

**Mechanisms by which *Xenorhabdus*
nematophila interacts with hosts using
integrated -omics approaches**

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

Nicholas C. Mucci
May 2022

Copyright © 2022 by Nicholas C. Mucci

All rights reserved.

ACKNOWLEDGEMENTS

I cannot believe I am writing up the acknowledgements section of my dissertation. The God's honest truth is that I never expected to get this far. When I joined, I kept telling myself to get through another year, then another year, and now I'm here. I am so deeply appreciative of my support systems that pushed me in the right direction all this time.

I am so grateful for the mentorship I received from both professors and colleagues over my time here. My committee members guided me through different technologies and protocols that I did not know existed before I entered the program, and pushed me on the best methods to approach the questions I was interested in. All my lab mates selflessly helped me by offering so many unique perspectives on my work and brought forth good vibes every time we got together. To Terra, thank you for your endless patience when I first joined the lab and pestered you about cloning every day. To Sarah, you always offered to help me prepare samples and you are forever my wet lab buddy. To Alex, oh my God you are the most brilliant peer I've had the pleasure of getting to know. You had the best ideas on so many of my projects here and are genuinely such a cool dude. I cannot wait to see you run a research institute. And to Heidi, there are not enough words to express my gratitude here. Tears are welling in my eyes as I think about how grateful I am to have you guide me through graduate school. From the bottom of my heart, I cannot thank you enough for your mentorship.

My family back in Boston always believed in me. Constant, unwavering support. I always joke that the most my parents know about DNA is the paternity test they administer in Maury. But the truth is they try to understand, and that is all I can ask. Mom, Dad, and Mimi engrained for me from a very young age about the importance of an education. They stressed how powerful a public education can be to transform the world, and that lesson is etched deep in my bones. I want to make a positive impact on this planet so badly. I see such a dark future ahead

and there is nothing I want more in this life to leave the world so much better when I go. I thank my family for listening, their love, and above all else, reminding me to not take life so seriously.

And to my chosen family, my queer family, you mean the world to me. I know I've been difficult at times, graduate school is a monster I've always prioritized and your patience has been so kind. I do not have siblings, but I treat my closest friends like they are mine. To Daniel, the first and closest queer friend I ever made, you are my rock. Through our Miranda/Charlotte relationship, you brought out my authentic self (and ground me). If we never met at UMass orientation week, I know I'd be the type of person we make fun of now. Thank you for your camaraderie. To Steven, my person, my ender pearl. As I have expressed constantly, you more than anyone else pushed me. You forced me to figure out what I wanted in this life. Tough, fair, correct. I consider you one of my life's mentors and I love you deeply. To Elliott (Mary Lou Bottoms), the first real friend I made in Tennessee. They've opened my eyes to the resilience of the queer spirit alive in the south and helped me see beyond my coastal elitism. Their selflessness and grit is woven deeply into their DNA. I hope I made a fraction of the impact on you that you have left on me.

Thanks for the memories.

ABSTRACT

Nearly all organisms exist in proximity to microbes. These microbes perform most of the essential metabolic processes necessary for homeostasis, forming the nearly hidden support system of Earth. Microbial symbiosis, which is defined as the long-term physical association between host and microbes, relies on communication between the microbial community and their host organism. These interactions among higher order organisms (such as animals, plants, and fungi) and their bacteria links metabolic processes between interkingdom consortia. Many questions on microbial behavior within a host remain poorly understood, such as the colonization efficiency among different microbial species, or how environmental context changes their behavior. We can utilize integrated -omics protocols to better characterize these complex interactions from their genetic blueprint to their metabolic functions. Many animal pathogenicity mechanisms are conserved among bacteria species regardless of their host, thus -omics investigations of host-pathogen interactions can be used to better understand the fundamental biology of host-pathogen interactions that subsequently leads to knowledge supporting translational research to mitigate harmful bacterial infections. The focus of this dissertation is to better understand the molecular biology underlying mutualism and antagonism using a simplified animal-microbe symbiosis model that I argue is foundational for understanding larger, more complex microbiome studies. I aim to elucidate the molecular mechanisms that confer the *Xenorhabdus-Steinernema* animal-microbe symbiosis using integrated -omics based approaches. This is a simplified one host, one microbe system with aspects of the lifecycle pertaining to mutualism (*Xenorhabdus* bacteria mutualistic colonization of *Steinernema* nematodes to find prey) and antagonism (*Xenorhabdus* release from *Steinernema* within prey insects to kill them, convert the insect cadaver nutrients, and reproduction of nematodes and bacteria). For this dissertation, I discuss the study of this system as two parts of the *Xenorhabdus* lifecycle. In Chapter 2, I focus on the broad chemical ecology

of the *Xenorhabdus-Steinernema* system by investigating how the nematode-bacterium pair metabolically alters the host chemical environment. In Chapter 3, I seek to better understand the specific biochemical mechanisms by which *Xenorhabdus* bacteria associates with *Steinernema* nematodes, with implications of this research extending to human pathogenic Proteobacteria. As such, the introductory chapter (Chapter 1) was constructed as a foundation for these studies, presenting where the “symbiomics” field has come and where it is likely going.

TABLE OF CONTENTS

CHAPTER 1: The application of multi-omics tools to study animal-microbe symbioses.....	1
Why should we study animal-microbe symbioses?	2
How has the –omics era revolutionized the symbiosis field, and what are the gaps in knowledge that can be filled with more widespread usage of these tools?	3
Microbial behavior and the divergency of microbial lifestyles: what –omics can tell us.....	6
Justification of the <i>Xenorhabdus-Steinernema</i> animal model system to study symbiosis.....	8
Appendix	11
Chapter 1 figures and tables	11
CHAPTER 2: Apex-predator nematodes and meso-predator bacteria consume their basal insect prey through discrete stages of chemical transformations	12
Abstract	14
Introduction	16
Results	19
Trophic analysis reveals <i>S. carpocapsae</i> nematodes directly feed on <i>X. nematophila</i> bacteria.....	19
<i>X. nematophila</i> transcriptional control of metabolic pathways.....	20
The metabolomic profile of EPNB-infected <i>G. mellonella</i> : an overview	23
Multivariate data analysis shows metabolic profile gradient corresponding to infection progression	24
Tricarboxylic acid (TCA) cycle metabolites abundances change significantly throughout the nematode lifecycle	30
Discussion	32
Materials and methods	37
Conventional nematode and aposymbiotic nematode production.....	37
In vitro controlled feeding experiment.....	38
In vivo feeding experiment and sample collection	39
Nematode lyophilization and trophic position analysis.....	40
Preparation for mass spectrometry	41
Metabolomics analysis	42
Bacterial strains, plasmids, and culture conditions	42
Microarray experiment and data analysis	43
Statistical analysis.....	43
Acknowledgments	44
Data and materials availability	44

Appendix	45
Chapter 2 figures and tables	45
Supplementary text (A-D).....	55
Chapter 2 supplementary figures and tables	64
CHAPTER 3: Integrated -omics investigation of the newly characterized Type 11 Secretion System (TXISS), a factor necessary for <i>Xenorhabdus nematophila</i> colonization of <i>Steinernema carpocapsae</i> nematodes	73
Abstract	75
Introduction	76
Results	78
Deletion of <i>nilB</i> and <i>nilC</i> causes global metabolome and proteome changes including impacts on peptidoglycan precursor and exopolysaccharide biosynthesis	78
Establishing a control for domain of unknown function (DUF) gene neighborhood co-occurrence analysis	81
Functional assessment of TXISS co-occurring proteins identifies a role of metal homeostasis, amino sugar metabolism, and DNA binding regulatory proteins	83
Controlled whole genome co-inheritance assay of TXISS homologs.....	84
Bioinformatic investigation of additional TXISS cargo and their structural prediction	88
Discussion	88
Materials and methods	94
<i>Proteomics and metabolomics sample preparation</i>	94
<i>LC-MS/MS and UPLC-HRMS analysis</i>	95
<i>Proteome database search analysis</i>	97
<i>Gene neighborhood and phylogenomics analysis</i>	98
Acknowledgements	100
Data and materials availability	100
Appendix	101
Chapter 3 figures and tables	101
Chapter 3 supplementary figures and tables	112
CHAPTER 4: Summary, conclusions, and future directions	127
REFERENCES	137
APPENDICES.....	150
Appendix 1: Nematocidal activity of a small molecule produced by <i>X. nematophila</i> (GKA-L-Arginine assays).....	151
Introduction	151
Results.....	152

Discussion	162
Materials and methods.....	163
Appendix 2: Summary of findings with Leucine-responsive regulatory protein (Lrp), which acts as a virulence switch in <i>Xenorhabdus nematophila</i>	165
Appendix 3: F1000 Review on “Unravelling the relationship between the tsetse fly and its obligate symbiont <i>Wigglesworthia</i> : transcriptomic and metabolomic landscapes reveal highly integrated physiological networks.”	168
Appendix 4: Manuscript contributions to “ <i>Cylindrospermopsis raciborskii</i> Virus and host: genomic characterization and ecological relevance” as part of my rotation in the Wilhelm Lab working under Dr. Robbie Martin.	169
Appendix 5: Manuscript contributions to “Synergy between T6SSs is an important component of <i>X. nematophila</i> infection of the insect host”	171
Appendix 6: “The value of an integrated statewide environmental report for Tennessee” White Paper summary as part of the Environmental Science and Policy Practicum course	175
Appendix 7: OUTgrads presidency summary.....	176
VITA.....	177

LIST OF TABLES

CHAPTER 2

Table 2.1. Summary of *in vitro* and *in vivo* trophic measurements.....45

Table S2.1: List of strains used in the microarray analysis.....71

Table S2.2: *G. mellonella* weight (g) upon sampling.....72

CHAPTER 3

Table 3.1: Top 50 results from the controlled whole genome co-inheritance assay organized by cellular localization.....107-109

Table S3.1: List of strains used in this chapter.....118

Table S3.2: Significantly differential proteins detected in the whole cell samples ordered by p-value significance (as determined by Students T-tests).....119-122

Table S3.3: Significantly differential proteins detected in the supernatant samples ordered by p-value significance (as determined by Students T-tests).....123-126

LIST OF FIGURES

CHAPTER 1

Figure 1.1: Lifecycle of the entomopathogenic nematode <i>Steinernema carpocapsae</i>	11
---	----

CHAPTER 2

Figure 2.1: Trophic analyses reveal <i>Steinernema</i> nematodes feed on <i>Xenorhabdus</i> bacteria...	46
---	----

Figure 2.2: Differentially expressed transcripts between <i>X. nematophila</i> mutants relative to wild type and their broad functional categorization.....	48
---	----

Figure 2.3: Key moments in the EPNB lifecycle mapped onto important molecules are indicative of the bioconversion of the insect cadaver.....	49
--	----

Figure 2.4: Distinct chemical environments occur during bioconversion of an insect cadaver by <i>S. carpocapsae</i> and <i>X. nematophila</i>	51
---	----

Figure 2.5: Hierarchical clustering analysis of detected metabolites revealed ten clusters, within each of which the metabolites displayed similar rates of change over the infection.....	52
--	----

Figure 2.6: Infection with <i>S. carpocapsae</i> IJs affects insect TCA cycle.....	54
--	----

Figure S2.1: Summary of the trophic study results and protocol.....	64
---	----

Figure S2.2: Venn diagram comparing the differentially expressed transcripts between the 5 microarrays analyzed.....	65
--	----

Figure S2.3: The top 15 VIPs contributing to component 1.....	66
---	----

Figure S2.4: Heatmap of metabolite clusters with pairwise correlation displayed.....	67
--	----

Figure S2.5: Amino acid abundances shift significantly throughout the lifecycle.....	68
--	----

Figure S2.6: PLS-DA plots including the nematode samples.....	70
---	----

CHAPTER 3

Figure 3.1: Differentially translated proteins between $\Delta SR1$ and WT show significant differences in metal binding, homeostasis, and metabolic pathways.....	101
--	-----

Figure 3.2. Predicted pathways for amino sugar metabolism leading to peptidoglycan (PG), exopolysaccharide (poly-N-acetylglucosamine (PNAG)), and lipopolysaccharide LipidA biosynthesis, and glycolysis pathways.....	103
Figure 3.3: Design for a novel control implementation technique for gene neighborhood studies.....	104
Figure 3.4: Functional assessment of protein sequences in the gene neighborhood of TXISS.....	105
Figure 3.5: rJC score distribution of TXISS.....	106
Figure 3.6: Secondary structures of 17 predicted TXISS cargo proteins.....	110
Figure 3.7: Tertiary structures of 17 predicted TXISS cargo proteins.....	111
Figure S3.1: Growth curves of strains used for sampling to conduct metabolome and proteome analyses.....	112
Figure S3.2: Significant metabolome differences between WT and $\Delta SR1$ indicate TXISS role of amino acid and amino sugar metabolism.....	113
Figure S3.3: Top pathways impacted by TXISS using significant ($P < 0.1$) metabolites.....	115
Figure S3.4: TXISS (DUF560) protein size distribution (in amino acids).....	116
Figure S3.5: Statistics from Blast2GO run.....	117
APPENDICES	
Figure A1.1: <i>Steinernema carpocapsae</i> GKA-L-Arginine assays.....	152
Figure A1.2: <i>Steinernema anatoliense</i> GKA-L-Arginine assays.....	154
Figure A1.3: <i>Heterorhabditis bacteriophora</i> GKA-L-Arginine assays.....	157
Figure A1.4: <i>Caenorhabditis elegans</i> GKA-L-Arginine assays.....	160
Figure A2.1: Lrp regulation in <i>E. coli</i> and <i>S. typhimurium</i>	166
Figure A2.2: Preliminary fluorescence data shows possible Lrp acetylation and regulation by Pat and CobB.....	167

Figure A5.1: Identification of candidate T6SS-1 secreted proteins by proteomics comparison of wild type and $\Delta T6SS-1$ supernatants.....173

Figure A5.2: Comparison of supernatant proteomics samples.....174

LIST OF ATTACHMENTS

CHAPTER 2

Trophic study results. Includes raw data, isocline calculations, and trophic level calculations. Also includes notes pertaining to measurements taken (Data S2.1.xlsx)

Microarray results. Includes every comparison between wild type (HGB800 or HGB007) and the mutants (described in Table S1). $2 < |\text{fold change signal strength}|$ between the two strains are shown, as are the signals for each of these genes (Data S2.2.xlsx)

Time course metabolomics known data. Includes the raw and normalized data, statistics on the normalized data (including t-test comparisons between uninfected vs. individual time points, t-test comparisons between uninfected vs. defined time phases, fold change differences, average metabolite abundance for each time point among the replicates taken, the standard deviation for the replicates, the standard error for the replicates, the CV for the replicates, and ANOVA results), and a detailed list of the features included in the hierarchical clustering analysis. The VIP metabolites derived from the PLS-DA plots for both the entire time course, and the entire time course with the addition of the input nematode IJs are also included. Data shown includes VIP scores for every detected metabolite and the top three components. The last three tabs include the time course metabolomics unknown data. Includes a list of the unidentified features, statistics on those features (including average metabolite abundance for each time point among the replicates taken, the standard deviation for the replicates, the standard error for the replicates, and the CV for the replicates), and a summary of those measurements (Data S2.3.xlsx)

Interpreting and analyzing unidentified detected metabolites from the metabolomics analysis. Includes m/z ratios and retention times for each metabolite that was found to have a signal throughout the entire time course, and keys to look up what the features were identified as (Data S2.4.xlsx)

CHAPTER 3

Whole cell proteomics results. Includes fold change and Student's t-test statistics for every protein detected in the analysis (Data S3.1.xlsx)

Supernatant proteomics results. Includes fold change and Student's t-test statistics for every protein detected in the analysis (Data S3.2.xlsx)

Metabolomics results. Includes raw and normalized data, as well as statistics (fold change, averages, standard deviation and error, and t-test results) comparing each strain and fraction (Data S3.3.xlsx)

Gene neighborhood analysis. Includes uncontrolled pfam list and separation of pfams that passed the FDR and pfams that failed the FDR (Data S3.4.xlsx)

Whole genome co-inheritance analysis. Includes uncontrolled pfam list and separation of pfams that were unique to TXISS and LptD, as well as a comparison to the gene neighborhood (Data S3.5.xlsx)

Biological function GO terms of TXISS co-occurring proteins. Darker orange colors correspond to categories that have more sequences with that GO term. Specificity is nested under broader categories at the top of the schematic, where more specific activities are near the bottom and connected to the top categories with lines (Data S3.6.png)

Molecular function GO terms of TXISS co-occurring proteins. Darker orange colors correspond to categories that have more sequences with that GO term. Specificity is nested under broader categories at the top of the schematic, where more specific activities are near the bottom and connected to the top categories with lines (Data S3.7.png)

CHAPTER 1: The application of multi-omics tools to study animal-microbe symbioses

Why should we study animal-microbe symbioses?

Almost all organisms on Earth interact with microbial species, where macroscopic organisms live in close association with microbes externally or internally. These interactions are ancient and have driven the course of life since before the evolution of eukaryotes, where it is theorized that one bacterium engulfed another and compartmentalized metabolic function. This, in turn, created the mitochondria and the chloroplast as bacterial organelles in major evolutionary milestones called endosymbiosis (1). Metabolic compartmentalization is, what many in the scientific community are beginning to appreciate, an incredibly strong evolutionary mechanism that drives speciation. Traditional evolutionary theory states natural selection through genetic mutations is the paramount driver of adaptation. However, as we begin to understand all the roles the estimated 10^{30} microbes on Earth perform, this microscopic support system is shaping the world around us behind the scenes (2). Through symbiosis, hosts acquire the whole genomes of their microbes and rely on them to perform essential metabolic processes (3).

Considering how essential microbes can be for a host's metabolism, it becomes important to consider that higher eukaryotes not only live within ecosystems but that they themselves represent a complex ecosystem. Most animals harbor diverse and dense microbial communities within themselves. In humans, it is estimated that there exists a 1:1 ratio of human cells to microbial cells (4). We are as much of our own life as we are "foreign" life. When the genetic catalogue of these microbes is considered, the ratio of human vs. microbe DNA shifts completely toward the microbes. There are an estimated 1000-1150 microbial species that have been identified in the human gut, with around 160 species being concurrently active in the gut per person (5). Each of these species has their own genetic potential, their own genome capable of performing metabolic tasks their human hosts have evolved to rely on. The Human

Microbiome Project (HMP) estimated 10 million unique microbial genes, a stark contrast to the estimated 30,000 genes found in the Human Genome Project (HGP) (6).

Today, microbes are considered essential for animal health, and microbial communication links metabolism between host and microbe. Microbes send and receive chemical signals to and from their hosts (and to each other). These signals are recognized by the host cells or by surrounding microbes to modulate their metabolic activity. Consequently, host and microbe gene expression are affected to induce a phenotype (7). Co-evolution between host-microbe and microbe-microbe interactions has shaped how hosts and microbes perceive and respond to the presence of one another

How has the –omics era revolutionized the symbiosis field, and what are the gaps in knowledge that can be filled with more widespread usage of these tools?

A major knowledge gap in the DNA-RNA-Protein paradigm is elucidating the role of genes, transcripts, and proteins of unknown function. Apart from playing important structural and metabolic roles in organisms, proteins are capable of perceiving and transmitting chemical signals between hosts and microbes that ultimately shape symbiotic behaviors. Proteins serve as the chemical signaling hubs between microbes associated with their hosts, modulating symbiotic behavior. It is estimated that 23% of the Protein Family (PFam) database are protein domains of unknown function (8). Traditional molecular biology techniques are essential to characterizing these proteins, yet the time it takes to properly assign function without strong hypotheses are extremely time consuming and labor intensive. Comparative genomics techniques can allow functional hypothesis generation based off sequence similarity. Transcriptomics (through RNA sequencing) and proteomics and metabolomics methodology (through the use of mass spectrometry) allow a broader picture on the types of molecules and genes the unknown protein interacts with. When integrated, multi-omics approaches can be

employed to provide multiple lines of evidence that contextualize the importance of a particular genes of unknown function. With these techniques becoming more widely available and utilized, the proverbial baton can be passed to molecular biological research in a quicker and more efficient manner than ever before. For example, shRNA was screened among patients in a genomic oncological study, where a specific tyrosine phosphatase was identified among multi-drug resistant patients as a possible biomarker target. A small molecule inhibitor was used to block the phosphatase, and thus halted growth of the drug-resistant tumors (9). Studies like these show a promising future for multi-omics research and how it will serve as a major contributor to more rapid scientific breakthroughs in biomedical research.

The central dogma of biology states that DNA is transcribed into RNA, which is translated into proteins. These proteins can produce metabolites, which induce phenotypes. Each of these steps from genotype to phenotype is highly regulated, and environmental context determines which parts of this process are activated or deactivated (10). With the development of “-omics” technology, we can probe every part of the central dogma with specific detail. As Araújo *et al.* (2017) describes, genomics can tell us what can happen, transcriptomics tells us what appears to be happening, proteomics can tell us what makes it happen, and metabolomics can tell us what has and is happening (11). Integrating these technologies together can revolutionize the ways we approach personalized health care (12). We can now take chemical inventories of ourselves and the world around us, which can enhance our grasp of animal-microbe symbioses. What genes are encoded by the partners? How are these genes transcribed under different environmental conditions? Are the transcripts encoded by both partners actually producing proteins, and how do they effect the chemical ecosystem they interact in (13)? We know the microbiome to be necessary for digestion and nutrient acquisition, conferring immunity to pathogens, and can even affect behavior. However, with all the diversity between human microbiomes, understanding how the microbiome modulates the human host phenotypes is incredibly difficult without effective integration of -omics technologies. We can

address these questions in a revolutionary way now in the 21st century, and it is an exciting time to be a life scientist.

While a major limitation to the use of -omics technologies has been their cost, each year high throughput DNA, RNA, protein, and metabolite measurements are getting cheaper and more accurate (14). As these measurements become more available, there is an apparent need for bioinformatic pipelines that allow life scientists to access, analyze, and visualize the substantial amount of data generated by -omics-based studies. Currently, all -omics datasets require their own independent bioinformatic pipelines for data processing prior to their integration with each other, which is a procedure warranting a separate tool (14). If -omic studies are designed and analyzed with consideration for future integration efforts, a more complete and holistic perspective of the biochemistry underpinning our natural world will become more accessible and that will propel basic science research into bold future directions.

Databases are being built to integrate multi-omics work more efficiently. Many bioinformatic pipelines exist on GitHub written in various coding languages such as Python, Perl, R, and MySQL. These pipelines can be difficult to navigate and execute for inexperienced users new to bioinformatic research. However, bioinformaticians are making integrating these pipelines simpler to use. KBase is a new software suite that allows scientists to easily collaborate on systems biology research, with a major focus on data integration from -omics sciences (15). The Global Natural Products Social Molecular Networking (GNPS) server is another prime example of an open-access, collaborative platform set on combining the research communities' mass spectrometry data to enhance natural product discovery (16). MetaboAnalyst 5.0 is a web server that can integrate a user genomic and metabolomic information and cross reference their data with the Kyoto Encyclopedia of Genes and Genomes (KEGG) and perform an array of statistical tests in a user-friendly manner (17). And KEGG itself is a massive repository capable of performing pathway mapping using a user's protein, metabolite, and/or transcript data (with many organism's genomic data ready and available),

allowing quick visualization of metabolism (18). Specific databases also exist for certain organisms, like Trematode.net and Nematode.net that serve as repositories for integrating sequencing data and functional annotation software for helminths (19). Together, -omics researchers are aiming to make their data and analytical software more accessible across disciplines.

Integrating multi-omics tools to study microbial symbioses has become increasingly common in the 2010s (20). The most common type of multi-omics analyses has been combining transcriptomics, genomics, and proteomics, although integrating metabolomics and epigenetic modification data has become increasingly popular (21). Mammalian species dominate the integrated -omics fields, with humans, mice, rats, and cattle among the top candidates for these studies. There is a dearth of studies involving species with less complex physiologies and microbiomes. These mammalian host microbiomes are diverse and interactions between microbes and their hosts are difficult to disentangle (22). Moreover, mammalian gut microbiomes often diverge greatly between individuals in a population, making understanding questions on what members are essential and crosstalk between members difficult to extricate (23). I argue in this dissertation that research on more simple animal models can provide valuable insight to the broader body of knowledge on host-microbe interactions, focusing on why certain molecules and genes are required for the microbial colonization and pathogenicity of hosts.

Microbial behavior and the divergency of microbial lifestyles: what – omics can tell us

Microbial behavior can shift depending on environmental context, where gene expression can induce rapid and dramatic changes in phenotype (24, 25). In most cases, production of virulence factors (e.g., toxins) or colonization factors (e.g., biofilms) are metabolically costly for bacteria, and expression is induced only when absolutely necessary to

save energy for the organism (26). Understanding how opportunistic pathogenic bacteria regulate these factors remains an important task for the study of human disease.

Integrated -omics technology can shed light on how and when in an organism's lifecycle it can activate colonization or pathogenicity mechanisms (27). As reviewed by Chaston and Douglas (2012), studies in the past decade have begun using these technologies to their fullest potential to construct testable hypotheses on the specificity and timing of host-associated microbial behavior (13). For example, transcriptomics and metabolomics were employed to investigate *Euprymna scolopes* squid host and *Vibrio fischeri* bacterial symbiont gene expression and their effects on microbial population size and host tissue localization. Transcriptomics enabled observations of the combined (host and microbe) patterns of global gene expression and revealed that these dramatically change depending on the time of day. At dawn (the time when most symbionts are expelled by the squid into water), host cytoskeletal genes and symbiont anaerobic respiration genes are increased relative to the pre-dawn levels. By nightfall, after the remaining bacteria have proliferated in the squid during the day, these host cytoskeletal and bacterial respiration genes decreased back to the pre-dawn levels (28). Importantly, the transcriptomics-based hypothesis that microbial respiration is elevated in the day relative to the night was further supported by metabolomics studies, in which mass spectrometry-based measurements on bacterial membrane lipids revealed a direct integration of host-derived lipids. Together, these data demonstrate that the remaining bacteria that were not expelled were incorporating host fatty acids coinciding with the morphological change of host tissues revealing coordination of symbiont respiration and host cell membrane reorganization to achieve optimal microbial population control.

In another example, Fukada *et al.* (2011) utilized metabolomics and genomics to study mammalian host defense responses against infectious disease (29). In a murine model co-inoculated with either *Bifidobacterium longum* or *B. adolescentis* (which are common Gram-positive mammalian microbiome species, including humans) and a Shiga-toxin producing

Escherichia coli strain, the mice with *B. adolescentis* died. Nuclear magnetic resonance (NMR)-based metabolomics of murine feces revealed that the surviving *B. longum* mice had a higher concentration of acetic acid relative to *B. adolescentis*. The hypothesis derived from these -omics studies, which was supported by experimental data, was that the acetic acid provides protection against the lethal Shiga toxin produced by *E. coli*. Genomics techniques were then used to determine and experimentally test which *Bifidobacterium* species can produce enough acetate to counter Shiga toxin.

Bacterial behavior (including the ability to colonize hosts) is modulated by environmental factors, including the presence of other community members, host-circadian rhythms, and environmental temperature. The examples noted above highlight the value of combined -omics technologies to help identify and parse the myriad parameters that influence host-microbe interactions, laying the foundation and guiding experimentation to assess the fundamental biology.

Justification of the *Xenorhabdus-Steinernema* animal model system to study symbiosis

It is impossible to fully understand anything in the natural world without considering how microbes interact with hosts. Nearly every aspect of biology relies on microbial signaling. The goal of this dissertation research was to better understand the molecular basis of how symbioses are established and maintained, with a focus on using recently developed multi-omics tools. With these technologies and approaches, we have the opportunity to better understand fundamental principles shaping animal-microbe symbiosis. Application of these tools, particularly in combination, to a broad range of biological systems is rapidly expanding, and the methods used to interpret -omics data in a biological context are still being developed. I opted to use the animal-microbe symbiosis between *Steinernema carpocapsae* nematodes and their obligate mutualist gut bacterium *Xenorhabdus nematophila* to ask broad questions

pertaining to symbiotic establishment using multi-omics tools. How does microbial behavior change over time depending on environmental context? What molecular mechanisms are involved in certain microbes ability to colonize a host while others cannot?

The tripartite relationship between the nematode host *S. carpocapsae*, its obligate mutualist bacterium, *X. nematophila*, and the insects they together infect offers an experimentally tractable system to study the chemical environment of an ecosystem with both antagonistic and mutualistic behaviors (30, 31). The soil-dwelling infective juvenile (IJ) stage nematodes will infect an insect by crawling through natural openings (Fig 1.1, Chapter 1 Appendix). Early infection, which ends when the insect is dead (as early as 24 hours post-infection), involves nematode infection of insect hemocoel and release of bacteria, which multiply rapidly and produce toxins to kill the insect and suppress the insect immune response. Middle infection, taking place between day 2 and day 8 post-infection, is characterized by *X. nematophila* converting and protecting the carcass into a nutrient dense medium to support nematode growth and development. Late infection, taking place between day 10 and day 16 post-infection, is characterized by nutrient depletion, where *X. nematophila* will colonize a new generation of IJ nematodes in search of more insect prey.

The *Xenorhabdus-Steinernema* pair is a binary interaction between two organisms, whereas many animals (including the nematode model *Caenorhabditis elegans*) interact with a consortium, often non-specific, comprising multiple bacterial species (32). With only two partners, the *Xenorhabdus-Steinernema* system is a very specific interaction, allowing molecular and chemical communications to be examined between just the pair. Another benefit is the genetic tractability of *Xenorhabdus* species where genes can have their activity altered to study bacterial phenotypes. As for genome availability, 40 sequenced bacterial genomes available on NCBI. *X. nematophila* molecular biology and behavior has been studied in depth with many publications utilizing mutant strains, so there is already a community of knowledge available (31, 33-36). However, questions remain that can be addressed using -omics

technologies, like how the bacteria convert the cadaver into a nutrient source, and what are the molecular requirements for the bacteria to colonize the nematode host? Recently, progress has been made to genetically manipulate the *Steinernema* animal host (37). The *Xenorhabdus*-*Steinernema* partners each can be cultured separately from each other, which allows us to associate varied bacterial strains at specific nematode life stages of interest. Both can grow quickly; *Xenorhabdus* can grow to stationary phase overnight in liquid culture (which is faster than *E. coli* (38)) and *Steinernema* eggs typically take less than a week to prepare (39). The nematode's transparency makes them (and their bacterial symbionts) easy to visualize using microscopy, and their small size allows living tissue observations to be made vs. fixed tissue assays common in other animal tissue observations (39). Nematoda constitute the most abundant phylum of animals on Earth, with around 80% of animals being nematodes, which makes studying them even more relevant as nematode-bacterial symbiosis is occurring over nearly the entire natural world and thus heavily impacts nutrient cycling globally (40). As a final point on the pair, *Steinernema* nematodes can be bought commercially and applied to agriculture. They parasitize and kill various garden pests (woodboring beetles, wax moths, weevils, worm species) and act as a biocontrol, which could be more beneficial for plant health than applying chemical pesticides (41, 42). They can also inhibit insect vectors for diseases like Zika, West Nile virus, and yellow fever (43, 44). Thus, understanding the extent of the pair's parasitic interactions with these pests economically relevant.

This model is an excellent candidate for understanding animal-microbe interactions and I use it to address questions relating to how life recycles unto itself under parasitism and the how bacteria colonize their animal hosts.

Appendix

Chapter 1 figures and tables

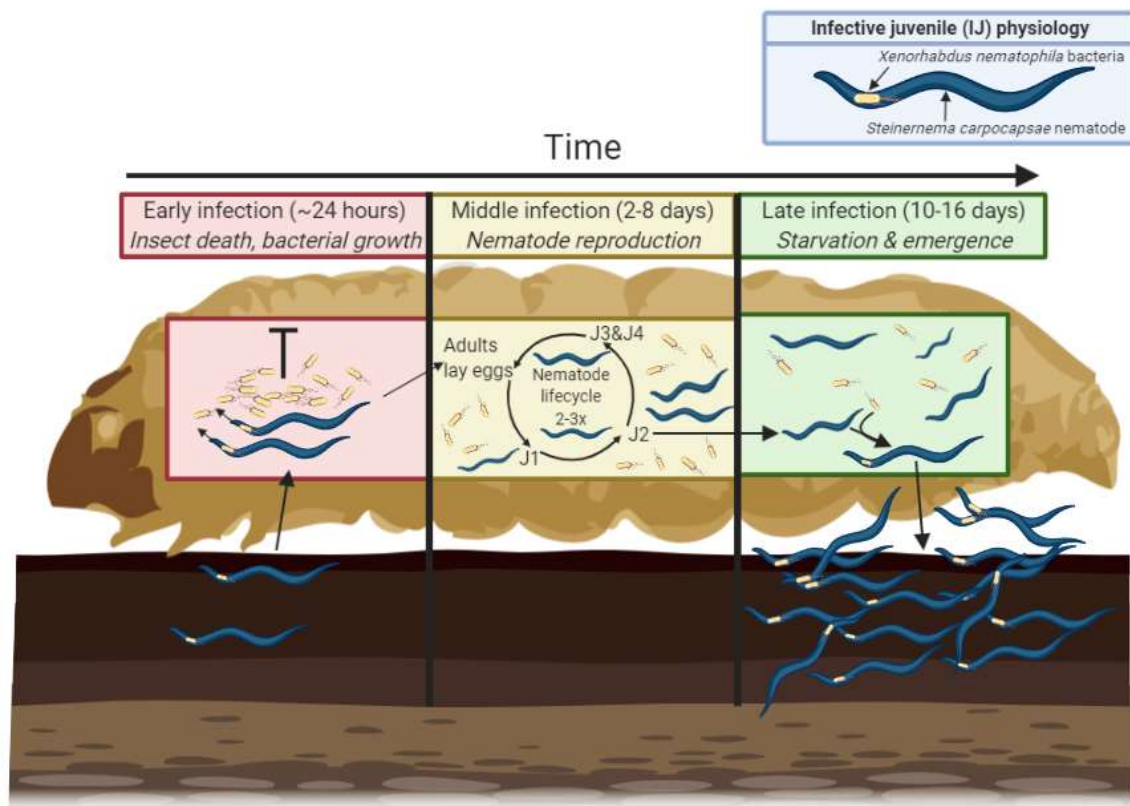


Figure 1.1: Lifecycle of the entomopathogenic nematode *Steinernema carpocapsae*.

Infective juvenile (IJ) physiology shown in top right. The 3 main phases of the lifecycle within the insect are shown in different colors, where the early infection stage is red, the middle infection stage is yellow, and the late infection stage is green. Created with BioRender.com.

CHAPTER 2: Apex-predator nematodes and meso-predator bacteria consume their basal insect prey through discrete stages of chemical transformations

This chapter will be submitted as a manuscript for publication as:

Apex-predator nematodes and meso-predator bacteria consume their basal insect prey through discrete stages of chemical transformations

Nicholas C. Mucci¹, Katarina A. Jones², Mengyi Cao³, Michael R. Wyatt II⁴, Shane Foye⁵, Sarah Kauffman¹, Michela Taufer⁴, Yoshito Chikaraishi⁶, Shawn Steffan^{5, 7}, Shawn Campagna^{2, 8}, Heidi Goodrich-Blair^{1, 3*}

¹Department of Microbiology, University of Tennessee-Knoxville, Knoxville, TN 37996, USA.

²Department of Chemistry, University of Tennessee-Knoxville, Knoxville, TN 37996, USA.

³Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706, USA.

⁴Department of Electrical Engineering and Computer Science, University of Tennessee-Knoxville, Knoxville, TN 37996, USA.

⁵Department of Entomology, University of Wisconsin-Madison, Madison, WI 53706, USA.

⁶Department of Biogeochemistry, Japan Agency for Marine-Earth Science and Technology, Yokosuka 237-0061, Japan.

⁷US Department of Agriculture, Agricultural Research Service, Madison, WI 53706, USA.

⁸Biological and Small Molecule Mass Spectrometry Core, University of Tennessee-Knoxville, Knoxville, TN 37996, USA.

*Corresponding author: Heidi Goodrich-Blair, hgblair@utk.edu (HGB)

Keywords: Animal-microbe symbiosis, metabolomics, transcriptomics, trophic hierarchies, interkingdom interactions, food web

Author contributions for the work presented in this chapter:

Conceptualization: NCM, MC, MRW, MT, SS, SC, HGB

Data Curation: NCM, KAJ, MC, MRW, YC, SS, SC

Formal Analysis: NCM, KAJ, MC, MRW, SF, MT, YC, SS, SC, HGB

Funding Acquisition: MC, HGB

Investigation: NCM, KAJ, MC, MRW, SF, SK, YC, SS

Methodology: NCM, KAJ, MC, MRW, SF, SK, MT, YC, SS, SC, HGB

Project Administration: NCM, MT, SS, SC, HGB

Resources: YC, MT, SS, SC, HGB

Software: NCM, KAJ, MRW, MT, SC

Supervision: MT, YC, SS, SC, HGB

Validation: MT, YC, SS, SC, HGB

Visualization: NCM, KAJ, MRW, MT, SS, HGB

Writing- Original Draft Preparation: NCM, KAJ, MC, SS, HGB

Writing- Review and Editing: NCM, KAJ, MC, MRW, SF, SK, MT, SS, SC, HGB

MC, SF, YC, and SS primarily focused on the trophic study aspects of the chapter, to which NCM did not contribute analysis for.

SK performed the sample preparation for the metabolomics. KAJ and SC processed the metabolomics samples and assisted with their analysis with NCM.

MRW and MT designed the hierarchical clustering analysis, conceptualized and designed with NCM.

Abstract

Microbial symbiosis drives physiological processes of higher order systems, including the acquisition and consumption of nutrients that support symbiotic partner reproduction. Metabolic analytics provide new avenues to examine how chemical ecology, or the conversion of existing biomass to new forms, changes over a symbiotic lifecycle. We applied these approaches to the nematode *Steinernema carpocapsae*, its mutualist bacterium, *Xenorhabdus nematophila*, and the insects they infect. The nematode-bacterium pair infects, kills, and reproduces in an insect until nutrients are depleted. To begin to understand the conversion of insect biomass over time into either nematode or bacterium biomass, we integrated information from trophic, metabolomics, and gene regulation analyses. Trophic analysis established bacteria as meso-predators and primary insect consumers. Nematodes hold a trophic position of 4.37, indicative of an apex predator, consuming bacteria and likely also other nematodes. Metabolic changes associated with bioconversion of *Galleria mellonella* insects were assessed using multivariate statistical analyses of metabolomics datasets derived from sampling over an infection time course. Statistically significant, discrete phases were distinguishable from each other, indicating the insect chemical environment changes reproducibly during bioconversion. Tricarboxylic acid cycle components and amino acids such as proline and leucine were significantly affected throughout the infection. A novel hierarchical clustering method was designed to probe molecular abundance fluctuation patterns over time, revealing distinct metabolite clusters that exhibit similar abundance shifts across the time course. Identification of these metabolites reveals the chemistry underlying the recycling of organic material during carnivory.

Importance

The processes by which organic life is consumed and reborn in a complex ecosystem were investigated through a multi-omics approach applied to the tripartite *Xenorhabdus* bacteria-

Steinernema nematode-*Galleria* insect symbiosis. Trophic analyses demonstrate the primary consumers of the insect are the bacteria, and in turn the nematode consumes the bacteria. This suggests the *Steinernema-Xenorhabdus* mutualism is a form of agriculture in which the nematode cultivates their bacterial food source by inoculating them into insect hosts.

Metabolomics analysis revealed a shift in biological material throughout progression of the lifecycle: active infection, insect death, and conversion of cadaver tissues into bacterial biomass and nematode tissue. We show that each phase of the lifecycle is metabolically distinct, with significant differences in tricarboxylic acid cycle and amino acid metabolism. Our findings demonstrate that symbiotic lifecycles can be defined by reproducible stage-specific chemical signatures, enhancing our broad understanding of metabolic processes that underpin a three-way symbiosis.

Introduction

Symbiotic interactions are ubiquitous in biological systems and have shaped the evolution of life (45). These long-term, intimate associations are driven by small molecule signaling between partners. Bacterial populations establish diverse and expansive metabolite-mediated signaling networks that control gene expression and downstream behaviors, such as biofilm formation and the production of host-interacting effectors (13, 46, 47). A common mechanism by which bacteria sense and transduce metabolic signals is through transcription factors whose DNA binding affinity or specificity is modulated by binding metabolite ligands. For instance, LysR-type transcription factors, which are conserved across proteobacteria, are characterized by a conserved N-terminal DNA-binding domain and a C-terminal domain that varies among LysR-type regulator homologs. The latter domain is responsible for ligand metabolite binding and dictates the response specificity of the transcription factor (48, 49). In a range of bacteria, LysR-type regulators modulate various phenotypes, including virulence, nutrient uptake and metabolic homeostasis, motility, quorum sensing, and antibiotic resistance (50). Another diverse family of transcription factors, feast/famine regulatory factors like leucine-responsive regulatory protein (Lrp), can detect nutrient levels by binding amino acids, which trigger Lrp multimerization and consequent changes in global transcriptional patterns (25).

Given the key function of metabolites in communicating information about intracellular and extracellular environmental conditions, examining their identities and abundances is critical to understanding biological systems. Metabolomics has enabled such studies and has been used to detect specific small molecules that drive essential cellular processes and inter-kingdom signaling (51, 52). Further, it is being applied to more complex ecosystems comprising multi-species microbiota colonizing a host (53). However, to date such studies primarily have been focused on binary conditional comparisons between treatments, or on single, snapshot sampling of complex interactions. Here, to gain insights into temporal changes in metabolic

pathways that occur in complex ecosystems, a longitudinal analysis of metabolic profiles was conducted in a closed ecosystem in which biomass is reproducibly converted from one type of living organism to another. The closed ecosystem comprised an individual insect infected with an entomopathogenic nematode (EPN) and bacterium (EPB) pair.

EPNs of the genera *Steinernema* and *Heterorhabditis* associate with mutualistic EPB in the genera *Xenorhabdus* and *Photorhabdus*, respectively. An infective juvenile (IJ) stage of EPN carry their mutualistic bacteria in their intestine as they dwell in the soil seeking insect hosts to infect. Upon infection, the bacteria are released into the insect blood cavity and together the nematode and bacterium kill and consume the insect for their own reproduction before developing into the bacteria-colonized infective stage again to repeat the cycle (30, 31). Entomopathogenic nematodes and bacteria (EPNB) have been applied as insecticide alternatives to promote agricultural productivity and to help prevent transmission of insect diseases like dengue and West Nile virus (43, 44).

Together, EPNB reproduce using the nutrients available from the insect carcass, but the details of this decomposition process are unknown. For instance, do the nematodes and bacteria both directly consume the insect, or do they form a consumption hierarchy? This information is essential to fully integrate 'omics information available for members of a community toward assigning their function and temporal dynamics within the complex ecosystem. Recent efforts to calculate trophic positions (TP) of microbes and integrate microbial processes within food webs has opened new avenues for addressing such questions. TP estimates are derived as the nitrogen isotopic ratio between glutamic acid and phenylalanine ($TP_{\text{glu-phe}}$). Glutamic acid becomes increasingly enriched in ^{15}N as it moves up the food chain, while phenylalanine ^{15}N enrichment is not affected by food chain positioning (54). The consumption hierarchy, or food web, of an ecosystem can be revealed by calculating $TP_{\text{glu-phe}}$ for individual community members.

In this study, consumption and bioconversion of the insect *Galleria mellonella* by the EPNB pair, *Steinernema carpocapsae* and its mutualistic bacterial symbiont, *Xenorhabdus nematophila* was examined from a metabolic perspective. *G. mellonella* is used for laboratory isolation and propagation of EPNB, is a model host to understand virulence of a variety of microbial pathogens, and has a characterized metabolome (55). The *S. carpocapsae*-*X. nematophila* pair was chosen due to the wealth of information available about them from molecular, cellular, and genetic studies (31). For example, it is known that *X. nematophila* bacterial effectors and natural products suppress insect immunity, kill insect blood cells, degrade insect tissues, and defend the insect cadaver from opportunistic competitors (56). Also, *X. nematophila* bacteria are essential for *S. carpocapsae* reproduction; in the absence of bacteria fewer nematode IJs emerge from insect cadavers after reproduction (57). Expression of effectors and physiological adaptation to changing host environments is controlled in *X. nematophila* by transcriptional regulators that are predicted to sense and respond to prevailing metabolic conditions (58). For instance, the LysR-type regulator LrhA that is necessary for *X. nematophila* virulence and controls expression of an extracellular phospholipase that is necessary for insect degradation (30, 58, 59) and the sigma factor RpoS that is necessary for colonizing the IJ stage of the nematode (60). The two-component system CpxRA and the leucine-responsive regulatory protein Lrp are both necessary for normal virulence and mutualism behaviors (56, 61, 62). NilR, a lambda-like repressor family transcription factor negatively regulates genes necessary for nematode colonization (63) and the two-component system OmpR/EnvZ negatively controls *X. nematophila* swarming motility behavior and exoenzyme production (64). Further, *X. nematophila* displays phenotypic heterogeneity with respect to behaviors important for adaptation to host environments. For instance, “primary form” [1°] *X. nematophila* can be distinguished from “secondary form” [2°] by its motility, antibiotic and natural products secretion, and hemolytic and lipolytic activities and additional phenotypic variants arise over the course of insect infection (65, 66).

