list of significant co-occurrences from the marine T11SS cluster 3. Instead, the DUF1194 domain (PF06707) co-occurs with marine T11SS with a similar level of ubiquity (102/145 queries) (See Chapter5DUF1194-DUF560Coocurrence.xlsx). In fact, the majority of all marine TXISS

sequences from host-isolated bacteria have at least one DUF1194/T11SS co-occurrence. The predicted RoseTTAFold structure (Baek *et al.*, 2021) of the DUF1194 domain is a mixed β-strand/α-helix. According to the Pfam database, DUF1194 occurs in 31 distinct protein architectures with other domains, including C-terminal autotransporter domain (PF03797) predicted to function in T5SS secretion and the CARDB domain (PF07705), which adopts a 7 β-strand structure, and DUF11 (PF01345), predicted to have an 8-stranded β-barrel structure, and PF18911, the known crystal structures of which also adopt a β-barrel (e.g. 1wgo and 2y72). Of the 102 DUF1194-containing sequences found in the cluster 3 co-occurrence list, 7 are predicted to be lipoproteins, 76 are predicted to be soluble Sec-secreted proteins, and 19 have no detectable signal peptide according to SignalP 6. Many genomic neighborhoods contained a T11SS and 2 or 3 distinct DUF1194 genes in close proximity. In the case of the algal symbiont *Rhodobacteraceae* PD-2, a single genomic locus contains two DUF1194 homologs, predicted to be lipidated (ETA49263.2) and unlipidated (ETA49262.2), respectively (See Chapter5DUF1194-DUF560Coocurrence.xlsx). 6 of the co-occurring DUF1194 sequences also annotated as having homology to von Willebrand factor type A domain or vWA (PF00092). Both DUF1194 and vWA have similar Rossman folds and may be playing similar molecular roles.

The experimentally verified T11SS cargo are characterized by the presence of a common 8-stranded β-barrel C-terminal structural domain, even when they lack significant sequence similarity to each other. To examine potential structural similarities between known cargo and the DUF1194-containing homologs predicted to be encoded by cluster 3 co-occurring genes, we focused on three adjacently encoded DUF1194 genes from *Ascidiaceihabitans donghaensis* (SPH20589/SPH20588/SPH20587) for structural analysis. PsiPred 4.0 predicted that these proteins have at least 5 predominantly hydrophobic β-strands, all separated by α-helix regions in a structure that also appears to be a variation on the classical α/β Rossman fold (Fig. 5.10A). RoseTTAFold

predictions of the tertiary structure of these DUF1194 proteins reveal globular proteins wherein 6 hydrophobic β-strands form a β-sheet at the core of the protein, which is protected by three α helices per side. For each protein, the last 11-13 C-terminal residues appeared to be disordered and their positions could not be estimated with a high degree of confidence (Fig. 5.10B).

Discussion:

Our data reinforce the idea that cluster 1 T11SS-dependent cargo appear to have roles in metal uptake, single carbon metabolism, and nutrient provision. Further, our data extend the experimentally verified T1SS cargo to include *Haemophilus parahaemolyticus* Pls. The structural prediction that this cargo protein has fused a functional Pls domain with a C-terminal 8-stranded β-barrel domain further supports the importance of this structure in directing cargo to the T11SS. However, the fact that known domains that adopt this structure (TbpBBD and lipoprotein C) were not represented among the genes significantly co-occurring with cluster 3 T11SS is intriguing, and suggests that novel T11SS targeting domains may await identification.

Our bioinformatic approach relied upon gene co-occurrence analysis, which uses genomic proximity as a proxy for coordinate regulation and potential interactions, since genes within a common functional pathway often cluster together within a genome. While this is not a foolproof method, particularly when the range of available genomes is limited, by using large genomic datasets we can begin to refine the hypothesis generating power of co-occurrence analysis. This study establishes *in silico* controls for genome neighborhood co-occurrence. These controls were implemented to reduce false positives and noise in the output datasets of putative T11SS co-occurring protein families. For instance, activities associated with protein translation and RNA

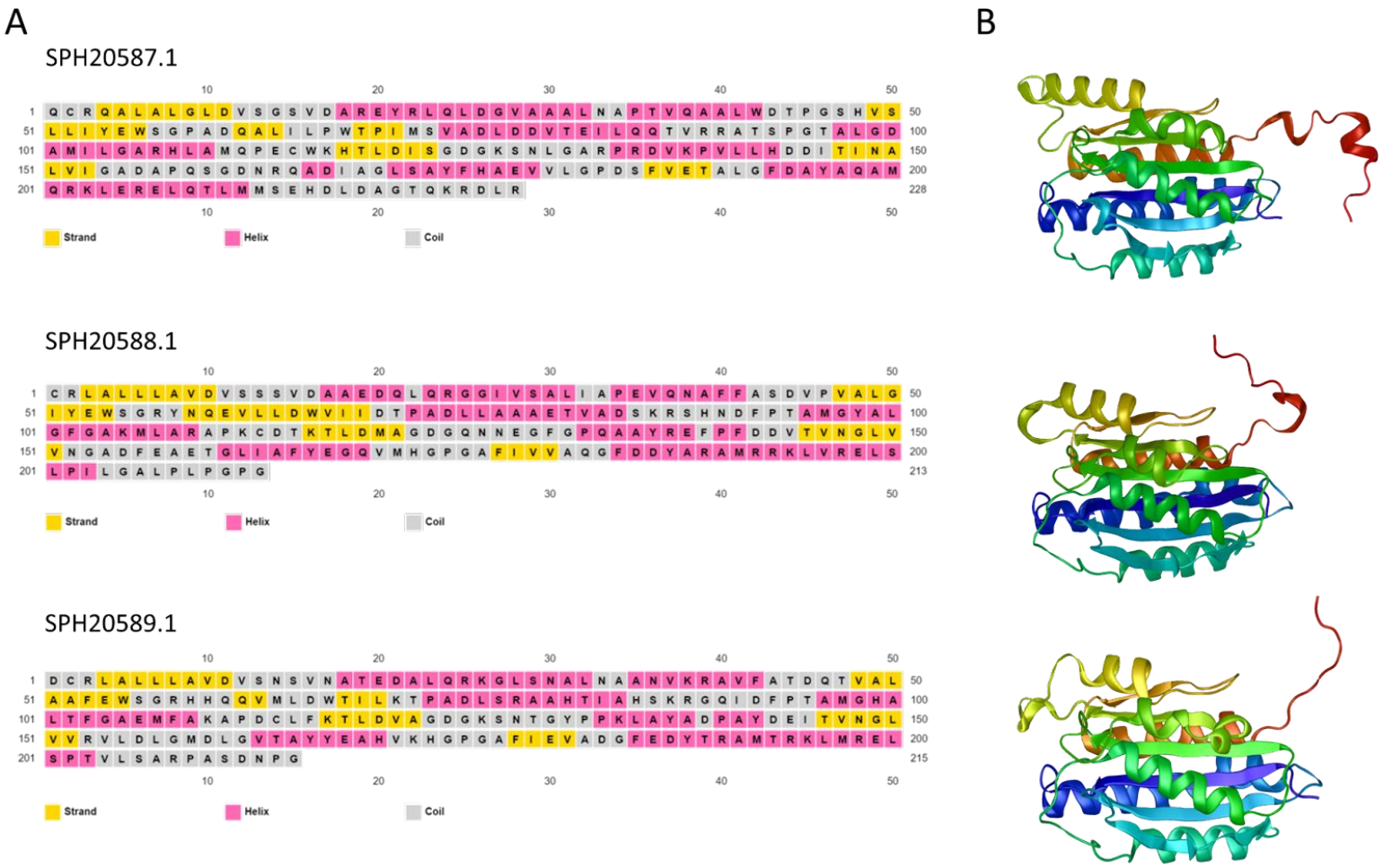


Figure 5.10 **Representative structures of DUF1194 proteins located within a T11SS genomic neighborhood in *Ascidiaceihabitans donghaensis*.** These three DUF1194 homologs are encoded adjacent to a T11SS OMP-encoding gene within a microbe isolated from a tunicate host. These proteins appear to be paralogs of each other based on sequence similarity. Signal peptides were predicted with SignalP 6 and trimmed from the sequences prior to submission to PsiPred **(A)** and RoseTTAFold **(B)**.

metabolism, which are essential cellular processes expected to be present in every cell, are not always useful co-occurrence signatures for elucidating T11SS function and components. Such general functions are not indicative of the specific functions of T11SS relative to any other OMP. Our work lays a groundwork that other researchers can build upon when designing controls for refining co-occurrence datasets. These non-specific control datasets decrease the impact of non-specific and ubiquitous co-occurrences, which could be considered a form of false negative, Type II error; such co-occurrences are not biologically relevant, but are not being excluded from the analysis because they do co-occur frequently with the query gene. However, removal of such co-occurrences can also lead to false positive, Type I error, because they can also result in removal of some biologically relevant co-occurrences. For example, there could be protein families that are important functional partners of the query protein, but that are removed from the dataset if they were important for both query and control proteins. This type of error occurred in our analysis. For example, TonB-dependent receptors, with which many T11SS-dependent cargo interact, were excluded from the cluster 1 T11SS analyses since these domains associate with other outer membrane proteins unrelated to T11SS. In a sense, this control narrows the focus from total co-occurrence to unique co-occurrence, and as such it is best suited towards applications in which the researcher seeks to differentiate protein families. Additionally, the technique can be further adapted and refined, depending on application. For instance, multiple non-specific gene neighborhood controls could be generated by randomly subsampling the list of proteins that are biophysically similar to the query proteins, allowing for increased stringency and measures of variance.

Domains which significantly co-occurred with T11SS genes from the predominantly animal-associated members of T11SS cluster 1 are consistent with previous observations (Grossman, Escobar, et al. 2022; Grossman, Mauer, et al. 2022; Hooda, Lai, and Moraes 2017). Transferrin binding protein B, Lactoferrin binding protein B, Heme receptor protein C, and Hemophilin are all

T11SS-dependent cargo involved in iron uptake from the environment, and in the controlled gene neighborhoods of T11SS proteins we found heme oxygenases and heme transporters that facilitate iron uptake (Hantke 2003). Additionally, many of the protein export genes that significantly co-occurred with T11SS supported previous observations. TonB and ExbD are essential to energize uptake of nutrients from known T11SS-dependent cargo like TbpB, LbpB, and hemophilin (Bateman et al. 2021; Hooda et al. 2016; Ollis and Postle 2012). TolA is essential to bring periplasmic nutrients into the cytoplasm of the cell (Levengood, Beyer, and Webster 1991). Signal peptidase II is essential for the generation of bacterial lipoproteins, which constitute a large percentage of T11SS-dependent cargo (Tjalsma et al. 1999). Finally, the observed co-occurrence with transposases, ribosome components, and tRNA-synthases is consistent with the hypothesis that T11SS are regularly associated with mobile genetic islands and can be horizontally acquired and contribute to fitness within a host environment (Dobrindt and Reidl 2000).

In addition to the expected iron uptake pathways, folate (vitamin B9) biosynthesis, folate-dependent enzymes, and one carbon metabolism via folate emerged as being associated with the host associated T11SS cluster 1 and the marine associated cluster 3. Folate, like heme, is an enzyme cofactor required for many diverse biological pathways. Folate is essential for thymidylate synthase (the third most common co-occurring gene from the marine cluster) to participate in nucleotide biosynthesis. However, unlike heme, folate does not incorporate a metallic ion and most bacteria synthesize their folate instead of scavenging it from the environment (Dawadi et al. 2017). Based on these observations, we do not favor the hypothesis that folate itself is acquired through a T11SS-dependent uptake mechanism, though this remains formally possible. Instead we interpret the close association of folate metabolism and T11SS to reflect the fact that folate biosynthesis itself depends on two different metal bearing enzyme cofactors, cobalamin (vitamin B12) and molybdopterin cofactors (MoCo) (Luk et al. 2013; Mendel 2013). Vitamin B12 is a cobalt bearing porphyrin which

can be methylated by folate into methylcobalamin, which drives one carbon metabolism and is required for MetH to convert homocysteine into the amino acid methionine (Old et al. 1990). Molybdenum/tungsten/selenium bearing molybdopterins are required for bacterial folate metabolism, iron-sulfur cluster biosynthesis, and methylotrophic metabolism by formate dehydrogenase (which also significantly co-occurred with T11SS) (Mendel 2013). Furthermore, in eukaryotes molybdopterin cofactors bind heme as a redox active cofactor (Ringel et al. 2013) , though this interaction is not well characterized in prokaryotes. Periplasmic molybdenum ions can be transported across the inner membrane through ABC transporters, either from low affinity (MolABC) or high affinity (ModABC) transporters (Zhong, Kobe, and Kappler 2020). Reduction of Molybdenum uptake leads to a virulence defect in Gram-negative bacteria (Périnet et al. 2016). Given the conserved role of T11SS-dependent cargo in heme and iron uptake, future studies should probe whether T11SS proteins that co-occur with cobalamin or molybdenum cofactor dependent pathways can facilitate uptake of either metallated co-factor. Regardless of the role these metallic cofactors may have, it is clear that single carbon metabolism and methyltropism unifies many of the significantly co-occurring genes around host associated T11SS.

The functional analysis of significantly co-occurring genes from the marine T11SS cluster 3 lacked a co-occurrence with TonB energization or heme/iron uptake. This may suggest that the cluster 3 T11SS have distinct functions or cargo relative to cluster 1. Clusters 1 and 3 shared the association with single carbon metabolism, folate biosynthesis, and folate dependent nucleotide biosynthesis. Unique to the marine cluster co-occurrence analysis, the most common characterized co-occurring gene was lactoylglutathione lyase (GloA), which performs the final step of the glyoxalase detoxification system. This system uses glutathione to detoxify aldehyde compounds produced by various metabolic processes, and GloA specifically uses metal ion cofactors (typically nickel) to regenerate the glutathione intermediate (C. Lee and Park 2017). Furthermore,

detoxification by the glyoxalase system is required for virulence by *Listeria monocytogenes* (Anaya-Sanchez et al. 2021). It is possible that the nickel requirement of this enzyme reflects a role for cluster 3 T11SS in facilitating nickel uptake. Alternatively, this co-occurrence may indicate that cluster 3 T11SS are more generally participating in an aldehyde producing process. Notably, aldehydes like methylglyoxal are exceptionally cytotoxic and can damage DNA, and proteins, so the increased co-occurrence of nucleotide and amino acid biosynthesis genes seen in the marine cluster may also be part of an aldehyde repair response. *In vivo* analysis of the *Rhodobacteraceae* species identified within this analysis may help examine the relationship between T11SS and the glyoxalase detoxification system.

By mining our gene co-occurrence datasets, we identified up to 141 homology groups of T11SS-dependent cargo, some of which displayed completely new architectures. All of the predicted cargo either have a C-terminal hydrophilic β-barrel or appear to be gene fragments, indicating that this particular structure may be important for T11SS function or specificity. The N-terminal domains of these proteins varied structurally, potentially reflecting distinct functionality. We demonstrated that bona fide T11SS-dependent cargo can be identified based on co-occurrence combined with structural domain predictions; the plasmin-sensitive surface protein (Pls) from the Gram negative organism *H. parahaemolyticus* relies on expression of a genomically-associated type eleven Pls secretor (TepS) to reach the bacterial cell surface.

Pls represents a novel N-terminal effector domain among T11SS cargo. This is particularly interesting, since this N-terminal effector domain is related to Pls proteins found in Gram positive organisms such as *Staphylococcus aureus* and *epidermidis*. In these organisms, Pls is a heavily glycosylated protein located on the cytoplasmic membrane surface of the cell where it contributes to biofilm formation, regulates/prevents adhesion to host proteins such as fibronectin, and is spontaneously

cleaved from the cell (Bleiziffer et al. 2017; Savolainen et al. 2001). Given that the function of Pls requires surface exposure, its horizontal transmission from a Gram positive organism to a Gram negative organism is complicated by the existence of the outer membrane. This hurdle seems to have been overcome via fusion of the Pls domain with the C-terminal TbpBBD domain necessary for T11SS-dependent secretion across the outer membrane. This C-terminal β-barrel structure, which is present in all characterized T11SS-dependent cargo (Grossman, Mauer, et al. 2022; Hooda, Lai, and Moraes 2017), is absent from gram-positive Pls homologs. We show here that the *H. parahaemolyticus* Pls, like its gram-positive counterparts, can be localized to the *E. coli* cell surface. Further, our observations suggest that *H. parahaemolyticus* Pls can be glycosylated, based on the fact that its migration in SDS-PAGE gels indicates a far larger size than expected. Our analyses revealed that when co-expressed with its T11SS partner, TepS, in *E. coli,* Pls predominantly localizes to the supernatant and is not tethered to the cell surface as would be expected of a lipoprotein. Our working model is that some Pls can reach the cell surface in the absence of T11SS, as has been observed for factor H binding protein (Fantappiè et al. 2017; Konar et al. 2015), and that in the presence of the T11SS TepS, Pls is secreted but then cleaved, releasing it into the extracellular milieu despite expression in a protease deficient strain of *E. coli*. It will be exciting in future studies to determine if TepS is directly responsible for Pls cleavage or if it alters the topology of Pls in order to make it susceptible to proteolysis.

Our finding that *H. parahaemolyticus* Pls is T11SS-dependent and may be cleaved from the cell surface in the process of secretion has important implications for efforts to use this protein to control human pathogens. *Actinobacillus pleuropneumoniae* encodes a close homolog of Pls (YP_001652736.1) that is expressed on the cell surface. It is being pursued as a candidate target for vaccine development, making it important to consider its translocation across the outer membrane and its potential localization in the extracellular milieu (Y. Cao et al. 2020). The Pls homologs from

A. pleuropneumonia and *Neisseria sicca* are annotated as colicin uptake proteins based on homology to the tandem repeat region of TolA from *E. coli* (Levengood, Beyer, and Webster 1991). However, in light of their C-terminal homology to other TXISS dependent cargo and localization to the outer membrane we conclude that this automatically-generated annotation is likely misleading. Instead, we propose that in gram-negative bacteria, the Pls domain has become fused to a T11SS-targeting domain to facilitate its secretion, where it functions, either on the cell surface or in the extracellular milieu, as an effector of host or other environmental interactions, such as adhesion to surfaces. The Pls-like proteins have repeat rich regions resembling *N*-linked bacterial glycosylation motifs (Meng et al. 2019), and the possible glycosylation of *H. parahaemolyticus* Pls is supported by our observations of its aberrant mobility in SDS-PAGE. It is possible that other T11SS-dependent cargo, including lipoproteins, may also be glycosylated. If so, such glycoproteins tethered to the cell surface would display glycans as surface antigens that may modulate host-microbe recognition and immune evasion (Schmidt, Riley, and Benz 2003; Tortorelli et al. 2022; Zhou and Wu 2009).