The goal of this study was to begin to understand the overall metabolic transformations, or bioconversion processes occurring within a closed yet complex biological ecosystem. ^{15}N isotopic enrichment analyses were performed to establish the relative trophic positions of the insect, *G. mellonella*, the nematode, *S. carpocapsae*, and the bacterium, *X. nematophila*, so that the relative roles in bioconversion of each ecosystem member could be established. Then, a metabolomics analysis using an ultra-high performance liquid chromatography high-resolution mass spectrometry (UHPLC-HRMS) metabolomics technique was conducted over a 16-day time course after *S. carpocapsae*-*X. nematophila* infection of *G. mellonella*, encompassing a complete bioconversion of insect tissues to the bacterial-colonized progeny IJs that emerged from the insect.

Results

Trophic analysis reveals *S. carpocapsae* nematodes directly feed on *X. nematophila* bacteria

The trophic identities of the entomopathogenic nematode (*Steinernema carpocapsae*), its bacterial symbiont (*Xenorhabdus nematophila*), and their host insect were measured empirically based on ^{15}N isotopic enrichment of amino acids (Table 2.1, Fig S2.1, and Data S2.1). We first established that the degree of ^{15}N -enrichment between the consumers (nematodes, bacteria) and their respective diets (e.g., agar growth media, bacteria, or the insect) is consistent with past studies of inter-trophic enrichment (Fig 2.1A, Fig S2.1, Supplementary Text 2A). With these data in hand, we next conducted *in vivo* trials involving insect cadavers (Fig 2.1B, Table 2.1, Fig S2.1, and Data S2.1). We found that the mean $\text{TP}_{\text{glu-phe}}$ of an uncolonized insect cadaver was 2.2 ± 0.02 ($N = 9$). When the insect was colonized by bacteria alone, the $\text{TP}_{\text{glu-phe}}$ of the insect-bacteria complex was 2.5 ± 0.03 ($N = 3$). This complex represented the blending of consumer and diet [as described in (67)], wherein the consumer (i.e., the *Xenorhabdus* bacterial population) was suffused within and throughout its diet (the

insect cadaver). Given that both the bacterial and insect biomass were available within the cadaver, this established the basis for the question as to what a developing nematode would consume/assimilate within the cadaver. The diet of the nematodes (i.e., the insect-bacterial complex) was measured at ~ 2.5 . Thus, if the nematodes within the cadaver fed randomly on all available substrates, the nematode $TP_{\text{glu-phe}}$ would be expected to be ~ 3.5 (i.e., $\sim 2.5 + 1.0$). However, the infective juveniles emerging from the cadavers registered a $TP_{\text{glu-phe}}$ of 4.6 ± 0.08 ($N = 5$), a full trophic level higher than expected, indicating that the *Steinernema* nematodes largely consume their mutualistic bacteria as they develop within an insect host.

Within developing nematodes, *Xenorhabdus* bacteria colonize the anterior intestinal caecum (25). To determine if this colonization influences the ability of *Steinernema* nematodes to consume *Xenorhabdus*, the $TP_{\text{glu-phe}}$ of adult *Steinernema* cultivated on lawns of either wild type (WT) *Xenorhabdus*, or a non-colonizing mutant ($\Delta SR1$) was assessed (68). Nematodes had approximately the same $TP_{\text{glu-phe}}$ regardless of the colonization proficiency of the bacterial diet ($TP_{\text{glu-ph}}$ of WT and $\Delta SR1$ were 2.9 ± 0.02 and 2.9 ± 0.04 , respectively) indicating that colonization of the anterior intestinal caecum is not required for nematode feeding on symbiotic bacteria (Data S2.1).

***X. nematophila* transcriptional control of metabolic pathways**

The trophic analyses described above establish *X. nematophila* bacteria as the linchpin organism in the closed ecosystem, responsible for direct consumption of the insect tissue and serving as a primary food source for its mutualistic host *S. carpocapsae*. To gain insights into the metabolic pathways utilized by *X. nematophila* in performance of these functions, the global regulons of several transcription factors were identified using an exploratory microarray analysis, portions of which have been reported elsewhere (Fig 2.2, Table S2.1, and Data S2.2) (69-71). Microarray analyses were conducted on mutants lacking genes encoding the transcription factors LrhA, RpoS, NilR, and Lrp, each of which has a defect in one or more

aspects of the *X. nematophila* lifecycle (58, 60, 62, 63). In addition, since the primary to secondary form phenotypic variation globally influences host-interaction phenotypes, the transcriptional profiles of these variants were examined and compared from a metabolic perspective (65, 66). The mutant and secondary form cells were each compared to their wild type parent or primary form, respectively, using a $2 < |\text{signal fold change}|$ significance cutoff for differences in transcript levels.

The number of differential transcripts is a fraction of the 3733 averaged total expressed chromosome ORFs among the strains. The number of genes with differential transcript abundance in wild type/mutant (or primary/secondary) comparisons was highest in the comparison of wild type (primary) with $\Delta lrpA$ at 396 genes (10.6%), followed by comparisons with Δlrp , secondary form, $\Delta rpoS$, and $\Delta nilR$ with 273 (7.3%), 159 (4.3%), 157 (4.2%), and 83 (2.2%) differentially abundant transcripts, respectively (Fig 2.2A). The proportion of genes categorized as being involved in metabolic activity varied amongst the strains. Through KEGG annotation, the highest proportion of differentially expressed genes categorized as metabolic was observed in the $\Delta rpoS$ strain, at 38.9%, with $\Delta lrpA$, Δlrp , secondary form, and $\Delta nilR$ having metabolic-related activities at 18.9%, 13.6%, 10.7%, and 9.8%, of the differentially expressed genes in the respective strain. Differential transcript overlap was observed while comparing the 5 strains (Fig S2.2). One of the largest overlaps between 2 strains was for the secondary form and Δlrp mutant, which overlap xenocoumacin biosynthesis genes *xcnB* and *xcnL*. These genes encode non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) that produce the xenocoumacin compound, which could play a role in interspecies competition in the cadaver (64). *X. nematophila* lacking *lrp* are phenotypically secondary form (62). Another large overlap was observed between Δlrp and $\Delta nilR$, which synergistically repress nematode colonization (63). The overlap in regulation includes a phosphotransferase sugar uptake system (XNC1_2826-2828) and components of a CRISPR system (XNC1_3717-3719, 3724). In some

X. nematophila strains, disruption of a specific CRISPR sRNA locus causes a nematode colonization defect (71, 72).

To assess the metabolic roles these differential transcripts play in insect tissue bioconversion, functional analysis of the transcripts was performed with KEGG annotation using BlastKOALA (KEGG Orthology and Links Annotation). Sequences were aligned against a nonredundant set of prokaryotic KEGG genes using BLAST searches (18). Consistent with the KEGG annotation analysis noted above, of the strains tested, $\Delta rpoS$, $\Delta lrhA$, Δlrp were the most strongly impacted with respect to metabolic pathway transcripts relative to the other mutant strains tested. These three mutants displayed differences from wild type in carbohydrate and amino acid metabolic regulation (Fig 2A). Branches of carbohydrate metabolism, like propanoate, pentose and glucuronate, and glyoxylate metabolism were impacted by the $\Delta lrhA$ and $\Delta rpoS$ mutations (Fig 2.2B). Inositol phosphate metabolism was impacted in $\Delta rpoS$ and Δlrp strains, indicating a possible role of inositol phosphate in nematode colonization. Pyruvate metabolism was impacted in the $\Delta lrhA$ strain, providing insights into metabolites necessary for pathogenicity. Butanoate metabolism, a branch of carbohydrate metabolism where the amino acid ornithine is converted into short-chain fatty acids, was commonly impacted for all mutant strains except $\Delta nilR$. Amino acid biosynthetic pathway transcripts were differently impacted among these three strains, with tyrosine, and the alanine, aspartate, and glutamate biosynthetic pathways similarly impacted. Histidine and valine metabolism was uniquely altered by the $\Delta rpoS$ mutation, glycine, serine, and threonine metabolism was uniquely altered by the $\Delta lrhA$ mutation, and phenylalanine metabolism was uniquely altered by the Δlrp mutation. *rpoS* is necessary for colonization, *lrhA* is necessary for virulence, and *lrp* impacts both phenotypes. These amino acid transcripts provide insight into the types of pathways utilized to achieve what occurs biologically. To investigate the impacts these pathways and others have on the *Xenorhabdus*-

Steinernema lifecycle, a time course metabolomics experiment was designed to measure the relative quantities of metabolites within them.

The metabolomic profile of EPNB-infected *G. mellonella*: an overview

Having established that *X. nematophila* bacteria consume infected insect tissue, and in turn the bacteria are consumed by reproducing and developing nematodes, the temporal dynamics of metabolic profiles associated with these processes were examined. *G. mellonella* were infected with *S. carpocapsae* infective juveniles colonized by *X. nematophila* bacteria. Weights of the whole insect samples were relatively similar (Table S2.2). Insects were sampled over a 16-day time course. As expected, insects began to die by Hour 24 after infection, and both living and dead insects were sampled at that time point. Within this 24 hour time frame *X. nematophila* begins by suppressing the insect immune response, after which it releases toxins and begins to reproduce rapidly. By the end of the first 24 hours post-infection all insects had succumbed to infection and were dead. Consistent with the initial degradation of the insect cadaver by bacteria predicted by the trophic analysis conducted above, the nematodes began to reproduce by Day 4, 3 days after the insects had died from infection (Fig 3). At Day 7 post-infection, insect cadavers were placed in a collection trap to encourage the emergence of progeny *S. carpocapsae* IJs. Adults and IJs were observed at this stage. This is when the 2nd generation of nematodes begin to emerge, and endotokia matricida occurs, in which some progeny hatch within and consume the mother. By Day 16 the insect cadavers were largely consumed, and most remaining IJs exited. The proportions of nematodes observed at these stages are consistent with previous observations, in which there are more adults and juveniles during the middle phase compared to IJs, followed by high numbers of IJs in the late phase (25). Based on these observations, we divided the metabolomics samples into 3 major time frames: 1) early infection, characterized by killing of the insect host and bacterial replication, 2)

middle infection, characterized by nematode reproduction and nutrient conversion of the cadaver, and 3) late infection, characterized by nutrient depletion and IJ emergence.

Metabolites were extracted from individual insects sampled over the infection time course and analyzed using an untargeted UHPLC-HRMS method. The untargeted metabolic profiling analysis revealed 13,748 spectral features. Through the mass spectrometric measurements, a total of 170 of these features were identified based on comparison to known exact mass-to-charge (m/z) ratio and retention times from a database of central energy metabolites (Data S2.3). Another 3,138 unidentified spectral features were included in the analysis and putatively annotated based on their exact masses compared to a *Xenorhabdus* secondary metabolite database (Data S2.3). This database serves as a rich repository to explore secondary metabolite temporal abundance shifts in the tripartite ecosystem of insect, bacteria, and nematode. Retention times and m/z ratios were run through the Metabolite Annotation and Gene Integration (MAGI) software (Supplementary Text 2B, Data S2.4) for a first run through of feature identification.

Multivariate data analysis shows metabolic profile gradient corresponding to infection progression

Partial least squares-discriminant analysis (PLS-DA) was performed to observe gross chemical environment changes over time of insect bioconversion to nematode-bacterium complex, when combining all detected metabolite data. A three-dimensional PLS-DA plot showed a progression of distinct metabolic profiles from uninfected insects (black circles) to insects in which bacteria and nematodes are reproducing (red and yellow gradients), and finally to fully consumed insects (green gradients) from which bacterial-colonized infective juvenile populations are emerging (Fig 2.4). Component 1 is 41.1% and contributes the most significantly to the separation observed in the PLS-DA plot.

To examine which metabolites are responsible for most of the variation represented by the PLS plots, Variable Importance in Projection (VIP) values for the top 3 components were calculated. VIP is a weighted sum of squares of the PLS loadings that considers the amount of explained Y-variation in each dimension. A VIP score >1 indicates that the metabolite significantly contributed to time point differentiation. Most of the VIP >1 metabolites exhibited a bimodal pattern, going from very low in the uninfected insect, to rising in the early bacterial replication phase, to dropping during the middle nematode reproduction phase, and finally rising very high in the late nutrient deplete phase (Fig S2.3 and Data S2.3). Overall, these metabolites were involved in nucleotide and nucleoside biosynthesis, NAD⁺ biosynthesis, and iron acquisition. These included the purine and pyrimidine metabolites 7-methylguanosine, guanine, deoxyinosine, uridine, deoxyuridine, and deoxycytidine. Other top metabolites include kynurenic acid and anthranilate, both of which are connected to tryptophan and NAD⁺ synthesis pathways. Of the top 15 VIPs, ascorbate was the only molecule to exhibit a decreased abundance over time, dropping from very high abundance to very low later in the time course. This vitamin is necessary for neuron development and could be salvaged from the cadaver to build the nematode nervous systems (73).

Kynurenic acid (VIP scores for Components 1-3: 1.984, 1.583, 1.4563, respectively) abundances increase significantly between uninfected and early phases, and again between middle and late phases, and abundances show a slight (though not significant) decline between early and middle phases (Fig 3). Kynurenic acid is a side product of the kynurenine pathway that produces NAD⁺ from tryptophan in organisms that lack the classical quinolinate phosphoribosyltransferase (QPRTase) pathway (74). *S. carpocapsae*, like *C. elegans*, lacks a standard QPRTase but encodes the uridine monophosphate phosphoribosyltransferase (*umps-1*) which synthesizes NAD⁺ from the kynurenine pathway (75). Kynurenic acid has immunomodulatory activity in mammalian systems: it is a reactive oxygen species scavenger,

can attenuate the inflammatory response to microbial antigens such as lipopolysaccharide, and can signal through G-protein coupled receptors and aryl-hydrocarbon receptors (76). Its higher abundance in early phase infection relative to uninfected may reflect its function in suppressing insect immunity or protection for reactive oxygen species. In contrast, the lower abundances in the middle phase, when nematodes are reproducing, relative to the late phase, may reflect its relationship to NAD⁺ biosynthesis; since kynurenic acid is a side product of the tryptophan-derived kynurenine pathway, its abundance is likely to be inversely correlated with other metabolic products of this pathway, including NAD⁺ (76).

Phenylacetic acid (VIP scores for Components 1-3: 2.1253, 1.7413, 1.5997, respectively) had significantly higher abundances in early relative to uninfected, significantly lower levels in middle versus early, and significantly higher in late versus middle (Fig 3). In bacteria, phenylacetic acid can be transformed into acetyl coA or succinyl coA and routed into the TCA cycle. *X. nematophila* has homologs of the enzymes necessary for this pathway. The first step is the conversion of phenylacetate to phenylacetyl coA is catalyzed by phenylacetate-coenzyme A ligase (PaaK), predicted to be encoded by XNC1_4627 (EC 6.2.1.30). *X. nematophila* encodes another putative AMP-dependent ligase (XNC1_4600 EC 6.2.1.30) that is negatively regulated by LrhA (transcript levels are 3X lower in wild type relative to the *lrhA* mutant: Data S2.2). These data indicate that phenylacetic acid may be a branchpoint metabolite that is diverted to different metabolic pathways, or accumulates depending on the dominant metabolism of the ecosystem. In *P. aeruginosa* phenylacetic acid accumulates at high cell density and inhibits the Type III Secretion System (T3SS), which delivers toxic effectors to host cells (77). *X. nematophila* does not have a T3SS but does have the evolutionarily-related flagellar export apparatus. The transcription factor Lrp positively regulates the flagellar regulon, as well as the gene encoding the XlpA lipase, an enzyme associated with the ability of *X. nematophila* to support *S. carpocapsae* reproduction (58, 59). Although not detected by

microarray analysis, quantitative reverse transcriptase analyses indicate that the transcription factor LrhA also positively regulates *xlpA* expression (58). The relatively higher abundance of phenylacetic acid at later stages of insect bioconversion is consistent with reduced routing through the bacterial TCA cycle (see below), and may inhibit expression and secretion of bioconversion enzymes, such as XlpA, by the bacteria.

2,3-dihydroxybenzchoate (VIP scores for Components 1-3: 1.7488, 1.3571, 1.2757, respectively) is the precursor of the iron-binding siderophore enterobactin, synthesized by the enzymes EntA and EntB. Iron appears to be available in insect cadavers based on previous studies in the insect model *M. sexta* (78), but its acquisition and uptake by bacteria and nematodes would be required for its use in essential cellular processes. *X. nematophila* does not encode *entA* or *entB*, and is not predicted to synthesize enterobactin (36), which suggests that other taxa or pathways may be responsible for 2,3-dihydroxybenzoate abundance shifts (79). Enterobactin can have ecological relevance within the insect cadaver; *Photorhabdus*, a close relative of *Xenorhabdus* and part of an EPNB complex with the nematode *Heterorhabditis*, does produce enterobactin that confers a competitive advantage against opportunistic bacteria within the insect cadaver.

To identify patterns and groups of metabolite abundance changes over time, hierarchical clustering was performed to reveal groups of metabolites that exhibit similar abundance changes over the time course of bioconversion. A dendrogram of all 170 identified metabolites was generated using the absolute value of the spearman correlation between molecular abundances, where distance between molecules is defined as $1-|rs|$ with rs as the spearman rank correlation between time course data points of said molecules (Fig 2.5A). Metabolite abundance averages were taken for the four time-phases defined: uninfected, early, middle, and late infection. Metabolite clusters were visualized in a heatmap that displays their pairwise correlation between each molecule combining all the time phases (Fig S2.4). A heatmap that

shows the metabolite clusters, separated by black bars, with the molecule trends in abundance change over time was generated (Fig 2.5B). This was measured as the log(rate of abundance change) for all 170 detected metabolites, comparing their mean abundance to the previous time phase (uninfected to early, early to middle, and middle to late). There were 10 total metabolite clusters identified, each with clear molecular abundance patterns in which the metabolites in that cluster exhibited similar rates of change together over the time phases.

Clusters of metabolites that exhibit similar change trends were examined to gain an understanding of very broad metabolic pathways affected at each time phase (Data S2.3). Clusters within each phase were compared to those in the previous and subsequent time phases. Relative to the previous uninfected phase, in the early infection phase, during which the insect is mounting an immune response and succumbing to infection and death, there are increased abundances of Clusters 4 and 5. These clusters contain metabolites involved in glutathione biosynthesis (glutamate, cysteine, pyroglutamic acid, and NADP⁺). Glutathione intermediates are increasing while glutathione itself (Cluster 3) is decreasing. Decreased abundance of glutathione paired with increased abundance of synthesis intermediates could be reflective of the failure of inhibition of insect self-protection from its own induced immune response, which includes reactive oxygen or nitrogen species molecules through the phenoloxidase melanization cascade. Glutathione is an antioxidant which protects against such reactive molecules, and its decreased abundance relative to its precursors may indicate that the EPNB complex has successfully suppressed the phenoloxidase cascade (80, 81). Also increasing during the early infection stage are several amino acids in Cluster 4, specifically those involved in tryptophan metabolism (tryptophan itself and its precursor indole). Tryptophan is a precursor to several physiologically important metabolites, including kynurenic acid, one of the VIP metabolites identified in this analysis, kynurenine pathway intermediates and products, including NAD⁺, serotonin, and indole compounds (82).

Metabolites that exhibited continuously decreasing abundances in the early phase relative to the uninfected phase were in Clusters 1, 3, 8, 9, and 10. These clusters contain many compounds, and the highest proportions are involved in purine and pyrimidine biosynthesis and ascorbate metabolism (myo-Inositol, UDP-glucose, UDP-glucuronate, and glucarate) (83). The decreased abundance of ascorbate pathway metabolites may be connected to reduced glutathione noted above; in *C. elegans*, ascorbate was synthesized by the nematode and the ascorbate-glutathione pathway could be used to detoxify reactive oxygen species (84). From the microarray data, generally the regulator mutants caused transcripts in ascorbate degradation (*nilR* and *Irp*, both affecting XNC1_2826, XNC1_2827, and XNC1_2828) and tryptophan (*Irha* and *rpoS*) metabolism to decrease. These data from the early phase represent the clusters of metabolites that are at the battle front between the insect immune system and infecting EPNB and support snapshot data that EPNB can suppress insect immunity and prevent production of reactive oxygen species.

The middle infection phase occurs after insect death and likely is dominated by decomposition of the insect by *X. nematophila* bacteria. During the transition between the early phase and the middle infection phase, two clusters (Clusters 8 and 10) exhibit an increase in rates of abundance changes. The metabolites in these clusters include several purine components (deoxyinosine, xanthosine, and inosine) as well as one of the only B vitamins detected in this screen: riboflavin (vitamin B₂). Other detected B vitamins, like biotin (vitamin B₇), pantothenate (vitamin B₅), and 4-pyridoxate (catabolic product of vitamin B₆), decreased in abundance during the middle phase. Other metabolites that decreased in the middle relative to the early phase include pyrimidine intermediates (UMP, CMP, CDP, and UDP) and ascorbate and sugar acid compounds, as well as amino acids (arginine, phenylalanine, tyrosine, tryptophan, cysteine, and methionine). This decrease is expected, as nucleotides and amino acids are being incorporated into DNA, RNA, and protein for bacterial and nematode biomass

accumulation. The identification of specific types of amino acids and pyrimidines may reflect that these compounds are more limited in abundance than others. Microarray analysis indicates widespread differential regulation, particularly in the secondary form, *lrp*, and *lrhA* mutants, in the pathways related to the decreasing abundance amino acids noted above. These data from the middle phase are reflective of the compounds that are being siphoned from the cadaver by the bacteria and incorporated by digestion into the nematode biomass and emphasize the importance of amino acid metabolism regulation in *X. nematophila* bacterial physiology.

As the insect cadavers entered the late infection phase, during which the dominant activity may be exponential expansion of nematode populations, nucleobases (guanine, thymine, and uracil in Cluster 3) and amino acids (aspartate, arginine, cysteine, and methionine in Clusters 1, 2, 4, and 9, respectively) steeply increased. This suggests these accumulating compounds are available for nematode DNA, RNA, and protein incorporation, but other factors such as overcrowding in the cadaver or lack of other necessary resources force the nematode to exit. These amino acids are also commonly found to biomarkers of decomposing animal tissue, with cysteine, leucine, and aspartate having the highest increase rate of abundance shifts in rat cadavers (85). Decreasing rate of change of metabolites relative to the previous middle phase included compounds involved with leucine metabolism and the TCA cycle. These could be more rate-limiting compounds, where their decreasing abundance could signal to the expanding nematode population that it is time to exit.

Tricarboxylic acid (TCA) cycle metabolites abundances change significantly throughout the nematode lifecycle

To identify significant metabolites that are important for infection progression, an ANOVA with post-hoc Tukey's HSD test was performed on metabolite abundances throughout the lifecycle. Two comparisons were examined: metabolite abundances from uninfected insects compared to individual time points, and individual time points compared to the next subsequent

time point. As summarized in Figure 6, TCA cycle components significantly ($p < 0.05$) fluctuate in relation to uninfected insects as well as between time phases, throughout the time course (Supplementary Text 2C). Additionally, significant abundance shifts in amino acid metabolism were identified. These trends, especially pertaining to proline and leucine biosynthesis, reveal the importance of insect bioconversion into building blocks essential for nematode development (Supplementary text 2D, Fig S2.5).

In the early phase of infection, while the insect is still alive and combatting bacteria and nematode invaders using innate immunity, several key TCA cycle intermediates are reduced in abundance relative to an uninfected insect (Fig 2.6). This is shown through significantly lower abundances of citrate in the Hour 12 and Hour 24 living insects compared to the uninfected insect and the Hour 24 dead insects. Although not significant, a similar trend is observed for two other TCA-related metabolites, malate and *sn*-glycerol-3-phosphate, which aids in NAD^+ regeneration through the glycerol phosphate shuttle, as well as NAD^+ , NADH, fumarate. This could mean these metabolites were diverted for the immune response, given the differences between the living and dead insects. As the infection progresses into a middle phase, citrate abundances shift, but generally are decreasing. Into the late phase on Day 10, malate, *sn*-glycerol phosphate, succinate, and citrate all drop, which could suggest carbon is being stored (rather than used) in the IJs before they exit the cadaver.

TCA cycle components were mostly in Cluster 6 (*sn*-glycerol-3-phosphate, NAD^+ , NADH, citrate and isocitrate) and Cluster 2 (fumarate, alpha-Ketoglutarate) (Fig 2.5). Student t-tests were utilized to determine additional significant components by comparing each time phase to the uninfected insects (Data S2.3). Acetyl-phosphate abundance was found to be approaching significantly high ($p < 0.1$) at the early phase and was significantly high ($p < 0.05$) at the middle and late phases. NAD^+ abundance was found to be approaching significantly low ($p < 0.1$) at the early and middle phases and was significantly low ($p < 0.05$) for the late phase.

Xenorhabdus spp. cannot synthesize NAD⁺ and requires nicotinate for growth (36). Generally, there is an overall trend of decreasing abundances for the detected TCA components over the lifecycle as the infection progresses, with the exceptions of acetyl-phosphate, FAD, and succinate.

The aforementioned *X. nematophila* mutants, relative to wild type displayed differences in transcripts involved in either pyruvate metabolism, glyoxylate metabolism, or the TCA cycle (Data S2.2). Most of these genes were differentially abundant in the $\Delta\rho hA$ and $\Delta\rho oS$ mutant backgrounds. *aceA* and *aceB* transcripts were found to be in lower abundance in the mutants relative to wild type, while *aceE* and *aceF* were found to be in higher abundance in the mutants relative to wild type. These genes are involved in the glyoxylate shunt which is a pathway utilized by many bacteria and nematodes in the absence of bountiful sugars to convert 2-carbon compounds into energy resources (86). $\Delta\rho oS$ mutants upregulate several succinate dehydrogenase genes which are necessary for oxidative phosphorylation. In *E. coli* these genes are induced in response to different environmental conditions like iron and heme availability (87).

Discussion

A comprehensive framework to understand how metabolism shifts during infection lifecycles of entomopathogenic nematodes was established. Physiologically, it was determined that *X. nematophila* bacteria consume insect tissues, while *S. carpocapsae* nematodes consume *X. nematophila*. The high TP_{glu-phe} of 4.5 observed in the nematodes emerging from an insect cadaver suggests the IJs potentially cannibalize previous generations of nematodes and/or feed upon bacteria that were, themselves, already feeding on previous generations of nematodes and bacteria, since if the colonizing nematodes were feeding on bacterial and insect biomass only, they would register at around 3.5. One form of cannibalism in which *S. carpocapsae* nematodes engage is endotokia matricida (or bagging) in which nematode eggs

hatch within and consume the mother, likely during the second generation of nematodes when nutrients are becoming depleted. However, we cannot rule out from our data that some other form of cannibalism occurs within the cadaver (88). A $TP_{\text{glu-phe}}$ of 4.6 is similar to many apex predators, such as large marine carnivores or the rare top predators observed in terrestrial ecosystems (54, 89). This underscores the importance of including microbes in studies of organismal trophic identity. In effect, the cadavers used in this study may represent microcosms of the broader communities and ecosystems in which they are embedded. The insect cadavers were, when alive, herbivores. To find multiple levels of carnivory within a single cadaver suggests that a nematode-colonized arthropod mirrors the trophic richness of the broader food-web. The interdigitation of microbial carnivores in a trophic hierarchy—here, nematodes and bacteria—is likely a much more common feature of food-webs than previously thought (67, 90). The microbial trophic identities reported in this study may necessitate a re-calibration of organismal niche concepts, but in so doing, will facilitate the unification of the macro- and microbiome in food web ecology.

It should be emphasized that it has been exceedingly uncommon to find higher-order consumers ($TP > 4.0$) in a community or ecosystem, given that apex predators feed upon other predators that have, themselves, had to find and subdue ‘lower’ carnivores (91). In classical food web ecology, apex predators are generally considered to be large, fierce, and rare vertebrates (92). However, perhaps the assumption that apex predators exist only within the province of large/fierce/rare vertebrates needs to be revisited. The high trophic positions exhibited by the nematodes in this study suggest that such obligate higher-order consumers are more common than previously thought, with multitudes of apex carnivores existing underfoot in many terrestrial ecosystems. Further, the nematodes can be viewed as farming their symbionts: acting as shepherds that bring their bacterial flock to a fresh insect pasture for harvesting of nutrients.

The trophic study described in this manuscript established the foundation to understand metabolic shifts occurring in the cadaver. With the time course metabolomics study we sought to better understand the process of bioconversion in the cadaver; how is the insect biomass being converted to bacterial and nematode biomass? Applying multivariate statistical tests to the infection metabolomics data set revealed distinct time phase clustering. The variance among the time phases seems to increase as infection progresses, as the healthy insects are degraded by the bacteria and turned into bacterial biomass and nematode tissue. We observed a bimodal pattern of abundances of the most significant metabolites from the PLS-DA. These metabolites may be signatures of the timing of bacterial and nematode development. Metabolite peaks occurring during the early phase when bacteria are multiplying may indicate byproducts of bacterial conversion of insect tissues. In turn, the drop in abundance of these metabolites in the second phase, when nematodes are reproducing, may indicate the consumption of these byproducts by the nematode. Finally, the second peak in the late phase of infection occurs when the nematodes are exiting the cadaver, leaving behind residual populations of bacteria that may remain physiologically active (Fig S2.3). We compared the infected insect samples to samples of *S. carpocapsae* nematode IJs that had emerged from an insect cadaver and that had been stored in water for several weeks. These emerged, water-stored, solely IJ samples had metabolic profiles that cluster away from the insect sample time phases (Fig S2.6), and were most similar to the late infected insect time phase. This is not surprising, given that the late insect samples are comprised primarily of hundreds of thousands of IJ nematodes that have developed within the cadaver and are close to emerging (Fig 3). However, it is important to note that the nematode IJ samples tested in this study had been stored in water for weeks, and were not derived from the same insects that had been sampled for the other time points. An important follow up study will be to monitor metabolic shifts occurring immediately preceding, during, and during aging following IJ emergence from insect

cadavers. Such a study would provide more detailed insights into IJ physiology during migration, aging prior to reinfection, environmental stress responses, and hunting for new prey.

Metabolic analysis revealed TCA cycle components were among the most statistically significant results, indicating that their use by the entomopathogenic nematodes is paramount to infection success and subsequent nematode propagation. Citrate metabolism is ubiquitous in many intracellular pathogens and contributes to virulence phenotypes in insect host models (93). Citrate reduces the virulence of the Gram-negative pathogenic bacterium *Pseudomonas aeruginosa*; citrate treatment caused a reduction of bacterial NADH levels are reduced upon treatment with citrate, and host-killing activity is abolished as a result (94). The authors of this study hypothesized that this could be due to decreased flux through the glyoxylate bypass, which has been found to activate the T3SS in this system (95). As mentioned, *X. nematophila* does not encode a T3SS, but does have the evolutionarily related flagellar export apparatus. Several *X. nematophila* glyoxylate bypass genes, were differentially transcribed between avirulent genetic mutants and WT. Glyoxylate was not detected in our screen, and whether flux through this pathway affects virulence should be investigated further. The significant abundance shifts of TCA metabolites during the middle phase may be indicative of the role of the TCA cycle in *S. carpocapsae* development. The TCA cycle plays an essential role in early embryogenesis in *C. elegans* (96, 97), in that repression of citrate synthase (encoded by *cts-1*) and cyclin-dependent kinase 1 (encoded by *cdk-1*), both of which have homologs in *S. carpocapsae*, halt *C. elegans* development. Neutral lipids are formed from *sn*-glycerol-3-phosphate and are the major energy reserve in the closely related *S. feltiae* nematodes (98). Fats are stored as lipid droplets in *C. elegans* dauer larvae intestines and serve as a starvation survival mechanism (99). The previously mentioned glyoxylate bypass forms carbohydrates from fatty acids and has been implicated in extending the lifecycle of *C. elegans* (100), highlighting another role of this TCA vs. glyoxylate switching that could be happening later in the lifecycle. Any indication that cholesterol is being synthesized from these intermediates can be attributed entirely to the

insect's wheat germ diet, since *X. nematophila*, *S. carpocapsae*, and *G. mellonella* cannot synthesize sterols but require them to grow (101).

Metabolites such as 2,3-dihydroxybenzoate that are synthesized by neither *X. nematophila* nor *S. carpocapsae* were found to increase over the infection, past insect death. The *X. nematophila* genome does not contain the genes that convert 2,3-dihydroxybenzoate to the siderophore enterobactin, which bind iron to create the ferric enterobactin (FeEnt) complex (102). However, *X. nematophila* encodes *fepB*, a periplasmic enterobactin binding protein, as well as *fepC*, *fepD*, and *fepG*, which encode proteins to transport the FeEnt into the cell. The increased abundance of 2,3-dihydroxybenzoate over time suggests other *G. mellonella* microbiome members that survive infection, wherein *Xenorhabdus* utilizes these community members to synthesize a compound that is paramount in iron extraction and could contribute to the overall fitness of this symbiosis.

Metabolic analysis also revealed the significance in proline throughout the lifecycle. Insect hemolymph is rich in proline and is used as a main fuel source in some species of flying insects because of its ability to oxidize carbohydrates (103). Proline can be a signal molecule inducing secondary metabolite biosynthesis in *Xenorhabdus* species. Several virulence factors and antibiotics are regulated via exogenously supplied proline to *Xenorhabdus* cultures (104). *Xenorhabdus* species may have evolved to use proline in insect hemolymph as a preferred amino acid source capable of enhancing the bacterium's virulence as well as protecting it from various stressors it encounters within a new insect host. Proline catabolism also has been implicated in promoting reactive oxygen species homeostasis and modulating innate immunity in *C. elegans*, highlighting a possible mechanism by which *S. carpocapsae* survives reactive oxygen species produced by the insect, bacterium, or themselves (105). Enhanced understanding of proline changes over time in the EPNB lifecycle highlights the multiple roles this amino acid is playing for both *Xenorhabdus* virulence and *Steinernema* protection and

reproduction. Additional amino acids exhibited similar abundance shifts over the lifecycle as detected by hierarchical clustering analysis. This machine learning technique can be improved with higher granularity of sample time points, which would strengthen the software developed for this study and allow its application to more complex and dynamic chemical environments.

We have described the food web and metabolic details of prey consumption in a parasitic infection by an apex predator nematode. Our work offers insight into the metabolism of parasitism and highlights the importance of including microbes as components of food chains. Through rigorous metabolic pathway reconstruction and multivariate statistics, these results suggest each phase of prey bioconversion is characterized by specific chemical signatures. Expanding on this initial identification of signature metabolic profiles, future targeted metabolomics experiments on EPNB have the potential to reveal a detailed picture of metabolic routes by which meso-predators such as *X. nematophila* consume their prey, and are themselves consumed by apex predators such as *S. carpocapsae* nematodes. This finding raises new interpretations of the *Steinernema-Xenorhabdus* symbiosis as a predator-prey relationship in which the nematode modulates prey (symbiont) consumption to ensure sufficient reserves are available for repopulating new environment. Our work adds to a growing scientific understanding of how symbioses, both mutualistic and parasitic, shape the chemical environments they inhabit.

Materials and methods

Conventional nematode and aposymbiotic nematode production

S. carpocapsae nematodes (All strain) were propagated through 5th instar larvae of insect *G. mellonella* and conventional IJs were collected in distilled water using a White trap, and stored at room temperature for <1.5 months (106). To generate aposymbiotic IJs, *X. nematophila* Δ SR1 mutant bacteria, which cannot colonize *S. carpocapsae*, were grown in Luria Broth (LB)

media overnight at 30°C on a cell culture wheel and 600µl of overnight bacterial culture were spread onto 10ml lipid agar plates to grow into a confluent lawn at 25°C for 48 hours.

Conventional IJs were surface-sterilized, seeded onto a $\Delta SR1$ mutant lawn on lipid agar plates (5000 IJs per 10ml media), and incubated at 25°C for 7 days in the dark for nematode reproduction. Aposymbiotic IJs were collected by water-trapping using distilled water and stored at room temperature in the dark (106).

In vitro controlled feeding experiment

To collect samples of bacteria feeding on terrestrial C3 plants and yeast-based media (Fig S1B), *X. nematophila* wild-type or $\Delta SR1$ mutant bacteria were grown in the dark in yeast soy broth (0.5% yeast extract, 3% tryptic soy broth, and 0.5% NaCl) modified from the bacterial growth media from (107) at 30°C on a cell culture wheel. Wild-type bacterial overnight cultures (5ml per condition per biological replicate) were collected into microfuge tubes, spun down at top speed (>15,000 RCF), and washed three times using 1x PBS buffer by resuspending and spinning down the bacterial pellets. Exactly 600µL of bacterial sample were spread onto yeast-soy lipid agar plates (0.5% yeast extract, 3% tryptic soy broth, 1.5% agar, 0.2% MgCl₂, 0.7% corn syrup, 0.4% soybean oil, supplemented with 40µg cholesterol per liter of media) and incubated for 48h at 25°C to grow into a confluent lawn. Bacterial lawns were washed off the agar plate using 1x PBS buffer, pelleted and washed as described above. Three individual tubes of bacterial culture, derived from colonies on three separate plates per strain, were used as three independent biological replicates. To grow nematodes using a controlled diet, approximately 5000 conventional IJs were surface-sterilized and seeded onto bacterial lawn grown on yeast-soy lipid agar plate as described above. Three individual yeast-soy lipid agar plates were used as three biological replicates for each bacterial condition. To collect first generation reproductive stage nematodes (which include both adult males and females), at three days post-inoculation with nematodes on bacterial lawns, the plates were flooded with 1x

PBS buffer to resuspend the nematodes. The nematode resuspension was collected in a glass cell culture tube and washed three times by resuspending in 1x PBS buffer. To collect second generation IJ progeny nematodes, water-traps were set up seven days post-seeding IJs on bacterial lawns. IJs that emerged from the plate into the water-traps were collected, allowed to settle by gravity, and washed three times in distilled water by repeated settling and resuspension.

In vivo feeding experiment and sample collection

To prepare insect controls, *G. mellonella* 5th instar larvae were injected with 10 μ L of either 1x PBS buffer, yeast-soy broth media, or nothing. Three insect larvae were prepared per condition as three biological replicates. To collect nematodes directly fed on *Galleria* insect tissues, *S. carpocapsae* axenic eggs were extracted from adult female nematodes grown on yeast-soy lipid agar plates. Approximately 6000 axenic eggs were seeded on each of the *Galleria*-tissue agar plates (20% (w/v) frozen *G. mellonella* insects cleaned, blended and filtered; 0.5% (w/v) NaCl; and 1.5% (w/v) agar, supplemented with 50mg/L Kanamycin). Mixed stages of nematodes were collected by flooding the *Galleria*-tissue agar with 1x PBS buffer to resuspend the nematodes, then washed in 1x PBS buffer for 3 times to separate nematodes from insect tissue debris. To establish controlled feeding experiments *in vivo* for bacteria and nematodes, *X. nematophila* overnight cultures (in yeast-soy broth) were diluted in 1x PBS buffer and approximately 10⁴ bacterial cells were injected with or without aposymbiotic nematodes (100 IJs per insect). Insect cadavers injected with bacteria only were directly lyophilized and used as insect-bacteria complex controls (see methods below). Insects with bacteria and nematodes co-injection mixture were used to collect IJ progenies by water-trapping, washing (three times in distilled water), and pelleting the IJ samples. Three to five insects were used for each experimental condition as biological replicates.

Nematode lyophilization and trophic position analysis

Nematodes from *G. mellonella* were collected by placing infected cadavers in modified White traps in which nematodes migrate into distilled water. Trapped nematodes were transferred to 15 ml Falcon test tubes and allowed to settle into a pellet at the bottom of the tube. Nematodes from plate cultivations were harvested by rinsing with sterile distilled water, transferred to Falcon test tubes, and allowed to settle. Samples of nematodes were stored in water at 10°C within 1-2 days until they were lyophilized. For lyophilization, water was decanted off of the sample until only the undisturbed pellet remained at the bottom of the test tube. The top of the test tube was covered with a Kimwipe held in place with a rubber band before lyophilization for >48 h in a Labconco Freezone lyophilizer. During this time, pressures fell below 0.2 millibar, and temperatures reached -50°C. Once the samples had been thoroughly lyophilized, they were removed from tubes using a laboratory spatula that was sterilized with ethanol and dried with a kimwipe after every use. Each individual sample was relocated into a sterile 1.5 ml microfuge tube and stored at room temperature for 1-3 months until shipment to Hokkaido, Japan for analysis.

Trophic position ($TP_{\text{glu-phe}}$) estimates were generated using the following equation:

$$TP = \frac{\delta^{15}N_{\text{glu}} - \delta^{15}N_{\text{phe}} - \beta}{\Delta_{\text{glu-phe}}} + \lambda$$

where $\delta^{15}N_{\text{glu}}$ represents the nitrogen isotopic ratio of glutamic acid, $\delta^{15}N_{\text{phe}}$ represents the nitrogen isotopic ratio of phenylalanine, β corrects for the difference in ^{15}N values between glutamic acid and phenylalanine within the primary producers of the food web (e.g. $\beta \sim 8.4\text{‰}$ for C3 plants), $\Delta_{\text{glu-phe}}$ represents the net trophic discrimination between glutamic acid and phenylalanine, and λ represents the basal trophic level (=1) of the food web (54). The trophic discrimination factor, $\Delta_{\text{glu-phe}}$ (referred to here as the $\text{TDF}_{\text{glu-phe}}$), represents the net intertrophic ^{15}N -discrimination between glutamic acid and phenylalanine. Significant differences between

known and observed TP values were examined using univariate ANOVA and nonparametric tests (paired Wilcoxon signed rank tests where data were heteroscedastic). Statistical significance among TDF values was accomplished using paired t-tests (108).

Metabolomics sample collection

As per normal infection protocols, 11 *G. mellonella* larvae (Grubco) were placed in the bottom of each of six 6 x 1.5 cm petri plates lined with 2 pieces of #1 filter paper. The filter paper was then inoculated with 1 ml of conventional *S. carpocapsae* IJ stage nematodes (carrying *X. nematophila* bacteria in their intestinal receptacle) to achieve a final average concentration of 10 IJ/ μ l. The nematodes naturally infect the insects by crawling or jumping onto them and burrowing themselves inside. At each specified time point (see below), one *G. mellonella* was taken from each of plates 1-5. All insect samples were flash frozen using a dry ice-ethanol bath, and subsequently stored at -80°C. The uninfected, Hour 1 post-infection and Hour 12 post-infection data points were taken of live *G. mellonella*. Since the *G. mellonella* were starting to succumb to the infection at Hour 24, living and dead insects were taken at this time point. At Day 7 post-infection, a water trap was set up to enable IJ emergence. IJs will slowly begin to exit the cadaver once the water trap has been established and pooling in the water. At Day 12, the last of the *G. mellonella* from plates 1-5 was used, so insects representing the Day 16 time point were taken entirely from plate 6. Input *S. carpocapsae* IJs, which were previously passaged through insects and stored in water in flasks at room temperature, and *X. nematophila* symbionts were also collected from the lab stocks and sent for analysis, approximately 50 μ l of settled IJ per sample, and a total of 4 samples were sent.

Preparation for mass spectrometry

For metabolite extraction, *G. mellonella* insects were equilibrated to -20°C for ~1 h, 300 μ l of extraction solution (40:40:20 acetic acid, methanol, and water) was added, and insects were ground using a pestle that fit snugly into the sample tube. All manipulations were performed in a

cold room and samples were processed in groups of 12. After grinding, to each tube an additional 1000 μ l of extraction solution was added and vortexed for 5-10 sec before being placed at -20°C for 20 min. Tubes were centrifuged at 16,200 x g for 5 min, and the supernatant was decanted to a clean tube. To the original insect sample an additional 200 μ l of extraction solution was added, mixed with a pipette tip, vortexed for 5-10 sec, and incubated at -20°C for 20 min. After pelleting this subsequent extraction, the supernatant was combined with the first supernatant sample. Samples were dried (Savant), resuspended, and randomized samples were analyzed consecutively by mass spectrometry using an established 25-minute method (109).