In addition to Pls, we identified in our co-occurrence network several other novel, candidate N-terminal handle/ligand-binding domains that are associated with the C-terminal T11SS-targeting domain. Several of these, from *Sphingomonas* and *Psychrobacter,* appear to be disordered N-terminal protein architectures, and these may require post translational modification to fold or may just remain flexible in order to facilitate flexible protein surface exposure. The *Sphingomonas* surface proteins identified here contained ~21 proline-threonine repeats predicted to remain disordered. Similar disordered linker regions are found in *Xanthomonas campestris* endoglucanase, where they confer greater elasticity than a standard linker and facilitate enzyme positioning around its substrate (Skaf, Polikarpov, and Stanković 2020). The *Psychrobacter* surface proteins contained ~12 proline-threonine-aspartic acid repeats (PTD motif) which has been previously identified in a family of surface proteins from the *Gracilibacteria* (BD1-5) (Sieber et al. 2019). These bacteria are believed to be

episymbionts based on their reduced genomes. Within these organisms the PTD bearing surface proteins are hypothesized to act as adhesins and as such may facilitate *Gracilibacteria*'s episymbiotic lifestyle. The eukaryotic microbe *Plasmodium falciparum* also encodes a PTD repeat rich surface protein used for binding host factors named sporozoite surface protein 2 (Sieber et al. 2019; Swearingen et al. 2016). All of these cargo architectures with N-terminal disordered regions are predominantly predicted to be lipoproteins, suggesting that these flexible regions help surface proteins extend beyond the cell surface and interact with their substrates.

Mining our dataset of significantly co-occurring genes from the marine cluster 3 for T11SS-dependent cargo revealed a near complete absence of the C- terminal β-barrel motif that is ubiquitous among thus-far characterized cargo proteins. This suggests that these proteins either seldom co-occur with their own cargo proteins, or that their cognate cargo proteins are not homologous to previously characterized cargo. Turning to the significantly co-occurring domains to find an alternative family of cargo proteins revealed that the DUF1194 domain co-occurred with cluster 3 T11SS exceptionally often. Furthermore, marine cluster 3 T11SS loci often had multiple, tandemly-encoded DUF1194 proteins, similar to the pattern observed for the animal-associated cluster 1 T11SS which had multiple adjacent genes encoding TbpBBD proteins. The function of DUF1194 is currently unknown. It is most commonly found within Alphaproteobacteria from the orders Rhodobacterales and Hyphomicrobiales. DUF1194 is predicted to fold into a globular protein with an α/β Rossman fold. Several of the DUF1194 proteins we detected also had homology to the very similar α/β Rossman fold domain annotated as a von Willebrand factor A (vWA). The vWA domain can be found in prokaryotes and eukaryotes, and is most often associated with serum glycoproteins and large multimeric protein complexes in mammalian systems such as those generated by clumping factor A on the surface of *S. aureus* (Peyvandi, Garagiola, and Baronciani 2011; Von Willebrand 1999). Several vWA domain-containing proteins are also known to

bind divalent metal cations such as MG^{2+} in a mechanism which has been termed metal ion-dependent adhesion sites (MIDAS) (Cantí et al. 2005). Our data do not allow us to conclude if DUF1194 proteins are a type of T11SS-dependent cargo, or if they contribute to a multimeric protein complex that interacts with T11SS proteins. However, the remarkable frequency of T11SS-DUF1194 co-occurrence suggests that they are very likely to interact or have functional overlap.

In summary, our findings have enhanced our understanding of potential effectors of the novel T11SS, as well as expanding our list of T11SS secreted cargo. We also have expanded the field of genomic co-occurrence analysis by establishing a protocol that can serve as a basis for more controlled and specific experimental designs. Future research should focus on generating an automated pipeline for controlling co-occurrences analyses, determining if Pls is required for host-colonization by *Haemophilus parahaemolyticus*, investigating if DUF1194 represents a novel T11SS cargo type, and exploring the relationship between one carbon metabolism and T11SS.

Materials and methods:

Gene neighborhood analysis

The T11SS genome neighborhoods (±6 genes around a T11SS gene) from cluster 1 (1111 queries) and cluster 3 (145 queries) (Grossman, Mauer, et al. 2022) were analyzed. T11SS genome neighborhoods were generated through Rapid ORF Detection & Evaluation Online (RODEO) (Tietz et al. 2017). As described in the main text, a control was developed to identify and remove non-specific genes from among those flagged as co-inherited or syntenic with T11SS (Fig 3.3). The primary control list was generated by extracting all Go term outer membrane proteins from proteobacteria within 1 amino acid of the median T11SS length and removing all DUF560 containing proteins. For both T11SS cluster analyses a random subset of control sequences were chosen equal to the number of queries being analyzed. Controlled protein sequences were

downloaded from UniProt (UniProt: the universal protein knowledgebase in 2021 2021) in February 2021. Co-occurring protein domains were annotated and summed up to determine frequency among queries and controls. False Discovery Rate was calculated by dividing frequency among T11SS queries and control sequences. All domains with and FDR > 0.1 were excluded from analysis. For cluster 1 all domains with fewer than 25 occurrences were excluded. For cluster 3 all domains with fewer than 5 co-occurrences were excluded.

Functional and structural analyses of co-occurring gene neighborhood domains

Protein sequences that contain domains that co-occur with T11SS from the gene neighborhood analysis were obtained via UniProt (UniProt: the universal protein knowledgebase in 2021 2021). KofamKOALA (Aramaki et al. 2020) and BlastKOALA (Kanehisa, Sato, and Morishima 2016) were used to assess function. KofamKOALA assigns KEGG orthology to user sequence data by HMMER/HMMSEARCH against the KEGG database. BlastKOALA is similar but uses a BLAST search to assign KEGG orthology. Genes were matched to brite hierarchies to assess general cellular function and KEGG pathway in order to assess potential interactions.

To identify potential cargo proteins all downloaded protein sequences were matched to UniRef50 clusters in order to compress potential homologs (Suzek et al. 2007). The reference sequences from these UniRef50 clusters where then subjected to a BLAST search against known T11SS cargo β-barrel and handle domains (nematode intestine localization protein (NilC) in *Xenorhabdus nematophila*, transferrin binding protein (TbpB) N-terminal and C-terminal in *Xenorhabdus nematophila*, heme receptor protein (HrpC) in *Xenorhabdus nematophila*, Cobalt receptor protein (CrpC) in *Xenorhabdus cabanillasii*, Hemophilin (Hpl) in *Haemophilus haemolyticus*, hemophilin (HphA) in *Acinetobacter baumanii*, factor H binding protein (fHbp) *Neisseria meningitidis*, and lactoferrin binding protein (LbpB) N-terminal and C-terminal in *Neisseria meningitidis* to search for homologs of T11SS

cargo with a low stringency E-value filter of <0.05 in order to find even distantly related homologs, generating a list of 141 potential cargo proteins from cluster 1 and a single potential cargo from cluster 3. MSA alignment was used to group the predicted cargo proteins into distinct architectures. 10 groups emerged, TbpB-like, LbpB-like, Hemophilin-like 1 and 2, fHbp-like, Pls-like, highly disordered N-terminus 1 and 2, *Psychrobacter* surface protein, and *Sphingomonas* surface protein. Structural data predictions for the 9 architectures which contained more than one sequence were generated through PSIPRED (Buchan and Jones 2019) for secondary structure and AlphaFold2 (Jumper et al. 2021; Mirdita et al. 2022) for tertiary structure.

Co-expression of Pls and its cognate T11SS protein TepS

All cultures were grown in glucose minimal media supplemented with 1% LB (Bhasin, Chaston, and Goodrich-Blair 2012). Plate based cultures were grown on glucose minimal plates (Bhasin, Chaston, and Goodrich-Blair 2012). For plasmid-based expression, chemically competent *Escherichia coli* strain BL21-DE3 (C43) were chosen for ease of transformation and their ability to non-toxically express membrane proteins (Dumon-Seignovert, Cariot, and Vuillard 2004; Miroux and Walker 1996). Strains of *Escherichia coli* were grown at 37°C. Where appropriate media was supplemented with ampicillin at a concentration of 150 μg/ml. Protein expression was induced at the midlog point of bacterial growth via addition of isopropyl β-d-1-thiogalactopyranoside (henceforth IPTG) at a concentration of 0.5mM.

The genes encoding Pls-FLAG and 6xHis-TepS were synthesized and cloned into Multiple cloning site 1 and Multiple cloning site 2 of pETDuet-1 respectively by Genscript, resulting in expression plasmids pETDuet-1/Pls-FLAG and pETDuet-1/6xHis-TepS/Pls-FLAG. Expression plasmids were transformed into *E. coli* via electroporation. Strains were grown in defined medium with 150 μg/ml ampicillin (Orchard and Goodrich-Blair 2004). Bacteria were subcultured into

100 ml of broth at an initial optical density at 600 nm (OD_{600}) of 0.1, grown for 6 h at 37°C to reach late logarithmic growth, and induced with 500 μM isopropyl-β-d-thiogalactopyranoside (IPTG) for 1 hour. To monitor surface exposure of Pls whole cells normalized to an OD600 of 1, rinsed 3x, and spotted in technical triplicate on nitrocellulose membranes. The remaining cells were lysed open via 30 seconds of sonication at ~500 root mean square volts (V_{rms}). Cell lysates were also spotted in technical triplicate on nitrocellulose membranes to enumerate total cellular Pls. To monitor extracellular secretion of Pls supernatant was collected from cultures after induction and filtered sterilized to remove all cells. Supernatant samples were ultracentrifuged at 150,000 relative centrifugal force (RCF) for 3 h to separate the soluble fraction from the insoluble component, composed of cell membrane components and membrane vesicles. The membrane vesicle fraction was suspended in protein sample buffer. 700 μl of each soluble supernatant fraction was precipitated via 10% trichloroacetic acid (TCA) and the resulting pellet was suspended in protein sample buffer (Koontz 2014b). Samples were analyzed by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting. All dot blots and Western blots were probed with rat anti-FLAG primary antibody and goat anti-rat secondary antibody conjugated to a 680nm fluorophore. Intensities were recorded for FLAG reactive bands. For every supernatant sample, a band from the Coomassie blue-stained gel was used as a loading control to normalize intensities of supernatant samples prior to analysis. Fold change of secretion was determined by dividing the amount of supernatant Pls detected in the Pls/TepS co-expression treatment by the amount detected in the Pls alone treatment. A Tukey's honestly significant difference (HSD) test was used for comparing fold change of secretion (Tukey 1949).

Acknowledgements:

We thank the HGB Lab members for feedback on the co-occurrence controls.

Chapter 6: Conclusions and Avenues for Expansion

Author: Alex Grossman

It is a well-known and cherished tenet of microbiology that the smallest of things all too often led to the largest discoveries. By developing "simplified" models and observing isolated cells we slowly uncover and describe the phenomena of increasingly complex and interconnected systems. I find beauty in this humble philosophy and believe that it stands at the heart of the studies described within this dissertation. On my path to describing the behavior of a single bacterium and a single animal, I uncovered a novel bacterial secretion system with conserved roles in facilitating host-microbe symbioses, that we termed the type eleven secretion system (henceforth T11SS). My predecessors in the Goodrich-Blair lab and I initially endeavored to describe the molecular mechanisms by which *Xenorhabdus nematophila* went about colonizing its nearly microscopic nematode hosts (0.4 - 5mm in length). This question led Drs. Heungens, Cowles, Bhasin, Chaston, and Goodrich-Blair to identify the nematode intestinal localization (*nil*) locus, which encoded what would eventually be described as the T11SS NilB and its cognate cargo surface lipoprotein NilC (Cowles and Goodrich-Blair 2004, 2008). Few homologs were known at that time, and all are encoded by microbes that also made their homes within host mucosa. When I set about describing this odd locus, with few predicted homologs, from a tiny nematode I did not expect to find something far larger. But with each experiment the apparent functional range of this protein family grew.

When I began working in the Goodrich-Blair lab I set about expanding this project in two directions. Diving deeper into *Xenorhabdus* biology, I wanted to determine what molecular roles T11SS loci were playing within this bacterium to facilitate symbiosis. Extending outward, I wanted to use the rich sequence databases available to see if T11SS might mediate other host-microbe symbioses such as those found in the human microbiome and environmentally relevant species.

Chapters 2, 3, and 4 all strike a balance of these two approaches by integrating *in vivo* experiments focusing on *Xenorhabdus* with a combination of *in vitro* and *in silico* exploration of diverse T11SS. Chapter 5 leaves behind *Xenorhabdus* to focus on bioinformatic prediction of novel T11SS-dependent cargo proteins and *in vivo* demonstration of one of these novel cargo proteins.

The type 11 secretion system as understood today is composed of two families of proteins, outer membrane secretors possessing a DUF560/SlipAM domain and cargo proteins possessing small hydrophilic β domains(Grossman, Mauer, et al. 2022; Hooda et al. 2016). These cargo proteins can be either lipidated surface proteins or soluble extracellular proteins. Furthermore, all experimentally characterized T11SS-dependent cargo contribute to host-microbe interactions in some manner, either by binding host metal chelators (transferrin, lactoferrin, hemoproteins), binding host-immune factors (factor H), or interacting with host surfaces (Bateman et al. 2021; Grossman, Escobar, et al. 2022; Grossman, Mauer, et al. 2022; Hooda, Lai, and Moraes 2017; da Silva et al. 2019). These secretor/cargo pairs are highly specific and often locally associated in the genome. All known T11SS-dependent cargo have a Sec-transposon signal peptide targeting them to Sec for transport across the inner membrane, though a subset of these cargo encodes weak Tat motifs that may indicate some flexibility in the mechanism of inner membrane translocation. Studies of the transferrin binding protein B (TbpB) have shown that T11SS are dependent on the periplasmic chaperone Skp, indicating that some cargo proteins remain unfolded while in the periplasm and only fold after secretion through the secretor (Huynh et al. 2021).

Broadly the T11SS secretors present within *Xenorhabdus* species fell into two categories, NilB and what would eventually come to be known as heme receptor protein B (HrpB/HrpB2). The heme receptor protein locus was named for its association with a predicted TonB-dependent heme uptake co-receptor, and data presented in chapter 3 later confirmed that HrpB secretes a heme

binding protein called hemophilin (HrpC). NilB is present in only a few *Xenorhabdus* species while HrpB is present in all *Xenorhabdus* species for which we have genome sequences (and in many other Proteobacteria). Chapter 2 of this dissertation describes how my co-authors and I used HrpB and its cognate cargo protein to experimentally demonstrate for the first time T11SS secretion of soluble extracellular proteins. We also utilized sequence similarity networking to systematically identify and categorize T11SS according to their ability to secret lipidated or unlipidated cargo proteins. This was made possible by a unique structural feature of T11SS-dependent cargo, the near ubiquitous presence of a small hydrophilic β-barrel domain (TbpBBD or lipoprotein C). This analysis resulted in 851 predicted T11SS/cargo pairs spread across 463 unique bacterial isolates. Homologs can be found throughout Proteobacteria including human pathogens, commensals, and proposed probiotics. Finally, we used phylogenetics to support the hypothesis that genomic acquisition and loss of T11SS genes has occurred through horizontal acquisition during the evolutionary history of *Xenorhabdus* and that these instances correlate with nematode host switching events, supporting a role for these proteins in host adaptation. Considering the cladistic correlation of NilB with bacteria that colonize *Steinernema* from clade 2 reflects the cladistic correlation of the CrpB T11SS with bacteria that colonize *Steinernema* from clade 5, it is possible that both of these acquisitions facilitated host switching. To test this hypothesis I would suggest future researchers delete the *crp* locus, and its adjacent B12 biosynthesis operon, within *Xenorhabdus cabanillasii* to see if either of these features individually or collectively impact the bacteria's ability to colonize its host nematode *Steinernema riobrave*.

Chapter 3 built upon this foundation by using hemophilin homologs, and their cognate T11SS from four different bacterial species, *X. nematophila*, *Haemophilus haemolyticus*, *Acinetobacter baumannii*, and *Xenorhabdus cabanillasii*. By co-expressing different combinations of secretor proteins and cargo proteins within *E. coli*, I demonstrated that T11SS secrete their cognate cargo proteins

with a high level of specificity, having little to no effect on non-cognate cargo secretion. This specificity is interesting because the tested T11SS/cargo pairs are relatively close homologs of each other with similar roles in heme acquisition. Strikingly, specificity was predominantly driven by the C-terminal β-barrel domain present in all known T11SS-dependent cargo. Secretion through a given T11SS could be conveyed to a non-cognate cargo by fusing the C-terminal domain of a cognate cargo to the non-cognate cargo. This combination of highly specific and molecularly programmable secretion has great potential as a molecular biology tool. The ability of T11SS to surface expose membrane anchored surface proteins on Gram-negative bacteria, could be exploited to experimentally surface expose desired antigenic epitopes on a cell surface.

In vitro characterization of the purified hemophilin proteins revealed variable binding affinity for metallated and unmetallated porphyrins that reflected differences in the lifestyles of the bacteria that encoded them. For example, within *H. haemolyticus* hemophilin is used to compete with other heme auxotrophic bacteria within the nasopharynx (Latham et al. 2020), and as such it has a very strong binding affinity for heme. Conversely, *X. nematophila* is capable of heme biosynthesis and may be provisioning heme to its mutualist host, and its hemophilin has a relatively weak binding affinity that shows no preference for metallated porphyrins over unmetallated. As part of chapter 3 we also define the hemophilin ligand binding domain and identify homologs that have diversified in this domain region, presumably evolving novel ligand binding domains while losing the capacity to bind heme. When examined using phylogenetic techniques, a pattern of evolution is revealed in which T11SS-dependent cargo are duplicated within the genome and then evolve into paralogs with distinct ligand binding domains. Within *Xenorhabdus* these evolved homologs (CrpC, HrpC2, HrpC3) all have reduced ligand binding domains that lack residues known to be essential for porphyrin binding in hemophilin from *H. haemolyticus* and *A. baumanii*. While we were unable to identify a high affinity ligand for CrpC, threading models of these reduced domains suggest that they are likely

interacting with a smaller molecule than heme. Observing the genomic context these paralogs does reinforce a connection with metals; CrpC is encoded adjacent to a cobalt porphyrin (cobalamin) uptake and synthesis locus, and HrpC2 is encoded adjacent to molybdopterin-dependent methylotrophic pathways.

I would suggest that future research into hemophilin examine the effects of HrpC and its divergent paralogs HrpC2 and HrpC3 on nematode host health and development. In order to better characterize potential interactions, I have a series of three potential follow up experiments. First, purified holo-HrpC and apo-HrpC from *X. nematophila* could be added to metal/heme starved bacterial cultures and nematode cultures in order to qualify how effective this hemophore is at scavenging and donating heme to diverse organisms. If hemophilin is provisioning heme to developing nematodes we would predict that holo-HrpC would augment nematode growth on limiting mediums while apo-HrpC would decrease nematode growth. Repeating this experiment using bacterial cultures could reveal how effective HrpC is at mediating inter-bacterial competition by isolating available heme. The second experiment could use FLAG-tagged HrpC within *X. nematophila* to track secretion *in vivo*. This would reveal the conditions for hemophilin secretion *in vivo*, which I would hypothesis to be optimal under metal starvation applied via bypyridyl. Additionally, spectroscopic analysis of the resulting supernatants and outer membrane vesicles (OMVs) could be used to determine if secreted hemophilin is predominantly apo-HrpC or holo-HrpC. If the hemophilin is being used to provision heme to the nematode host we would expect to see mostly holo-HrpC, particularly within OMVs where hemophilin would not be able to interact with its co-receptor for use by the bacterium. Alternatively, if the hemophilin is predominantly being used to extract heme from a host derived chelator we would expect to see predominantly apo-HrpC in these fractions. Finally, I would propose an experiment within the Goodrich-Blair labs newest model bacterium *Xenorhabdus griffiniae* (Xg9) to knock out HrpC and HrpC2 separately and in

tandem. *X. griffiniae* colonizes the genetically tractable nematode *S. hermaphroditum,* and this pairing is being developed into a powerful model system for symbiosis research. These hemophilin-like protein mutants would allow us to determine if hemophilin plays similar roles within *Xenorhabdus* species outside of *X. nematophila*, if the divergent paralog HrpC2 has any effect on host nematode colonization, and would facilitate the use of elemental scanning microscopy within *S. hermaphroditum* nematodes to determine how exactly these genes impact the distribution of metal ions within a nematode host. If element detecting microscopy can link the HrpC2 mutant to the distribution of any particular metal or metal cofactor (ie cobalamin or molybdopterins) within the nematode it may help identify the ligand binding function of the entire CrpC/HrpC2 protein family.

In chapter 4, I return to the proteins that inspired this journey: NilB and NilC. Previous research from the Goodrich-Blair lab had suggested that NilC was not located on the cell surface, which was at odds with my hypothesis that NilB was surface exposing the lipoprotein NilC (Cowles and Goodrich-Blair 2004). Using a combination of cell surface dot blots and flow cytometry my co-authors and I demonstrated that the transposition of NilC from the periplasm to the cell surface is a carefully regulated event that could only be observed when expression of NilB and NilC was de-repressed by deletion of their regulatory protein NilR (Cowles and Goodrich-Blair 2006). This careful regulation of surface antigen exposure may help *X. nematophila* adapt to different host organisms, since this bacterium participates in a tripartite association. While surface adherent NilC is required for colonization of their mutualistic host nematodes, it may be detrimental when growing within insect prey. Furthermore, when de-repressed some amount of NilC was capable of reaching the cell surface in the absence of NilB, though it was significantly less than when NilB was present, suggesting that NilB may not be required for surface exposure and instead acts to regulate the timing and intensity of NilC surface exposure. NilC was quite resistant to our attempts at crystallization for

protein structure, but biochemical and structural analysis revealed that it folds stably into two domains and has an intrinsically disordered N-terminal region.

NilC has no homologs in the publically available sequence databases used, though the N-terminal domain has some weak homology to carbohydrate hydrolases leading us to hypothesize that it may be interacting with mucins within the nematode intestine. To better understand the mucoid environment occupied by NilC-expressing *X. nematophila* bacteria, we used a lectin library to characterize the glycans present in several nematode species. We identified a WGA-reactive glycan which was present in nematodes *S. carpocapsae* and *S. scapterisci,* colonized by NilC-encoding bacteria, but not in a nematode, *S. feltiae,* colonized by a bacterium that lacks NilC. WGA is a lectin that binds glycans with a terminal N-acetyl glucosamine or N-acetyl neuraminic acid. Furthermore, WGA lectin was capable of blocking effective *X. nematophila* colonization of the anterior intestinal cecum of *S. carpocapsae*, suggesting that access to this glycan is important for optimal colonization. We hypothesized that if NilC is responsible for binding this glycan we could similarly block colonization by adding purified NilC. Unexpectedly the purified NilC increased colonization by *X. nematophila*. These data led us to a revised working model that NilC enzymatically liberates glycans from either the host mucins or a bacterial biofilm in order to mediate interactions between the two. Future studies may attempt to isolate and identify the substrate bound by NilC using resin embedded with purified WGA and NilC to assay nematode extracts for reactive compounds.

Additionally, in chapter 4 we use multi-omics (metabolomic and proteomic) analysis of a *nil* locus mutant to examine the impact of NilB and NilC on *X. nematophila* physiology. We identified several pathways which were significantly affected; Amino sugar/cell wall biosynthesis and polysaccharide/biofilm biosynthesis was generally downregulated in the absence of the *nil* locus, while glycolysis and LPS biosynthesis were generally upregulated. These results indicated that loss of

the *nil* genes disrupted the cell envelope in a way that altered composition of the cell wall, outer membrane, and potential biofilm layers. The alterations to the outer membrane and extracellular polysaccharides are consistent with our model that surface exposed NilC is interacting with a carbohydrate or glycan moiety, however the alterations to amino sugar/cell wall biosynthesis revealed new possibilities for NilC function. Like the anterior intestinal cecum of *S. carpocapsae*, the bacterial cell wall contains *N*-acetyl glucosamine with which NilC may interact. Exposing purified NilC to purified bacterial cell wall saccules revealed that at high concentrations peptidoglycan bound NilC. Since NilC is predominantly located within the periplasm when not de-repressed, we extended our model to include the idea that periplasmic NilC may bind the cell wall in circumstances where it has not been translocated across the outer membrane by NilB. A cell wall interaction may also explain why NilC that had been mutated to be unlipidated is retained within the cell instead of being secreted like other soluble T11SS-dependent cargo. I suggest that future research on the function of NilC within *X. nematophila* begin with mass spectrometry-based co-immunoprecipitation using purified NilC and purified WGA. By embedding these proteins into resin beads and placing them in columns we could look for glycan binding interactions by passing nematode and insect homogenates over the columns to isolate the retained sugar moieties and glycoproteins. This analysis could be further refined using a library of glycosylase enzymes to experimentally abrogate potential substrate binding.

Finally, in chapter 5 we develop a novel protocol for bioinformatically controlling genome neighborhood co-occurrence analysis and use this protocol to explore two distinct clusters of T11SS. Analyzing the largest cluster of T11SS, which are predominantly encoded by host-associated bacteria, revealed broadly conserved association with cofactor metabolism. Porphyrin metabolism and heme oxygenase were the most common reflecting systems for incorporation of host derived iron/heme. There was also a significant co-occurrence with folate biosynthesis and several folate

dependent processes like one carbon metabolism, and nucleotide biosynthesis. Further strengthening the connection to cofactor metabolism, folate biosynthesis and several folate dependent pathways are dependent on metallated cofactors such as the cobalt carrying cobalamin molecule (vitamin B12) and the molybdenum carrying molybdopterin (MoCo) (Mendel 2013). Additionally, the genomic neighborhoods containing these T11SS genes often encoded genes common to mobile genetic elements, suggesting that T11SS are regularly moved on mobile genetic elements such as phage and transposons (Dobrindt and Reidl 2000). This co-occurrence supports phylogenetic observations in chapter 2 and 3 which established horizontal T11SS acquisitions by *Xenorhabdus* species.

Mining our data set of genes that significantly co-occur with T11SS from the animal-associated clusters revealed 141 distinct UniRef50 clusters predicted to be T11SS dependent cargo (Suzek et al. 2007). These predicted cargo proteins fell into 10 different general architectures distinguished by their unique N-terminal domains. Several of the architectures were well studied representatives like Transferrin binding protein B, lactoferrin binding protein B, factor H binding protein, and hemophilin, however several of the others were entirely new. Three novel cargo architectures featured elongated disordered N-terminal domains seemingly optimized for flexible surface lipoproteins. One novel cargo architecture had elongated α-helical N-terminal domains composed of highly polar EEARKA repeat motifs homologous to the Plasmin surface protein (Pls) from *Staphylococcus aureus*. Since these Pls-like cargo proteins are predicted to be lipid anchored to the membrane, their T11SS may be essential for exposure of surface glycoproteins. Additionally, in *Staphylococcus* Pls acts as a virulence factor by modulating biofilm formation and binding host surfaces, so these Gram negative Pls homologs may be another T11SS-dependent symbiosis factor (Bleiziffer et al. 2017; Savolainen et al. 2001; Zhou and Wu 2009). To validate our cargo predictions, we chose a representative Pls homolog and its cognate type eleven secretor from *Haemophilus*

parahaemolyticus and demonstrated secretion when expressed in *E. coli*. Future research may map these 10 cargo architecture back onto the original cluster 1 T11SS sequence similarity network described within chapter 2 and utilize Markov clustering analyses such as the Fisher's exact test to statistically demonstrate correlations between cargo type and subcluster formation.

This same co-occurrence analysis was used on a second T11SS cluster predominantly encoded by marine bacteria, both pelagic and invertebrate-associated. No co-occurrence with porphyrin or heme metabolism genes was detected this time, indicating that this pathway was not participating in iron acquisition for these organisms. However, the co-occurrence with folate biosynthesis, and folate dependent pathways such as one carbon metabolism was still present. Additionally, the most common co-occurring pathway was the glyoxalase detoxification system. This system uses metalloenzymes (predominantly Ni^{2+} dependent) to detoxify metabolically derived aldehydes, potentially indicating a role for these T11SS in an aldehyde producing metabolic pathway (Anaya-Sanchez et al. 2021; C. Lee and Park 2017). Notably, no significantly co-occurring genes of the marine T11SS cluster possessed the small hydrophilic β-barrel domains ubiquitous to cargo proteins for the main animal associated T11SS. Indicating that these T11SS either seldom co-occur with their cargo, or their cargo proteins are very different from those previously characterized. In place of the missing cargo, the most common co-occurrence was with DUF1194 proteins. This domain is uncharacterized to date and is predicted to encode a von Willebrand factor like Rossman fold. These folds are typically associated with glycoproteins and assembly of large multimeric complexes (Peyvandi, Garagiola, and Baronciani 2011). These DUF1194 proteins are predicted to include both lipidated and soluble representative and often times a single T11SS loci will contain two or three DUF1194 genes, similarly to known TbpBBD containing cargo proteins. Further studies are needed to experimentally determine if these represent a new family of T11SS or if they play some other role mechanistic role.

Future research into bioinformatically predicted cargo proteins has many directions it could potentially expand into. First, I would suggest utilizing a strain of *Haemophilus parahaemolyticus* that encodes the Pls lipoprotein to generate a Pls mutant. These strains could then be used to determine if Pls is playing a role in biofilm formation (via crystal violet assay) or in oropharyngeal colonization potential (within a murine model system) similar to Pls proteins from Gram-positive *Staphylococcus* species. These strains would also be helpful for determining if the T11SS TepS is responsible for cleaving Pls off of the cell membrane to solubilize it, or if that phenotype was a side effect of exogenous expression within *E. coli*. Similar experiments could be undertaken for the *Psychrobacter* surface proteins and *Sphingomonas* surface proteins in order to learn more about their potential *in vivo* function. Regarding the DUF1194 co-occurrence identified with T11SS from cluster 3, I would suggest using gene synthesis to FLAG-tag and clone the T11SS from *Ascidiaceihabitans donghaensis* (SPH20591) and the DUF1194 protein (SPH20589) into pETDuet-1, while tagging and cloning the DUF1194 proteins (SPH20588) and (SPH20587) into pCOLADuet-1. This would allow simultaneous expression of all four proteins within *E. coli* Bl21 C43. Monitoring localization of these proteins would reveal if they are being exported from the cell in a T11SS-dependent manner and would also allow us to look for multimeric complexes forming between the DUF1194 proteins. Application of short and medium length crosslinking molecules like DSS may be necessary to isolate multimeric complexes intact. I would hypothesize that these proteins form long multimeric linkers similar to the Rossman fold proteins used as adhesins by the surface protein Clumping Factor A within *S. aureus* (Peyvandi, Garagiola, and Baronciani 2011).

Synthesizing all of the finding described here, we have established a diverse Proteobacterial secretion system used to surface expose or export proteins that can be used to overcome nutritional immunity and humoral immunity via ligand binding, or to interact with host surfaces/mucins via carbohydrate-protein binding or protein-protein binding. These cargo proteins can be post-

translationally modified to expand their functional range even further (glycosylation and or cleavage) and their secretion can be tightly regulated based on prevailing environmental conditions. Perhaps most importantly, T11SS secretion is highly specific and that specificity can be re-constructed using structural motifs. The field of bacterial surface lipoprotein study is relatively new (Kinkead et al. 2018; Konovalova and Silhavy 2015) and having a molecular biology technique available to engineer surface lipoproteins may be invaluable for studying these proteins and general synthetic biology. Multiple lines of evidence suggest that T11SS are regularly acquired horizontally and, in that sense, they may behave like a "pathogenicity island" by allowing an organism to rapidly adapt to a host-organism. However, it is inaccurate to call these true pathogenicity islands since T11SS-dependent cargo are not unique to pathogens. Within *X. nematophila* the horizontal acquisition of the *nil* locus most likely allowed the bacterium to expand its host range into a new mutualistic host. Additionally, hemophilin homologs appear to play distinctive functional roles, depending on the expressing organisms. Within the human commensal *H. haemolyticus,* hemophilin mediates inter-bacterial competition by excluding *H. influenzae* through metal sequestration (Atto et al. 2020). In the human pathogen *A. baumanii,* hemophilin, which has a distinctive pattern of heme coordination with stronger heme binding affinity compared to *H. haemolyticus* hemophilin, contributes to pathogenesis in a murine model, presumably by virtue of its ability to extract metal from host chaperones (Bateman et al. 2021). Finally, in *X. nematophila* hemophilin is predicted to have an *A. baumanii* hemophilin-like ligand coordination pattern but with weaker affinity, did not impact virulence in an insect host and did not detectably slow competitor bacterial growth. Instead, *X. nematophila* hemophilin appeared to augment the growth of conspecific bacteria and supported host-animal fitness in vitro. Our findings demonstrated that hemophilin from *X. nematophila* had nearly equivalent affinity for metallated and unmetallated porphyrins, suggesting that this protein may be acting as a chaperon and exchanging heme with higher affinity hemoproteins. Future studies will

attempt to determine if the fitness impact of T11SS-dependent hemophilin export is a result of bacterial provisioning of heme to their mutualistic host organisms. On top of the diversity of T11SS-dependent cargo already described, our studies suggest that there are likely many more to discover. Intriguing candidates for experimental validation include the DUF1194 protein family. Since all previously discovered T11SS-dependent cargo have some role in host-microbe association I hypothesize that identifying and characterizing novel T11SS/cargo pairs will continue to provide valuable insights into the molecular foundations of symbiosis. In this way, something as small as the nanomachinery of a secretion system can inform hypotheses about biology at the far larger scales of ecology and microbiome assembly.

Bibliography:

Aballay, Alejandro, P Yorgey, and Frederick M Ausubel. 2000. "Salmonella Typhimurium Proliferates and Establishes a Persistent Infection in the Intestine of Caenorhabditis Elegans." *Current biology : CB* 10(23): 1539–42.

Adams, P A, and M C Berman. 1980. "Kinetics and Mechanism of the Interaction between Human Serum Albumin and Monomeric Haemin." *Biochemical Journal* 191(1): 95–102. https://doi.org/10.1042/bj1910095.

Akhurst, R J. 1983. "Neoaplectana Species: Specificity of Association with Bacteria of the Genus Xenorhabdus." *Experimental Parasitology* 55(2): 258–63. http://dx.doi.org/10.1016/0014-4894(83)90020-6.

Alamos, Pamela et al. 2018. "Functionality of TRNAs Encoded in a Mobile Genetic Element from an Acidophilic Bacterium." *RNA biology* 15(4–5): 518–27.

Alexeyev, M F, I N Shokolenko, and T P Croughan. 1995. "New Mini-Tn5 Derivatives for Insertion Mutagenesis and Genetic Engineering in Gram-Negative Bacteria." *Canadian journal of microbiology* 41(11): 1053–55.

Almagro Armenteros, José Juan et al. 2019. "SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks." *Nature Biotechnology* 37(4): 420–23. http://dx.doi.org/10.1038/s41587-019-0036-z.

Altekar, G, S Dwarkadas, John P Huelsenbeck, and F Ronquist. 2004. "Parallel Metropolis Coupled Markov Chain Monte Carlo for Bayesian Phylogenetic Inference." *Bioinformatics* 20(3): 407–15. http://dx.doi.org/10.1093/bioinformatics/btg427.

Altschul, Stephen F et al. 1990. "Basic Local Alignment Search Tool." *Journal of molecular biology* 215(3): 403–10.

———. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25(17): 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

Anaya-Sanchez, Andrea, Ying Feng, John C Berude, and Daniel A Portnoy. 2021. "Detoxification of Methylglyoxal by the Glyoxalase System Is Required for Glutathione Availability and Virulence Activation in Listeria Monocytogenes." *PLoS pathogens* 17(8): e1009819.

Anyanful, Akwasi, Kirk A Easley, Guy M Benian, and Daniel Kalman. 2009. "Conditioning Protects C. Elegans from Lethal Effects of Enteropathogenic E. Coli by Activating Genes That Regulate Lifespan and Innate Immunity." *Cell host & microbe* 5(5): 450–62.

Aramaki, Takuya et al. 2020. "KofamKOALA: KEGG Ortholog Assignment Based on Profile HMM and Adaptive Score Threshold." *Bioinformatics* 36(7): 2251–52. https://doi.org/10.1093/bioinformatics/btz859.

Arnoux, Pascal et al. 2000. "Functional Aspects of the Heme Bound Hemophore HasA by Structural Analysis of Various Crystal Forms." *Proteins: Structure, Function, and Bioinformatics* 41(2): 202–10. https://doi.org/10.1002/1097-0134(20001101)41:2%3C202::AID-PROT50%3E3.0.CO.

Ascoli, Franca, Maria Rosaria Rossi Fanelli, and Eraldo B T - Methods in Enzymology Antonini.