Metabolomics analysis

An established untargeted metabolomics method utilizing ultra-high performance liquid chromatography coupled to high resolution mass spectrometry (UHPLC-HRMS) (Thermo Scientific, San Jose, CA, USA) was used to analyze water-soluble metabolites (110). A Synergi 2.6 μ m Hydro RP column 100 Å, 100 mm x 2.1 mm (Phenomenex, Torrance, CA) and an UltiMate 3000 pump (Thermo Fisher) were used to carry out the chromatographic separations prior to full scan mass analysis by an Exactive Plus Orbitrap MS (Thermo Fisher). HPLC grade solvents (Fisher Scientific, Hampton, NH, USA) were used. Chromatographic peak areas for each detected metabolite were integrated using an open-source software package, Metabolomic Analysis and Visualization Engine (MAVEN) (110, 111). Area under the curve (AUC) was used for further analyses.

Bacterial strains, plasmids, and culture conditions

Table S1 lists strains used for this study with references where they were originally published. Unless specifically mentioned, *E. coli* was grown in LB broth or on LB plates at 37°C; *X. nematophila* were grown in LB broth or on LB plates supplemented with 0.1% pyruvate at 30°C and kept in the dark. Where appropriate, the following antibiotic concentrations were used:

ampicillin, 150 µg/ml for *E. coli* and 50 µg/ml for *X. nematophila*; chloramphenicol, 30 µg/ml; erythromycin, 200 µg/ml; kanamycin, 50 µg/ml and streptomycin, 25 µg/ml. *E. coli* donor strain S17 (λ pir) or Δ asd strain BW29427 was used to conjugate plasmids into *X. nematophila*.

Microarray experiment and data analysis

Wild-type and mutant (as described in Table S1) *Xenorhabdus* bacterial cultures were grown overnight in 3 ml of LB supplemented with 0.1% pyruvate and appropriate antibiotics in culture tubes at 30°C on a roller, subcultured 1:100 into 30 ml of LB supplemented with 0.1% pyruvate and 50 g/ml ampicillin in 125 ml glass flasks and grown for 12 hours to early stationary phase (OD 2-2.1) at 30°C at 150 rpm on a shaker. 1 ml of each culture was used to extract total RNA using Qiagen RNeasy Mini Kit, and on-column DNA digestion was performed using Qiagen RNase-Free DNase Set according to manufacturer's protocol (Qiagen, Valencia, CA). The RNA purity was tested by measuring 260 nm/280 nm and 260 nm/230 nm ratios in TE buffer and the values should be over 1.8. RNA integrity was verified by running 2 µg of RNA samples on 1% denaturing agarose gel. The samples were then submitted to Roche NimbleGen for processing and microarray analysis. Gene signals for *IrhA*, *lrp*, and secondary form *X. nematophila* were compared to their isogenic parent wild type strain HGB800 using a 2-fold change average signal strength cutoff. The *rpoS* mutant was compared to its isogenic parent wild type strain HGB007 using the same significance cutoff. Genes were annotated via the Magnifying Genomes (MaGe) microbial genome annotation system (112), the STRING database (113), as well as through BlastKOALA (114).

Statistical analysis

PLS-DA plots were generated in MetaboAnalyst 4.0 on August 3rd, 2020. VIP scores were calculated for each component. When multiple components are used to calculate the feature importance, the average of the VIP scores are used. The other importance measure is based on the weighted sum of PLS-regression. The weights are a function of the reduction of the sums of

squares across the number of PLS components (17). Samples were normalized before processing through MetaboAnalyst based on insect weight. Data was log transformed and pareto scaling was applied. Two-way ANOVA with multiple comparisons and Tukey post-hoc tests were completed by taking individual time point metabolite abundances and comparing their means to the uninfected insect model and each other. Student t-tests were performed on individual metabolites by comparing uninfected sample metabolite abundance to each time phase metabolite abundance (early, middle, and late infection). Relevant metabolic pathways were identified in MetaboAnalyst's "Pathway Analysis" module using *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Escherichia coli* as KEGG pathway libraries (18).

Acknowledgments

We thank Terra Mauer for her help in insect rearing. We thank Xiaojun Lu for sample preparation and initial processing of the microarray data. We thank Jordan Rogerson for processing the metabolomics samples prior analysis, and Hector Castro for use of the mass spectrometry facilities. We thank Jennifer Heppert and Nadia St. Thomas for the endotokia matricida image of *Steinernema carpocapsae* used in figure 3.

Data and materials availability

All data are available in the main text or the supplementary materials. Additional detail on the hierarchical clustering analysis can be found at: <http://doi.org/10.5281/zenodo.3962081>

Appendix

Chapter 2 figures and tables

Table 2.1. Summary of *in vitro* and *in vivo* trophic measurements

Consumer ^a	Diet type ^a	TDF _{glu-phe} ^b	TP _{expected} ^b	TP _{glu-phe} ^b
<i>in vitro</i> growth conditions				
None	Yeast-soy lipid agar (YE-YS)	NA	1.0	1.0
<i>Xenorhabdus</i>	Yeast-soy broth (YE-YS)	6.53	2.0	1.9
<i>Steinernema</i> (adults)	<i>Xenorhabdus</i> on yeast-soy lipid agar (YE-YS)	6.96	3.0	2.9
<i>Steinernema</i> (IJJs)	<i>Xenorhabdus</i> on yeast-soy lipid agar (YE-YS)	8.02	3.0	3.0
<i>in vivo</i> growth conditions				
None	<i>Galleria</i> (base of food-chain)	NA	2.0	2.2
None	<i>Galleria</i> + PBS buffer (positive control)	NA	2.0	2.2
None	Yeast-soy broth (YE-YS)	NA	2.0	2.2
<i>Xenorhabdus</i> (measured as <i>Galleria</i> - <i>Xenorhabdus</i> complex)	<i>Galleria</i>	NA	2.5	2.5
<i>Steinernema</i> IJJs	<i>Galleria</i> - <i>Xenorhabdus</i> complex	NA	3.5	4.6

^aEach organism in the ecosystem was assessed for its trophic position as a consumer under controlled *in vitro* conditions or *in vivo* within an insect cadaver. None indicates that the condition tested was diet only, no consumer.

^bIndividual trophic discrimination factors (TDF), expected trophic positions (TP_{expected}), and measured TP (TP_{glu-phe}) are given for each consumer diet.

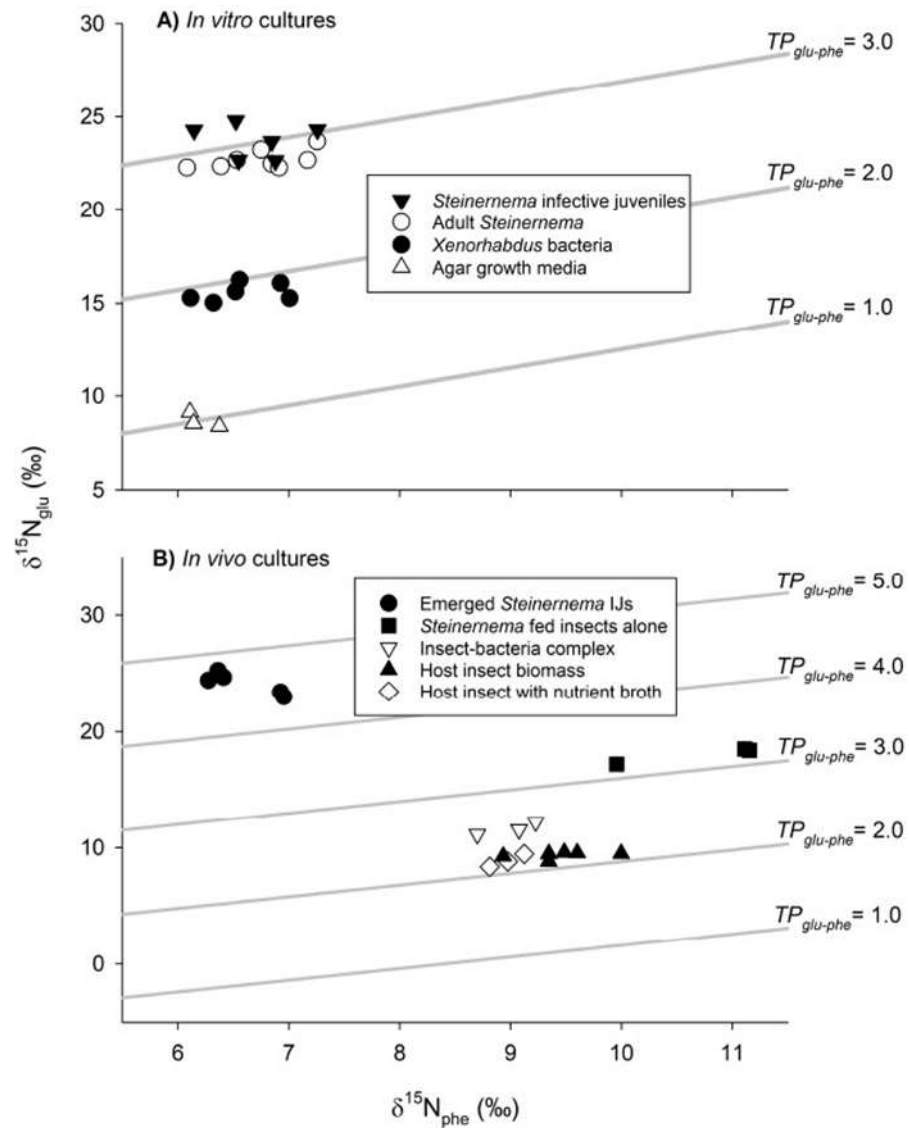


Figure 2.1: Trophic analyses reveal *Steinernema* nematodes feed on *Xenorhabdus*

bacteria (A) In vitro and (B) In vivo. Trophic isoclines are represented via numeric $TP_{\text{glu-phe}}$ ‰ ratios. Specific bacterial cultures or animals are displayed as the different shapes shown in the figure legends.

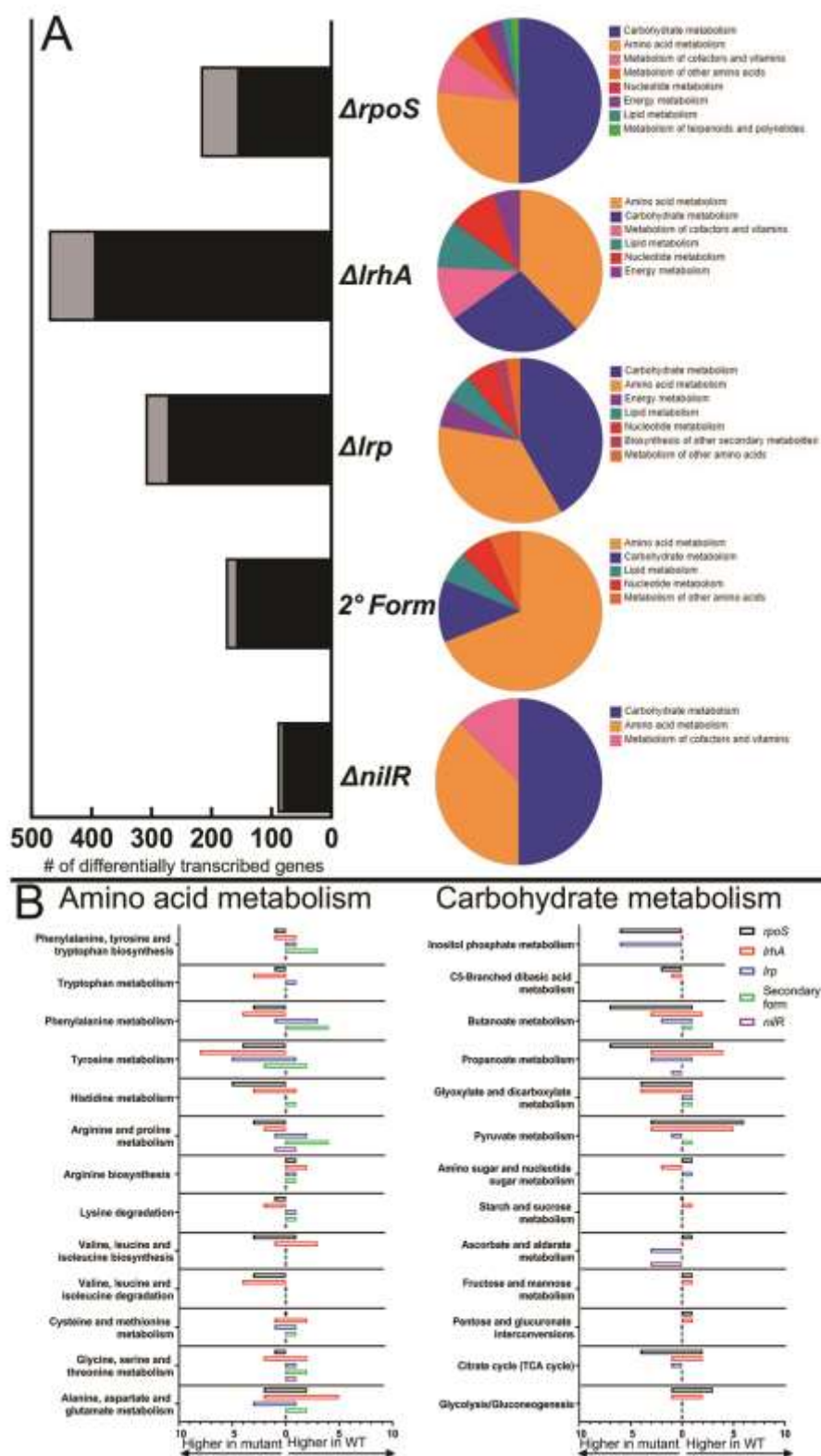


Figure 2.2: Differentially expressed transcripts between *X. nematophila* mutants relative to wild type and their broad functional categorization.

A) Quantification of the number of differentially expressed transcripts and how many are considered metabolic (light grey), as determined by KEGG annotation. |Signal fold change| > 2 was used as a cutoff for significance. BlastKOALA functional categorization of the differential metabolic transcripts are adjacent. The color legend is organized by having the most common category listed first. B) Breakdown of the specific amino acid and carbohydrate metabolism pathways that were affected by the mutations, compared among each strain, with # of genes listed. Positive genes represent transcripts higher in the mutant relative to WT, negative genes represent transcripts lower in the mutant relative to WT.

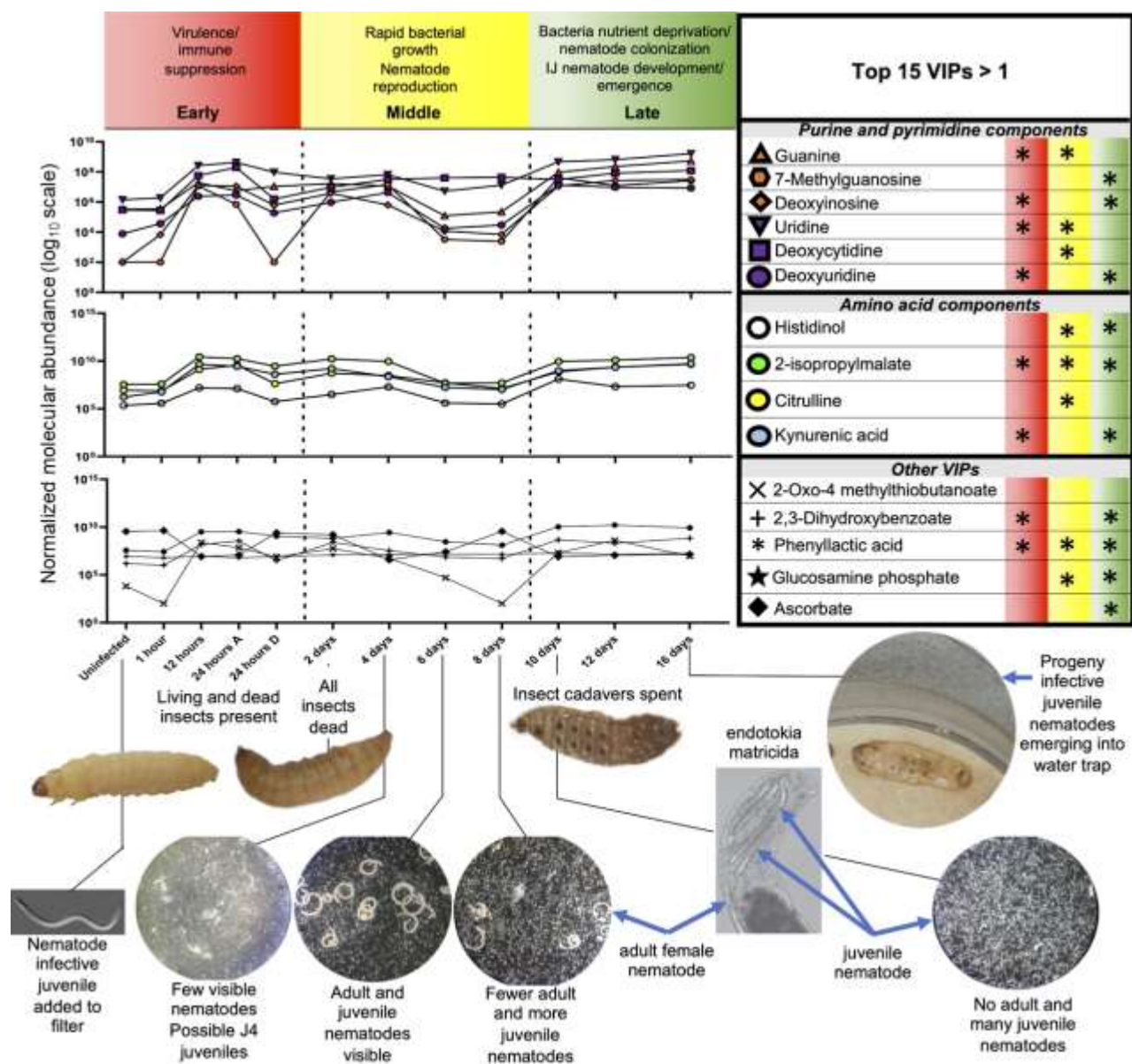


Figure 2.3: Key moments in the EPNB lifecycle mapped onto important molecules are indicative of the bioconversion of the insect cadaver.

The top 15 VIPs>1 metabolite averaged relative abundances were grouped together into 3 categories: purine and pyrimidine components, amino acid components, and other important molecules. Asterisks next to the metabolites represent significant difference in metabolite abundance ($P<0.05$) from t-tests comparing the uninfected stage to the time stages (with red representing early phase, yellow representing middle phase, and green representing late phase). Relative metabolite abundance in log scale is displayed on the y-axis of the line graphs, and dashed lines are used to define the early, middle, and late stages of infections as defined in this study, the general characteristics of which are described. Images of representative living, dead, spent, and water-trapped insects are shown, with lines connecting each to the approximate time frame of bioconversion they represent. In addition, representative images of nematodes are shown (not to scale). Images shown for days 4, 6, 8, and 10 were taken of mixed nematode populations from cadavers dissected at those time points of the metabolomics experiment. An example of an adult female nematode undergoing endotokia matricida, in which juvenile nematodes hatch within and consume the mother, is shown. Blue arrows are used to indicate adult female nematodes and juvenile nematodes in select images.

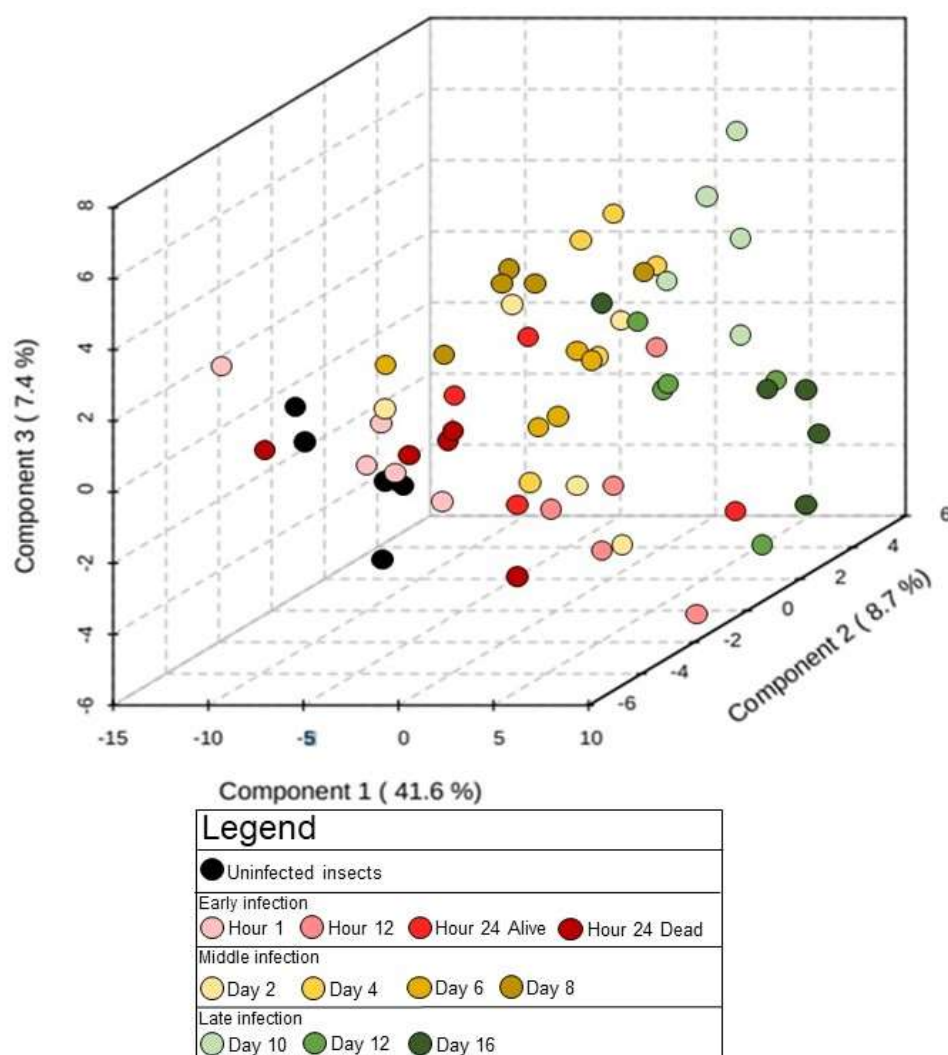


Figure 2.4: Distinct chemical environments occur during bioconversion of an insect cadaver by *S. carpocapsae* and *X. nematophila*. A) Three-dimensional partial least squares-discriminant analysis (PLS-DA) of time course infection metabolic profiles grouped according to stage of infection: uninfected (black) and early (red gradient), middle (yellow gradient), and late infected insects (green gradient). Components contributing to the separation of the profiles are listed (in %) on the axes. B) The top 15 VIPs contributing to component 1 are listed, where the relative abundance shifts over the time course shown in a heatmap on the right. The numbers on top of the heatmap show the time phases: 0 (uninfected), 1 (early), 2 (middle), and 3 (late).

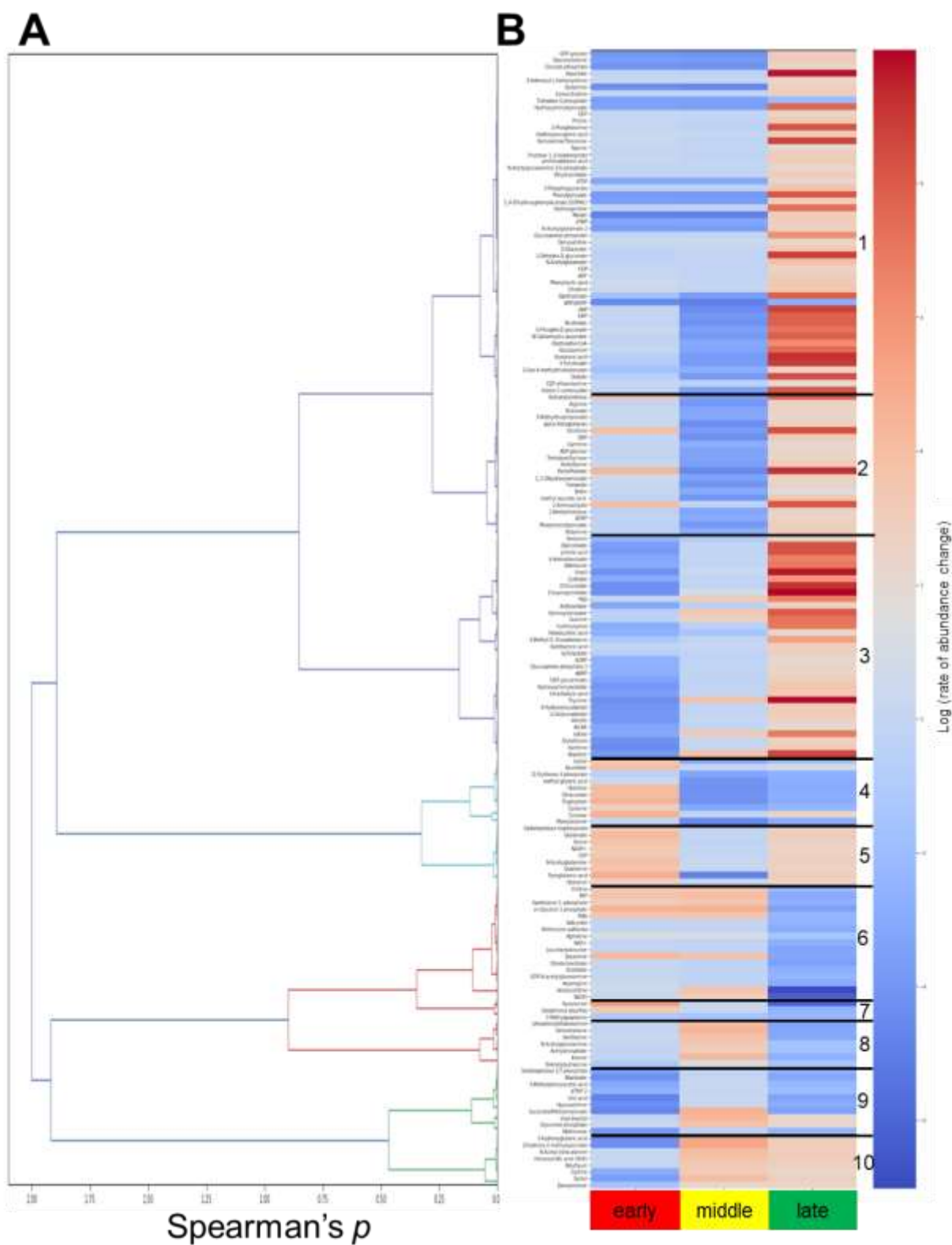


Figure 2.5: Hierarchical clustering analysis of detected metabolites revealed ten clusters, within each of which the metabolites displayed similar rates of change over the infection.

A) Dendrogram corresponding to spearman correlation values for each metabolite. B) Identified metabolite clusters as depicted by numbers 1-10 with similar log(rate of change) over the lifecycle for the time phase, compared to the previous time phase: early compared to uninfected (early), middle compared to early (middle) and late compared to middle (late). Metabolites with a red gradient exhibit an increased molecular abundance shift and metabolites with a blue gradient exhibit a decreased molecular abundance shift.

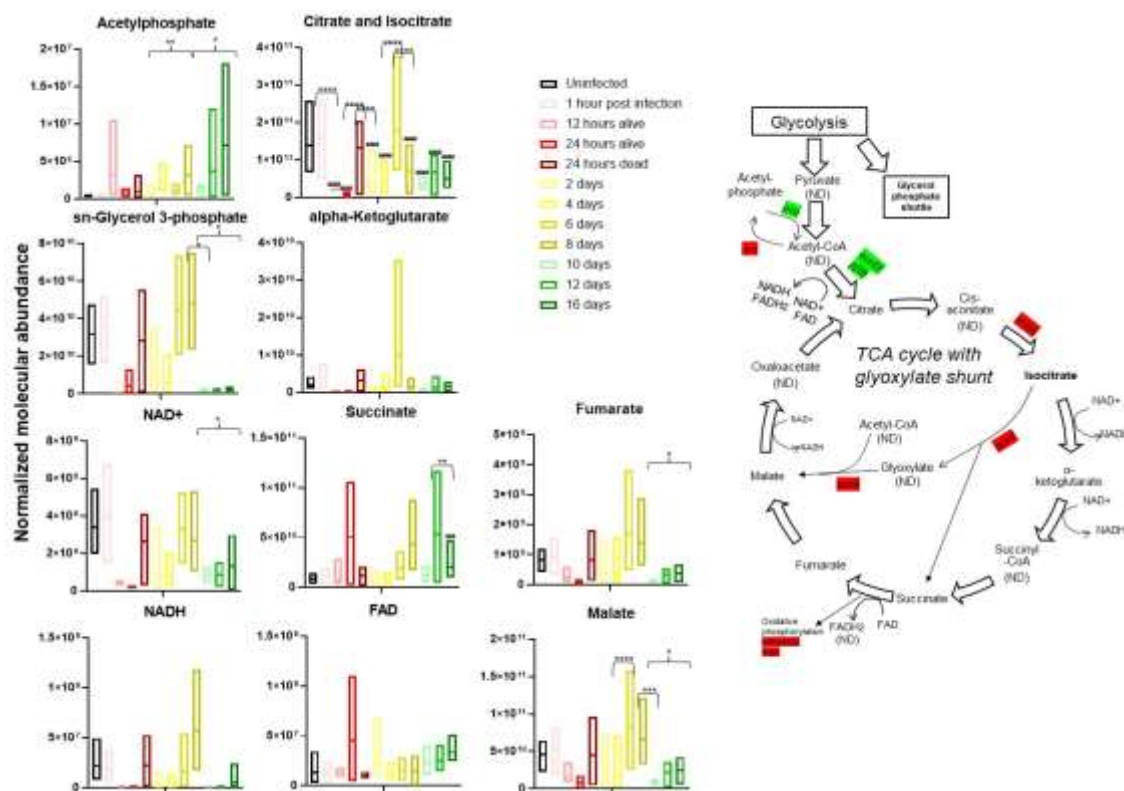


Figure 2.6: Infection with *S. carpocapsae* IJs affects insect TCA cycle. Normalized molecular abundance box plots throughout the lifecycle are shown for all detected metabolites involved in the TCA cycle. Box plot colors represent which time phase the individual plots belong to for: uninfected (black), early infection (red gradient, going from earliest, 1hr, to latest, 24 hours dead, time points), middle infection (yellow gradient, days 2-8), and late infection (green gradient, days 10-16). Lines in the middle of the boxes indicate the mean molecular abundance. Brackets indicate t-test significant abundance shifts between time phases and the uninfected insects, where $*=p<0.05$, $**=p<0.01$, $***=p<0.001$, and $****=p<0.0001$. Parentheses with *s indicate two-way ANOVA with post-hoc Tukey test significance between one time point and the next subsequent time point. #s above time points indicates two-way ANOVA with post-hoc Tukey significance between a time point and the uninfected insect. ND indicates not detectable. Highlighted genes were detected as significant for the microarray in the $\Delta IrhA$ and $\Delta rpoS$ strains. Green indicates transcript abundance is higher in the WT background, red indicates transcript abundance is higher in the mutant background.

Supplementary text (A-D)

A) *In vitro* trophic analysis trials

This past work relied upon compound-specific isotopic analyses of select amino acid pools, particularly the degree of ^{15}N -enrichment between two amino acids—glutamic acid (glu) and phenylalanine (phe). The differential enrichment between these two amino acids provides a measure of inter-trophic enrichment, which is largely attributable to an organism's assimilation of dietary amino acids (115, 116). Such inter-trophic enrichment has been referred to as the trophic discrimination factor ($\text{TDF}_{\text{glu-phe}}$), and in carefully controlled feeding studies among diverse consumer groups in the Animalia, Fungi, and Bacteria, the $\text{TDF}_{\text{glu-phe}}$ has averaged approximately 7.2‰ (67, 106). Here, following controlled-feeding *in vitro* trials, the nematodes and bacteria were shown to have both registered $\text{TDF}_{\text{glu-phe}}$ values in line with past findings (Fig 2.1A). Specifically, the mean ($\pm$ SE) $\text{TDF}_{\text{glu-phe}}$ value exhibited by nematodes cultured on bacterial lawns was $7.41 \pm 0.22\text{‰}$ ($N = 14$). When parsed by nematode stage, the $\text{TDF}_{\text{glu-phe}}$ values of adult and infective juvenile nematodes were, respectively, $6.96 \pm 0.16\text{‰}$ ($N = 8$) and $8.02 \pm 0.36\text{‰}$ ($N = 6$). Nematodes fed exclusively a diet of homogenized insect biomass produced a TDF of $7.38 \pm 0.05\text{‰}$ ($N = 3$), $\text{TP}=3.2$, exactly one trophic level about the insect homogenate which was determined to be $\text{TP}=2.2$ (Data S2.1). Collectively, the nematode TDF was $7.40 \pm 0.18\text{‰}$ ($N = 17$). The bacterial symbiont, *Xenorhabdus*, which had been cultured on agar growth media, registered a $\text{TDF}_{\text{glu-phe}}$ of $6.53 \pm 0.20\text{‰}$ ($N = 6$). The mean $\text{TDF}_{\text{glu-phe}}$ across both the nematodes and bacteria in this food-chain was $7.18 \pm 0.16\text{‰}$, which did not represent a significant departure from the generalized 7.2‰ $\text{TDF}_{\text{glu-phe}}$ benchmark ($t_{22} = -0.14$, $P = 0.893$). Given the degree of inter-trophic enrichment exhibited in the nematodes and bacteria, these consumer groups were consistent with the enrichment patterns of heterotrophs across terrestrial, marine, and freshwater systems, allowing for trophic position estimation using established isotopic protocols (67, 106, 115-117).

Using compound-specific isotopic analysis of amino acids, the trophic identities of consumers and their respective diets within the *in vitro* food-chain were measured. At the base of the food-chain, the agar growth media registered a trophic position ($TP_{\text{glu-phe}}$) of $1.0 \pm 0.04\text{‰}$ ($N = 3$), and the bacteria feeding upon the agar registered at $1.9 \pm 0.03\text{‰}$ ($N = 6$), which represented approximately one trophic level higher than their diet. Correspondingly, the adult and infective juvenile nematodes that had fed upon the bacteria registered, respectively, at $2.90 \pm 0.02\text{‰}$ ($N = 8$) and $3.0 \pm 0.06\text{‰}$ ($N = 6$), which, as predicted, was exactly one trophic level higher than their diet. The homogenate of insect biomass was measured at $2.2 \pm 0.02\text{‰}$ ($N = 6$), and the nematodes feeding exclusively on this homogenate registered at $3.2 \pm 0.01\text{‰}$ ($N = 3$), which again demonstrated that when the nematodes consumed a given diet, they registered one trophic level higher. The *in vitro* food-chain effectively compartmentalized each consumer group and thereby provided a means to confirm that when the nematodes or bacteria consumed a given diet, their isotopic compositions enriched consistently and produced predictable trophic position estimates.

B) Interpreting and analyzing unidentified detected metabolites

Rational

We were able to accurately match 170 metabolites to their database of known retention times (rt) and mass-to-charge ratios (m/z) from our time course metabolomics experiment, where we took nematode infected insects and analyzed their metabolomes over 16 days. There remained 3,138 unidentified spectral features that they could not match in their database, due to the limitations of the data such as the untargeted UHPLC-HRMS method employed and that signal strength for some of the features are really weak and hard to track. Of the 3,138 features, we took 649 features and analyzed them using the Metabolite Annotation and Gene Integration (MAGI) tool (Data S2.4). The reason why so many features were cut out is that is that these 649

had signals consistently throughout the 16 days, while those trimmed would be missing values at certain points throughout the time course.

MAGI is a tool for metabolite identification using the data we acquired, which is m/z and rt for the unidentified spectral features. They have an additional module that was not employed because there was difficulty using the whole *X. nematophila* genome, where it can match specific genes to these metabolites to point to the culprits that are synthesizing these products. I was only able to run certain sections of the genome through for some reason that I couldn't identify, so I opted to just do the metabolite identification analysis.

Since we performed MS1 level analysis on our samples, molecular networking using the Global Natural Products Social Molecular Networking (GNPS) server (in my opinion, the gold standard for this analysis) was not available to use, as they require MS2 level data. Thus, in an ideal world, we would repeat this experiment at higher granularity using tandem mass spectrometry techniques to get more accurate results on these secondary metabolites. m/z varies extremely due to ionization of the metabolites, and rt is completely dependent on the chromatographer machinery used to separate out the metabolites. So, take everything I outline below with a grain of salt, and use this resource as more of a hypothesis generator than firm data to go by. Take what you find and subject these metabolites to a more targeted MS approach to find the answers you seek. Or experimentally test some of the hypotheses you derive here.

Last point: MAGI also performs an analysis of closely related compounds that are found in the same reaction network as the compounds. This data was removed from the final list, as this analysis is hand shaky enough. We can be as confident as possible with the data in hand with this analysis and I wanted to make this as simple as possible

How to read the table

This is a little more involved on the user than I anticipated. I was able to batch search international chemical IDs (InChIKeys) using this the PubChem Identifier Exchange Service (<https://pubchem.ncbi.nlm.nih.gov/identexchange/identexchange.cgi>). But there are multiple entries for the same key (I think that defeats the purpose of having a unique key, but what do I know). These are chemical synonyms, so they are most likely the same chemical with just different names. The issue with batch translating them is that this causes all of the results to not be perfectly aligned to their m/z and rt values, defeating the purpose of translating the IDs in the first place.

Here is what each column from the MAGI results tab means:

A (label): The program combined the m/z and rt values listed in columns B and C. These two measures combined were used to analyze identify the metabolites.

B (original_mz): mass to charge ratios for a given unknown metabolite. This is the reading of the mass of the ions and the charge of the ions that are read by the mass spectrometer, presented as a ratio.

C (rt_peak): retention times (in minutes) for a given unknown metabolite. This is a measure of how long it takes for the metabolite to pass through the liquid chromatography column before it is read by the mass spectrometer.

D (compound_score): The program scores each match for a particular m/z and rt using metric called compound_score. The scoring algorithm is part of a program called Pactolus which scores the mass fragmentation patterns against a database of ion fragmentation trees.

Basically, the higher the score, the more confident they are in a match. I'd only refer to this if there are multiple matches for the same m/z and rt , trust the highest compound score. I sorted the spreadsheet by highest compound score, but don't discard the lower scores.

E (original_compound): These are the matches we care about. For a given m/z and rt , these are the metabolite identifiers. PubChem is the best source for looking these up, either by Googling the IDs directly or putting them into their translation service. As I explain below, it will be easiest to search for your compound of interest's international chemical identifier key (InChI) and then search this list to see if it appears.

How to identify a metabolite of interest with an example analysis

1. You are interested in how dextrin shifts over the time course infection.
2. Google "Dextrin InChIKey" and find the PubChem page on it, the key is:
FYGDTMLNYKFZSV-MRCIVHHJSA-N
3. Search the Magi Results tab (using control+F) for the InChIKey you just found, if there is a hit, great! If not, it was not detected in the analysis. As I mentioned earlier, try researching other InChIKeys for similar metabolites, as I've found multiple keys pop up when searching for what we're looking for. As for dextrin, it's here with an m/z of 504.172 and an rt of 106.333. Several other InChIKeys are here with the same m/z and rt values. These compounds are similar to dextrin, but again these repeated values have to do with the inaccuracy of the method.
4. Take the m/z and rt values, and search for them in the stats tab. You could also search the unidentified_features_list tab if you want to know the individual replicate values. However, I usually just the averages of the replicates tab for initial inquiries. Control+F 504.172 and we see that in column B, and the associated 106.333 is in column C.
5. Once the row has been identified, graph out the values from uninfected (time 0) to Day 16, and see what the trends are:
6. From here, you can do more advanced stats if you wish using the individual replicate values found in the unidentified_features_list tab, like t-tests comparing the individual times (or time phases of early/middle/late) to the uninfected insects. Or comparing the

time to the next subsequent time (like is the drop from uninfected to 1 hour significant?).

Some values may exist for the nematode samples, so those can be analyzed as well.

C) Implication for the bacterial acetylome

Acetyl-coA is an important node in metabolism, and metabolites and transcripts that could affect its abundance were detected in our screens. Acetyl-coA connects glycolysis, the TCA cycle, fatty acid, amino acid, and secondary metabolite pathways, and acetate dissimilation (118). RpoS and LrhA increased expression of *aceF*, *pflB*, *pta*, and *ackA* transcripts. Pta-AckA comprise the acetate dissimilation (excretion) pathway (118). Coordinated elevation of these enzymes is predicted to result in lower levels of acetyl-coA and higher levels of acetyl-phosphate and acetate, which is excreted by the bacteria and potentially available for use by the nematodes. Acetyl-phosphate is a phosphoryl donor for some response regulators and can be a donor for protein acetylation. Protein acetylation, a ubiquitous post-translational modification in prokaryotes and eukaryotes, is involved in regulation of many different bacterium-host interactions like chemotaxis, replication, and acid resistance, as well as regulating bacterial DNA-binding and protein stability (119). Acetyl-phosphate was detected in the metabolome and generally increased over the infection, as well as being a VIP>1 metabolite for components 1 and 2. Acetate freely diffuses across membranes and can be incorporated into biomass of both bacteria and nematodes via the glyoxylate shunt (120). *pflB* is predicted to encode the pyruvate-formate lyase (PFL) enzyme involved in conversion of pyruvate and CoA into formate and acetyl-coA and is greatly (>7 |fold change|) lower in abundance in the $\Delta rpoS$ and $\Delta lrhA$ mutants relative to wild type. PflB converts glucose to formate, and up to one-third of the carbon procured from glucose is converted through this enzyme in *E. coli* (121). PFL condenses acetyl-CoA and formate, allowing for the microbes to use acetate and formate (fermentation products) as the sole carbon sources (122). The wild type *Xenorhabdus* may be able to use these carbon sources during the late phase, while these $\Delta rpoS$ and $\Delta lrhA$ mutants may not. RpoS and LrhA

negatively regulate Acs, which is the acetate assimilation pathway (118). In *E. coli*, Acs activity is inhibited by acetylation of a conserved lysine by acetyl-phosphate and its abundance is negatively regulated by the small RNA SdhX (123). RpoS and LrhA also both negatively influence the abundance of the TCA cycle transcripts *acnA* and *aceAB*, a lack of which is predicted to result in accumulation of citrate. Citrate and isocitrate progressively decrease in abundance over the infection cycle and these combined data might indicate that citrate produced and accumulated by *X. nematophila* bacteria may be a provision for nematodes, consumed during reproduction.

D) Analysis of amino acid abundances found proline fluctuates throughout the infection

The amino acid pyroglutamic acid, which is a precursor to glutamate, significantly rises (as determined through the two-way ANOVA with post-hoc Tukey tests) in abundance during the late phase, at Day 16 compared to the uninfected insect and between Day 12 and Day 16 (Fig S2.5). Glutamate is a precursor for proline metabolism (124). Proline abundances fluctuated significantly throughout the time course. Compared to uninfected insects, proline levels were significantly lower at Day 10 and higher at Day 16, but otherwise were not significantly different. When comparing each time point to the previous, there was a significantly higher level of proline in dead insects at Hour 24 relative to living insects at the same time point. However, in dead insects between Hour 24 and Day 2 the levels of proline dropped again. Thereafter there was a cycle of increase and decrease in proline abundance (Day 4<Day 6, Day 8>Day 10, Day 12<Day 16), with an overall rise throughout the middle and late phases. Additional metabolites in the arginine/proline/polyamine metabolism pathways that were significantly different among samples based on student t-tests (Data S2.3), included ornithine, which was significantly high at the early and middle infection stages compared to uninfected

insects, and hydroxyproline, which was significantly high at early, middle, and late infection stages compared to uninfected insects. Hierarchical clustering analysis reveals Cluster 1 (proline and hydroxyproline) and Cluster 5 (glutamate and pyroglutamic acid) contain these compounds. These cluster trends differ in the early phase, where Cluster 1 decreases while Cluster 5 increases, possibly indicating that these compounds are getting converted into each other.