1981. "Preparation and Properties of Apohemoglobin and Reconstituted Hemoglobins." In *Hemoglobins*, Academic Press, 72–87. https://www.sciencedirect.com/science/article/pii/0076687981761159.

Atto, Brianna et al. 2020. "In Vitro Probiotic Potential of Hemophilin-Producing Strains of Haemophilus Haemolyticus." *bioRxiv*: 2020.01.02.893487. http://biorxiv.org/content/early/2020/01/03/2020.01.02.893487.abstract.

Ayres, Daniel L et al. 2012. "BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics." *Systematic Biology* 61(1): 170–73. https://doi.org/10.1093/sysbio/syr100.

Bacic, Antony, I Kahane, and B M Zuckerman. 1990. "Panagrellus Redivivus and Caenorhabditis Elegans: Evidence for the Absence of Sialic Acids." *Experimental parasitology* 71(4): 483–88.

Baek, Minkyung et al. 2021. "Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network." *Science (New York, N.Y.)* 373(6557): 871–76.

Bahadur, Raj, Pavan Kumar Chodisetti, and Manjula Reddy. 2021. "Cleavage of Braun's Lipoprotein Lpp from the Bacterial Peptidoglycan by a Paralog of l,d-Transpeptidases, LdtF." *Proceedings of the National Academy of Sciences of the United States of America* 118(19).

Bai, Xiaodong et al. 2013. "A Lover and a Fighter: The Genome Sequence of an Entomopathogenic Nematode Heterorhabditis Bacteriophora." *PloS one* 8(7): e69618.

Banno, Masaki et al. 2017. "Development of a Sugar-Binding Residue Prediction System from Protein Sequences Using Support Vector Machine." *Computational biology and chemistry* 66: 36–43.

Barbercheck, Mary. 2008. "Field Manual of Techniques in Invertebrate Pathology: Application and Evaluation of Pathogens for Control of Insects and Other Invertebrate Pests, 2nd Edn - Edited by L. Lacey & H. K. Kaya." *Entomologia Experimentalis et Applicata* 129(1): 115. https://doi.org/10.1111/j.1570-7458.2008.00751.x.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. "Gephi: An Open Source Software for Exploring and Manipulating Networks." *Proceedings of the International AAAI Conference on Web and Social Media* 3(1 SE-Demonstration Papers): 361–62. https://ojs.aaai.org/index.php/ICWSM/article/view/13937.

Bateman, Thomas J et al. 2021. "A Slam-Dependent Hemophore Contributes to Heme Acquisition in the Bacterial Pathogen Acinetobacter Baumannii." *Nature Communications* 12(1): 6270. https://doi.org/10.1038/s41467-021-26545-9.

Bendtsen, Jannick D, Lars Kiemer, Anders Fausbøll, and Søren Brunak. 2005. "Non-Classical Protein Secretion in Bacteria." *BMC Microbiology* 5(1): 58. https://doi.org/10.1186/1471-2180-5-58.

Berg, Maureen et al. 2016. "Assembly of the Caenorhabditis Elegans Gut Microbiota from Diverse Soil Microbial Environments." *The ISME Journal* 10(8): 1998–2009. https://doi.org/10.1038/ismej.2015.253.

———. 2019. "TGFβ/BMP Immune Signaling Affects Abundance and Function of C. Elegans Gut Commensals." *Nature Communications* 10(1): 604. https://doi.org/10.1038/s41467-019-08379-8.

van den Berg, Susanne, Per-Åke Löfdahl, Torleif Härd, and Helena Berglund. 2006. "Improved Solubility of TEV Protease by Directed Evolution." *Journal of Biotechnology* 121(3): 291–98. https://www.sciencedirect.com/science/article/pii/S0168165605004840.

Berndt, Volker et al. 2019. "Metabolome and Transcriptome-Wide Effects of the Carbon Storage Regulator A in Enteropathogenic Escherichia Coli." *Scientific reports* 9(1): 138.

Bertini, I., Gray, H.B., Stiefel, E.I. and Selverstone Valentine, J. 2007. *Biological Inorganic Chemistry: Structure and Reactivity*. Sausalito: University Science Books.

Bhasin, Archna, John M Chaston, and Heidi Goodrich-Blair. 2012. "Mutational Analyses Reveal Overall Topology and Functional Regions of NilB, a Bacterial Outer Membrane Protein Required for Host Association in a Model of Animal-Microbe Mutualism." *Journal of Bacteriology* 194(7): 1763–76. https://doi.org/10.1128/JB.06711-11.

Van Bibber, M, C Bradbeer, N Clark, and J R Roth. 1999. "A New Class of Cobalamin Transport Mutants (BtuF) Provides Genetic Evidence for a Periplasmic Binding Protein in Salmonella Typhimurium." *Journal of bacteriology* 181(17): 5539–41.

Bird, A F, and R J Akhurst. 1983. "The Nature of the Intestinal Vesicle in Nematodes of the Family Steinernematidae." *International Journal for Parasitology* 13(6): 599–606. http://dx.doi.org/10.1016/s0020-7519(83)80032-0.

Bito, Tomohiro et al. 2013. "Vitamin B12 Deficiency in Caenorhabditis Elegans Results in Loss of Fertility, Extended Life Cycle, and Reduced Lifespan." *FEBS open bio* 3: 112–17. https://pubmed.ncbi.nlm.nih.gov/23772381.

Blatch, Gregory L, and Michael Lässle. 1999. "The Tetratricopeptide Repeat: A Structural Motif Mediating Protein-Protein Interactions." *BioEssays* 21(11): 932–39. http://dx.doi.org/10.1002/(sici)1521-1878(199911)21:11%3C932::aid-bies5%3E3.0.co.

Bleiziffer, Isabelle et al. 2017. "The Plasmin-Sensitive Protein Pls in Methicillin-Resistant Staphylococcus Aureus (MRSA) Is a Glycoprotein." *PLOS Pathogens* 13(1): e1006110. https://doi.org/10.1371/journal.ppat.1006110.

Boes, Adrien, Samir Olatunji, Eefjan Breukink, and Mohammed Terrak. 2019. "Regulation of the Peptidoglycan Polymerase Activity of PBP1b by Antagonist Actions of the Core Divisome Proteins FtsBLQ and FtsN." *mBio* 10(1).

Bonifassi, E et al. 1999. "Gnotobiological Study of Infective Juveniles and Symbionts of Steinernema Scapterisci: A Model to Clarify the Concept of the Natural Occurrence of Monoxenic Associations in Entomopathogenic Nematodes." *Journal of invertebrate pathology* 74(2): 164–72.

Boraston, Alisdair B et al. 2002. "Differential Oligosaccharide Recognition by Evolutionarily-Related Beta-1,4 and Beta-1,3 Glucan-Binding Modules." *Journal of molecular biology* 319(5): 1143–56.

Borg, Annika J E, Alexander Dennig, Hansjörg Weber, and Bernd Nidetzky. 2021. "Mechanistic Characterization of UDP-Glucuronic Acid 4-Epimerase." *The FEBS journal* 288(4): 1163–78.

Borths, Elizabeth L, Kaspar P Locher, Allen T Lee, and Douglas C Rees. 2002. "The Structure of Escherichia Coli BtuF and Binding to Its Cognate ATP Binding Cassette Transporter." *Proceedings of the National Academy of Sciences* 99(26): 16642–47.

https://doi.org/10.1073/pnas.262659699.

Bouchery, T et al. 2013. "The Symbiotic Role of Wolbachia in Onchocercidae and Its Impact on Filariasis." *Clinical Microbiology and Infection* 19(2): 131–40. https://www.sciencedirect.com/science/article/pii/S1198743X14602482.

Bovian, Prosper. 1938. "Association between Nematodes and Insects." *Nature* 141(3573): 758. https://doi.org/10.1038/141758b0.

Brandt, Julia P, and Niels Ringstad. 2015. "Toll-like Receptor Signaling Promotes Development and Function of Sensory Neurons Required for a C. Elegans Pathogen-Avoidance Behavior." *Current biology : CB* 25(17): 2228–37.

Brenner, Sydney. 1974. "The Genetics of Caenorhabditis Elegans." *Genetics* 77(1): 71–94.

———. 2009. "In the Beginning Was the Worm …." *Genetics* 182(2): 413–15. https://doi.org/10.1534/genetics.109.104976.

Brenner, Sydney, F Jacob, and M Meselson. 1961. "An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis." *Nature* 190: 576–81.

Bright, Monika, and Silvia Bulgheresi. 2010. "A Complex Journey: Transmission of Microbial Symbionts." *Nature reviews. Microbiology* 8(3): 218–30.

Buchan, Daniel W A, and David T Jones. 2019. "The PSIPRED Protein Analysis Workbench: 20 Years On." *Nucleic acids research* 47(W1): W402–7.

Buecher, E J, and I Popiel. 1989. "Liquid Culture of the Entomogenous Nematode Steinernema Feltiae with Its Bacterial Symbiont." *Journal of nematology* 21(4): 500–504.

Burmester, Thorsten, and Thomas Hankeln. 2007. "The Respiratory Proteins of Insects." *Journal of insect physiology* 53(4): 285–94.

Butland, Gareth et al. 2005. "Interaction Network Containing Conserved and Essential Protein Complexes in Escherichia Coli." *Nature* 433(7025): 531–37.

Calder, Joshua T, Nicholas D Christman, Jessica M Hawkins, and David L Erickson. 2020. "A Trimeric Autotransporter Enhances Biofilm Cohesiveness in Yersinia Pseudotuberculosis but Not in Yersinia Pestis." *Journal of bacteriology* 202(20).

Calmettes, Charles et al. 2012. "The Structural Basis of Transferrin Sequestration by Transferrin-Binding Protein B." *Nature Structural & Molecular Biology* 19(3): 358–60. https://doi.org/10.1038/nsmb.2251.

Cantí, C et al. 2005. "The Metal-Ion-Dependent Adhesion Site in the Von Willebrand Factor-A Domain of Alpha2delta Subunits Is Key to Trafficking Voltage-Gated Ca2+ Channels." *Proceedings of the National Academy of Sciences of the United States of America* 102(32): 11230–35.

Cao, Mengyi et al. 2017. "High Levels of the Xenorhabdus Nematophila Transcription Factor Lrp Promote Mutualism with the Steinernema Carpocapsae Nematode Host." *Applied and environmental microbiology* 83(12).

Cao, Yurou et al. 2020. "Genome-Wide Screening of Lipoproteins in Actinobacillus Pleuropneumoniae Identifies Three Antigens That Confer Protection against Virulent

Challenge." *Scientific reports* 10(1): 2343.

Capella-Gutiérrez, Salvador, José M Silla-Martínez, and Toni Gabaldón. 2009. "TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics (Oxford, England)* 25(15): 1972–73.

Castelletto, Michelle L, Spencer S Gang, and Elissa A Hallem. 2020. "Recent Advances in Functional Genomics for Parasitic Nematodes of Mammals." *The Journal of experimental biology* 223(Pt Suppl 1).

Chang, Dennis Z et al. 2019. "A Core Set of Venom Proteins Is Released by Entomopathogenic Nematodes in the Genus Steinernema." *PLOS Pathogens* 15(5): e1007626. https://doi.org/10.1371/journal.ppat.1007626.

Chapuis, E et al. 2009. "Manifold Aspects of Specificity in a Nematode-Bacterium Mutualism." *Journal of evolutionary biology* 22(10): 2104–17.

Chaston, John M et al. 2011. "The Entomopathogenic Bacterial Endosymbionts Xenorhabdus and Photorhabdus: Convergent Lifestyles from Divergent Genomes." *PLoS ONE* 6(11): e27909. http://dx.doi.org/10.1371/journal.pone.0027909.

Chaston, John M, and Heidi Goodrich-Blair. 2010. "Common Trends in Mutualism Revealed by Model Associations between Invertebrates and Bacteria." *FEMS microbiology reviews* 34(1): 41–58.

Chaston, John M, Kristen E Murfin, Elizabeth A Heath-Heckman, and Heidi Goodrich-Blair. 2013. "Previously Unrecognized Stages of Species-Specific Colonization in the Mutualism between Xenorhabdus Bacteria and Steinernema Nematodes." *Cellular Microbiology* 15(9): 1545–59. https://doi.org/10.1111/cmi.12134.

Chavez-Dozal, Alba A, Clayton Gorman, C Phoebe Lostroh, and Michele K Nishiguchi. 2014. "Gene-Swapping Mediates Host Specificity among Symbiotic Bacteria in a Beneficial Symbiosis." *PLOS ONE* 9(7): e101691. https://doi.org/10.1371/journal.pone.0101691.

Chen, C J et al. 1996. "Identification and Purification of a Hemoglobin-Binding Outer Membrane Protein from Neisseria Gonorrhoeae." *Infection and immunity* 64(12): 5008–14. https://pubmed.ncbi.nlm.nih.gov/8945539.

Chen, Ming-Hui et al. 2012. "Roseivivax Isoporae Sp. Nov., Isolated from a Reef-Building Coral, and Emended Description of the Genus Roseivivax." *International journal of systematic and evolutionary microbiology* 62(Pt 6): 1259–64.

Choe, Andrea et al. 2012. "Ascaroside Signaling Is Widely Conserved among Nematodes." *Current biology : CB* 22(9): 772–80.

Choi, Kyoung-Hee et al. 2005. "A Tn7-Based Broad-Range Bacterial Cloning and Expression System." *Nature methods* 2(6): 443–48.

Choi, Kyoung-Hee, and Herbert P Schweizer. 2006. "Mini-Tn7 Insertion in Bacteria with Secondary, Non-GlmS-Linked AttTn7 Sites: Example Proteus Mirabilis HI4320." *Nature protocols* 1(1): 170–78.

Chomicki, Guillaume, E Toby Kiers, and Susanne S Renner. 2020. "The Evolution of Mutualistic

Dependence." *Annual Review of Ecology, Evolution, and Systematics* 51(1): 409–32. https://doi.org/10.1146/annurev-ecolsys-110218-024629.

Chu, Byron C et al. 2010. "Siderophore Uptake in Bacteria and the Battle for Iron with the Host; a Bird's Eye View." *BioMetals* 23(4): 601–11. https://doi.org/10.1007/s10534-010-9361-x.

Ciche, Todd A et al. 2008. "Cell Invasion and Matricide during Photorhabdus Luminescens Transmission by Heterorhabditis Bacteriophora Nematodes." *Applied and environmental microbiology* 74(8): 2275–87.

Ciche, Todd A, S B Bintrim, A R Horswill, and J C Ensign. 2001. "A Phosphopantetheinyl Transferase Homolog Is Essential for Photorhabdus Luminescens to Support Growth and Reproduction of the Entomopathogenic Nematode Heterorhabditis Bacteriophora." *Journal of bacteriology* 183(10): 3117–26.

Ciche, Todd A, and Paul W Sternberg. 2007. "Postembryonic RNAi in Heterorhabditis Bacteriophora: A Nematode Insect Parasite and Host for Insect Pathogenic Symbionts." *BMC developmental biology* 7: 101.

Ciezki, Kristin et al. 2017. "R-Type Bacteriocins in Related Strains of Xenorhabdus Bovienii: Xenorhabdicin Tail Fiber Modularity and Contribution to Competitiveness." *FEMS Microbiology Letters* 364(1): fnw235. https://doi.org/10.1093/femsle/fnw235.

Clarke, David J. 2020. "Photorhabdus: A Tale of Contrasting Interactions." *Microbiology (Reading, England)* 166(4): 335–48.

Cock, P J A et al. 2009. "Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics." *Bioinformatics* 25(11): 1422–23. http://dx.doi.org/10.1093/bioinformatics/btp163.

Cohen, Jon, William Powderly, and Steven Opal. 2017. "Preface to the Fourth Edition." *Infectious Diseases*: xiv. http://dx.doi.org/10.1016/b978-0-7020-6285-8.00276-8.

Collins, Kevin M et al. 2016. "Activity of the C. Elegans Egg-Laying Behavior Circuit Is Controlled by Competing Activation and Feedback Inhibition." *eLife* 5.

Constable, Edwin C, and Catherine E Housecroft. 2019. "The Early Years of 2,2'-Bipyridine-A Ligand in Its Own Lifetime." *Molecules (Basel, Switzerland)* 24(21): 3951. https://pubmed.ncbi.nlm.nih.gov/31683694.

Cook, Daniel E, Stefan Zdraljevic, Joshua P Roberts, and Erik C Andersen. 2017. "CeNDR, the Caenorhabditis Elegans Natural Diversity Resource." *Nucleic acids research* 45(D1): D650–57.

Coolon, Joseph D et al. 2009. "Caenorhabditis Elegans Genomic Response to Soil Bacteria Predicts Environment-Specific Genetic Effects on Life History Traits." *PLoS genetics* 5(6): e1000503.

Couillault, Carole, and Jonathan J Ewbank. 2002. "Diverse Bacteria Are Pathogens of Caenorhabditis Elegans." *Infection and immunity* 70(8): 4705–7.

Cowles, Charles E, and Heidi Goodrich-Blair. 2004. "Characterization of a Lipoprotein, NilC, Required by Xenorhabdus Nematophila for Mutualism with Its Nematode Host." *Molecular Microbiology* 54(2): 464–77. http://dx.doi.org/10.1111/j.1365-2958.2004.04271.x.

———. 2006. "NilR Is Necessary for Co-Ordinate Repression of Xenorhabdus Nematophila

Mutualism Genes." *Molecular microbiology* 62(3): 760–71.

———. 2008. "The Xenorhabdus Nematophila NilABC Genes Confer the Ability of Xenorhabdus Spp. To Colonize Steinernema Carpocapsae Nematodes." *Journal of Bacteriology* 190(12): 4121–28. https://doi.org/10.1128/JB.00123-08.

Craig, Nancy L. 1996. "Transposon Tn7." *Current topics in microbiology and immunology* 204: 27–48.

Crenn, Klervi et al. 2016. "Silicimonas Algicola Gen. Nov., Sp. Nov., a Member of the Roseobacter Clade Isolated from the Cell Surface of the Marine Diatom Thalassiosira Delicatula." *International journal of systematic and evolutionary microbiology* 66(11): 4580–88.

Crick, F H, L Barnett, Sydney Brenner, and R J Watts-Tobin. 1961. "General Nature of the Genetic Code for Proteins." *Nature* 192: 1227–32.

D'Enfert, C, C Chapon, and A P Pugsley. 1987. "Export and Secretion of the Lipoprotein Pullulanase by Klebsiella Pneumoniae." *Molecular Microbiology* 1(3): 107–16. https://doi.org/10.1111/j.1365-2958.1987.tb00534.x.

Darriba, Diego, Guillermo L Taboada, Ramón Doallo, and David Posada. 2012. "JModelTest 2: More Models, New Heuristics and Parallel Computing." *Nature Methods* 9(8): 772. http://dx.doi.org/10.1038/nmeth.2109.

Dawadi, Surendra, Shannon L Kordus, Anthony D Baughn, and Courtney C Aldrich. 2017. "Synthesis and Analysis of Bacterial Folate Metabolism Intermediates and Antifolates." *Organic Letters* 19(19): 5220–23. https://doi.org/10.1021/acs.orglett.7b02487.

Dearth, Stephen P et al. 2018. "Metabolome Changes Are Induced in the Arbuscular Mycorrhizal Fungus Gigaspora Margarita by Germination and by Its Bacterial Endosymbiont." *Mycorrhiza* 28(5–6): 421–33.

Dent, Alecia T, Susana Mouriño, Weiliang Huang, and Angela Wilks. 2019. "Post-Transcriptional Regulation of the Pseudomonas Aeruginosa Heme Assimilation System (Has) Fine-Tunes Extracellular Heme Sensing." *Journal of Biological Chemistry* 294(8): 2771–5555. https://doi.org/10.1074/jbc.RA118.006185.

Denver, Dee R et al. 2016. "Genome Skimming: A Rapid Approach to Gaining Diverse Biological Insights into Multicellular Pathogens." *PLOS Pathogens* 12(8): e1005713. https://doi.org/10.1371/journal.ppat.1005713.

Dickinson, Daniel J, and Bob Goldstein. 2016. "CRISPR-Based Methods for Caenorhabditis Elegans Genome Engineering." *Genetics* 202(3): 885–901.

Dillman, Adler R et al. 2012. "An Entomopathogenic Nematode by Any Other Name." *PLoS pathogens* 8(3): e1002527.

———. 2015. "Comparative Genomics of Steinernema Reveals Deeply Conserved Gene Regulatory Networks." *Genome Biology* 16(1). http://dx.doi.org/10.1186/s13059-015-0746-6.

Diogo, Jesica, and Ana Bratanich. 2014. "The Nematode Caenorhabditis Elegans as a Model to Study Viruses." *Archives of virology* 159(11): 2843–51.

Dirksen, Philipp et al. 2016. "The Native Microbiome of the Nematode Caenorhabditis Elegans: Gateway to a New Host-Microbiome Model." *BMC Biology* 14(1): 38.

https://doi.org/10.1186/s12915-016-0258-1.

———. 2020. "CeMbio - The Caenorhabditis Elegans Microbiome Resource." *G3 Genes|Genomes|Genetics* 10(9): 3025–39. https://doi.org/10.1534/g3.120.401309.

Dobrindt, Ulrich, and Joachim Reidl. 2000. "Pathogenicity Islands and Phage Conversion: Evolutionary Aspects of Bacterial Pathogenesis." *International Journal of Medical Microbiology* 290(6): 519–27. https://www.sciencedirect.com/science/article/pii/S143842210080017X.

Dobrogosz, W J. 1968. "Effect of Amino Sugars on Catabolite Repression in Escherichia Coli." *Journal of bacteriology* 95(2): 578–84.

Douglas, Angela E. 2010. *The Symbiotic Habit*. Princeton University Press. https://books.google.com/books?id=gaAG1I50OkAC.

Drace, Kevin, and Creg Darby. 2008. "The HmsHFRS Operon of Xenorhabdus Nematophila Is Required for Biofilm Attachment to Caenorhabditis Elegans." *Applied and environmental microbiology* 74(14): 4509–15.

Dreyer, Jönike, Antoinette P Malan, and Leon M T Dicks. 2018. "Bacteria of the Genus Xenorhabdus, a Novel Source of Bioactive Compounds." *Frontiers in Microbiology* 9. http://dx.doi.org/10.3389/fmicb.2018.03177.

Duchaud, Eric et al. 2003. "The Genome Sequence of the Entomopathogenic Bacterium Photorhabdus Luminescens." *Nature Biotechnology* 21(11): 1307–13. https://doi.org/10.1038/nbt886.

Dumon-Seignovert, Laurence, Guillaume Cariot, and Laurent Vuillard. 2004. "The Toxicity of Recombinant Proteins in Escherichia Coli: A Comparison of Overexpression in BL21(DE3), C41(DE3), and C43(DE3)." *Protein expression and purification* 37(1): 203–6.

Dutky, Samson R. 1959. "Insect Microbiology." *Advances in applied microbiology* 1: 175–200.

Dutky, Samson R, and Walter S Hough. 1955. "Note on a Parasitic Nematode from Codling Moth Larvae, Carpocapsa Pamonetta (Lepidoptera, Olethreutidae)."

Easom, Catherine A, Susan A Joyce, and David J Clarke. 2010. "Identification of Genes Involved in the Mutualistic Colonization of the Nematode Heterorhabditis Bacteriophora by the Bacterium Photorhabdus Luminescens." *BMC microbiology* 10: 45.

Ebert, Dieter. 2013. "The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission." *Annual Review of Ecology, Evolution, and Systematics* 44(1): 623–43. https://doi.org/10.1146/annurev-ecolsys-032513-100555.

Edgar, R C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research* 32(5): 1792–97. http://dx.doi.org/10.1093/nar/gkh340.

Egler, Monique, Cornelia Grosse, Gregor Grass, and Dietrich H. Nies. 2005. "Role of the Extracytoplasmic Function Protein Family Sigma Factor RpoE in Metal Resistance of Escherichia Coli." *Journal of Bacteriology* 187(7): 2297–2307. https://doi.org/10.1128/JB.187.7.2297-2307.2005.

Ehlers, Ralf-Udo. 2001. "Mass Production of Entomopathogenic Nematodes for Plant Protection." *Applied microbiology and biotechnology* 56(5–6): 623–33.

Einhauer, A, and A Jungbauer. 2001. "The FLAG Peptide, a Versatile Fusion Tag for the Purification of Recombinant Proteins." *Journal of biochemical and biophysical methods* 49(1–3): 455—465. https://doi.org/10.1016/s0165-022x(01)00213-5.

Eisenbeis, Simone et al. 2008. "NagA-Dependent Uptake of N-Acetyl-Glucosamine and N-Acetyl-Chitin Oligosaccharides across the Outer Membrane of Caulobacter Crescentus." *Journal of bacteriology* 190(15): 5230–38.

El-Danasoury, Heba, and Javier Iglesias-Piñeiro. 2017. "Performance of the Slug Parasitic Nematode Phasmarhabditis Hermaphrodita under Predicted Conditions of Winter Warming." *Journal of pesticide science* 42(2): 62–66.

El-Gebali, Sara et al. 2019. "The Pfam Protein Families Database in 2019." *Nucleic Acids Research* 47(D1): D427–32. https://doi.org/10.1093/nar/gky995.

Ellis, H M, and H R Horvitz. 1986. "Genetic Control of Programmed Cell Death in the Nematode C. Elegans." *Cell* 44(6): 817–29.

Engel, Yvonne et al. 2017. "The Global Regulators Lrp, LeuO, and HexA Control Secondary Metabolism in Entomopathogenic Bacteria." *Frontiers in microbiology* 8: 209.

Eswar, Narayanan et al. 2006. "Comparative Protein Structure Modeling Using Modeller." *Current protocols in bioinformatics* Chapter 5: Unit-5.6.

Etzold, Sabrina, and Nathalie Juge. 2014. "Structural Insights into Bacterial Recognition of Intestinal Mucins." *Current opinion in structural biology* 28: 23–31.

Fantappiè, Laura et al. 2017. "Some Gram-Negative Lipoproteins Keep Their Surface Topology When Transplanted from One Species to Another and Deliver Foreign Polypeptides to the Bacterial Surface." *Molecular & Cellular Proteomics* 16(7): 1348 LP – 1364. http://www.mcponline.org/content/16/7/1348.abstract.

Felli, Isabella C., and Roberta Pierattelli, eds. 2015. *Intrinsically Disordered Proteins Studied by NMR Spectroscopy. Advances in Experimental Medicine and Biology*. 1st Edn. Springer Cham.

Ffrench-Constant, Richard H et al. 2003. "Photorhabdus: Towards a Functional Genomic Analysis of a Symbiont and Pathogen." *FEMS Microbiology Reviews* 26(5): 433–56. https://doi.org/10.1111/j.1574-6976.2003.tb00625.x.

Figueroa-Bossi, Nara et al. 2014. "RNA Remodeling by Bacterial Global Regulator CsrA Promotes Rho-Dependent Transcription Termination." *Genes & development* 28(11): 1239–51.

Fire, Andrew et al. 1998. "Potent and Specific Genetic Interference by Double-Stranded RNA in Caenorhabditis Elegans." *Nature* 391(6669): 806–11. https://doi.org/10.1038/35888.

Frézal, Lise et al. 2019. "Noda-Like RNA Viruses Infecting Caenorhabditis Nematodes: Sympatry, Diversity, and Reassortment." *Journal of Virology* 93(21): e01170-19. https://doi.org/10.1128/JVI.01170-19.

Fruchterman, Thomas M J, and Edward M Reingold. 1991. "Graph Drawing by Force-Directed Placement." *Software: Practice and Experience* 21(11): 1129–64. http://dx.doi.org/10.1002/spe.4380211102.

Fu, Zhen, Yuxiang Li, Axel A Elling, and William E Snyder. 2020. "A Draft Genome of a Field-

Collected Steinernema Feltiae Strain NW." *Journal of nematology* 52: 1–7.

Gay, N J, V L Tybulewicz, and J E Walker. 1986. "Insertion of Transposon Tn7 into the Escherichia Coli GlmS Transcriptional Terminator." *The Biochemical journal* 234(1): 111–17.

"Genome Sequence of the Nematode C. Elegans: A Platform for Investigating Biology." 1998. *Science* 282(5396): 2012–18.

Gerlt, John A et al. 2015. "Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1854(8): 1019–37. http://dx.doi.org/10.1016/j.bbapap.2015.04.015.

———. 2017. "Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence–Function Space and Genome Context to Discover Novel Functions." *Biochemistry* 56(33): 4293–4308. http://dx.doi.org/10.1021/acs.biochem.7b00614.

Gerrard, John G et al. 2003. "Photorhabdus Species: Bioluminescent Bacteria as Emerging Human Pathogens?" *Emerging infectious diseases* 9(2): 251–54.

Gerrard, John G, Nicholas R Waterfield, Renu Vohra, and Richard H Ffrench-Constant. 2004. "Human Infection with Photorhabdus Asymbiotica: An Emerging Bacterial Pathogen." *Microbes and infection* 6(2): 229–37.

Gianquinto, Eleonora et al. 2019. "Interaction of Human Hemoglobin and Semi-Hemoglobins with the Staphylococcus Aureus Hemophore IsdB: A Kinetic and Mechanistic Insight." *Scientific Reports* 9(1): 18629. https://doi.org/10.1038/s41598-019-54970-w.

Giovannoni, Stephen J. 2017. "SAR11 Bacteria: The Most Abundant Plankton in the Oceans." *Annual Review of Marine Science* 9(1): 231–55. https://doi.org/10.1146/annurev-marine-010814-015934.

Goldstein, Bob. 2016. "Sydney Brenner on the Genetics of Caenorhabditis Elegans." *Genetics* 204(1): 1–2.

Gómez-Santos, Nuria et al. 2019. "A TonB-Dependent Transporter Is Required for Secretion of Protease PopC across the Bacterial Outer Membrane." *Nature Communications* 10(1): 1360. https://doi.org/10.1038/s41467-019-09366-9.

Goodrich-Blair, Heidi. 2007. "They've Got a Ticket to Ride: Xenorhabdus Nematophila–Steinernema Carpocapsae Symbiosis." *Current Opinion in Microbiology* 10(3): 225–30. https://www.sciencedirect.com/science/article/pii/S1369527407000537.

Gouge, Dawn H, and Jennifer L Snyder. 2006. "Temporal Association of Entomopathogenic Nematodes (Rhabditida: Steinernematidae and Heterorhabditidae) and Bacteria." *Journal of Invertebrate Pathology* 91(3): 147–57. https://www.sciencedirect.com/science/article/pii/S0022201105002442.

Goya, María Eugenia et al. 2020. "Probiotic Bacillus Subtilis Protects against α-Synuclein Aggregation in C. Elegans." *Cell reports* 30(2): 367-380.e7.

Green, Erin R, and Joan Mecsas. 2016. "Bacterial Secretion Systems: An Overview." *Microbiology Spectrum* 4(1): 4.1.13. https://doi.org/10.1128/microbiolspec.VMBF-0012-2015.

Grossman, Alex S, Cristian A Escobar, et al. 2022. "A Surface Exposed, Two-Domain Lipoprotein

Cargo of a Type XI Secretion System Promotes Colonization of Host Intestinal Epithelia Expressing Glycans ." *Frontiers in Microbiology* 13. https://www.frontiersin.org/article/10.3389/fmicb.2022.800366.

Grossman, Alex S, Terra J Mauer, et al. 2022. "A Widespread Bacterial Secretion System with Diverse Substrates." *mBio* 12(4): e01956-21. https://doi.org/10.1128/mBio.01956-21.

Guan, Rongjin et al. 2004. "Structural Basis for Peptidoglycan Binding by Peptidoglycan Recognition Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 101(49): 17168–73.

Guérardel, Y et al. 2001. "The Nematode Caenorhabditis Elegans Synthesizes Unusual O-Linked Glycans: Identification of Glucose-Substituted Mucin-Type O-Glycans and Short Chondroitin-like Oligosaccharides." *The Biochemical journal* 357(Pt 1): 167–82.

Hantke, Klaus. 2003. "Is the Bacterial Ferrous Iron Transporter FeoB a Living Fossil?" *Trends in microbiology* 11(5): 192–95.

Hapeshi, Alexia, and Nicholas R Waterfield. 2017. "Photorhabdus Asymbiotica as an Insect and Human Pathogen." *Current topics in microbiology and immunology* 402: 159–77.

Hariadi, Nurul I et al. 2015. "Comparative Profile of Heme Acquisition Genes in Disease-Causing and Colonizing Nontypeable Haemophilus Influenzae and Haemophilus Haemolyticus." *Journal of clinical microbiology* 53(7): 2132–37.

Hartley, Cathryn J, Peter E Lillis, Rebecca A Owens, and Christine T Griffin. 2019. "Infective Juveniles of Entomopathogenic Nematodes (Steinernema and Heterorhabditis) Secrete Ascarosides and Respond to Interspecific Dispersal Signals." *Journal of invertebrate pathology* 168: 107257.

Hashmi, S, G Hashmi, and R Gaugler. 1995. "Genetic Transformation of an Entomopathogenic Nematode by Microinjection." *Journal of invertebrate pathology* 66(3): 293–96.

Head, Brian P, Abiola O Olaitan, and Alejandro Aballay. 2017. "Role of GATA Transcription Factor ELT-2 and P38 MAPK PMK-1 in Recovery from Acute P. Aeruginosa Infection in C. Elegans." *Virulence* 8(3): 261–74.

Heppert, Jennifer K et al. 2022. "Nematodes as Models for Symbiosis." *CABI Books*. https://doi.org/10.1079/9781789248814.0005.

Heungens, Kurt, Charles E Cowles, and Heidi Goodrich-Blair. 2002. "Identification of Xenorhabdus Nematophila Genes Required for Mutualistic Colonization of Steinernema Carpocapsae Nematodes." *Molecular Microbiology* 45(5): 1337–53. http://dx.doi.org/10.1046/j.1365-2958.2002.03100.x.

Hieb, W F, E L Stokstad, and M Rothstein. 1970. "Heme Requirement for Reproduction of a Free-Living Nematode." *Science (New York, N.Y.)* 168(3927): 143–44.

Hinnebusch, B J, Robert D Perry, and T G Schwan. 1996. "Role of the Yersinia Pestis Hemin Storage (Hms) Locus in the Transmission of Plague by Fleas." *Science (New York, N.Y.)* 273(5273): 367–70.

Hodgkin, J, P E Kuwabara, and B Corneliussen. 2000. "A Novel Bacterial Pathogen,

Microbacterium Nematophilum, Induces Morphological Change in the Nematode C. Elegans." *Current biology : CB* 10(24): 1615–18.

Hoffman, Casandra L, Jonathan Lalsiamthara, and Alejandro Aballay. 2020. "Host Mucin Is Exploited by Pseudomonas Aeruginosa To Provide Monosaccharides Required for a Successful Infection." *mBio* 11(2).

Holden, Victoria I, and Michael A Bachman. 2015. "Diverging Roles of Bacterial Siderophores during Infection." *Metallomics : integrated biometal science* 7(6): 986–95.

Hood, M Indriati, and Eric P Skaar. 2012. "Nutritional Immunity: Transition Metals at the Pathogen–Host Interface." *Nature Reviews Microbiology* 10(8): 525–37. https://doi.org/10.1038/nrmicro2836.

Hooda, Yogesh et al. 2016. "Slam Is an Outer Membrane Protein That Is Required for the Surface Display of Lipidated Virulence Factors in Neisseria." *Nature Microbiology* 1(4). http://dx.doi.org/10.1038/nmicrobiol.2016.9.

Hooda, Yogesh, Christine Chieh-Lin Lai, and Trevor F Moraes. 2017. "Identification of a Large Family of Slam-Dependent Surface Lipoproteins in Gram-Negative Bacteria." *Frontiers in Cellular and Infection Microbiology* 7. http://dx.doi.org/10.3389/fcimb.2017.00207.

Hooper, L V, and J I Gordon. 2001. "Glycans as Legislators of Host-Microbial Interactions: Spanning the Spectrum from Symbiosis to Pathogenicity." *Glycobiology* 11(2): 1R-10R.

House, H L, H E Welch, and T R Cleugh. 1965. "Food Medium of Prepared Dog Biscuit for the Mass-Production of the Nematode DD136 (Nematoda; Steinernematidae)." *Nature* 206(4986): 847. https://doi.org/10.1038/206847a0.

Hudson, Kieran L et al. 2015. "Carbohydrate-Aromatic Interactions in Proteins." *Journal of the American Chemical Society* 137(48): 15152–60.

Huson, Daniel H, and Celine Scornavacca. 2012. "Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks." *Systematic Biology* 61(6): 1061–67. http://dx.doi.org/10.1093/sysbio/sys062.

Hussa, Elizabeth A, and Heidi Goodrich-Blair. 2012. "Rearing and Injection of Manduca Sexta Larvae to Assess Bacterial Virulence." *Journal of visualized experiments : JoVE* (70): e4295.

Huynh, Minh Sang et al. 2021. "Reconstitution of Surface Lipoprotein Translocation Reveals Slam as an Outer Membrane Translocon in Gram-Negative Bacteria." *bioRxiv*: 2021.08.22.457263. http://biorxiv.org/content/early/2021/08/23/2021.08.22.457263.abstract.