Microarray analysis of Δlrp , $\Delta nilR$, $\Delta rpoS$, and secondary form *X. nematophila* mutants reveal similarities to each other in disrupted proline gene regulation compared to their wild type/primary form counterparts. *putA*, predicted to encode 1-pyrroline-5-carboxylate dehydrogenase, an enzyme involved in the conversion of proline to glutamate, had a lower transcript abundance ($2 < |\text{fold change}|$) in the Δlrp and $\Delta rpoS$ strains, and had a higher transcript abundance in the secondary form strain. PutA is necessary for virulence in *Pseudomonas aeruginosa* and protects the bacterium against oxidative stress (125). XNC1_2468 has a lower transcript abundance in the Δlrp , $\Delta rpoS$, and secondary form mutant backgrounds. This gene encodes a spermidine N1-acetyltransferase as part of polyamine biosynthesis from ornithine and putrescine. XNC1_2274 and XNC1_3619 are FAD-dependent oxidoreductases involved downstream of proline metabolism in putrescine utilization and have higher transcript abundance in the Δlrp , $\Delta nilR$, and $\Delta rpoS$ mutant backgrounds. These genes are mostly involved in downstream proline metabolism, specifically polyamine biosynthesis. Pathogenic gram-negative bacteria exploit polyamine-related processes of their host for growth and proliferation, using these host molecules for toxin activity, biofilm production, and limiting host immune responses (126, 127). XNC1_2154 is predicted to encode an enzyme with L-aspartate-2-oxoglutarate aminotransferase activity and has lower transcript abundance in the secondary form and Δlrp mutant backgrounds.

Amino acids leucine, isoleucine, and phenylalanine amino acids were significantly higher in abundance at the late stage of infection. These are essential amino acids which *C. elegans*, and presumably *S. carpocapsae*, must acquire from its diet (i.e. bacteria) (128). Several *X. nematophila* genes, including *ilvC* and *ilvI*, involved in leucine/isoleucine biosynthesis were differentially transcribed in the $\Delta lrhA$ (lower transcript abundance) and $\Delta rpoS$ (lower transcript abundance *ilvC* and higher transcript abundance *ilvI*) mutant backgrounds, relative to wild type. $\Delta rpoS$ mutants had higher levels of *leuA* transcript, a gene predicted to encode a 2-isopropylmalate synthase, a leucine precursor which is increases throughout the lifecycle. $\Delta lrhA$ increased transcript abundances of *fadA*, *fadB*, *fadI*, and *fadJ* which participate in the conversion of leucine and isoleucine into fatty acids and acetyl-CoA. In the Δlrp and $\Delta rpoS$ mutant backgrounds, *mmsA* had a higher transcript abundance, another gene that encodes a methylmalonate-semialdehyde dehydrogenase that participates in the breakdown of leucine and isoleucine.

Chapter 2 supplementary figures and tables

	A	B	C	D	E	F	G
Trophic position (TP)	$1.0 \pm 0.04\text{‰}$ (N = 3)	$1.9 \pm 0.03\text{‰}$ (N = 6)	$2.90 \pm 0.02\text{‰}$ (N = 8) adult $3.0 \pm 0.06\text{‰}$ (N = 7) IJ	$2.2 \pm 0.02\text{‰}$ (N = 9)	$2.5 \pm 0.03\text{‰}$ (N = 3)	$3.2 \pm 0.01\text{‰}$ (N = 3)	$4.6 \pm 0.08\text{‰}$ (N = 5)
Apex-predator							
Meso-predator							
Primary consumer							
Diet							
	<i>in vitro</i>			<i>in vivo</i>			

Figure S2.1: Summary of the trophic study results and protocol. Red dashed boxes

indicate the sample for which trophic position ($TP_{\text{glu-phe}}$) was measured. A-C represent the *in vitro* controls using yeast-based media, to establish that the trophic discriminator factors are consistent with previous findings (see Supplementary Text A for more information) A) $TP_{\text{glu-phe}}$ of the yeast media alone, used as a control for bacterial diet. B) $TP_{\text{glu-phe}}$ of the bacteria feeding on the yeast media. C) $TP_{\text{glu-phe}}$ of the adult or infective juvenile (IJ) nematodes feeding on bacteria grown on yeast media using colonizing (WT) and non-colonizing ($\Delta SR1$) *Xenorhabdus nematophila* bacteria. Similar values were obtained, regardless of bacterial strain, indicating that colonization proficiency does not affect $TP_{\text{glu-phe}}$. D-F represent controls measuring the $TP_{\text{glu-phe}}$ of uninfected insects (D), insects infected with bacteria (E), or infective juvenile nematodes emerging from an insect homogenate agar plate. G) The *in vivo* experimental testing using the entomopathogenic-bacterium pair and displays the $TP_{\text{glu-phe}}$ of the infective juvenile nematodes emerging from an insect cadaver in which they had developed in the presence of their bacterial symbiont.

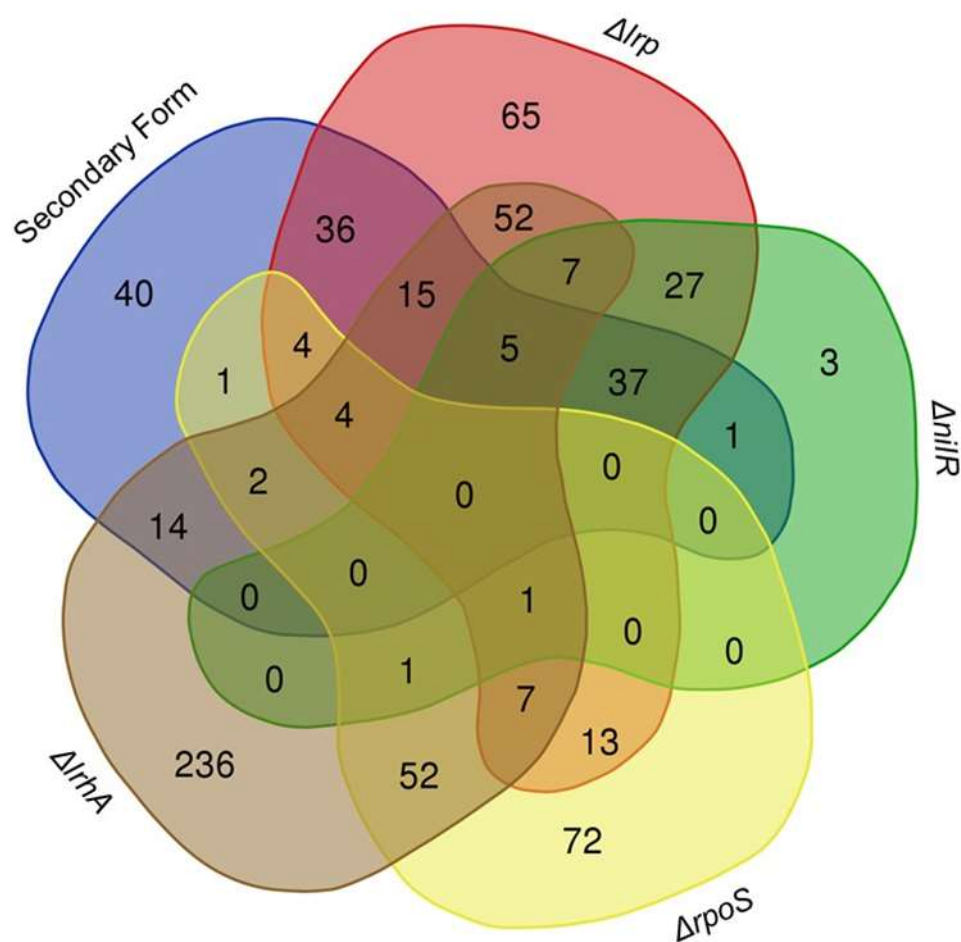


Figure S2.2: Venn diagram comparing the differentially expressed transcripts between the 5 microarrays analyzed. Visualization provided from:

<http://bioinformatics.psb.ugent.be/webtools/Venn/>

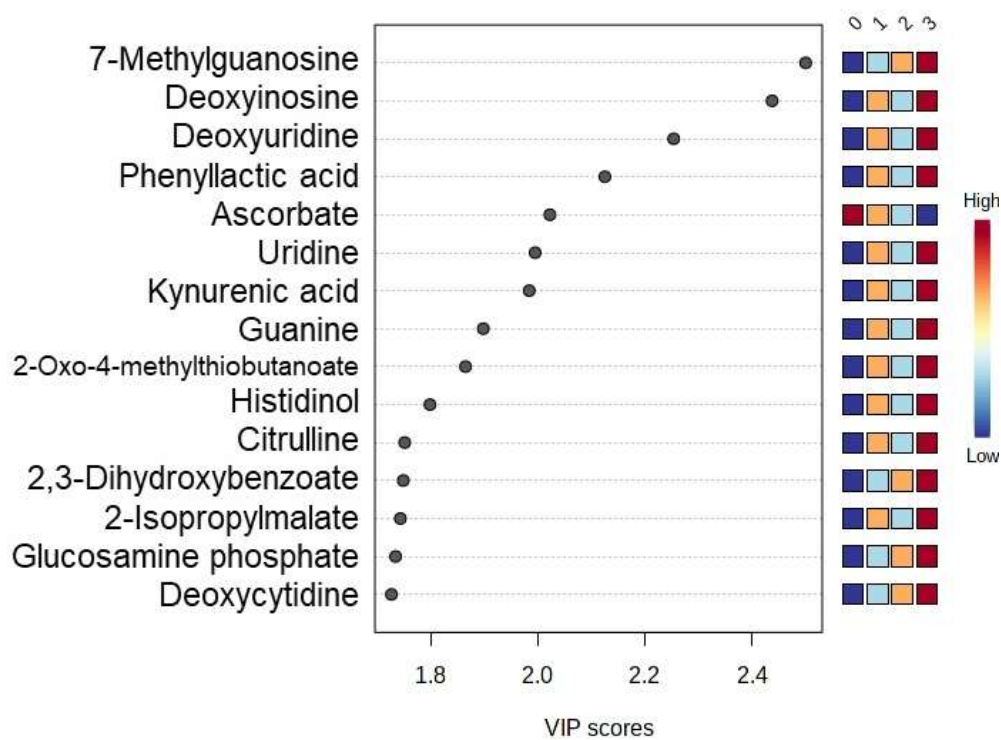


Figure S2.3: The top 15 VIPs contributing to component 1. The relative abundance shifts over the time course shown in a heatmap on the right. The numbers on top of the heatmap show the time phases: 0 (uninfected), 1 (early), 2 (middle), and 3 (late).

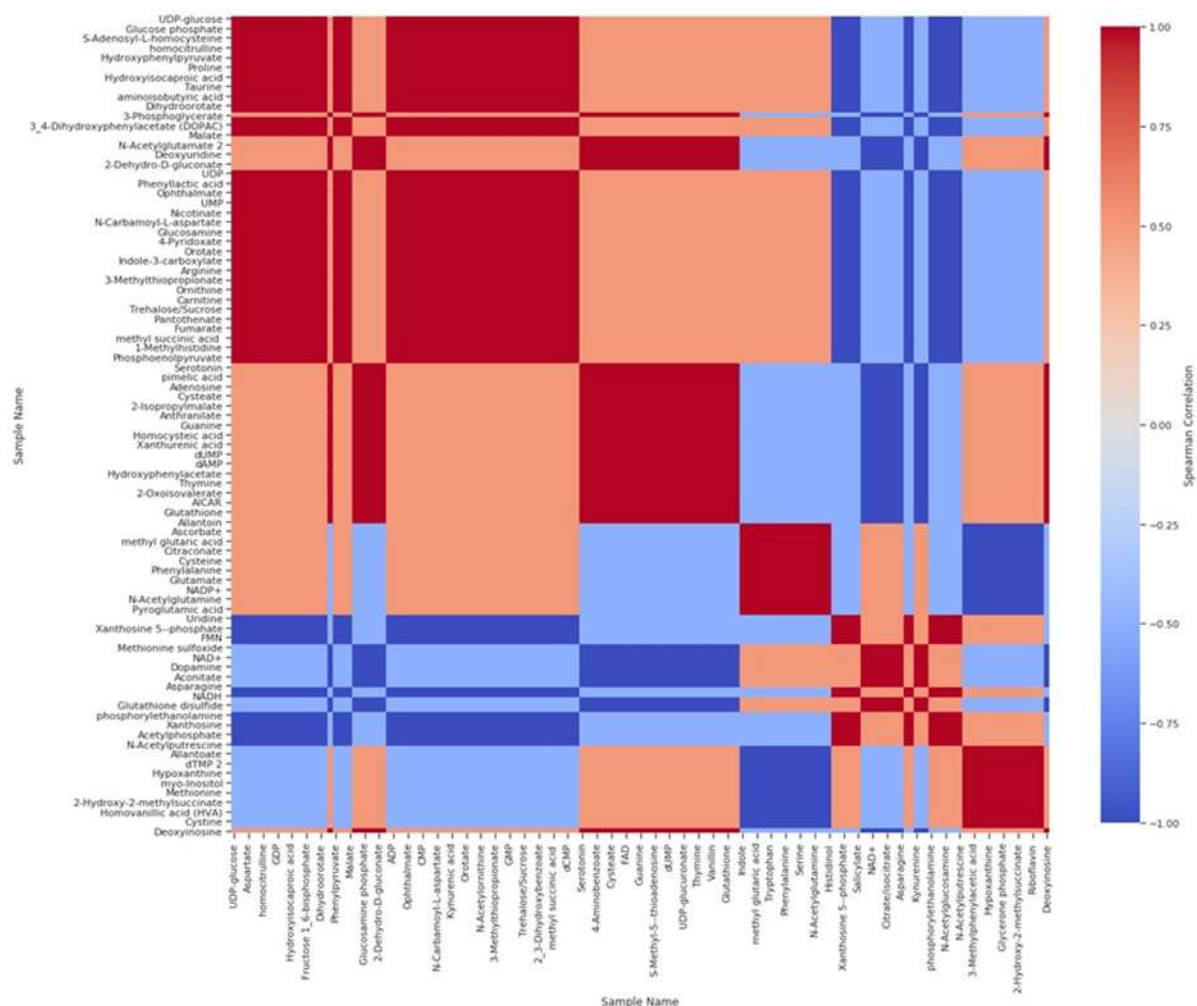


Figure S2.4: Heatmap of metabolite clusters with pairwise correlation displayed. Red indicates positive correlation between indicated metabolites, blue indicates negative correlation.

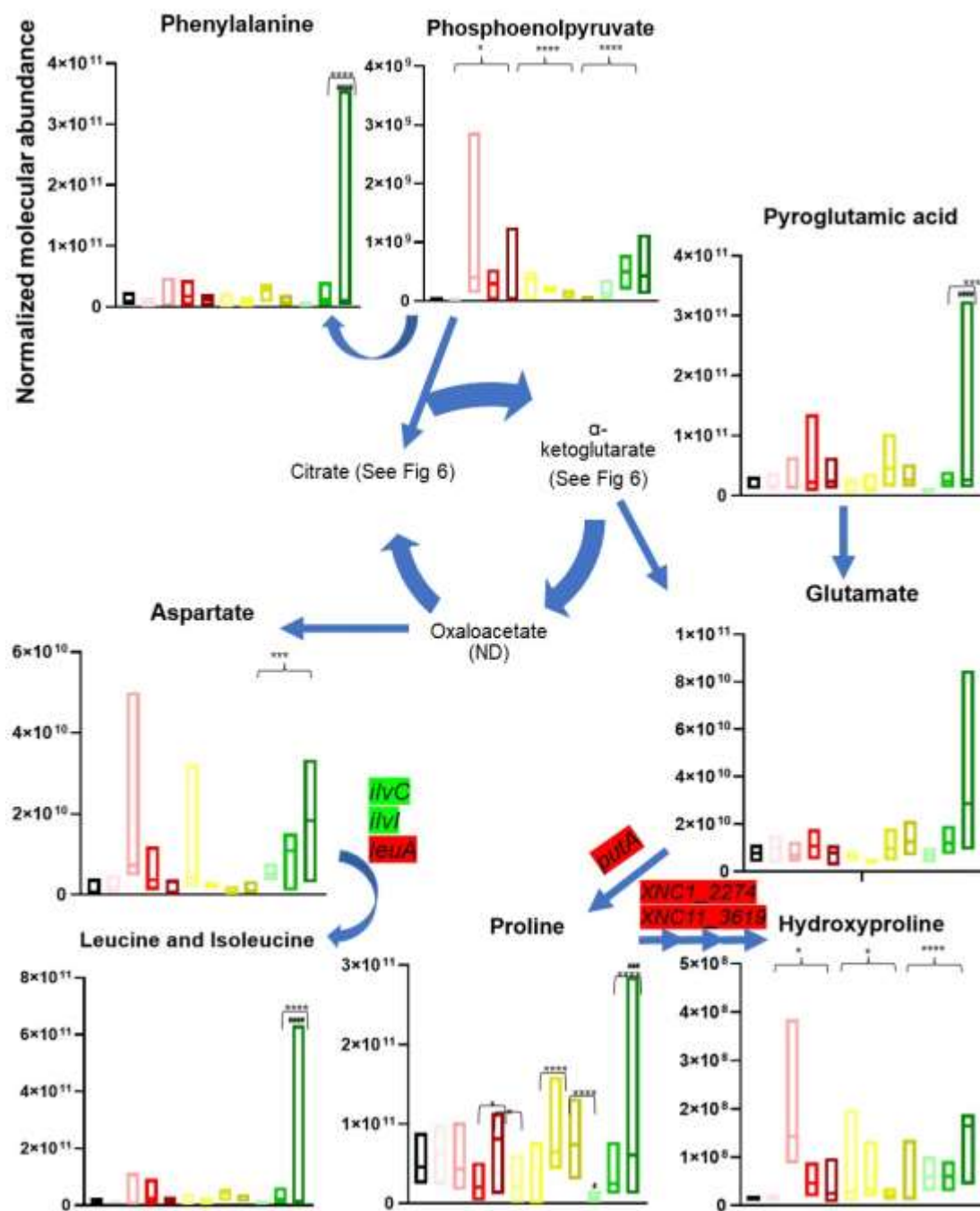


Figure S2.5: Amino acid abundances shift significantly throughout the lifecycle.

Normalized molecular abundance box plots throughout the lifecycle are shown for all detected metabolites involved in the amino acid biosynthesis. Box plot colors represent which time phase the individual plots belong to for: uninfected (black), early infection (red gradient, going from earliest, 1hr, to latest, 24 hours dead, time points), middle infection (yellow gradient, days 2-8), and late infection (green gradient, days 10-16). Lines in the middle of the boxes indicate the mean molecular abundance. Brackets indicate t-test significant abundance shifts between time phases and the uninfected insects, where $*=p<0.05$, $**=p<0.01$, $***=p<0.001$, and $****=p<0.0001$. Parentheses with * indicate two-way ANOVA with post-hoc Tukey test significance between one time point and the next subsequent time point. #s above time points indicates two-way ANOVA with post-hoc Tukey significance between a time point and the uninfected insect. ND indicates not detectable. Highlighted genes were detected as significant for the microarray in multiple strain backgrounds (See Supplementary Text C for more detail). Green indicates transcript abundance is higher in the WT background, red indicates transcript abundance is higher in the mutant background.

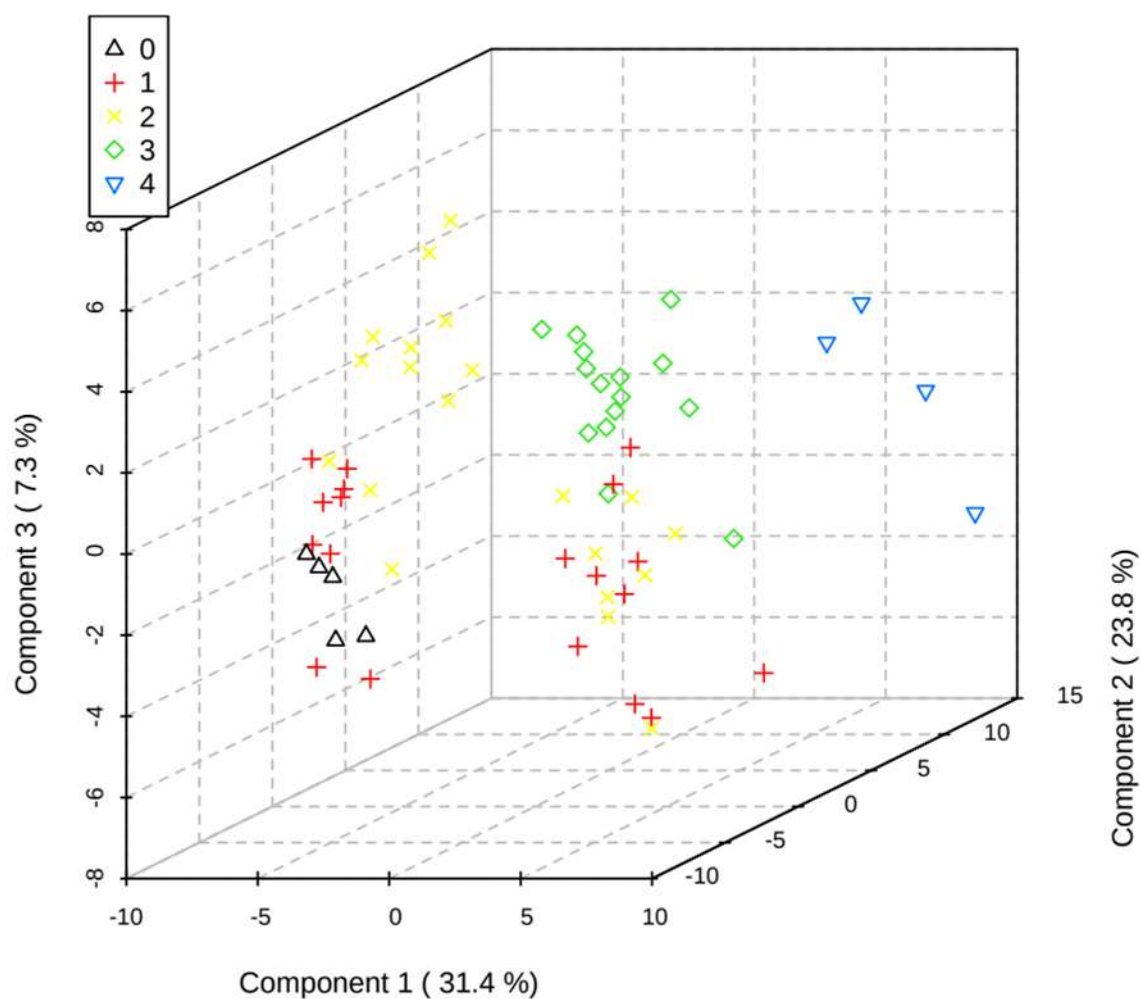


Figure S2.6: PLS-DA plots including the nematode samples. Groups are: uninfected insects (black triangles), early phase infected insects (red cross), middle phase infected insects (yellow xs), late phase infected insects (green diamonds), and input nematode IJs (blue upside down triangles). VIP metabolites contributing to the separation of these phases are available at Data S2.3.

Table S2.1: List of strains used in the microarray analysis.

Strain	Description	Reference
HGB007	Amp ^r ; <i>X. nematophila</i> wildtype ATCC 19061	ATCC
HGB151	Amp ^r ; Kan ^r ; $\Delta rpoS::kan$; HGB007	Vivas <i>et al.</i> , 2001 (60)
HGB800	Amp ^r ; <i>X. nematophila</i> wildtype ATCC 19061	ATCC
HGB1059	Amp ^r ; Kan ^r ; HGB800; <i>lrp-2::kan</i>	Cowles <i>et al.</i> , 2006 (63); Cowles <i>et al.</i> 2007 (62)
HGB1320	Amp ^r ; Kan ^r ; HGB800; $\Delta lrhA2$	Richards and Goodrich-Blair, 2010 (59)
HGB1061	Amp ^r ; HGB800 secondary form	Cowles <i>et al.</i> , 2006 (63)
HGB1103	Amp ^r ; Kan ^r ; HGB800 $\Delta nilR17::kan$	Cowles <i>et al.</i> , 2006 (63)

Relevant publications detailing how strains were constructed are listed.

Table S2.2: *G. mellonella* weight (g) upon sampling.

Group s	Un- infecte d	1 hour post- infectio n	12 hour s alive	24 hour s alive	24 hour s dead	2 day s	4 day s	6 day s	8 day s	10 day s	12 day s	16 days (Plat e 6)
Plate 1	0.21	0.20	0.29	0.22	0.26	0.2 4	0.2 6	0.1 8	0.1 8	0.1 5	0.2 3	0.05
Plate 2	0.19	0.18	0.29	0.22	0.21	0.1 2	0.2 1	0.1 4	0.1 4	0.1 4	0.1 1	0.10
Plate 3	0.21	0.19	0.22	0.15 (Plat e 6)	0.21	0.2 3	0.2 1	0.1 3	0.1 7	0.1 7	0.1 9	0.06
Plate 4	0.20	0.26	0.22	0.23	0.28	0.1 9	0.2 1	0.1 6	0.2 4	0.2 3	0.2 5	0.15
Plate 5	0.13	0.18	0.15	0.13 (Plat e 6)	0.13	0.1 6	0.1 9	0.1 5	0.1 3	0.1 4	0.2 0	0.20

Time points taken are listed.

CHAPTER 3: Integrated -omics

investigation of the newly characterized

Type 11 Secretion System (TXISS), a

factor necessary for *Xenorhabdus*

nematophila colonization of

Steinernema carpocapsae nematodes

Excerpts from this chapter were submitted as a manuscript for publication. Only the portions of which I had direct experimental contributions to (the proteomics and metabolomics analysis) are written in this chapter. The manuscript submitted to Frontiers in Microbiology special issue on Bacterial Secretion Systems, Volume II on 10/22/21 as:

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

Alex S. Grossman^{1¶}, Cristian A. Escobar-Bravo ^{2¶}, Erin J. Mans^{1¶}, Nicholas C. Mucci¹, Terra J. Mauer², Katarina A. Jones⁴, Cameron Moore¹, Paul E. Abraham³, Robert L. Hettich³, Liesel Schneider⁵, Shawn Campagna⁴, Katrina T. Forest^{2*}, and Heidi Goodrich-Blair^{1,2*}

¹University of Tennessee-Knoxville, Department of Microbiology, Knoxville, TN

²University of Wisconsin-Madison, Department of Bacteriology, Madison, WI, 53706

³Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge TN, USA

⁴Department of Chemistry, University of Tennessee-Knoxville, Knoxville, TN 37996, USA

⁵Department of Animal Sciences, University of Tennessee-Knoxville, Knoxville, TN 37996, USA

[¶]These authors contributed equally

*Co-corresponding authors

Author contributions for the work presented in this chapter:

Conceptualization: TJM, HGB

Data Curation: NCM, KAJ, PEA

Formal Analysis: NCM, HGB

Funding Acquisition: RLH, SC, HGB

Investigation: NCM, TJM, KAJ, PEB, HGB

Methodology: NCM, TJM, KAJ, PEB, SC, HGB

Project Administration: NCM, PEB, RLH, SC, HGB

Resources: RLH, SC, HGB

Software: NCM, KAJ, PEB

Supervision: PEB, RLH, SC, HGB

Validation: NCM, TJM, KAJ, PEB, RLH, SC, HGB

Visualization: NCM, KAJ, PEB, HGB

Writing- Original Draft Preparation: NCM, HGB

Writing- Review and Editing: NCM, TJM, KAJ, PEB, RLH, SC, KTF, HGB

Portions of this chapter will be submitted as a manuscript for publication to BMC Genomics as:

Bioinformatic investigation of effectors and substrates for the novel type 11 secretion system (TXISS)

Nicholas C. Mucci^{1*}, Alex S. Grossman^{1*} Heidi Goodrich-Blair¹

¹University of Tennessee-Knoxville, Department of Microbiology, Knoxville, TN

*These authors contributed equally

Author contributions for the work presented in this chapter:

Conceptualization: NCM, ASG, HGB

Data Curation: NCM, ASG

Formal Analysis: NCM, ASG

Funding Acquisition: HGB

Investigation: NCM, ASG, HGB

Methodology: NCM, ASG, HGB

Project Administration: HGB

Resources: NCM, ASG, HGB

Software: NCM, ASG

Supervision: HGB

Validation: NCM, ASG, HGB

Visualization: NCM

Writing- Original Draft Preparation: NCM, HGB

Writing- Review and Editing: NCM, ASG, HGB

Abstract

The recently identified type 11 secretion system (TXISS) is a symbiosis factor mediating mutualistic colonization of *Xenorhabdus nematophila* bacteria in *Steinernema carpocapsae* nematodes and pathogenesis in an array of human pathogens. However, we have yet to understand the extent of TXISS cargo proteins or what effectors influence TXISS expression or activity; of the over 3,000 TXISS family proteins, only 7 have demonstrated cargo. This suggests there is still much research to be done on the cargo's functional potential and the cellular processes these systems interact with. We employed a multi-omics approach to understand the impact of TXISS and TXISS cargo on *X. nematophila* colonization because we hypothesized the cargo could be moving small molecules or proteins that are binding to/cleaving off small molecules from the nematode. Proteomic and metabolomic analyses of wild-type *X. nematophila* compared to those lacking TXISS revealed differences in metal-dependent processes in cell wall and polysaccharide metabolic pathways. Next, we used a gene neighborhood and whole genome co-phylogeny analyses to broadly evaluate TXISS protein family co-occurrence, as genes within a common functional pathway often cluster together within a genome and these techniques can be used to probe this phenomenon. Bioinformatics controls were designed and implemented in both approaches. The novel techniques developed in this chapter enabled removal from the experimental results ubiquitous, non-specific co-occurring domains. This enabled identification of proteins specific to TXISS, relative to other similar types of proteins. The controlled co-occurrence lists suggest roles for TXISS and its cargo in metal homeostasis, heme biosynthesis, and biofilm formation, consistent with previous hypotheses. Utilizing sequence similarity from co-occurring proteins, 17 predicted cargo were identified with 16 having a C-terminal β -barrel with structurally varied N-termini, suggesting a structural specificity for proteins surface exposed by TXISS. The data presented in this chapter have enhanced our understanding of why TXISS is necessary for bacterial colonization by *X.*

nematophila, as well how to better narrow down functional potential from gene co-occurrence studies.

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 (8). This means roughly a quarter of the pfam database remain uninvestigated, and these domains make up proteins of unknown function that demand to be studied to better characterize basic biological processes. When DUFs have their functions defined by molecular biology techniques, this allows DUFs to be annotated and taken off the list. Molecular biology techniques are necessary to properly assign protein function, yet they are time consuming and labor intensive and so cannot 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 (129). Computational research on protein family function allows the proverbial baton to be passed to molecular biological research.

The domain of unknown function 560 (DUF560) protein family recently was defined as the type 11 bacterial secretion system (TXISS) (130). This outer membrane protein (OMP) family has homologs in a diverse array of Proteobacteria, including human pathogens in the *Neisseria*, *Haemophilus*, and *Moraxella* genera (131, 132). In the gamma-proteobacterium *X. nematophila*, TXISS is necessary for colonization of the mutualistic nematode host *Steinernema carpocapsae* (68). The nematode intestinal localization (*nil*) locus contains *nilB* (TXISS family) and *nilC* (outer membrane-associated lipoprotein), both of which are necessary for colonization. TXISS mediates secretion of both lipidated and soluble protein cargo across the outer membrane. Known TXISS-dependent effectors include co-receptors for metal-containing compounds heme, transferrin, and lactoferrin (131, 132). However, the specific range of cargo proteins and potential regulators that modulate TXISS activity has not been identified yet.

Increasing the functional potential of possible candidate TXISS-cargo pairs can allow us to modify TXISS activity by pointing to the specific metabolic pathways they are involved in.

Available sequence data have skyrocketed since the first bacterial genome was sequenced in 1995, growing nearly exponentially to having over 200,000 bacterial and archaeal genomes deposited onto online databases (133). 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 to assign function to proteins yet to be characterized. Gene neighborhood studies analyze the regions around a gene of unknown function, and this region containing nearby genetic partners could illuminate the roles of the unknown gene (134). Whole genome co-inheritance is another method of identifying unknown protein function by examining the entire genomic contexts with which these proteins co-occur with, not just particular regions. When two protein families are consistently found together in genomes it can be an indicator of connected or common function, through a term colloquially titled “guilty-by-association” (135). When one of the two protein families has a known function, this can help provide clues about the possible function of the other. Local and whole genome co-occurrence analyses can be used in tandem to identify potential function and drive hypothesis development that can be tested by focused experimental molecular experiments. However, one flaw in the current methods used to derive co-occurrence lists is the lack of controls that remove genes that would commonly co-occur with many other genes, regardless of function. For example, commonly inherited, deeply ancestral domains, like transcription and translation machinery, can encapsulate the top results but do not yield useful functional information about the specific query gene under investigation. 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.

Here, we combine “-omics” based approaches to characterize the TXISS family and their cargo’s functional potential. We probed the cellular function and potential interacting partners of one specific TXISS, *X. nematophila* NilB and its cargo protein, NilC by analyzing differences in metabolic and protein abundances in the presence and absence of this protein pair. We next used and adapted new bioinformatics approaches, for use on the entire known TXISS family to reveal global themes in the cellular processes with which these systems interact.

Results

Deletion of *nilB* and *nilC* causes global metabolome and proteome changes including impacts on peptidoglycan precursor and exopolysaccharide biosynthesis

X. nematophila requires TXISS (NilB and the outer membrane-associated lipoprotein NilC) to colonize its nematode host *S. carpocapsae* (68, 136). During the nematode infective juvenile (IJ) stage of its lifecycle, *S. carpocapsae* consumes bacterial cells in the insect cadaver and the outgrowth of those cells fill a specific nematode gut region. Our working model is that NilB is a TXISS_{OMP} that facilitates surface exposure of the lipoprotein NilC, which is a host-interaction effector. We suggest that NilC may be helping to acquire a host-derived nutrient, similar to the function of other TXISS lipoprotein effectors (e.g., transferrin and lactoferrin binding proteins) to confer colonization. To gain insights into the identities of nutrients that may be acquired by NilC, and downstream physiological effects, we compared the proteomes and metabolomes of *X. nematophila* with the locus encoding *nilB* and *nilC*, which is called Symbiosis Region 1 (SR1), and a mutant knocking that locus out (see methods for more detail, for this comparison referred to as WT and Δ SR1 respectively, though it should be noted that *nilB* and *nilC* gene expression is de-repressed in the “WT”) (Table S3.1). We grew the *X. nematophila* cells in minimal medium with glucose as a carbon source (Fig S1) and harvested cells and

supernatant at OD₆₀₀ ~0.6 for processing and analysis. Whole cell and supernatant samples were analyzed by liquid chromatography tandem mass spectrometry for proteomic analysis and ultra high performance liquid chromatography high resolution mass spectrometry for metabolomic analysis.

In the proteomics comparison of WT and $\Delta SR1$, a total of 3,336 proteins were detected, 1742 proteins in whole cell samples (Data S3.1) and 1594 proteins in the supernatant samples (Data S3.2). Of the 3,336 total proteins detected, 61 were considered to be significantly different in abundance between $\Delta SR1$ and WT based on their passing a Student's T-test filter ($P < 0.05$) and a fold change filter ($|FC| > 1$) (Table S3.2; Table S3.3). All detected proteins are represented by volcano plots taking these filters into consideration (Fig 3.1). A total of 85 features were detected in the metabolomics set and comparisons were made between the metabolomes of the 2 strains using partial least square discriminant analysis (PLS-DA) to determine the separation of the strain metabolic profiles (Fig S3.2, Data S3.3). For both the supernatant and whole cell fractions, there is clear separation of the metabolic profiles between $\Delta SR1$ and WT. T-tests were applied to determine significant metabolites between strains and 18 metabolites were significant ($P < 0.1$) for the whole cell fraction and 10 metabolites were significant ($P < 0.1$) for the supernatant fraction (Fig S3.3).

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-glucuronate, and from the PLS-DA analysis ($VIP > 1$), the nucleotide amino sugar UDP-*N*-acetylglucosamine, the precursor to the exopolysaccharide poly-*N*-acetylglucosamine (PNAG) (Fig 3.2; Fig S3.2). UDP-glucuronate is a central intermediate in the synthesis of precursors for exopolysaccharide and cell wall polysaccharide biosynthesis. The proteome revealed that a putative UDP-glucuronate epimerase (XNC1_2486), the substrate of which is predicted to be UDP-glucuronate, has lower abundance in WT relative to $\Delta SR1$, consistent with the elevated UDP-glucuronate abundances detected in the former (137). 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 (138). Consistent with this metabolic connection, according to the PLS-DA analysis, relative to $\Delta SR1$, whole cell samples of WT had higher abundances of the glycolytic intermediates 3-phosphoglycerate and fructose 1,6-bisphosphate, as well as UDP-glucuronate and UDP-glucose, lipopolysaccharide precursors. The proteomic data also indicated differences between the WT and $\Delta SR1$ strains in polysaccharide and glycolytic processes. 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 $\Delta SR1$ mutant relative to WT. In addition, while LptG (XNC1_4255), a component of the ABC transporter that moves lipopolysaccharide across the inner membrane was at higher abundance in wild type relative to $\Delta SR1$. 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 significantly higher ($P < 0.05$, |fold change difference between strains| > 1) in the WT whole cell fraction relative to $\Delta SR1$. FtsB, a cell division protein that regulates peptidoglycan biosynthesis was lower in WT relative to $\Delta SR1$ (139). WT had higher abundance of a PgaA (a.k.a. HmsH) homolog, the secretin for the exopolysaccharide PNAG (140, 141) while the mutant had elevated abundance of CsrA, a negative regulator of PNAG biofilm formation (142, 143), and XNC1_4381, predicted to encode a lipoprotein with poly- β -1,6-N-acetylglucosamine de-N-acetylase activity, similar to PgaB/HmsF (144). 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 (140, 145). Overall, the combined proteomics data indicate that the flux of amino sugar intermediates toward peptidoglycan, exopolysaccharide, and lipopolysaccharide structures is altered in the $\Delta SR1$ mutant relative to WT (Fig S3.2).

Biofilm formation is regulated by the amount of c-di-GMP, which is affected by the pool of purine nucleotides (146). Among the metabolites that were differentially abundant between the WT and $\Delta SR1$ were those involved in purine metabolism: Xanthosine, AICAR, guanine, and guanosine were detected at significantly higher abundances in the WT background whole cell fraction (Fig S3.3). 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 $\Delta SR1$ 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 (147).

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

Previous studies had revealed that TXISS-cargo pairs are involved in bacterial cell metal acquisition or immune evasion. The results described above, indicating that a TXISS-cargo protein pair is involved in bacterial cell amino-sugar metabolism, suggest that TXISS systems may function in diverse metabolic activities, not just metal acquisition and immune modulation. To gain insights into the potential breadth of biological processes in which TXISS-cargo pairs function, we sought to identify genes that co-occur with TXISS OMPs. While TXISS is ubiquitous among Proteobacteria, they are not universally conserved. Even within a species, some strains might have them and some might not. This makes gene co-occurrence very interesting to observe when considering strain specificity. We utilized the Rapid ORF Description & Evaluation Online (RODEO) software, which obtains a gene neighborhood of 6 genes upstream and downstream of the target gene under default parameters (148).

We sought to ensure that the results we obtained were unique to the TXISS protein family and not due to common features of proteins that localize to the outer membrane. Toward this end, we developed and implemented a control process in which we trimmed the RODEO output list by removing genes that were also identified in the genome neighborhoods of similarly sized outer membrane proteins, identified based on their categorization with the gene ontology (GO) term: outer membrane (GO0019867) (Fig S3.3). The median protein size of all TXISS proteins is 481 amino acids. We initially considered using mean protein size, but due to left-skewed distribution, median was selected instead to reduce bias from the smaller sized proteins (Fig S3.4). We extracted ORFs with 0.2% of this size, with the range selected based on the fact that this created a list of OMP protein accession IDs that is similar in number to the TXISS OMP query list. The protein accession IDs obtained were limited to the Proteobacteria phylum, where TXISS predominantly occurs (130). Once this control list was obtained, duplicates from the control list that appeared in the query list were removed so there was no overlap in the gene neighborhoods acquired. A random sample of this control list was taken equal to the number of protein accession IDs in the original query list (Fig 3.3).

The control set of OMP accession IDs were then used as a query for the genome neighborhood network function of RODEO, using the same parameters as the original query list. A false discovery rate (FDR) was applied by taking the number of co-occurrences of a given domain in the control list and dividing by the number of co-occurrences of the same domain in the original query list, where 5% was applied as a cutoff (Data S3.4). FASTA sequences were harvested for proteins that contain 1 or more domains from the list of domains that passed the FDR. The number of co-occurrences ranged from 928 domain co-occurrences down to 1 domain co-occurrence, and we chose to extract FASTA sequences that contained a domain using 25 appearances in the domain list as a significance cutoff. After observing a histogram of domain totals, <25 was when the domain categorizations began to steeply rise. >25 also

seemed appropriate for domains appearing in a large number of gene neighborhoods. Using 25 as a cutoff, this represents (in terms of frequency) the top 11.5% co-occurring domains detected. These sequences were then run through several functional annotation software modules, such as the KOALA family of pipelines from KEGG, eggNOG-mapper, and Blast2GO (Fig 3.4, Fig S3.5, Data S3.6, and Data S3.7) (114, 149, 150).

Functional assessment of TXISS co-occurring proteins identifies a role of metal homeostasis, amino sugar metabolism, and DNA binding regulatory proteins

Functional annotation of protein sequences is a protocol to batch search protein activity bioinformatically. There is a range of software tools available for this task and we utilized several techniques to analyze protein sequences that co-occur with TXISS domain containing proteins. KofamKOALA is a technique that uses HMMER/HMMSEARCH on a set of protein sequences against the curated Kyoto Encyclopedia of Genes and Genomes (KEGG) ortholog (KO) database (151). This analysis allowed us to group the co-occurring proteins into several major categories representative of the top 15 sequence categories. Pathogenicity island integration (transposases, tRNA modification GTPases, and tryptophanyl-tRNA synthetases) was observed in the results. tRNA genes are frequently located adjacent to pathogenicity islands indicating their importance, and they have been hypothesized to function as an anchor for horizontally acquired DNA, and some bacteriophages can use tRNA genes as their target to incorporate into a host genome (152). This suggests that TXISS is horizontally acquired. YidC/Oxa1 insertases were also detected, suggesting a required chaperon for TXISS membrane insertion (153). Metal binding proteins (hemoglobin receptor proteins, iron complex receptor proteins, and formate dehydrogenase [FdhD] proteins) and stimuli sensor proteins (transmembrane sensors, and RNA polymerases sigma-19 and -70 factors) were also in the top 15 sequence categories (Fig 3.4A). Sigma-19 (or FecI) is considered an “iron starvation sigma” and has been shown to cause expression of genes for uptake of ferric citrate (154). We also employed BlastKOALA to align protein sequences against a nonredundant set of prokaryotic

KEGG genes using BLAST searches, which uses sequence similarity instead of HMM searching (114). This further establishes the idea that TXISS co-occurs with sensor proteins, particularly a high proportion of transporters that are part of the “signaling and cellular processing” (transporters, other secretion systems) and “environmental information processing” (ABC transporters, 2-component systems) categories. This analysis also pointed to several metabolic pathways that could be related to TXISS activity, including folate cofactor metabolism, amino acids (glycine, serine, threonine, cysteine, and methionine), and carbohydrates (glyoxylate, pyruvate, and propanoate) (Fig 3.4B).

To increase the scope of functional annotation software, we opted to batch search function using the eggNOG-mapper tool which assigns orthologs using the EggNOG database (149). This tool uses precomputed orthologous groups from the EggNOG database to determine “fine-grained orthologs”, which are phylogenetically refined orthology assignments not acquired as a result of a duplication event. Clusters of orthologous genes (COG) classification from this analysis indicated that ORFs that co-occur with TXISS OMPs are involved in inorganic ion transport, inorganic ion transport coupled with signal transduction mechanisms, transcription and translation, and lipid transport (Fig 3.4C). Similarly, the Blast2GO pipeline (Fig S3.5, Data S3.6, and Data S3.7) assigned TXISS co-occurring ORFs to the inorganic ion and lipid transporters and binding proteins classes, as well as an emphasis on nucleotide and DNA binding. Several classes of hydrolases and transferases were annotated as well, including pyrophosphatases which energetically couple hydrolysis of pyrophosphate to hydrogen and sodium ion pumping. Metabolic GO terms were nested primarily within macromolecule and nitrogen-containing compound biosynthesis.

Controlled whole genome co-inheritance assay of TXISS homologs

The data described above provided insights into co-occurring genes that are encoded in narrow genomic proximity of TXISS OMPs. However, the TXISS OMPs Slam-1 is encoded at a

distance from the gene encoding its cargo protein, transferrin binding protein (TbpB) (131). Therefore, we sought to eliminate the genomic association restriction to expand the list of potential TXISS cargo or other interacting proteins, which is a technique called gene co-inheritance. We employed a phylogenomics approach using the PhyloCorrelate software which can detect protein families (not individual ORFs) with comparable phylogenetic distributions across 28315 bacterial species from the Genome Taxonomy Database (GTDB) (155, 156).

To focus our attention on the most significantly co-inherited domains, we applied the runs-adjusted Jaccard overlap coefficient (rJC) metric. This is a model-based approach that reduces phylogenetic redundancy, and therefore reduces the presence of commonly occurring pfams from co-occurrence outputs. The phylogenetic redundancy is reduced by compaction of repeated co-inheritance patterns that occur within certain lineages (157). rJCs are calculated for each pairwise comparison of phylogenic profiles across the 28315 species within the dataset. The two gene profiles are compacted by collapsing gene comparisons of presence-presence or absence-absence for the JC calculation if they appear in the same clade (155). So, if a gene is co-inherited with another gene, but the genomes with which they appear are phylogenetically adjacent to each other, the gene is counted once not twice in the JC. This compression thus limits long “runs” of genes that are frequently co-inherited, and ancestral, deeply conserved genes are limited. Once runs-adjusted, the Jaccard coefficient is calculated as the intersection between the probed data sets (the TXISS family query against the other 11340 protein families within 27372 genomes comprising the GTDB), where the higher the rJC value is, the more that family is co-inherited with the TXISS family query.

We initially opted to use 3 different stringency filters, where $rJC > 0.45$ was the most stringent and $rJC > 0.35$ was the least restrictive. These values were chosen as they are equal intervals and $rJC < 0.35$ families start to rapidly climb in number and thus assumed to be more ubiquitously co-inherited and less useful for analysis. However, the initial results showed that

the top candidates co-inherited with TXISS were tRNA synthetases, helicases, and other protein translation machinery. Under the rJC metric, these ancestral, ubiquitous genes present in nearly every bacterial genome were reduced in the results, but still represented a significant signal of co-inheritance, masking the more meaningful co-occurrence data. To address this issue, we layered onto the rJC metric an additional control, comprising an essentially random protein family co-occurrence to which the experimental data set could be compared, to remove common features and leave protein families that distinctively co-occur with TXISS. As a positive control, we assessed the range of rJC that included known bona fide TXISS-associated protein families. Out of 7099 TXISS co-occurring protein families in the rJC distribution, TonB-dependent receptors (PF00593), which TXISS cargo proteins interact with, ranked at #177 and TbpB_B_D (PF01298), the C-terminal β -barrel domain present in the TXISS cargo protein, ranked at #1805 (131). This could suggest that TXISS primarily functions through nutrient acquisition through structurally similar TonB-dependent uptake systems, while the variability of the cargo's C-terminal is broad across Proteobacteria. We used the TbpB_B_D protein family ranking as an approximate upper bound, rounding up to 2,000, of co-inherited pfams (as measured by rJC values). This new approach, which considers the co-inheritance distribution of known interacting partners, was examined over the stringent tiered $rJC > 0.45$, 0.40 , and 0.35 employed earlier.

Next, we created a control protein family co-occurrence list as a negative control (to remove non-specific co-occurrences from the list of 7099 TXISS co-occurrence protein families). We opted to use the LPS transport system D (LptD, PF04453) family. Like TXISS, this family is widely distributed among the Proteobacteria, is localized to the outer membrane where it facilitates lipoprotein surface exposure (158), and has an rJC distribution profile similar to TXISS. This profile is bimodal, with a peak at < 0.05 and again around 0.30 , and reaches the lowest levels around 0.40 . While the median size of LptD homologs is larger than that of TXISS

OMP homologs (785 amino acids compared to 469 amino acids for TXISS), the functional and localization similarities were deemed more important than size for the purposes of creating a control group for the identification of specific co-inherited protein families. The top 2000 co-inherited protein families of LptD were acquired and this list was compared to the TXISS list. Any protein family that appeared in both lists was trimmed, leaving a list of 531 that were deemed specific to TXISS (Data S3.5). This new list contains TbpB_B_D, but not TonB or TonB-dependent receptors.

The top hits for TXISS whole genome co-inherited proteins are membrane and biofilm assembly (peptidoglycan-binding, lipid biosynthesis, the control group LptD, inositol phosphate, and LolB), metal binding domains (oxidoreductases, SecB, alcohol dehydrogenase, and metal-dependent hydrolase), heme biosynthesis (HemX and UroD), and transcriptional regulators (Fis regulator and DNA-binding domains). The list was further organized by cellular localization using 3 major categories: predicted secreted proteins (which could represent TXISS cargo), predicted cytoplasmic domains (which could represent effectors that influence TXISS expression), and predicted periplasmic domains (which could represent a TXISS interactor). Annotations were assigned using GO terms based on searches using the GOMiner platform, which utilizes a computational discovery-based approach to assign GO terminology using gene neighbors (159). Of the top 50 co-inherited domains, 47 occur in *Xenorhabdus* genomes as determined through searching protein families in the GTDB (156). The domains not present in *Xenorhabdus* genomes are DUF1178 and DUF815, and poly(3-hydroxybutyrate) depolymerase C-terminus. This new list of domains serves as an expansion of possible TXISS cargo and interactors, increasing the breadth of our understanding of this novel secretion system.

Bioinformatic investigation of additional TXISS cargo and their structural prediction

To experimentally evaluate our predictions that the controlled gene co-occurrence yielded new TXISS cargo proteins, we investigated the structures of proteins that contained at least 1 domain and assessed their similarity to experimentally tested TXISS cargo. FASTA sequences were harvested for proteins that contain 1 or more domains from the list of domains that passed the FDR. These sequences were then BLAST searched to known TXISS cargo domains. 50% sequence identity was used as a cutoff comparing our list to the β -barrel and handle domains of: NilC, TbpB, HrpC, CrpC, Haemophilin, Factor H binding protein, and LbpC. This created a new list of 17 potential cargo proteins, more than doubling our original list of cargo proteins.

Using PSIPred 4.0 (160), the secondary structures of these 17 proteins were examined (Fig 3.6). Nearly each protein is predicted to have a long, disordered N-terminal region. Most are β -strand rich, consisting of ~5-12 strand residues parsed out between coils, similar to HrpC and NilC which were used for comparison. However, 4 of the 17 proteins have long helix repeat regions. Using AlphaFold (161), the tertiary structures of these 17 proteins were also examined (Fig 3.7). 16 of the 17 have a C-terminal β -barrel. We employed LipoP prediction for distinguishing lipoproteins (162), which identifies lipoproteins by locating lipidated cysteine residues. From this prediction, we determined 8 of the 17 potential cargos are lipoproteins.

Discussion

Protein secretion mechanisms are critically important for the colonization efficiency and pathogenicity of bacteria (163). These nanomachines are essential for bacterial adaptation to a broad range of environmental conditions, including host cell surfaces. They can be reengineered for vaccine development as well, such as the Type 3 secretion system which can deliver antigens (164). Still, open questions remain on how secretion systems and their

molecular cargo's activity can be modulated, especially for recently discovered secretion systems like TXISS. TXISS is a characterized symbiosis factor that is distributed across Proteobacteria, including several human pathogens (130). Understanding what factors affect cargo secretion and what cargo activity is paramount to characterizing how TXISS contributes to host-associated phenotypes. Here, we employed a combination of genomic, proteomic, metabolomic, structural, and molecular techniques to further characterize the recently identified type eleven secretion system (TXISS). This work focused on using a combination of bioinformatics methods using -omics data to elucidate more possible TXISS homologs, their cargo, and their cargo's functional potential. TXISS cargo appears to have roles in metal homeostasis which can be linked to biofilm metabolism and nutrient provision.

TXISS systems appear to be associated with several categories proteins and metabolites that TXISS could be moving or interacting with. Previous research has found that TXISS cargo are coreceptors for TonB-dependent nutrient uptake systems, and our research has expanded the types of nutrients that TXISS cargo could be transporting (130). Amino sugar, peptidoglycan, and metal-related proteins and metabolites were significantly affected by the presence vs. absence of TXISS. As discussed above, the altered polysaccharide pathway and glycolysis metabolite homeostasis in the $\Delta SR1$ strain relative to its isogenic SR1+ strain may underlie the change in PgaA abundance, and may be transduced through the RNA-binding protein CsrA (carbon storage regulator), which is more abundant in $\Delta SR1$ 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 (142, 143, 165, 166). CsrA negatively regulates PgaA translation (167, 168). Further, 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 (169). One caveat of our work is that the metabolomics and proteomics analyses were conducted on cells that lack the transcription factor NilR, a global regulator of many genes, not just *nilB* and *nilC* (63). As such, the metabolites and proteins that were detected as differentially abundant in the presence or absence of SR1 may only have been apparent due to the absence of NilR regulation in both strains. We noted two differentially abundant proteins encoded by genes that, like *nilB* and *nilC*, are NilR repressed as detected by microarray analysis (170). XNC1_1277 and XNC1_4587 had lower and higher abundance, respectively in the $\Delta SR1 \Delta nilR$ strain, relative to the SR1+ $\Delta nilR$ strain. XNC1_1277 (YfiA/RaiA) is predicted to encode a ribosome stabilizing factor. *E. coli* YfiA/RaiA is a carbon-starvation inducible ribosome hibernation factor that sequesters ribosomes in an inactive state, reducing growth and promoting resistance to stress (171, 172). XNC1_4587 is predicted to encode a secreted, DUF930 domain containing protein, and no function has been ascribed to homologs to date. However, it is encoded adjacent to genes predicted to encode a periplasmic protein of unknown function, and a trehalose-specific phosphotransferase system. This genomic organization is consistent with homologs in other organisms (e.g., *Mycoplasma mycoides* subsp. *mycoides* NC_021025; MMS_A0824). Overall, 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.

We sought to follow up on TXISS cargo activity probed in the proteomics and metabolomics analyses by using gene co-occurrence techniques. Gene co-occurrence tells us which protein sequences are near each other as measured by proximity, either encoded nearby the gene of interest or found somewhere else in the genome. Genes within a common functional pathway often cluster together within a genome, so this was a technique we employed to find more potential TXISS cargo and their function not just within the *Xenorhabdus* genus, but across Proteobacteria. For this work, we established *in silico* controls for gene neighborhood

and gene co-inheritance assays. These controls were implemented to reduce false positives and noise in the output datasets of putative TXISS co-occurring ORFs or protein families. For instance, activities associated with protein translation and RNA metabolism, which are essential cellular processes expected to be present in every cell, are not useful co-occurrence signatures for elucidating TXISS function and components. Similarly, functions that are broadly involved in outer membrane homeostasis are likely to be associated with TXISS, as they comprise integral outer membrane proteins. However, such general functions are not indicative of the specific functions of TXISS relative to any other OMP. Our work lays groundwork upon which other researchers can build for incorporating thoughtfully designed, tailored controls for refining co-occurrence data. However, one caution is that incorporating such controls may cause some biologically interesting data to be lost. For example, there could be protein families that are important partners, but that would be removed from the dataset if they were important for both query and control proteins. Indeed, we found that TonB and TonB-dependent receptors, which some TXISS homologs interact with, was trimmed out under both co-occurrence analyses. Characterized TXISS cargo have been shown to be coreceptors for TonB-dependent metal uptake, so thus trimming these out of the final results loses crucial information (173). However, this is foundational work that can be expanded upon and improved. Multiple random samples of gene neighborhood controls can be automatically incorporated into traditional gene neighborhood analyses instead of the 1 random sample described in this work. GO terminology can be swapped out for a different tool to understand localization, like KEGG ontology. Swiss-Prot, which is a manually annotated and reviewed protein database, can be utilized instead of the entirety of UniProtKB which includes hundreds of millions more automatically annotated proteins.

Protein domains that co-occur with TXISS are consistent with previous data and hypotheses. A subset of characterized TXISS are heme receptor proteins. In the controlled

gene neighborhood of TXISS, heme oxygenases, hemoglobin utilization, and FeoB which facilitates iron uptake (174), were among the top 20 protein categories. Uroporphyrinogen biosynthesis proteins HemX and UroD, which are involved in the generation of heme biosynthesis intermediates (175), were among the top protein categories for the controlled genome co-inheritance assay. These proteins function at a branchpoint early in heme biosynthesis that can instead lead to Vitamin B12 (cobalamin), which has been found to increase the volume of *S. carpocapsae* nematodes when grown with *X. nematophila* subcultured with B12 (176). Several more B12 biosynthetic enzymes (PF01923, PF11775, PF02654, PF07683, PF02965) are co-inherited with TXISS. Perhaps these data support the concept that EPNB-associated TXISS cargo activity can modulate the predation of their host, where under some conditions, heme is acquired from the environment for iron scavenge whereas in other conditions, B12 can be procured to grow the nematode brigade. In addition to iron, molybdopterin cofactors metabolism came up in both analyses. Bacteria require molybdenum for iron-sulfur cluster biosynthesis and other metabolic pathways (177), and in eukaryotes (this interaction is not well characterized in prokaryotes) molybdopterin cofactors bind heme as a redox active cofactor (178). Molybdenum is transported through ABC transporters into the cell from the environment, either from low affinity (MolABC) or high affinity (ModABC) transporters (179), and reduction of Molybdenum uptake leads to a virulence defect in Gram-negative bacteria (180). TXISS cargo could be moving involved in either molybdenum uptake or metabolism into the molybdenum cofactor. Future studies should probe whether TXISS increases secretion of molybdenum transporters or see if TXISS mutants are deficient in molybdenum uptake/molybdenum cofactor biosynthesis. TXISS also co-occurs with cell membrane recognition and assembly proteins (such as PGAP1 found in lipases, SPFH found in lipid raft proteins, transmembrane sensors, and inositol monophosphatase which synthesizes phosphatidylinositol) that use metal cofactors. These data suggest that TXISS cargo could work to maintain membrane integrity amidst host cell environmental conditions, which is in line with

NilB and NilC being necessary for bacterial colonization of *S. carpocapsae* (68). TXISS also co-occurs with various nutrient biosynthetic proteins (namely, folate, amino acids, and carbohydrates). Together, this data could suggest TXISS and TXISS cargo interact with the host cell surfaces to scavenge metals for their own nutrient provisions and promote colonization efficiency through maintaining membrane integrity.

We used the gene co-occurrence analyses to identify a set of potential cargo for TXISS. Clustering the list of co-occurring protein families to known cargo domains (such as the HrpC β -barrel and the CrpC handle domains which are strand rich regions adjacent to the barrels), we identified a short list of 17 potential cargo. 16 of the 17 cargos have a C-terminal β -barrel, indicating a structural specificity for TXISS to recognize cargo to surface expose. The N-terminals of these proteins varied structurally, which could be performing host-associated activities. To determine if our search of potential TXISS cargo was accurate, we have begun investigating Plasmin-sensitive surface protein (Pls) to see if it could be localized to the bacterial cell surface in a TXISS-dependent manner. Like all other known TXISS cargo proteins (130), it is predicted to have a C-terminal β -barrel. BLAST searching Pls reveals similarities to proteins that bind erythrocytes, which contain heme. Heme receptor proteins in *X. nematophila* are TXISS homologs and have been biochemically characterized. In addition to Pls, a colicin transporter in *Haemophilus haemolyticus* was found in our bioinformatics screen. Colicin is a deadly toxin for bacteria closely related to *E. coli* and are transported through TonB-dependent transporters or the Tol system to degrade DNA or RNA, degrade peptidoglycan, or through making ion channels to destabilize the cells (181). A homolog of this colicin transporter is present in the genome of *Actinobacillus pleuropneumonia* (TolA2), a Gram-negative bacterium that causes a highly contagious and deadly disease in swine (182). This colicin transporter is being developed as a vaccine target against *A. pleuropneumonia*, making its translocation to the outer membrane an integral part of modulating this bacterium's pathogenicity. Thus, our

bioinformatics screen has several more candidates to test their secretion by TXISS. We recently started testing aimed at elucidating if PIs is surface exposed by TXISS. Preliminary results indicate that PIs secretion significantly increases when expressed with TXISS (data not shown), indicating the robustness of our cargo search method.

Additionally, bioinformatic motif searches could expand the list of potential TXISS cargo. The Multiple Em for Motif Elicitation tool (MEME) and the Gapped Local Alignment of Motifs (GLAM2) tools could identify repeating motifs in our potential and experimentally characterized TXISS cargo proteins (183, 184). The outputs derived from these tools can be used with the Motif Alignment & Search Tool (MAST) using sequence patterns to find additional cargo sequences (185).

In summary, our findings have enhanced our understanding of potential effectors of the novel TXISS, as well as expanding our list of cargo that TXISS secretes. We also have expanded the field of gene co-occurrence for unknown protein function analysis. With the techniques described in this chapter, all future gene co-occurrence studies should use these protocols as a basis for their work or expand the designs. Controls can then be automatically integrated into these software packages for ease of use. Future opportunities for discovery involving TXISS involvement in host attachment via biofilms and metal acquisition will be explored.

Materials and methods

Proteomics and metabolomics sample preparation

HGB 1495 and HGB1496 (Table S3.1) were struck from frozen stocks onto LB pyruvate plates and incubated at 30°C overnight. Three biological replicates for each strain were inoculated in 50mL of minimal medium (136) in a sterile 500mL flask, and incubated at 30°C, 200rpm to an OD600 of 0.6. According to Bhasin et al. 2012, growth in this medium to an OD600 of 0.6

should maximize *nilB* expression and took 55 hours. Cell lysis was monitored at 260-280 Abs. The 50mL culture was divided in two at this point, with one set of samples prepared for the Campagna lab for metabolomic processing and one set for the Hettich lab for proteomic processing. For proteomic analysis, cell pellets were suspended in SDS lysis buffer (2% in 100 mM of NH_4HCO_3 , 10 mM DTT). Samples were physically disrupted by bead beating (0.15 mm) @ 8k rpm for 5 min. Crude lysates were boiled 5 min at 90 °C. Cysteines were blocked by adjusting each sample to 30 mM iodoacetamide and incubated in the dark for 15 min at room temperature. Proteins were precipitated using a chloroform/methanol/water extraction. Dried protein pellets were resuspended 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 ug 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 hr at 37 °C. The peptide mixture was adjusted to 0.5% FA to precipitate SDC. Hydrated ethyl acetate was added to each sample at a 1:1 [v:v] ratio three times to effectively remove SDC. Samples were then placed in a SpeedVac Concentrator (Thermo Fischer Scientific) to remove ethyl acetate and further concentrate the sample. The peptide-enriched flow through was quantified by BCA assay, desalted on RP-C18 stage tips (Pierce Biotechnology) 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 0.1M formic acid in 4:4:2 acetonitrile:water:methanol according to the procedure described previously (186).

LC-MS/MS and UPLC-HRMS analysis

All samples were analyzed on a Q Exactive Plus mass spectrometer (Thermo Fischer Scientific) coupled with a with a Proxeon EASY-nLC 1200 liquid chromatography (LC) pump (Thermo Fisher Scientific). 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 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 was 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 1×10^6 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 was 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 1×10^6 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-performance liquid chromatography coupled to high resolution mass spectrometry (UPLC-HRMS) (Thermo Scientific, San Jose, CA, USA) was used to analyze water-soluble metabolites (Metabolomic Analysis via Reversed-Phase Ion-Pairing Liquid Chromatography Coupled to a Stand Alone Orbitrap Mass Spectrometer). A Synergi 2.6 μm Hydro RP column 100 Å, 100 mm x 2.1 mm (Phenomenex, Torrance, CA) and an UltiMate 3000 pump (Thermo Fisher) were used to carry out the chromatographic separations prior to full scan mass analysis by an Exactive Plus Orbitrap MS (Thermo Fisher). HPLC grade solvents (Fisher Scientific, Hampton, NH, USA) 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.

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) (33) to which common contaminate 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 were 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, a 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 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, retention time window for extracted ion chromatogram was 3 min, time window to get the apex for 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 log2-transformed. Using the freely available software Perseus (<http://www.perseus-framework.org>), 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.

Gene neighborhood and phylogenomics analysis

The TXISS gene neighborhood (as measured by +/- 6 genes to a TXISS gene) and TXISS co-inherited genes (as measured by gene appearances within the entire genome of which a TXISS gene occurs within) was analyzed. TXISS Cluster 1 (130) protein sequences were evaluated through Rapid ORF Detection & Evaluation Online (RODEO) (148) for the gene neighborhood analysis. 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 TXISS (Fig 3.3). Controlled protein sequences (as determined by size similarity to TXISS, outer membrane localization using GO, and distributed throughout Proteobacteria) were downloaded from UniProt (187) in February 2021. PhyloCorrelate, which detects pairs of genes with similar phylogenetic distributions, was employed for the whole genome co-inheritance analysis using the PF04575 (pfam number for TXISS) search term. Gene pair filters were set to very low confidence scores for initial inquiries to detect every possible gene that is co-inherited with TXISS, and then further refined using their rJC scoring algorithm. rJC is a metric that calculates the Jaccard coefficient between gene pairs, which returns a frequency at which genes are co-inherited with each other. Prior to Jaccard coefficient calculations, the gene pairs are runs-adjusted, meaning a phylogenetic correction is applied to restrict highly conserved genes. Domains that appear in proteins that are frequently encoded in the same genomic context as TXISS (e.g., TonB-dependent receptors and TbpB_B_D) were located in the rJC distribution. Next, the number of co-occurring domains was rounded to the nearest thousand (2000). This list of 2000 co-occurring TXISS domains was then compared to a control list based on the predicted co-occurring domains of LptD (a protein family of similar localization, function, and phylogenetic distribution to TXISS). Duplicates between the two lists were removed, yielding a TXISS co-occurring domain list of 1531 protein families. Domains for each protein family were examined, as were the GO term enrichments for each set of domains.

Functional and structural analyses of co-occurring gene neighborhood domains

Protein sequences that contain domains that co-occur with TXISS from the gene neighborhood analysis were obtained via UniProt (187). KofamKOALA (151), BlastKOALA (114), and eggNOG-mapper were used (149) 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. eggNOG also uses fast orthology assignment for user sequence data against their annotation database, but takes into account orthologs that may have arisen from duplication events and restricts them in the results. Blast2GO (150, 188) was also used for additional analysis of co-occurring protein GO terms, where proteins were blasted, mapped, annotated, and clustered together for similarities using default program settings. All downloaded protein sequences were then subjected to a BLAST search against known TXISS 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*, cobalamin receptor protein (CrpC) in *Xenorhabdus cabanillasii*, Haemophilin, heme receptor protein (HrpC) in *Acinetobacter baumannii*, 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 TXISS cargo based on 50% sequence similarity, generating a list of 17 proteins. Structural data predictions for these 17 proteins were generated through PSIPRED (160) for secondary structure and AlphaFold (161) for tertiary structure. PSIPRED uses position specific iterated-BLAST outputs against their designed neural network to analyze secondary structure. AlphaFold also uses a machine learning method to accurately determine the attraction and repulsion between the amino acids to predict structure. Tertiary protein structure was visualized in ChimeraX (189).

Acknowledgements

We thank Hector Castro for use of the mass spectrometry facilities. We thank Jordan Rogerson for her initial help in metabolomics sample preparation and analysis. We thank the HGB Lab members for feedback on the co-occurrence controls.

Data and materials availability

All data are available in the main text or the supplementary materials.

Appendix

Chapter 3 figures and tables

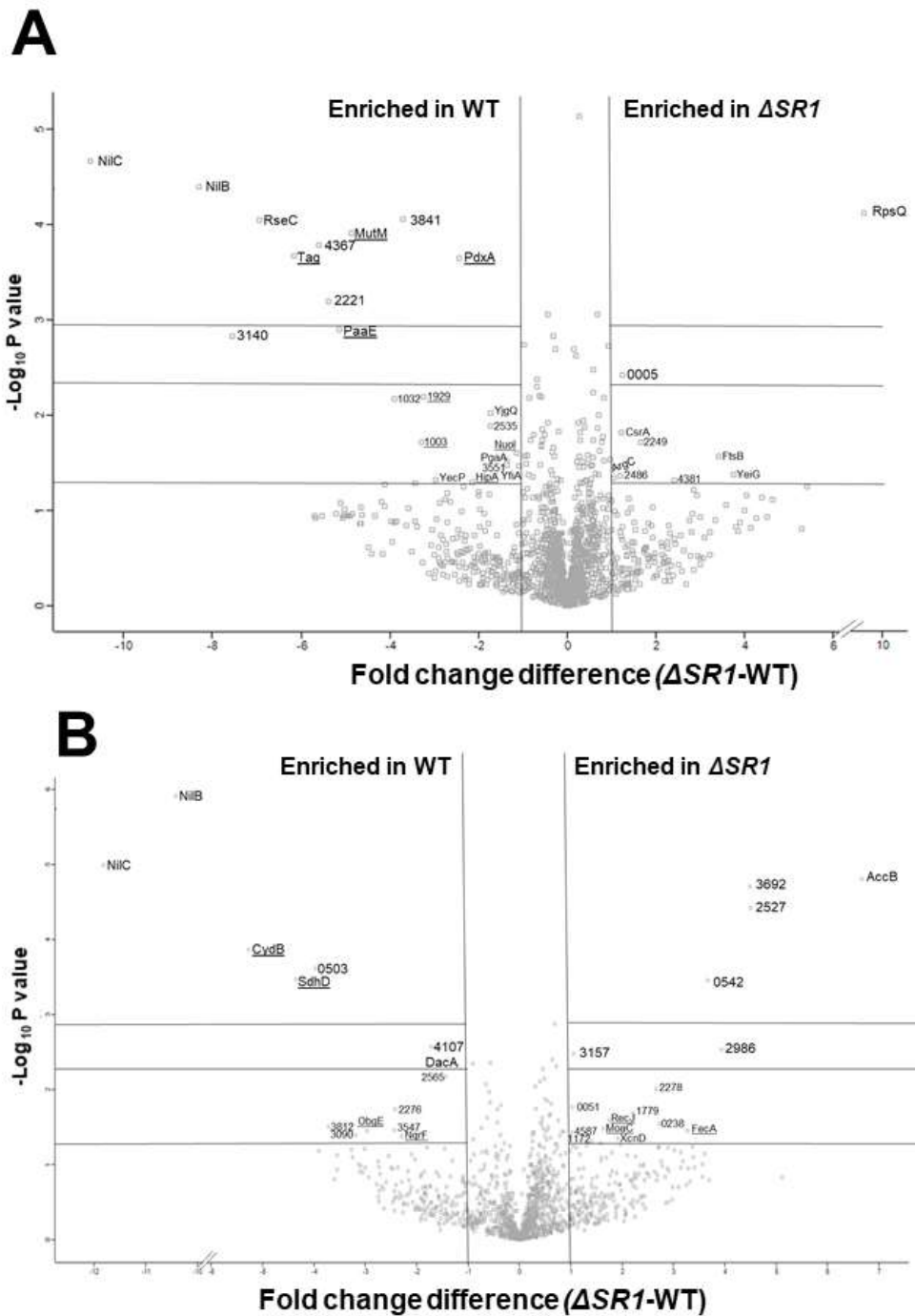


Figure 3.1: Differentially translated proteins between $\Delta SR1$ and WT show significant differences in metal binding, homeostasis, and metabolic pathways.

Volcano plot representations of proteins differentially present in *X. nematophila* wild type and $\Delta SR1$ 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 $\Delta SR1$ 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.

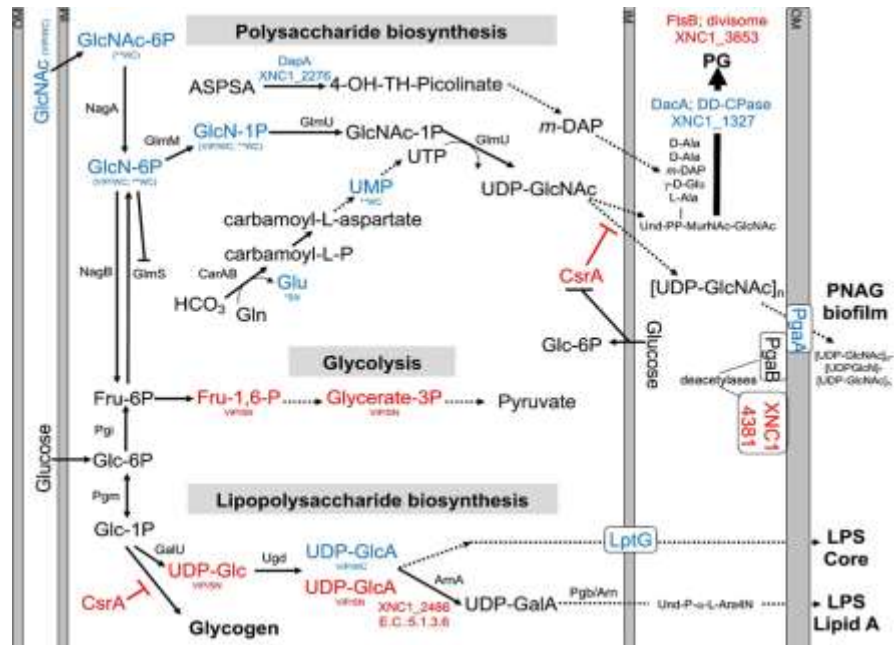


Figure 3.2. Predicted pathways for amino sugar metabolism leading to peptidoglycan (PG), exopolysaccharide (poly-N-acetylglucosamine (PNAG)), and lipopolysaccharide LipidA biosynthesis, and glycolysis pathways. Blue and red font text indicates higher and lower abundance of the indicated metabolites and proteins in WT relative to $\Delta SR1$. 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 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. Abbreviations: N-Acetyl-Glucosamine (GlcNAc-6P); Glucosamine-6-phosphate (GlcN-6P); Glucosamine-1-phosphate (GlcN-1P); N-Acetyl-Glucosamine-1-phosphate (GlcNAc-1P), Uridine diphosphate N-acetyl glucosamine (UDP-GlcNAc); Undecaprenyl (Und); N-Acetyl-Muramic Acid (MurNac); Fructose-6-phosphate (Fru-6P); Fructose 1,6 bisphosphate (Fru-1,6-P); 3-Phosphoglycerate (Glycerate-3P); Glucose-6-phosphate (Glc-6-P); Glucose-1-phosphate (Glc-1P); Uridine diphosphate glucose (UDP-Glc); Uridine diphosphate glucuronate (UDP-GlcA); Uridine diphosphate galacturonate (UDP-GalA); L-4-aminoarabinose (L-Ara4N). Dashed lines indicate multiple steps in the indicated pathway.

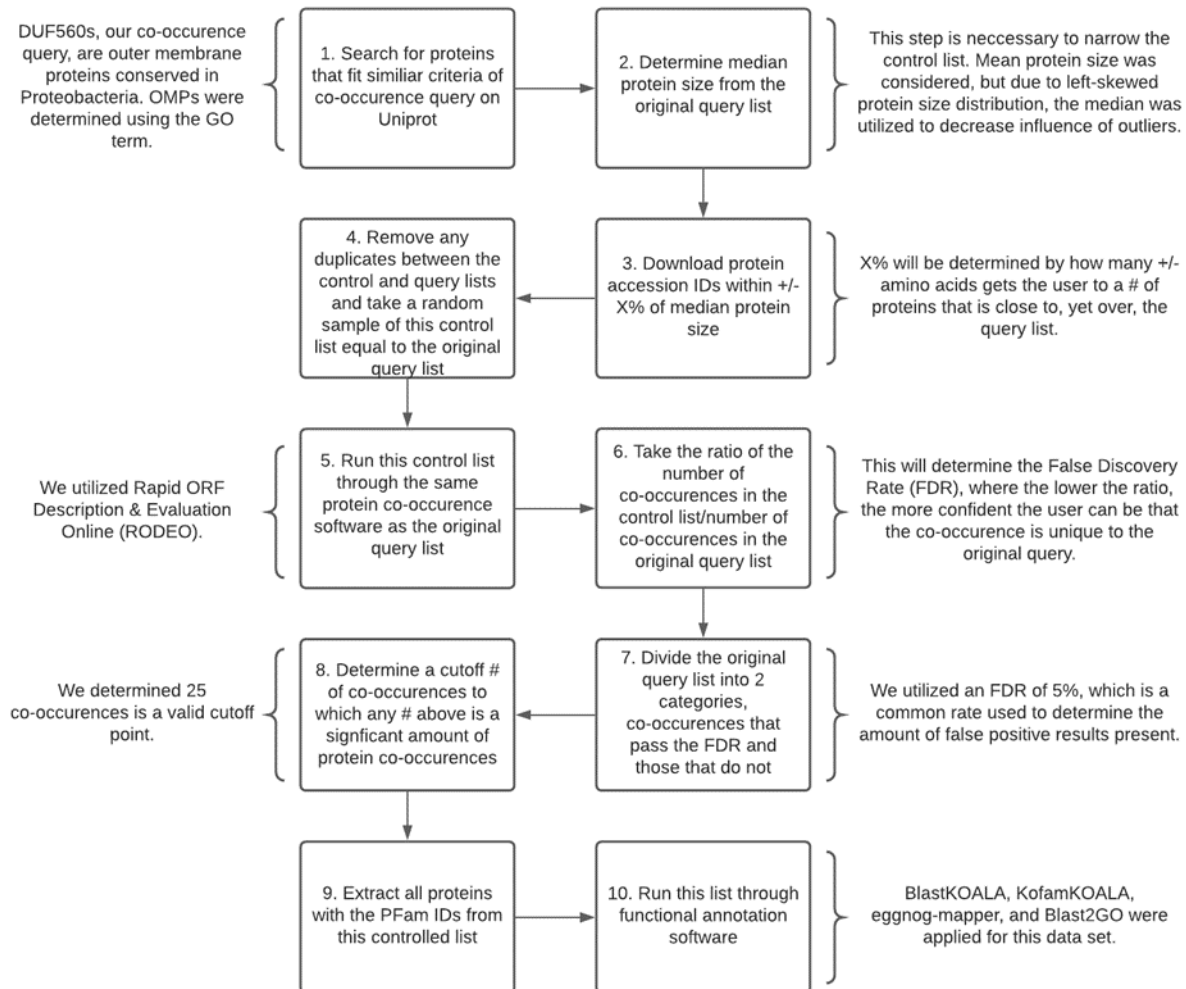


Figure 3.3: Design for a novel control implementation technique for gene neighborhood

studies. Schematic diagram with commentary from our study on how to run a gene neighborhood analysis with an implemented control, developed for this study. Created in Lucidchart, www.lucidchart.com.

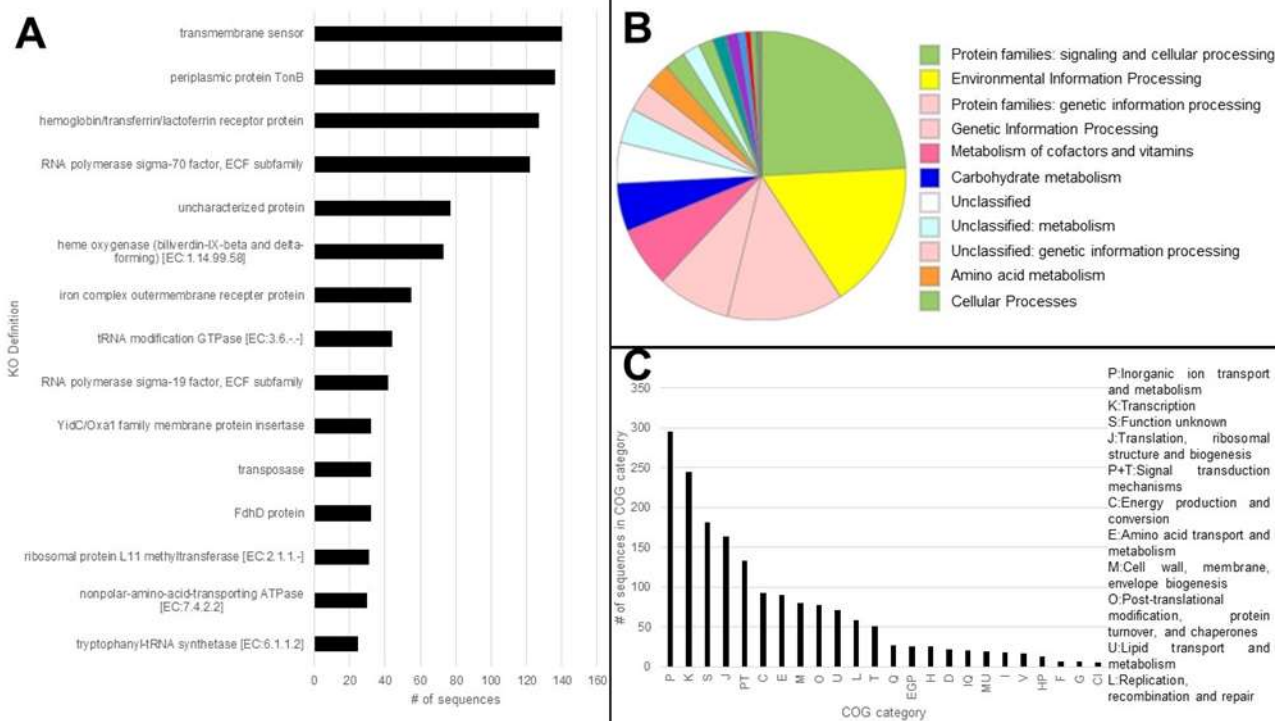


Figure 3.4: Functional assessment of protein sequences in the gene neighborhood of TXISS. These proteins possess 1 or more domains that passed the false discovery rate (FDR) filter, excluding the domains of other TXISS, TonB dependent plugs, and TbpB_B_D (as those are already molecularly characterized cargo domains). Sequences were functionally annotated using A) KofamKOALA (KO definition shown for sequences >20), B) BlastKOALA, and C) eggNOG-mapper (sequences with COG <5 were trimmed). Sequence categories >50 are listed.

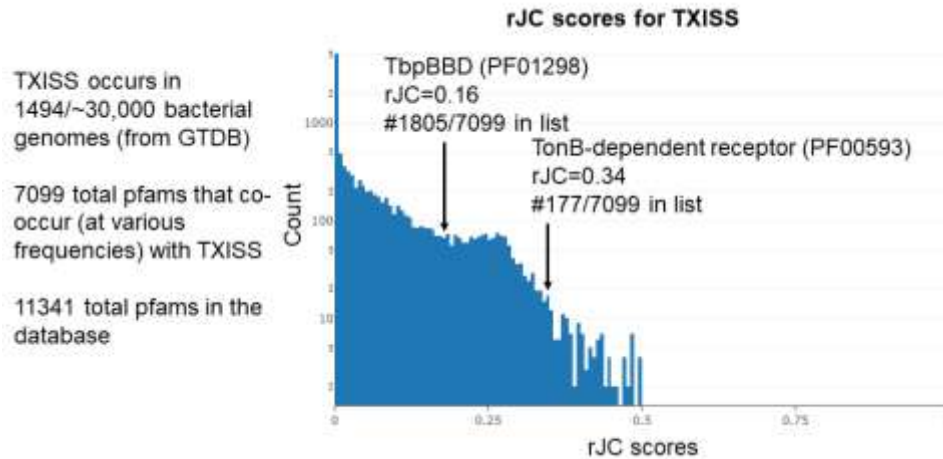


Figure 3.5: rJC score distribution of TXISS. Histogram displays domains that co-occur with TXISS, where a higher rJC corresponds to more frequently co-inherited domains. Two pfams used for further controlling of this list were selected, PF01298 and PF00593. Their frequency of co-inheritance is displayed as a number out of 7099, which is the total number of pfams that co-occur with TXISS. Using TbpBBD's inheritance frequency with TXISS, 2000 domains was chosen as a frequency cutoff to compare to LptD genome co-inheritance.

Table 3.1: Top 50 results from the controlled whole genome co-inheritance assay organized by cellular localization.

Pfam that co-occurs with TXISS	Description of co-occurring pfam	rJC for TXISS
PF04453	LPS transport system D	0.356
PF03668	P-loop ATPase protein family	0.296
PF05036	Sporulation related domain	0.289
PF02556	Preprotein translocase subunit SecB	0.288
PF02955	Prokaryotic glutathione synthetase, ATP-grasp domain	0.286
PF02954	Bacterial regulatory protein, Fis family	0.285
PF13193	AMP-binding enzyme C-terminal domain	0.284
PF04751	Protein of unknown function (DUF615)	0.283
PF00459	Inositol monophosphatase family	0.279
PF12626	Polymerase A arginine-rich C-terminus	0.278
PF00027	Cyclic nucleotide-binding domain	0.277
PF03550	Outer membrane lipoprotein LolB	0.277
PF01145	SPFH domain / Band 7 family	0.274
PF13414	TPR repeat	0.273
PF01266	FAD dependent oxidoreductase	0.273
PF00384	Molybdopterin oxidoreductase	0.271
PF01435	Peptidase family M48	0.270
PF11898	Domain of unknown function (DUF3418)	0.269
PF01925	Sulfite exporter TauE/SafE	0.268

Table 3.1 continued

PF02545	Maf-like protein	0.267
PF06026	Ribose 5-phosphate isomerase A (phosphoriboisomerase A)	0.267
PF00313	'Cold-shock' DNA-binding domain	0.265
PF01488	Shikimate / quinate 5-dehydrogenase	0.264
PF00990	Diguanylate cyclase, GGDEF domain	0.264
PF13147		0.264
PF09335	SNARE associated Golgi protein	0.263
PF00920	Dehydratase family	0.263
PF04354	ZipA, C-terminal FtsZ-binding domain	0.263
PF03799	Cell division protein FtsQ	0.262
PF02896	PEP-utilising enzyme, TIM barrel domain	0.261
PF02152	Dihydroneopterin aldolase	0.259
PF08240	Alcohol dehydrogenase GroES-like domain	0.258
PF00563	EAL domain	0.258
PF00391	PEP-utilizing enzyme, mobile domain	0.256
PF00899	ThiF family	0.256
PF04002	RadC-like JAB domain	0.256
PF02541	Ppx/GppA phosphatase family	0.254
PF04359	Protein of unknown function (DUF493)	0.254
PF06676	Protein of unknown function (DUF1178)	0.254
PF00999	Sodium/hydrogen exchanger family	0.253
PF06850	PHB de-polymerase C-terminus	0.253
PF02597	ThiS family	0.251

Table 3.1 continued

PF01208	Uroporphyrinogen decarboxylase (URO-D)	0.249
PF04375	HemX, putative uroporphyrinogen-III C-methyltransferase	0.249
PF12849	PBP superfamily domain	0.248
PF12804	MobA-like NTP transferase domain	0.247
PF03109	ABC1 family	0.247
PF05673	Protein of unknown function (DUF815)	0.247
PF03872	Anti sigma-E protein RseA, N-terminal domain	0.246
PF13561	Enoyl-(Acyl carrier protein) reductase	0.245

Descending rJC distribution with the top rows representing the highest rate of co-inheritance.

Pfam gene ontology (GO) terms were used to establish cellular localization. Yellow highlight refers to pfams with secreted or plasma membrane annotations, which could mean this is a TXISS cargo domain. Pink highlight refers to pfams with cytoplasmic annotations, which could represent effectors that influence TXISS expression Blue highlight refers to pfams with periplasmic annotations, which could mean this domain participates in TXISS interactions. No highlight refers to pfams with none of these annotations.



Figure 3.6: Secondary structures of 17 predicted TXISS cargo proteins. NiIC and HrpC are shown for comparison. Box colors represent predicted residue architecture: yellow (strand), pink (helix), and gray (coil). Box outlines represent predicted disordered regions (blue) and whether or not they can potentially bind proteins. PDB files generated with PSIPRED 4.0 (160).