———. 2022. "Reconstitution of Surface Lipoprotein Translocation through the Slam Translocon." *eLife* 11.

Ikeda, Takanori et al. 2007. "Influence of Lactic Acid Bacteria on Longevity of Caenorhabditis Elegans and Host Defense against Salmonella Enterica Serovar Enteritidis." *Applied and environmental microbiology* 73(20): 6404–9.

Irazoqui, Javier E, Jonathan M Urbach, and Frederick M Ausubel. 2010. "Evolution of Host Innate Defence: Insights from Caenorhabditis Elegans and Primitive Invertebrates." *Nature Reviews Immunology* 10(1): 47–58. https://doi.org/10.1038/nri2689.

Ivanova, Elena P et al. 2004. "Sulfitobacter Delicatus Sp. Nov. and Sulfitobacter Dubius Sp. Nov., Respectively from a Starfish (Stellaster Equestris) and Sea Grass (Zostera Marina)." *International journal of systematic and evolutionary microbiology* 54(Pt 2): 475–80.

Jackson, Lydgia et al. 2010. "Transcriptional and Functional Analysis of the Neisseria Gonorrhoeae Fur Regulon." *Journal of Bacteriology* 192(1): 77–85. https://doi.org/10.1128/JB.00741-09.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *PLOS ONE* 9(6): e98679. https://doi.org/10.1371/journal.pone.0098679.

Jansen, Ruud, Jan D A van Embden, Wim Gaastra, and Leo M Schouls. 2002. "Identification of Genes That Are Associated with DNA Repeats in Prokaryotes." *Molecular microbiology* 43(6): 1565–75.

Joyce, Susan A, and David J Clarke. 2003. "A HexA Homologue from Photorhabdus Regulates Pathogenicity, Symbiosis and Phenotypic Variation." *Molecular microbiology* 47(5): 1445–57.

Jumper, John et al. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature* 596(7873): 583–89.

Kamath, Ravi S, and Julie Ahringer. 2003. "Genome-Wide RNAi Screening in Caenorhabditis Elegans." *Methods (San Diego, Calif.)* 30(4): 313–21.

Kämpfer, Peter et al. 2017. "Xenorhabdus Thuongxuanensis Sp. Nov. and Xenorhabdus Eapokensis Sp. Nov., Isolated from Steinernema Species." *International Journal of Systematic and Evolutionary Microbiology* 67(5): 1107–14. http://dx.doi.org/10.1099/ijsem.0.001770.

Kanehisa, Minoru, Yoko Sato, and Kanae Morishima. 2016. "BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences." *Journal of molecular biology* 428(4): 726–31.

Kaplan, Fatma et al. 2012. "Interspecific Nematode Signals Regulate Dispersal Behavior." *PLOS ONE* 7(6): e38735. https://doi.org/10.1371/journal.pone.0038735.

Kelley, Lawrence A et al. 2015. "The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis." *Nature Protocols* 10(6): 845–58. http://dx.doi.org/10.1038/nprot.2015.053.

Kelley, Lawrence A, and Michael J E Sternberg. 2009. "Protein Structure Prediction on the Web: A Case Study Using the Phyre Server." *Nature Protocols* 4(3): 363–71. http://dx.doi.org/10.1038/nprot.2009.2.

Kergunteuil, Alan et al. 2016. "Biological Control beneath the Feet: A Review of Crop Protection against Insect Root Herbivores." *Insects* 7(4).

Kessner, Linda et al. 2016. "Draft Genome Sequence of Aliiroseovarius Crassostreae CV919-312, the Causative Agent of Roseovarius Oyster Disease (Formerly Juvenile Oyster Disease)." *Genome announcements* 4(2).

Khan, Fida A, Melanie A Fisher, and Rashida A Khakoo. 2007. "Association of Hemochromatosis with Infectious Diseases: Expanding Spectrum." *International Journal of Infectious Diseases* 11(6): 482–87. https://www.sciencedirect.com/science/article/pii/S1201971207000811.

Kim, Il-Hwan et al. 2017. "The Insect Pathogenic Bacterium Xenorhabdus Innexi Has Attenuated Virulence in Multiple Insect Model Hosts yet Encodes a Potent Mosquitocidal Toxin." *BMC genomics* 18(1): 927.

Kim, Jiyeun Kate et al. 2014. "Purine Biosynthesis, Biofilm Formation, and Persistence of an Insect-Microbe Gut Symbiosis." *Applied and environmental microbiology* 80(14): 4374–82.

Kim, Sam Kyu, Yolanda Flores-Lara, and Patricia S Stock. 2012. "Morphology and Ultrastructure of the Bacterial Receptacle in Steinernema Nematodes (Nematoda: Steinernematidae)." *Journal of Invertebrate Pathology* 110(3): 366–74. https://www.sciencedirect.com/science/article/pii/S0022201112001176.

Kim, Young-Ok et al. 2014. "Ascidiaceihabitans Donghaensis Gen. Nov., Sp. Nov., Isolated from the Golden Sea Squirt Halocynthia Aurantium." *International journal of systematic and evolutionary microbiology* 64(Pt 12): 3970–75.

Kim, Younghoon, and Eleftherios Mylonakis. 2012. "Caenorhabditis Elegans Immune Conditioning with the Probiotic Bacterium Lactobacillus Acidophilus Strain NCFM Enhances Gram-Positive Immune Responses." *Infection and immunity* 80(7): 2500–2508.

Kinkead, Lauren C et al. 2018. "Bacterial Lipoproteins and Other Factors Released by Francisella Tularensis Modulate Human Neutrophil Lifespan: Effects of a TLR1 SNP on Apoptosis Inhibition." *Cellular microbiology* 20(2).

Kohl, Paul, Franz G Zingl, Thomas O Eichmann, and Stefan Schild. 2018. "Isolation of Outer Membrane Vesicles Including Their Quantitative and Qualitative Analyses BT - Vibrio Cholerae: Methods and Protocols." In ed. Aleksandra E Sikora. New York, NY: Springer New York, 117–34. https://doi.org/10.1007/978-1-4939-8685-9_11.

Konar, Monica et al. 2015. "A Mutant Library Approach to Identify Improved Meningococcal Factor H Binding Protein Vaccine Antigens." *PLOS ONE* 10(6): e0128185. https://doi.org/10.1371/journal.pone.0128185.

Konopka, James B. 2012. "N-Acetylglucosamine (GlcNAc) Functions in Cell Signaling." *Scientifica* 2012.

Konovalova, Anna, and Thomas J Silhavy. 2015. "Outer Membrane Lipoprotein Biogenesis: Lol Is Not the End." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370(1679).

Koontz, Laura. 2014a. "Chapter One - TCA Precipitation." In *Laboratory Methods in Enzymology: Protein Part C*, ed. Jon B T - Methods in Enzymology Lorsch. Academic Press, 3–10. http://www.sciencedirect.com/science/article/pii/B978012420119400001X.

———. 2014b. "TCA Precipitation." *Methods in enzymology* 541: 3–10.

Kozyraki, Renata, Pierre Verroust, and Olivier Cases. 2022. "Chapter Four - Cubilin, the Intrinsic Factor-Vitamin B12 Receptor." In *Vitamin B12*, ed. Gerald B T - Vitamins and Hormones Litwack. Academic Press, 65–119. https://www.sciencedirect.com/science/article/pii/S008367292200005X.

Krokan, Hans E, and Magnar Bjørås. 2013. "Base Excision Repair." *Cold Spring Harbor perspectives in biology* 5(4): a012583.

Kumazaki, Kaoru et al. 2014. "Crystal Structure of Escherichia Coli YidC, a Membrane Protein Chaperone and Insertase." *Scientific Reports* 4(1): 7299. https://doi.org/10.1038/srep07299.

Lacey, L A et al. 2015. "Insect Pathogens as Biological Control Agents: Back to the Future." *Journal of invertebrate pathology* 132: 1–41.

Lasica, Anna M, Miroslaw Ksiazek, Mariusz Madej, and Jan Potempa. 2017. "The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function ." *Frontiers in Cellular and Infection Microbiology* 7. https://www.frontiersin.org/articles/10.3389/fcimb.2017.00215.

Latham, Roger D et al. 2020. "A Heme-Binding Protein Produced by Haemophilus Haemolyticus Inhibits Non-Typeable Haemophilus Influenzae." *Molecular Microbiology* 113(2): 381–98. https://doi.org/10.1111/mmi.14426.

Lee, Changhan, and Chankyu Park. 2017. "Bacterial Responses to Glyoxal and Methylglyoxal: Reactive Electrophilic Species." *International journal of molecular sciences* 18(1).

Lee, Ming-Min, and Patricia S Stock. 2010. "A Multigene Approach for Assessing Evolutionary Relationships of Xenorhabdus Spp. (γ-Proteobacteria), the Bacterial Symbionts of Entomopathogenic Steinernema Nematodes." *Journal of Invertebrate Pathology* 104(2): 67–74. http://dx.doi.org/10.1016/j.jip.2010.01.005.

Levengood, S K, W F Jr Beyer, and R E Webster. 1991. "TolA: A Membrane Protein Involved in Colicin Uptake Contains an Extended Helical Region." *Proceedings of the National Academy of Sciences of the United States of America* 88(14): 5939–43.

Lillard, James W Jr, Scott W Bearden, Jacqueline D Fetherston, and Robert D Perry. 1999. "The Haemin Storage (Hms+) Phenotype of Yersinia Pestis Is Not Essential for the Pathogenesis of Bubonic Plague in Mammals." *Microbiology (Reading, England)* 145 (Pt 1: 197–209.

Lopes-Marques, Mónica et al. 2014. "Basal Gnathostomes Provide Unique Insights into the Evolution of Vitamin B12 Binders." *Genome biology and evolution* 7(2): 457–64. https://pubmed.ncbi.nlm.nih.gov/25552533.

Low, Kristin E, and P Lynne Howell. 2018. "Gram-Negative Synthase-Dependent Exopolysaccharide Biosynthetic Machines." *Current Opinion in Structural Biology* 53: 32–44. https://www.sciencedirect.com/science/article/pii/S0959440X1830037X.

Lu, Dihong et al. 2017. "Activated Entomopathogenic Nematode Infective Juveniles Release Lethal Venom Proteins." *PLOS Pathogens* 13(4): e1006302. https://doi.org/10.1371/journal.ppat.1006302.

Luck, Ashley N et al. 2016. "Heme Acquisition in the Parasitic Filarial Nematode Brugia Malayi." *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 30(10): 3501–14.

Luk, Louis Y P et al. 2013. "Unraveling the Role of Protein Dynamics in Dihydrofolate Reductase Catalysis." *Proceedings of the National Academy of Sciences of the United States of America* 110(41): 16344–49.

Maddocks, Sarah E, and Petra C F Oyston. 2008. "Structure and Function of the LysR-Type Transcriptional Regulator (LTTR) Family Proteins." *Microbiology (Reading, England)* 154(Pt 12):

3609–23.

Maher, Abigail M D et al. 2021. "Competition and Co-Existence of Two Photorhabdus Symbionts with a Nematode Host." *Microbial ecology* 81(1): 223–39.

Mantel, N. 1966. "Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration." *Cancer chemotherapy reports* 50(3): 163–70.

Marmont, Lindsey S et al. 2017. "PelA and PelB Proteins Form a Modification and Secretion Complex Essential for Pel Polysaccharide-Dependent Biofilm Formation in Pseudomonas Aeruginosa." *The Journal of biological chemistry* 292(47): 19411–22. https://pubmed.ncbi.nlm.nih.gov/28972168.

Martens, Eric C et al. 2003. "Xenorhabdus Nematophila Requires an Intact IscRSUA-HscBA-Fdx Operon to Colonize Steinernema Carpocapsae Nematodes." *Journal of bacteriology* 185(12): 3678–82.

Martens, Eric C, and Heidi Goodrich-Blair. 2005. "The Steinernema Carpocapsae Intestinal Vesicle Contains a Subcellular Structure with Which Xenorhabdus Nematophila Associates during Colonization Initiation." *Cellular microbiology* 7(12): 1723–35.

Martens, Eric C, Kurt Heungens, and Heidi Goodrich-Blair. 2003. "Early Colonization Events in the Mutualistic Association between Steinernema Carpocapsae Nematodes and Xenorhabdus Nematophila Bacteria." *Journal of Bacteriology* 185(10): 3147–54. https://doi.org/10.1128/JB.185.10.3147-3154.2003.

Martens, Eric C, Frances M Russell, and Heidi Goodrich-Blair. 2005. "Analysis of Xenorhabdus Nematophila Metabolic Mutants Yields Insight into Stages of Steinernema Carpocapsae Nematode Intestinal Colonization." *Molecular microbiology* 58(1): 28–45.

Matthews, Benjamin J, and Leslie B Vosshall. 2020. "How to Turn an Organism into a Model Organism in 10 'easy' Steps." *The Journal of experimental biology* 223(Pt Suppl 1).

Mauer, Terra J, Alex S Grossman, Katrina T Forest, and Heidi Goodrich-Blair. 2020. "Beyond Slam: The DUF560 Family Outer Membrane Protein Superfamily SPAM Has Distinct Network Subclusters That Suggest a Role in Non-Lipidated Substrate Transport and Bacterial Environmental Adaptation." *bioRxiv*. https://www.biorxiv.org/content/early/2020/01/21/2020.01.20.912956.

McGinnis, S, and Thomas L Madden. 2004. "BLAST: At the Core of a Powerful and Diverse Set of Sequence Analysis Tools." *Nucleic Acids Research* 32(Web Server): W20–25. http://dx.doi.org/10.1093/nar/gkh435.

McMullen II, John, and Patricia S Stock. 2014. "In Vivo and In Vitro Rearing of Entomopathogenic Nematodes (Steinernematidae and Heterorhabditidae)." *Journal of visualized experiments : JoVE* 91.

McMullen, John G et al. 2017. "Fitness Costs of Symbiont Switching Using Entomopathogenic Nematodes as a Model." *BMC Evolutionary Biology* 17(1): 100. https://doi.org/10.1186/s12862-017-0939-6.

Mendel, Ralf R. 2013. "The Molybdenum Cofactor." *The Journal of biological chemistry* 288(19): 13165–72.

Meng, Qingyun et al. 2019. "Probing Peptide Substrate Specificities of N-Glycosyltranferase Isoforms from Different Bacterial Species." *Carbohydrate research* 473: 82–87.

Micsonai, András et al. 2015. "Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy." *Proceedings of the National Academy of Sciences of the United States of America* 112(24): E3095-103.

Miller, J. H. 1972. *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Miller, Mark A, Wayne Pfeiffer, and Terri Schwartz. 2010. "Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees." *2010 Gateway Computing Environments Workshop (GCE)*. http://dx.doi.org/10.1109/gce.2010.5676129.

Mirdita, Milot et al. 2022. "ColabFold: Making Protein Folding Accessible to All." *Nature methods* 19(6): 679–82.

Miroux, B, and J E Walker. 1996. "Over-Production of Proteins in Escherichia Coli: Mutant Hosts That Allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels." *Journal of molecular biology* 260(3): 289–98.

Mistry, Jaina et al. 2021. "Pfam: The Protein Families Database in 2021." *Nucleic Acids Research* 49(D1): D412–19. https://doi.org/10.1093/nar/gkaa913.

Mitkowski, Paweł et al. 2019. "Structural Bases of Peptidoglycan Recognition by Lysostaphin SH3b Domain." *Scientific reports* 9(1): 5965.

Miyata, Sachiko, Jakob Begun, Emily R Troemel, and Frederick M Ausubel. 2008. "DAF-16-Dependent Suppression of Immunity during Reproduction in Caenorhabditis Elegans." *Genetics* 178(2): 903–18.

Montalvo-Katz, Sirena et al. 2013. "Association with Soil Bacteria Enhances P38-Dependent Infection Resistance in Caenorhabditis Elegans." *Infection and immunity* 81(2): 514–20.

Moore, Rebecca S, Rachel Kaletsky, and Coleen T Murphy. 2019. "Piwi/PRG-1 Argonaute and TGF-β Mediate Transgenerational Learned Pathogenic Avoidance." *Cell* 177(7): 1827-1841.e12.

Moraes, Trevor F et al. 2021. "Slam Polynucleotides and Polypeptides and Uses Thereof." : 1–36. https://patents.google.com/patent/US20210379176A1/en.

Morran, Levi T et al. 2016. "Nematode-Bacteria Mutualism: Selection within the Mutualism Supersedes Selection Outside of the Mutualism." *Evolution; international journal of organic evolution* 70(3): 687–95.

Morris, Robert et al. 2017. "A Neuropeptide Modulates Sensory Perception in the Entomopathogenic Nematode Steinernema Carpocapsae." *PLOS Pathogens* 13(3): e1006185. https://doi.org/10.1371/journal.ppat.1006185.

Mucci, Nicholas C et al. 2022. "Apex Predator Nematodes and Meso-Predator Bacteria Consume Their Basal Insect Prey through Discrete Stages of Chemical Transformations." *mSystems* 7(3): e0031222.

Mudgal, Richa, Sankaran Sandhya, Nagasuma Chandra, and Narayanaswamy Srinivasan. 2015. "De-

DUFing the DUFs: Deciphering Distant Evolutionary Relationships of Domains of Unknown Function Using Sensitive Homology Detection Methods." *Biology Direct* 10(1): 38. https://doi.org/10.1186/s13062-015-0069-2.

Muller, H J. 1964. "The Relation of Recombination to Mutational Advance." *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 1(1): 2–9. https://www.sciencedirect.com/science/article/pii/0027510764900478.

Murali, Ranjani, and Robert B Gennis. 2018. "Functional Importance of Glutamate-445 and Glutamate-99 in Proton-Coupled Electron Transfer during Oxygen Reduction by Cytochrome Bd from Escherichia Coli." *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1859(8): 577–90. https://www.sciencedirect.com/science/article/pii/S0005272818300884.

Murfin, Kristen E, Ming-Min Lee, et al. 2015a. "Xenorhabdus Bovienii Strain Diversity Impacts Coevolution and Symbiotic Maintenance with Steinernema Spp. Nematode Hosts." *mBio* 6(3): e00076-15. https://doi.org/10.1128/mBio.00076-15.

———. 2015b. "Xenorhabdus Bovienii Strain Diversity Impacts Coevolution and Symbiotic Maintenance with Steinernema Spp. Nematode Hosts." *mBio* 6(3): e00076–e00076. https://doi.org/10.1128/mBio.00076-15.