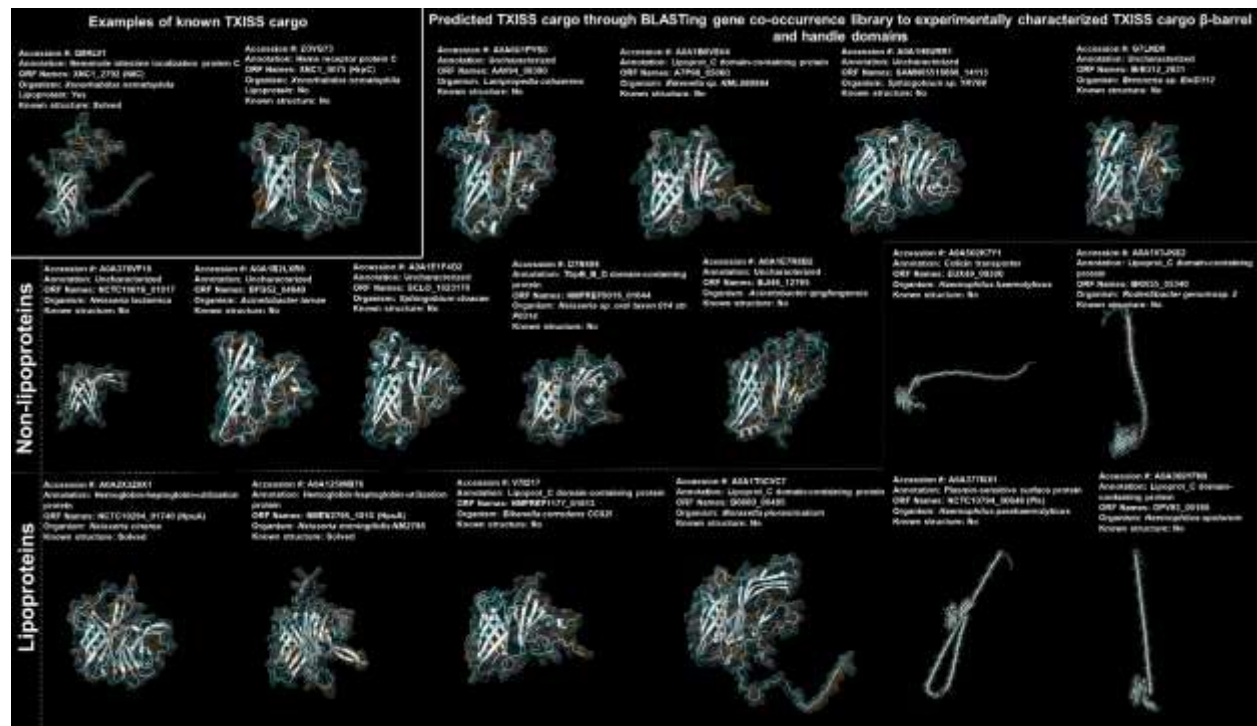


Figure 3.7: Tertiary structures of 17 predicted TXISS cargo proteins. NilC and HrpC are shown for comparison. Organization based on whether or not the proteins were predicted to be a lipoprotein. PDB files generated with AlphaFold. Visualized in ChimeraX (189).

Chapter 3 supplementary figures and tables

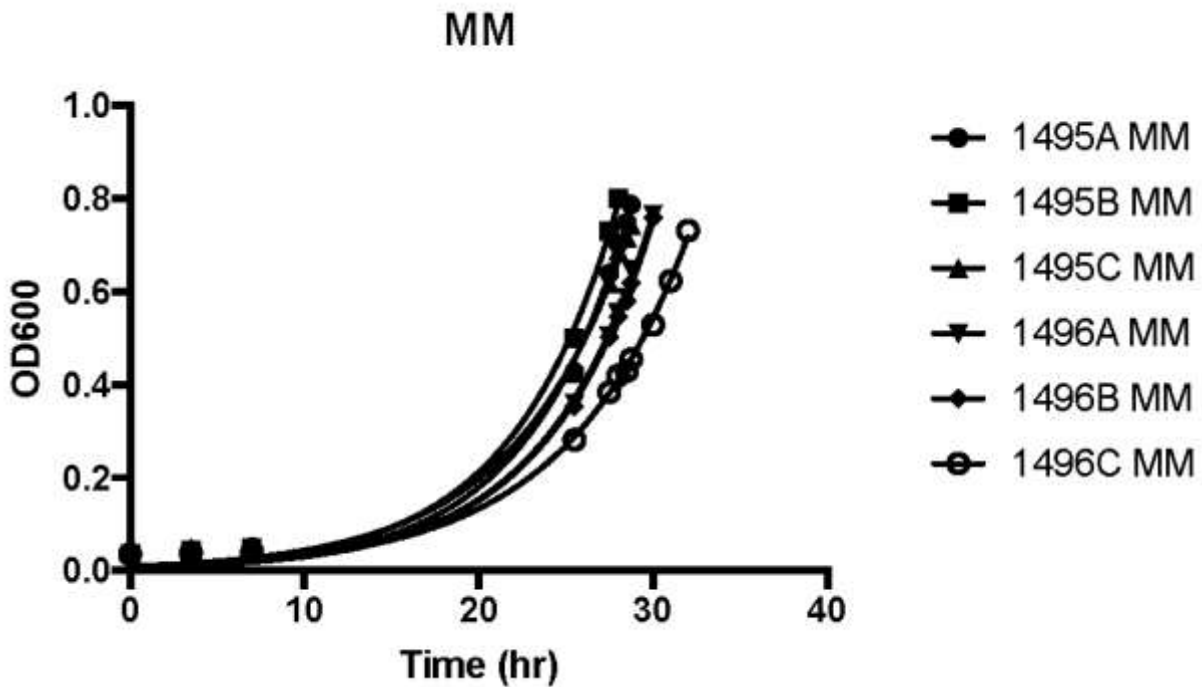


Figure S3.1: Growth curves of strains used for sampling to conduct metabolome and proteome analyses. Triplicate cultures (A, B, and C) of each strain (1495: DeltaSR1 attTn7::eTn7; 1496: DeltaSR1 attTn7::Tn7-SR1) were grown in defined medium with glucose (no casamino acids) in 500 ml flasks with 100 ml media for >24 h until $OD_{600} \cong 0.6$ when they were harvested for proteomics and metabolomics analyses.

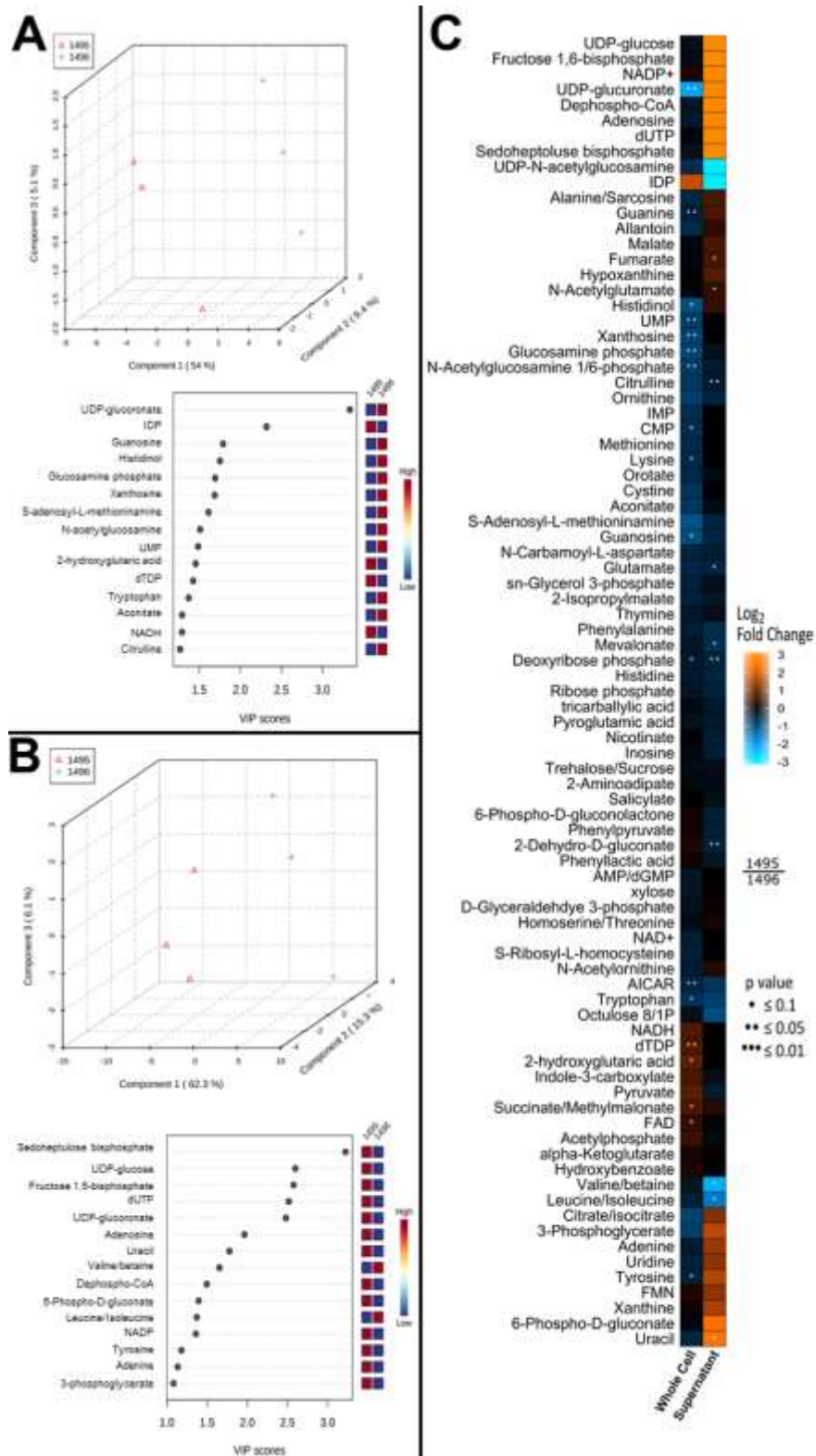


Figure S3.2: Significant metabolome differences between WT and Δ SR1 indicate TXISS role of amino acid and amino sugar metabolism.

A) Whole cell metabolomics partial least squares-discriminant analysis (PLS-DA) between WT and $\Delta SR1$ 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 $\Delta SR1$ and the associated VIP scores for the metabolites that contribute to the separation of metabolic profiles. C) Heatmap of all metabolites detected in the screen with fold change and student t-test p-values listed between strains for both the whole cell and supernatant analyses.

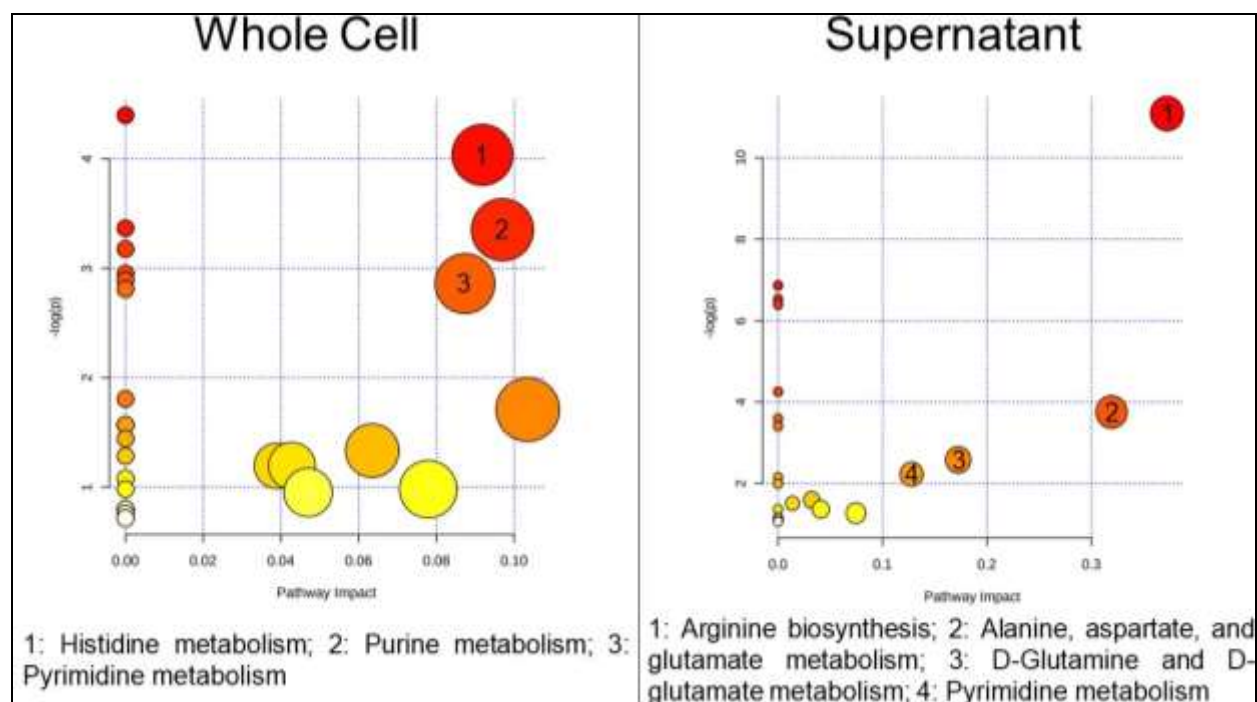


Figure S3.3: Top pathways impacted by TXISS using significant ($P < 0.1$) metabolites. All matched pathways according to the p values from the pathway enrichment analysis and pathway impact values from the pathway topology analysis in MetaboAnalyst for the whole cell (left) and supernatant (right) fractions. Relevant pathways, as determined by a pathway impact threshold of 0.08 and a $-\log(p)$ threshold of 2, are numbered and listed along the bottom of each respective fraction.

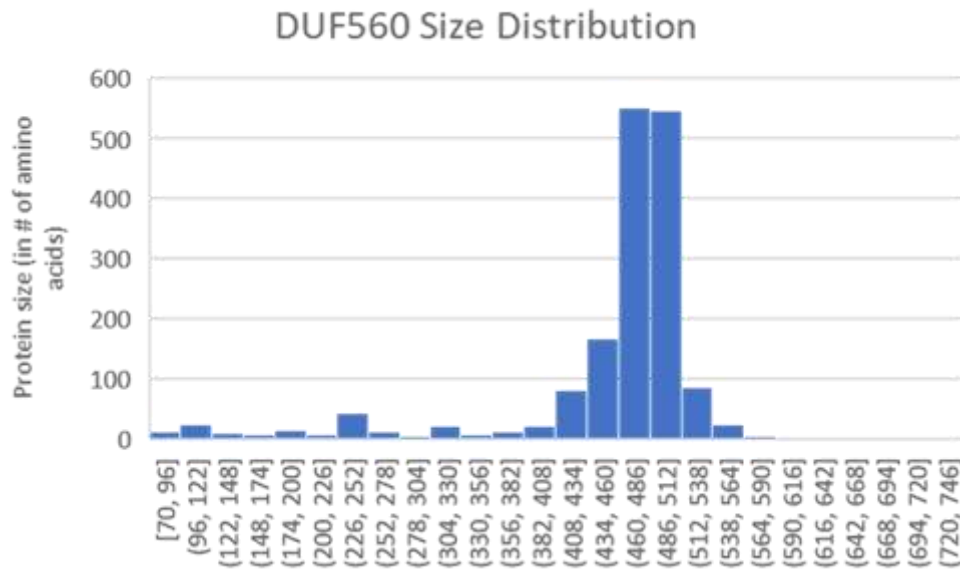


Figure S3.4: TXISS (DUF560) protein size distribution (in amino acids). After examination of the size distribution, median was chosen as an appropriate measure of representative protein size over mean due to left-skewed distribution.

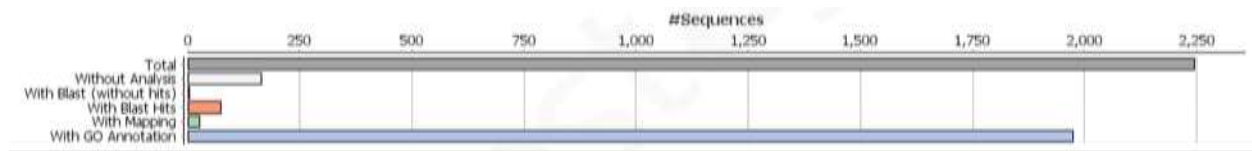


Figure S3.5: Statistics from Blast2GO run. The top row represents the total number of query proteins. The 2nd row represents the number of sequences that were not analyzed. The 3rd row represents sequences that were BLAST searched and returned no hits. The 4th and 5th rows represent sequences that had BLAST hits, but could not be mapped or receive a GO annotation. The 6th row represents the number of sequences that received GO category. Roughly 88% of sequences were classified within a GO category.

Table S3.1: List of strains used in this chapter.

Strain	Description	Reference
HGB1783	<i>Escherichia coli</i> S17-1 λ pir	Bhasin <i>et al.</i> 2012 (136)
HGB1495	<i>X. nematophila</i> Δ nilR16:: <i>Str</i> Δ SR1-7:: <i>kan</i> <i>att</i> Tn7:: <i>empty</i> Tn7 <i>kefA</i> ::pJMC001-GFP (from HGB1783)	Bhasin <i>et al.</i> 2012 (136)
HGB1496	<i>X. nematophila</i> Δ nilR16:: <i>Str</i> Δ SR1-7:: <i>kan</i> <i>att</i> Tn7:: <i>Tn7-SR1</i> <i>kefA</i> ::pJMC001-GFP (from HGB1783)	Bhasin <i>et al.</i> 2012 (136)
HGBXXXX	<i>Escherichia coli</i> BL21 DE3 pETDuet- 1/MCS1::PlsCtermFLAG	This study
HGBXXXX	<i>Escherichia coli</i> BL21 DE3 pETDuet- 1/MCS2::FLAG-TXISSP/Pls-FLAG	This study

Bhasin *et al.* 2012 details how strains were constructed. See methods for *pIs* strain construction.

Table S3.2: Significantly differential proteins detected in the whole cell samples ordered by p-value significance (as determined by Students T-tests).

ProteinID	Student's T-test Fold Difference $\Delta SR1$ (HGB1495) vs. WT (HGB1496)	Student's T-test p-value $\Delta SR1$ (HGB1495) vs. WT (HGB1496)	Annotation via UniProt, STRING, and PaperBLAST
XNC1_0327 ID:1120623 RpsQ	9.58264	7.58E-05	30S ribosomal subunit protein S17
XNC1_2053 ID:1123413 YeiG	3.74479	0.042141	S-formylglutathione hydrolase; formaldehyde detoxification
XNC1_3853 ID:1124183 FtsB	3.40784	0.02679	Essential cell division protein. Septal cell wall synthesis inhibition for cell division coordination.
XNC1_4381 ID:1122482	2.41503	0.047031	Xylanase/chitin deacetylase
XNC1_2249 ID:1123478	1.66335	0.019203	Hypothetical protein; putative exported protein; Trypsin-like serine protease superfamily
XNC1_0005 ID:1124557	1.23376	0.003795	Putative secreted protein

Table S3.2 continued

XNC1_1263 ID:1121121 CsrA	1.22092	0.015185	Carbon storage regulator, post-translational activator of flhDC expression, regulates biofilm formation, RNA-binding A key translational regulator that binds mRNA to regulate translation initiation and/or mRNA stability. Mediates global changes in gene expression, shifting from rapid growth to stress survival by linking envelope stress, the stringent response and the catabolite repression systems
XNC1_2486 ID:1123580	1.18659	0.043023	Putative polysaccharide deacetylase lipoprotein; UDP-glucuronate 4-epimerase
XNC1_0268 ID:1120592 ArgC	1.05026	0.045164	N-acetyl-gamma-glutamylphosphate reductase
XNC1_2792 ID:1121766 NilC	-10.558	3.92E-05	Internal control
XNC1_2789 ID:1123731 NilB	-8.29132	3.97E-05	Internal control
XNC1_3140 ID:1121900	-7.5495	0.001465	Aminoglycoside 3'-N-acetyltransferase
XNC1_3270 ID:1121939 RseC	-6.94633	8.85E-05	Regulator of sigma E (sigma 24) factor. May play a role in reduction of the SoxR iron-sulfur cluster. Part of the rseD-rpoE-rseA-rseB-rseC operon

Table S3.2 continued

XNC1_4499 ID:1122518 Tag	-6.15121	0.000211	tag 3-methyl-adenine DNA glycosylase I, constitutive
XNC1_4367 ID:1122468	-5.59392	0.000165	<i>tusC</i> tRNA processing enzyme?
XNC1_2221 ID:1121532 TacA	-5.37064	0.000635	TacA antitoxin
XNC1_4620 ID:1122566 PaaE	-5.12934	0.001247	Putative phenylacetic acid degradation NADH oxidoreductase paaE 2Fe-2S, Complete proteome, Iron, Iron-sulfur, Metal-binding, Oxidoreductase
XNC1_0165 ID:1122678 MutM	-4.86999	0.000123	Formamidopyrimidine DNA glycosylase, also acts on 5-formyluracil and 5-hydroxymethyluracil
XNC1_1032 ID:1122989	-3.90096	0.00662	Uncharacterized, has GO: lipid metabolism; putative lipase
XNC1_3841	-3.70846	8.63E-05	Restriction modification system DNA specificity domain:Filamentation induced by cAMP protein Fic
XNC1_1003 ID:1120967	-3.2993	0.019022	Hypothetical, catalytic activity/iron-sulfur cluster binding.
XNC1_1929 ID:1121399	-3.24001	0.006381	Hypothetical, magnesium and manganese binding. Hydrolase activity.

Table S3.2 continued

XNC1_2143 ID:1123449 YecP	-2.96338	0.047181	Putative methyltransferase with S-adenosyl-L-methionine-dependent methyltransferase domain
XNC1_4064 ID:1122336 PdxA	-2.42829	0.000224	Pyridoxine (vitamin B6) biosynthesis 4-hydroxythreonine-4-phosphate dehydrogenase metal binding
XNC1_4231 ID:1124363 HipA	-2.14062	0.049747	Toxin in toxin/antitoxin pair
XNC1_2535 ID:1123595	-1.73266	0.013004	Hypothetical, intein-mediated protein splicing
XNC1_4255 ID:1124379 YjgQ	-1.72416	0.009504	LptG; LPS transport, inner membrane protein
XNC1_3551	-1.36046	0.032782	Hypothetical protein, near sugar fermentation stimulation protein B
XNC1_2841 ID:1123750 YcdS	-1.35837	0.029872	HmsH; PgaA; Putative outer membrane protein with transferase domain. PNAG secretion and biosynthesis.
XNC1_2818 ID:1121781 Nuol	-1.15099	0.024956	NADH dehydrogenase I subunit L
XNC1_1277 ID:1123061 YfiA	-1.09532	0.033991	Ribosome associated factor, stabilizes ribosomes against dissociation

Rows are separated between higher in the $\Delta SR1$ mutant background (white highlight) and higher in the WT background (gray highlight).

Table S3.3: Significantly differential proteins detected in the supernatant samples ordered by p-value significance (as determined by Students T-tests).

ProteinID	Student's T-test Fold Difference $\Delta SR1$ (HGB1495) vs. WT (HGB1496)	Student's T-test p-value $\Delta SR1$ (HGB1495) vs. WT (HGB1496)	Annotation via UniProt, STRING, and PaperBLAST
XNC1_0794 ID:1120877 AccB	6.66447	1.54E-05	Acetyl-CoA carboxylase, biotin carboxyl carrier protein subunit
XNC1_2527 ID:1121673	4.50272	3.76E-05	Type VI secretion system effector (Hcp1 family)
XNC1_3692 ID:1124109 Rhs	4.48753	1.94E-05	Rhs family protein
XNC1_2986 ID:1121851	3.93102	0.002872	Myo-inositol utilization transcriptional regulator lolR, or phosphosugar-binding transcriptional regulator, RpiR family
XNC1_0542 ID:1120757	3.66832	0.000342	Putative toxin of the neighboring ParE antitoxin
XNC1_3741 ID:1122218 FecA	3.27566	0.035184	Iron(III) dicitrate transport protein
XNC1_0238 ID:1120576	2.72792	0.028309	Dihydrodipicolinate synthase
XNC1_2278 ID:1121563	2.6597	0.009691	DNA-binding protein Cro/Ci family. Possible toxin from toxin/antitoxin system.

Table S3.3 continued

XNC1_1779 ID:1121334	2.20412	0.020734	Putative surface protein; some similarity to a UDP-N-acetylglucosamine acyltransferase
XNC1_1708 ID:1123250 XcnD	1.90942	0.04427	Xenocoumacin synthesis acyl-CoA dehydrogenase
XNC1_1136 ID:1121043 RecJ	1.74231	0.02568	ssDNA exonuclease
XNC1_1449 ID:1121211 MoaC	1.63047	0.032527	Molybdenum cofactor biosynthesis protein C; Together with MoaA, is involved in the conversion of 5'- GTP to cyclic pyranopterin monophosphate (cPMP or molybdopterin precursor Z)
XNC1_1172 ID:1123031	1.38513	0.048399	AMP-dependent synthetase/ligase
XNC1_3157 ID:1121905 YfcG	1.05986	0.003298	Putative glutathione S-transferase enzyme with thioredoxin-like domain
XNC1_0051 ID:1122631	1.0226	0.016938	Inner membrane protein yjch,
XNC1_4587 ID:1124526	1.01995	0.03691	Hypothetical. Predicted secreted (SPI). DUF930 domain containing protein. Found in Rhizobia.
XNC1_2792 ID:1121766 NilC	-11.8249	1.03E-05	Internal control
XNC1_2789 ID:1123731 NilB	-10.4337	1.35E-06	Internal control

Table S3.3 continued

XNC1_1414 ID:1121188 CydB	-5.28305	0.000135	Cytochrome bd-I ubiquinol oxidase subunit 2
XNC1_1406 ID:1121180 SdhD	-4.35286	0.000339	Succinate dehydrogenase, hydrophobic subunit, cytochrome b556 with SdhC; Membrane-anchoring subunit of succinate dehydrogenase
XNC1_0503 ID:1120736 RimP	-3.96888	0.00024	Ribosome maturation factor RimP
XNC1_3812 ID:1124161 OsmY	-3.72302	0.030845	Osmotically inducible protein Y
XNC1_3090 ID:1123880 GpA	-3.1867	0.039814	Replication gene A protein
XNC1_0493 ID:1120729 ObgE	-2.97453	0.035261	Putative GTP-binding protein with nucleoside triP hydrolase domain. Plays a role in the stringent response, perhaps by sequestering 50S ribosomal subunits and decreasing protein synthesis.
XNC1_3547 ID:1122102	-2.43868	0.03406	Putative exported protein. Homolog of <i>Burkholderia pseudomallei</i> Type IV secretion system conjugative transfer protein X994_308. Linked with a possible CsrA homolog.

Table S3.3 continued

XNC1_2276 ID:1123486 DapA	-2.42024	0.01801	4-hydroxy-tetrahydrodipicolinate synthase
XNC1_3032 ID:1123839 NqrF	-2.29449	0.041516	Na(+)-translocating NADH-quinone reductase subunit F. Accepts electrons from NADH and reduces ubiquinone-1 to ubisemiquinone by a one-electron transfer pathway. Note that ubiquinone mutants have reduced glycogen storage levels.
XNC1_4107 ID:1124319	-1.7161	0.002672	Putative capsid protein of prophage CP-933C (major head protein)
XNC1_2565 ID:1123613 YgiW/VisP BOF family protein	-1.42966	0.006652	Found to increase biofilm surface coverage when deleted in <i>Aggregatibacter</i>
XNC1_1327 ID:1123076 DacA	-1.19536	0.004716	D-alanyl-D-alanine carboxypeptidase. Removes C-terminal D-alanyl residues from sugar-peptide cell wall precursors.

Rows are separated between higher in the $\Delta SR1$ mutant background (white highlight) and higher in the WT background (gray highlight).

CHAPTER 4: Summary, conclusions, and future directions

The work described in this dissertation investigated mechanisms of bacterial virulence and colonization using integrated -omics approaches. To examine these phenomena, I employed the *Xenorhabdus nematophila* bacterium, which is an obligate mutualist with its nematode host *Steinernema carpocapsae*. Together, they infect, kill, and reproduce within insect hosts. This system serves as a simple model to study the poorly understood mechanisms of microbial pathogenicity and establishment within animals. I have made the case that investigating these mechanisms using a simple system has intellectual merits. The integrated -omics studies on this simple system I employed in this dissertation accomplished 2 main goals. In Chapter 2, I demonstrated the processes that organic life is consumed and recycled through carnivory, which highlighted the importance of microbes in food chains. In Chapter 3, I expanded our functional understanding of a secretion system that transports symbiosis factors, leading to potential new drug targets in the fight against human pathogens.

I applied metabolomics to monitor the changing chemical ecology of an insect cadaver as it is parasitized and consumed by an entomopathogenic nematode-bacterium (EPNB) pair. This was a refined time course experiment on a closed chemical ecosystem, allowing for rigorous statistical methods to be employed. I interpreted the metabolomics dataset in the context of the trophic hierarchy, established by collaborators, in which the bacteria consume the insect tissue, while the nematodes consume the bacterial biomass and each other. The tractability of this system is what allowed us to interpret these results, enabling us to separate out the parts of this hierarchy to understand who is producing what chemical signals. I also incorporated bacterial transcriptomics datasets based on well-characterized regulatory mutants to help reveal the biochemical pathways by which *X. nematophila* converts the insect tissue. The strains that were virulence-defective ($\Delta\rho A$, $\Delta\rho$, and the secondary form bacteria) mostly impacted transcription of amino acid metabolism transcripts, while the colonization-defective strains ($\Delta\rho S$ and $\Delta n i l R$) mostly impacted carbohydrate metabolism (Figure 2.2). The specific

pathways in amino acid metabolism that were most impacted by these mutants were tyrosine, phenylalanine, and proline metabolism. For the most impacted carbohydrate pathways, inositol phosphate, pyruvate, and propanoate metabolism were affected. This indicated what types of metabolites could be necessary for *X. nematophila* to kill the insect and then colonize the next generations of nematodes in the cadaver. Taken together with the metabolomics data, these analyses revealed a distinct metabolic profile for each phase of the EPNB lifecycle within insects, indicating that the conversion of an insect to bacterial and nematode biomass occurs through an ordered and reproducible process.

There are specific metabolites that make each phase of the lifecycle unique, whether they be from the insect amounting an immune response against the infection, or from the bacteria converting the insect tissue into nutrients to be consumed by the nematodes. Environmental context affects the chemicals *Xenorhabdus* contributes interacts with. While the insect is still alive, kynurenic acid significantly increases, which could be acting as an insect immunosuppressive compound. After the insect is dead, the siderophore precursor 2,3-dihydroxybenzoate significantly increases through the late phase. This could be indicative of other insect microbiome members that are kept alive (despite the antimicrobials *X. nematophila* secretes) to harvest iron from the cadaver for the EPNB pair. Another stage-specific trend observed was the rise, and then significant drop in *sn*-glycerol 3-phosphate in the late phase, which could be indicative of the types of metabolites the emerging nematodes are storing for their journey for more prey. This is coupled with the significant rise of acetylphosphate, a metabolite used for the protein acetylation post-translational modification, which could be used to prepare the bacteria for colonization of the new generation of nematodes. Combined with previously published genomics data on *S. carpocapsae* and *X. nematophila*, the metabolomics combined with the transcriptomics described in Chapter 2 led to additional findings. For example, XNC1_4600 transcript abundance was 3-times higher in the virulence-deficient Δ *lrhA*

mutant compared to WT. This transcript encodes a protein at the branchpoint of phenylacetic acid formation, which was one of the significantly higher abundant metabolites at the late phase. This branchpoint can turn phenylacetic acid into acetyl coA, and we observed reduced routing through the TCA at the late phase consistent with this finding. Given that we know phenylacetic acid accumulation in *P. aeruginosa* (also a Gram-negative Gammaproteobacteria) inhibits effector transport to host cells (77), and comparing its genome and transcriptome to *X. nematophila*, we hypothesized that expression of *X. nematophila* bioconversion enzymes may be inhibited at the late phase due to phenylacetic acid accumulation.

To strengthen our understanding of how life is consumed and recycled during parasitism, I propose this model be probed further. Achieving greater granularity between time points is necessary to assess the trends described in Chapter 2. For example, hourly time points can be taken during the first 24 hours of the lifecycle to understand how the EPNB pair is defeating the insect. I was able to derive several theories on the mechanisms by which insect immunity is suppressed; the hierarchical clustering analysis we designed found increased abundance in metabolites involved in glutathione metabolism coupled with decreased abundance in metabolites involved in ascorbate metabolism in the first 24 hours up until insect death. This suggests that the EPNB pair is inhibiting the phenoloxidase (PO) cascade while protecting themselves from reactive oxygen species, consistent with previous data on the insect PO response after infection with *X. nematophila* (190, 191). Knowing that *X. nematophila* is coordinating this evasion of host immune response can help us understand broader concepts in invertebrate immune responses. When exactly is the response activated, and how does *X. nematophila* coordinate gene expression at that time point to avoid it? Trends like these deserve to be better identified with more precise measurements, as this would enhance our understanding of how *Xenorhabdus* is parasitizing insects, as well as how many more invertebrate pathogens avoid and suppress the immune response. This higher granularity

screen could be coupled with repeating the analysis at the MS2 (MS/MS) level, as this initial screen was done at the MS1 level.

The reasoning for expansion to tandem mass spectrometry is to enable molecular networking analyses via the Global Natural Products Social Molecular Networking (GNPS) server (16). This tool allows community wide sharing of raw or processed mass spectra data and creates massive public spectral libraries, which could identify many more metabolites from our screen. MS2 would permit access to a global database of metabolites and could widen the scope of the discoveries found in Chapter 2. Ideally, this would spark collaborations between other metabolomics researchers both within and outside of our field, which gains visibility of this study system and allow research to answer biological questions that are broadly applicable and of potential use for the benefit of society and the environment. Other *Xenorhabdus* researchers are already adding to and utilizing the database. Wolff *et al.* (2018) and Tobias *et al.* (2017) used GNPS to discover *Xenorhabdus* secretes tilivallines, which are natural products that bind to DNA and cause cytotoxicity (192, 193). This chemical class of compounds are also secreted by the Gram-negative Proteobacteria *Klebsiella oxytoca*, which is a resident of the human microbiome but can cause antibiotic-associated hemorrhagic colitis after penicillin treatment (194). It is important to note that none of molecular networking studies have considered the entire lifecycle in their experimental approaches, and this is crucial because bacteria change their metabolite secretions depending on environmental context. (192, 193, 195). Applying *Xenorhabdus* metabolome studies to the GNPS has led to natural product discovery which could have wide reaching health implications on chemical compounds secreted by a wide variety of Proteobacteria (See Appendix 1 for data I collected regarding a natural product produced by *X. nematophila* that has nematocidal activity).

Additionally, *S. carpocapsae* IJ samples were taken and compared to the time course. These samples were input IJs, which means they had been surviving in water-filled flasks for

several months after being passaged through and harvested from spent insect cadavers. The PLS-DA plot with these data included show their metabolic profiles were most similar to the late phase, relative to the early and middle phase. This was as expected, since the IJs emerge from the late phase insect cadavers. However, as IJs age, we also expect the metabolic profiles to change. Possibly, if the analysis had focused on the IJs immediately after they had emerged from the spent insect cadaver, the metabolic profiles of late phase insects and IJs would have been overlapping. The metabolic profiles of the input IJs, and output IJ samples over time of storage in water would yield insights into the metabolic changes occurring during aging in the absence of nutrient influx, the processes by which parasites use nutrient reserves during quiescent phases, and the tradeoffs between energy expenditure for hunting and for survival over the potentially long time periods between prey.

In Chapter 3, I explored the mutualism aspect of the lifecycle, and in a broader context, how bacteria colonize animal hosts. Previous work in the Goodrich-Blair Lab established that the genomic Symbiosis Region 1 (SR1, comprised of *nilB* and *nilC*) locus is necessary for *X. nematophila* to colonize *S. carpocapsae* (68, 196). We hypothesized that this could be due to small molecule and/or protein transport mediated by NilB and NilC proteins. We designed an integrated -omics experiment combining proteomics, metabolomics, and genomics to test this hypothesis. When probing the proteome and metabolome, we found that SR1 impacts cell wall, amino sugar, and secreted polysaccharide metabolic pathways. Specifically, we found that poly-N-acetylglucosamine (PNAG) biofilm formation may be perturbed by the absence of SR1. We hypothesize that NilC remains periplasmic when the bacterium is not in the nematode intestinal environment, where it can affect biofilm assembly in other environments. Once in the nematode, NilB surface exposes NilC, and NilC could be importing amino sugars necessary for colonization.

NilB is part of the DUF560 protein family (pfam) that comprise a new type of secretion system, now titled the Type 11 Secretion System (TXISS) (130). Consistent with its membership in this family, NilB can facilitate surface exposure of the lipoprotein NilC (130). TXISS is well distributed among Proteobacteria, with homologs in human pathogen genera *Haemophilus*, *Pasteurella*, and *Neisseria* (131). In the work presented in Chapter 3, I sought to characterize the functional prospects of TXISS homologs and their cargo using publicly available tools to analyze co-occurring gene neighborhood networks and whole genome co-inheritance patterns. I developed a novel refinement to these tools by controlling for those co-occurring pfam lists that are unique to TXISS relative to other similar but functionally unrelated families of proteins. This reduced the number of false positives to refine the results as a method of identifying more specialized interactions.

I controlled the gene neighborhood of TXISS by creating a parallel query of gene neighborhoods associated with a group of proteins that are distributed amongst Proteobacteria, were localized to the outer membrane (using Gene Ontology terminology) and were roughly the same size (in amino acids) as the median TXISS homolog. A false discovery rate (FDR) was calculated where an $FDR < 5\%$ meant this pfam was unique to TXISS, while an $FDR > 5\%$ could mean this pfam is more universally co-occurring and could commonly be found in gene neighborhoods of any protein with this similarity profile (size, distribution, and localization). In total, 91 pfams were trimmed from the original list of 173 and thus were not included in the subsequent functional and structural analyses. Pfams that were trimmed using this control, such as the FecR (PF04773), LysR (PF03466), ABC transporter (PF00005), and amino acid permeases (PF13520 and PF00324), may be widely distributed among Proteobacteria, but still be potential interactors with TXISS and were trimmed out using this control method. Post-control, pfams co-occurring within the TXISS gene neighborhood were heme oxygenases, hemoglobin utilization, and iron uptake domains. This is consistent with our understanding of

TXISS cargo interaction with metal-carrying compounds; TXISS surface presents host metal acquisition proteins and this cargo also function as coreceptors for TonB-dependent metal uptake (130-132, 173). Our knowledge of the TXISS metal homeostasis role and the extent of which it contributes to host-associated phenotypes is growing and should be examined more in future studies.

For the whole-genome co-inheritance assay, I also developed a control by comparing the results of TXISS co-inheritance to a similarly distributed protein family with comparable function, structure, and localization, the lipopolysaccharide transport system D (LptD, PF04453). In addition to metal homeostasis, co-inherited pfams included biofilm and cell membrane assembly domains. Specifically, peptidoglycan interactions were found in this co-inheritance assay; these were PF05036, peptidoglycan binding proteins that binds to glycans, and PF03799, the cell division protein FtsQ. PF03550, which is the lipoprotein localization factor LolB, is structurally similar to the sigma factor regulatory protein RseB, both of which have 11-stranded antiparallel β -sheets (197). RseC appears in the same operon as RseB and was found to be statistically more abundant in the WT fraction compared to $\Delta SR1$ (Table S3.2). RseA (PF03872) is also in the same operon and was in the top co-inherited pfams (Table 3.1). The RseB sigma factor forms a complex with SigmaE and has been shown to cause a colonization defect in *X. nematophila* in a signature-tagged mutagenesis screen (196). There are also links to this system with copper homeostasis (198). Like the proteomic and metabolomic work and previous work done by Hooda *et al.* (2017) (131), this suggests TXISS is important for colonization in more species than *X. nematophila*, and more TXISS homologs to be analyzed for their ability to mediate attachment to host cell surfaces. This work outlined in Chapter 3 both serves as a novel technique for the genomics field and for defining the potential scope of our functional understanding of a newly characterized secretion system.

The structural analysis of the co-occurring proteins yielded a short list of potential TXISS cargos/substrates within the Proteobacteria phylum. Secondary and tertiary structure predictions indicate that nearly all of them had a C-terminal β -barrel (besides A0A378VF18, an uncharacterized protein in *Neisseria lactamica*), indicating that this domain organization could be important for TXISS recognition to surface expose the cargo. Most of the potential cargo clustered to TbpBBD and Lipoprotein C domains, which are 8-stranded β -barrel domains. Hooda *et al.* (2016) first suggested the β -barrel domains could be a possible TXISS targeting domain (132). To determine if our prediction method could accurately identify potential TXISS cargo proteins, we assessed if the plasmin-sensitive surface protein (PIs) could be transported by its putative cognate TXISS OMP. We chose this pair because the function of PIs homologs is known; they function as erythrocyte binding proteins or cell adhesion proteins (199, 200). [Update will be provided at defense.]

Biochemically characterizing the ability for TXISS to surface expose the other 16 potential cargo proteins found in the bioinformatic cargo search would further strengthen the robustness of the method developed and employed. Of interest, a colicin transporter homolog (A0A502K7Y1 in *Haemophilus haemolyticus*) was found as a possible TXISS cargo. Colicins, which are toxins produced to drive competition among bacterial strains in the environment, impact bacterial virulence and colonization of hosts (201). Moreover, homologs of this colicin transporter have been identified as a possible vaccine target against *Actinobacillus pleuropneumonia* which is a deadly pathogen of swine (182). Thus, TXISS could be transporting virulence factors in a number of mammalian pathogens.

In addition to the cargo search we employed in Chapter 3, motif searching tools can be employed to identify more cargo beyond our method of BLAST searching the co-occurrence list to known cargo domains. The Multiple Em for Motif Elicitation (MEME) suite has several tools available that can identify motifs conserved among known or putative TXISS cargo (183). Once

identified, we can use the motif patterns in TXISS domains identified by these tools to search protein databases using the Motif Alignment & Search Tool (MAST) (202). With the methods outlined in this work, we have the ability to pinpoint functionality for a novel secretion system that could be used to combat human pathogenic bacteria. Knowing what molecules these proteins bind to and what regulators control their activity can provide insight into how we can modulate their host-associated phenotypes.

Problems in the life sciences that have long puzzled molecular biologists can be approached using integrated -omics techniques. In symbiosis, which occurs with nearly every organism on Earth, we can now take chemical inventories (the sets of molecules and metabolites being secreted by each partner) of each partner to understand how these partnerships are established. These chemical inventories can reveal the interkingdom molecular communication that is conferring these symbioses. This developing field (which could be considered symbiomics), in which combinatorial -omics technologies are applied in an integrated way to study symbiosis, is growing rapidly alongside the application of this technology. Using the *Xenorhabdus-Steinernema* EPNB pair, I investigated the mechanisms conferring symbiosis from both an antagonism and mutualism angle. This foundational research on a simple host-microbe model system will have larger implications for human health and disease, agriculture, and natural resource management. Using the research provided in this dissertation, we have enhanced the growing scientific understanding of how symbioses and the symbiotic factors that confer them shape biochemical environments.

REFERENCES

1. Lazcano A, Pereto J. 2017. On the origin of mitosing cells: A historical appraisal of Lynn Margulis endosymbiotic theory. *J Theor Biol* 434:80-87.
2. Stolz JF. 2017. Gaia and her microbiome. *FEMS Microbiol Ecol* 93.
3. Margulis L. 1993. Origins of species: acquired genomes and individuality. *Biosystems* 31:121-5.
4. Sender R, Fuchs S, Milo R. 2016. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol* 14:e1002533.
5. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59-65.
6. Human Microbiome Project C. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207-14.
7. McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Loso T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* 110:3229-36.
8. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res* 49:D412-D419.
9. Dardaei L, Wang HQ, Singh M, Fordjour P, Shaw KX, Yoda S, Kerr G, Yu K, Liang J, Cao Y, Chen Y, Lawrence MS, Langenbucher A, Gainor JF, Friboulet L, Dagogo-Jack I, Myers DT, Labrot E, Ruddy D, Parks M, Lee D, DiCecca RH, Moody S, Hao H, Mohseni M, LaMarche M, Williams J, Hoffmaster K, Caponigro G, Shaw AT, Hata AN, Benes CH, Li F, Engelman JA. 2018. SHP2 inhibition restores sensitivity in ALK-rearranged non-small-cell lung cancer resistant to ALK inhibitors. *Nat Med* 24:512-517.
10. Gottlieb G. 1998. Normally occurring environmental and behavioral influences on gene activity: from central dogma to probabilistic epigenesis. *Psychol Rev* 105:792-802.
11. Araujo AM, Carvalho M, Carvalho F, Bastos ML, Guedes de Pinho P. 2017. Metabolomic approaches in the discovery of potential urinary biomarkers of drug-induced liver injury (DILI). *Crit Rev Toxicol* 47:633-649.
12. Canuel V, Rance B, Avillach P, Degoulet P, Burgun A. 2015. Translational research platforms integrating clinical and omics data: a review of publicly available solutions. *Brief Bioinform* 16:280-90.
13. Chaston J, Douglas AE. 2012. Making the most of "omics" for symbiosis research. *Biol Bull* 223:21-9.
14. Jendoubi T. 2021. Approaches to integrating metabolomics and multi-Omics data: a primer. *Metabolites* 11:184.
15. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia JM, Chia JM, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devold S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, et al. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36:566-569.

16. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crusemann M, Boudreau PD, Esquenazi E, Sandoval-Calderon M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrew K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34:828-837.
17. Chong J, Xia J. 2020. Using MetaboAnalyst 4.0 for metabolomics data analysis, interpretation, and integration with other omics data. *Methods Mol Biol* 2104:337-360.
18. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45:D353-D361.
19. Martin J, Tyagi R, Rosa BA, Mitreva M. 2018. A multi-omics database for parasitic nematodes and trematodes. *Methods Mol Biol* 1757:371-397.
20. Quinn RA, Navas-Molina JA, Hyde ER, Song SJ, Vazquez-Baeza Y, Humphrey G, Gaffney J, Minich JJ, Melnik AV, Herschend J, DeReus J, Durant A, Dutton RJ, Khosroheidari M, Green C, da Silva R, Dorrestein PC, Knight R. 2016. From sample to multi-omics conclusions in under 48 Hours. *mSystems* 11.
21. Krassowski M, Das V, Sahu SK, Misra BB. 2020. State of the field in multi-omics research: from computational needs to data mining and sharing. *Front Genet* 11:610798.
22. Coyte KZ, Rakoff-Nahoum S. 2019. Understanding competition and cooperation within the mammalian gut microbiome. *Curr Biol* 29:R538-R544.
23. Nishida AH, Ochman H. 2018. Rates of gut microbiome divergence in mammals. *Mol Ecol* 27:1884-1897.
24. Smits WK, Kuipers OP, Veening JW. 2006. Phenotypic variation in bacteria: the role of feedback regulation. *Nat Rev Microbiol* 4:259-71.
25. Cao M, Goodrich-Blair H. 2020. *Xenorhabdus nematophila* bacteria shift from mutualistic to virulent Lrp-dependent phenotypes within the receptacles of *Steinernema carpocapsae* insect-infective stage nematodes. *Environ Microbiol* 22:5433-5449.
26. Beitelshies M, Hill A, Jones CH, Pfeifer BA. 2018. Phenotypic variation during biofilm formation: implications for anti-biofilm therapeutic design. *Materials (Basel)* 11.
27. Khan MM, Ernst O, Manes NP, Oyler BL, Fraser IDC, Goodlett DR, Nita-Lazar A. 2019. Multi-omics strategies uncover host-pathogen interactions. *ACS Infect Dis* 5:493-505.
28. Wier AM, Nyholm SV, Mandel MJ, Massengo-Tiasse RP, Schaefer AL, Koroleva I, Splinter-Bondurant S, Brown B, Manzella L, Snir E, Almabrazi H, Scheetz TE, Bonaldo Mde F, Casavant TL, Soares MB, Cronan JE, Reed JL, Ruby EG, McFall-Ngai MJ. 2010. Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. *Proc Natl Acad Sci U S A* 107:2259-64.
29. Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, Tobe T, Clarke JM, Topping DL, Suzuki T, Taylor TD, Itoh K, Kikuchi J, Morita H, Hattori M, Ohno H. 2011. Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* 469:543-7.
30. Richards GR, Goodrich-Blair H. 2009. Masters of conquest and pillage: *Xenorhabdus nematophila* global regulators control transitions from virulence to nutrient acquisition. *Cell Microbiol* 11:1025-33.
31. Stock SP. 2019. Partners in crime: symbiont-assisted resource acquisition in *Steinernema entomopathogenic* nematodes. *Curr Opin Insect Sci* 32:22-27.

32. Zhang F, Berg M, Dierking K, Felix MA, Shapira M, Samuel BS, Schulenburg H. 2017. *Caenorhabditis elegans* as a Model for Microbiome Research. *Front Microbiol* 8:485.
33. Chaston JM, Murfin KE, Heath-Heckman EA, Goodrich-Blair H. 2013. Previously unrecognized stages of species-specific colonization in the mutualism between *Xenorhabdus* bacteria and *Steinernema* nematodes. *Cell Microbiol* 15:1545-59.
34. Hillman K, Goodrich-Blair H. 2016. Are you my symbiont? Microbial polymorphic toxins and antimicrobial compounds as honest signals of beneficial symbiotic defensive traits. *Curr Opin Microbiol* 31:184-190.
35. Dreyer J, Malan AP, Dicks LMT. 2018. Bacteria of the genus *Xenorhabdus*, a novel source of bioactive compounds. *Front Microbiol* 9:3177.
36. Martens EC, Russell FM, Goodrich-Blair H. 2005. Analysis of *Xenorhabdus* nematophila metabolic mutants yields insight into stages of *Steinernema carpocapsae* nematode intestinal colonization. *Mol Microbiol* 58:28-45.
37. Cao M, Schwartz HT, Tan C-H, Sternberg PW. 2021. The entomopathogenic nematode *Steinernema hermaphroditum* is a self-fertilizing hermaphrodite and a genetically tractable system for the study of parasitic and mutualistic symbiosis. *Genetics* doi:10.1093/genetics/iyab170.
38. Pletnev P, Osterman I, Sergiev P, Bogdanov A, Dontsova O. 2015. Survival guide: *Escherichia coli* in the stationary phase. *Acta Naturae* 7:22-33.
39. Murfin KE, Chaston J, Goodrich-Blair H. 2012. Visualizing bacteria in nematodes using fluorescent microscopy. *J Vis Exp* doi:10.3791/4298.
40. van den Hoogen J, Geisen S, Routh D, Ferris H, Traunspurger W, Wardle DA, de Goede RGM, Adams BJ, Ahmad W, Andriuzzi WS, Bardgett RD, Bonkowski M, Campos-Herrera R, Cares JE, Caruso T, de Brito Caixeta L, Chen X, Costa SR, Creamer R, Mauro da Cunha Castro J, Dam M, Djigal D, Escuer M, Griffiths BS, Gutierrez C, Hohberg K, Kalinkina D, Kardol P, Kergunteuil A, Korthals G, Krashevskaya V, Kudrin AA, Li Q, Liang W, Magilton M, Marais M, Martin JAR, Matveeva E, Mayad EH, Mulder C, Mullin P, Neilson R, Nguyen TAD, Nielsen UN, Okada H, Rius JEP, Pan K, Peneva V, Pellissier L, Carlos Pereira da Silva J, et al. 2019. Soil nematode abundance and functional group composition at a global scale. *Nature* 572:194-198.
41. Enrique Cabanillas H. 2003. Susceptibility of the boll weevil to *Steinernema riobrave* and other entomopathogenic nematodes. *J Invertebr Pathol* 82:188-97.
42. Kapranas A, Maher AMD, Griffin CT. 2017. The influence of organic matter content and media compaction on the dispersal of entomopathogenic nematodes with different foraging strategies. *Parasitology* 144:1956-1963.
43. da Silva WJ, Pilz-Junior HL, Heermann R, da Silva OS. 2020. The great potential of entomopathogenic bacteria *Xenorhabdus* and *Photorhabdus* for mosquito control: a review. *Parasit Vectors* 13:376.
44. Javal M, Terblanche JS, Conlong DE, Malan AP. 2019. First screening of entomopathogenic nematodes and fungus as biocontrol agents against an emerging pest of sugarcane, *Cacosceles newmannii* (Coleoptera: Cerambycidae). *Insects* 10.
45. Margulis L. 1981. *Symbiosis in cell evolution : life and its environment on the early Earth*. W. H. Freeman, San Francisco.
46. Camilli A, Bassler BL. 2006. Bacterial small-molecule signaling pathways. *Science* 311:1113-6.
47. Worthington RJ, Richards JJ, Melander C. 2012. Small molecule control of bacterial biofilms. *Org Biomol Chem* 10:7457-74.
48. Maddocks SE, Oyston PCF. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology (Reading)* 154:3609-3623.
49. Budnick JA, Sheehan LM, Ginder MJ, Failor KC, Perkowski JM, Pinto JF, Kohl KA, Kang L, Michalak P, Luo L, Heindl JE, Caswell CC. 2020. A central role for the transcriptional

- regulator VtIR in small RNA-mediated gene regulation in *Agrobacterium tumefaciens*. *Sci Rep* 10:14968.
50. Abdelhamed H, Ramachandran R, Narayanan L, Ozdemir O, Cooper A, Olivier AK, Karsi A, Lawrence ML. 2020. Contributions of a LysR transcriptional regulator to *Listeria monocytogenes* virulence and identification of its regulons. *J Bacteriol* 202.
 51. Tang J. 2011. Microbial metabolomics. *Curr Genomics* 12:391-403.
 52. Phelan VV, Moree WJ, Aguilar J, Cornett DS, Koumoutsis A, Noble SM, Pogliano K, Guerrero CA, Dorrestein PC. 2014. Impact of a transposon insertion in *phzF2* on the specialized metabolite production and interkingdom interactions of *Pseudomonas aeruginosa*. *J Bacteriol* 196:1683-93.
 53. Dong Y, Brewer GJ. 2019. Global metabolic shifts in age and Alzheimer's Disease mouse brains pivot at NAD⁺/NADH redox sites. *J Alzheimers Dis* 71:119-140.
 54. Chikaraishi Y, Steffan SA, Ogawa NO, Ishikawa NF, Sasaki Y, Tsuchiya M, Ohkouchi N. 2014. High-resolution food webs based on nitrogen isotopic composition of amino acids. *Ecol Evol* 4:2423-49.
 55. Killiny N. 2018. Generous hosts: why the larvae of greater wax moth, *Galleria mellonella* is a perfect infectious host model? *Virulence* 9:860-865.
 56. Herbert EE, Goodrich-Blair H. 2007. Friend and foe: the two faces of *Xenorhabdus nematophila*. *Nat Rev Microbiol* 5:634-46.
 57. Mitani DK, Kaya HK, Goodrich-Blair H. 2004. Comparative study of the entomopathogenic nematode, *Steinernema carpocapsae*, reared on mutant and wild-type *Xenorhabdus nematophila*. *Biological Control* 29:382-391.
 58. Richards GR, Herbert EE, Park Y, Goodrich-Blair H. 2008. *Xenorhabdus nematophila* *IrhA* is necessary for motility, lipase activity, toxin expression, and virulence in *Manduca sexta* insects. *J Bacteriol* 190:4870-9.
 59. Richards GR, Goodrich-Blair H. 2010. Examination of *Xenorhabdus nematophila* lipases in pathogenic and mutualistic host interactions reveals a role for *xlpA* in nematode progeny production. *Appl Environ Microbiol* 76:221-9.
 60. Vivas EI, Goodrich-Blair H. 2001. *Xenorhabdus nematophilus* as a model for host-bacterium interactions: *rpoS* is necessary for mutualism with nematodes. *J Bacteriol* 183:4687-93.
 61. Herbert Tran EE, Andersen AW, Fau - Goodrich-Blair H, Goodrich-Blair H. *CpxRA* influences *Xenorhabdus nematophila* colonization initiation and outgrowth in *Steinernema carpocapsae* nematodes through regulation of the *nil* locus.
 62. Cowles KN, Cowles CE, Richards GR, Martens EC, Goodrich-Blair H. 2007. The global regulator *Lrp* contributes to mutualism, pathogenesis and phenotypic variation in the bacterium *Xenorhabdus nematophila*. *Cell Microbiol* 9:1311-23.
 63. Cowles CE, Goodrich-Blair H. 2006. *nilR* is necessary for co-ordinate repression of *Xenorhabdus nematophila* mutualism genes. *Molecular microbiology* 62:760-771.
 64. Park D, Forst S. 2006. Co-regulation of motility, exoenzyme and antibiotic production by the *EnvZ-OmpR-FlhDC-FlhA* pathway in *Xenorhabdus nematophila*. *Mol Microbiol* 61:1397-412.
 65. Givaudan A, Lanois A. 2000. *flhDC*, the flagellar master operon of *Xenorhabdus nematophilus*: requirement for motility, lipolysis, extracellular hemolysis, and full virulence in insects. *J Bacteriol* 182:107-15.
 66. Cambon MC, Parthuisot N, Pagès S, Lanois A, Givaudan A, Ferdy J-B. 2019. Selection of bacterial mutants in late infections: when vector transmission trades off against growth advantage in stationary phase. *mBio* 10:e01437-19.
 67. Steffan SA, Chikaraishi Y, Dharampal PS, Pauli JN, Guedot C, Ohkouchi N. 2017. Unpacking brown food-webs: animal trophic identity reflects rampant microbivory. *Ecol Evol* 7:3532-3541.

68. Cowles CE, Goodrich-Blair H. 2008. The *Xenorhabdus nematophila* *niABC* genes confer the ability of *Xenorhabdus* spp. to colonize *Steinernema carpocapsae* nematodes. *J Bacteriol* 190:4121-8.
69. Hussa EA, Casanova-Torres AM, Goodrich-Blair H. 2015. The global transcription factor Lrp controls virulence modulation in *Xenorhabdus nematophila*. *J Bacteriol* 197:3015-25.
70. Lu X. 2012. Examining pathogenic and mutualistic regulatory networks in the bacterium *Xenorhabdus nematophila*. UW-Madison PhD thesis.
71. Lu X, Goodrich-Blair H, Tjaden B. 2011. Assessing computational tools for the discovery of small RNA genes in bacteria. *RNA* 17:1635-47.
72. Veessenmeyer JL, Andersen AW, Lu X, Hussa EA, Murfin KE, Chaston JM, Dillman AR, Wassarman KM, Sternberg PW, Goodrich-Blair H. 2014. *niID* CRISPR RNA contributes to *Xenorhabdus nematophila* colonization of symbiotic host nematodes. *Mol Microbiol* 93:1026-42.
73. May JM, Qu ZC, Meredith ME. 2012. Mechanisms of ascorbic acid stimulation of norepinephrine synthesis in neuronal cells. *Biochem Biophys Res Commun* 426:148-52.
74. Lee Y, Jeong H, Park KH, Kim KW. 2020. Effects of NAD(+) in *Caenorhabditis elegans* models of neuronal damage. *Biomolecules* 10.
75. McReynolds MR, Wang W, Holleran LM, Hanna-Rose W. 2017. Uridine monophosphate synthetase enables eukaryotic de novo NAD(+) biosynthesis from quinolinic acid. *J Biol Chem* 292:11147-11153.
76. Wirthgen E, Hoeflich A, Rebl A, Gunther J. 2017. Kynurenic acid: the janus-faced role of an immunomodulatory tryptophan metabolite and its link to pathological conditions. *Front Immunol* 8:1957.
77. Wang J, Dong Y-H, Zhou T, Liu X, Deng Y, Wang C, Lee J, Zhang L-H. 2013. *Pseudomonas aeruginosa* cytotoxicity is attenuated at high cell density and associated with the accumulation of phenylacetic acid. *PloS one* 8:e60187.
78. Ciche TA, Blackburn M, Carney JR, Ensign JC. 2003. Photobactin: a catechol siderophore produced by *Photobacterium luminescens*, an entomopathogen mutually associated with *Heterorhabditis bacteriophora* NC1 nematodes. *Appl Environ Microbiol* 69:4706-13.
79. Cambon MC, Lafont P, Frayssinet M, Lanois A, Ogier JC, Pages S, Parthuisot N, Ferdy JB, Gaudriault S. 2020. Bacterial community profile after the lethal infection of *Steinernema-Xenorhabdus* pairs into soil-reared *Tenebrio molitor* larvae. *FEMS Microbiol Ecol* 96.
80. Clark KD, Lu Z, Strand MR. 2010. Regulation of melanization by glutathione in the moth *Pseudoplusia includens*. *Insect Biochem Mol Biol* 40:460-7.
81. Crawford JM, Portmann C, Zhang X, Roeffaers MB, Clardy J. 2012. Small molecule perimeter defense in entomopathogenic bacteria. *Proc Natl Acad Sci U S A* 109:10821-6.
82. Bosi A, Banfi D, Bistoletti M, Giaroni C, Baj A. 2020. Tryptophan metabolites along the microbiota-gut-brain axis: an interkingdom communication system influencing the gut in health and disease. *Int J Tryptophan Res* 13:1178646920928984.
83. Smirnoff N. 2018. Ascorbic acid metabolism and functions: A comparison of plants and mammals. *Free radical biology & medicine* 122:116-129.
84. Patananan AN, Budenholzer LM, Pedraza ME, Torres ER, Adler LN, Clarke SG. 2015. The invertebrate *Caenorhabditis elegans* biosynthesizes ascorbate. *Archives of biochemistry and biophysics* 569:32-44.
85. Pesko BK, Weidt S, McLaughlin M, Wescott DJ, Torrance H, Burgess K, Burchmore R. 2020. Postmortomics: the potential of untargeted metabolomics to highlight markers for time since death. *OMICS* 24:649-659.

86. Maloy SR, Nunn WD. 1982. Genetic regulation of the glyoxylate shunt in *Escherichia coli* K-12. *J Bacteriol* 149:173-80.
87. Park SJ, Tseng CP, Gunsalus RP. 1995. Regulation of succinate dehydrogenase (sdhCDAB) operon expression in *Escherichia coli* in response to carbon supply and anaerobiosis: role of ArcA and Fnr. *Mol Microbiol* 15:473-82.
88. Baliadi Y, Yoshiga T, Kondo E. 2004. Infectivity and post-infection development of infective juveniles originating via endotokia matricida in entomopathogenic nematodes. *Applied Entomology and Zoology* 39:61-69.
89. Chikaraishi Y, Ogawa NO, Doi H, Ohkouchi N. 2011. 15N/14N ratios of amino acids as a tool for studying terrestrial food webs: a case study of terrestrial insects (bees, wasps, and hornets). *Ecological Research* 26:835-844.
90. Steffan SA, Dharampal PS. 2019. Undead food-webs: integrating microbes into the food-chain. *Food Webs* 18:e00111.
91. Lindeman RL. 1991. The trophic-dynamic aspect of ecology. *Bulletin of Mathematical Biology* 53:167-191.
92. Colinvaux PA, Barnett BD. 1979. Lindeman and the ecological efficiency of wolves. *The American Naturalist* 114:707-718.
93. Martino GP, Perez CE, Magni C, Blancato VS. 2018. Implications of the expression of *Enterococcus faecalis* citrate fermentation genes during infection. *PLoS One* 13:e0205787.
94. Perinbam K, Chacko JV, Kannan A, Digman MA, Siryaporn A. 2020. A shift in central metabolism accompanies virulence activation in *Pseudomonas aeruginosa*. *mBio* 11.
95. Chung JC, Rzhepishevskaya O, Ramstedt M, Welch M. 2013. Type III secretion system expression in oxygen-limited *Pseudomonas aeruginosa* cultures is stimulated by isocitrate lyase activity. *Open Biol* 3:120131.
96. Hada K, Hirota K, Inanobe A, Kako K, Miyata M, Araoi S, Matsumoto M, Ohta R, Arisawa M, Daitoku H, Hanada T, Fukamizu A. 2019. Tricarboxylic acid cycle activity suppresses acetylation of mitochondrial proteins during early embryonic development in *Caenorhabditis elegans*. *J Biol Chem* 294:3091-3099.
97. Rahman MM, Rosu S, Joseph-Strauss D, Cohen-Fix O. 2014. Down-regulation of tricarboxylic acid (TCA) cycle genes blocks progression through the first mitotic division in *Caenorhabditis elegans* embryos. *Proc Natl Acad Sci U S A* 111:2602-7.
98. Wright DJ, Grewal PS, Stolinski M. 1997. Relative importance of neutral lipids and glycogen as energy stores in dauer larvae of two entomopathogenic nematodes, *Steinernema carpocapsae* and *Steinernema feltiae*. *Comp Biochem Physiol B Biochem Mol Biol* 118:269-73.
99. Mak HY. 2012. Lipid droplets as fat storage organelles in *Caenorhabditis elegans*: Thematic Review Series: lipid droplet synthesis and metabolism: from yeast to man. *Journal of lipid research* 53:28-33.
100. Edwards CB, Copes N, Brito AG, Canfield J, Bradshaw PC. 2013. Malate and fumarate extend lifespan in *Caenorhabditis elegans*. *PLoS One* 8:e58345.
101. Jing X, Behmer ST. 2020. Insect sterol nutrition: physiological mechanisms, ecology, and applications. *Annu Rev Entomol* 65:251-271.
102. Raymond KN, Dertz EA, Kim SS. 2003. Enterobactin: an archetype for microbial iron transport. *Proc Natl Acad Sci U S A* 100:3584-8.
103. Teulier L, Weber JM, Crevier J, Darveau CA. 2016. Proline as a fuel for insect flight: enhancing carbohydrate oxidation in hymenopterans. *Proc Biol Sci* 283.
104. Crawford JM, Kontnik R, Clardy J. 2010. Regulating alternative lifestyles in entomopathogenic bacteria. *Curr Biol* 20:69-74.
105. Tang H, Pang S. 2016. Proline catabolism modulates innate immunity in *Caenorhabditis elegans*. *Cell Rep* 17:2837-2844.

106. McMullen JG, 2nd, Stock SP. 2014. In vivo and in vitro rearing of entomopathogenic nematodes (Steinernematidae and Heterorhabditidae). *J Vis Exp* doi:10.3791/52096:52096.
107. Buecher EJ, Popiel I. 1989. Liquid culture of the entomogenous nematode *Steinernema feltiae* with its bacterial symbiont. *J Nematol* 21:500-4.
108. Steffan SA, Chikaraishi Y, Currie CR, Horn H, Gaines-Day HR, Pauli JN, Zalapa JE, Ohkouchi N. 2015. Microbes are trophic analogs of animals. *Proc Natl Acad Sci U S A* 112:15119-24.
109. Lu W, Clasquin MF, Melamud E, Amador-Noguez D, Caudy AA, Rabinowitz JD. 2010. Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. *Anal Chem* 82:3212-21.
110. Clasquin MF, Melamud E, Rabinowitz JD. 2012. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr Protoc Bioinformatics Chapter* 14:Unit14 11.
111. Melamud E, Vastag L, Rabinowitz JD. 2010. Metabolomic analysis and visualization engine for LC-MS data. *Anal Chem* 82:9818-26.
112. Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, Lajus A, Pascal G, Scarpelli C, Medigue C. 2006. MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res* 34:53-65.
113. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47:D607-D613.
114. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726-731.
115. Yoshito C, Yuichiro K, Nanako OO, Hiroshi K, Naohiko O. 2007. Metabolic control of nitrogen isotope composition of amino acids in macroalgae and gastropods: implications for aquatic food web studies. *Marine Ecology Progress Series* 342:85-90.
116. Chikaraishi Y, Ogawa N, Kashiyaama Y, Takano Y, Suga H, Tomitani A, Miyashita H, Kitazato H, Ohkouchi N. 2009. Determination of aquatic food-web structure based on compound-specific nitrogen isotopic composition of amino acids: trophic level estimation by amino acid $\delta^{15}\text{N}$. *Limnology and oceanography* 7:740-750.
117. Steffan SA, Chikaraishi Y, Horton DR, Ohkouchi N, Singleton ME, Miliczky E, Hogg DB, Jones VP. 2013. Trophic hierarchies illuminated via amino acid isotopic analysis. *PLoS One* 8:e76152.
118. Wolfe AJ. 2015. Glycolysis for microbiome generation. *Microbiol Spectr* 3.
119. Ren J, Sang Y, Lu J, Yao YF. 2017. Protein acetylation and its role in bacterial virulence. *Trends Microbiol* 25:768-779.
120. Alaimo JT, Davis SJ, Song SS, Burnette CR, Grotewiel M, Shelton KL, Pierce-Shimomura JT, Davies AG, Bettinger JC. 2012. Ethanol metabolism and osmolarity modify behavioral responses to ethanol in *C. elegans*. *Alcohol Clin Exp Res* 36:1840-50.
121. Doberenz C, Zorn M, Falke D, Nannemann D, Hunger D, Beyer L, Ihling CH, Meiler J, Sinz A, Sawers RG. 2014. Pyruvate formate-lyase interacts directly with the formate channel FocA to regulate formate translocation. *J Mol Biol* 426:2827-39.
122. Zelcbuch L, Lindner SN, Zegman Y, Vainberg Slutskin I, Antonovsky N, Gleizer S, Milo R, Bar-Even A. 2016. Pyruvate formate-lyase enables efficient growth of *Escherichia coli* on acetate and formate. *Biochemistry* 55:2423-6.
123. De Mets F, Van Melderden L, Gottesman S. 2019. Regulation of acetate metabolism and coordination with the TCA cycle via a processed small RNA. *Proc Natl Acad Sci U S A* 116:1043-1052.

124. Watford M. 2008. Glutamine metabolism and function in relation to proline synthesis and the safety of glutamine and proline supplementation. *J Nutr* 138:2003S-2007S.
125. Zheng R, Feng X, Wei X, Pan X, Liu C, Song R, Jin Y, Bai F, Jin S, Wu W, Cheng Z. 2018. PutA Is required for virulence and regulated by PruR in *Pseudomonas aeruginosa*. *Front Microbiol* 9:548.
126. Di Martino ML, Campilongo R, Casalino M, Micheli G, Colonna B, Prosseda G. 2013. Polyamines: emerging players in bacteria-host interactions. *Int J Med Microbiol* 303:484-91.
127. Shah P, Swiatlo E. 2008. A multifaceted role for polyamines in bacterial pathogens. *Mol Microbiol* 68:4-16.
128. Zecic A, Dhondt I, Braeckman BP. 2019. The nutritional requirements of *Caenorhabditis elegans*. *Genes Nutr* 14:15.
129. Mudgal R, Sandhya S, Chandra N, Srinivasan N. 2015. De-DUFing the DUFs: Deciphering distant evolutionary relationships of Domains of Unknown Function using sensitive homology detection methods. *Biol Direct* 10:38.
130. Grossman AS, Mauer TJ, Forest KT, Goodrich-Blair H. 2021. A widespread bacterial secretion system with diverse substrates. *mBio* 12:e0195621.
131. Hooda Y, Lai CCL, Moraes TF. 2017. Identification of a large family of Slam-dependent surface lipoproteins in Gram-negative bacteria. *Front Cell Infect Microbiol* 7:207.
132. Hooda Y, Lai CC, Judd A, Buckwalter CM, Shin HE, Gray-Owen SD, Moraes TF. 2016. Slam is an outer membrane protein that is required for the surface display of lipidated virulence factors in *Neisseria*. *Nat Microbiol* 1:16009.
133. Zhang Z, Wang J, Wang J, Wang J, Li Y. 2020. Estimate of the sequenced proportion of the global prokaryotic genome. *Microbiome* 8:134.
134. Rogozin IB, Makarova KS, Murvai J, Czabarka E, Wolf YI, Tatusov RL, Szekely LA, Koonin EV. 2002. Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res* 30:2212-23.
135. Aravind L. 2000. Guilt by association: contextual information in genome analysis. *Genome Res* 10:1074-7.
136. Bhasin A, Chaston JM, Goodrich-Blair H. 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* 194:1763-1776.
137. Borg AJE, Dennig A, Weber H, Nidetzky B. 2021. Mechanistic characterization of UDP-glucuronic acid 4-epimerase. *FEBS J* 288:1163-1178.
138. Dobrogosz WJ. 1968. Effect of amino sugars on catabolite repression in *Escherichia coli*. *J Bacteriol* 95:578-84.
139. Boes A, Olatunji S, Breukink E, Terrak M. 2019. Regulation of the peptidoglycan polymerase activity of PBP1b by antagonist actions of the core divisome proteins FtsBLQ and FtsN. *mBio* 10.
140. Drace K, Darby C. 2008. The hmsHFRS operon of *Xenorhabdus nematophila* is required for biofilm attachment to *Caenorhabditis elegans*. *Appl Environ Microbiol* 74:4509-15.
141. Low KE, Howell PL. 2018. Gram-negative synthase-dependent exopolysaccharide biosynthetic machines. *Curr Opin Struct Biol* 53:32-44.
142. Berndt V, Beckstette M, Volk M, Dersch P, Bronstrup M. 2019. Metabolome and transcriptome-wide effects of the carbon storage regulator A in enteropathogenic *Escherichia coli*. *Sci Rep* 9:138.
143. Parker A, Cureoglu S, De Lay N, Majdalani N, Gottesman S. 2017. Alternative pathways for *Escherichia coli* biofilm formation revealed by sRNA overproduction. *Mol Microbiol* 105:309-325.

144. Calder JT, Christman ND, Hawkins JM, Erickson DL. 2020. A trimeric autotransporter enhances biofilm cohesiveness in *Yersinia pseudotuberculosis* but not in *Yersinia pestis*. *J Bacteriol* 202.
145. Couillault C, Ewbank JJ. 2002. Diverse bacteria are pathogens of *Caenorhabditis elegans*. *Infect Immun* 70:4705-7.
146. Kim JK, Kwon JY, Kim SK, Han SH, Won YJ, Lee JH, Kim CH, Fukatsu T, Lee BL. 2014. Purine biosynthesis, biofilm formation, and persistence of an insect-microbe gut symbiosis. *Appl Environ Microbiol* 80:4374-82.
147. Krokan HE, Bjoras M. 2013. Base excision repair. *Cold Spring Harb Perspect Biol* 5:a012583.
148. Tietz JI, Schwalen CJ, Patel PS, Maxson T, Blair PM, Tai HC, Zakai UI, Mitchell DA. 2017. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat Chem Biol* 13:470-478.
149. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44:D286-93.
150. Conesa A, Gotz S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics* 2008:619832.
151. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36:2251-2252.
152. Dobrindt U, Reidl J. 2000. Pathogenicity islands and phage conversion: evolutionary aspects of bacterial pathogenesis. *Int J Med Microbiol* 290:519-27.
153. Kumazaki K, Kishimoto T, Furukawa A, Mori H, Tanaka Y, Dohmae N, Ishitani R, Tsukazaki T, Nureki O. 2014. Crystal structure of *Escherichia coli* YidC, a membrane protein chaperone and insertase. *Sci Rep* 4:7299.
154. Angerer A, Enz S, Ochs M, Braun V. 1995. Transcriptional regulation of ferric citrate transport in *Escherichia coli* K-12. Fecl belongs to a new subfamily of sigma 70-type factors that respond to extracytoplasmic stimuli. *Mol Microbiol* 18:163-74.
155. Tremblay BJ, Lobb B, Doxey AC. 2021. PhyloCorrelate: inferring bacterial gene-gene functional associations through large-scale phylogenetic profiling. *Bioinformatics* doi:10.1093/bioinformatics/btaa1105.
156. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. 2021. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* doi:10.1093/nar/gkab776.
157. Cokus S, Mizutani S, Pellegrini M. 2007. An improved method for identifying functionally linked proteins using phylogenetic profiles. *BMC Bioinformatics* 8 Suppl 4:S7.
158. Botos I, Majdalani N, Mayclin SJ, McCarthy JG, Lundquist K, Wojtowicz D, Barnard TJ, Gumbart JC, Buchanan SK. 2016. Structural and functional characterization of the LPS transporter LptDE from Gram-negative pathogens. *Structure (London, England : 1993)* 24:965-976.
159. Alborzi SZ, Ritchie DW, Devignes M-D. 2018. Computational discovery of direct associations between GO terms and protein domains. *BMC Bioinformatics* 19:413.
160. Buchan DWA, Jones DT. 2019. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res* 47:W402-W407.
161. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstern S, Silver

- D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583-589.
162. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* 12:1652-62.
 163. Costa TR, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, Waksman G. 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol* 13:343-59.
 164. Carleton HA, Lara-Tejero M, Liu X, Galan JE. 2013. Engineering the type III secretion system in non-replicating bacterial minicells for antigen delivery. *Nat Commun* 4:1590.
 165. Willias SP, Chauhan S, Lo CC, Chain PS, Motin VL. 2015. CRP-Mediated Carbon Catabolite Regulation of *Yersinia pestis* Biofilm Formation Is Enhanced by the Carbon Storage Regulator Protein, CsrA. *PLoS One* 10:e0135481.
 166. Pourciau C, Pannuri A, Potts A, Yakhnin H, Babitzke P, Romeo T. 2019. Regulation of iron storage by CsrA supports exponential growth of *Escherichia coli*. *mBio* 10.
 167. Figueroa-Bossi N, Schwartz A, Guillemardet B, D'Heygere F, Bossi L, Boudvillain M. 2014. RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. *Genes Dev* 28:1239-51.
 168. Wang X, Dubey AK, Suzuki K, Baker CS, Babitzke P, Romeo T. 2005. CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*. *Mol Microbiol* 56:1648-63.
 169. Shanmugam M, El Abbar F, Ramasubbu N. 2015. Transcriptome profiling of wild-type and *pga*-knockout mutant strains reveal the role of exopolysaccharide in *Aggregatibacter actinomycetemcomitans*. *PLoS One* 10:e0134285.
 170. Mucci NC, Jones KA, Cao M, Wyatt MR, Foye S, Kauffman S, Taufer M, Chikaraishi Y, Steffan S, Campagna S, Goodrich-Blair H. 2021. Chemical ecology of a tripartite symbiosis. *bioRxiv* doi:10.1101/2021.02.13.430781:2021.02.13.430781.
 171. Polikanov YS, Blaha GM, Steitz TA. 2012. How hibernation factors RMF, HPF, and YfiA turn off protein synthesis. *Science* 336:915-8.
 172. Prossliner T, Gerdes K, Sorensen MA, Winther KS. 2021. Hibernation factors directly block ribonucleases from entering the ribosome in response to starvation. *Nucleic Acids Res* 49:2226-2239.
 173. Noinaj N, Guillier M, Barnard TJ, Buchanan SK. 2010. TonB-dependent transporters: regulation, structure, and function. *Annu Rev Microbiol* 64:43-60.
 174. Hantke K. 2003. Is the bacterial ferrous iron transporter FeoB a living fossil? *Trends Microbiol* 11:192-5.
 175. Choby JE, Grunenwald CM, Celis AI, Gerdes SY, DuBois JL, Skaar EP. 2018. *Staphylococcus aureus* HemX modulates glutamyl-tRNA reductase abundance to regulate heme biosynthesis. *mBio* 9.
 176. Akduman N, Lightfoot JW, Roseler W, Witte H, Lo WS, Rodelsperger C, Sommer RJ. 2020. Bacterial vitamin B12 production enhances nematode predatory behavior. *ISME J* 14:1494-1507.
 177. Mendel RR. 2013. The molybdenum cofactor. *J Biol Chem* 288:13165-72.
 178. Ringel P, Krausze J, van den Heuvel J, Curth U, Pierik AJ, Herzog S, Mendel RR, Kruse T. 2013. Biochemical characterization of molybdenum cofactor-free nitrate reductase from *Neurospora crassa*. *J Biol Chem* 288:14657-14671.
 179. Zhong Q, Kobe B, Kappler U. 2020. Molybdenum enzymes and how they support virulence in pathogenic bacteria. *Front Microbiol* 11:615860.
 180. Perinet S, Jeukens J, Kukavica-Ibrulj I, Ouellet MM, Charette SJ, Levesque RC. 2016. Molybdate transporter ModABC is important for *Pseudomonas aeruginosa* chronic lung infection. *BMC Res Notes* 9:23.

181. Jakes KS, Cramer WA. 2012. Border crossings: colicins and transporters. *Annu Rev Genet* 46:209-31.
182. Cao Y, Gao L, Zhang L, Zhou L, Yang J, Deng L, Zhao J, Qi C, Liu J. 2020. Genome-wide screening of lipoproteins in *Actinobacillus pleuropneumoniae* identifies three antigens that confer protection against virulent challenge. *Sci Rep* 10:2343.
183. Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Res* 43:W39-49.
184. Frith MC, Saunders NF, Kobe B, Bailey TL. 2008. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol* 4:e1000071.
185. Bailey TL, Gribskov M. 1998. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14:48-54.
186. Dearth SP, Castro HF, Venice F, Tague ED, Novero M, Bonfante P, Campagna SR. 2018. Metabolome changes are induced in the arbuscular mycorrhizal fungus *Gigaspora margarita* by germination and by its bacterial endosymbiont. *Mycorrhiza* 28:421-433.
187. UniProt C. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49:D480-D489.
188. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674-6.
189. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci* 30:70-82.
190. Eom S, Park Y, Kim Y. 2014. Sequential immunosuppressive activities of bacterial secondary metabolites from the entomopathogenic bacterium *Xenorhabdus nematophila*. *J Microbiol* 52:161-8.
191. De Lerma Barbaro A, Gariboldi MB, Mastore M, Brivio MF, Giovannardi S. 2019. In vivo effects of a Pro-PO System inhibitor on the phagocytosis of *Xenorhabdus nematophila* in *Galleria mellonella* larvae. *Insects* 10.
192. Wolff H, Bode HB. 2018. The benzodiazepine-like natural product tilivalline is produced by the entomopathogenic bacterium *Xenorhabdus eapokensis*. *PLoS One* 13:e0194297.
193. Tobias NJ, Wolff H, Djahanschiri B, Grundmann F, Kronenwerth M, Shi YM, Simonyi S, Grun P, Shapiro-Ilan D, Pidot SJ, Stinear TP, Ebersberger I, Bode HB. 2017. Natural product diversity associated with the nematode symbionts *Photorhabdus* and *Xenorhabdus*. *Nat Microbiol* 2:1676-1685.
194. Unterhauser K, Pörtl L, Schneditz G, Kienesberger S, Glabonjat RA, Kitsera M, Pletz J, Josa-Prado F, Dornisch E, Lembacher-Fadum C, Roier S, Gorkiewicz G, Lucena D, Barasoain I, Kroutil W, Wiedner M, Loizou JI, Breinbauer R, Díaz JF, Schild S, Högenauer C, Zechner EL. 2019. *Klebsiella oxytoca*; enterotoxins tilimycin and tilivalline have distinct host DNA-damaging and microtubule-stabilizing activities. *Proceedings of the National Academy of Sciences* 116:3774.
195. Ernst M, Kang KB, Caraballo-Rodriguez AM, Nothias LF, Wandy J, Chen C, Wang M, Rogers S, Medema MH, Dorrestein PC, van der Hooft JJJ. 2019. MolNetEnhancer: Enhanced molecular networks by integrating metabolome mining and annotation tools. *Metabolites* 9.
196. Heungens K, Cowles CE, Goodrich-Blair H. 2002. Identification of *Xenorhabdus nematophila* genes required for mutualistic colonization of *Steinernema carpocapsae* nematodes. *Mol Microbiol* 45:1337-53.
197. Hirano S, Asamizu S, Onaka H, Shiro Y, Nagano S. 2008. Crystal structure of VioE, a key player in the construction of the molecular skeleton of violacein. *J Biol Chem* 283:6459-66.

198. Egler M, Grosse C, Grass G, Nies DH. 2005. Role of the extracytoplasmic function protein family sigma factor RpoE in metal resistance of *Escherichia coli*. *J Bacteriol* 187:2297-307.
199. Bleiziffer I, Eikmeier J, Pohlentz G, McAulay K, Xia G, Hussain M, Peschel A, Foster S, Peters G, Heilmann C. 2017. The Plasmin-sensitive protein Pls in methicillin-resistant *Staphylococcus aureus* (MRSA) is a glycoprotein. *PLoS Pathog* 13:e1006110.
200. Huesca M, Peralta R, Sauder DN, Simor AE, McGavin MJ. 2002. Adhesion and virulence properties of epidemic Canadian methicillin-resistant *Staphylococcus aureus* strain 1: identification of novel adhesion functions associated with plasmin-sensitive surface protein. *J Infect Dis* 185:1285-96.
201. Micenkova L, Bosak J, Kucera J, Hrala M, Dolejsova T, Sedo O, Linke D, Fiser R, Smajs D. 2019. Colicin Z, a structurally and functionally novel colicin type that selectively kills enteroinvasive *Escherichia coli* and *Shigella* strains. *Sci Rep* 9:11127.
202. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202-8.
203. Li Z, Zhu D, Shen Y. 2018. Discovery of novel bioactive natural products driven by genome mining. *Drug Discov Ther* 12:318-328.
204. Booyesen E, Rautenbach M, Stander MA, Dicks LMT. 2021. Profiling the production of antimicrobial secondary metabolites by *Xenorhabdus khoisanus* J194 under different culturing conditions. *Front Chem* 9:626653.
205. Herbert Tran EE, Goodrich-Blair H. 2009. CpxRA contributes to *Xenorhabdus nematophila* virulence through regulation of *IrhA* and modulation of insect immunity. *Applied and Environmental Microbiology* 75:3998.
206. Qin R, Sang Y, Ren J, Zhang Q, Li S, Cui Z, Yao YF. 2016. The bacterial two-hybrid system uncovers the involvement of acetylation in regulating of *Lrp* activity in *Salmonella typhimurium*. *Front Microbiol* 7:1864.
207. Zhang K, Zheng S, Yang JS, Chen Y, Cheng Z. 2013. Comprehensive profiling of protein lysine acetylation in *Escherichia coli*. *J Proteome Res* 12:844-51.
208. Bing X, Attardo GM, Vigneron A, Aksoy E, Scolari F, Malacrida A, Weiss BL, Aksoy S. 2017. Unravelling the relationship between the tsetse fly and its obligate symbiont *Wigglesworthia*: transcriptomic and metabolomic landscapes reveal highly integrated physiological networks. *Proceedings Biological sciences* 284:20170360.
209. Martin RM, Moniruzzaman M, Mucci NC, Willis A, Woodhouse JN, Xian Y, Xiao C, Brussaard CPD, Wilhelm SW. 2019. *Cylindrospermopsis raciborskii* virus and host: genomic characterization and ecological relevance. *Environ Microbiol* 21:1942-1956.

APPENDICES

Appendix 1: Nematocidal activity of a small molecule produced by *X. nematophila* (GKA-L-Arginine assays)

Author contributions:

Nicholas Mucci: Experimental design, data collection and analysis, writeup.

Sarah Kauffman: Experimental design and sample prep.

Jessica DeCuyper: Data analysis and writeup.

Publication status: To be determined

Introduction

Genome mining aids in the discovery of natural products that could be applied pharmaceutically and agriculturally (203). Using genome mining techniques on the Gram-negative bacterium *Xenorhabdus nematophila* could lead to novel secondary metabolite discovery. Such metabolites have been identified in *Xenorhabdus* species, such as xenocoumacins and xenoamicins which have antimicrobial activity (204). The transcriptional regulator LysR homolog A-encoding gene *lrhA* in the bacterium negatively regulates the cupin locus (58, 205). This locus encodes metabolic biosynthetic machinery and synthesizes the compound Gamma-Keto-Acyl-Arginine and Proline (GKA-L-Arginine and GKA-L-Proline). Previous phenotypic assays indicated that both compounds affected nematode growth and development. This experiment aimed to see how more biologically relevant doses similar to those found in nature of GKA-L-Arginine to see how it affects nematodes. We hypothesize that the compound will have similar results to previous experiments where the compound inhibits nematode egg hatching and length, even when exposed to lower concentrations of the compound.

Results

All graphs show mean $\pm$ standard deviation between 3 technical replicates.