Murfin, Kristen E, John M Chaston, and Heidi Goodrich-Blair. 2012. "Visualizing Bacteria in Nematodes Using Fluorescent Microscopy." *JoVE* (68): e4298. http://dx.doi.org/10.3791/4298.

Murfin, Kristen E, Daren R Ginete, Farrah Bashey, and Heidi Goodrich-Blair. 2018. "Symbiont-Mediated Competition: Xenorhabdus Bovienii Confer an Advantage to Their Nematode Host Steinernema Affine by Killing Competitor Steinernema Feltiae." *Environmental microbiology*.

Murfin, Kristen E, Amy C Whooley, Jonathan L Klassen, and Heidi Goodrich-Blair. 2015. "Comparison of Xenorhabdus Bovienii Bacterial Strain Genomes Reveals Diversity in Symbiotic Functions." *BMC Genomics* 16(1): 889. https://doi.org/10.1186/s12864-015-2000-8.

Nakata, A, M Amemura, and Kozo Makino. 1989. "Unusual Nucleotide Arrangement with Repeated Sequences in the Escherichia Coli K-12 Chromosome." *Journal of bacteriology* 171(6): 3553–56.

Narita, Shin-ichiro, and Hajime Tokuda. 2011. "Overexpression of LolCDE Allows Deletion of the Escherichia Coli Gene Encoding Apolipoprotein N-Acyltransferase." *Journal of bacteriology* 193(18): 4832–40.

Nealson, K H, and J W Hastings. 1979. "Bacterial Bioluminescence: Its Control and Ecological Significance." *Microbiological reviews* 43(4): 496–518.

Nielsen, Henrik, J Engelbrecht, S Brunak, and Gunnar Von Heijne. 1997. "Identification of Prokaryotic and Eukaryotic Signal Peptides and Prediction of Their Cleavage Sites." *Protein Engineering Design and Selection* 10(1): 1–6. http://dx.doi.org/10.1093/protein/10.1.1.

Noinaj, Nicholas, Maude Guillier, Travis J Barnard, and Susan K Buchanan. 2010. "TonB-Dependent Transporters: Regulation, Structure, and Function." *Annual review of microbiology* 64: 43–60. https://pubmed.ncbi.nlm.nih.gov/20420522.

O'Donnell, Michael P et al. 2020. "A Neurotransmitter Produced by Gut Bacteria Modulates Host Sensory Behaviour." *Nature* 583(7816): 415–20. https://doi.org/10.1038/s41586-020-2395-5.

Ogier, Jean-Claude, Sylvie Pagès, Marie Frayssinet, and Sophie Gaudriault. 2020. "Entomopathogenic Nematode-Associated Microbiota: From Monoxenic Paradigm to Pathobiome." *Microbiome* 8(1): 25.

Oh, S, G W Go, E Mylonakis, and Y Kim. 2012. "The Bacterial Signalling Molecule Indole Attenuates the Virulence of the Fungal Pathogen Candida Albicans." *Journal of applied microbiology* 113(3): 622–28.

Old, Iain G, Danielle Margarita, Robert E Glass, and Isabelle Saint Girons. 1990. "Nucleotide Sequence of the MetH Gene of Escherichia Coli K-12 and Comparison with That of Salmonella Typhimurium LT2." *Gene* 87(1): 15–21. https://www.sciencedirect.com/science/article/pii/037811199090490I.

Ollis, Anne A, and Kathleen Postle. 2012. "ExbD Mutants Define Initial Stages in TonB Energization." *Journal of molecular biology* 415(2): 237–47.

Orchard, Samantha S, and Heidi Goodrich-Blair. 2004. "Identification and Functional Characterization of a Xenorhabdus Nematophila Oligopeptide Permease." *Applied and environmental microbiology* 70(9): 5621–27. https://doi.org/10.1128/AEM.70.9.5621-5627.2004.

Ostan, Nicholas K H et al. 2017. "Lactoferrin Binding Protein B - a Bi-Functional Bacterial Receptor Protein." *PLoS pathogens* 13(3): e1006244.

Ostberg, Karen et al. 2013. "Conserved Regions of Gonococcal TbpB Are Critical for Surface Exposure and Transferrin Iron Utilization." *Infection and Immunity* 81(9): 3442–50. https://doi.org/10.1128/IAI.00280-13.

Packer, Jonathan S et al. 2019. "A Lineage-Resolved Molecular Atlas of C. Elegans Embryogenesis at Single-Cell Resolution." *Science (New York, N.Y.)* 365(6459).

Palmer, Tracy et al. 2021. "A Holin/Peptidoglycan Hydrolase-Dependent Protein Secretion System." *Molecular Microbiology* 115(3): 345–55. https://doi.org/10.1111/mmi.14599.

Parker, Ashley et al. 2017. "Alternative Pathways for Escherichia Coli Biofilm Formation Revealed by SRNA Overproduction." *Molecular microbiology* 105(2): 309–25.

Périnet, Simone et al. 2016. "Molybdate Transporter ModABC Is Important for Pseudomonas Aeruginosa Chronic Lung Infection." *BMC Research Notes* 9(1): 23. https://doi.org/10.1186/s13104-016-1840-x.

Peters, Arne, and Ralf-Udo Ehlers. 1994. "Susceptibility of Leatherjackets (Tipula Paludosa and Tipula Oleracea; Tipulidae; Nematocera) to the Entomopathogenic Nematode Steinernema Feltiae." *Journal of Invertebrate Pathology* 63(2): 163–71. https://www.sciencedirect.com/science/article/pii/S0022201184710317.

Peters, Joseph E, and Nancy L Craig. 2001. "Tn7: Smarter than We Thought." *Nature Reviews Molecular Cell Biology* 2(11): 806–14. https://doi.org/10.1038/35099006.

Pettersson, Mats E, and Otto G Berg. 2007. "Muller's Ratchet in Symbiont Populations." *Genetica* 130(2): 199–211.

Peyvandi, Flora, Isabella Garagiola, and Luciano Baronciani. 2011. "Role of von Willebrand Factor in the Haemostasis." *Blood transfusion = Trasfusione del sangue* 9 Suppl 2(Suppl 2): s3-8.

Pitt, Jason N et al. 2019. "WormBot, an Open-Source Robotics Platform for Survival and Behavior Analysis in C. Elegans." *GeroScience* 41(6): 961–73.

Pogoutse, Anastassia K, and Trevor F Moraes. 2017. "Iron Acquisition through the Bacterial Transferrin Receptor." *Critical Reviews in Biochemistry and Molecular Biology* 52(3): 314–26. http://dx.doi.org/10.1080/10409238.2017.1293606.

Poinar, George O. 1966. "The Presence of Achromobacter Nematophilus in the Infective Stage of a Neoaplectana Sp. (Steinernematidae: Nematoda)." *Nematologica* 12(1): 105–8. http://dx.doi.org/10.1163/187529266x00068.

———. 1990. "Taxonomy and Biology of Steinernematidae and Heterorhabditidae." In , 21–62.

Poinar, George O, and P S Grewal. 2012. "History of Entomopathogenic Nematology." *Journal of nematology* 44(2): 153–61.

Poinar, George O, and R Leutenegger. 1968. "Anatomy of the Infective and Normal Third-Stage Juveniles of Neoaplectana Carpocapsae Weiser (Steinernematidae: Nematoda)." *The Journal of parasitology* 54(2): 340–50.

Poinar, George O, and Gerard M Thomas. 1965. "A New Bacterium, Achromobacter Nematophilus Sp. Nov. (Achromobacteriaceae: Eubacteriales) Associated with a Nematode." *International Journal of Systematic and Evolutionary Microbiology* 15(4): 249–52. https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-15-4-249.

———. 1966. "Significance of Achromobacter Nematophilus Poinar and Thomas (Achromobacteraceae: Eubacteriales) in the Development of the Nematode, DD-136 (Neoaplectana Sp. Steinernematidae)." *Parasitology* 56(2): 385–90.

———. 1967. "The Nature of Achromobacter Nematophilus as an Insect Pathogen." *Journal of Invertebrate Pathology* 9(4): 510–14. https://www.sciencedirect.com/science/article/pii/0022201167901310.

Poinar, George O, Gerard M Thomas, and Roberta Hess. 1977. "Characteristics of the Specific Bacterium Associated With Heterorhabditis Bacteriophora (Heterorhabditidae: Rhabditida)." *Nematologica* 23(1): 97–102. https://brill.com/view/journals/nema/23/1/article-p97_12.xml.

Posey, J E, and F C Gherardini. 2000. "Lack of a Role for Iron in the Lyme Disease Pathogen." *Science* 288(5471): 1651–53.

Pourciau, Christine et al. 2019. "Regulation of Iron Storage by CsrA Supports Exponential Growth of Escherichia Coli." *mBio* 10(4).

Provencher, S W, and J Glöckner. 1981. "Estimation of Globular Protein Secondary Structure from Circular Dichroism." *Biochemistry* 20(1): 33–37.

Qi, Huang, and Palmer Tracy. 2017. "Signal Peptide Hydrophobicity Modulates Interaction with the Twin-Arginine Translocase." *mBio* 8(4): e00909-17. https://doi.org/10.1128/mBio.00909-17.

Quillin, Sarah, Adam Hockenberry, Michael Jewett, and Steven H Seifert. 2018. "Neisseria Gonorrhoeae Exposed to Sublethal Levels of Hydrogen Peroxide Mounts a Complex Transcriptional Response." *mSystems* 3(5): e00156-18. https://doi.org/10.1128/mSystems.00156-18.

Quiocho, F A. 1986. "Carbohydrate-Binding Proteins: Tertiary Structures and Protein-Sugar Interactions." *Annual review of biochemistry* 55: 287–315.

Rae, Robbie G, Maria Tourna, and Michael J Wilson. 2010. "The Slug Parasitic Nematode Phasmarhabditis Hermaphrodita Associates with Complex and Variable Bacterial Assemblages That Do Not Affect Its Virulence." *Journal of invertebrate pathology* 104(3): 222–26.

Rambaut, Andrew. 2018. "FigTree v1.4.4." http://tree.bio.ed.ac.uk/software/figtree/.

Rao, Anita U, Lynn K Carta, Emmanuel Lesuisse, and Iqbal Hamza. 2005. "Lack of Heme Synthesis in a Free-Living Eukaryote." *Proceedings of the National Academy of Sciences* 102(12): 4270–75. https://doi.org/10.1073/pnas.0500877102.

Ratnappan, Ramesh et al. 2016. "RNAi-Mediated Gene Knockdown by Microinjection in the Model Entomopathogenic Nematode Heterorhabditis Bacteriophora." *Parasites & Vectors* 9(1): 160. https://doi.org/10.1186/s13071-016-1442-4.

El Rayes, Jessica et al. 2021. "Disorder Is a Critical Component of Lipoprotein Sorting in Gram-Negative Bacteria." *Nature chemical biology* 17(10): 1093–1100.

Richards, Gregory R, and Heidi Goodrich-Blair. 2009. "Masters of Conquest and Pillage: Xenorhabdus Nematophila Global Regulators Control Transitions from Virulence to Nutrient Acquisition." *Cellular microbiology* 11(7): 1025–33.

Ringel, Phillip et al. 2013. "Biochemical Characterization of Molybdenum Cofactor-Free Nitrate Reductase from Neurospora Crassa." *The Journal of biological chemistry* 288(20): 14657–71.

Rodríguez-Alonso, Raquel et al. 2020. "Structural Insight into the Formation of Lipoprotein-β-Barrel Complexes." *Nature chemical biology* 16(9): 1019–25.

Rogozin, Igor B et al. 2002. "Connected Gene Neighborhoods in Prokaryotic Genomes." *Nucleic acids research* 30(10): 2212–23.

Ronquist, Fredrik et al. 2012. "MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space." *Systematic Biology* 61(3): 539–42. http://dx.doi.org/10.1093/sysbio/sys029.

Rougon-Cardoso, Alejandra et al. 2016. "The Genome, Transcriptome, and Proteome of the Nematode Steinernema Carpocapsae: Evolutionary Signatures of a Pathogenic Lifestyle." *Scientific Reports* 6(1): 37536. https://doi.org/10.1038/srep37536.

Ruiz, Natividad, Luisa S Gronenberg, Daniel Kahne, and Thomas J Silhavy. 2008. "Identification of Two Inner-Membrane Proteins Required for the Transport of Lipopolysaccharide to the Outer Membrane of Escherichia Coli." *Proceedings of the National Academy of Sciences of the United States of America* 105(14): 5537–42.

Safarian, Schara et al. 2016. "Structure of a Bd Oxidase Indicates Similar Mechanisms for Membrane-Integrated Oxygen Reductases." *Science (New York, N.Y.)* 352(6285): 583–86.

Samuel, Buck S et al. 2016. "Caenorhabditis Elegans Responses to Bacteria from Its Natural Habitats." *Proceedings of the National Academy of Sciences* 113(27): E3941–49. https://doi.org/10.1073/pnas.1607183113.

Savolainen, K et al. 2001. "Expression of Pls, a Gene Closely Associated with the MecA Gene of

Methicillin-Resistant Staphylococcus Aureus, Prevents Bacterial Adhesion in Vitro." *Infection and immunity* 69(5): 3013–20.

Schabenberger, Oliver. 2005. "Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models." *Proceedings of the Thirtieth Annual SAS users Group International Conference*.

Schelkle, Bettina et al. 2018. "Caenorhabditis Elegans Predation on Bacillus Anthracis: Decontamination of Spore Contaminated Soil with Germinants and Nematodes ." *Frontiers in Microbiology* 8. https://www.frontiersin.org/articles/10.3389/fmicb.2017.02601.

Schmidt, M Alexander, Lee W Riley, and Inga Benz. 2003. "Sweet New World: Glycoproteins in Bacterial Pathogens." *Trends in microbiology* 11(12): 554–61.

Schneider, Muriel C et al. 2006. "Functional Significance of Factor H Binding to Neisseria Meningitidis." *Journal of immunology (Baltimore, Md. : 1950)* 176(12): 7566–75.

Schwartz, Christopher J et al. 2001. "IscR, an Fe-S Cluster-Containing Transcription Factor, Represses Expression of Escherichia Coli Genes Encoding Fe-S Cluster Assembly Proteins." *Proceedings of the National Academy of Sciences* 98(26): 14895–900. https://doi.org/10.1073/pnas.251550898.

Schweppe, Devin K. et al. 2015. "Host-Microbe Protein Interactions during Bacterial Infection." *Chemistry & Biology* 22(11): 1521–30. http://www.sciencedirect.com/science/article/pii/S107455211500383X.

Schwiesow, Leah et al. 2018. "Control of Hmu Heme Uptake Genes in Yersinia Pseudotuberculosis in Response to Iron Sources ." *Frontiers in Cellular and Infection Microbiology* 8. https://www.frontiersin.org/article/10.3389/fcimb.2018.00047.

Science Gateway. "Protein Molecular Weight Calculator." https://www.sciencegateway.org/tools/proteinmw.htm.

Scott, A L et al. 2012. "Filarial and Wolbachia Genomics." *Parasite immunology* 34(2–3): 121–29.

Serra, Lorrayne et al. 2019. "Hybrid Assembly of the Genome of the Entomopathogenic Nematode Steinernema Carpocapsae Identifies the X-Chromosome." *G3 Genes|Genomes|Genetics* 9(8): 2687–97. https://doi.org/10.1534/g3.119.400180.

Shah, Mohammad Manjur, and Mohammad Mahamood. 2017. "Introductory Chapter: Nematodes - A Lesser Known Group of Organisms."

Shanmugam, Mayilvahanan, Faiha El Abbar, and Narayanan Ramasubbu. 2015. "Transcriptome Profiling of Wild-Type and Pga-Knockout Mutant Strains Reveal the Role of Exopolysaccharide in Aggregatibacter Actinomycetemcomitans." *PloS one* 10(7): e0134285.

Shannon, P. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome Research* 13(11): 2498–2504. http://dx.doi.org/10.1101/gr.1239303.

Shapira, Michael. 2009. "Stress Effects on Immunity in Vertebrates and Invertebrates." *Stress - From Molecules to Behavior*: 207–27. https://doi.org/10.1002/9783527628346.ch11.

Shapira, Michael, and Man-Wah Tan. 2008. "Genetic Analysis of Caenorhabditis Elegans Innate

Immunity." *Methods in molecular biology* 415: 429–42.

Shapiro-Ilan, David et al. 2013. Mass Production of Beneficial Organisms: Invertebrates and Entomopathogens *Production of Entomopathogenic Nematodes*.

Shields, Robert C, and Paul A Jensen. 2019. "The Bare Necessities: Uncovering Essential and Condition-Critical Genes with Transposon Sequencing." *Molecular oral microbiology* 34(2): 39–50.

Shimell, Mary Jane, Edwin L Ferguson, Steven R Childs, and Michael B O'Connor. 1991. "The Drosophila Dorsal-Ventral Patterning Gene tolloid</Em> Is Related to Human Bone Morphogenetic Protein 1." *Cell* 67(3): 469–81. https://doi.org/10.1016/0092-8674(91)90522-Z.

Shishiniova, M, L Budurova, and D Gradinarov. 1998. "Steinernema Carpocapsae (Weiser, 1955) (Nematoda, Rhabditida) - New Species for Entomopathogenic Fauna of Bulgaria." *Experimental Pathology and Parasitology* 1: 30–35.

Sicard, M et al. 2004. "When Mutualists Are Pathogens: An Experimental Study of the Symbioses between Steinernema (Entomopathogenic Nematodes) and Xenorhabdus (Bacteria)." *Journal of evolutionary biology* 17(5): 985–93.

Sieber, Christian M K et al. 2019. "Unusual Metabolism and Hypervariation in the Genome of a Gracilibacterium (BD1-5) from an Oil-Degrading Community." *mBio* 10(6).

Sierra, Gonzalo. 1957. "A Simple Method for the Detection of Lipolytic Activity of Micro-Organisms and Some Observations on the Influence of the Contact between Cells and Fatty Substrates." *Antonie van Leeuwenhoek* 23(1): 15–22. https://doi.org/10.1007/BF02545855.

Silva-Rohwer, Amelia R et al. 2021. "CsrA Enhances Cyclic-Di-GMP Biosynthesis and Yersinia Pestis Biofilm Blockage of the Flea Foregut by Alleviating Hfq-Dependent Repression of the HmsT MRNA." *mBio* 12(4): e0135821.

da Silva, Ronni A G et al. 2019. "Variant Signal Peptides of Vaccine Antigen, FHbp, Impair Processing Affecting Surface Localization and Antibody-Mediated Killing in Most Meningococcal Isolates." *Frontiers in Microbiology* 10: 2847. https://www.frontiersin.org/article/10.3389/fmicb.2019.02847.

Singh, Swati et al. 2014. "Microbial Population Dynamics in the Hemolymph of Manduca Sexta Infected with Xenorhabdus Nematophila and the Entomopathogenic Nematode Steinernema Carpocapsae." *Applied and environmental microbiology* 80(14): 4277–85.

Skaf, Munir Salomão, Igor Polikarpov, and Ivana M Stanković. 2020. "A Linker of the Proline-Threonine Repeating Motif Sequence Is Bimodal." *Journal of molecular modeling* 26(7): 178.

Snyder, Holly et al. 2007. "New Insights into the Colonization and Release Processes of Xenorhabdus Nematophila and the Morphology and Ultrastructure of the Bacterial Receptacle of Its Nematode Host, Steinernema Carpocapsae." *Applied and Environmental Microbiology* 73(16): 5338–46. http://dx.doi.org/10.1128/aem.02947-06.

Somvanshi, Vishal S et al. 2010. "Photorhabdus Phase Variants Express a Novel Fimbrial Locus, Mad, Essential for Symbiosis." *Molecular microbiology* 77(4): 1021–38.

Spiridonov, Sergei E et al. 2004. "Phylogenetic Relationships within the Genus Steinernema

(Nematoda: Rhabditida) as Inferred from Analyses of Sequences of the ITS1-5.8S-ITS2 Region of RDNA and Morphological Features." *Nematology* 6(4): 547–66. http://dx.doi.org/10.1163/1568541042665304.

Sprouffske, Kathleen, and Andreas Wagner. 2016. "Growthcurver: An R Package for Obtaining Interpretable Metrics from Microbial Growth Curves." *BMC Bioinformatics* 17(1): 172. https://doi.org/10.1186/s12859-016-1016-7.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30(9): 1312–13. http://dx.doi.org/10.1093/bioinformatics/btu033.

Stiernagle, Theresa. 2006. "Maintenance of C. Elegans." *WormBook : the online review of C. elegans biology*: 1–11.

Stock, Patricia S. 2005. "Insect-Parasitic Nematodes: From Lab Curiosities to Model Organisms." *Journal of invertebrate pathology* 89(1): 57–66.

———. 2019. "Partners in Crime: Symbiont-Assisted Resource Acquisition in Steinernema Entomopathogenic Nematodes." *Current opinion in insect science* 32: 22–27.

Stohl, Elizabeth A, Alison K Criss, and Steven H Seifert. 2005. "The Transcriptome Response of Neisseria Gonorrhoeae to Hydrogen Peroxide Reveals Genes with Previously Uncharacterized Roles in Oxidative Damage Protection." *Molecular Microbiology* 58(2): 520–32. https://doi.org/10.1111/j.1365-2958.2005.04839.x.

van Stokkum, I H et al. 1990. "Estimation of Protein Secondary Structure and Error Analysis from Circular Dichroism Spectra." *Analytical biochemistry* 191(1): 110–18.

Stoll, Norman R. 1953. "Axenic Cultivation of the Parasitic Nematode, Neoaplectana Glaseri, in a Fluid Medium Containing Raw Liver Extract." *The Journal of Parasitology* 39(4): 422–44. http://www.jstor.org/stable/3274285.

Stolz, John F. 2017. "Gaia and Her Microbiome." *FEMS microbiology ecology* 93(2).

Stoy, Kayla S, Amanda K Gibson, Nicole M Gerardo, and Levi T Morran. 2020. "A Need to Consider the Evolutionary Genetics of Host–Symbiont Mutualisms." *Journal of Evolutionary Biology* 33(12): 1656–68. https://doi.org/10.1111/jeb.13715.

Stubbendieck, Reed M, Hongjie Li, and Cameron R Currie. 2019. "Convergent Evolution of Signal-Structure Interfaces for Maintaining Symbioses." *Current Opinion in Microbiology* 50: 71–78. https://www.sciencedirect.com/science/article/pii/S1369527419300554.

Styer, Katie L et al. 2008. "Innate Immunity in Caenorhabditis Elegans Is Regulated by Neurons Expressing NPR-1/GPCR." *Science* 322(5900): 460–64.

Sugar, Darby Renneckar et al. 2012. "Phenotypic Variation and Host Interactions of Xenorhabdus Bovienii SS-2004, the Entomopathogenic Symbiont of Steinernema Jollieti Nematodes." *Environmental microbiology* 14(4): 924–39.

Sulston, J E, E Schierenberg, J G White, and J N Thomson. 1983. "The Embryonic Cell Lineage of the Nematode Caenorhabditis Elegans." *Developmental biology* 100(1): 64–119.

Suzek, Baris E et al. 2007. "UniRef: Comprehensive and Non-Redundant UniProt Reference

Clusters." *Bioinformatics* 23(10): 1282–88. https://doi.org/10.1093/bioinformatics/btm098.

Swati, Singh et al. 2015. "Role of Secondary Metabolites in Establishment of the Mutualistic Partnership between Xenorhabdus Nematophila and the Entomopathogenic Nematode Steinernema Carpocapsae ." *Applied and Environmental Microbiology* 81(2): 754–64. https://doi.org/10.1128/AEM.02650-14.

Swearingen, Kristian E et al. 2016. "Interrogating the Plasmodium Sporozoite Surface: Identification of Surface-Exposed Proteins and Demonstration of Glycosylation on CSP and TRAP by Mass Spectrometry-Based Proteomics." *PLoS pathogens* 12(4): e1005606.

Tailford, Louise E, Emmanuelle H Crost, Devon Kavanaugh, and Nathalie Juge. 2015. "Mucin Glycan Foraging in the Human Gut Microbiome ." *Frontiers in Genetics* 6. https://www.frontiersin.org/articles/10.3389/fgene.2015.00081.

Tan, L, and P S Grewal. 2001. "Pathogenicity of Moraxella Osloensis, a Bacterium Associated with the Nematode Phasmarhabditis Hermaphrodita, to the Slug Deroceras Reticulatum." *Applied and environmental microbiology* 67(11): 5010–16.

Tan, Man-Wah et al. 1999. "Pseudomonas Aeruginosa Killing of Caenorhabditis Elegans Used to Identify P. Aeruginosa Virulence Factors." *Proceedings of the National Academy of Sciences of the United States of America* 96(5): 2408–13.

Thaler, Jacques-Olivier, S Baghdiguian, and Noël Boemare. 1995. "Purification and Characterization of Xenorhabdicin, a Phage Tail-like Bacteriocin, from the Lysogenic Strain F1 of Xenorhabdus Nematophilus." *Applied and environmental microbiology* 61(5): 2049–52.

Thappeta, Kishore Reddy Venkata et al. 2020. "R-Type Bacteriocins of Xenorhabdus Bovienii Determine the Outcome of Interspecies Competition in a Natural Host Environment." *Microbiology (Reading, England)* 166(11): 1074–87.

Tietz, Jonathan I et al. 2017. "A New Genome-Mining Tool Redefines the Lasso Peptide Biosynthetic Landscape." *Nature chemical biology* 13(5): 470–78.

Tjalsma, H et al. 1999. "The Role of Lipoprotein Processing by Signal Peptidase II in the Gram-Positive Eubacterium Bacillus Subtilis. Signal Peptidase II Is Required for the Efficient Secretion of Alpha-Amylase, a Non-Lipoprotein." *The Journal of biological chemistry* 274(3): 1698–1707.

Torres-Barragan, Andrea, Alonso Suazo, Wayne G Buhler, and Yasmin J Cardoza. 2011. "Studies on the Entomopathogenicity and Bacterial Associates of the Nematode Oscheius Carolinensis." *Biological Control* 59(2): 123–29. https://www.sciencedirect.com/science/article/pii/S1049964411001630.

Tortorelli, Giada et al. 2022. "Cell Surface Carbohydrates of Symbiotic Dinoflagellates and Their Role in the Establishment of Cnidarian-Dinoflagellate Symbiosis." *The ISME journal* 16(1): 190–99.

Troemel, Emily R et al. 2006. "P38 MAPK Regulates Expression of Immune Response Genes and Contributes to Longevity in C. Elegans." *PLoS genetics* 2(11): e183.

———. 2008. "Microsporidia Are Natural Intracellular Parasites of the Nematode Caenorhabditis Elegans." *PLOS Biology* 6(12): e309. https://doi.org/10.1371/journal.pbio.0060309.

Tukey, John W. 1949. "Comparing Individual Means in the Analysis of Variance." *Biometrics* 5(2): 99–114. http://www.jstor.org/stable/3001913.

"UniProt: The Universal Protein Knowledgebase in 2021." 2021. *Nucleic Acids Research* 49(D1): D480–89. https://doi.org/10.1093/nar/gkaa1100.

Vaidya, Gaurav, David J Lohman, and Rudolf Meier. 2011. "SequenceMatrix: Concatenation Software for the Fast Assembly of Multi-Gene Datasets with Character Set and Codon Information." *Cladistics* 27(2): 171–80. http://dx.doi.org/10.1111/j.1096-0031.2010.00329.x.

Vallenet, David et al. 2017. "MicroScope in 2017: An Expanding and Evolving Integrated Resource for Community Expertise of Microbial Genomes." *Nucleic acids research* 45(D1): D517–28.

Veesenmeyer, Jeff L et al. 2014. "NilD CRISPR RNA Contributes to Xenorhabdus Nematophila Colonization of Symbiotic Host Nematodes." *Molecular microbiology* 93(5): 1026–42.

Vivas, E I, and Heidi Goodrich-Blair. 2001. "Xenorhabdus Nematophilus as a Model for Host-Bacterium Interactions: RpoS Is Necessary for Mutualism with Nematodes." *Journal of bacteriology* 183(16): 4687–93.

Wang, Xin et al. 2005. "CsrA Post-Transcriptionally Represses PgaABCD, Responsible for Synthesis of a Biofilm Polysaccharide Adhesin of Escherichia Coli." *Molecular microbiology* 56(6): 1648–63.

Wang, Yan et al. 2016. "Structural Basis for Translocation of a Biofilm-Supporting Exopolysaccharide across the Bacterial Outer Membrane." *The Journal of biological chemistry* 291(19): 10046–57. https://pubmed.ncbi.nlm.nih.gov/26957546.

Ward, Jordan D. 2018. "Spotlight on CRISPR in Strongyloides Parasitic Nematodes." *Trends in parasitology* 34(1): 6–9.

Wasala, Sulochana K et al. 2019. "Variable Abundance and Distribution of Wolbachia and Cardinium Endosymbionts in Plant-Parasitic Nematode Field Populations ." *Frontiers in Microbiology* 10. https://www.frontiersin.org/articles/10.3389/fmicb.2019.00964.

Weber, Jacob J, Michael R Kanost, and Maureen J Gorman. 2020. "Iron Binding and Release Properties of Transferrin-1 from Drosophila Melanogaster and Manduca Sexta: Implications for Insect Iron Homeostasis." *Insect biochemistry and molecular biology* 125: 103438.

Webster, J. M., GenHui Chen, KaiJi Hu, and JianXiong Li. 2002. Entomopathogenic nematology *Bacterial Metabolites*. Wallingford: CABI Publishing.

Wei, Xing, Christopher J Potter, Liqun Luo, and Kang Shen. 2012. "Controlling Gene Expression with the Q Repressible Binary Expression System in Caenorhabditis Elegans." *Nature methods* 9(4): 391–95.

Weis, W I, and K Drickamer. 1996. "Structural Basis of Lectin-Carbohydrate Recognition." *Annual review of biochemistry* 65: 441–73.

Wheeler, Kelsey M et al. 2019. "Mucin Glycans Attenuate the Virulence of Pseudomonas Aeruginosa in Infection." *Nature microbiology* 4(12): 2146–54.

Whitney, J C, and P Lynne Howell. 2013. "Synthase-Dependent Exopolysaccharide Secretion in Gram-Negative Bacteria." *Trends in Microbiology* 21(2): 63–72. https://www.sciencedirect.com/science/article/pii/S0966842X12001771.

Wilkinson, Paul et al. 2009. "Comparative Genomics of the Emerging Human Pathogen Photorhabdus Asymbiotica with the Insect Pathogen Photorhabdus Luminescens." *BMC genomics* 10: 302.

Von Willebrand, E A. 1999. "Hereditary Pseudohaemophilia." *Haemophilia : the official journal of the World Federation of Hemophilia* 5(3): 223–31; discussion 222.

Willias, Stephan P et al. 2015. "CRP-Mediated Carbon Catabolite Regulation of Yersinia Pestis Biofilm Formation Is Enhanced by the Carbon Storage Regulator Protein, CsrA." *PloS one* 10(8): e0135481.

Wilson, Marlena M, and Harris D Bernstein. 2016. "Surface-Exposed Lipoproteins: An Emerging Secretion Phenomenon in Gram-Negative Bacteria." *Trends in microbiology* 24(3): 198–208.

Wilson, Michael J, and Robbie Rae. 2015. "Phasmarhabditis Hermaphrodita as a Control Agent for Slugs." In *Nematode Pathogenesis of Insects and Other Pests: Ecology and Applied Technologies for Sustainable Plant and Crop Protection*, , 509–21.

Winkle, Matthias et al. 2021. "DpaA Detaches Braun's Lipoprotein from Peptidoglycan." *mBio* 12(3).

Xu, J, S Lohrke, I M Hurlbert, and R E Hurlbert. 1989. "Transformation of Xenorhabdus Nematophilus." *Applied and environmental microbiology* 55(4): 806–12.

Xu, J, M E Olson, M L Kahn, and R E Hurlbert. 1991. "Characterization of Tn5-Induced Mutants of Xenorhabdus Nematophilus ATCC 19061." *Applied and environmental microbiology* 57(4): 1173–80.

Xu, Jimin, and Ronald E Hurlbert. 1990. "Toxicity of Irradiated Media for Xenorhabdus Spp." *Applied and Environmental Microbiology* 56(3): 815–18. http://dx.doi.org/10.1128/aem.56.3.815-818.1990.

Yadav, Shruti, Upasana Shokal, Steven Forst, and Ioannis Eleftherianos. 2015. "An Improved Method for Generating Axenic Entomopathogenic Nematodes." *BMC Research Notes* 8(1). http://dx.doi.org/10.1186/s13104-015-1443-y.

Yokoyama, Katsushi et al. 2006. "Feast/Famine Regulatory Proteins (FFRPs): Escherichia Coli Lrp, AsnC and Related Archaeal Transcription Factors." *FEMS microbiology reviews* 30(1): 89–108.

Yukl, Erik T et al. 2010. "Kinetic and Spectroscopic Studies of Hemin Acquisition in the Hemophore HasAp from Pseudomonas Aeruginosa." *Biochemistry* 49(31): 6646–54. https://doi.org/10.1021/bi100692f.

Zallot, Rémi, Nils O Oberg, and John A Gerlt. 2018. "'Democratized' Genomic Enzymology Web Tools for Functional Assignment." *Current Opinion in Chemical Biology* 47: 77–85. http://dx.doi.org/10.1016/j.cbpa.2018.09.009.

———. 2019. "The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways." *Biochemistry* 58(41): 4169–82. http://dx.doi.org/10.1021/acs.biochem.9b00735.

Zeuthen, Louise Hjerrild, Lisbeth Nielsen Fink, and Hanne Frokiaer. 2008. "Epithelial Cells Prime the Immune Response to an Array of Gut-Derived Commensals towards a Tolerogenic Phenotype through Distinct Actions of Thymic Stromal Lymphopoietin and Transforming

Growth Factor-Beta." *Immunology* 123(2): 197–208.

Zhang, Chong-Xing et al. 2009. "Serratia Nematodiphila Sp. Nov., Associated Symbiotically with the Entomopathogenic Nematode Heterorhabditidoides Chongmingensis (Rhabditida: Rhabditidae)." *International journal of systematic and evolutionary microbiology* 59(Pt 7): 1603–8.

Zhang, Fan et al. 2017. "Caenorhabditis Elegans as a Model for Microbiome Research ." *Frontiers in Microbiology* 8. https://www.frontiersin.org/articles/10.3389/fmicb.2017.00485.

Zhang, Hao-Ran et al. 2021. "Roles of Iron Limitation in Phytoplankton Dynamics in the Western and Eastern Subarctic Pacific ." *Frontiers in Marine Science* 8. https://www.frontiersin.org/article/10.3389/fmars.2021.735826.

Zhang, Ke Y et al. 2012. "Heterorhabditidoides Rugaoensis n. Sp. (Rhabditida: Rhabditidae), a Novel Highly Pathogenic Entomopathogenic Nematode Member of Rhabditidae." *Journal of nematology* 44(4): 348–60.

Zhang, Liwen et al. 2016. "Insights into Adaptations to a Near-Obligate Nematode Endoparasitic Lifestyle from the Finished Genome of Drechmeria Coniospora." *Scientific Reports* 6(1): 23122. https://doi.org/10.1038/srep23122.

Zhang, Zheng et al. 2020. "Estimate of the Sequenced Proportion of the Global Prokaryotic Genome." *Microbiome* 8(1): 134.

Zhong, Qifeng, Bostjan Kobe, and Ulrike Kappler. 2020. "Molybdenum Enzymes and How They Support Virulence in Pathogenic Bacteria ." *Frontiers in Microbiology* 11. https://www.frontiersin.org/articles/10.3389/fmicb.2020.615860.

Zhou, Meixian, and Hui Wu. 2009. "Glycosylation and Biogenesis of a Family of Serine-Rich Bacterial Adhesins." *Microbiology (Reading, England)* 155(Pt 2): 317–27.

Zugasti, Olivier, and Jonathan J Ewbank. 2009. "Neuroimmune Regulation of Antimicrobial Peptide Expression by a Noncanonical TGF-Beta Signaling Pathway in Caenorhabditis Elegans Epidermis." *Nature immunology* 10(3): 249–56.

Vita:

Alex Grossman grew up in the town of Wind Gap Pennsylvania, where he attended school at Pen Argyl Area High School. After graduating, Alex moved to Cleveland Ohio in order to attend Case Western Reserve University to study biology, anthropology, and chemistry. His undergraduate honors research was performed in the laboratory of Dr. Michael Benard and focused on the sensory ecology of green frog tadpoles. While studying at Case Western, Alex attended the Microbiology REU program at University of Tennessee Knoxville and immediately became fascinated with the bacterial world and all its mechanistic complexity. Alex chose to attend the University of Tennessee, Knoxville to pursue a Doctor of Philosophy degree in microbiology. He performed his dissertation research in the lab of Dr. Heidi Goodrich-Blair where he worked to characterize the bacterial type eleven secretion system, its diverse cargo proteins, and the roles they both play in facilitating host-microbe symbiosis. He is incredibly grateful to all the mentors and friends who have brought wisdom, knowledge, and happiness into his life.