Steinernema carpocapsae GKA-L-Arginine assays

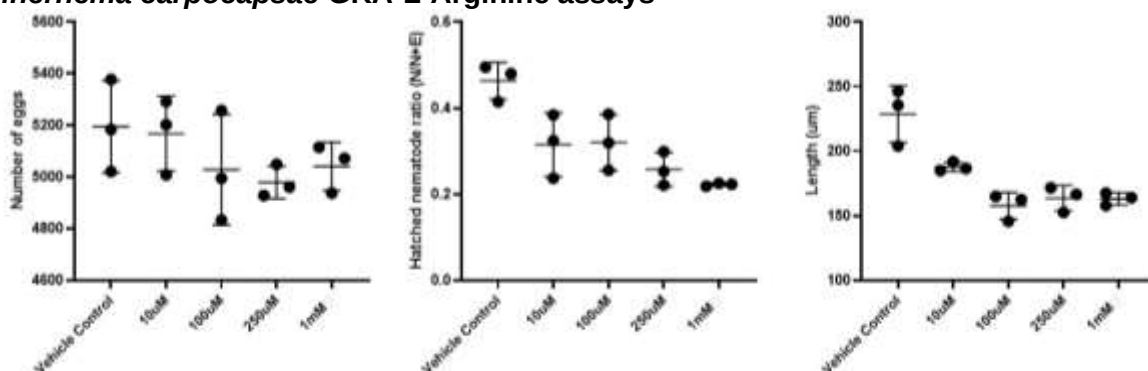


Figure A1.1: *Steinernema carpocapsae* GKA-L-Arginine assays. Left panel: Egg counts on Day 0 to ensure each well contained the same number of eggs and thus nematodes came into contact with the compound as frequently in every treatment. Middle panel: Ratio of eggs hatched on Day 4 to determine if the compound inhibits egg hatching. Right panel: Nematode length on Day 4 to determine if the compound decreases nematode length.

Table A1.1: Day 0 egg counts

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	0.9994	0.9980
Vehicle Control vs. 100uM	0.6632	0.4927
Vehicle Control vs. 250uM	0.4422	0.2904
Vehicle Control vs. 1mM	0.7226	0.5572
10uM vs. 100uM:	0.7820	
10uM vs. 250uM	0.5597	
10uM vs. 1mM	0.8339	
100uM vs. 250uM	0.9936	
100uM vs. 1mM	>0.9999	
250uM vs. 1mM	0.9845	

Conclusion: Roughly the same amount of *S. carpocapsae* eggs were put into each well.

Table A1.2: Day 4 hatched vs. unhatched nematode ratio

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	<u>0.0332</u>	<u>0.0170</u>
Vehicle Control vs. 100uM	<u>0.0402</u>	<u>0.0207</u>
Vehicle Control vs. 250uM	<u>0.0042</u>	<u>0.0020</u>
Vehicle Control vs. 1mM	<u>0.0013</u>	<u>0.0006</u>
10uM vs. 100uM:	>0.9999	
10uM vs. 250uM	0.6538	
10uM vs. 1mM	0.2460	
100uM vs. 250uM	0.5842	
100uM vs. 1mM	0.2072	
250uM vs. 1mM	0.9095	

Conclusion: GKA-L-Arginine causes a significant decrease in the hatching of *S. carpocapsae* nematodes when compared to the vehicle control.

Table A1.3: Panel 3: Day 4 nematode length calculations

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	<u>0.0134</u>	<u>0.0066</u>
Vehicle Control vs. 100uM	<u>0.0002</u>	<u>0.0001</u>
Vehicle Control vs. 250uM	<u>0.0004</u>	<u>0.0002</u>
Vehicle Control vs. 1mM	<u>0.0004</u>	<u>0.0002</u>
10uM vs. 100uM:	0.0716	
10uM vs. 250uM	0.1763	
10uM vs. 1mM	0.1646	
100uM vs. 250uM	0.9721	
100uM vs. 1mM	0.9792	
250uM vs. 1mM	>0.9999	

Conclusion: GKA-L-Arginine causes a significant decrease in the length of *S. carpocapsae* nematodes when compared to the vehicle control.

Steinernema anatoliense GKA-L-Arginine assays

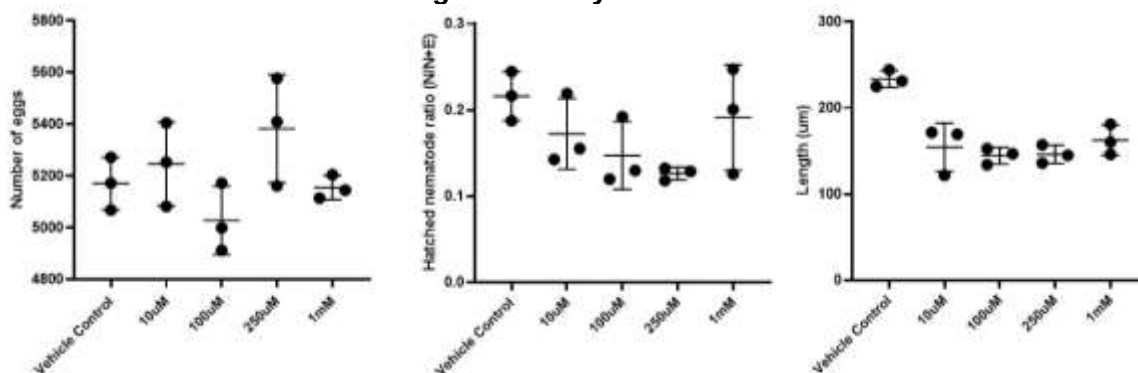


Figure A1.2: *Steinernema anatoliense* GKA-L-Arginine assays. Left panel: Egg counts on Day 0 to ensure each well contained the same number of eggs and thus nematodes came into contact with the compound as frequently in every treatment. Middle panel: Ratio of eggs hatched on Day 4 to determine if the compound inhibits egg hatching. Right panel: Nematode length on Day 4 to determine if the compound decreases nematode length.

Table A1.4: Panel 1: Day 0 egg counts

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	0.9609	0.9068
Vehicle Control vs. 100uM	0.7304	0.5661
Vehicle Control vs. 250uM	0.4038	0.2599
Vehicle Control vs. 1mM	>0.9999	0.9998
10uM vs. 100uM:	0.3765	
10uM vs. 250uM	0.7612	
10uM vs. 1mM	0.9255	
100uM vs. 250uM	0.0693	
100uM vs. 1mM	0.8024	
250uM vs. 1mM	0.3403	

Conclusion: Roughly the same amount of *S. anatoliense* eggs were put into each well.

Table A1.5: Day 4 hatched vs. unhatched nematode ratio

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	0.6665	0.4961
Vehicle Control vs. 100uM	0.2785	0.1679
Vehicle Control vs. 250uM	0.1103	0.0602
Vehicle Control vs. 1mM	0.9336	0.8539
10uM vs. 100uM:	0.9312	
10uM vs. 250uM	0.6290	
10uM vs. 1mM	0.9743	
100uM vs. 250uM	0.9641	
100uM vs. 1mM	0.6618	
250uM vs. 1mM	0.3289	

Conclusion: GKA-L-Arginine does not significantly affect the hatching of *S. anatoliense* nematodes when compared to the vehicle control.

Table A1.6: Day 4 nematode length calculations

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	<u>0.0013</u>	<u>0.0006</u>
Vehicle Control vs. 100uM	<u>0.0005</u>	<u>0.0003</u>
Vehicle Control vs. 250uM	<u>0.0006</u>	<u>0.0003</u>
Vehicle Control vs. 1mM	<u>0.0029</u>	<u>0.0014</u>
10uM vs. 100uM:	0.9466	
10uM vs. 250uM	0.9724	
10uM vs. 1mM	0.9713	
100uM vs. 250uM	>0.9999	
100uM vs. 1mM	0.6833	
250uM vs. 1mM	0.7512	

Conclusion: GKA-L-Arginine causes a significant decrease in the length of *S. anatoliense* nematodes when compared to the vehicle control.

VIDEO NOTES FOR *S. anatoliense* (taken by Jessica DeCuyper): Not all of the videos in the data set would have the same nematode observations. Ranging from most active to least active data sets, I would rank set 5 as the most active, set 3 as second active, set 2, set 1, and then set 6 as the least active. Set 5 appeared to have the most active and healthy nematodes based

on their movement and the overall plump appearance of the hatched nematodes. The nematodes in this data set also appeared to have all of the nematodes moving. Set three nematodes did have the majority of the nematodes moving, with a few exceptions of inhibited nematodes. Their activity level and appearance were very close to that of set five. Set 2 was ranked as middle movement since 2.3 had the majority of the nematodes moving. However, this was not consistently seen in 2.1 or 2.2, where only a few nematodes were observed moving. Set 1 was ranked after set 2 since fewer nematodes were observed moving compared to the ones present in set 2. Set 1 also had more visibly truncated appearance in the nematodes compared to set 2. Set 6 was ranked last in movement and seemed to be the most inhibited of all of the video sets. Almost all nematodes in set 6 were observed as inhibited in movement, or the very few that were seen moving had a small twitching movement.

Nick edit- After unblinding the data, most to least active: Set LB, EtOH, 10uM, 250uM, 100uM.

Set 4 (Jessica did not list set 4 [1mM], but based off the notes she took, they seem the least active of all sets.

Heterorhabditis bacteriophora GKA-L-Arginine assays

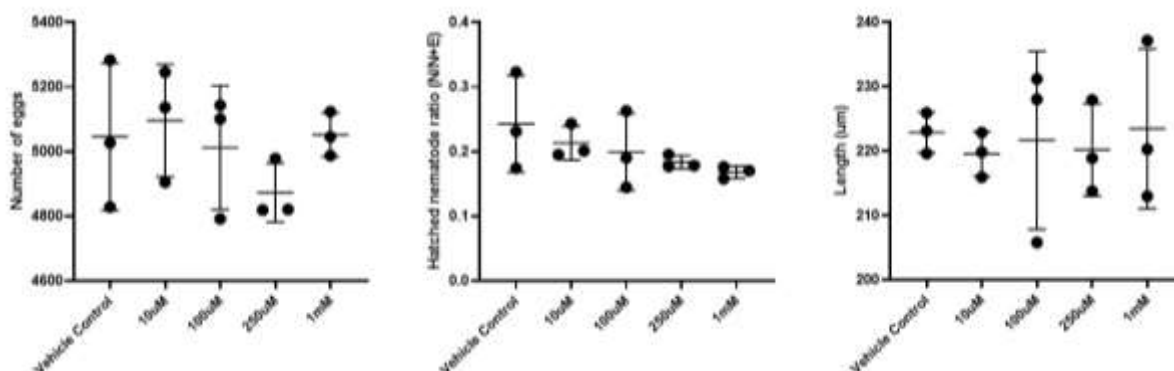


Figure A1.3: *Heterorhabditis bacteriophora* GKA-L-Arginine assays. Left panel: Egg counts on Day 0 to ensure each well contained the same number of eggs and thus nematodes came into contact with the compound as frequently in every treatment. Middle panel: Ratio of eggs hatched on Day 4 to determine if the compound inhibits egg hatching. Right panel: Nematode length on Day 4 to determine if the compound decreases nematode length.

Table A1.7: Day 0 egg counts

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	0.9951	0.9860
Vehicle Control vs. 100uM	0.9989	0.9966
Vehicle Control vs. 250uM	0.6938	0.5253
Vehicle Control vs. 1mM	>0.9999	>0.9999
10uM vs. 100uM:	0.9665	
10uM vs. 250uM	0.4853	
10uM vs. 1mM	0.9969	
100uM vs. 250uM	0.8269	
100uM vs. 1mM	0.9980	
250uM vs. 1mM	0.6699	

Conclusion: Roughly the same amount of *H. bacteriophora* eggs were put into each well.

Table A1.8: Day 4 hatched vs. unhatched nematode ratio

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	0.9222	0.8333
Vehicle Control vs. 100uM	0.7572	0.5978
Vehicle Control vs. 250uM	0.5199	0.3559
Vehicle Control vs. 1mM	0.3158	0.1942
10uM vs. 100uM:	0.9948	
10uM vs. 250uM	0.9223	
10uM vs. 1mM	0.7359	
100uM vs. 250uM	0.9920	
100uM vs. 1mM	0.9091	
250uM vs. 1mM	0.9923	

Conclusion: GKA-L-Arginine does not significantly affect the hatching of *H. bacteriophora* nematodes when compared to the vehicle control.

Table A1.9: Day 4 nematode length calculations

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	0.9905	0.9738
Vehicle Control vs. 100uM	0.9998	0.9993
Vehicle Control vs. 250uM	0.9959	0.9882
Vehicle Control vs. 1mM	>0.9999	0.9999
10uM vs. 100uM:	0.9983	
10uM vs. 250uM	>0.9999	
10uM vs. 1mM	0.9827	
100uM vs. 250uM	0.9996	
100uM vs. 1mM	0.9991	
250uM vs. 1mM	0.9913	

Conclusion: GKA-L-Arginine causes a significant decrease in the length of *H. bacteriophora* nematodes when compared to the vehicle control.

VIDEO NOTES FOR *H. bacteriophora* (taken by Jessica DeCuyper): Overall comparison between the videos was that set 5.2 had the most movement compared to the others. The videos in set 5 (5.1-5.3) did have quite a bit of movement with the videos in set 3 (3.1-3.3) coming close behind. Set five videos also appeared to have more hatched nematodes

compared to the others. Set 2 videos (2.1-2.3) seemed to have the least amount of movement in them, with the majority of the nematodes appearing dead. An overall ranking of movement from highest to lowest would be set 5, set 3, set 4, set 1, and set 2 with the least amount.

Nick edit- After unblinding the data, an overall ranking of movement from highest to lowest would be LB, EtOH, 100uM, 250uM, and 1mM with the least amount (Jessica forgot to list 10 [set 6] but those observations look in line with what we saw and could be placed in between 100uM and EtOH).

Caenorhabditis elegans GKA-L-Arginine assays

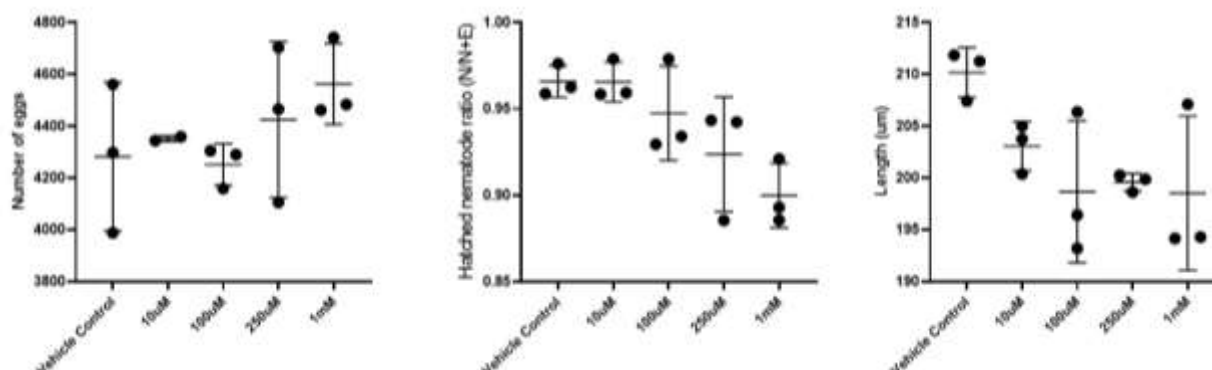


Figure A1.4: *Caenorhabditis elegans* GKA-L-Arginine assays. Left panel: Egg counts on Day 0 to ensure each well contained the same number of eggs and thus nematodes came into contact with the compound as frequently in every treatment. Middle panel: Ratio of eggs hatched on Day 4 to determine if the compound inhibits egg hatching. Right panel: Nematode length on Day 4 to determine if the compound decreases nematode length.

Table A1.10: Day 0 egg counts

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	0.9959	0.9888
Vehicle Control vs. 100uM	0.9997	0.9991
Vehicle Control vs. 250uM	0.9170	0.8303
Vehicle Control vs. 1mM	0.5237	0.3670
10uM vs. 100uM:	0.9836	
10uM vs. 250uM	0.9948	
10uM vs. 1mM	0.8082	
100uM vs. 250uM	0.8486	
100uM vs. 1mM	0.4317	
250uM vs. 1mM	0.9266	

Conclusion: Roughly the same amount of *C. elegans* eggs were put into each well.

Table A1.11: Day 4 hatched vs. unhatched nematode ratio

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	>0.9999	>0.9999
Vehicle Control vs. 100uM	0.8394	0.7042
Vehicle Control vs. 250uM	0.2074	0.1202
Vehicle Control vs. 1mM	0.0274	0.0138
10uM vs. 100uM:	0.8460	
10uM vs. 250uM	0.2117	
10uM vs. 1mM	0.0280	
100uM vs. 250uM	0.6838	
100uM vs. 1mM	0.1320	
250uM vs. 1mM	0.6791	

Conclusion: High doses (1mM) of GKA-L-Arginine significantly affects the hatching of *C. elegans* nematodes when compared to the vehicle control.

Table A1.12: Day 4 nematode length calculations

Comparison	1 way ANOVA with multiple comparisons (Tukey) p-value	1 way ANOVA with multiple comparisons (Dunnett's) p-value
Vehicle Control vs. 10uM	0.4118	0.2662
Vehicle Control vs. 100uM	0.0853	0.0457
Vehicle Control vs. 250uM	0.1219	0.0671
Vehicle Control vs. 1mM	0.0807	0.0431
10uM vs. 100uM:	0.7828	
10uM vs. 250uM	0.8959	
10uM vs. 1mM	0.7741	
100uM vs. 250uM	0.9992	
100uM vs. 1mM	>0.9999	
250uM vs. 1mM	0.9986	

Conclusion: GKA-L-Arginine causes a significant decrease in the length of *C. elegans* nematodes when compared to the vehicle control.

Discussion

Of the nematode species tested, GKA-L-Arg had negative effects on nematode growth and development (as measured by the number of eggs hatched and size of the fully grown nematodes) on *S. carpocapsae*, *S. anatoliense*, and *C. elegans*. In all species tested, there was a neurological phenotype as determined by movement speed and thrashing when the compound was administered. Note: *S. carpocapsae* and *C. elegans* not shown due to inconsistencies with filming method compared to the more streamlined *S. anatoliense* and *H. bacteriophora*, although I do have those video files.

During the later phases of the life cycle after the insect is dead, we would expect more GKA-L-Arginine to be produced. This is because *IrhA* activity is expected to decrease since most of the *IrhA* regulatory pathway we know of includes virulence factors against the insect, and *IrhA* negatively regulates the locus that produces the enzymes that synthesize GKA-L-Arginine. As a result, the bacteria could be producing the compound as means of nematode population control. Slowly the population increase would be beneficial for the pair, where less nematodes hatching means the ones that do hatch can sequester more nutrients from the insect cadaver. The ones that survive can get fatter, which would be more useful during the stages of starvation that follow until they find more insect prey. Nematode size being affected could mean that the nutrients scavenged from the cadaver would be better distributed among the population, with less larger worms taking more nutrients. This would increase the fitness of the nematode population as a whole. The neurological phenotypes could be a method the bacteria have evolved to deploy as a means of increasing time spent in the cadaver. Less nematode movement means less bacteria are getting consumed, which allows them more time to convert the cadaver into nutrients.

Outside of *S. carpocapsae*, another *Steinernema* species and *C. elegans* were also affected by the compound with similar phenotypes. Thus, GKA-L-Arg can have wider effects than what was studied here. Even at low, seemingly biologically relevant concentrations, this

compound could be applied as a nematode biocontrol. It could be used to slow population growth. More species will need to be tested to support this hypothesis.

Materials and methods

To test how the compound interacts with nematodes, we prepared nematode eggs (*Steinernema carpocapsae*, *Steinernema feltiae*, *Steinernema anatoliense*, *Steinernema hermaphroditum*, *Steinernema scapterisci*, *Caenorhabditis elegans*, and *Heterorhabditis bacteriophora*.) Bleaching time to make the eggs axenic was reduced by 50% of standard protocols, as we felt this reduced the stress put on the eggs. GKA-L-Arg was diluted using 50% ethanol to 10X desired final concentration. The ethanol-diluted GKA-L-Arg was then added to axenic eggs at 4 different final concentrations: 1 mM, 250 μ M, 100 μ M, and 10 μ M. For a control, we added the same volume of 50% ethanol to the eggs. We also had a control where no ethanol was added to the eggs (data not shown). For each treatment, there were 3 replicates in 3 separate wells of a 24 well plate. Images were taken on a Keyence microscope (BZ-X710) at 24 hour time points, starting at $t=0$ and continuing for 4 days. 4X images of the whole well were taken, as well as 49 20X images starting at the center of the well, and subsequently stitched together. Pictures were then blinded to be quantified. Videos were also recorded everyday to monitor nematode movement. 30 second clips were taken and blinded. The videos best suited for quantification were for *S. anatoliense* and *H. bacteriophora* because the camera stayed in position for the record, and thus those were used for qualitative observations. Data obtained from day 4 was used in all quantifications. ImageJ was used to measure nematode length. Egg hatching was quantified using the equation:

$$\text{Hatched nematode ratio} = N/N+E$$

N= nematodes, E= eggs. Dark circular structures were categorized as intact eggs. Statistics were calculated on Graphpad Prism platform. All statistical tests used were one way ANOVA with post-hoc Dunnett's test or post-hoc Tukey test.

S. hermaphroditum and *S. scapterisci* experienced dumpy phenotypes after several egg preps of each. At about day 3 or 4, all the nematodes (including our ethanol and no ethanol controls) turned dumpy. Because of this, we had to stop testing on these species. Additionally, *S. feltiae* could not hatch when administered ethanol treatment. Again, several egg preps were performed. Because of this, *S. feltiae* could not be used for these assays either, as this vehicle was used for every other protocol performed. This should be investigated further as to why certain species of nematodes can handle EtOH and why others cannot.

Other assays ran on both GKA-L-Arg and GKA-L-Pro were virulence assays in *Manduca sexta*, growth curves on *Bacillus subtilis*, *Escherichia coli*, and *Xenorhabdus nematophila*, and ring of inhibition assays on those bacterial species. None of these experiments produced significant results and are thus not shown here.

Appendix 2: Summary of findings with Leucine-responsive regulatory protein (Lrp), which acts as a virulence switch in *Xenorhabdus nematophila*

Publication status: to be submitted when additional investigations on Lrp regulation by Cameron Moore are completed as:

Nutritional conditions influence *Xenorhabdus nematophila* Lrp-dependent virulence modulation

Mengyi Cao^a, Cameron Moore^b, Luella Allen-Waller^b, Nicholas Mucci^b, Sarah J. Kauffman^b, Jennifer Heppert^b, Ellen Gough^a, Heidi Goodrich-Blair^{a,b}

^a Department of Bacteriology, University of Wisconsin-Madison, WI 53706, USA

^b Department of Microbiology, University of Tennessee-Knoxville, TN 37996, USA

I investigated the regulatory mechanisms that could be modulating Lrp activity (Figure A2.1) as part of my early graduate work. Lrp in *X. nematophila* controls the virulence switch from mutualism to antagonism once the bacterium reaches the prey insect (25). Pat is a phosphotransacetylase that acetylates Lrp in *Escherichia coli* and *Salmonella typhimurium* (206, 207). Acetylation causes Lrp to become inactive. CobB is a sirtuin that uses NAD⁺ as a substrate to deacetylate Lrp, thus increases Lrp activity. In *X. nematophila*, when Lrp is low, virulence of the bacterium increases. High Lrp in *X. nematophila* has been linked to mutualism phenotypes. As part of this work, I generated the Δpat strain, as well as introducing fluorescent Lrp reporters in the $\Delta cobB$ and Δpat backgrounds. Open questions remain, such as confirming $\Delta cobB$ and Δpat strains have phenotypes to the constitutively high and low Lrp strains. We also should confirm if Lrp is acetylated, to see if Lrp activity is modulated by Pat and CobB, and observe if NAD⁺ and acetate alter Lrp activity.

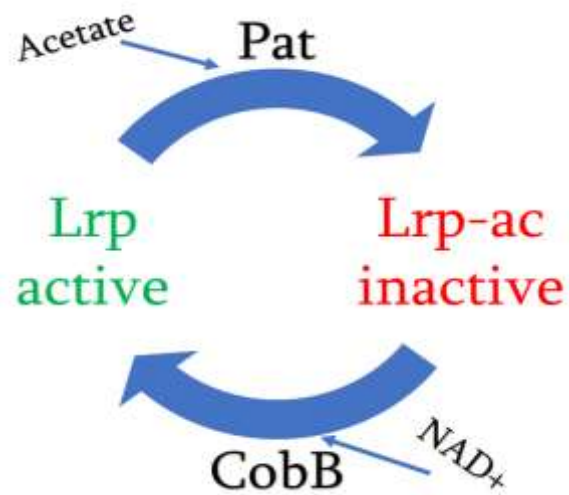


Figure A2.1: Lrp regulation in *E. coli* and *S. typhimurium*. Unacetylated, active Lrp is acetylated by protein transacetylase (Pat), decreasing the amount of active Lrp. CobB, a sirtuin, can use NAD⁺ as a substrate to deacetylate Lrp, returning it to an active state. We hypothesized this model was applicable to *X. nematophila*.

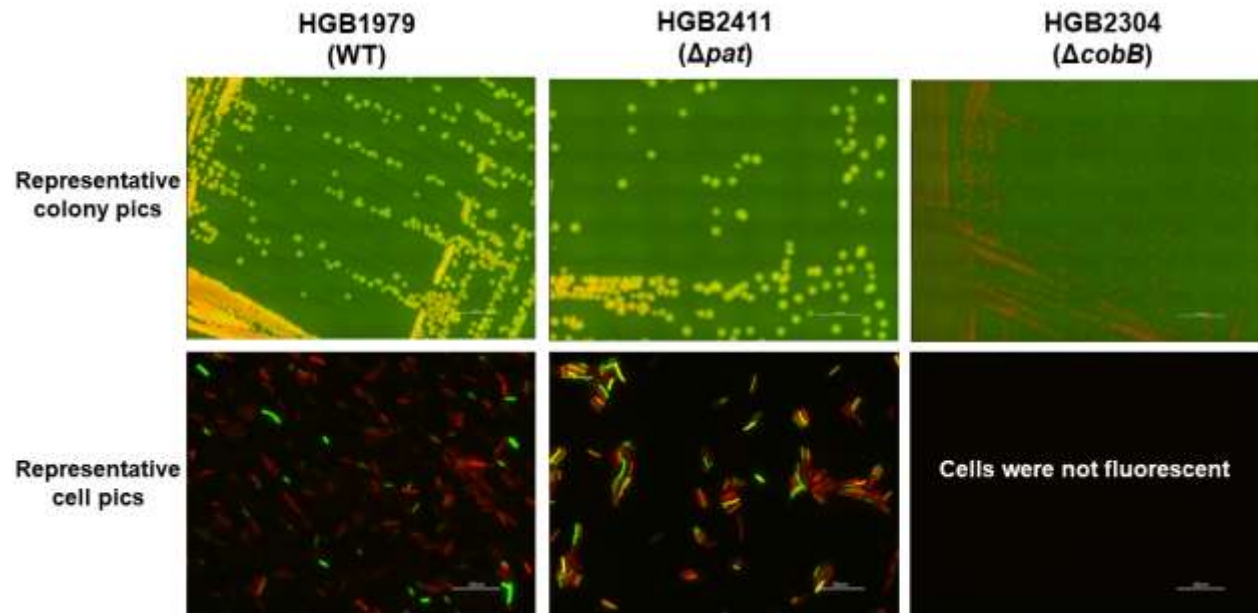


Figure A2.2: Preliminary fluorescence data shows possible Lrp acetylation and regulation by Pat and CobB. Strains are HGB1979 (WT *X. nematophila* with fluorescent reporter), HGB2411 (Δpat *X. nematophila* with fluorescent reporter), and HGB 2304 ($\Delta cobB$ *X. nematophila* with fluorescent reporter). Low Lrp levels are linked to red colony and cell phenotypes in the $\Delta cobB$ mutant. High Lrp levels are linked to the green colony and cell phenotypes in the Δpat mutant.

Appendix 3: F1000 Review on “Unravelling the relationship between the tsetse fly and its obligate symbiont *Wigglesworthia*: transcriptomic and metabolomic landscapes reveal highly integrated physiological networks.”

Publication status: This review is available to read at: <https://doi.org/10.1098/rsph.2017.0360>

“Mutually beneficial relationships between symbionts and hosts are mediated through biochemical “conversations”. This study integrates RNA-seq and mass spectrometry to characterize this dialogue between the tsetse fly and its obligate gut endosymbiont *Wigglesworthia glossinidia*. Both host and symbiont transcripts were sequenced from bacteriome tissue, where the symbionts are housed. These transcript profiles were then compared to midgut tissue samples from a prior study to identify transcripts enriched in the bacteriome. These included those encoding tsetse fly PGRP-LB, a protein that protects the symbiont from host immunity factors, a C-type lectin, and transporters that facilitate movement of metabolites across the bacteriocyte membranes. Metabolites from symbiotic flies were compared to those from flies cured of their symbionts. These data, coupled with the transcriptomic analysis, indicated that the presence of the *Wigglesworthia* symbiont is critical for homeostasis of host carbon, nucleotide, lipid, and amino acid metabolism, predominantly due to symbiont provisioning of vitamin cofactors that are essential for these processes. The combined transcriptomic and metabolomic datasets provides a detailed picture of the deeply inter-dependent physiologies of animal and bacterial cells that co-exist in a symbiosis. As the authors point out, such knowledge can be exploited to manipulate symbioses for the benefit of society.”

I worked on this summary with Heidi as part of my initial literature review to prepare my 2nd chapter. This manuscript helped me conceptualize what a paper integrating metabolomic and transcriptomic data could look like so we took the time to carefully craft this F1000 review (208). This was published on June 28th, 2017.

Appendix 4: Manuscript contributions to “*Cylindrospermopsis raciborskii* Virus and host: genomic characterization and ecological relevance” as part of my rotation in the Wilhelm Lab working under Dr. Robbie Martin.

Publication status: Submitted and accepted in Environmental Microbiology (209).

Paper abstract: *Cylindrospermopsis (Raphidiopsis) raciborskii* is an invasive, filamentous, nitrogen-fixing cyanobacterium that forms frequent blooms in freshwater habitats. While viruses play key roles in regulating the abundance, production and diversity of their hosts in aquatic ecosystems, the role(s) of viruses in the ecology of *C. raciborskii* is almost unexplored. Progress in this field has been hindered by the absence of a characterized virus–host system in *C. raciborskii*. To bridge this gap, we sequenced the genome of CrV-01T, a previously isolated cyanosiphovirus, and its host, *C. raciborskii* strain Cr2010. Analyses suggest that CrV-01T represents a distinct clade of siphoviruses infecting, and perhaps lysogenizing, filamentous cyanobacteria. Its genome contains unique features that include an intact CRISPR array and a 12 kb inverted duplication. Evidence suggests CrV-01T recently gained the ability to infect Cr2010 and recently lost the ability to form lysogens. The cyanobacterial host contains a CRISPR-Cas system with CRISPR spacers matching protospacers within the inverted duplication of the CrV-01T genome. Examination of metagenomes demonstrates that viruses with high genetic identity to CrV-01T, but lacking the inverted duplication, are present in *C. raciborskii* blooms in Australia. The unique genomic features of the CrV/Cr2010 system offers opportunities to investigate in more detail virus–host interactions in an ecologically important bloom-forming cyanobacterium.

My specific contribution to this work was the assembly of the *C. raciborskii* host genome using metagenomic lake data samples. I tested a variety of different assemblers, including CLC Genomics Workbench and metaSPADEs, to reduce the contigs as much as possible. The final

resulting host genome (after trimming for quality reads) was 3.55 Mb in size comprising of 89 contigs. Subsequent analysis of the contents of the genome and comparison to the CRISPR spaces in the assembled viral genome were performed by Robbie Martin in the Wilhelm Lab, as my rotation in the lab came to a close by the time I finished the assembly.

Appendix 5: Manuscript contributions to “Synergy between T6SSs is an important component of *X. nematophila* infection of the insect host”

Publication status: The manuscript was originally submitted to Applied and Environmental Microbiology in 2020 and is currently in revision.

Paper abstract (as it stands): “The type VI secretion system (T6SS) acts as a molecular weapon, delivering toxic effector molecules into prey cells. Agar plate-based experiments indicate that these systems function in bacterial competition, but T6SS function in ecologically relevant settings is less well known. The gram-negative bacterium *Xenorhabdus nematophila* and its nematode host *Steinernema carpocapsae*, engage in a mutualistic relationship and together parasitize insects. Inside the insect, *X. nematophila* is a pathogen and must overcome immune defenses present in the insect blood system. Here, we assessed the role of two T6SSs, T6SS-1 and T6SS-2, in *X. nematophila* ATCC 19061 during interaction with the agricultural insect pest, *Manduca sexta*. We discovered that in *X. nematophila* both T6SSs are necessary for normal cytolytic activity against *M. sexta* immune cells. The T6SS-2 deletion mutant had attenuated virulence as well as overactivation of the prophenoloxidase (proPO) immune cascade. These T6SS-2 mutant virulence and proPO suppression defects were rescued when the T6SS-1 locus was also deleted, suggesting an unexpected coordination of activity between the two systems. We searched for possible T6SS effectors of host interaction using a proteomic and a comparative genomic analysis. The former discovered possible novel T6SS effectors and suggests a role of *X. nematophila* T6SS in redox homeostasis. The genomic analysis revealed widespread conservation of T6SSs across the *Xenorhabdus* genus and significant variation in number of T6SS clusters and effector chaperones at both the species and strain levels. Our study reveals new links between T6SS, immune system interactions, and cluster coordination in the *Xenorhabdus* lifecycle.”

I was responsible for the proteomic analysis between T6SS strains and WT. I analyzed the proteomes between groups, using a $|\text{fold change difference}| > 1.0\text{-fold}$ and $P < 0.05$ signal cutoff between protein abundances in the T6SS mutants and WT. I generated volcano plots that showed the differentially expressed proteins between strain backgrounds (Figure A5.1), including those between $\Delta T6SS$ strains (not shown). I also generated the hierarchical clustering analysis between all of the strains to compare their all of their protein profiles together (Figure A5.2). Dr. Hillman took this statistical analysis I performed to interpret why certain protein abundances are up in the different strain backgrounds.

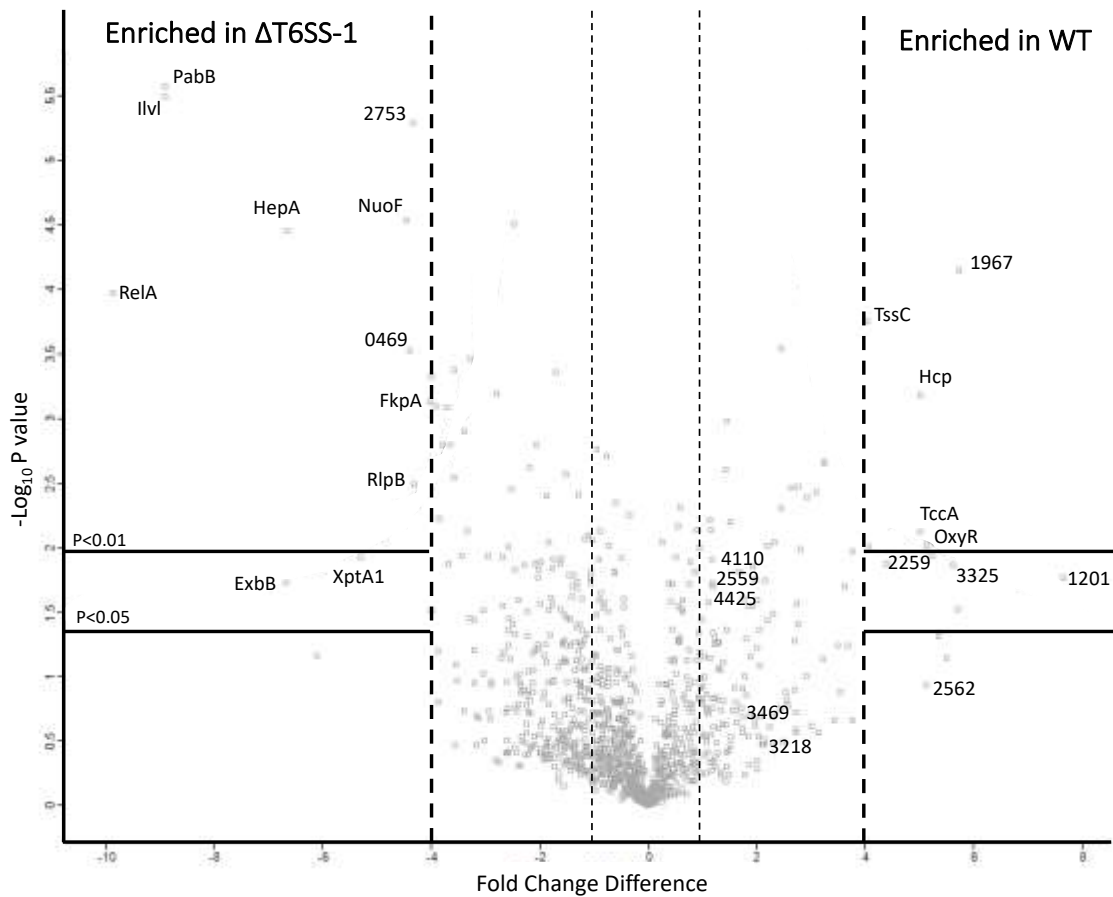
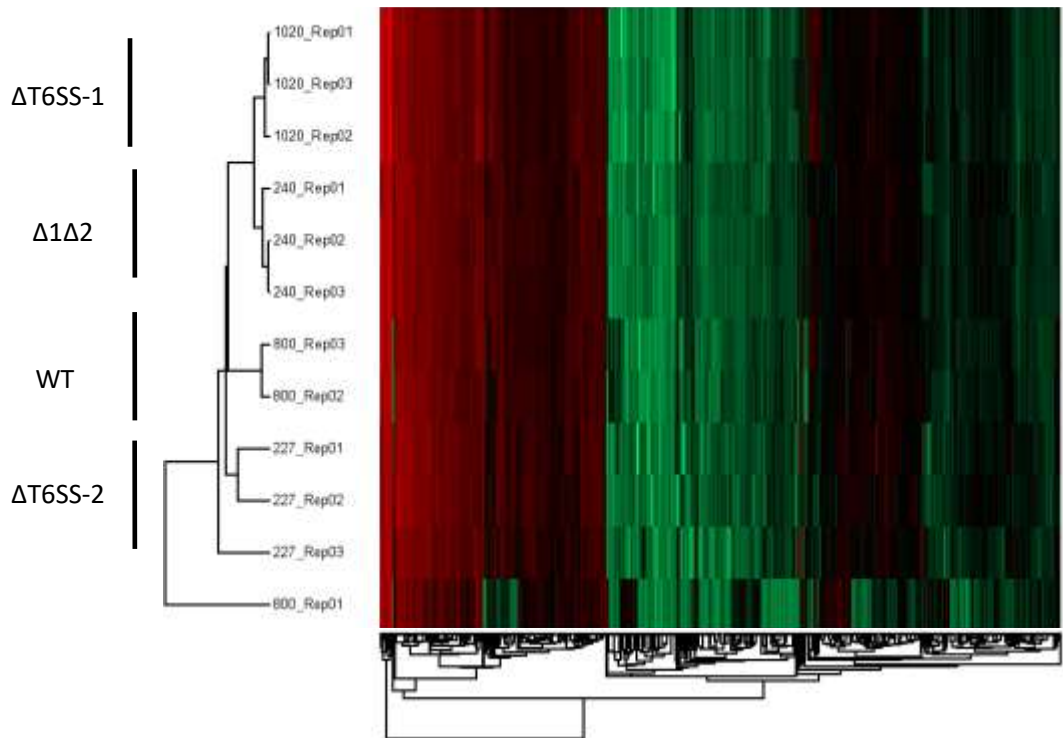


Figure A5.1: Identification of candidate T6SS-1 secreted proteins by proteomics

comparison of wild type and Δ T6SS-1 supernatants. Volcano plot statistical significance ($-\log_{10}$ P value; Student's *t*-test *P* values, *n*=3 biological replicates) of differences between wild type and Δ T6SS-1 using LC-MS/MS analysis. The $\log_2$ of the ratios of protein signals between wild type and the Δ T6SS-1 are plotted against the Student's *t*-test *P* values. Proteins significantly increased in abundance in wild type compared to the Δ T6SS-1 mutant ($|\text{Fold change difference}| > 1.0\text{-fold}$, $P < 0.05$) are depicted. Two categories of proteins are labeled, proteins related to T6SS and/or enriched above 4-fold. Proteins are labeled either with XNC1 locus identifier or annotated function.

A



B

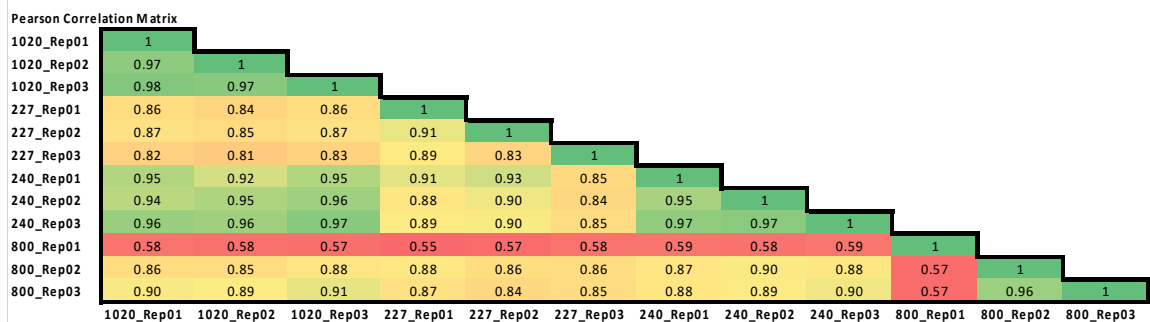


Figure A5.2: Comparison of supernatant proteomics samples. (A) The proteins present in the culture supernatant of three biological replicates were extracted from wild-type *X. nematophila* and each of the T6SS mutants and subsequently analyzed using LC/MS-MS. (B) Samples from each replicate showed good biological reproducibility (average coefficient of variation ~15%), with the exception of one wild-type sample as an outlier, average coefficient of variation ~40%.

Appendix 6: “The value of an integrated statewide environmental report for Tennessee” White Paper summary as part of the Environmental Science and Policy Practicum course

Publication status: See <https://consci.utk.edu/wp-content/uploads/2020/09/The-value-of-an-integrated-statewide-environmental-report-for-Tennessee.pdf> for the full paper. Publishing was delayed due to COVID-19 taking priority in statewide policy.

Project summary: “The environment is constantly changing, affecting the structure and function of ecosystems and impacting people’s quality of life. Different agencies within the state of Tennessee monitor and report environmental data, but there is no single, easy-to-use, centralized report summarizing the range of observations made. We describe the need for an integrated environmental report for the state. State of the environment reporting has been done at the national level, but is less common for states. However, a clear precedent for state-level reporting in Tennessee has been set by an influential annual report on the state of the state’s economy. We identify the steps necessary to synthesize a counterpart state of the Tennessee environment report. We also offer examples of the kind of environmental indicators it would be important for such a report to consider, presented in the form of an exemplar report. An annual state of the environment report will help Tennessee promote a sustainable and resilient future.”

This paper was set to be published in the White Paper series at the Howard H. Baker Jr. Center for Public Policy was part of a class project for EEB 504. I was responsible for coming up with the original paper concept with our small interdisciplinary team of 3 graduate students (and Dr. Paul Armsworth), as well as writing the specific section on the air quality and sections of the water quality data. We worked together as a team to draft and edit the final product.

Appendix 7: OUTgrads presidency summary

OUTgrads is a graduate student networking organization aimed at building a queer community on campus. Graduate school can be isolating, and queer students do not typically have many opportunities to meet each other and socialize. This is especially at the 2nd most LGBTQ+ unfriendly university in the country (see here: <https://www.wbir.com/article/news/education/ut-ranks-as-second-most-lgbtq-unfriendly-university/51-d57ae861-f695-4139-bc2a-db623b0c31c4>).

My role as President (Fall 2018-Summer 2021) had been to organize social events to connect these students. This has led to growing OUTgrads membership and creating a large interdisciplinary community that meets monthly. When I joined there were roughly 5-8 active members, and by February 2020, there were around 25-30. My favorite event I held was at Central Cinema in November 2019. Dr. Chris Holmlund at UTK gave a lecture on the classic 1974 John Waters' movie "Female Trouble", followed by an airing of the movie. It was truly an honor to bring together a community of queer graduate students.

Due to the COVID-19 pandemic, socials have halted, but my last role as President has been to establish a motivated executive board. This board will expand the goals of the organization and hopefully lead to drastic changes to improve queer graduate student experiences at UTK.

VITA

Nicholas C. Mucci was born in Boston, Massachusetts, U.S.A., in 1995. They were raised in Revere, Massachusetts. After graduating from Revere High School, they attended the University of Massachusetts Amherst where they received a B.Sc. degree in Biology with a minor in Information Technology in 2017. They then moved to Knoxville to pursue this degree.